Linear programming in planning and executing the exploration phase of mining

Tommi Kauppinen
Linear programming in planning and executing the exploration phase of mining

Tommi Kauppinen

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the auditorium f239a of the school on 15th December 2017 at 12 noon.

Aalto University
School of Electrical Engineering
Department of Electrical Engineering and Automation
Supervising professor
Associate Professor Quan Zhou, Aalto University, Finland

Thesis advisor
Docent Kai Zenger, Aalto University, Finland

Preliminary examiners
Professor Holger Paulick, Oulu University, Finland
Professor Xue Li, University of Queensland, Australia

Opponent
Professor, Executive Director Pertti Lamberg, Keliber Oy, Finland

Aalto University publication series
DOCTORAL DISSERTATIONS 228/2017

© 2017 Tommi Kauppinen

ISBN 978-952-60-7729-1 (printed)
ISBN 978-952-60-7730-7 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Unigrafia Oy
Helsinki 2017

Finland
Abstract
The doctoral dissertation consists of analysis of exploration phase of mining on different levels. The dissertation is of the field of systems technology. Firstly, methods for drillcore analysis are studied. The main emphasis is on rapid, contactless and online analysis of drillcore by Raman spectroscopy and Laser-Induced Fluorescence (LIF). The main result is the introduction of a Non-Negative Least Squares (NNLS) method for studying Raman spectra in order to partly automate the analysis of measurements. Because this in turn makes the analysis of drillcores more rapid, the research can alleviate the negative economic and ecological impact of exploratory phase of mining activities.
Secondly, the mathematical methods are extended to help geologists in search for ore. This goal is achieved by measuring homogenised pellet samples with Raman spectroscopy and analysing the results with Data Envelopment Analysis (DEA). This creates opportunities for more efficient exploration as geologist can use the results to identify ore zones.
Thirdly, the sustainable exploration phase of mining is discussed and suitable mathematical method based on DEA is introduced. The mathematical method takes into account different economic, ecological and social factors and it can be used to alleviate negative social or ecological impacts the exploration phase of mining may have.
As a conclusion, Non-negative least squares and Data Envelopment Analysis are introduced as useful tools for developing the exploration phase of mining to become more economically, ecologically and socially acceptable.

Keywords exploration, non-negative least squares, data envelopment analysis, Raman analysis, mineral concentration, rock breakage, sustainability
Tiivistelmä
Toiseksi matemaattisia menetelmiä jatkokehitetään soveltumaan geologien apuvälineiksi heidän toimiessaan malminetsintäpäikällä. Tavoitteen tehtävänä robotaan käytännöllä homogenisoitujen pelletinäytteittä mittaavaa Raman-spektroskopiaa ja Data Envelopment Analysis (DEA, suom. tehokkuusanalyysi) -menetelmä. Tämä tutkimushejaan tuottaa välineitä tehokkaampaan malminetsintänä nopeuttamalla geologien tekemää malmiesintymien tunnistamista.
Kolmanneksi tarkastellaan malminetsinnän kokonaiskäsittevyyttä, jonka pohjalta esitellään malminetsinnän tekemää tehokkuusanalyysiin perustuvaa matemaattinen mallinnusväline. Väline ottaa huomioon joitain malminetsinnän taloudellisia, sosiaalisia ja ekologisia vaikutuksia ja sitä voidaan käyttää vähentämään haitallisia sosiaalisia ja ekologisia vaikutuksia, joita malminetsintään usein liittyy. Yhteenvetona voidaan esittää pienimmän nelisumman merkittäviä positiivisille luvuille ja tehokkuusanalyysi käyttökoelpoisina matemaattisina välineinä malminetsintävaiheen kehittämiseksi taloudellisesti, sosiaalisesti ja ekologisesti kestävämmäksi.

Avainsanat: malminetsintä, pienin nelisumma, tehokkuusanalyysi, Raman-analyysi, mineraalikonsentraatio, kiven kovuus, kestävyys

ISSN-L: 1799-4934 ISSN (painettu): 1799-4934 ISSN (pdf): 1799-4942
Julkaisupaikka: Helsinki Painopaikka: Helsinki Vuosi: 2017
I would like to thank my instructor Docent Kai Zenger for his impeccable interest in my doctoral studies. Also, I thank Dr. Olli Haavisto for his initial instructorship during the years 2012-2013. I also thank Ass. Prof. Quan Zhou for the supervision of my thesis.

I would also like to thank M. Sc. Navid Khajehzadeh for his help with many of the studies we conducted together. Of the departments other doctoral candidates, I would also like to thank D. Sc. Phillip Nguyen, M. Sc. Sergey Samokhin, M. Sc. Christian Giovanelli and D. Sc. Ville Liimatainen for the many off-the-topic conversations we had together.

TEKES Green Mining –project helped me start my doctoral studies during the years 2012-2014 and Alfred Kordelin fund funded my research for the duration of the year 2015. After this funding, Renlund foundation funded my research during the year 2016, as did The Finnish Foundation for Automation, funding the finishing of this dissertation. For these instances, I am deeply grateful.

The industrial collaboration with First Quantum Minerals Kevitsa and Agnico-Eagle Kittilä mines was important for the data acquisition of this research and without their contribution this thesis would not have been possible. I would also like to thank Dr. Heikki Häkkänen for the help in laser-based measurements we made at the University of Jyväskylä. Heikki’s help was uncontested in doing the measurements and without him the measurements would have most certainly taken a longer time with an increased noise-to-signal ratio.

I also had the pleasure to visit Brisbane, Australia at Julius Kruttschnitt Minerals Research Centre (JKMRC) for 4 months as a visiting researcher during 2013-2014. I kindly thank Prof. Emmanuel Manlapig and Dr. John Jackson as well as Dr. Janine Lay for arranging my stay and welcoming me to contribute on the research done in their institute.

Finally, my deepest gratitude goes to my wife Anna and our child Aava. You have together been a beacon without which I could not have found my way. Of all the mornings, days, evenings and nights we have spent together on this somewhat weird adventure I thank you both.

Helsinki, 5th of October 2017
Tommi Kaapro Kauppinen
Contents

Acknowledgements .......................................................................................................... 1
List of Abbreviations and Symbols .................................................................................. 4
List of Publications .......................................................................................................... 6
Author’s Contribution ....................................................................................................... 7
1. Introduction .................................................................................................................... 8
   1.1 The main hypotheses made and problems assessed ................................................. 9
   1.2 Outline of the work .................................................................................................. 9
   1.3 Publications on linear programming in the exploration phase ......... 11
      1.3.1 Publications on Raman spectroscopy and analysis with ore ...12
      1.3.2 Publications on Laser-induced fluorescence in case of ore ..13
      1.3.3 Publications on Sustainability in mining ........................................ 14
      1.3.4 Publications on Data Envelopment Analysis used in mining .14
2. Solving for abundance and concentration ................................................................ 15
   2.1 Laser-based rock drillcore analysis ....................................................................... 15
      2.1.1 LIF and Raman ......................................................................................... 16
   2.2 Laser-based rock drillcore analysis ....................................................................... 17
      2.2.1 Mathematics of Non-negative least squares .............................................. 17
      2.2.2 Concentration and variance ....................................................................... 17
   2.3 Experimental work and data used in Raman measurements .................. 18
      2.3.1 Measurement apparatus ............................................................................. 19
      2.3.2 Preparing the samples for line-scanning ...................................................... 19
      2.3.3 Preparing the samples for grid-scanning ...................................................... 20
      2.3.4 Point-count method .................................................................................. 20
3. Efficiency analysis .................................................................................................... 21
   3.1 Data Envelopment Analysis (DEA) ................................................................. 21
      3.1.1 Mathematical presentation of DEA ............................................................. 22
   3.2 Rock breakage with DEA .................................................................................... 23
      3.2.1 Methods ................................................................................................. 23
      3.2.2 Quotient model ...................................................................................... 24
## List of Abbreviations and Symbols

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>data matrix</td>
</tr>
<tr>
<td>AA</td>
<td>Atomic Absorption spectroscopy</td>
</tr>
<tr>
<td>b</td>
<td>vector of inverse values of wavelengths in the sample spectrum</td>
</tr>
<tr>
<td>BCC</td>
<td>Banker-Charnes-Cooper DEA model</td>
</tr>
<tr>
<td>c</td>
<td>abundance and/or concentration (as a percentage of the total concentration)</td>
</tr>
<tr>
<td>CCR</td>
<td>Charnes-Cooper-Rhodes DEA model</td>
</tr>
<tr>
<td>D</td>
<td>diagonal matrix consisting of the eigenvalues of $AA^T$</td>
</tr>
<tr>
<td>DEA</td>
<td>Data Envelopment Analysis</td>
</tr>
<tr>
<td>DMU</td>
<td>Decision Making Unit (in Data Envelopment Analysis)</td>
</tr>
<tr>
<td>E</td>
<td>matrix of eigenvectors</td>
</tr>
<tr>
<td>$E_1$</td>
<td>vector of efficiency scores associated with DEA model number 1</td>
</tr>
<tr>
<td>$E_2$</td>
<td>vector of efficiency scores associated with DEA model number 2</td>
</tr>
<tr>
<td>$E_q$</td>
<td>vector defining a quotient model for DEA</td>
</tr>
<tr>
<td>$f_1$</td>
<td>Sum of weighted input variables for a given DMU</td>
</tr>
<tr>
<td>$f_2$</td>
<td>Sum of weighted input variables for a given DMU</td>
</tr>
<tr>
<td>ICP</td>
<td>Inductively Coupled Plasma</td>
</tr>
<tr>
<td>LHD</td>
<td>Load Haul Dumper</td>
</tr>
<tr>
<td>LIBS</td>
<td>Laser-Induced Breakdown Spectroscopy</td>
</tr>
<tr>
<td>LIF</td>
<td>Laser-Induced Fluorescence</td>
</tr>
<tr>
<td>LOI</td>
<td>Loss of Ignition</td>
</tr>
<tr>
<td>LP</td>
<td>Linear Programming</td>
</tr>
<tr>
<td>MLA</td>
<td>Mineral Liberation Analysis</td>
</tr>
</tbody>
</table>
NNLS  Non-Negative Least Squares
PCA  Principal Component Analysis
\( R_{\text{ref}} \)  matrix of different wavelengths in all the reference spectra
\( S \)  diagonal matrix consisting of the eigenvalues of \( AA^T \)
\( U \)  the eigenvectors of \( AA^T \)
\( u_o \)  variable denoting variable returns to scale assumption of BCC DEA modelling
\( u_s \)  weight associated with \( s \)th output variable in a DEA model
UV  Ultra Violet
\( V \)  the eigenvectors of \( A^T A \)
\( v_m \)  weight associated with \( m \)th input variable in a DEA model
\( x_{m0} \)  input variable \( m \) for DMU \( o \) in a DEA model
\( x_{\text{sample}} \)  vector of different wavelengths in the sample spectrum
XRD  X-Ray Diffraction
XRF  X-ray Fluorescence
\( y_{so} \)  output variable \( s \) for DMU \( o \) in a DEA model
List of Publications

This doctoral dissertation consists of a summary and of the following publications which are referred to in the text by their numerals


Author’s Contribution

**Publication 1:** Data Envelopment Analysis as a tool for the exploration phase of mining

The journal article was completely planned and written by the author.

**Publication 2:** Sustainability in the exploration phase of mining: a data envelopment analysis approach

The conference paper was completely planned and written by the first author, second author gave valuable comments.

**Publication 3:** Data fusion on Rock breakage using Data Envelopment Analysis

The conference paper was planned and written by the first author, but the abstract and introduction was reviewed and commented by the other authors.

**Publication 4:** Raman analysis: detection of minerals and measurement of concentration using reference spectra

The conference paper was planned and written by the first author. The second and third author gave comments on the manuscript while the fourth author assisted in planning the measurement setup.

**Publication 5:** Laser-induced fluorescence images and raman spectroscopy studies on rapid scanning of rock drillcore samples

The journal article was a joint effort with the second author, who wrote a substantial part of the manuscript, while the first author was mainly responsible for preparing the calibration technique for analysing the results.

**Publication 6:** Optimisation algorithms in the case of mineral detection using Raman analysis

The conference paper was planned and written by the first author. However, it was influenced by the second author who assisted in writing the manuscript. The third author gave a contribution by designing the measurement setup.
1. Introduction

There is a growing demand for metals globally. Rising global demand for consumer goods needs more efficient extraction industry. Therefore, the author presents this doctoral study, which was done in order to point out the possibilities for reducing inefficiencies in the exploration phase of mining and finding more sustainable ways of exploring metals.

The study of geology is analysis of rock samples. This doctoral dissertation is heavily based on the previously built geological knowledge of rocks and the minerals therein. The thesis is concerned with laser-based methods for analyzing rocks and especially the linear programming methods employed. This study is also limited to the exploration phase of mining, which relates to exploration both before mine starts to operate as well as during the mining operations.

The rock samples received during the exploration phase of mining are acquired through diamond drilling and are called rock drillcores. The exploration phase is a costly and time-consuming process. This thesis proposes new tools to optimise the exploration phase of mining by the assistance of laser-based methods and mathematics.

The main laser-based method, which is studied in detail, is Raman analysis. The individual minerals in a rock drillcore can be identified using Raman analysis and ready-made online libraries. Furthermore, another method called Laser-Induced Fluorescence (LIF) for mineral analysis of rock drillcores is discussed in this thesis.

The linear programming methods which are under scrutiny are Non-Negative Least Squares (NNLS) and Data Envelopment Analysis (DEA). Furthermore, Principal Component Analysis (PCA) is used to verify some of the results acquired by linear programming. Even though we note that these methods are not novel, the way they are employed is, and therefore this thesis gives the mining industry new tools for optimising the exploration phase of mining.

To make it short, this doctoral dissertation seeks to find data analytical means to produce useful results on mineralogical findings. The results received in this doctoral dissertation are as follows:

- NNLS is a possible tool to solve for abundance alongside point-count method
- NNLS can be used to solve for concentration and is more accurate than the point-count method
- LIF and Raman can be combined for industrial scale online scanning for rock drillcores
Rock breakage can be studied with DEA
- Rock drillcores can be successfully studied with DEA in the exploration phase of mining
- Sustainability of the exploration phase of mining can be effectively studied with DEA

1.1 The main hypotheses made and problems assessed in the doctoral study

This doctoral dissertation is based on the following hypotheses:
- In Raman analysis, a combination of known library spectra can be used to model the sample spectrum
- Laser-induced fluorescence can be used in a mine to identify minerals after it has been calibrated with Raman analysis
- Data Envelopment Analysis can be used to model non-linear changes in rock breakage by using a so-called quotient model
- Data Envelopment Analysis can assist the work of geologists by showing abundance of interesting minerals in the rock drillcore to be analysed
- Data Envelopment Analysis can be used to produce a simple measure for sustainability when all the relevant indicators are included into the analysis

The main research questions for this doctoral study are as follows:
- How is the exact abundance and concentration of minerals solved using Non-Negative Least Squares for rock drillcores?
- How and how efficiently can the Laser-Induced fluorescence and Raman be combined for discovering the minerals in rock drillcores for a certain mine?
- How can Data Envelopment Analysis be used to formulate a model for rock breakage?
- Is it possible to alleviate the work of geologists in the exploration phase of mining using Data Envelopment Analysis?
- Can Data Envelopment Analysis be used to model sustainability of the exploration phase of mining? If so, what type of model should one use?

1.2 Outline of the work

The doctoral study was begun by analysis of Raman measurements done on rock drillcores of Agnico-Eagle Kittilä gold mine (Kauppinen et al. 2013). As the initial discussion only concerned itself with a suitable mathematical method for detecting minerals with Raman spectroscopy, the main result found was the Non-negative least squares (NNLS) method for studying the spectral data yielded by Raman analysis, to solve for abundance of minerals in an ore sample. The results were preliminary, however, but showing the potential of mathematical methods in the study of rock drillcores. The NNLS method is mainly a data-fitting tool for finding out the exact mixture of spectra in each measurement step. When referring to an “ore”, a specific type of rock with a comparably high
economic potential is referred to. It should also be noted that alongside whole drillcores, also powdered pellet samples, made of drillcores, were used in the measurements where applicable. Two different sources for rock drillcore and powdered pellet samples were used in the study, namely Kittilä mine, Suurikuusikko deposit (Agnico-Eagle Finland) and Kevitsa mine (First Quantum Minerals), Finland.

The study continued by analysing properly the possibility of using the preliminary results to solve for abundance and also finding the concentration of different minerals, using spectral analysis of Raman results. Abundance measurement (or qualitative description) is different to concentration measurement (or quantitative result) in that abundance yields the existence of minerals in an ore sample, while concentration measurement gives the amount of minerals in the sample. Kauppinen et al. (2014b) concluded that there are high uncertainties in measuring the concentration directly from one spectrum. Therefore, an average of thousand to ten thousand spectra should be used, as this brings down the variance of the results. With thousand measurements per sample, the variance of the results can be controlled adequately. In this thesis, the concentration measurement from spectral analysis has been done, to the best knowledge of the author, for the first time. Discussion has now only dealt with the precision (or random error) of the results, but the accuracy (or systematic error, also called trueness) should also be considered. It seems that with very high or low concentration values there is an systematic error (Kauppinen et al. 2014b). However, with moderate concentration values, also the accuracy is of useful standard.

The next step in the doctoral thesis was to bring together the results found in the research group by combining Raman analysis with Laser-Induced Fluorescence (LIF) image processing (Kauppinen et al. 2014c). The result of using digital camera to extract LIF images from rock drillcore enabled a state-of-the-art, industrial scale application of LIF and Raman analysis. While doing this work, the fusion of different types of data was identified as the most important research question for the future.

At this stage Data Envelopment Analysis (DEA) was discovered as a data fusion method for analysis. DEA can be seen as a versatile tool for data fusion and a thorough study was done on different ways DEA could be used to fusion information related to exploration phase of mining, as described below. The first step was to study rock breakage with DEA (Kauppinen et al. 2014a), which led to a finding of a new tool for the study of rock hardness. After promising results with the Australian partners, the author continued with the colleagues in Finland by studying the sustainability of exploration phase of mining with DEA (Kauppinen & Khajehzadeh 2015), producing a simple but effective measurement tool for sustainability. The final part of this dissertation, bringing together the preliminary studies using DEA, seeks to show that DEA is an excellent tool for the exploration phase of mining, helping geologists to log drillcores more carefully (Kauppinen 2016, see also Figure 1). The way DEA was employed in these applications was in all three cases a novelty and reveals the value of applying DEA in the exploration phase of mining.
Figure 1 introduces the main result from Kauppinen (2016). In the Figure 1, y-axis shows the proportional amount of valuable minerals found in the different depths. The exact depth is shown by x-axis. Valuable minerals are here pyrite and arsenopyrite minerals, as the samples used are from Kittilä gold mine, where the gold is received from these minerals. The initial measurement was done using Raman analysis, while the final result (Figure 1) was found using Data Envelopment Analysis on the measurements. The drillcores used are from routine drilling setup at Agnico-Eagle Kittilä gold mine, and the drillhole intersects an ore zone. The results found coincide well with standard geochemical analyses. These matters are discussed further in coming sections.

This doctoral dissertation has the following structure: first, a literature review is presented in Section 1.3. Second, Section 2 discusses the theory for computing abundance and concentration for rock drillcores. Section 2 also shows the method of combining Laser-Induced Fluorescence and Raman analysis. Third, Data Envelopment Analysis and its applications in this dissertation are given in Section 3. Fourth, Sections 4 and 5 discuss data preparation for linear programming and mathematics concerning linear programming, respectively. Fifth, Section 6 gives the results of the dissertation, and the conclusions drawn in the dissertation are presented in Section 7.

1.3 Publications on linear programming in the exploration phase of mining

The history of linear programming in the exploration phase of mining dates back at least to the seventies (e.g. Rogado 1975). The topic of this dissertation is therefore not new, although the research has not been very active until recently.
This is due to the developments in the computer-based mathematical methods of finding ore (Computers & Geosciences 2011).

At present, the publications on the use of linear programming in the exploration phase of mining is mostly concerned with mineral prospectivity mapping (Carranza & Hale 2002, Lewkowski et al. 2010). In this context, Data Envelopment Analysis, which forms a considerable part of this dissertation, has been used as well (Hosseini & Abedi 2015). However, the discussion on semantic interoperability of geodata (or the different types of geological data received by measurements, see Ma et al. 2010) is a topic that falls outside the scope of this dissertation.

It can be noted that linear programming has not been used concisely for the exploration phase of mining outside mineral prospectivity mapping. In this dissertation, linear programming is used to develop both laser-based techniques (Laser-induced fluorescence and Raman analysis) as well as the exploration phase of mining in logging the drillcores (Kauppinen 2016). Linear programming is also used to study sustainability (Kaupinen and Khajehzadeh 2015), as well as measuring rock breakage (Kaupinen et al. 2014a). In all of these areas, linear programming can be called a novel method for exploration phase of mining, seldom seen in the scientific literature (e.g. Zhang et al. 2012).

1.3.1 Publications on Raman spectroscopy and analysis with ore

Raman spectroscopy and analysis of minerals is a heavily researched topic. The main areas of interest are classification of minerals with Raman, stand-alone applications and identification of specific materials of interest, like diamonds (e.g. Lamprecht et al. 2007). The most important paper in the field considering this study is the one by Haskin et al. (1997), developing the point-count method for determining the abundance of minerals in an ore sample.

The classification of minerals can be either qualitative or quantitative. As the process of studying Raman spectra has developed, there is an increasing interest to produce quantitative applications. This type of analysis is described in the doctoral study at hand as well. To mention some key studies, the qualitative methods for ore analysis have been reviewed by Das & Agrawal (2012) and developed by Das & Hendry (2011). The quantitative methods have been researched by Biasio et al. (2012) and Dörfer et al. (2009). The latter describes a similar method as in this doctoral study but deems it unsuitable for further analysis, considering more involved methods in detail. Therefore, it is of interest to show how linear programming can be used to directly measure abundance and even concentration of minerals in ore samples (Kaupinen et al. 2014b, Kauppinen et al. 2013). This has been attempted before with the help of Principal Component Analysis (Van de Sompel et al. 2012), but the computation is much more demanding than with the methods introduced in this study.

Going through the scientific literature for Raman spectroscopy, during the last few years a clear need for stand-alone applications has developed. There are many studies conducted on Raman to be used on automated vehicles sent for Mars expeditions, e.g. Thiele et al. (2007). Less prestigious but still important use for stand-alone Raman measurement is in mines and exploration sites,
where automated mineral analysers could be used to increase the effectiveness of the exploration. The design of this type of automated analysers has been attempted (Ishikawa & Gulick 2013), but it is still under discussion how the actual machinery should be implemented. This is where the doctoral study at hand can improve the current state-of-the-art, by showing how the design of semi-automated Raman analyser could be done (Kauppinen 2016).

Raman can be used for more specific purposes. The diamonds were mentioned (Lamprecht et al. 2007), but Raman spectroscopy can be used for transition metals as well (Ren et al. 2007). Goncharov et al. (2001) studies extensively the high-pressure application for finding metals. However, Raman spectroscopy has relatively long measurement time and is susceptible to auto-fluorescence emission of the minerals (Khajehzadeh et al. 2017). Moreover, in order for a sample to be Raman active, it must have vibrations (or rotations) that result in a change in the polarizability of the molecule or material. Therefore, Raman scattering in metallic substances is very weak, and cannot be detected without special conditions (Goncharov and Struzhkin, 2003). This type of development work shows that even though there are weaknesses in Raman analysis, it is under constant scientific scrutiny and can therefore be considered a viable tool for mineral analysis, as long as the deficits such as those mentioned above are taken into account. Overall, Raman spectroscopy is useful for different mining applications, even if considering the fact that minerals which have a very high metal content are not Raman responsive.

1.3.2 Publications on Laser-induced fluorescence in case of ore

The fluorescence or luminescence spectrum of minerals have been studied extensively in laboratory conditions (see e.g. Waychunas (1988), Reisfeld et al. (1996), Gaft et al. (1998), Gaft et al. (2008) and references therein). Bozlee et al. (2005) analyzed the LIF spectra of natural minerals and rocks, including calcite, gypsum, fluorapatite, and other minerals using remote (long distance) measurement configuration.

Since the interpretation of the LIF spectra is difficult, a typical approach for industrial applications is to create a reference file with spectral fingerprints of the relevant rock or ore types in the given mining environment. This reference library can then be used to classify the spectra measured from new samples. Thus the LIF analysis is best suited for cases where the number of different rock or ore types is limited and preferably well known in advance (Broicher 1999).

Consequently, industrial applications of LIF in mining and minerals processing have concentrated on dilution control, assessment of ore types and quality control. Applications for classification of phosphorous iron ore and waste rock from belt conveyors and load haul dumper (LHD) buckets have been presented by Broicher (1999). Nienhaus and Bayer (2003) described an online LIF analyzer and its application for classification of hard coal, potash ores and lignites. Phosphor monitoring at an iron ore mine and dilution control at a copper operation based on LIF were discussed by Nienhaus et al. (2003). However, quantitative results describing the performance of the systems were not given.
To the best knowledge of the author, LIF studies specifically on drillcores have not been published. This is most likely due to the more qualitative nature of LIF analysis. However, the combination of LIF with other analysis methods (e.g. Raman, as is done in this doctoral study, see Kauppinen et al. 2014c) should improve the overall analysis results.

1.3.3 Publications on Sustainability in mining

The sustainability in mining context has been of growing interest in research. However, in the doctoral study the issue is discussed by introducing a new way to measure sustainability. Hassan et al. (2014) gives a wide review of sustainability research, where extractive industries are included as one sub-branch. Hassan et al. (2014) show how important the sustainability framework has become for different industries, showing the significant growth in interest towards sustainability.

The main topic of interest here is the measurement of sustainability. Fonseca et al. (2012, 2013) discuss in critical manner the state-of-the-art corporate responsibility measurement systems. It is therefore of interest to try and create a more transparent system of measurement. In this doctoral study, Data Envelopment Analysis is proposed as such a measurement system (Kauppinen & Khajehzadeh 2015). Other measures have also been proposed in the literature, but they all suffer from the lack of ease in using them as Fonseca et al. (2013) show in their research (see also Fuisz-Kehrbach 2015). There have also been attempts to move towards qualitative analysis from the traditional triple-bottom line (economic, ecological and social bottom line, Bond 2014). However, these research topics are beyond the scope of this doctoral study.

1.3.4 Publications on Data Envelopment Analysis used in mining applications

Data Envelopment Analysis (DEA) builds on the traditional linear programming theory to measure efficiency (Farrell 1957). It is therefore of interest to discuss whether mining industry or research could benefit from using DEA as a tool for modelling the exploration phase of mining. The research of this topic forms the main part of the doctoral study at hand. DEA is used to study rock breakage, which is a novel method in the area (Kauppinen et al. 2014a); It is also used to study geology of a rock, more precisely pyrite and arsenopyrite minerals for producing gold therein (Kauppinen 2016); and finally, to measure sustainability (Kauppinen & Khajehzadeh 2015).

It is therefore of interest to note that DEA has not been used in mining activities, except for the few studies mentioned here. Hosseini & Abedi (2015) study prospectivity mapping using DEA and Tsolias (2011) discusses using a specific and more complex DEA method in performance assessment of mining operations. This doctoral study shows that the complexity is unnecessary and even the basic DEA modelling can be used to produce plausible results for e.g. sustainability measurement (Kauppinen & Khajehzadeh 2015).
2. Solving for abundance and concentration

In Raman analysis, when the number of spectra which resemble a certain mineral in spectral analysis is known, the abundance can be easily established by using the point-count method (Haskin et al. 1997). The non-negative least squares (NNLS) method used in the dissertation is a novel and state-of-the-art means of arriving to abundance and concentration of a rock drillcore sample. Furthermore, the measuring of the concentration from spectral data is shown to find the concentration of minerals suitable for Raman analysis with improved precision compared to the point-count method. In addition, NNLS and Raman analysis can be used in combination with Laser-Induced Fluorescence (LIF), enabling an analysis for the contents of a rock drillcore.

2.1 Laser-based rock drillcore analysis

As a well-established method, Raman analysis can be considered for analysing mineral composition of ores (for detailed description of the measurement set-ups, see Section 2.3). For this purpose, Raman scattering is introduced. Materials have their individual scattering patterns when excited by light. Measuring Raman scattering is called Raman spectroscopy (Das & Agrawal 2012). Raman spectroscopy is based on the shifts in wavelength that are caused by a laser pointed on the material. The radiation from the material can be used to deduct its chemical contents. In many parts of the world, Raman analysis is being used for ore analysis, e.g. by De Biasio et al. (2012) and by Pinzaru & Onac (2009).

In this study, the Raman analysis is used as a rapid and contactless method for on-line mineral analysis. The motivation is to show that Raman analysis can be used in designing more automated mineral analysis which enables fast detection of minerals from a drillcore sample. The study shows that designing this type of automation will enable more sustainable solutions to mineral exploration. However, the limitations of Raman analysis should also be considered as well. The laser used for Raman scattering has a very small diameter (0.1 μm) and consequently the drillcore cannot be scanned as a whole in an effective time frame.

There are three possible processes for Raman scattering: Rayleigh scattering, Anti-Stokes scattering and Stokes scattering (Smith & Dent 2005). The Rayleigh
process is the most intense of the possible modes of scattering, but it does not involve energy change. This is why in this study the Raman scattering recorded is Stokes scattering, even though anti-Stokes scattering could be preferred as well. In the future, the anti-Stokes scattering may well be adopted to deal with fluorescence interference, but this option falls outside the scope of the thesis.

2.1.1 LIF and Raman

The possibility of using other methods besides Raman was also addressed, and as a result a calibration technique for Laser-Induced Fluorescence (LIF) with Raman analysis was introduced (Kauppinen et al. 2014c). However, before the selection of these two methods of measurement, the study also tested Laser-Induced Breakdown Spectroscopy (LIBS) and X-ray Fluorescence (XRF). However, with these methods the analysis is based on elemental and not mineralogical findings while Raman and LIF are both mineralogical methods for analysis. Also, considering rock drillcores, the speed at which the analysis can be done is also important. Other mineralogical methods would include e.g. X-Ray Diffraction (XRD), but these were not considered because of the lack of suitable measurement setup.

In Laser-induced Fluorescence (LIF) the activated areas of rock drillcore are called impurities. These are mainly transition metals or rare earth element ions (Hawthorne & Waychunas 1988). Different colours in LIF imaging signify different minerals and LIF is a well-studied method with which to identify certain types of minerals. Until this dissertation (Kauppinen et al. 2014c) the LIF analysis techniques were still point-wise, slow and eventually inappropriate for rapid on-line scanning. However, by using off-the-shelf digital cameras the research was able to show how LIF imaging could be done faster but still reliably. The merger of Raman spectroscopy and LIF was a novel idea introduced.

In Laser-induced Fluorescence (LIF) the activated areas of rock drillcore are called impurities. These are mainly transition metals or rare earth element ions (Hawthorne & Waychunas 1988). Different colours in LIF imaging signify different minerals and LIF is a well-studied method with which to identify certain types of minerals. Until this dissertation (Kauppinen et al. 2014c) the LIF analysis techniques were still point-wise, slow and eventually inappropriate for rapid on-line scanning. However, by using off-the-shelf digital cameras the research was able to show how LIF imaging could be done faster but still reliably. The merger of Raman spectroscopy and LIF was a novel idea introduced.

The combined LIF and Raman measurement setup was needed because the speed at which Raman can measure whole drillcores is very limited. However, with LIF the speed at which results are obtained can be significantly increased. The LIF imaging process included analysis of images taken with ordinary digital camera. The images were analysed by the help of Raman spectra. First, the surface areas of drillcores with visible transitions between minerals were photographed with a digital camera and line-scanned with Raman. Second, the Raman line-scans were analysed to find which minerals were present with the help of Non-negative least squares (NNLS). Third, the results of the Raman analysis were compared to the LIF images taken, and the different colours present in these images were established as the minerals given by Raman analysis. The combination of LIF imaging technique and Raman analysis was able to reliably establish the qualitative abundance of different minerals at the surface of a drillcore.
2.2 Non-negative least squares

For fitting the curve on sample spectrum (see e.g. Figure 8) there are plenty of possible methods. However, the simplest one should be selected and hence Non-negative least squares (NNLS) was selected. NNLS has been discussed in scientific journals in relation to Raman analysis, but before this thesis it has been found flawed as a method for finding the best fit (Dörfer et al. 2009). Here the NNLS method is described in detail as well as its usage in finding out the quantitative concentration of a given mineral in the sample.

2.2.1 Mathematics of Non-negative least squares

At first, nonlinear optimisation was used because some of the sample spectra displayed a property of having multiple reference spectra merged together. Later this same phenomenon was used to gather information about the concentration of different minerals. Before attending to NNLS, many algorithms were tested. From these, the nonlinear optimisation and optimisation with PCA has been discussed in Kauppinen et al. (2013).

The interest is on the positive spectra because negative spectra make no physical sense. The optimisation algorithm should therefore include a non-negativity constraint for the given reference spectra fitted into sample spectrum. In numerical linear algebra, non-negativity constraints often arise in the least squares problem, which is called Non-Negative Least Squares (NNLS) (Chen and Plemmons 2009). The current NNLS algorithms have been established by Lawson and Hanson (1995). A variation of Lawson and Hanson’s algorithm is available as lsqnonneg in the Matlab software.

The optimisation problem is defined as follows. There exists a vector $c \in R^n$, which defines the values at which the different reference spectra (1, ..., n) are included into the model for the measured sample spectrum. The sample spectrum studied is given by a vector $x_{\text{sample}}$ of measured Raman counts as a function of wavenumber. The intensities of the reference spectra as a function of wavenumber are included into the columns of matrix $R_{\text{ref}}$. The NNLS problem is therefore given as

$$\min_{c} \|x_{\text{sample}} - R_{\text{ref}}c\|^2, \ c \geq 0$$  \hspace{1cm} (1)

where $\|\cdot\|$ denotes the $L_2$ norm. The main interest here is to solve $c$ which defines the coefficients at which different reference spectra are summed up (to form the sample spectrum). This NNLS method was found adequate for dealing with the curve fitting problem.

2.2.2 Concentration and variance

The concentration measurement means finding out the concentration of different minerals using only the sample spectra. The problems with this approach have been discussed widely, beginning from Haskin et al. (1997). However, by using the NNLS method and high number of measurement points per sample, one can decrease the variance in the result to a point where defining the exact
concentration becomes possible. This way, the uncertainty in the concentration measurement is only dependent on the variance. One may note that the measurement process becomes comparably slow: in order to control the variance, measurements must be repeated from thousand to ten thousand times.

In Kauppinen et al. (2014b), the NNLS concentration method was introduced and it was discovered to yield results closer to laboratory measurements than point-count method. This is because the Raman analysis done by NNLS concentration can take into account the occurrence of Raman laser hitting multiple grains at the same time, which by observing the measurement results can be established to happen often. The NNLS concentration method is able to procure a combination of minerals for a measured spectrum, not labelling each spectrum as only one mineral, as the point-count method does. The Raman method is slow, taking approximately fifteen minutes for one thousand measurements, but with pulverised and homogenised samples it can work fast enough so that thousand or even ten thousand measurements per sample can be done at the exploration sites.

2.3 Experimental work and data used in Raman measurements

Raman analysis was conducted on rock drillcore samples gathered from two mines: Agnico-Eagle Kittilä Mine, Finland; and Kevitsa Mine of FQM Kevitsa Mining Oy, Finland. Kittilä mine (Suurikuusikko deposit) is concentrated on gold mining, while Kevitsa mine (Kevitsa deposit) has been found for mining of nickel, copper and other valuable minerals. In Kittilä mine there is abundance of sulphide and gangue, as three fourths of gold is present in arsenopyrite mineral and the rest in pyrite mineral, free gold is very rare. Kevitsa deposit holds nickel, copper, cobalt, platinum, palladium and gold. Kevitsa intrusion is comprised of ultramafic olivine pyroxenites and peridotites in the northeast, gabbro in the west and central sections and predominantly granophyre in the south.

The experimental work was mainly done in the University of Jyväskylä, where Dr. Heikki Häkkänen runs a state-of-the-art Raman measurement facility. The Raman and Laser-Induced Fluorescence (LIF) measurements were usually done in matter of two to three days, both for drillcore samples and for pulverized pellet samples. The reason for the duration of measurements was mainly not the speed of the measurement but the different measurement techniques tested in the process of finding the most useful ones.

The main measurement techniques for Raman were

1. Line-scanning for pieces of drillcore with a structure of interest in accordance with LIF measurements, to produce comparable results with LIF images, and
2. Grid-scanning (usually 32 times 32 points) for pulverized pellet samples, in order to reduce the variance found in single measurement of a homogenized and pulverized ore in the pellet sample.
The samples used were received from two different mines operating in Finland. First mine was Kittilä mine, Suurikuusikko deposit (Agnico-Eagle Finland) and the second mine was Kevitsa mine (First Quantum Minerals), Finland. The drillcores were mainly from Kevitsa mine, while the pellet samples were mainly from the Kittilä mine.

2.3.1 Measurement apparatus

A laser operating at 532 nm and with power of 100 mW was used for Raman measurements (see Figure 2). The measurements were performed by moving the sample in front of the measurement area of the spectrograph. The Raman measurements were performed with a homebuilt Raman spectrometer in a backscattering geometry. Samples were excited with laser at wavelength of 532 nm (Changchun New Industries Optoelectronics Tech.Co., Ltd, MLL-III-532 nm-200 mW). The excitation light was attenuated to 100 mW and focused into a sample with a microscope objective (Zeiss, 10x, 0.30 N.A.) and the backscattered light was collected with the same objective. Rayleigh scattering was attenuated with a 532 nm Razor Edge ultrasteep long-pass edge filter (Semrock) allowing recording of Raman spectrum down to 100 (1/cm). Raman scattering was dispersed with 1200 g/mm grating in a 0.5 m imaging spectrograph (Acton, SpectraPro 2500i). The signal was detected with a CCD camera (Andor, Newton DU940N-BV).

![Figure 2. Measurement apparatus for Raman measurements (Kauppinen 2016).](image)

2.3.2 Preparing the samples for line-scanning

The preparation of samples for line-scanning was done by first halving the drillcores, then wiping away the dust and drying the moisture from the surface before measurements. The samples were scanned from the round side, from a section of a drillcore that has an interesting structure with a line scanning of length 1-3 cm. The measurements were usually taken 0.001 cm (or 1 μm) apart. The speed of measurement was approximately one second per point scanned, with a diameter of 0.0001 cm (or 0.1 μm) for the Raman laser.
2.3.3 Preparing the samples for grid-scanning

First the samples of one meter of rock drillcore (halved) were each pulverised and homogenised. Although for the purpose of this study the samples were received ready for spectroscopy, they can also be pulverised and homogenised on site. The pulverised samples were pressed together for the Raman measurement (see Figure 3). Each measured sample represents a one meter of rock drillcore, in a form of homogenised and pulverised sample.

The Raman spectra were collected by 32×32 grid on the surface of the pulverised sample. The total area covered was 0.64cm×0.64cm in size. The speed of measurement was approximately one second per grid point, with a diameter of 0.0001 cm (or 0.1 μm) for the Raman laser.

2.3.4 Point-count method

Haskin et al. (1997) introduced a point-count method for Raman analysis. It is a method computing the number of spectra that represent a certain mineral. If there are a hundred spectra, where ten of them show a clear peak on a wavelength of a certain mineral, it can be concluded that ten percent of the sample spectra exhibit that given mineral.

Point-count method can be compared to the Non-Negative Least Square (NNLS) method proposed in this thesis. The main difference is that instead of absolute concentration of one mineral for each spectrum (as point-count method does), NNLS produces a group of minerals with a percentage concentration for each spectrum, which makes it more precise in describing the contents of a sample.
3. Efficiency analysis

Efficiency analysis (or productivity analysis) is usually discussed in connection to econometrics, where the concept is widely used (Cooper et al. 2006). Here we use efficiency analysis for data fusion. This line of research stems from the fact that efficiency analysis provides tools for data fusion on many different levels. First we note that it can be used in data analysis concerning rock breakage, then with the same method we can study the rock drillcores, and finally, still with the same method the sustainability of exploration phase can be discussed. The usefulness and novelty of DEA in the exploration phase of mining gives grounds to claim DEA as a promising method, showing high potential for the application areas described.

3.1 Data Envelopment Analysis (DEA)

DEA is a non-parametric method for data fusion. The main advantage of DEA compared to other data fusion methods is that it can fuse different types of data into a single dimension with ease (Cooper et al. 2006). Non-parametric means that DEA is a method that allows the functional form of a fit to data to be obtained without any statistical constraints. In other words, DEA results do not have any statistical properties, like mean or variance.

DEA is used to evaluate decision-making units (DMUs). DMUs can be any sort of measurables depicting similar variables. The different aspects of each DMU can be taken into account as long as these aspects can be presented numerically.

The history of DEA begins from the year 1957, when Farrell solved a Linear Programming (LP) problem for input-output data set of companies (Farrell 1957). After the discovery of DEA it has grown in popularity, but its usage has gained pace especially in the last fifteen years.

Usually it is common to refer to Charnes et al. (1978, 1979) and Banker et al. (1984) when discussing the waypoints of DEA development. Charnes et al. (1978, 1979) introduced the basis for DEA modelling, called CCR model after its protagonists Charnes, Cooper and Rhodes. CCR model gives all the DMUs under comparison an efficiency score from zero to one and can deal with both technical and economic efficiency. Technical efficiency refers to a DEA modelling of a data set which does not include production costs. If production costs were included, one would consider economic efficiency.
Banker et al. (1984) introduced yet another DEA model, called BCC model after its founders Banker, Charnes and Cooper. BCC model is a similar model as CCR model, but it takes a different view on returns to scale. BCC model views the high input and output values as a positive aspect for a DMU, thus giving it a better efficiency, while CCR model has neutral view on the magnitude of input and output values of a DMU, regarding the ratio on these values as the only measure for efficiency.

In recent years, the number of academic publications using DEA has grown, and it is viewed alongside with more traditional regression methods. Also, the number of possible models has grown, encompassing now as many as 20-50 models (Cooper et al. 2006), depending on the definition of a model. However, in this work the only DEA models considered are CCR and BCC models.

### 3.1.1 Mathematical presentation of DEA

Efficiency is defined as unit output per unit input, as follows:

\[
\text{Efficiency} = \frac{\text{output}}{\text{input}} = \frac{u_1y_{1o} + \cdots + u_s y_{so}}{v_1x_{1o} + \cdots + v_m x_{mo}}
\]  

(2)

In the definition, for DMU \(o\), outputs \(y\) and inputs \(x\) are given weights \(u\) and \(v\), respectively. By linear programming, the weights are chosen as to maximise the efficiency score for all DMUs \(o = 1, \ldots, n\), while the other DMUs apply limits to the weights \(u\) and \(v\). For example, the basic input oriented CCR DEA model for DMUs \(o = 1, \ldots, n\) can be given as follows:

\[
\begin{align*}
\max & \quad u_1y_{1o} + \cdots + u_s y_{so} \\
\text{s.t.} & \quad v_1x_{1o} + \cdots + v_m x_{mo} = 1 \\
& \quad (u_1y_{ij} + \cdots + u_s y_{sj}) \leq (v_1x_{ij} + \cdots + v_m x_{mj}), j = 1, \ldots, n \\
& \quad u_1, \ldots, u_s \geq 0 \\
& \quad v_1, \ldots, v_m \geq 0
\end{align*}
\]  

(3)

The model given above is a DEA CCR model. For BCC model we need to modify the problem statement. For BCC model we have

\[
\begin{align*}
\max & \quad u_1y_{1o} + \cdots + u_s y_{so} - u_0 \\
\text{s.t.} & \quad v_1x_{1o} + \cdots + v_m x_{mo} = 1 \\
& \quad u_1y_{ij} + \cdots + u_s y_{sj} - u_0 \leq v_1x_{ij} + \cdots + v_m x_{mj}, j = 1, \ldots, n \\
& \quad u_1, \ldots, u_s \geq 0 \\
& \quad v_1, \ldots, v_m \geq 0
\end{align*}
\]  

(4)

Here \(u_0\) denotes the variable returns to scale assumption of BCC modelling. It should also be noted that both equations above represent input-oriented DEA
models, due to the fact that it is the outputs that are maximised. The output-oriented DEA models however, minimise the inputs.

The weights for variables do not need to be defined as they are chosen optimally by the algorithm, in the sense that every DMU receives as high efficiency score as possible. This is how the different measurement techniques are further fused into a meaningful index of performance, giving us with relatively simple formulation the overall view of the problem at hand.

3.2 Rock breakage with DEA

One of the ways to quantify rock hardness is to study its chemical properties and compare the results to known values of rock breakage. In the case that this works, it eases the study of rock hardness considerably. The rock breakage is studied with DEA to reveal whether it can be used to model the breakage. The analysis introduces a certain quotient model which can take into account non-linearities in the modelling of rock breakage.

To extract valuable minerals or ore from the host rock, one has to break it. This is why ore ‘hardness’ is a property that is taken into account in designing a mineral processing flowsheet and managing a minerals processing operation (Kauppinen et al. 2014a). Ore hardness can be measured directly but this is time and expense intensive. Faster and cheaper indicators or ‘proxies’ for ore hardness, rather than direct measurements, are being researched in the industry. This is why the DEA was used to study rock breakage:

1. To alleviate the direct measurement of the rock ‘hardness’ by the help of proxies.
2. To test the Data Envelopment Analysis (DEA) in solving for rock breakage.

In this analysis, 98 data points were used for range of variables (total of 18), including chemical compositions determined by x-ray fluorescence (XRF), inductively coupled plasma (ICP) and atomic absorption spectroscopy (AA), and rock hardness as determined by the Point Load technique and Loss on Ignition (LOI). The analysis of this data was carried out by choosing all the possible combinations of two, four and six variables, modelling each selection with DEA. Moreover, a quotient DEA model was used to model non-linear behaviour. The model which best fitted the measured rock hardness was then selected, and the best proxies for rock hardness were assumed to be the variables which appeared in the best fit. Best fit was determined by simple square error.

3.2.1 Methods

The output oriented CCR-DEA model is used to solve for a vector of efficiency scores. The selection of the model was done a priori without any experimental work and other DEA models could also be considered. The output oriented CCR-DEA can be derived from input oriented CCR DEA model. The derivation yields:
min \sum_{i=1}^{m} v_i x_{i0} + \cdots + v_m x_{mo} \\
\text{s.t.} \\
u_1 y_{10} + \cdots + u_s y_{so} = 1 \\
(v_1 x_{1j} + \cdots + v_m x_{mj}) \geq (u_1 y_{1j} + \cdots + u_s y_{sj}), j = 1, ..., n \tag{5} \\
u_1, ..., u_s \geq 0 \\
v_1, ..., v_m \geq 0 \\

Where \( u \) and \( v \) are weights assigned to input and output variables, \( y \) includes all the output variables and \( x \) all the input variables (Cooper et al. 2006).

The method uses known values for rock hardness as the x-axis and the solved efficiency scores with given variables as y-axis. The Matlab code created tests the different combinations of variables in the given data, adding up to twenty thousand tested combinations of variables for each model. The chosen combination of variables is the one with least variance from a linear model for rock breakage, when the rock hardness is known for all data points.

### 3.2.2 Quotient model

The use of DEA modelling in rock breakage is a novelty and therefore the reasons behind choosing this data fusion method should be discussed. One may note that the DEA, eventhough it is linear programming method, can be used to produce nonlinear dependencies. This makes it superior to basic linear regression modelling.

To model for nonlinear dependencies, a quotient model \( E_q \) is introduced, of form

\[
E_q = \frac{E_1}{E_2} \tag{6}
\]

where \( E_1 \) and \( E_2 \) are defined as vectors of efficiency scores of output-oriented CCR-DEA models number 1 and 2, respectively. Therefore, the quotient model is an element-wise division of two different models.

The manner in which the quotient model is used in studying rock breakage suggests it is a relationship between input variables. Next an assumption is made that the output variables are dummy variables equal to one, making the measure of efficiency (see Eq. (2)) equal to \( \frac{1}{f_1} = \frac{1}{\sum v_i x_y} \) and similarly for \( f_2 \). Let there be a relationship between the ratio of input variables which defines the quotient model, namely \( \frac{f_2}{f_1} \). This relationship can be seen as defining a group of possible solutions, where DEA modelling is included as a subset. Mathematically this can be expressed as:
Therefore, the DEA solution belongs to the set of possible quotient model solutions for the ratio of input variables, where the input variables are defined as

$$\frac{f_1}{f_2} \leq \frac{1}{\frac{1}{f_1}} \max \frac{1}{f_2} \quad \text{s.t. ...}$$

(7)

Therefore, the DEA solution belongs to the set of possible quotient model solutions for the ratio of input variables, where the input variables are defined as

$$f_1 = \sum v_i x_i$$

(8)

and similarly for $$f_2$$.

The quotient model creates a ratio relationship between the variables. Above this is demonstrated to show a relationship akin to ratio of factors making rock easier to break and factors making the rock harder to break. This implies that DEA results, including the variables chosen from 18 variables, can be interpreted in the context of the factors making rock easier and harder to break. There can be other possible contexts of interpretation which fall outside the scope of this study.

Quotient model therefore makes it possible to not only model nonlinear dependencies (as there is a denominator in the model) but also to understand the reasons behind why certain variables become chosen for the model (by testing all the possible combinations). Some factors make the rock harder to break and others make it easier to break. Quotient model was shown to be a possible model for interpreting the model variables in this context.

### 3.3 DEA to assist geologists in logging drillcores

After developing the CCR model, one may introduce BCC model, which is used here for the computations. This is done as explained previously (see 3.1.1).

Figure 4 shows graphically how CCR and BCC efficiency frontiers differ from each other for a data of 200 data points (Kauppinen 2016). The data is random data, including one input and one output variable. Graphically $$u_0$$ marks the difference between the efficiency frontiers of CCR and BCC models (see Figure 4b). The red line denotes the BCC efficiency frontier, while the blue line denotes the CCR efficiency frontier. One can see from Figure 4b that the BCC and CCR frontiers are considerable different from each other. The points on the efficiency frontier always have an efficiency of one. Therefore the BCC model gives better efficiency scores for the data set under study. Also, the rest of the points score a better efficiency scores as the distance from the efficiency frontier is somewhat reduced.
The Eq. (4) for BCC model in Section 3.1.1 is an input oriented model. The output-oriented model can be understood graphically in one input and one output case (see Figures 4a and 4b) as measuring the distance from the efficiency frontier along the y-axis, while input oriented model does this along the x-axis.

In the case of one input and one output, one can measure the individual efficiency score of a DMU in a simple manner. This is done by taking the total vertical distance to the x-axis from a given point, and dividing the result by the distance that a vertical line through the given point has to the efficiency frontier. Eq. (9) introduces the output oriented BCC DEA model (see also Eq. (5)). One can note that the difference between the CCR and BCC model is exactly the variable $u_o$. Therefore $u_o$ varies for different DMUs $o$ (See Figure 4b).

The optimisation process is as follows: the DMU 1 is chosen, the optimization is solved for DMU 1. This is then repeated for DMU 2 and so on, until also the DMU $n$ has an efficiency score. Effectively, we have $n$ efficiency scores, one for each DMU.
3.3.1 Data Envelopment Analysis compared to Principal Components Analysis

In this subsection, DEA is compared to Principal Component Analysis (PCA, see also Shlens, 2005), which can be used to construct a method comparable to certain DEA models. Later it is also shown how PCA can be used to make an algorithm producing the same results as output-oriented BCC DEA model, but in less calculation time.

To begin with, the modelling scheme for PCA is introduced, which generates the same efficiency frontier as the BCC DEA model. The main reason in using PCA is to speed up the process of finding local maxima. Performing PCA is in practice quite simple. After (1) organising the data into a \( p \times n \) matrix (\( p \) rows are the input and output variables, in this order, and \( n \) columns are the different DMUs) and (2) subtracting the mean of each row, (3) the eigenvectors of the covariance can be calculated. Eigenvectors can be solved for the symmetrical matrix \( A^T A \) from the Eq. (10).

\[
A \overset{p \times n}{\times} U = E \overset{p \times p}{\times} D \overset{p \times p}{\times} V^T \quad (10)
\]

where \( A \) is the \( p \times n \) data matrix, \( E \) is the matrix of eigenvectors and \( D \) is a diagonal matrix consisting of the eigenvalues of \( A^T A \) (Shlens 2005).

One can also use the Singular Value Decomposition (SVD) to arrive to eigenvectors. Any \( p \times n \) matrix \( A \) can be written as (Shlens 2005)

\[
A = U S V^T \quad (11)
\]

where \( U \) equals the eigenvectors of \( A A^T \), \( V \) equals the eigenvectors of \( A^T A \), \( S \) is a diagonal matrix consisting of the eigenvalues of \( A A^T \), \( V^T V = I \), and therefore

\[
A A^T = U S^2 U^T \quad (12)
\]

After calculating the eigenvectors, which can also be called the principal component axes, it can be noted that these axes are orthogonal to each other and useful in the context of finding the efficiency frontier of DEA. As it stands, the second principal component axis can be directly used to measure the highest point on the efficiency frontier (for a graphical presentation for the one output

\[
\begin{align*}
\min_{v, \lambda, u_0} & \quad v_1 x_{10} + \cdots + v_m x_{m0} - u_0 \\
\text{s.t.} & \quad u_1 y_{10} + \cdots + u_k y_{k0} = 1 \\
& \quad (v_1 x_{1j} + \cdots + v_m x_{mj}) \geq (u_1 y_{1j} + \cdots + u_k y_{kj}) + u_0, j = 1, \ldots, n \quad (9)
\end{align*}
\]

\[ u_1, \ldots, u_k \geq 0 \]

\[ v_1, \ldots, v_m \geq 0 \]
and one input data, see Figures 4a and 4b, and for the second principal component axis, Figure 5). This is done by finding the maximum value of all the points in to the direction of the axis of the second principal component.

Following pseudocode can be formulated for PCA, finding the DEA efficiency frontier:

1. Find the numerical value of each DMU in the direction of $n$th Principal Component, where $n$ is the total number of input and output variables in the problem statement.
   a. Pick the point with the highest value; store it.
   b. Check if the point picked is higher or equally high on all the output axes and lower on all the input axes to other points; label the other points.
2. The point picked in 1.a is always more efficient compared to the points labelled in 1.b; the points labelled are not needed in finding the efficiency frontier and are omitted from the search.
3. Store the point found in 1.a and omit it from the rest of the search.
4. Check if there are points left to analyse;
   a. If no, then stop.
   b. If yes, continue (returning to 1.).

![Figure 5. PCA space for data with one input and one output (see the short arrow pointing upwards for the second principal component, Kauppinen 2016).](image)

The pseudocode introduced is meant for output-oriented BCC DEA modelling. If there would be a need to make a different DEA model, different pseudocode would be needed.

The process of searching the local maximum in principal component space produces the DEA efficiency frontier. It should be noted that the pseudocode above produces a DEA BCC efficiency frontier. The algorithm always picks up the highest value measured on the $n$th (or second, in case of one input and one output) axis of principal component space, stores it and after this ignores it in
further calculation. The algorithm picks up the next highest value and continues
the iteration. After all the points have been analysed, the program terminates,
with the stored efficiency frontier values.

After receiving the efficiency frontier values, another algorithm can be built
which finds the individual efficiency scores for all the points. This can be done
by vector calculus: the algorithm measures the distance of the points from the
efficiency frontier, and divides this distance with the total distance from the x-
axis to the efficiency frontier (see Figure 4). It should be noted that this proce-
dure produces efficiency scores for a DEA BCC model with output orientation.

Regarding the case with \( n \) input and output variables in total, one can note the
following. The direction of \( n \)th principal component is crucial, because the effi-
ciency frontier will be formed with the use of this direction, regardless of other
factors. Therefore, the principal component selected for the analysis is not al-
ways \( n \)th, but more precisely \((n-k)\)th, because one selects the output variable
with least variance, and input variables should not be selected as possible direc-
tions. Variable \( k \) is chosen as follows. The variance of each \( p \) rows of \( A \) is com-
puted, with the idea of approximating the variance of principal components.
Now \( k \) can be defined as the number of input variables with less variance than
the output variable with least variance of all the output variables.

After forming the direction properly, the algorithm is able to discern points
that fall on the efficiency frontier. This is because the algorithm always picks the
next highest point to the direction chosen, if that point is not omitted from the
search (see pseudocode above). However, the pseudocode needs to be run \( m \)
times for \( n \)th dimension case, where \( m \) is the number of output variables in \( A \).
It should also be noted that \( k \) is defined independently for each run \( i \), with the
following definition: \( k \) in the run \( i \) is the same as in the run \( i-1 \), but there is an
addition of one plus the number of input variables with less variance than the
\( i \)th output variable but more variance than \((i-1)\)th output variable. In the first
run, \( k \) is solved as given in the previous paragraph.

3.3.2 Preparing the data for DEA

DEA models have by definition both input and output variables. In the case of
this study, only two output variables (pyrite and arsenopyrite content) are con-
sidered. These output variables were selected because Kittilä mine has all its
gold in pyrite and arsenopyrite, free gold being very rare. The only input variable
taken into account is the depth from which the drillcore is excavated. It is fasci-
nating however to understand the possibility of using more input variables, or
for example the costs as input variables, even though in this thesis these options
are left unused. It can also be argued that DEA modelling works also with
greater number of inputs and outputs than in the case of Kittilä mine studied.

Haskin et al. (1997) introduced a point-count method, which is used here as
follows: Pyrite mineral is given a window of 350-375 nm, all the highest peaks
found in this sector classify the single measurement as representing pyrite. On
the other hand, the highest peaks found at wavelengths 330-350 nm are classi-
fied as arsenopyrite.
3.4 Sustainability using DEA

This thesis shows that DEA can also be used to measure sustainability. It is a topic of discussion where there are many new publications, offering different conceptualisations and measures for sustainability (e.g. Hassan et al., 2014). All of this discussion leads to a dispersed field of study, offering a daunting task for any instance for creating a mutually agreed on and coherent measure for sustainability. Therefore, this study seeks to show how a simple but effective measure can be developed for sustainability (Kauppinen & Khajehzadeh, 2015).

The most common conceptualization used is to divide sustainability into economic, ecological and social sustainability and this scheme is used here as well. In addition, the measure used is DEA, which has the same outline as before while studying rock breakage and geology. This should give further confidence on the method used as it has a proven track record of producing reliable results. Usually, DEA is used for efficiency measurement but here efficiency is defined as a measure of three dimensions of sustainability. This type of simple framework makes the measurement of sustainability of exploration transparent and helps develop the sustainability of the exploration phase.

As mentioned beforehand, there are multiple sustainability schemes already available for mining sector. The most prominent sustainability measurement practice in Finland is the one formulated by The Finnish Network for Sustainable Mining, who have created through meticulous discussions with the local stakeholders a detailed measurement tool of factors affecting sustainability, giving each measurement subsection a classification from AAA (the best) to C (the worst). In global context, these types of measures have been developed in other countries as well, most notably in Canada, where a measurement system functioning as an example to the Finnish system was agreed upon. However, it is worth to note that these tools have a very detailed accounting processes whereas here the study strives for simplicity. One can note that there is a line of discussion in literature giving criticism on the complexity of measurement (Giurco & Cooper 2012). This thesis continues along the lines of the given critique, providing a simple measure for sustainability.

This study develops a new approach for sustainability, comprising just the exploration phase of mining. The basic idea is introduced in Figure 6, showing the exploration phase as a system with different inputs and outputs. It can be said that the DEA is used as a modelling tool for sustainability (Kauppinen & Khajehzadeh 2015). The objective of this research is to establish sustainability scores for the exploration phase of different exploration projects. In DEA, it is of paramount importance to choose the model correctly, but even more important is the question which variables to use. The type of DEA model used is somewhat dependent on the data used, so it is a two-stage process: first, a suitable set of data should be gathered and second, a suitable model should be selected.
This study on sustainability uses artificial data, which has been verified using other sources (Kauppinen & Khajehzadeh 2015). Because the data is artificial, however, it is of no real interest to discuss the values of the data as such (for example, no variance is included) but just to inform the reader that the author has formulated the data keeping in mind that it should be easy to gather from a real exploration site, both in commencing one as well as sustaining its functions. The DEA approach does not require any weights for different decision variables, but the selection of variables themselves is crucial, therefore a short discussion on the variables selected follows.

In this study, the chosen variables are as follows: investment; future profit; surface disturbance; environmental programs; disturbance of World heritage sites or indigenous peoples; and salaries. One can only select six variables because of the small number of exploration sites covered. If there were more exploration sites included in the analysis then the amount of variables selected could also be higher (see e.g. Cooper et al. 2006).

The economic variables are investment and future profit. The investment is treated as an input variable because the company that owns mine A has to make an investment in order to start exploring on a given site. However, they are also entitled to make profit and that is why the output variable is selected to be future profit. The profit is discounted, meaning that if the mine operates for 10 years after commencing the exploration, each profit from each year needs to be discounted to the moment where the investment was made, to make the two values comparable.

The ecological variables are surface disturbance and environmental programs. Surface disturbance was selected among many other options as the input variable for ecological values, as it is seen as a possible proxy attribute for many other ecological factors. The decision in selecting the ecological variable was hard. In
the end, the decision was made between three variables, namely greenhouse gas emissions, water contamination and surface disturbance. Of these, surface disturbance is easiest to measure and can possibly work as a proxy for GHG-emissions and water contamination, so surface disturbance was selected as ecological input variable. The selection of the output variable was also difficult, but for different reasons. The question needed to be asked was: how can mining benefit ecology? This was seen as a possibility to promote ecological development programs. Of course, it would be better to measure these programs by their impact, but this is difficult at the moment. That is why the measure chosen for ecological output variable was the monetary value spent into ecological development programs.

The social variables chosen were the disturbance of World heritage sites or indigenous peoples as the input variable, and salaries as the output variable. Social input variable is a simple yes or no fact if there has been disturbance or not, which is a simplification, giving more weight to the possible disturbances reported. If there has been disturbance, the value is 1, if there hasn’t been any disturbance, the value is 0. It is important to note that this makes the disturbance very unwanted, as it gives an otherwise zero input variable a value clearly larger than zero. Also, the output variable is the total amount of salaries paid in the exploration site to the personnel working there. This is a useful selection for output variable as it measures the positive social impact of the exploration.
4. Data preparation for linear programming

The data in raw form need preparation before the described processes can be applied to it. Here the different cases of data preparation for use in the dissertation are described and the data preparation is given, which is done with help of Matlab (without describing the linear programming Matlab functions themselves).

4.1 Data for Raman measurements

Data received from the Raman measurements needs normalization, baseline correction and high and low frequencies need to be tuned so that the different wavelengths measured are in line with every spectrum applied to them. These data preparation techniques were all done as part of the Matlab code which solves for abundance (Kauppinen et al. 2013), concentration (Kauppinen et al. 2014b), LIF measurements calibrated by Raman (Kauppinen et al. 2014c) and finally efficiency in DEA (Kauppinen 2016). In all these cases, the data preparation is similar. Before any preprocessing can occur, the raw data containing all the necessary data was read in memory for use in Matlab.

4.1.1 Normalisation, baseline correction and wavelength regularisation

Normalisation and baseline correction were done before any other techniques applied to raw data. Baseline correction was done by algorithm given by Eilers and Boelens (2005) to remove the baseline from the reference spectra. Normalisation was done to remove the mean and to scale each reference spectrum to unit variance.

After this preprocessing, the tail and head of the Raman spectra in question were treated. The highest values and the lowest values were cut off, so that only spectral values between 188-1200 of Raman shift (cm⁻¹) were used. This process was done for both measured spectra and the reference spectra (Downs 2006). Furthermore, the spacing of measured points was regularised so that the measured wavelength points and the reference wavelength points were in the same
wavelengths. After normalisation, baseline correction and wavelength regularisation the data was ready for further analysis, with techniques described in the dissertation.

### 4.2 Data for Laser-Induced Fluorescence measurements

For LIF imaging, average hue values of the known areas of interest in LIF images are computed (Kauppinen et al. 2014c). The LIF imaging data received does not need preprocessing. However, in order to arrive to LIF results, the LIF method should be calibrated by Raman. The method of calibration is line scanning of the area of interest and therefore the data preparation is similar to the Section 4.1.

Figure 7 gives the results of the calibration. Figure 7 (a) shows the picture taken in visible light while 7 (b) shows the zoomed in view on the area of interest in UV light. Finally, Figures 7 (c) and (d) show the minerals found by LIF and Raman in a 1 cm line scanned. It can be seen that the red hue range (Figure 7(c)) and bluish white hue range (Figure 7 (d)) coincide with Raman measurement results for albite and magnesite, respectively.

![Figure 7](image)

**Figure 7.** (a) Drillcore sample from Kevitsa mine, Finland, depth 18m. Specified region was irradiated by UV laser, (b) LIF image of the same region, Raman spectroscopy scanned 1 cm line on the drillcore surface. Albite (red) and Magnesite (bluish white). (c) Position of Albite detected by Raman and LIF along the same scanned line,(d) position of Magnesite detected by Raman and LIF along the same scanned line. (Kauppinen et al. 2014c).
4.3 Data for sustainability measurements

In sustainability measurements, the data preparation was not needed, as the data was only used for Data Envelopment Analysis (DEA). DEA is versatile in using different types of data, and binary as well as continuous metrics could be used in the same model (Kauppinen & Khajehzadeh 2015). Despite of this, the model was tested to yield plausible results before using the modelled results.
5. Linear programming

Linear programming methods are an intriguing choice to analyse spectral data. In literature (e.g. Dörfer et al. 2009) these linear models are seen as weak or outright wrong way to analyse spectral data, although some recent publications show the potential of linear programming in analyzing spectral data (e.g. Hosseini & Abedi 2015). All of the articles in this dissertation make use of linear programming (also called linear optimisation) and it is therefore appropriate to discuss the method accordingly. This chapter reviews the linear programming (LP), its history and modern day uses as well as the exact formulation of different ways of solving LP problems used in this dissertation.

LP can be defined shortly as finding the minimum (or maximum) solution from the set of possible solutions for a given problem which introduces the mathematical dependencies between the variables included in the problem. The minimum (or maximum) solution set is called the objective function, under constraints given by the LP problem statement. In linear programming, the objective function and constraints can only be linear. Constraints are defined as equalities or inequalities. To solve a linear programming problem, one needs an algorithm which selects from the set of possible solutions the minimum (or maximum) solution.

5.1 History

The first generalised algorithm to solve for a system of linear inequalities was introduced in the beginning of the 19th century (Fourier 1827). However, the first modern scientist to propose a general linear programming problem and its solution was Kantorovich (1939). In the Second World War this method was developed extensively. Moreover, Koopmans (1942) formulated linear programs of classical economic problems. Simplex method, which is a highly successful method for solving linear programs, was first introduced by Hitchcock (1941), but the final formulation was given by Dantzig (1947).

Building on this work, John von Neumann (1948) introduced the dual or duality theory for linear programming problems. The later developments included gradual development in both computer computation power and algorithms used to solve for LP problems. In 1984, Naredha Karmarkar developed an interior-
point method (Karmarkar 1984), which is in use today. It can be noted, that even this dissertation uses a variant of the interior-point method.

5.2 Canonical and standard form

Linear programs are problems that can be expressed in canonical form as

\[
\begin{align*}
\text{maximise} & \quad c^T x \\
\text{subject to} & \quad Ax \leq b \\
\text{and} & \quad x \geq 0
\end{align*}
\]

where \( x \) represents the vector of variables (to be determined), \( c \) and \( b \) are vectors of (known) coefficients, \( A \) is a (known) matrix of coefficients and \((.)^T\) is the matrix transpose.

5.2.1 Standard form

Standard form can also be used. In matrix form it can be given as

\[
\begin{align*}
\text{max} & \quad c^T x \\
\text{subject to} & \quad Ax \leq b \land x \geq 0
\end{align*}
\] (14)

Other forms can be written in the way of the standard form.

5.3 NNLS by standard form

The formulation of NNLS is the same as previously, namely

\[
\begin{align*}
\text{min} & \quad \| x_{\text{sample}} - R_{\text{ref}} \|^2, c \geq 0.
\end{align*}
\] (1)

One may note that adding the norm into the problem takes one further away from the standard form, effectively measuring the distance from zero by a formed vector and not by scalar as in (14).

5.4 Derivation of the CCR DEA model using standard form

\[
\text{Efficiency} = \frac{\text{output}}{\text{input}} = \frac{u_1y_{\text{io}} + \ldots + u_5y_{\text{so}}}{v_1x_{\text{io}} + \ldots + v_mx_{\text{mo}}}
\]

Efficiency is defined as unit output per unit input. To analyse efficiency further, one may create a fractional linear optimisation problem, as follows.
This holds for each Decision-Making unit (DMU) \( o=1,...,n \). Above is a fractional problem, which can be modified into a Linear Programming (LP) form by defining

\[
\begin{align*}
&\text{max} \quad u_1y_{1o} + \cdots + u_5y_{5o} \\
&\text{s.t.} \quad \frac{u_1y_{1j} + \cdots + u_5y_{5j}}{v_1x_{1j} + \cdots + v_5x_{5j}} \leq 1, j = 1, \ldots, n \\
&\quad u_1, \ldots, u_5 \geq 0 \\
&\quad v_1, \ldots, v_5 \geq 0
\end{align*}
\] (15)

One gets

\[
\begin{align*}
&\text{max} \quad u_1y_{1o} + \cdots + u_5y_{5o} \\
&\text{s.t.} \quad v_1x_{1o} + \cdots + v_5x_{5o} = 1 \\
&\quad (u_1y_{1j} + \cdots + u_5y_{5j}) \leq (v_1x_{1j} + \cdots + v_5x_{5j}), j = 1, \ldots, n \\
&\quad u_1, \ldots, u_5 \geq 0 \\
&\quad v_1, \ldots, v_5 \geq 0
\end{align*}
\] (3)

Now we can treat this as a standard LP-problem. Solving for each DMU \( o=1,\ldots,n \), the efficiency scores \( E_o, o=1,\ldots,n \) are found. This method of calculation is called DEA CCR modelling.

By knowing the DEA CCR model it can be further developed into DEA BCC model (see Section 3.1.1, Section 3.3 and Kauppinen 2016).
6. Results

The following results were achieved. The main points raised were

1. The use of linear programming in finding the abundance and concentration in an ore sample;
2. Combining Raman and LIF measurements to provide fast and reliable results from a drillcore sample; and
3. Use of DEA in the study of rock breakage, drillcores and sustainability.

6.1 NNLS as a possible tool to solve for abundance

Kauppinen et al. (2013) discusses the measurement of mineral abundance using Raman spectroscopy. The results achieved are given for both cases of minimising the square and absolute value of the difference between sample spectrum and the reference spectra.

In general, the methods proved to be adequate in explaining the sample spectra. It was found that also nonlinear optimisation is able to discern useful information for the user. Nonlinear algorithm is able to discern two or more minerals from a given sample. The more complex cases, containing multiple peaks on many different wavenumbers, were beyond the scope of this study because of the limited array of reference spectra used.

6.1.1 Kittilä sample spectrum no. 4

The study also examines a case of combining the reference spectra to form a single sample spectrum. The sample spectrum chosen holds both characteristics of rutile and ankerite, which are not important minerals for Kittilä, but useful for establishing the mathematical model used. There are clear differences between different methods, namely; linear, non-linear and Principal Component Analysis (PCA) -aided optimization, in discerning the peaks. This is due to difference in the optimisation routines but also in the use of absolute values and squares in the cost function. However, the double peaks of rutile as well as the single peak of ankerite were clearly identified by the optimization. The minor errors in the optimization are due to either an error in measuring the sample spectra or in the small amount of reference spectra used to identify the sample spectra.
The best fit for the double peak of rutile as well as the single peak of ankerite is given by the linear combination of the reference spectra. The clearest difference is the ankerite peak, where the Non-negative least squares (NNLS) optimisation clearly follows better the sample spectrum whereas nonlinear optimisation with absolute values fails somewhat in discerning the peak. However, even the nonlinear optimisation with absolute values finds the peak, only that it does not reach the intensity of the original sample spectrum. When examining the optimal values for the optimisation methods used, it can be concluded that the nonlinear optimisation routine identifies the ankerite peak as siderite, which does not fit the data as well as ankerite peak. Using PCA-aided optimisation could be a more cost-efficient way of arriving to the same results as NNLS. However, PCA fails to discern between the main minerals in the sample. For example, with PCA-aided optimization, the ankerite peak is not found. The same kind of problem arises throughout the sample set and PCA cannot be recommended for use in discerning the main minerals.

6.1.2 Existence and concentration of minerals

The question now arises if one can use the solved abundance as relative concentrations of different minerals in the sample. Haskin et al. (1997) analyse in depth the different factors which might have an effect on the peak intensities of the Raman measurement. In their work, for our measurement setup, the most heavy burden placed upon such interpretation of abundance are both compositional heterogeneities within grains encountered by the laser beam and mineral transparency to both incident beam and scattered radiation. In Haskin et al. (1997), mineral standard deviations about mean peak heights are typically 40-60 percent.

The substitution of abundance as concentration is therefore dubious. Only the existence of minerals can be observed as a possible topic of discussion with these experiments. To measure concentrations, the experimental setup should be modified as to make sure that the measurements only have a minimal effect on the relative concentrations of different minerals. Designing such an experimental setup is considered in the Section 6.2.

6.2 NNLS used to solve for concentration

The NNLS can be used for a more in depth analysis of a sample. An example of NNLS optimisation is shown in Figure 8. It is seen that there is both pyrite peak (left) and ankerite/magnesite peak (right) in the Figure 8, as well as bornite peak in the centre (Kauppinen et al. 2014b).
Figure 8. The linear combination of reference spectra (red) and an average of one hundred sample spectra (blue), sample no. 2 (Kauppinen et al. 2014b).

Figure 9. Concentration of main minerals as a function of position (10 times 10 grid), sample no. 2 (Kauppinen et al. 2014b).

Figure 9 shows more detailed data on ankerite/magnesite and pyrite concentrations. When using NNLS one can find multiple minerals in the same position. Most notably this is shown in grid points 70-100 where both ankerite/magnesite and pyrite are present in high amounts. This is an anomaly compared to the point-count method. Most likely multiple minerals appear in the same measurement point because the Raman laser hits two grains at once. NNLS method is therefore more useful than the point-count method as it can take into account the situation where two or more grains are hit at the same time.

The eight pulverised samples (1-8) from Kittilä gold mine were all measured with three different methods: Mineral Liberation Analysis (MLA), point-count and NNLS concentration. The results of all the measurements are shown in Table 2. MLA, measured in a commercial laboratory setup, is used as reference values to compare point-count and NNLS concentration values to each other.
It can be seen that NNLS concentration values follow the MLA results better than the point-count method. There are deviations, however. Notably the NNLS result on sample no. 8 deviates from the MLA result (see Table 1). This can be because of some mistake in the measurement setup, e.g. dust on the sample. As both point-count and calculated concentration failed in recognising the pyrite in sample no. 8, the analysis conducted only refers to samples from 1 to 7.

Table 1. The total amount of pyrite measured with MLA, point-count and NNLS, where NNLS is measured directly from the sample spectra (Kauppinen et al. 2014b).

<table>
<thead>
<tr>
<th>No. of sample</th>
<th>MLA results for pyrite (%)</th>
<th>Point-count results for pyrite (%)</th>
<th>NNLS concentration for pyrite (%)</th>
<th>Variance of NNLS Concentration (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33.7</td>
<td>59</td>
<td>34</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>16.9</td>
<td>23</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>34.5</td>
<td>66</td>
<td>36</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>10.6</td>
<td>21</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>21.3</td>
<td>51</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1.1</td>
<td>19</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>0.5</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>68.0</td>
<td>10</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

Moreover, low concentrations are also in doubt for both point-count and NNLS concentration (see sample no. 6 and 7). However, for other samples the NNLS concentration is closer to the MLA results, compared to the point-count method (see sample no. 1, 2, 3, 4 and 5).

Referring to Table 1, the uncertainties are presented as a variance and not standard deviations. As it stands, standard deviations would be notably higher, at worst from 0 to 100 percentage. However, as the results look quite promising, one may note that with more measurements per sample the concentrations can be identified with less uncertainty.

The Raman analysis method described here can also be considered for on-site analysis of rock drillcore. In this context, the speed of the measurement is of the essence. For ten thousand measurements per one sample, the measurement would take two and half hours per sample. For some on-site applications this is too slow. However, it is much faster than laboratory analysis (e.g. MLA) for which it might take weeks or even months to receive the results. There is also a balance between speed and precision for Raman analysis. For thousand measurements per sample one may find the speed increased to only fifteen minutes per sample, but this has an adverse effect on the precision of the measurement.

6.3 LIF and Raman combined for industrial scale online scanning for rock drillcores

The calibration of LIF by Raman was discovered to be an effective means of finding minerals in LIF imaging (see also Figure 7 in Section 4.2). LIF imaging technique is also much faster than Raman analysis, providing a possibility for online scanning of ore samples. However, there were some issues which needed
more study. As discussed before, some of the sample spectra may display a property of having a multiple reference spectra merged together in Raman analysis, and this was the case also in calibrating LIF with Raman analysis. The NNLS optimization routine was used to find a linear combination of reference spectra which matched the given sample spectrum.

Let us examine a case of combining the reference spectra to form a single sample spectrum more closely (Kauppinen et al. 2014c). With a more varied sample spectrum, there are many peaks visible. When the individual measurements are studied, every sample spectrum clearly presents one, two or three different minerals. This is due to fitting of reference spectra (see Kauppinen et al. (2013)). If we use point-count method, the mineral most clearly presented is the one that is chosen to present the whole data in that specific point of measurement.

Why the NNLS method is then used and not the simpler point-count method? The answer is three-fold (Kauppinen et al. 2014c). First, the NNLS method is fast enough to form the results needed for any experimental setup on industrial or smaller scale. Second, NNLS method is backward compatible with point-count method as we discussed in Section 6.2. Third, the results achieved through NNLS method, although here used only for point-count, can be further analysed to present more varied Raman data on the samples measured.

One has to also go back to the hypothesis of this dissertation which stated that in Raman analysis, there is a certain combination of known library spectra which can be used to model the sample spectrum. One should note that the accuracy of NNLS method depends on the number of reference spectra used. If a spectrum with a certain peak is not available in the reference spectra, it cannot be fitted. The reference spectra used in calibration of LIF with Raman were 77 in total, collected from RRUFF database (Downs 2006).

6.3.1 LIF colours

In this section focus is on the information that luminescence provides about the minerals. Kauppinen et al. (2014c) simply implies that in the UV illumination, luminescent minerals show strongly coloured fluorescence and makes use of the idea of mineral exploration using LIF colours. LIF technique needs to be calibrated using Raman spectroscopy. Small-scale and sporadic Raman measurements were performed on the regions containing large quantities of distinct minerals based on their exposed fluorescent colours. After scanning the drill-core samples with Raman spectroscopy, the dominant red colours were associated to Albite (NaAlSi3O8), green colours to anthophyllite ((Mg, Fe)7Si8O22(OH)2) and bluish white colours to Quartz (SiO2).

The incorporation of Raman spectroscopy results into LIF images is done as follows: a photograph was taken of one piece of the drillcore sample from a depth of 18 m below the ground level (See Figure 7, Section 4.2), and another photograph was taken of the same area specified after being illuminated by UV laser. Raman method scanned 1 cm of the specimen. Each Raman spot size is 10 μm and each pixel of LIF image is about 15 μm. According to Raman results, the red colour was associated to Albite and the bluish white colour to Magnesite (MgCO3) (which is similar to that of Quartz (SiO2), mentioned earlier).
Exposed colours of the fluorescent minerals in the LIF images are noticeably following the Raman spectroscopy results since Albite has been detected by Raman technique at the same position in which associated LIF colour was revealed. The same results were reached for Magnesite, which similarly show the correspondence between Raman results and LIF colours. Compatibility of Raman spectroscopy results and LIF analysis has been investigated through multiple drillcore samples from various depths.

Due to the different activators in minerals, similar minerals may fluoresce with dissimilar colours in different mines. Moreover, different minerals may have fluorescence with identical colours because of analogous activators (Kauppinen et al. 2014c). Based on this, one can see the importance of Raman spectroscopy in calibrating the analysis of LIF images, since LIF analysis technique is incapable of distinguishing between the minerals with same colour.

6.3.2 Abundance

As discussed in the study (Kauppinen et al. 2014c), average hue values of the known areas of interest in LIF images are calculated. Having an average value of the target colour, a tolerance is specified in hue dimension and finally a mask is employed. By using the hue component of the HSV colour space, original LIF image is masked for red and bluish-white colours.

6.4 Rock breakage studied with DEA

The results showing the different models for rock breakage chosen as the best by the algorithm introduced are given in Table 2. These results were calculated using the DEA CCR model with output orientation, as this was found to be the best model to use in the preliminary analysis. This model was then used throughout the analysis. The best model of CCR-OO models is clearly the quotient model of 4 variables in the divisor and 4 variables in the dividend (model no. 4, see also Figure 10). Figure 10 shows that there is a clear correlation with Efficiency scores and rock hardness (A*b). This provides confidence in the method used.

DEA technique has not been used for rock breakage before. However, if one assumes the combination of attributes identified by technique described to indeed have the highest correlation with rock A*b for the ore studied, the results are found useful.

1. The DEA technique described may provide mining professionals guidance in which variables (or ‘proxies’) to take into account while solving for rock hardness.
2. The DEA technique has correctly identified some attributes as proxies which are known to have an effect on rock hardness. This demonstrates consistency with established practice (e.g. minerals containing metals are generally harder).
3. The DEA technique revealed the proxies in automated manner, giving faster results than hand-picked proxies would be.
4. The DEA technique also found some attributes which are not currently used as proxies (e.g. Sulphur content).

The best model found (see Table 2, model 4) only has XRF elemental analysis components. This would effectively mean that only XRF assays, which are cheaper than many other geophysical characterizations (such as ICP, inductively coupled plasma), are needed to assess $A^*b$. This could mean that XRF assays could be used as preliminary means to solve for $A^*b$, while giving significant savings in cost and time for mining companies, and revealing possible areas for more detailed analysis.

**Table 2.** The results of the study on rock breakage*.

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Description</th>
<th>Variable 1</th>
<th>Variable 2</th>
<th>Variable 3</th>
<th>Variable 4</th>
<th>Variable 5</th>
<th>Variable 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CCR-oo, 2 variables</td>
<td>Magnesium Oxide (XRF)</td>
<td>Point Load</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CCR-oo/CCR-oo, 2 variables, dividend</td>
<td>Magnesium Oxide (XRF)</td>
<td>Point Load</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCR-oo/CCR-oo, 2 variables, divisor</td>
<td>Aluminium Oxide (XRF)</td>
<td>Sodium Oxide (XRF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CCR-oo, 4 variables</td>
<td>Copper (XRF)</td>
<td>Titanium Oxide (XRF)</td>
<td>Loss of Ignition</td>
<td>Point Load</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CCR-oo/CCR-oo, 4 variables, dividend</td>
<td>Copper (XRF)</td>
<td>Sulphur (XRF)</td>
<td>Iron Oxide (XRF)</td>
<td>Loss of Ignition</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCR-oo/CCR-oo, 4 variables, divisor</td>
<td>Aluminium Oxide (XRF)</td>
<td>Potassium Oxide (XRF)</td>
<td>Manganese Oxide (XRF)</td>
<td>Fosforus Oxide (XRF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CCR-oo, 6 variables</td>
<td>Gold (AA)</td>
<td>Copper (XRF)</td>
<td>Sulphide (ICP)</td>
<td>Titanium Oxide (XRF)</td>
<td>Loss of Ignition</td>
<td>Point Load</td>
</tr>
<tr>
<td>6</td>
<td>CCR-oo/CCR-oo, 6 variables, dividend</td>
<td>Gold (AA)</td>
<td>Copper (XRF)</td>
<td>Sulphide (ICP)</td>
<td>Titanium Oxide (XRF)</td>
<td>Loss of Ignition</td>
<td>Point Load</td>
</tr>
<tr>
<td></td>
<td>CCR-oo/CCR-oo, 6 variables, divisor</td>
<td>Sulphur (XRF)</td>
<td>Sulphite (XRF)</td>
<td>Aluminium Oxide (XRF)</td>
<td>Calcium Oxide (XRF)</td>
<td>Potassium Oxide (XRF)</td>
<td>Strontium Oxide (XRF)</td>
</tr>
</tbody>
</table>

*The variables include chemical compositions determined by x-ray fluorescence (XRF), inductively coupled plasma (ICP) and atomic absorption spectroscopy (AA), hardness as determined by the Point Load technique and Loss on Ignition (LOI). There were 98 data points.*
Figure 10. Output oriented DEA CCR model per Output oriented DEA CCR model (quotient model), 8 variables.

6.5 Rock drillcores studied with DEA in the exploration phase of mining

The main result achieved in the study of rock drillcores with combined Raman analysis and Data Envelopment Analysis (DEA) is that the DEA was found as a reliable method to assist geologists to log drillcores. The speed of scanning pulverized and homogenized sample of one meter of drillcore is approximately fifteen minutes, giving a possibility to perform the scanning at the exploration site. This in turn helps geologists log drillcores more effectively as they can concentrate on the parts of drillcore which have a high amount of minerals of interest. Combined DEA and Raman analysis therefore provide an auxiliary means of arriving to a decision and checking the results received by the trained eye for the geologists.

Dissertation is contributing on the different means of producing on-line exploration data. Therefore, the ideas presented cannot yet be ones that already have been tested. Besides, it is a convincing argument that with one or two ship containers a needed analytical and mechanical facility for Raman measurements could be established near the drilling site.

6.5.1 DEA compared to PCA

PCA produces similar results compared to DEA. However, a model can also be built which introduces the mineralogical data with an ideal case (which has a high content of pyrite and arsenopyrite near the surface (in low depth)). The PCA-related algorithm cannot handle the ideal case, as it becomes numerically unstable for the reason that the ideal case is numerically of higher order compared to the rest of the data points. This is just one example how PCA-related algorithms need to be built on case-by-case basis to handle the different phenomena, whereas DEA (here output-oriented BCC model) has more robust
functioning for different numerical orders. It should also be noted that PCA model, including the pseudocode given in 3.3.1, finds an efficiency frontier for only one type of DEA model, namely for output-oriented BCC model.

The computing times of DEA and PCA models can be compared. The mineralogical data produces 1.4 sec for the DEA model and 0.7 sec for the PCA-related algorithm. By testing it was concluded that the PCA-related algorithm is always faster than linear programming that DEA uses. However, one can note that the time it takes to compute DEA or PCA model is very small (approximately 1 sec) compared to the time it takes to scan one sample (15 minutes). This shows how little difference the faster computation time makes in the time it takes to make the total analysis.

6.5.2 The DEA modelling

The Figure 1 in Section 1.2 shows the efficiency frontier formed by DEA modelling, when there is an ideal case included in the analysis. The differences are small compared to the indicator elements measured of the ten meters rock drill-core, and the results generally exhibit the same behaviour in the rock drillcores studied, compared to the combined Raman and DEA modelling.

6.6 Sustainability of the exploration phase of mining studied with DEA

The results using Charnes-Cooper-Rhodes (CCR) output oriented model are shown in Figure 11. These results reflect the fact that some exploration sites are more effective in the sense of sustainability. Half of the exploration sites receive a sustainability score of nearly one or exactly one. This relates to the fact that the number of exploration sites (9) is not high compared to the number of variables (6): there are many dimensions for an exploration sites to reach effectiveness (Kauppinen & Khajehzadeh 2015). The DEA modelling could be improved by increasing the exploration sites under observation. For example, the Site 1 receives a sustainability score of 1, even though it has a social input variable of 1. Explanation is that Site 1 also has a small surface disturbance (ecological input variable) and moderate investment (economic input variable) with plausible discounted future profit (economic output variable) and a high total salary (social output variable). This is a notion which explains the behaviour of efficiency scores in DEA modelling: the dimensions in which Site 1 is efficient are enough to counteract the one dimension in which site 1 is inefficient.
Results

Figure 11. Results achieved using artificial data (Kauppinen & Khajehzadeh 2015).

There should always be considerably more exploration sites included in the DEA analysis of sustainability than in this study with artificial data. This provides means to take into account the strengths and weaknesses of different exploration sites. However, there are already some differences to be seen. Site 4 receives a very low sustainability score (see Figure 11) because the values for economic, ecological and social inputs are high and respectively, the outputs are low. The future profit (economic output variable) is second lowest in the set of exploration sites studied, but investment (economic input variable) is the highest. Similar tendency is found for other input and output variables, making the Site 4 clearly the worst exploration site in this set of exploration sites studied.
This thesis focused on linear programming as a tool for exploration. The study was able to show how linear programming can be utilized in the analysis of rock drillcores, with Laser-Induced Fluorescence (LIF) and Raman spectroscopy; analysis of rock breakage; assisting geologists logging the drillcores; and sustainability measurement. All of these aspects can be covered with just a few methods, adding to the usefulness of the results.

These main contributions of the thesis can be listed as follows.

1. To use Non-negative least squares (NNLS) as a tool to solve for abundance alongside point-count method and to show that NNLS is more accurate in solving for concentration compared to the point-count method.

2. Second, NNLS was used to produce an industrial scale online scanning method utilizing laser-induced fluorescence (LIF) and Raman concurrently for rock drillcores.

3. The applicability of Data Envelopment Analysis (DEA) to the exploration phase of mining was shown on three different levels: how rock breakage can be studied with DEA; DEA can help geologists log rock drillcores; and to measure sustainability with DEA.

NNLS method provides a new tool for measuring abundance and concentration. Haskin et. al. (1997) proposed point-count method twenty years ago. This method is still reliable but it does not utilize the computation power of modern computers. NNLS on the other hand is shown in this thesis to include point-count method in the results, but further developing these results with help of more detailed data gathered of the ore samples.

LIF can be utilized to first scan an ore sample to find the important transition areas of minerals in the sample before using LIF imaging. Furthermore, Raman spectroscopy can then be used to identify the minerals appearing in colour under fluorescent light and therefore to calibrate LIF imaging technique. The identification with Raman uses the aforementioned results achieved using NNLS. After the identification, LIF imaging can be used to rapidly scan the surface of the ore samples, identifying the minerals therein.

DEA can be used to identify the main factors having an effect in the rock breakage. This is achieved by creating DEA model of all the possible combinations of different factors which can have an effect, and then choosing the model that best fits the data measured. Rock breakage can be analysed rather with DEA than
other linear models, because a certain type of DEA modelling can include also nonlinear factors into the model. The modelling for rock breakage with DEA can be fully automated procedure, reducing the time and resources needed to identify the main factors for certain type of rock.

The geologists logging drillcores face a challenging task of finding the drillcores where there are ore deposits. The assistance of combined Raman spectroscopy and DEA can help geologists to log drillcores more reliably and more effectively, by showing the depth at which the minerals occur. The process can be fully automated, but at this stage it should be noted that geological knowledge and a trained eye are still needed to verify the results achieved. It is of importance however, that auxiliary methods like Raman and DEA can help the geologist to arrive at the correct results.

Sustainability is paramount in the current societal discussion on mining. In this thesis it is shown that sustainability can be measured with easy metrics, not including a lot of paperwork and secondary variables. The exploration phase of mining was here reduced to six key variables which were compressed with DEA to one dimensional measure of sustainability. This does not mean that all three aspects of sustainability, namely the social, environmental or economic, would not have been included, but that state-of-the-art measurement of sustainability is easy and reliable way of showing the public the development in the field of sustainable mining.

As a conclusion, the study shows how effective linear programming methods like NNLS and DEA are when they are applied to exploration. The thesis also shows that industrial scale application for analysing rock samples can be built, where laser-based methods provide data to linear models. The linear programming can also be used to design reliable models for more sustainable decision making.
References


References


References


