Contents lists available at ScienceDirect

Computers & Geosciences

Research paper

Data Envelopment Analysis as a tool for the exploration phase of mining

CrossMark

Tommi Kauppinen

Aalto University School of Electrical Engineering, Department of Electrical Engineering and Automation, PO Box 15500, 00076 Aalto, Finland

A B S T R A C T

The exploration of mining has often been limited by time-consuming methods of analysis. This paper introduces Data Envelopment Analysis (DEA) as a new tool for the exploration phase of mining. DEA is a non-parametric method for data fusion, and it is used alongside with the on-site Raman analysis. Ten meters of halved rock drillcore from the Kittil mine (Suurikuusikko deposit) were pulverised and homogenised, thus ensuring that each meter had a representative sample. These 10 samples, one for each meter, were subsequently measured with a grid measurement (32×32 measurement each) using the Raman setup. All the data points were analysed using the point-count method. After identifying the frequency at which potentially valuable minerals appear in the samples, this information was analysed using DEA. The study ends by presenting an efficiency score for each meter of drillcore. These efficiency scores enable geologists to judge more rapidly which parts of the drillcore must be logged more carefully. In addition, Principal Component Analysis (PCA) is discussed as an alternative for producing similar results to DEA.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, expediting the exploration phase of mining has sparked considerable interest in the mining industry due to the fact that it is highly time-consuming process. One needs to cut through bedrock with a diamond drill, analyse the drillcores obtained, and then decide the feasibility of establishing a mine, as well as the exact mining location. The exploration phase depletes large amounts of both natural and economic resources in the phase of extraction of drillcores and their analysis. However, this is the sole way to acquire evidence of ore as well as estimate the amount available in the rock (Computers & Geosciences, 2011). Moreover, after the mine has been started, the mine planning includes drilling new holes to ascertain where the deposit is located.

Possessing accurate information on the contents of the drillcores is vital for the mining industry (Ma et al., 2010). The ore excavated is completely dependent on the analysis carried out on the drillcores. Time is also of the essence, even though mining occurs over a period of years. By improving the speed of the analysis of drillcore samples, geologists can more quickly provide accurate information to decision-makers.

This paper focuses on affordable but precise methods of analysing drillcores. These methods might speed up the process for the geologist from the phase of acquiring the drillcores from the

ground to the next phase of passing on their findings. The phases involved are: drilling, analysing and presenting the findings. This paper concentrates on the analysis part of the exploration. A new method is proposed to help geologists analyse the drillcore faster but still in a reliable manner.

Analysing the drillcore samples is a time-consuming process requiring one to two months to acquire the laboratory results. The methods to circumvent the laboratory analysis using on-site analysis are further described elsewhere (see Kauppinen et al. (2013, 2014a)) but here it might be noted that by using on-site analysis, the time required to produce reliable results is measured in hours, not months. One of the candidates for on-site analysis is the Raman analysis (see e.g. Ishikawa and Gulick (2013)), which is used in this paper to produce data from rock drillcores. In addition, geologists need to analyse the rock when it is extracted from the ground. This analysis is called logging. Logging is hard to quickly and reliably execute, hence the methodology presented in this paper. By the method introduced, geologists can produce results for the decision-makers from a high number of drillholes in a matter of days.

The object of research in this paper is to show that Data Envelopment Analysis (DEA) can be used to help geologists log the drillcores more efficiently (for mineral prospectivity mapping with DEA, see Hosseini and Abedi (2015)). The case studied is Kittilä mine in Finland, but the general application for drillcores is also discussed. DEA is a non-parametric efficiency measure. It is used here to fusion the spectral data received from Raman

E-mail address: tommi.kauppinen@aalto.fi

experiments into a single efficiency score available for the use of the geologists. Furthermore, the mathematics of DEA is discussed and its linkages to other data fusioning methods (see e.g. Abedi et al. (2012) and Chauhan et al. (2016)) are considered. This paper also shows that DEA can be considered to be a reliable tool even though it is by essence non-parametric.

2. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric method for evaluating Decision Making Units (DMUs). Typically, DEA is used to evaluate branches of a companies or medical facilities. The first contributions for DEA were made by Farrell (1957), who established the foundations for DEA by introducing the comparative efficiency of companies using a linear programming (LP) problem for input–output data set.

Ever since Farrell's discovery, the DEA method has grown in popularity. Charnes et al. (1978) published a paper in the field of operations research. In this and the following paper, Charnes et al. (1979), presented the Charnes–Cooper–Rhodes (CCR) model, which was and still is the first basic model for DEA. The CCR model evaluates the efficiency of the weighted sums of the input–output ratio. In addition, the efficiency scores of DMUs are given between [0,1] in CCR modelling, which is even today the basic premise for many DEA models. After the CCR model, Banker et al. (1984) introduced the Banker–Charnes–Cooper (BCC) model. In BCC modelling, variable returns to scale of DMUs define the efficiency frontier, while the CCR model assumes constant returns to scale.

The DEA method has grown in popularity in academic publications worldwide. Therefore, the scientific community has possibly already accepted or is in the process of accepting the DEA method as a valid tool for decision making. At present, this method includes many different models for different usages. For example, Cooper et al. (2006) describe 20–50 different models, while the exact number is related to the definition of a model. Of course, for the user of the DEA method, this poses a new problem of finding the appropriate model from a selection of models.

In this paper, output-oriented BCC modelling is used to ascertain the efficiency scores for ten (10) consecutive samples of rock drillcore. BCC modelling assumes varying returns to scale, meaning that even smaller quantities of ore can show a high efficiency score. Considering drillcore analysis, this is highly acceptable: the cases where there are multiple minerals in small quantities might also be a valuable area for geologists to study in more detail.

2.1. Introduction to mathematics of data envelopment analysis

Data Envelopment Analysis is a linear programming (LP) based technique for measuring the relative performance of data points (see e.g. Kauppinen and Khajehzadeh (2015)). The usual measure of efficiency, i.e.:

$$\text{efficiency} = \frac{\text{output}}{\text{input}} = \frac{u_1 y_{10} + \cdots + u_s y_{s0}}{v_1 x_{10} + \cdots + v_m x_{m0}} \tag{1}$$

is often inadequate due to existence on multiple inputs and outputs related to the data in question. In Eq. (1), u and v are weights for output and input variables y and x, respectively.

From Eq. (1), a fractional linear optimisation problem can be formulated, by giving limits to the weights u and v and for the efficiency as a whole. Then one has (for decision-making units (DMUs) $o = 1, ..., n$)

$$\max_{v,u} \frac{u_1 y_{1o} + \cdots + u_s y_{so}}{v_1 x_{1o} + \cdots + v_m x_{mo}}$$

s. t.

$$\frac{u_1 y_{1j} + \cdots + u_s y_{sj}}{v_1 x_{1j} + \cdots + v_m x_{mj}} \leq 1, j = 1, ..., n \tag{2}$$

$$u_1, ..., u_s \geq 0$$

$$v_1, ..., v_m \geq 0$$

Eq. (2) is a fractional problem, which can be modified into Linear Programming (LP) form by defining $v_1 x_{1o} + \cdots + v_m x_{mo} = 1$. The result in Eq. (3) is treated as a problem akin to a standard Linear Programming (LP) problem, also known as the input oriented CCR DEA model for DMUs $o = 1, ..., n$.

$$\max_{v,u} u_1 y_{1o} + \cdots + u_s y_{so}$$

s. t.

$$v_1 x_{1o} + \cdots + v_m x_{mo} = 1$$

$$(u_1 y_{1j} + \cdots + u_s y_{sj}) \leq (v_1 x_{1j} + \cdots + v_m x_{mj}), j = 1, ..., n \tag{3}$$

$$u_1, ..., u_s \geq 0$$

$$v_1, ..., v_m \geq 0$$

After developing the CCR model, yet another model can be introduced, known as the BCC model, which is used in this study for the actual computations. The multiplier form of input oriented BCC model can be developed from the Eq. (3) by adding a variable u_o to the cost function. u_o is free in sign and it is substracted from the cost function otherwise equal to the CCR model (see Eq. (4)).

Fig. 1 shows graphically how CCR and BCC efficiency frontiers differ form each other for a 200 data points. The data is random data, including one input and one output variable.

Graphically u_o marks the difference between the efficiency frontiers of CCR and BCC models (see Fig. 1), and therefore u_o varies for different DMUs o. The points on the efficiency frontier always have an efficiency of one. Furthermore, the rest of the points score a better efficiency scores as the distance from the efficiency frontier is somewhat reduced. Therefore, the BCC model gives higher efficiency scores for the data set under study
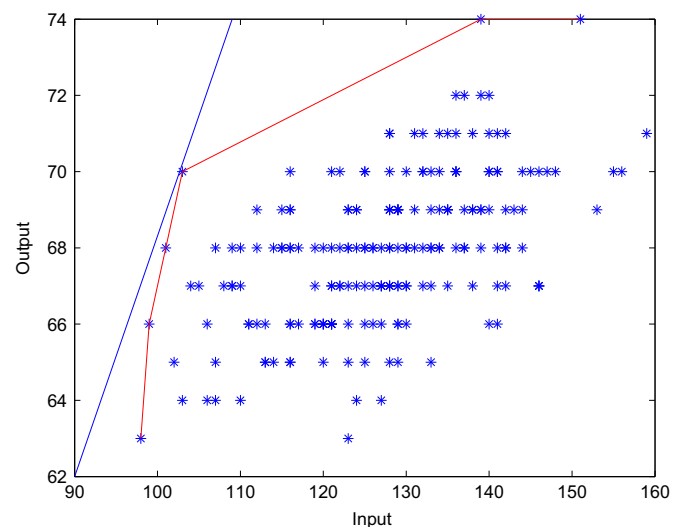

Fig. 1. Efficiency frontiers for both DEA CCR model (blue) and DEA BCC model (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\max_{v,u,u_o}\frac{u_1y_{1o} + \cdots + u_sy_{so} - u_o}{v_1x_{1o} + \cdots + v_mx_{mo}}$$

s. t.

$$\frac{u_1y_{1j} + \cdots + u_sy_{sj} - u_o}{v_1x_{1j} + \cdots + v_mx_{mj}} \le 1, j = 1, \ldots, n \tag{4}$$

$$u_1, \ldots, u_s \ge 0$$

$$v_1, \ldots, v_m \ge 0$$

The Eq. (4) is similar to fractional problem of Eq. (2), if one ignores the u_o variable. In the same way as before, the problem can be transformed into a LP-problem by assigning condition for the denominator to be equal to one (see Eq. (5))

$$\max_{v,u,u_0} u_1y_{1o} + \cdots + u_sy_{so} - u_o$$

s. t.

$$v_1x_{1o} + \cdots + v_mx_{mo} = 1$$

$$\left(u_1y_{1j} + \ldots + u_sy_{sj}\right) - u_o \le \left(v_1x_{1j} + \ldots + v_mx_{mj}\right), \quad j = 1, \ldots, n \tag{5}$$

$$u_1, \ldots, u_s \ge 0$$

$$v_1, \ldots, v_m \ge 0$$

The Eq. (5) is an input oriented model. The output-oriented model can be understood graphically in one input and one output case (see Fig. 1) 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, output oriented efficiency score of a DMU in the following manner: first, by taking the total vertical distance to the x-axis from a given point, and second, by dividing the result by the distance that a vertical line through the given point has to the efficiency frontier.

Eq. (6) introduces the output oriented BCC DEA model. The difference here is that the inputs are minimised whereas before the outputs were maximised. This change can most easily be understood graphically form Fig. 1, as the output-oriented model seeks to minimise the input variables of DMU o, creating a possibility to measure the actual efficiency along the y-axis as explained previously.

$$\min_{v,u,u_o} v_1x_{1o} + \cdots + v_sx_{so} - u_o$$

s. t.

$$u_1y_{1o} + \cdots + u_my_{mo} = 1$$

$$(v_1x_{1j} + \cdots + v_mx_{mj}) \ge (u_1y_{1j} + \cdots + u_sy_{sj}) + u_o, j = 1, \ldots, n \tag{6}$$

$$u_1, \ldots, u_s \ge 0$$

$$v_1, \ldots, v_m \ge 0$$

In the actual DEA program that is used for calculations in this study, a model known as envelopment form of DEA BCC model is used. This far the discussion has been limited only on the multiplier, or dual (see Matousek and Gärtner (2007)) form for the BCC model (Cooper et al., 2006, p. 91). It should be noted that mathematically, the envelopment form and multiplier form produce exactly the same results.

To develop an envelopment form of DEA BCC model from the above equations, an extra condition similar to u_o has to be given, namely $\sum_{j=1}^{n} \lambda_j = 1$, where λ is a column vector with all elements non-negative (see Eq. (7)). This is equivalent to the aforementioned difference by which the BCC differs from CCR model, imposing a convexity condition on allowable ways in which the observations for the n DMUs may be combined (Cooper et al., 2006,

p. 91).

In this study, the envelopment form of output oriented BCC model is used to produce efficiency scores for different parts of rock drillcore. For the envelopment form of BCC model with an output orientation (Cooper et al., 2006, p. 93), the linear programming problem can be given as follows

$$\max_{\eta_B,\lambda}(\eta_B)$$

s. t.

$$X\vec{\lambda} \le \vec{x}_0$$

$$\eta_B\vec{y}_0 - Y\vec{\lambda} \le 0$$

$$\vec{e}\,\vec{\lambda} = 1, \vec{\lambda} \ge 0, \tag{7}$$

In Eq. (7), η_B is a scalar, $X = (x_j) \in \mathbf{R}^{m \times n}$ and $Y = (y_j) \in \mathbf{R}^{s \times n}$ are the given input and output data sets, $\lambda \in \mathbf{R}^n$ and e is a row vector with all elements equal to 1. x_0 and y_0 stand for input and output values of DMU $o = 1, \ldots, n$, respectively. Therefore, there are n optimisation problems to be solved. It is also worth of note that x_0 and y_0 are vectors of length m and s, respectively. For example, if $X = (x_j) \in \mathbf{R}^{1 \times n}$, as is the case with the input data used in this paper, then x_0 is a scalar.

The optimisation process is as follows: the DMU 1 is chosen, the optimisation as in Eq. (7) is solved, and the efficiency score η_B is saved for DMU 1. This is then repeated for DMU 2 and so on, until also the DMU n has an efficiency score η_B. Effectively, n efficiency scores are reached, one for each DMU. These efficiency scores are optimal in the sense of Eq. (1), namely having as high value as the comparison with the other DMUs allow. For the DEA model solved in this paper for rock drillcores the following holds: $m = 1$, $s = 2$, $n = 11$.

2.2. 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. In the conclusion section the PCA and DEA methods are compared on a more general level.

A modelling scheme for PCA is introduced which generates the same efficiency frontier as the BCC DEA model. Reason for using PCA is to speed up the process of finding local maxima. One can perform PCA in practice as follows. 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) substracting the mean of each row, (3) the eigenvectors of the covariance can be calculated.

Eigenvectors can be solved for the symmetrical matrix AA^T from the equation

$$\underset{p \times n}{A}\ \underset{n \times p}{A^T} = EDE^T \tag{8}$$

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 AA^T (Shlens, 2005).

After calculating the eigenvectors, which can also be called the principal component axes, one may note 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 and one input data, see Fig. 2). This is done by finding the maximum value of all the points in to the direction of the axis of
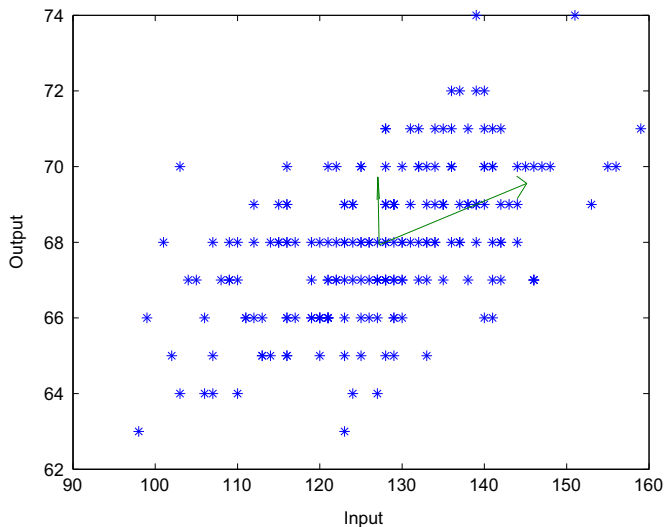
Fig. 2. PCA space for data with one input and one output (see the arrow pointing upwards for the second principal component).

the second principal component.

When the first maximum is found, it can be stored. After concluding whether the point dominates other points, the second highest value can be searched for. The term dominates means that the point found is higher on the output axis and lower on the input axis than some other point, and is therefore always more efficient. The described process of searching the local maximum in principal component space produces the DEA efficiency frontier.

The algorithm used to produce the BCC efficiency frontier using PCA-related functions as follows. The algorithm always picks up the highest value measured on the second axis of principal component space, stores it and after this ignores it in further calculation. If the maximum point found is higher on the output axis and lower on the input axis than some points, those points can also be ignored in further calculation, as they cannot be on the efficiency frontier. Then the algorithm again picks up the highest value and continues the iteration. After all the other points beside the last efficiency frontier point have been analysed, the program terminates, with the stored efficiency frontier values.

The points on the efficiency frontier always have the efficiency score of one. After receiving the efficiency frontier values, the algorithm continues by finding the individual efficiency scores of the other 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.

3. Raman measurements

Raman analysis was conducted on pulverised and homogenised rock drillcore samples gathered from Agnico-Eagle Kittilä Mine, Finland. Kittilä mine (Suurikuusikko deposit) is concentrated on gold mining. 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.

In total, 10,240 (ten thousand two hundred and forty) Raman spectra were analysed from Kittilä mine. The sample spectra were measured in a 32×32 grid on the surface of pulverised and homogenised samples.
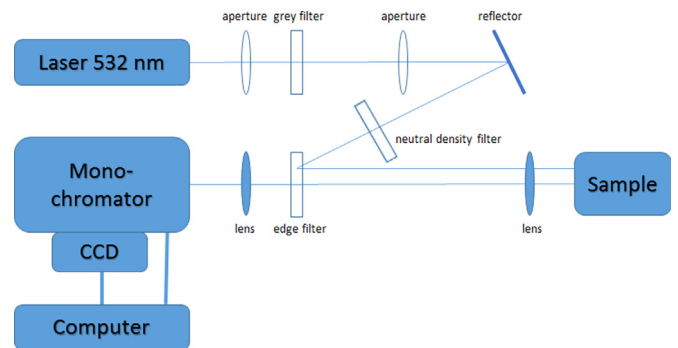

Fig. 3. Measurement apparatus for Raman measurements (see also Kauppinen et al. (2014b)).

3.1. Measurement apparatus

A laser operating at 532 nm and with power of 100 mW was used for Raman measurements (see Fig. 3). The measurements were performed by moving the sample in front of the measurement area of the spectrograph in a grid of 32 by 32 measurement points. The Raman measurements were performed with a home-built 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 lter (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 an CCD camera (Andor, Newton DU940N-BV).

3.2. Preparing the samples

First the samples were 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 (see Fig. 4) were pressed together for the Raman measurement. 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 0.64 cm × 0.64 cm in size. With 10 samples studied, the total

Fig. 4. Pulverised sample made into a pellet (on the left) and tools used to produce one (see also Kauppinen et al. (2014a)).

Raman spectra measured was 10,240. The speed of measurement was approximately one second per grid point, giving approximately seventeen minutes per meter of drillcore.

3.3. Data used in computation

The data from Raman spectroscopy was digitised as shown in Fig. 3, and loaded into computer memory for use in Matlab. Normalisation and baseline correction were performed before any other techniques applied to raw data. Baseline correction was done by an algorithm given by Eilers and Boelens (2005), to remove the baseline from the measured spectra. Normalisation was done to remove the mean and to scale each measured spectrum to unit variance. The data preparation is done with Matlab code, as is all the further analysis described herein.

3.4. Point-count method

Haskin et al. (1997) introduced a point-count method for Raman analysis. Point-count is a method of computing the number of spectra that represent a certain mineral. If in hundred measured spectra there are 10 that are classified as showing a clear peak on a wavelength of a certain mineral, then according to the point-count method it is safe to conclude that 10 percent of that sample has the given mineral.

Here the point-count method is used as follows: first, pyrite mineral is given a window of 350–375 nm, all the spectra showing the highest peak inside this window are classified as representing pyrite, and second, all the spectra showing the highest peak at wavelengths 330–350 nm are classified as arsenopyrite.

3.5. Preparing the data for DEA

DEA models have by definition both input and output variables. Here two output variables (pyrite and arsenopyrite content) are considered. These output variables are selected because most of the gold in Kittilä mine is present in pyrite or arsenopyrite, free gold being very rare. The only input variable taken into account is the depth from which the drillcore is excavated. This input variable is selected to take into account the effort needed to reach a given mineral deposit. It is fascinating however to understand the possibility of using more input variables, or for example the costs as input variables, eventhough in this paper 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.

4. Results

Table 1 shows the number of spectra that belong to either pyrite or arsenopyrite group (hereby called 'hits'). These hits can then be further analysed with DEA-BCC output-oriented modelling to arrive at efficiency scores. It should also be noted that an 'ideal' case with 500 hits for pyrite and 100 hits for arsenopyrite is inserted to give a baseline for DEA model. From Table 1 it can be seen that during the 10 m of drillcore studied there is a great variation in pyrite and arsenopyrite content.

4.1. DEA compared to PCA

As can be seen from Fig. 5, using the mineralogical data produces coinciding efficiency scores for both DEA BCC output oriented model and PCA-related algorithm.

A model can also be built which introduces the mineralogical data with an ideal case, having high content of pyrite and

Table 1
Drillcore samples measured with Raman from Kittilä mine: the number of arsenopyrite and pyrite spectra in the samples, with a combined maximum of 1024 for each meter.

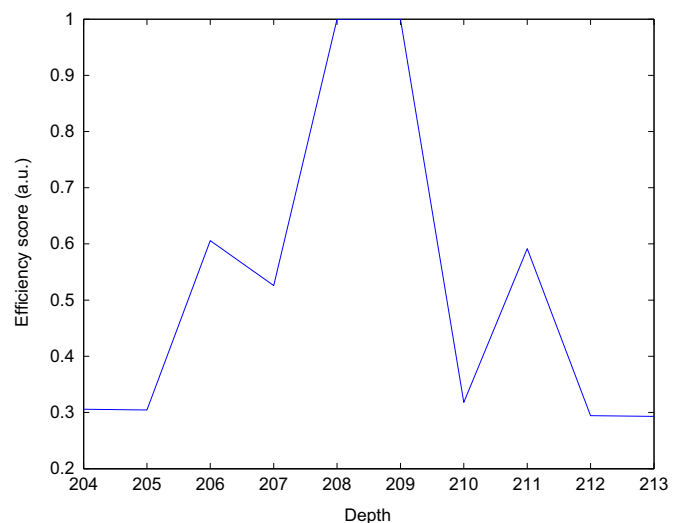| Drillcore	X (m=1) Depth (m)	Y (s=1) No. of arsenopyrite	Y (s=2) No. of pyrite
1	204	3	5
2	205	3	11
3	206	6	10
4	207	2	50
5	208	10	67
6	209	6	96
7	210	3	23
8	211	6	20
9	212	3	12
10	213	3	6
Average	208.5	4.5	30
Ideal	1	100	500

Fig. 5. Efficiency scores for both output-oriented BCC DEA and PCA-related algorithm.

arsenopyrite near the surface (low depth). However, the PCA-related algorithm cannot handle the ideal case, as it becomes numerically unstable. 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 (output-oriented BCC) is more robust in its handling of different problems. Fig. 6 shows the efficiency frontier formed by DEA modelling, when there is an ideal case (see Table 1) included in the analysis.

The computing times of different models can be compared. The mineralogical data produces 1.4 s for the DEA model and 0.7 s for the PCA-related algorithm. By testing it was concluded that the PCA-related algorithm is always faster than linear programming that DEA uses. To reduce the time difference in the favour of DEA, on could use significantly larger data. Moreover, if the multiplier form, or the dual (see Matousek and Gärtner (2007)) formulation of DEA is used, DEA algorithm should become comparatively faster with a large data set.

The Principle Component Analysis (PCA) modelling showed its computational strength against linear programming which the DEA method is based on. However, this is a subsidiary point in the studies where computational speed is only a minor factor affecting the results, as in the study of rock drillcore. It is more important to consider the DEA theory as a complete set of tools for different data, whereas PCA-related algorithms are more dependent on the model developed on a case-by-case basis. Therefore, the DEA can
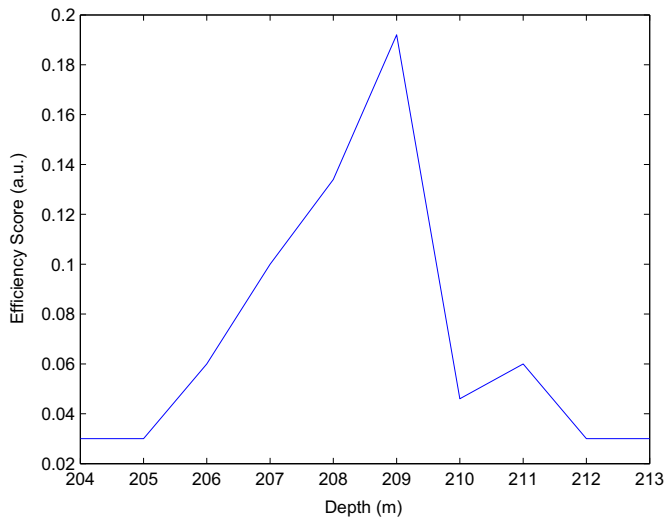
Fig. 6. Efficiency score for each meter of drillcore given by BCC modelling.

Fig. 8. The combined pyrite and arsenopyrite hits at each depth.

be seen as being slower, but more reliable a way of analysing efficiency, compared to PCA-related algorithms.

4.2. The DEA modelling

Figs. 6–8 present the main finding of this paper. One can see from the figures that they are somewhat similar. Fig. 6 shows the DEA efficiency scores of 10 m of drillcore, and Fig. 7 gives indicator elements, which are the current method for indicating whether there is pyrite or arsenopyrite present in the sample. For comparison, also the total number of hits for each depth is shown in Fig. 8.

It can be seen that the small efficiency scores do not differ from the findings made with indicator elements. The peak values, however, of pyrite and arsenopyrite are different in Figs. 6 and 8 compared to Fig. 7. In Figs. 6 and 8, the peak is at 209 mof depth, but in Fig. 7 the peak is at 208 m of depth. This can be because these samples actually have pyrite and/or arsenopyrite which the indicator elements do not show, or because the Raman analysis conducted contains some error in labelling minerals pyrite or arsenopyrite. However, most likely is that the sample at depth 209 m has a high efficiency score and high number of hits because it contains a lot of pyrite and only a little arsenopyrite. This gives higher efficiency than indicator elements (e.g. Arsenic, As) would dictate.
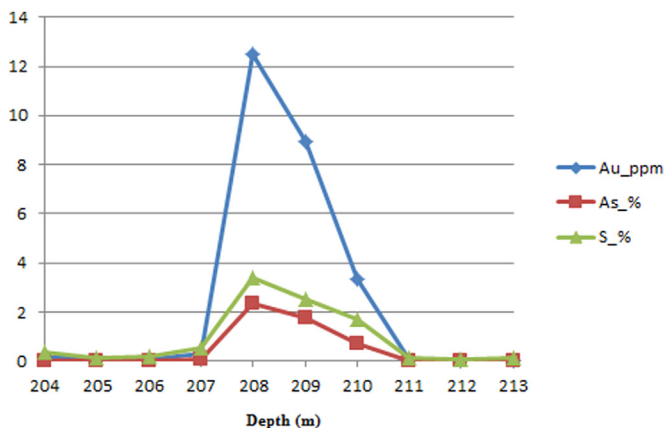
5. Conclusion

After having examined the results, there is little doubt that the Data Envelopment Analysis (DEA) is a proper tool for the exploration phase of mining. Although the example in this study was simple and non-complex, including depth, pyrite content and arsenopyrite content, it showed the potential of introducing multidimensional problems a two dimensional solution (Fig. 6). The number of outputs and inputs can be decided in different cases using geological knowledge of the mineralogy of the deposit. It should be of interest to geologists to introduce variables to DEA which reflect the mineralogy of the given mine.

5.1. DEA modelling of drillcores

The on-site Raman analysis gives more data, both more accurate and larger amount, than the study of indicator elements at the moment can. Also, using DEA brings greater confidence in the findings, and helps in logging as geologists can concentrate on the parts with a higher amount of valuable minerals. The results formulated with DEA can be trusted, and, as was shown, validated using other data analysis methods as PCA.

It was shown that the DEA method combined with the on-site Raman analysis can help geologists log the drillcore more precisely and faster. This is because the data provided combines all the important mineralogical aspects into one dimension (the efficiency score). By comparing the values in this dimension, geologists can make a faster decision of which drillcores to log more carefully. Moreover, the decision-making of the geologist utilising DEA method becomes easier, as the DEA can include many drillcores from different drillholes, comparing them to each other. Third, leading the exploration phase becomes easier with the fusioned data that DEA provides. The DEA results solved in this paper compared well with the current method of indicator elements.

Data Envelopment Analysis can be used as a new exploration tool for mining. In particular, the application of DEA in the logging of drillcores was found plausible in the case of undefined number of outputs and inputs. It is also worth to observe that almost any type of DEA model (here DEA BCC output oriented model) can potentially be used to produce meaningful efficiency scores. Furthermore, the output and input variables can be more varied than in the case presented in this paper. These points offer abundant possibilities for further research.

Fig. 7. The amount of trace elements for the 10 m of drillcore studied.

Acknowledgements

References

Abedi, M., Norouzi, G.H., Bahroudi, A., 2012. Support vector machine for multi-classification of mineral prospectivity areas. Comput. Geosci. 46, 272–283.

Banker, R., Charnes, R.F., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag. Sci. 30, 1078–1092.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. Eur. J. Oper. Res. 2, 429–444.

Charnes, A., Cooper, W.W., Rhodes, E., 1979. Short communication; measuring the efficiency of decision making units. Eur. J. Oper. Res. 3, 339.

Chauhan, S., Rühaak, W., Khan, F., Enzmann, F., Mielke, P., Kersten, M., Sass, I., 2016. Processing of rock core microtomography images: using seven different machine learning algorithms. Comput. Geosci. 86, 120–128.

Computers & Geosciences, 2011. Geocomputation of mineral exploration targets. Comput. Geosci. 37, 1907–1916 (editorial).

Cooper, W.W., Seiford, L.M., Tone, K., 2006. Introduction to Data Envelopment Analysis and its Uses: With DEA Solver Software and References. Springer, New York.

Eilers, P., Boelens, H., 2005. Baseline correction with a symmetric least squares smoothing.

Farrell, M.J., 1957. The measurement of productive efficiency. J. R. Stat. Soc. 120 (3), 253–290.

Haskin, L.A., Wang, A., Rockow, K.M., Jolliff, B.L., Korotev, R.L., Viskupic, K.M., 1997. Raman spectroscopy for mineral identification and quantification for in situ planetary surface analysis: a point count method. J. Geophys. Res. 102, 19293–19306.

Hosseini, S.A., Abedi, M., 2015. Data envelopment analysis: a knowledge-driven method for mineral prospectivity mapping. Comput. Geosci. 82, 111–119.

Ishikawa, S.T., Gulick, V.C., 2013. An automated mineral classifier using raman spectra. Comput. Geosci. 54, 259–268.

Kauppinen, T., Haavisto, O., Häkkänen, H., 2013. Optimisation algorithms in the case of mineral detection using raman analysis. In: Proceedings IFAC Symposium on Automation in Mining, Mineral and Metal Industries -Conference. Vol. 1, pp. 40–44.

Kauppinen, T., Haavisto, O., Khajehzadeh, N., Häkkänen, H., 2014a. Raman analysis: detection of minerals and measurement of concentration using reference spectra. In: Proceedings of IMPC2014-Conference. Santiago, Chile.

Kauppinen, T., Khajehzadeh, N., 2015. Sustainability in the exploration phase of mining: a data envelopment analysis approach. In: Proceedings of IFAC Symposium on Automation in Mining, Mineral and Metal Industries -Conference.

Kauppinen, T., Khajehzadeh, N., Haavisto, O., 2014b. Laser-induced fluorescence images and raman spectroscopy studies on rapid scanning of rock drillcore samples. Int. J. Miner. Process. 132 (10), 26–33.

Ma, X., Wu, C., Carranza, Emmanuel, J.M., Schetselaar, Ernst, M., van der Meer, Freek, D., Liu, G., Wang, X., Zhang, X., 2010. Development of a controlled vocabulary for semantic interoperability of mineral exploration geodata for mining projects. Comput. Geosci. Vol. 36, pp. 1512–1522.

Matousek, J., Gärtner, B., 2007. Understanding and Using Linear Programming. Springer, Berlin Heidelberg New York.

Shlens, J., 2005. A Tutorial on Principal Component Analysis. (Online). Available: ⟨http://www.cs.cmu.edu/elaw/papers/pca.pdf⟩.