

# Big Data on Decision Making in Energetic Management of Copper Mining

C. Lagos, R. Carrasco, G. Fuertes, S. Gutiérrez, I. Soto, M. Vargas

## Carolina Lagos\*

Facultad de Administración y Economía,  
Universidad de Santiago de Chile  
Av. Libertador Bernardo O'Higgins 3363,  
Santiago, Chile  
\*Corresponding author: carolina.lagos@usach.cl

## Raúl Carrasco

1. Departamento de Ingeniería Eléctrica,  
Universidad de Santiago de Chile  
Av. Ecuador 3519, Santiago, Chile  
2. Departamento de Matemáticas y Ciencias de  
la Computación, Universidad de Santiago de  
Chile  
Las Sophoras 173, Santiago, Chile  
raul.carrasco.a@usach.cl

## Guillermo Fuertes

Departamento de Ingeniería Industrial,  
Universidad de Santiago de Chile  
Av. Ecuador 3769, Santiago, Chile  
guillermo.fuertes@usach.cl

## Sabastián Gutiérrez

1. Facultad de Ciencias Económicas y  
Administrativas,  
Universidad Central de Chile  
Lord Cochrane 417, Santiago, Chile  
2. Facultad de Ingeniería, Universidad Andres  
Bello  
Antonio Varas 840, Providencia, Santiago, Chile  
sebastian.gutierrez@ucentral.cl

## Ismael Soto

Departamento de Ingeniería Eléctrica,  
Universidad de Santiago de Chile  
Av. Ecuador 3519, Santiago, Chile  
ismael.soto@usach.cl

## Manuel Vargas

Facultad de Ingeniería, Universidad Andres  
Bello,  
Santiago, Chile  
manuel.vargas@unab.cl

**Abstract:** It is proposed an analysis of the related variables with the energetic consumption in the process of concentrate of copper; specifically ball mills and SAG. The methodology considers the analysis of great volumes of data, which allows to identify the variables of interest (tonnage, temperature and power) to reach to an improvement plan in the energetic efficiency. The correct processing of the great volumen of data, previous imputation to the null data, not informed and out of range, coming from the milling process of copper, a decision support systems integrated, it allows to obtain clear and on line information for the decision making. As results it is establish that exist correlation between the energetic consumption of the Ball and SAG Mills, regarding the East, West temperature and winding. Nevertheless, it is not observed correlation between the energetic consumption of the Ball Mills and the SAG Mills, regarding to the tonnages of feed of SAG Mill. In consequence, From the experimental design, a similarity of behavior between two groups of different mills was determined in lines process . In addition, it was determined that there is a difference in energy consumption between the mills of the same group. This approach modifies the method presented in [1].<sup>a</sup>

**Keywords:** Copper mining, energetic efficiency, big data, process management.

<sup>a</sup>Reprinted (partial) and extended, with permission based on License Number 3962080057504 © [2016] IEEE, from "Computers Communications and Control (ICCCC), 2016 6th International Conference on".

## 1 Introduction

Market conditions, global competence and care of environment have created the urgent need to improve the energetic efficiency. So, efficient management of energetic resources for big industrial customers shaping up a critical aspect. It is important for industries to economize in the

use of energy and make it sustainable in long terms. It has emerged a significant market and a technological opportunity for the tools and technologies that allow an efficient management of energetic resources.

On the other hand, the quality and velocity of information interchange, requires a cooperation commitment among all the participants in the collecting process. Among the reasons for a non-efficient flux of information are: (1) unfulfilling relations among the different levels that receive and process the information, (2) lack of preventive maintenance programs in the measurement equipment and (3) intrinsic factors of the process.

Data science and its predictive analysis its getting more and more important for the production of goods and services [2], with applications from data monitoring for the energetic consumption [3], system for the prediction of maintenance [4, 5]. Decision Support Systems (DSS) can represent a solution, when human intervention is necessary [6–8], by using of techniques in the improvement of safety [9], and also, in the prediction for the analysis of natural resources consumption [10] and in models of optimization of processes [11]. This document discusses techniques and tools by consumption of energy data in mining, and the possible contribution that can do in the data analysis, for the decision making in the energetic efficiency. Filip (2008) concludes that “the development and application of intelligent DSS can help enterprise cope with problems and uncertainty and complexity, to increase efficiency and competitiveness in production networks” [6].

The optimization and the control of the minerals processing plant, such as copper, has generate new research lines. With the increasing availability of data basis coming from the copper productive process, the amount of data gathered at an exponential rhythm. This creates and opportunity for the mining companies to analyse and optimized their operations observing those data. Then as productive operations become bigger, with more variables, it is crucial for the direction and the area managers, to exploit the information in an efficient and timely way. To make this happen, it is important the synchronization among all the areas, through joint planning of all the areas participating in the data collection, and shared information [12]. Other related works related with copper mining can be found in reference [13–15].

A previous preliminary work can be found in reference [1], that Data analysis methods related to energetic consumption in copper mining.

The main contribution of this paper is the generation of a systematic model for the analyses of the data and the use of statistical tools, allowing improvement of the energetic efficiency.

This organization of the paper is: section 2, describes the problem of the sales of the mineral, knowing the energetic costs of the process and the evolution on time of the sales; section 3, details the methodology and the different focuses of big data; section 4, shows the design and analysis of results of the processing of big data, coming from the concentration process of copper, for the energetic efficiency and final conclusions of the research.

## 2 Description of the problem

With the purposes of improving the decision making, data mining can be considered a tool of quantitative analysis, that still needs more exploration for the energetic management. In this sense, Data mining and knowledge discovery in data basis have been attracting a significant amount of research [16], particularly to be used in the mining processes for its high data level [17].

The energetic costs hav been a relevant factor in the structure of costs of the mining processes. It is so, that the energy expenditure (electricity and fuel) it is the order of 15% of the operational cost (2008-2011), which increase affected negatively the operational margins, since a maximum of 65% in 2006, up to a 22% in 2014 [18]. So for a better understanding of the costs of

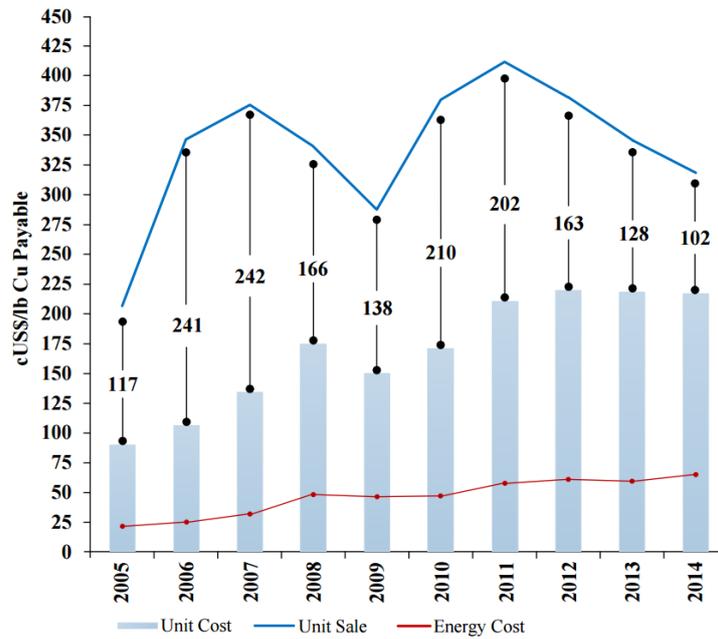


Figure 1: Costs and unit sales in the copper mining in Chile between 2005 and 2014

the process, it is required to study more deeply the behavior of the management of the energetic efficiency, with the intention of founding the optimum energetic consume [19].

On the other hand, analyzing the unit costs of exploitation (dollar cents per pund of payable copper), it is observed that costs have increased, but not in the same direction than sales, which is appreciated comparing costs and unit sales in Figure 1. Besides the same graphic shows the decrease to a little more than the half of the average operational process, from 210 cUS\$/lb for the year 2010 up to 102 cUS\$/lb in 2014 [18].

The previous graphic, also accounts for the decrease in the operational margins that have experienced the mining companies in the last years, due to the increase in costs and less copper price.

### 3 Problem formulation

Mining production companies as Codelco, have been collecting and storing increasingly more data of their productive processes, which offer an enormous potential as source of new knowledge [20], given the complexity of the data and the mining process, this document is used to data analysis [21, 22]. The data mining for the analysis of processes efforts to establish correlations between the key indicators a data mining for analysis of process strives to establish correlations between indicators of performance and the consume in the process. If the correlations can establish with the enough precision, it is possible to predict the behavior of energy, given the current consumption and the attributes of the process. In this study, were used several techniques:

1. Markov Chain Monte Carlo

This paper uses the Markov chain Monte Carlo (MCMC) algorithms for imputing data, are a widely used and well-studied methodology that can be used to draw samples from posterior distributions [23]. The sampling is efficient for estimating parameters [24].

2. Canonical correlation analysis

Canonical correlation analysis (CCA) is a known multivariate analysis method, the objective is to find and quantify the correlations between two sets of multidimensional variables [25]. In this study the CCA, it is used as a right tool to establish the correlation between the energetic consumption as a dependant variable, and the independant variables of the Ball Mills and SAG Mill. CCA can be used recursively depending on the amount of dependant and independant variables that can be related, forming possible variables and canonical correlations. [26, 27], uses CCA in industrial processes to perform the troubleshooting. On the other hand [28], proposes a model of a recursive and adaptative process, to improve the precision of variables of the process in time.

### 3. Design and Analysis of Experiments

The design and analysis of experiments are classical statistic models whose objective is to find out if some determined factors influence in a variable of interest, if exists influence of any factor and quantify the effect it has on it. The use of the design and analysis of experiments allow us to measure the effect of the mil as a factor on the energetic consume that is the variable of interest of this work, achieving quantify the effect it has the mil factor on the energetic consumption [29]. In this sense it is used in different problems of energy [23, 30, 31].

## 4 Design and analysis of results

The analysis of the unit of the article is the production process of copper in the biggest state company of Chile, in the Division Codelco Chuquicamata. The data analysis, is done through a three phases procedure: Imputation of data for the missing information; Study to correlate between the variables that influence in the energetic consumption through a multivariable model of Canonical Correlation; and a comparison between means of consumption of the mills to verify potential significant differences between them.

### 4.1 Variables of study

The study is performed on six mills that operate in the process of the concentrator: Mill of Ball 16A (MOBO 16A); Mill of Ball 16B (MOBO 16B); Mill of Ball 17A (MOBO 17A); Mill of Ball 17B (MOBO 17B), Mill SAG 16 (SAG 16) and Mill SAG 17 (SAG 17), as shown in Figure 2. The defined variables for the analysis of the energetic consumption are divided in two groups. In the first one, defined as mills MOBO 16A, MOBO 16B and SAG 16 and the second group is defined as: MOBO 17A y MOBO 17B and SAG 17, for both cases, the variable of study are: Consumed energy, West Temperature, East Temperature, Winding Temperature and the Tonnage, according with the defined variables with the Maintenance Management of Codelco Chuquicamata.

### 4.2 Previous Analysis

In Fig. 3 it is presented the potential of, MOBO 16B. It is observed periods without information and the presence of atypical data in the last three years, which should affect the analysis to look for solutions to the proposed problem.

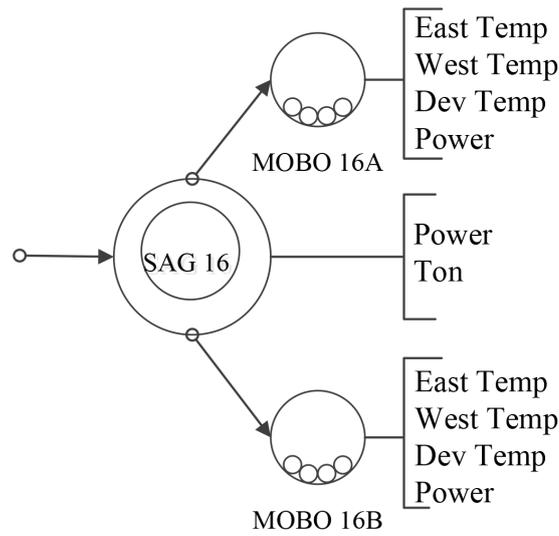


Figure 2: Measure Variables in the SAG mill and the mills of Ball MOBO 16A, MOBO 16B, in the concentrate process

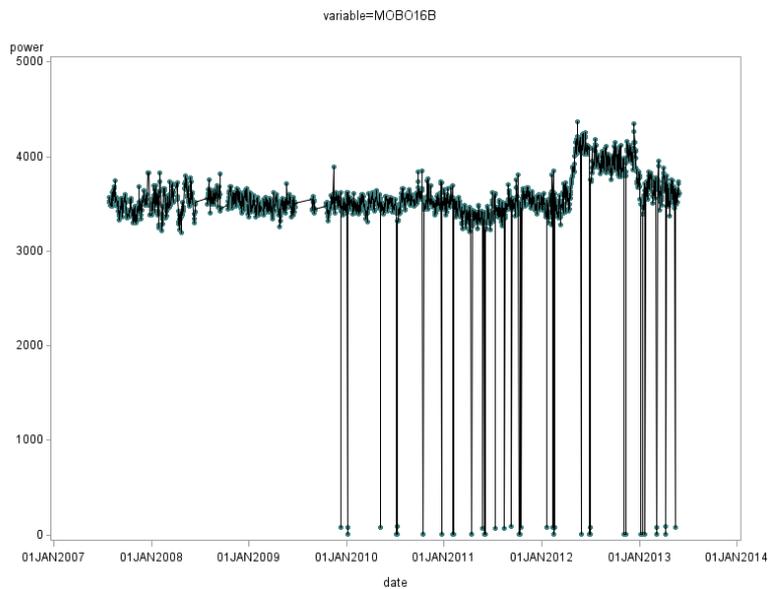


Figure 3: Power of MOBO 16B

As example in Fig. 4 it is shown the values of the power variable of the MOBO 16B. In this graphic are appreciated the outlier values that will be correct thanks to the MCMC procedure. Fig. 5 shows the dispersal among variables (power, tonnage, east temperature, west temperature and winding temperature), at all times, it is observed that data do not distribute normally since the Kernel distribution does not adjust to the normal distribution.

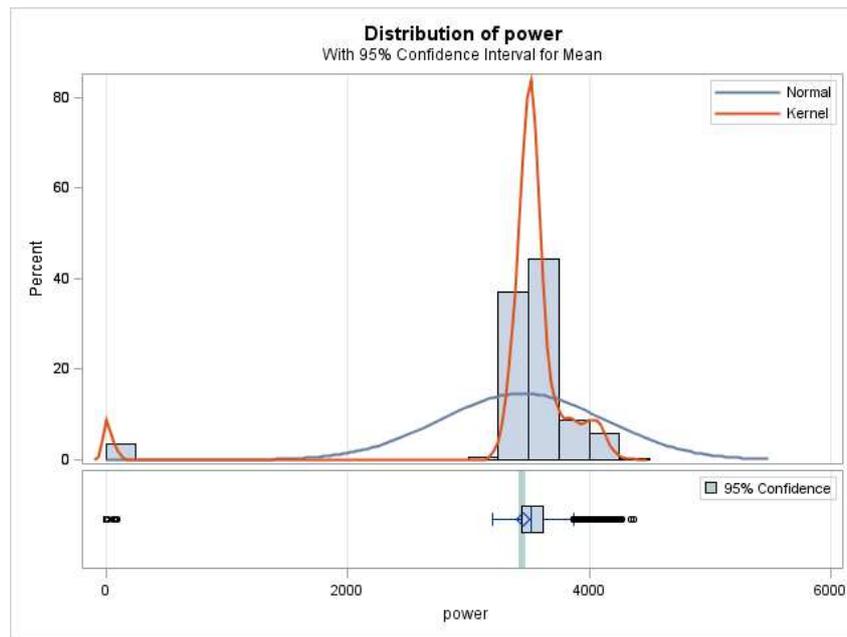


Figure 4: Distribution of the power variable in MOBO 16B

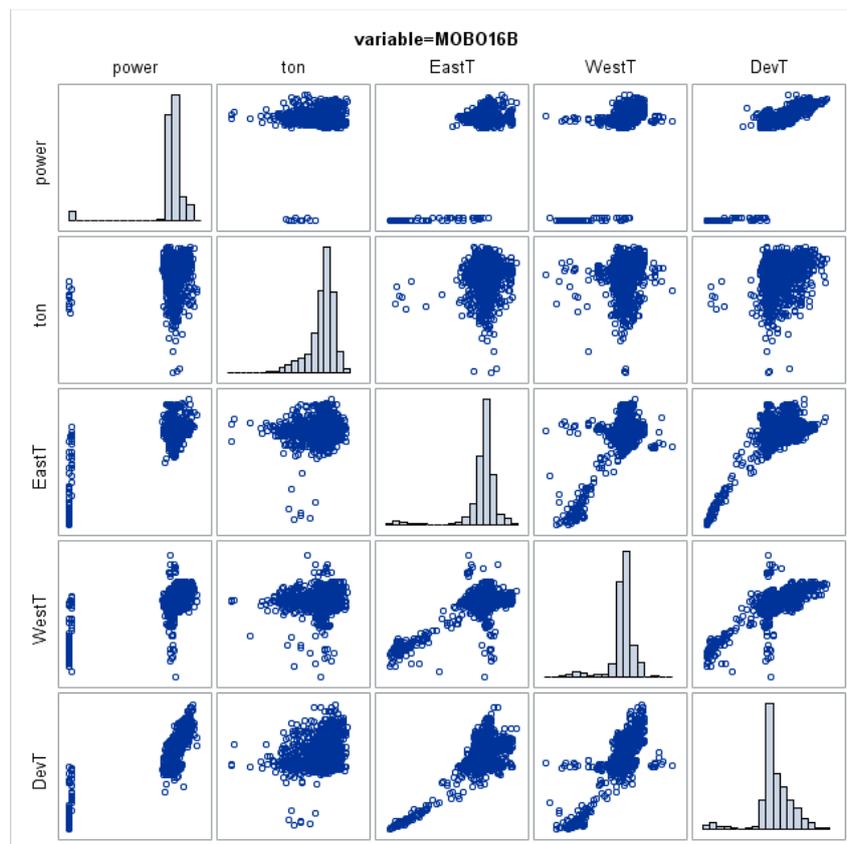


Figure 5: Plot of the correlations of the variables of the MOBO 16B

According to Table 1, a 10.77% of observations possess at least one missing data, so, the multiple imputation is used to recover the lost information and correct the atypical data.

Table 1: Group of MOBO 16B data, existing and non existing

Group	East T.	West T.	Wind T.	Power	Frequency	Percentage
1	X	X	X	X	1905	89.23%
2	X	X	X	.	60	2.81%
3	X	X	.	X	1	0.05%
4	X	.	X	X	1	0.05%
5	X	.	.	X	3	0.14%
6	.	X	X	X	1	0.05%
7	.	X	.	X	1	0.05%
8	.	.	.	X	1	0.05%
9	.	.	.	.	162	7.59%

### 4.3 Analysis of imputation

In the case of the previous analysis and of imputation, it is presented the case of the mill of the MOBO 16B, while in the canonical correlation, it is worked with the group of mills of sector 16 (of Ball and SAG); and for the experimental design. For the analysis it is used the software SAS v9.3.

Though it does not exist all the information, and being a great number of atypical data in the period of three years, it is proceed to the imputation, not only for the problem of the missing data, but also "to correct" these atypical data (a 14% of the observations that have at least one missing data). For the imputation it must be deliver, before being used, initial values for the mean vector and the variance and covariance matrix to be able to estimate the coefficients. The matrixes are obtained from the bootstrap procedure, as it is shown shown on table 2:

Table 2: Descriptive Statistical of the variables in study

Variable	EastT.	WestT.	DevT.	Power
Average	47.076154	35.790675	73.20456	3395.297468

Likewise, it is presented the existing correlation between the independant variables or explanatory disposed in a double entry matrix, as it is shown on table 3:

Table 3: Matrix of correlation

Variable	EastT.	WestT.	DevT.	Power
EastT.	102.449985	19.919678	41.713204	3271.145603
WestT.	19.919678	81.730596	51.081713	2880.435635
DevT.	41.713204	51.081713	236.074203	9127.414414
Power	3271.145603	2880.435635	9127.414414	570276

While the procedure is repeated, new mean, variance and covariance matrixes are generated, to estimate the coefficients that constitute the process, this is repeated until the difference between parameters of an iteration with the next one being almost null.

### 4.4 Test Hypothesis

Below are shown on table 4, the coefficients obtained by the MCMC method, with its corresponding confidence intervals.

Table 4: Test Statistical

Variable	Coefficient	Est. Error	Confidence Interval (5%)	Value $t$	$p$ -value
EastT.	47.104421	0.22795	[46.657; 47.552]	206.64	< 0.0001
WestT.	35.764552	0.204824	[35.362; 36.167]	174.61	< 0.0001
DevT.	73.163543	0.346796	[72.482;73.482]	210.97	< 0.0001
Power	3394.130311	16.935236	[3360.873; 3427.873]	200.42	< 0.0001

It can be observed that each of these values is significant, what can be verified with the hypothesis test  $H_0 : \beta_i = 0$  vs  $H_1 : \beta_i \neq 0$ , and the test statistical  $t$ -student, which decision criteria is to reject the hypothesis if  $p$ -value < 0.05, or if the coefficient value is within the confidence interval. In this case both are true, thus it demonstrate that this procedure is quite efficient and reliable to generate these new data basis.

#### 4.5 Application of the Canonical Correlation

The use of the canonical correlation, allows not only to search a relation from variable to variable, but also the, between one of them and a set of them, as well as between two complete groups. Below it is shown the equation of correlation for the mills 16 (Ball A and B, and SAG), structured in variables and abbreviated as it is seen in parenthesis: Power of MOBO 16A (P16A), MOBO 16B (P16B), SAG 16 (P16S), East Temperatures, MOBO 16A (TE16A), West (TO16A), Winding (TD16A) and of MOBO 16B (TE16B, TO16B y TD16A), and of the Tonnage (TON16S), then:

$$P16A + P16B + P16S = TE16A + TO16A + TD16A + TE16B + TO16B + TD16B + TON16S$$

So, the process implies to create as much canonical correlations as it is needed to achieve the minimum of present variables between the dependant and the independant (min(# Var. Dependents, # Var. Independants)), that in this case corresponds to 3. Table 5, shows the data obtained through this procedure.

Table 5: Canonical Correlation

	Canonical Correlation	Own value	Proportion	Accumulated
1	0.861034	2.8667	0.9239	0.9239
2	0.435197	0.2336	0.0753	0.9992
3	0.051116	0.0026	0.0008	1.0000

In this case, the canonical correlations are given for each group created inside the program, explains the relation between both groups of variables, that is to say, between the dependants and the independant, where it can be remarked that the first results in having a quite high value in comparison with the others, but to rectify how many groups must be considered, it can be performed the test of Bartlett-Lawley, which results can be appreciated below on Table 6:

Table 6: Statistical of the Bartlett-Lawley Test

	Aprox. F Value.	GL	$p$ -value
1	210.57	21	< 0.0001
2	39.74	12	< 0.0001
3	1.11	5	0.3505

This test, for the first line, works on the null hypothesis of  $H_0 : \rho_0 > \rho_1 = \rho_2 = \rho_3 = 0$  (with

$\rho_0 = 1$ ), while the second line works on  $H_0 : \rho_1 > \rho_2 = \rho_3$ , and so on. Thus, it can be observed that the two first hypothesis are rejected, analyzing the test by means of p-value  $< 0.05$ , but it is not the same for the last one, since it does not exist evidence to reject it. This implies that the two first correlations result significant and reliable to be taken in consideration in the study. The canonical correlation allows to modelate the information of the mills, with coefficient. The canonical correlation model allows for each of the variables within each group; to these made-up groups, are nominated as dependant and independant variables.

On one hand, the SAS program, gives two coefficients for each variable in each group, ones called raw coefficients, and the others standardized. The first ones correspond to the real coefficients and, at the same time, common of each variable, that is to say, are interpreted as it were a linear regression, this is, for each unit that increases some value, the response will increase / decrease in  $\beta_i$  units. By the other hand the standardized coefficients are given only with the purpose of facilitating its interpretation, and to establish also a comparison between variables. Avoiding differences between magnitudes or intervals these have.

Table 7: Raw Coefficients and dependant variables

Raw Coefficients	Dependant 1	Dependant 2	Dependant 3
Power MOBO 16A	0.000420051	0.002660012	-0.000485457
Power MOBO 16B	0.0011697651	0.0011781961	-0.000018758
Power SAG 16	-0.049858651	-0.059131666	0.88859719798

Table 8: Raw Coefficients and independant variables

Raw Coefficients	Dependant 1	Dependant 2	Dependant 3
Ton. SAG 16	-0.000117482	-0.000269535	-0.001644198
East Temp MOBO 16A	0.0160939058	0.0454415296	0.0501536296
West Temp MOBO 16A	0.0129238742	0.0075762637	0.0630824701
Dev Temp MOBO 16A	0.0106182145	-0.077979648	-0.005456832
East Temp MOBO 16B	0.0146624833	0.0514678025	-0.004535215
West Temp MOBO 16B	0.0065554753	0.0514678025	-0.034625719
Dev Temp MOBO 16B	0.0467922347	0.0392654956	0.0003705063

Table 9: Standardized Coefficients and dependant variables

Standardized Coefficients	Dependant 1	Dependant 2	Dependant 3
Power MOBO 16A	0.1931	-1.2226	-0.2231
Power MOBO 16B	0.8838	0.8902	-0.0142
Power SAG 16	-0.0577	-0.0685	1.0259

Table 10: Standardized Coefficients and independant variables

Standardized Coefficients	Dependant 1	Dependant 2	Dependant 3
Ton. SAG 16	-0.0368	-0.0845	-0.5155
East Temp MOBO 16A	0.1005	0.2837	0.3131
West Temp MOBO 16A	0.1562	0.0916	0.7626
Dev Temp MOBO 16A	0.1963	-1.4415	-0.1009
East Temp MOBO 16B	0.1481	0.4138	-0.0458
West Temp MOBO 16B	0.0595	0.4668	-0.3140
Dev Temp MOBO 16B	0.7170	0.6017	0.0057

In tables 7 to 10, it can be found important differences between the variables and the mills, with which it can be aroused the equations obtained in the analysis, considering that one of the groups (equation), it is discarded by the results obtained from the test of Bartlett-Lawley. Below are presented two equations, of which it is chosen the second one that represents the greater correlation among the powers of the mills, temperature and tonnage.

Equation 1:

$$\begin{aligned}
& 0.000420051 * P16A + 0.0011697651 * P16B - 0.049858651 \\
& * P16S = 0.0160939058 * TE16A + 0.0129238742 * TO16A \\
& + 0.01061821145 * TD16A + 0.146624833 * TE16B \\
& + 0.0065554753 * TO16B + 0.0467922347 * TD16B \\
& - 0.000117482 * TON16S
\end{aligned}$$

Equation 2:

$$\begin{aligned}
& - 0.002660012 * P16A + 0.0011781961 * P16B \\
& - 0.059131666 * P16S = 0.0454415296 * TE16A \\
& + 0.0075762637 * TO16A - 0.077979648 \\
& * TD16A + 0.0409671048 * TE16B + 0.0514678025 \\
& * TO16B + 0.0392654956 * TD16B - 0.000269535 * TON16S
\end{aligned}$$

#### 4.6 Analysis of the design of experiments

For this case, it is worked with all the mills, both SAG and the ball 16 and 17, but only with the powers, since mills will be measured as factors. The design of the experiment allows to measure the effect of the mill as factor on the total energetic consumption, this to find some difference among the processes, that are expected to be totally independant among the mills.

In the above ANOVA table 11, it is considered the effect of the process, that is to say, it is worked with experimental designs nested, where the mills of Ball A, B and SAG 16, belong to the first group of the process, and the remaining to the second group. Thus, it is concluded that it does not exist difference between mills 16 and 17 in the behaviour, but it exists in the consumption between the mills belonging to a same process (all this concludes starting from the test of the Fisher, by means of its  $p$ -value, and considering a significance of  $\alpha = 0.05$ ). Due the process in study, it starts in the SAG mills, the energy consumption of the ball mills, it should require less consumption. However, there is no difference in the behaviour of the groups of ball mills 16 and SAG 16, respect of the ball mills 17 and SAG 17, which is consistent with the electric energy consumption for both groups of mills.

Table 11: ANOVA the model

FV	GL	SC	MC	Valor F	p-value
Model	5	33306519304	6661303861	23023.6	< 0.0001
Error	12804	3704525732	289326		
Total	12809	37011045036			

In conclusion, if there is a difference in the process variable or group, as observed in Table 12.

Table 12: paired sample MOBO 16

Variable	Means	MOBO 16A v/s MOBO 16B	stat <i>t</i>	<i>p</i> -value
Power	90.2936	A>B	6.89	< 0.0001
EastT.	-7.1599	A<B	-34.63	< 0.0001
WestT.	2.3912	A>B	7.60	< 0.0001
DevT.	18.6154	A>B	59.00	< 0.0001

#### 4.7 Complementary Analysis

The procedure performed with the canonical correlations, does not result to be very different than the factorial analysis. These correlations are given by SAS, and are shown below on Table 13:

Table 13: Correlations and dependant variables

Correlation	Dependant 1	Dependant 2	Dependant 3
Power MOBO 16A	0.7096	-0.7046	-0.0071
Power MOBO 16B	0.9874	0.1440	-0.0652
Power SAG 16	0.1680	-0.1513	0.9741

It is observed a strong correlation of the powers of the mills Ball A and Ball B respect the first group of mills, so it exists a relation between the energies that are consumed in both mills, from this it can be rescue that the work performed in the process inside mill A, it has certain relation with the one done by mill B, what can be due to the type of mineral and/or the tonnage that is given from the SAG mill, which is divided among the ball mills.

Below are given on Table 14, the correlations of the independant variables in relation with the set of variables of temperatures and tonnage.

Table 14: Correlations and independant variables

Correlations	Independant 1	Independant 2	Independant 3
Ton. SAG 16	0.0775	-0.1130	-0.5006
East Temp MOBO 16A	0.6693	-0.1055	0.2723
West Temp MOBO 16A	0.1045	0.0018	0.7730
Dev Temp MOBO 16A	0.8109	-0.5198	-0.0089
East Temp MOBO 16B	0.4764	0.3272	0.1090
West Temp MOBO 16B	0.4805	0.2050	-0.2313
Dev Temp MOBO 16B	0.9219	0.0663	-0.2332

It is observed that a great part of the variables show a strong correlation with the group 1, excepting the tonnage of the mill SAG 16 and the temperature of the west engine of mill of ball 16 A. The fact is that a great part of the factors that act in the milling process (since this group of independant variables are strongly related with group 1 of the dependant variables), have a strong impact over the energetic consumption of the three mills of the process 16.

All the variables considered in this set are from mill of ball A and B, which was expected due to the fact that the first set of dependant variables is strongly related with the power of this two mills.

Other important set to remark (despite of being discarded by the Bartlett-Lawley test) is the third set of variables, since this only considers the energy consumed by the mill SAG 16 for the set of dependant variables. For the tonnage, it only works with the correlations of variables belonging to the mill SAG 16.

The above mentioned (for the first and third set of variables analyzed) can be corroborated with the following values given, where are shown the dependant variables related to the sets of independant variables and viceversa, and are shown below on Table 15, in which are found similar results to the previous presented, with the exception of the third set of variables where it is demonstrated why it was discarded by the Bartlett-Lawley test.

Table 15: Correlations and independant variables

Correlations	Independant 1	Independant 2	Independant 3
Power MOBO 16A	0.6110	-0.3066	-0.0004
Power MOBO 16B	0.8502	0.0627	0.0033
Power SAG 16	0.1446	-0.0658	0.0498

As it was expected, the powers of the mills Ball A and Ball B, have a strong relation with the set of independant variables of the first group, and thus the same independant variables present a strong correlation with the first set of dependant variables, giving in both cases, very similar values to the previous, and are shown below on Table 16.

Table 16: Correlations and dependant variables

Correlations	Dependant 1	Dependant 2	Dependant 3
Ton. SAG 16	0.0667	-0.0492	-0.0256
East Temp MOBO 16A	0.5763	-0.0459	0.0139
West Temp MOBO 16A	0.0900	0.0008	0.0395
Dev Temp MOBO 16A	0.6983	-0.2262	-0.0005
East Temp MOBO 16B	0.4102	0.1424	0.0056
West Temp MOBO 16B	0.4137	0.0892	-0.0118
Dev Temp MOBO 16B	0.7938	0.0289	-0.0119

It is concluded from the above table, that there is a strong relation between the temperatures (except the west engine of the mill of Ball 16A) with the energetic consumption of the mills, mainly in the Mill of Ball 16A and Mill of Ball 16B. Thus, it can be worked without problems for future analysis, discarding those that do not possess any relation with them, since these explain the behaviour of consumption in this machines together.

## Conclusions

Despite awareness of the wide potential of the energetic efficiency, there are few mining companies, that make decisions based on the analysis of data coming from their productive processes, due to obstacles such as the complexity of the process and the lack of clear information. A database that integrates models for the energetic efficiency, it is a great input due to the non existence of them in the industrial exploitation of copper. The great volume of data coming from the copper concentrate process, the ball mills and the SAG mills, must be extracted, filtrated and transformed, for the variables related with the energetic consumption. Big data processing, allows to identify the variables of interest (tonnage, temperature and power), and the relation that these variables have with the other variables of the process. For the analysis of the data it is used the statistical tools, arriving to a model of improvement in the energetic efficiency.

The main contribution of this paper, is obtain a model to improve energy efficiency, through the integrate DSS and the data analysis using different statistical tools such as Markov Chain Monte Carlo, Canonical correlation analysis and design and analysis of experiments. It is expected that the results are applicable in different areas of the production process for large-scale mining.

There is an urgency in the creation and implementation of models to improve the energy efficiency in the mining sector, due to the restriction, environmental challenges and energy cost.

## Acknowledgment

The authors acknowledge the financial support of the "Center for Multidisciplinary Research on Signal Processing" (CON-ICYT/ACT1120 Project), the USACH/DICYT 061413SG Project, "Continuous visible light communication Access point for man to machine interaction and detection in high risk mining environments", CORFO code 14IDL2-29919 and "Cooperation agreement between the Universidad de Santiago de Chile and Corporación Nacional del Cobre de Chile, División Chuquicamata", and the support given by Professor Francisco Torres Aviles (R.I.P).

## Bibliography

- [1] C. Lagos, F. Cordova, S. Gutierrez, G. Fuertes, and R. Carrasco, Data analysis methods related to energetic consumption in copper mining . A test case in Codelco, *Computers Communications and Control (ICCCC), 2016 6th International Conference on*, IEEE Xplore, e-ISSN 978-1-5090-1735-5, doi: 10.1109/ICCCC.2016.7496768, 1241–247, 2016.
- [2] J. A. Aloysius, H. Hoehle, S. Goodarzi, and V. Venkatesh (2016), Big data initiatives in retail environments: Linking service process perceptions to shopping outcomes, *Annals of Operations Research*, 1–27.
- [3] G. Zucker, J. Malinao, U. Habib, T. Leber, A. Preisler, F. Judex (2014), Improving energy efficiency of buildings using data mining technologies, *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, 2664–2669.
- [4] T. Mokfi, M. Almaenejad, and M. M. Sedighi, A Data Mining Based Algorithm to Enhance Maintenance Management: A Medical Equipment Case Study, *2011 First International Conference on Informatics and Computational Intelligence*, 74–80, IEEE, dec 2011.
- [5] P. Bastos, I. d. S. Lopes, and L. Pires (2012), A maintenance prediction system using data mining techniques, *World Congress on Engineering 2012*, 1448–1453.

- 
- [6] F. Filip (2008); Decision support and control for large-scale complex systems, *Annual Reviews in Control*, 32(1): 61–70, 2008.
- [7] A. Kaklauskas (2015), *Biometric and Intelligent Decision Making Support*, vol. 81, Springer International Publishing AG, 2015.
- [8] M. M. Polycarpou and Y. Ohta (2007), Nonlinear Fault Diagnosis of Dynamical Systems: An Intelligent Control Framework, *IFAC Proceedings Volumes*, 40(9):1-1.
- [9] L. Martín, L. Baena, L. Garach, G. López, and J. de Oña (2014), Using data mining techniques to road safety improvement in Spanish roads, *Procedia-Social and Behavioral Sciences*, 160:607–614.
- [10] M. A. Khan, M. Z. Islam, and M. Hafeez (2012), Evaluating the performance of several data mining methods for predicting irrigation water requirement, *Proceedings of the Tenth Australasian Data Mining Conference*, Australian Computer Society, 134:199–207.
- [11] C. Gröger, F. Niedermann, and B. Mitschang (2012), Data mining-driven manufacturing process optimization, *Proceedings of the World Congress on Engineering*, 3:4–6.
- [12] N. Altintas and M. Trick (2014), A data mining approach to forecast behavior, *Annals of Operations Research*, 216:3–22.
- [13] R. Carrasco, M. Vargas, M. Alfaro, I. Soto, and G. Fuertes (2015); Copper Metal Price Using Chaotic Time Series Forecasting, *IEEE Latin America Transactions*, 13(6):1961–1965,
- [14] F. Seguel, R. Carrasco, M. Alfaro, P. Adasme, and I. Soto, A Meta-heuristic Approach for Copper Price Forecasting, *Information and Knowledge Management in Complex Systems*, 449: 156–165, 2015.
- [15] C. Lagos, G. Fuertes, R. Carrasco, S. Gutierrez, M. Vargas, and R. Rodrigues, Facing the data analysis complexity for the energetic efficiency management at great copper, *IEEE ICA/ACCA 2016*, (Curicó), 1–6, 2016.
- [16] U. Fayyad, G. Piatesky-Shapiro, and P. Smyth (1996), From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, v17(3):37-46.
- [17] J. Ding, Q. Chen, T. Chai, H. Wang, and C.-Y. Su (2009), Data mining based feedback regulation in operation of hematite ore mineral processing plant, *2009 American Control Conference, IEEE*, 907–912.
- [18] D. Rojas, E. Castillo, and J. Cantalloppts (2015), *Caracterización de los costos de la gran minería del cobre*, Tech. Rep., Comisión Chilena del Cobre, 2015.
- [19] D. Hodouin, S.-L. Jämsä-Jounela, M. Carvalho, and L. Bergh (2001), State of the art and challenges in mineral processing control, *Control Engineering Practice*, 9(9):995–1005.
- [20] Codelco, Memoria Anual 2015, tech. rep., Codelco, Santiago of Chile, 2015.
- [21] Y. Y. Wen, W. M. Huang, J. Wu, Y. Chen, J. Q. Song (2013), Water consumption analysis system based on data mining, *Applied Mechanics and Materials*, 241:1093–1097.
- [22] Y. Liu, J. Zhao, and Z. Wang (2015), Identifying determinants of urban water use using data mining approach, *Urban Water Journal*, 12(8):618–630.

- 
- [23] R. C. Leme et al. (2014); Design of experiments applied to environmental variables analysis in electricity utilities efficiency: The Brazilian case, *Energy Economics*, 45:111–119.
- [24] Y. Liu and C. Li (2016), Complex-valued Bayesian parameter estimation via Markov chain Monte Carlo, *Information Sciences*, 326:334–349.
- [25] X. Xing, K. Wang, T. Yan, Z. Lu (2016), Complete canonical correlation analysis with application to multi-view gait recognition, *Pattern Recognition*, 50: 107–117.
- [26] Z. Chen, K. Zhang, S. X. Ding, Y. A. Shardt, and Z. Hu (2016), Improved canonical correlation analysis-based fault detection methods for industrial processes, *Journal of Process Control*, 41:26–34.
- [27] Z. Chen, S. X. Ding, K. Zhang, Z. Li, and Z. Hu (2016), Canonical correlation analysis-based fault detection methods with application to alumina evaporation process, *Control Engineering Practice*, 46:51–58.
- [28] L. Shang, J. Liu, K. Turksoy, Q. Min Shao, A. Cinar (2015), Stable recursive canonical variate state space modeling for time-varying processes, *Control Engineering Practice*, 36: 113–119.
- [29] D. C. Montgomery (2013); *Design and Analysis of Experiments Eighth Edition*, John Wiley & Sons, Inc., 8th ed., 2013.
- [30] J. Tang, G. Gong, H. Su, F. Wu, and C. Herman (2016); Performance evaluation of a novel method of frost prevention and retardation for air source heat pumps using the orthogonal experiment design method, *Applied Energy*, 169:696–708.
- [31] G. S. dos Reis et al., (2016); The use of design of experiments for the evaluation of the production of surface rich activated carbon from sewage sludge via microwave and conventional pyrolysis, *Applied Thermal Engineering*, 93: 590–597.