Water Management in Iron Ore Mining: Regression Models for Optimizing Water Use in Mining Complexes

Renata Andries[1], Renato Silva Júnior[2], Guilherme Alves[3]

[1]*Vale Technological Institute, Botafogo Beach, 186, Rio de Janeiro, RJ, Brazil, renata.andries@vale.com, ORCID 0009-0005-2002-512X*

[2]*Vale Technological Institute, Boaventura da Silva Street, 955, Belém, PA, Brazil, renato.silva.junior@itv.org, ORCID 0000-0001-8875-6299*

[3]*Botafogo Beach, 186, Rio de Janeiro, RJ, Brazil, guilherme.alves@vale.com, ORCID 0009-0008-3638-4655*

Abstract

Mining plays a crucial role in economic development by providing raw materials that drive social progress. In this context, water is a transversal and indispensable element at all stages. However, with the increasing demand for water and climate change, efficiency in the use of water resources has become a priority. Therefore, it is essential to develop tools that enable effective water resource management in mining, promoting cost reduction, mitigating water risks, and meeting environmental and social requirements, as well as creating competitive advantages for the sector through transparency and attracting investments.

The innovative approach of this work lies in the development of regression models to analyze the relationship between iron ore production, mineral beneficiation method, rainfall seasonality, and water demand. Using data from six Brazilian mining complexes over 89 months, the research seeks to establish correlations that can guide strategic decisions and increase water efficiency. Additionally, it is important to highlight that there are few studies in the literature that quantify water use in relation to production, differentiating Operational Water, Total Intake, and Reuse.

The main findings indicate that simple linear regression (SLR) is more effective for analyzing Operational Water and Reuse, while generalized Poisson linear regression (GPLR) presents lower errors for Total Intake. The research also reveals that both Total Intake and Reuse have lower correlation with the production variable. This is because total intake is more related to the volume of dewatering water than to the ore processing itself. In other words, most of the dewatering volume is returned to the environment without use and is not part of the operational water computation. Regarding the reuse portion, it is mainly related to units with dams, i.e., units with wet beneficiation. Furthermore, it occurs to a greater extent when there is robust water infrastructure capable of treating/reusing non-new water volumes.

The applications of this work are vast, including the planning and analysis of water use in mining enterprises. The developed models can be used as tools to increase transparency, attract investments, and create competitive advantages. The implications include promoting sustainable practices, reducing operational costs, and mitigating environmental impacts, contributing to water security and the sustainability of the mining sector.

Keywords: Water, water resources, sustainability, regression models, mining

Introduction

Mining is essential for economic and social development, with water being a critical input throughout all mining phases. In 2023, Brazil exported 378.5 million tons of iron ore, valued at US$30.5 billion, and consumed an average of 0.305 m^3 of new water per ton of ROM (IBRAM 2023, 2024). Consequently, it is estimated that the sector utilized approximately 115.4 million m^3 of new water in its production processes that year.

However, freshwater consumption limits are rapidly approaching (GERTEN *et al.*, 2013) or the available volume may have already been exceeded (Grafton *et al.* 2013; Rosa *et al.* 2019). The increasing water demand is driven by population and economic growth across various sectors, including industry, agriculture, livestock, energy, and mining. The World Resources Institute reports that a quarter of the global population lives in countries with extreme water stress, with over 1 billion people projected to face this situation by 2050.

Considering the reliance on water in iron ore mining, the rising water demand over the years, and the increasing frequency of global water scarcity events, it is crucial to thoroughly understand the operational water balance and actual water demand in mining projects. This understanding enables the prioritization of sustainable sources, the selection of water management strategies, the proper planning of necessary water infrastructure, and the enhancement of freshwater use efficiency, thereby preparing the sector for potential shortages or changes in supply conditions.

Water estimates for iron ore processing are limited in the literature, particularly when accounting for various components of the operational water balance, such as water capture for drawdown, freshwater, and reused water. Additionally, beneficiation methods (dry, wet or natural moisture processing) must be differentiated when calculating the project's overall water demand, as each method has distinct water requirements and infrastructure, influencing the feasibility of water reuse.

In the book "Perspectives and Advances in Water Resources Management in Mining" (ANA and IBRAM, 2024) the average specific water use per ton produced for different products was calculated by dividing the annual volume declared in grant applications by the annual ore production. However, this granted volume does not account for water capture needed for drawdown or the water used in the production process. Additionally, this indicator overlooks reused water, which constitutes most of the water consumption in an iron ore plant (approximately 80%) and treats all iron ore mining methods (natural moisture and dry processing) as similar.

NORTHEY *et al.*, (2019) analyzed 359 public mining reports for various minerals and found water withdrawals ranging from 0.13 m^3 to 17.29 m^3 per ton of processed ore. This research supports the assertion that it is not feasible to group or model a range of water use for different production processes. Furthermore, the indicator used in the research refers to total water withdrawal, not the new water used in operations. Therefore, in the case of iron ore, the water quantification presented by the author is more related to the water demand for drawdown than the operational use of new water.

The objective of this work is to develop a tool to quantify the water demand required for an iron ore mining project, from extraction to the final product. This tool will consider the total water collected, water used in the production process, and reused water. Linear regression models will be employed to compare water demand relative to the volume of raw ore processed (ROM) for different mineral processing methods. The following sections present the theoretical framework, methods, results, and conclusions necessary for understanding and executing the project.

Methods

The methodological steps of this study are as follows: 1) Data collection, including the definition of samples and variables; 2) Statistical treatment and analysis; 3) Correlation analysis; 4) Development of linear regression models (both simple and generalized); and 5) Model diagnostics. The following sections provide detailed descriptions of each step.

Database, variables and sample

The primary data utilized in this study encompass operational water monitoring and mining reports from six Brazilian iron ore mining complexes over a period of 89 months (January 2017 to May 2024). The analyzed complexes include Serra Sul (PA/BR), Serra Leste (PA/BR), Serra Norte (PA/BR), Itabira (MG/BR), Vargem Grande (MG/BR), and Paraopeba (MG/BR), resulting in a total of 2,670 observations. The variables incorporated into the linear regression model are Operational water, Total water withdrawn, Reused water, Plant feed, and Type of mineral processing.

Statistical treatment and analysis

During the analyzed period, operations experienced several events, including production shutdowns, changes in the production process, alterations in data consolidation personnel, and variations in water volume estimates. These factors increase the likelihood of data variability due to the absence of a verification or standardization process. Therefore, it is crucial to analyze and process the data, identifying and removing potential outliers to enhance data reliability before utilizing them in the proposed models.

Outliers were detected using the Interquartile Range (IQR) rule, which employs values estimated by regression methods (Jeong *et al.* 2017) to define acceptance or rejection limits for measured values. The normality of the data will be assessed using the Shapiro-Wilk test to determine whether parametric or non-parametric hypothesis tests should be applied. Visualizations such as boxplots, histograms, and scatter plots will be generated to examine the data set.

Correlation analysis

To understand the correlation between the variables in the model, Pearson's correlation coefficient will be employed. This coefficient is a bivariate measure of the association (strength) of the relationship between two variables (Paranhos *et al.* 2014). It ranges from -1 to 1, where the sign indicates the direction (positive or negative) of the relationship, and the value indicates the strength of the relationship. A perfect correlation (-1 or 1) signifies that the value of one variable can be precisely determined by knowing the value of the other (Elian 1988; Paranhos *et al.* 2014). Conversely, a correlation of zero indicates no linear relationship between the variables.

Simple linear regression model (SLR)

According to *(Elian 1988)*, linear regression is a global method and is based on the use of only one equation to explain the relationship between the variables studied (dependent and independent). The simple linear regression model (SLR) expected for the present study is given by the equation below.

$$y_i = \beta_0 + \beta_1 \,.\, X_1 + \varepsilon_i$$

Where y_i is the i-th value of the response variable, β_0 e β_1 are the parameters (regression coefficients), X_1 is the i-th value of the predictor variable and ε_i is the random error term.

Generalized linear model (GLM)

When aiming to associate a dependent variable with independent variables, linear modeling is commonly employed. However, a limitation of linear models is that the dependent variable must follow a normal distribution Akaike (1974). Therefore, it is necessary to seek an alternative method to satisfactorily associate the dependent and independent variables.

According to Dobson (2001), the generalized linear model (GLM) allows for the adjustment of regression models for univariate response data that follow a distribution from the exponential family. The exponential family includes distributions such as normal, binomial, Poisson, geometric, negative binomial, exponential, gamma, and inverse normal.

The generalized linear model with Poisson distribution is given by the equation below:

$$\ln(y_i) = \beta_0 + \beta_1 \,.\, X_1 + \beta_2 \,.\, X_2 + \dots + \beta_n \,.\, X_n + \varepsilon_i$$

Where y_i is the i-th value of the response variable, β_0, β_1 e β_n are the parameters (regression coefficients), X_1, X_2 e X_n are known constants and ε_i is the random error term.

Table 1 *Descriptive statistics of the variables Operational water, Total water withdrawn and Reused water (m^3) and Production unit (t).*

Complex	Variable	Mean	Median	σ	Minimum	Maximum
Itabira Complex	Operational water (m^3)	782,288	706,551	389,185	176,184	2,134,528
	Water Withdrawn (m^3)	1,239,970	1,177,568	418,597	607,152	2,610,282
	Reuse water (m^3)	5,248,472	5,121,209	742,584	3,942,365	6,855,672
	Production unit (t)	4,194,529	4,209,416	470,050	3,056,534	5,270,583
Paraopeba Complex	Operational water (m^3)	484,760	509,734	185,883	194,557	1,009,908
	Water Withdrawn (m^3)	1,243,860	1,202,686	263,204	356,785	1,859,243
	Reuse water (m^3)	1,154,011	790,026	921,491	162,860	2,916,920
	Production unit (t)	1,694,792	1,414,756	897,347	385,308	3,442,221
Vargem Complex Grande	Operational water (m^3)	860,612	877,598	359,375	82,480	1,602,300
	Water Withdrawn (m^3)	1,815,025	1,761,688	266,792	1,403,991	2,662,421
	Reuse water (m^3)	4,447,241	1,589,233	5,415,777	0	18,295,521
	Production unit (t)	3,306,432	3,261,708	1,454,959	125,848	6,023,899
Serra Leste Complex	Operational water (m^3)	18,834	18,054	8,159	4,494	38,574
	Water Withdrawn (m^3)	19,654	20,664	8,791	4,494	40,645
	Reuse water (m^3)	0	0	0	0	0
	Production unit (t)	429,488	410,497	95,215	73,559	573,849
Serra Norte Complex	Operational water (m^3)	645,536	606,621	271,821	168,415	1,221,619
	Water Withdrawn (m^3)	1,563,351	1,544,904	581,663	386,026	5,185,788
	Reuse water (m^3)	2,856,345	2,841,263	818,965	1,127,990	6,475,531
	Production unit (t)	9,687,420	9,646,836	2,458,428	4,669,465	13,838,155
Serra Sul Complex	Operational water (m^3)	86,213	82,216	27,926	34,392	151,884
	Water Withdrawn (m^3)	580,961	369,284	569,485	34,392	1,795,160
	Reuse water (m^3)	5,513	2,123	7,717	0	34,665
	Production unit (t)	6,038,252	6,447,528	2,098,740	934,899	9,643,820

Model diagnosis

The selection of the optimal regression model will be based on statistical criteria, including the coefficient of determination (R^2), standard error (σ), root mean-square deviation (RMSD), and Akaike information criterion (AIC). RMSD measures the difference between the values predicted by a model and the observed values, calculated as the square root of the mean of the squared errors. The AIC method, proposed by Akaike (1974), addresses model identification from the perspective of statistical decision theory, facilitating the selection of the most appropriate loss function for model adjustment.

Results

Statistical treatment and analysis

Statistical analysis of the data was conducted by calculating the means, medians, standard deviations, minimums, and maximums (Table 1). Comparing the standard deviation (σ) with the mean of the variables (Operational water, Total water withdrawn and Reused water) reveals significant variability within the complexes, as well as for plant feed. This variability can be attributed to the dynamic nature of operations, which experience fluctuations in production due to market demand, climatic seasonality, process maturity, changes in production routes, and other factors.

The Shapiro Wilk test was used to analyze the normality of the data. Most of the variables in the complexes do not follow a normal distribution. Although a non-normal distribution of most of the variables was observed, it is possible to perform a regression model with the data from the 6 complexes to analyze the relationship between the variables and the Production Unit. Linear regression assumes that the residuals (errors) of the model follow a normal distribution, not necessarily the independent or dependent variables. Furthermore, if the residuals do not follow a normal distribution, transformations in the variables (such as logarithm or square root) can be considered to improve the normality of the residuals.

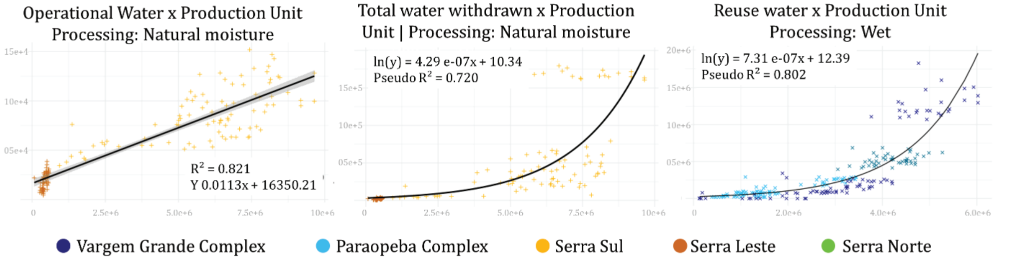

Figure 1 *Model behavior: 1) SLR Operational Water processing at natural moisture, 2) GLM Total water withdrawn at natural moisture and 3) GLM Reuse water at wet processing.*

Table 2 *Summary of analysis of SLR and Poisson GLM. Gray color indicates the best-fitting model.*

Operational Water (m^3) x Production Unit (t)										
Proc.	Model	β0	σ (β0)	β1	σ (β1)	R^2	σ (Res.)	p-value	AIC	RMSD
H	SLR	-1.01E+05	8.60E+04	7.70E-02	8.61E-03	0.485	1.96E+05	< 2e-16	1.87E+03	2.13E+04
	GLM	1.28E+01	2.03E-04	2.24E-07	5.59E-11	0.410	2.33E+07	< 2e-16	2.33E+07	2.61E+03
NM	SLR	1.64E+04	2.08E+03	1.13E-02	4.29E-04	0.821	1.69E+04	< 2e-16	2.49E+03	1.38E+03
	GLM	1.00E+01	7.37E-04	2.04E-07	1.16E-10	0.790	9.41E+05	< 2e-16	9.43E+05	1.65E+03
W	SLR	2.59E+05	3.99E+04	1.53E-01	1.24E-02	0.407	2.74E+05	< 2e-16	4.79E+03	1.84E+04
	GLM	1.22E+01	5.93E-04	1.21E-07	5.56E-11	0.490	4.98E+06	< 2e-16	4.98E+06	1.53E+03
Total capture (m^3) x Production Unit (t)										
Proc.	Model	β0	σ (β0)	β1	σ (β1)	R^2	σ (Res.)	p-value	AIC	RMSD
H	SLR	4.19E+05	2.22E+05	1.18E-01	2.22E-02	0.249	5.07E+05	8.64E-07	2.03E+03	5.50E+04
	GLM	1.40E+01	1.25E-04	5.64E-08	3.75E-11	0.090	2.39E+07	<2e-16	2.39E+07	2.33E+04
NM	SLR	-7.34E+04	4.45E+04	1.14E-01	9.17E-03	0.505	3.62E+05	< 2e-16	3.41E+03	2.95E+04
	GLM	1.03E+01	4.96E-04	4.29E-07	6.77E-11	0.720	2.48E+07	<2e-16	2.49E+07	7.26E+02
W	SLR	1.23E+06	5.76E+04	8.27E-02	1.79E-02	0.088	3.95E+05	6.70E-06	4.79E+03	1.84E+04
	GLM	1.35E+01	3.68E-04	7.58E-08	3.53E-11	0.280	1.17E+07	<2e-16	1.17E+07	4.41E+02
Reuse (m^3) x Production Unit (t)										
Proc.	Model	β0	σ (β0)	β1	σ (β1)	R^2	σ (Res.)	p-value	AIC	RMSD
H	SLR	3.59E+06	3.52E+05	-7.55E-02	3.52E-02	0.051	8.02E+05	3.48E-02	2.11E+03	8.70E+04
	GLM	1.24E+01	1.37E-04	7.31E-07	3.12E-11	0.802	1.67E+08	<2e-16	1.67E+08	2.33E+04
NM	SLR	-1.25E+03	6.35E+02	1.21E-03	1.31E-04	0.362	5.17E+03	<2e-16	2.13E+03	4.21E+02
	GLM	4.54E+00	6.88E-03	5.83E-07	8.92E-10	0.620	5.11E+05	<2e-16	5.11E+05	7.26E+02
W	SLR	-2.61E+06	3.49E+05	2.07E+00	1.09E-01	0.623	2.39E+06	<2e-16	5.74E+03	1.61E+05
	GLM	1.51E+01	2.56E-04	-2.65E-08	2.60E-11	0.050	1.87E+07	<2e-16	1.87E+07	2.33E+02

Correlation analysis

The linear correlation matrix was generated by calculating Pearson's correlation coefficients between the dependent and independent variables used in the study. A stronger correlation was observed among the variables associated with the natural moisture beneficiation type. This can be explained by the fact that, in this type of beneficiation, the primary use of water is for ore processing itself. Generally, in natural moisture processing, water is used more for particulate control than for ore processing. In other words, water use is more related to the mining area, roads, and piles that need to be sprayed.

Regarding operational water, an R^2 greater than 0.6 was obtained when comparing the production unit, suggesting that this variable can be used as a dependent variable in a model explaining the use of operational water in relation to the plant's feed. However, the same correlation is not observed for total water withdrawal or reuse. This is because total water withdrawal is more related to the volume of drawdown water than to ore processing itself. In other words, most of the water withdrawn is returned to the environment without being used and is not included in the calculation of operational water.

The potential for water reuse is mainly associated with units that have dams, i.e., units with wet processing. Additionally, reuse is more prevalent when there is a robust water infrastructure capable of treating and reusing non-new water volumes.

Simple Linear Regression Model (SLR) and Generalized Linear Regression Model (GLM)

Below is the graphical representation of the models with the highest R^2 values (Fig. 1), a comprehensive presentation of the results (Table 2), and a summary of the results analysis.

- Operating Water | Hybrid (H): The SLR model is considered the best due to its higher R^2 and lower AIC, despite the GLR model having a lower RMSD.
- Operating Water | Natural Moisture (NM): The SLR model is the best due to its higher R^2, lower AIC, and lower RMSD.
- Operating Water | Wet (W): The SLR model is preferred due to its lower AIC, despite the GLR model having a higher R^2 and lower RMSD.
- Total Water Withdrawal | Hybrid (H): The SLR model is the best due to its higher R^2 and lower AIC, despite the GLR model having a lower RMSD.
- Total Water Withdrawal | Natural Moisture (NM): The GLR model is the best due to its higher R^2 and lower RMSD, despite the SLR model having a lower AIC.
- Total Water Withdrawal | Wet (W): The GLR model is the best because it has the highest R^2 and lowest RMSD, despite the SLR model having a lower AIC.
- Reused Water | Hybrid (H): The GLR model is the best because it has the highest R^2 and lowest RMSD, despite the SLR model having a lower AIC.
- Reused Water | Natural Moisture (NM): The SLR model is the best because it has the lowest AIC and RMSD, despite the GLR model having a higher R^2.
- Reused Water | Wet (W): The SLR model is the best because it has the highest R^2 and lowest AIC, despite the GLR model having a lower RMSD.

Conclusion

Regarding the models studied, it was found that simple linear regression (SLR) presents the greatest gain when analyzing the Operational Water and Reuse coefficient. However, for Total Water Withdrawn, the generalized linear Poisson regression (GLPR) models generally presented smaller errors.

For both Total Water Withdrawn and Reuse, a lower correlation was observed with the Production Unit variable (Plant feed). This is because Total Water Withdrawn is more related to the volume of drawdown water than to the processing of the ore itself. In other words, most of the drawdown volume is returned to the environment unused and is not part of the calculation of operational water.

The potential for reuse, on the other hand, is mainly related to units with dams, i.e., units with wet processing. Additionally, it occurs to a greater degree when there is a robust water infrastructure capable of treating/reusing non-new water volumes.

Finally, the models proposed in this work, both the SLR model and the GLPR model, present an important technical-scientific contribution, as they can be used as useful tools for planning and analyzing water use in iron ore mining projects. Furthermore, there are no studies in the literature that estimate water use in relation to production by dividing the uses into Operational Water, Total Water Withdrawn, and Reuse.

References

Akaike H (1974) A New Look at the Statistical Model Identification, 6th edn. Boston

ANA (National Water and Sanitation Agency), IBRAM (Brazilian Mining Institute) (2024) Perspectives and Advances in Water Resources Management in Mining. Brasília (DF)

Dobson A (2001) An Introduction to Generalized Linear Models, Second Edition. Chapman and Hall/CRC

Elian SN (1988) Regression analysis. São Paulo

Grafton RQ, Pittock J, Davis R, *et al* (2013) Global insights into water resources, climate change and governance. Nat Clim Chang 3:315–321. https://doi.org/10.1038/nclimate1746

IBRAM (Brazilian Mining Institute) (2023) Mining in Numbers 2023. https://ibram.org.br/wp-content/uploads/2024/02/mineracao-em-numero-2023.pdf. Accessed 5 Aug 2024

IBRAM (Brazilian Mining Institute) (2024) ESG Results for Brazil Mining 2024. https://ibram.org.br/wp-content/uploads/2024/07/IBRAM-ESG_Resultados23_ColetivadeImprensa-1.pdf. Accessed 5 Aug 2024

Jeong J, Park E, Han WS, *et al* (2017) Identifying outliers of non-Gaussian groundwater state data based on ensemble estimation for long-term trends. J Hydrol (Amst) 548:135–144. https://doi.org/10.1016/j.jhydrol.2017.02.058

Northey SA, Mudd GM, Werner TT, *et al* (2019) Sustainable water management and improved corporate reporting in mining. Water Resour Ind 21:100104. https://doi.org/10.1016/j.wri.2018.100104

Paranhos R, Figueiredo Filho DB, Rocha EC da, *et al* (2014) Unraveling the Mysteries of the Pearson Correlation Coefficient. Leviathan (São Paulo) 66. https://doi.org/10.11606/issn.2237-4485.lev.2014.132346

Rosa L, Chiarelli DD, Tu C, *et al* (2019) Global unsustainable virtual water flows in agricultural trade. Environmental Research Letters 14:114001. https://doi.org/10.1088/1748-9326/ab4bfc