

Development of a Model to Predict Efficiency Parameters in Industrial Processes

Jorge Duarte Forero^{1*}, Guillermo E. Valencia¹ and Luis G. Obregon²

¹Department of Mechanical Engineering, KAI Research Group, Universidad del Atlantico, Barranquilla
jorgeduarte@mail.uniatlantico.edu.co, guillermoevalencia@mail.uniatlantico.edu.co

²Department of Chemical Engineering, Sustainable Chemical and Biochemical Processes Research Group, Universidad del Atlantico, Barranquilla Colombia; luisobregon@mail.uniatlantico.edu.co

Abstract

Background/Objectives: The implementation of regression and correlation analysis as a tool for the modeling of industrial-scale milling processes, as a base for optimization processes that allow establishing a production regime with the highest consumption rate and therefore, a better process performance. **Methods:** Using the multivariate regression methods, a mathematical model to establish a correlation between the consumption rate and the factors that influence it was developed; in this process, a series of statistical analysis was developed to select the factors that generate an effect on the behavior of the consumption rate, at the same time, to determinate the best fit for the experimental data measured, using a statistical tool for the operations required. **Findings:** From the obtained results, it is shown that from the set of initial variables studied, only two (the humidity and the production), have a degree of correlation enough to be taken in account into post-processing operations; besides, it was found that the behavior of the consumption rate can be approximated as a linear combination of the relevant factors, with a better accuracy when compared to more complex fits, which in turn generates a model easy to handle in optimization processes **Application:** To develop a tool for the prediction of the behavior of consumption rate in milling processes, which will be used for the establishment of a more efficient production regime, where a better planification yields a lower energy consumption, with the same production level.

Keywords: Fitting Processes, Operation Characteristics, Performance, Regression Methods, Statistical Model

1. Introduction

Milling processes are defined as those that involve breaking down, separating, sizing, or classifying aggregate material; these actions are done for 1. To increase the surface area of a solid, 2. To manufacture of a solid with a desired grain size and 3. To do pulping of resources¹. Though the milling processes are present in a wide variety of engineering applications, the economic activity that uses these to their full extent, is the construction; actually, concrete (a composite material composed of fine and coarse aggregate bonded together with a fluid cement paste that hardens over time) is the preferred choice in the construction of buildings and houses, due to its low cost and high resistance to environmental factors (as the rain and dust). As a result, the milling

process is now part of huge industrial complexes that devote specifically to the fabrication and processing of concrete.

However, with the apparition of standards that regulate the energetic and environmental issues associated with production (such as ISO standards), the latter has become a variable of special care into the industrial processes. Therefore, a huge quantity of studies was developed in recent years, with the sole purpose of modeling the influence of operational variables on the process efficiency, to find the most efficient combination of these variables and improve the overall performance¹⁻⁴. However, this raises the following question: Which is the best way to model the efficiency? If a simple and robust mathematical model is required to predict it, then a linear or quasi-linear model based in a low quantity of indepen-

*Author for correspondence

dent variables would be the best option, though it could be not able to accurately predict the efficiency⁵.

Regarding the said issue, modern statistics allows to evaluate the degree of approximation between a certain model and the experimental data used to estimate it; in other words, these statistical methods measure how good is the fit of the model. One of these methods and also a common term in scientific research is the Regression Analysis; this is a statistical methodology that encloses proceedings related to study and quantify the relationship between a certain set of variables and an output, being the efficiency in this case. Due to this, the Regression Analysis is widely used to study the behavior of a system modeled by using experimental data and to determine points of interest inside the model, even if very little information about the system is known⁶⁻¹². Besides these facts, a strong point of this method is that the final results, consists of a mathematical model for the studied system, with coefficients for the involved variables, which are calculated through linear regression methods in order to reduce to a minimal value the difference between the experimental data and the predicted values of the model.

In this article, a methodology based in the Regression Analysis is used to develop a model that allows predicting efficiency parameters, particularly consumption rate of an industrial-level milling process, from experimental data, processed through a statistical package, such as Minitab, due to the number of operations required to do this process; many of these operations are comprised of a series of intermediate statistical processes used to sort and discard non-relevant variables and to obtain a robust model that can predict accurately the consumption rate and at the same time, a model easy to handle in future optimization processes.

2. Methodology

2.1 Purpose and Description of the Study

The main purpose of the present study is to formulate a mathematical model that allows predicting the behavior of a performance parameter, such as the consumption rate (a widely used energetic indicator), when any of the variables that comprise the model is changed¹³. To do so, a set of experimental data was measured in an industrial-grade mill available in a factory that elaborates concrete for the construction

sector. This set contains the data related to the variables taken into account, which were the humidity of the raw material, the production volume and the non-used hours of work, measured along the first half of the current year (2018). These variables were used into a regression analysis, to statistically define their relevance and to quantify the numerical terms involved into the model, accompanying these variables, as well to validate the results given by the modeled consumption rate, when compared with the experimental data. As these operations require the application of some statistical concepts and operations that would represent huge time consumption if these were done by hand, software package Minitab was used as a tool for the statistical operations needed.

2.2 Fundamental Theory and Equations

To improve the efficiency and performance of industrial processes, it is necessary to establish a fundamental equation that allows correlating an output of interest and the factors that influence it. Using the variables (or factors) x_1, x_2, \dots, x_n and the output y , a process can be described or modeled as shown in the Equation 1:

$$y = f(x_1, x_2, \dots, x_n) \quad (1)$$

Where is stated that these variables would affect the desired output^{5,13}. However, Equation 1 cannot be used in practical applications (it is only a mathematical formulation). To worsen things, many processes are dependent on a huge number of factors, which makes process modeling a hard operation. Given this, two issues have to be solved: The first is related to the possibility to reduce the Equation 1 into another with less independent variables, such as Equation 2:

$$y = f(x_1, x_2) \quad (2)$$

Where only the influent factors remain in the equation; in the other hand, the second issue deals with the adequate mathematical expression for Equation 2, as there are some cases when a linear combination of variables, is not enough to accurately fit the model to the experimental data. Fortunately, actual statistics allow solving these two issues through the application of the correlation and regression analyses, respectively. In the correlation analysis, a sample data is evaluated in order to verify (using a hypothesis test) the existence of an

association, commonly linear between two quantitative variables¹³; that is to say, to test the null hypothesis that there is no association between an independent variable x and the dependent (output) variable y , assuming that the data follows a normal distribution of probability (which is easily verified using a normality test). This relationship is validated through the estimation of the Pearson coefficient of correlation (r), which measures the degree and direction of this relationship and it is calculated by the Equation 3:

$$r = \frac{\sum x_i y_i - n(\bar{x} \bar{y})}{n \sigma_x \sigma_y} \quad (3)$$

Where x_i and y_i denote each individual value of the variables and y respectively; in the other hand, \bar{x} and \bar{y} are the mean of each set of values, and σ_x , σ_y are the standard deviation of the x and y values. Besides that, given that correlation analysis is an type of hypothesis test, a statistic value F is calculated for each variable, which is compared with Fisher's probability distribution F to obtain the so-called P-value, which defines a variable studied as statistically relevant or not (for a given reliability, usually 95%) by comparing with the critical P-value, as in the Equation 4:

$$P(F_{data}) \leq P(F_{critical}) \quad (4)$$

Where the critical P-value is defined by the specified reliability of the analysis⁵.

While correlation analysis allows discarding unused variables from the modeling, the regression adjusts the predicted values to a model that expresses the response as a function of the remaining variables and their interaction; in its most basic form (a linear expression), when applied to two factors x_1 and x_2 , it is obtained the model from the Equation 5:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 \quad (5)$$

Where coefficients α is called regression coefficients, which are commonly determined through advanced mathematical methods, such as least squares⁵. In addition to these coefficients, the regression analysis results in the value R^2 , which is interpreted in general terms as the proportion of the variability in the data "explained" by the model obtained (that is, the closeness of the model to the data obtained), whose value ranges between 0 and 1, with 1 is a perfect fit to

the data. Finally, the accuracy of the model developed can be calculated through the residuals, defined as the difference between the values measured and those predicted by the model, as shown in Equation 6:

$$e_{Residual} = y_{real} - y_{model} \quad (6)$$

3. Result and Discussion

As said before, the main objective of the study was to develop a model able to predict the behavior of the consumption rate in a milling process, using as input some variables measured directly from the process; to do so, a series of statistical tools were applied in order to isolate the variables that have a statistically significant influence over the studied output. Therefore, the results can be divided according to the statistical and mathematical tools used in each step, which end in the definition and validation of the model developed.

3.1 Correlation Study for the Process Variables

As a first step in the process, a correlation study was done to all three independent variables (humidity, production volume and non-used hours of work) with respect to the consumption rate, to verify their degree of influence over the latter; Figures 1 to 3 show in a graphical form the results of the study. From these, it can be seen the responses in each case differ between each other: The humidity relates with the consumption rate in a linear form with a positive slope (a directly proportional relation), while the production relates using a negative slope (an inversely proportional relation). However, this is not the case

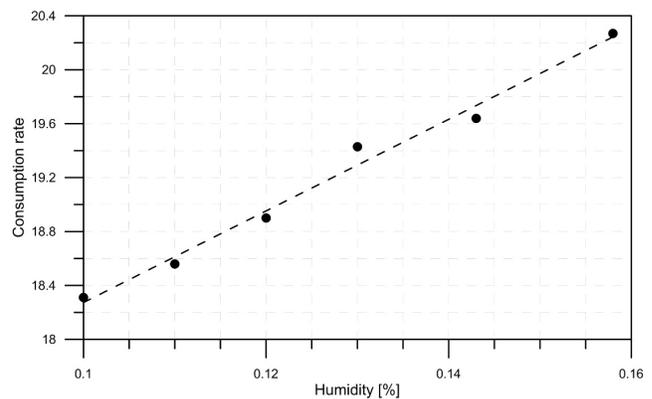


Figure 1. Scatter plot of consumption rate versus humidity.

for the non-used hours of work: as shown in Figure 3, the values have a high dispersion and cannot be correlated with the consumption rate. Therefore no relationship can be found between these variables. To reassure this finding, Table 1 shows the values of the Pearson correlation coefficient for all three variables,

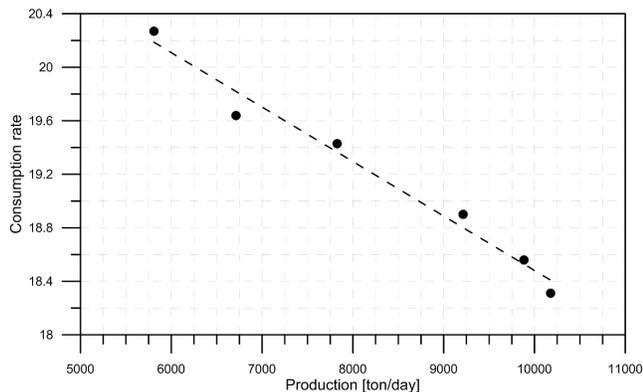


Figure 2. Scatter plot of consumption rate versus production.

together with their P-value; by comparison with the critical value (0.05), it is verified that from these variables, only two, the production and the humidity, exert a significative influence over the consumption rate, and therefore the non-used time can be discarded from the analysis¹³.

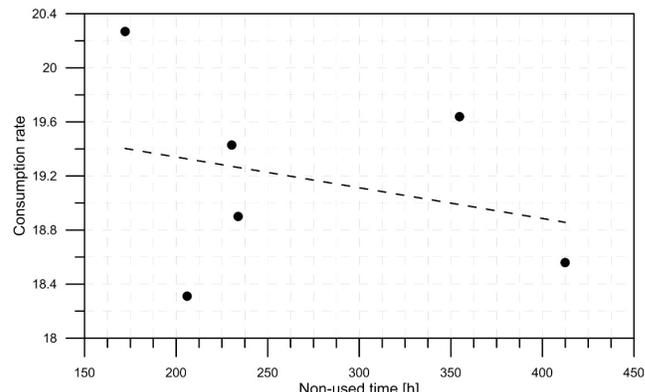


Figure 3. Scatter plot of consumption rate versus non-used time.

Table 1. Results of the correlation analysis done in Minitab

Variable		Humidity	Production	Non-used time
Production	Pearson's r	-0.988	-	-
	P-value	0.000	-	-
Non-used time	Pearson's r	-0.223	0.240	-
	P-value	0.671	0.647	-
Consumption rate	Pearson's r	0.993	-0.998	-0.291
	P-value	0.000	0.000	0.575

3.2 Regression Model and Validation

Once the relevant variables were selected (Humidity and Production volume), the next step involved the elaboration of the regression model; however, from Figures 1 and 2, some characteristics of the model can be guessed. From these figures, it is shown that the consumption rate presents a linear behavior concerning each variable; therefore, initial guessing for the regression model would be a linear combination of each factor. With this in mind, an initial linear formulation was introduced to the software, and the results show that the experimental data can be fitted to the expression of the form shown in Equation 7:

$$CR = 17.05 + 24.4 [Humidity] + 0.000116 [Production] \quad (7)$$

With an R² coefficient of 98.87%, which represents a good fit; a plot of the behavior of consumption rate calculated by Equation 6, is shown in Figure 4. Finally, two things need to be verified: The accuracy of the developed model and the initial assumption on the correlation analysis (the fact that the data must follow a normal probability distribution). For the first issue, Figure 5 shows the residuals calculated at the experimental points, which in the worst case, does not even exceed the 1% and it goes in concordance with the R² coefficient. For the second issue, a normality test was done to the values of the consumption rate; as shown in Figure 6, the measured data has a normal behavior, which validates the application of the correlation analysis and therefore, the developed model.

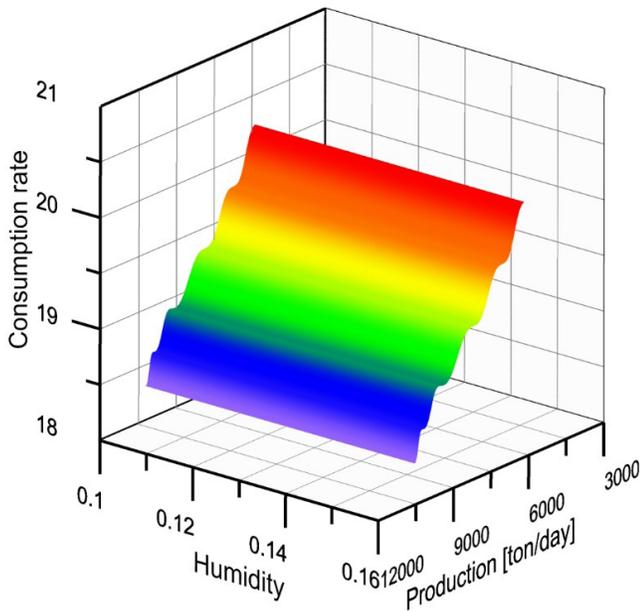


Figure 4. 3D plot of the developed model.

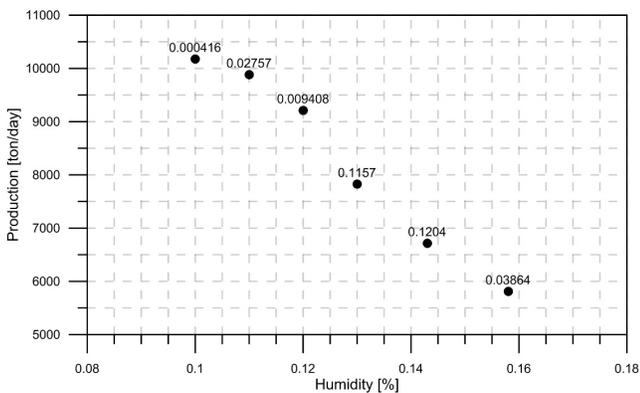


Figure 5. Residual plot for the developed model.

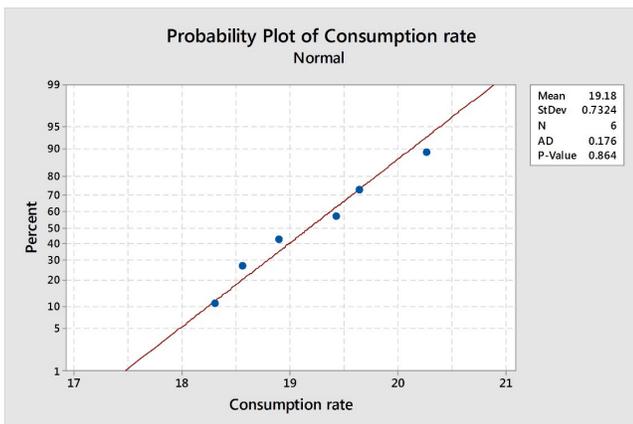


Figure 6. Results of the normality test done in Minitab.

4. Conclusions

From the present study, it can be concluded that the application of statistical methods constitute a powerful tool in the study and modeling of industrial processes, especially those where it would be normally difficult to establish a mathematical analysis that involves all the variables implicated within the process; this kind of situations, many times end in a waste of time and resources, as some of the variables involved in a process, do not have an appreciable influence over a specific factor. In the studied case, the correlation analysis made evident that non-used time does not influence the efficiency indicator and therefore can be neglected without an appreciable loss of accuracy and at the same time it allowed to predict (up to some extent) the behavior of the model, based on the individual study of the interaction between each variable and the output; therefore, it can be used to obtain hints about the mathematical model to use in regression analyses and as a tool to simplify said model by discarding variables unused properly during the modelling phase. Finally, the regression analysis allowed establishing a mathematical formulation of the influence exerted by the studied variables on the efficiency indicator (the consumption rate). As the residuals generated from the model showed an accuracy error lower than 1% with respect to the experimental data, it is verified the usefulness of these statistical methods (and the statistical software packages), as a solution to model (and by extension, optimize) industrial processes using experimental data as their starting point.

5. Acknowledgment

Acknowledgments for the support provided by Universidad del Atlantico, Barranquilla - Colombia through of Efficient Energy Management Research Group Kai, with J. Duarte - G. Valencia and Sustainable Chemical and Biochemical Processes Research Group, with L. Obregon.

6. References

1. Pecat O, Rentsch R, Brinksmeier E. Influence of milling process parameters on the surface integrity of CFRP. *Procedia CIRP*. 2012; 1:466–70. <https://doi.org/10.1016/j.procir.2012.04.083>

2. Campatelli G, Lorenzini L, Scippa A. Optimization of process parameters using a Response Surface Method for minimizing power consumption in the milling of carbon steel. *Journal of Cleaner Production*. 2014; 66:309–16. <https://doi.org/10.1016/j.jclepro.2013.10.025>
3. Vazquez E, Rodríguez CA, Elías-Zu-iga A, Ciurana J. An experimental analysis of process parameters to manufacture metallic micro-channels by micro-milling. *International Journal of Advanced Manufacturing Technology*. 2010; 51(9–12):945–55. <https://doi.org/10.1007/s00170-010-2685-4>
4. Wang H, Qin X, Ren C, Wang Q. Prediction of cutting forces in helical milling process. *International Journal of Advanced Manufacturing Technology*. 2012; 58(9–12):849–59. <https://doi.org/10.1007/s00170-011-3435-y>
5. Triola M. *Estadística*. 10th Ed. Mexico: Pearson Educacion de Mexico, S.A; 2009.
6. Woodside AG. Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. *Journal of Business Research*. 2013; 66(4):463–72. <https://doi.org/10.1016/j.jbusres.2012.12.021>
7. Lambourne K, Tomporowski P. The effect of exercise-induced arousal on cognitive task performance: A meta-regression analysis. *Brain Research*. 2010; 1341:12–24. PMID: 20381468. <https://doi.org/10.1016/j.brainres.2010.03.091>
8. Tonidandel S, LeBreton JM. Relative importance analysis: A useful supplement to Regression Analysis. *Journal of Business and Psychology*. 2011; 26(1):1–9. <https://doi.org/10.1007/s10869-010-9204-3>
9. Bel G, Fageda X, Warner ME. Is private production of public services cheaper than public production? A meta-regression analysis of solid waste and water services. *Journal of Policy Analysis and Management*. 2010; 29(3):553–77. <https://doi.org/10.1002/pam.20509>
10. Fernandes BS, Gama CS, Maria Cereser K, Yatham LN, Fries GR, Colpo G, de Lucena D, Kunz M, Gomes FA, Kapczinski F. Brain-derived neurotrophic factor as a state-marker of mood episodes in bipolar disorders: A systematic review and meta-regression analysis. *Journal of Psychiatric Research*. 2011; 45(8):995–1004. PMID: 21550050. <https://doi.org/10.1016/j.jpsychires.2011.03.002>
11. Logan SW, Robinson LE, Wilson AE, Lucas WA. Getting the fundamentals of movement: A meta-analysis of the effectiveness of motor skill interventions in children. *Child: Care, Health and Development*. 2012; 38(3):305–15. PMID: 21880055. <https://doi.org/10.1111/j.1365-2214.2011.01307.x>
12. Cohen P, West SG, Aiken LS. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge. 1984.
13. Camargo C, García C, Duarte Forero JE, Rincon A. Modelo estadístico para la caracterización y optimización en bombas periféricas. *Revista Científica Ingeniería y Desarrollo*. 2017; 36(1):18–39. <https://doi.org/10.14482/inde.36.1.10939>