Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 177 (2020) 444–449

Procedia Computer Science

www.elsevier.com/locate/procedia

The 6th International workshop on Big Data and Networks Technologies (BDNT 2020)

November 2-5, 2020, Madeira, Portugal

Artificial neural network to predict the performance of the phosphoric acid production

Ahmed Bichri[a,b*], Mohamed Anouar Kamzon[a], Souad Abderafi[a]

[a]MOSE2PI, Mohammadia Engineering School, Mohammed V University in Rabat, Av. IbnSina, B.P. 765, Agdal, Rabat, 10090, Morocco
[b]OCP SA, Jorf Lasfar, Morocco

Abstract

The performance of the phosphoric acid production process is based on an increasing amount of heterogeneous data. This performance depends on several parameters that must be controlled at the unit of attack and filtration of the process. From the description of the attack unit, we found that a considerable flow of data can be generated that needs to be integrated for a parametric sensitivity analysis. After processing these data, we followed the artificial neural network method. This method allowed us to develop a model that predicts the yield of phosphoric acid at different operating parameters, with a very satisfactory error. Then the reliable model was exploited to evaluate the relative importance input parameters on the phosphoric acid production.

Keywords: Big data ; Operating parameters ; Phosphoric acid performance ; Artificial neural network ; Performance.

1. Introduction

Typically, the phosphoric acid needed for the fertilizer industry is produced by the wet process [1]. The optimization of this process is largely based on both technical and empirical knowledge of process behavior. Various

* Corresponding author. Tel.: +212-666-28-23-92.
E-mail address: a.bichri@ocpgroup.ma

10.1016/j.procs.2020.10.060

phenomenological modelling and process design work has been reported in the literature to improve the understanding and efficiency of existing processes [2, 3]. However, given the large number of parameters that need to be analyzed requires the use of statistical modelling [4], which to date has not been applied to the field of phosphoric acid production. This statistical modeling consists in representing a system with the "Black Box" approach by mathematically linking the inputs of the system to its outputs without worrying about the knowledge of the phenomena involved. This procedure requires the operation of a large industrial data base involving problems for their analysis. It is therefore necessary to manage these Big Data, correctly, in order to be able to derive rigorous and reliable interpretations; especially since there are advanced techniques and technologies that can be adopted to deal with Big Data problems [5]. Among the different approaches that exist is the Artificial Neural Network (ANN) method, which has been successfully applied to various disciplines including computer science, science, engineering, medicine, environment, agriculture, mining, technology, climate, business, arts and nanotechnology, as shown in a recent study on the state of the art in ANN applications, evaluating its performance in the different fields studied [6].

In this work, we will focus on the application of this ANN method to analyze the parametric sensitivity on the yield of phosphoric acid, by exploiting the industrial data relating to the phosphoric acid production unit. Integrating such a large and heterogeneous amount of data is in itself a challenge that must be overcome before accurate models can actually be developed.

2. Process and database

The phosphate pulp coming from the laundromat via the pipeline feeds a thickener to increase its solid concentration from 55 to 65%. Then, this pulp is converted into phosphoric acid and calcium sulphate dihydrate (gypsum) in a reactor of OCP/Jorf Lasfar - Line E with the addition of a mixture of sulfuric acid and recycled phosphoric acid from the washing of the gypsum cake formed. The process is characterized by its annular reactor (Jacobs) composed of an attack tank formed of nine compartments and a maturation tank of four compartments. The phosphate pulp is introduced into a compartment, and mixed with a mixture of sulfuric and phosphoric acid in well-identified compartments, then cooled in a flash cooler before feeding the maturation tank allowing an additional residence time to the crystals in order to increase the filterability of the slurry at the level of two filters separating phosphoric acid and gypsum. At the reactor level, phenomena occurring in the phosphoric acid production process involve intense interactions between solid and liquid components, with sensitivity to a certain number of parameters such as the characteristics of the minerals, the compositions and flow rates of the reagents and the operating conditions of the process. The main challenge of this step is to produce an aqueous solution of phosphoric acid with a high P_2O_5 yield and negligible chemical losses in P_2O_5 (Soluble P_2O_5, unattacked P_2O_5, syncrystallized P_2O_5).

Based on industrial data for the first half of 2019, a large database has been obtained, the number of which is equal to 5057.This database reports, by post, the monitoring of the attack and filtration process of phosphoric acid. The different operating parameters considered for these two units are grouped in Table 1. It is a heterogeneous database that has to be analyzed. At first, we began with a database exploration, for a treatment that allowed us to eliminate missing and aberrant values. The identification and cleaning of aberrant values is an important point for the processing of large databases. There are different algorithms that can identify the aberrant values of a system, we have chosen the quartile algorithm which is a common method for detecting aberrant values. The principle of this method is recently described by Xiaojun et al. [7]. By following this method, we obtained a new database of 1236 which was exploited for the modelling of the studied process. This modelling was carried out, using artificial neural networks (ANN), following the procedure described below.

Table 1. Parameters of the attack and filtration unit considered in the database

1	2	3	4	5	6
Pulp density output Thickener	Sulphuric acid flow rate	Recycled phosphoric acid flow rate A	Recycled phosphoric acid flow rate B	Pulp flow rate	Temperature compartment 09
7	**8**	**9**	**10**	**11**	**12**
Input temperature Flash cooler	Temperature compartment 10	Delta T input and output Flash cooler	Vacuum flash cooler	Slurry flow rate filter A	Slurry flow rate filter B
13	**14**	**15**	**16**	**17**	**18**
Gypsum flow rate Filter A	Gypsum flow rate Filter B	Product acid flow rate A	Product acid flow rate B	Vacuum Filter A	Vacuum Filter B
19	**20**	**21**	**22**	**23**	**24**
Pulp density	Slurry density	Slurry Solid Rate	$SO_4^{2-}$of the slurry	Density Filtrate slurry	Product acid density filter A
25	**26**	**27**	**28**	**29**	**30**
Product acid density filter B	% P_2O_5 Product acid filter A	% P_2O_5 Product acid filter A	Product acid solid rate filter A	Product acid solid rate filter B	SO_4^{2-} Product acid filter A
31	**32**	**33**	**34**	**35**	**36**
SO_4^{2-} Product acid filter B	P_2O_5 (WS) Filtrer A	P_2O_5 (SYN) Filtrer A	P_2O_5 (Unat) Filtrer A	CaO Filtrer A	P_2O_5 (WS) Filtrer B
37	**38**	**39**	**40**	**41**	
P_2O_5 (SYN) Filtrer B	P_2O_5 (Unat) Filtrer B	CaO Filtrer B	Pulp Solid Rate Decanter Outlet	Dry phosphate flow rate	

3. ANN methodology

An artificial neural network, analogous to the behavior of the biological neural structure, is an effective empirical modelling tool for the approximation of non-linear functions, pattern recognition and classification problems. Neural networks perform the correlation without requiring a mathematical description of how the output depends on the input, giving neural networks an advantage over traditional methods of estimating the function. Instead, neural networks learn from examples of input-output data sets provided to them. A neural network is a set of nodes (neurons) connected by connections (synapses). Two or more neurons can be combined in one layer. A network can have one or more layers, such as an input layer, one or more hidden layers and an output layer. Each connection is associated with a scalar weight that modifies the signal strength. The neuron function consists of summing the weighted inputs and transmitting the sum to the neuron by a non-linear transfer function. This function has many different forms [8], we choose, a "sigmoid tangent" function for the hidden layer, because it is continuous and relatively easy to calculate (just like its derivative). It maps the outputs far from extremes and introduces a non-linear behavior to the network, its equation is given by:

$$\mathrm{tansig}(x) = \frac{2}{1+e^{-2x}} \quad \text{and} \quad x_i = \sum w_{i,j} y_i + b_i \qquad (1)$$

Where, x is the sum of the input weights and a bias, b_i can also be used, which is another neuron parameter, added to the weighted neuron inputs.

After building the neural network, we move on to the training stage. The latter is an optimization strategy that estimates weights and biases so that the output of the network is as close as possible to the target values and the neural network will learn the relationship between the inputs and outputs of a dynamic system. The purpose of network training is to minimize the error between the actual outputs and the network outputs by changing the weights and biases of a network, called a training algorithm. After each iteration, the network outputs are compared to the actual target values until the performance function is satisfied. This function is given by the mean square error (MSE):

$$\mathrm{MSE}=\frac{1}{n}\sum(V_{exp}-V_{cal})^2 \quad (2)$$

Where: n is the amount of data; V_{exp} and V_{cal} are the experimental and calculated yield values of P_2O_5, respectively.

There are different learning algorithms to train neural networks; the fastest learning algorithm is Levenberg-Marquardt's algorithm [9]. It was used to optimize network weights.

4. Results and discussion

The network was formed by exploiting the industrial database mentioned above, considering as input the 41 independent parameters and as a dependent variable the phosphoric acid yield. 70% of these data are used for training, 15% for validation and 15% for testing. The analysis was carried out using the MATLAB software; and by choosing the tansig transfer function and the Levenberg-Marquardt algorithm for training. We sought to minimize the number of nodes in the hidden layer in order to have minimal complexity. Several tests were carried out, each time changing the number of neurons in the hidden layer. The best validation performance was obtained for a mean square error value of 2.2×10^{-6} completed after 68 epochs (Figure 1). The optimal structure of the implemented neural network corresponding to this performance is shown in Figure 2. This figure shows that the identified neural network has eight neurons in the hidden layer and one neuron in the output layer, with a tangent-sigmoid and linear transfer function in the hidden and output layers, respectively.

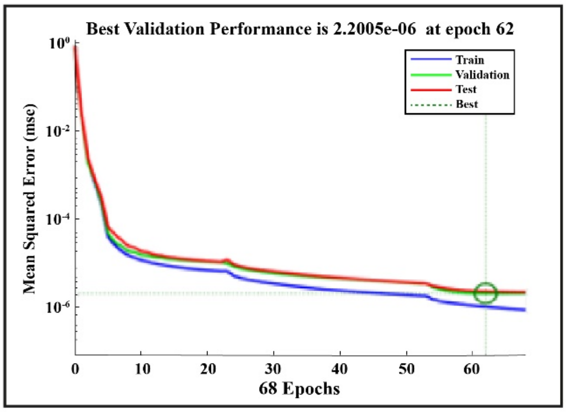

Fig. 1. Satisfaction of the performance function

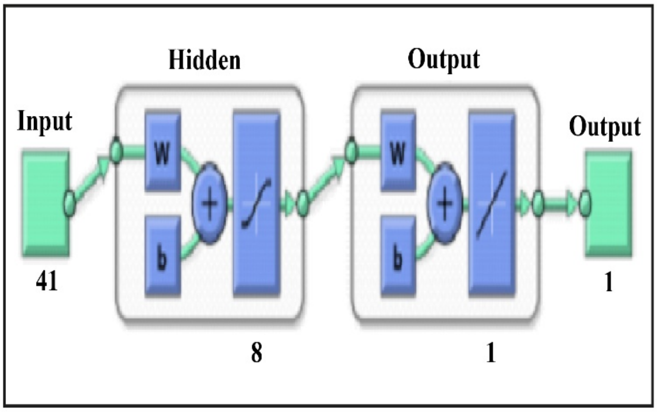

Fig. 2. Neural network structure

The validation of the model was also verified by comparing the calculated values with the industrial ones. This comparison is shown in Figure 3. It shows that the output follows the targets very well, for training (R = 0.99999), validation (R = 0.99998), and testing (R = 0.99998). These values may be equivalent to a total response of R equal to 0.99999. All these results allow us to deduce that the ANN has produced a good model in terms of precision and reliability. In this case, the simulation can be used to enter new inputs.

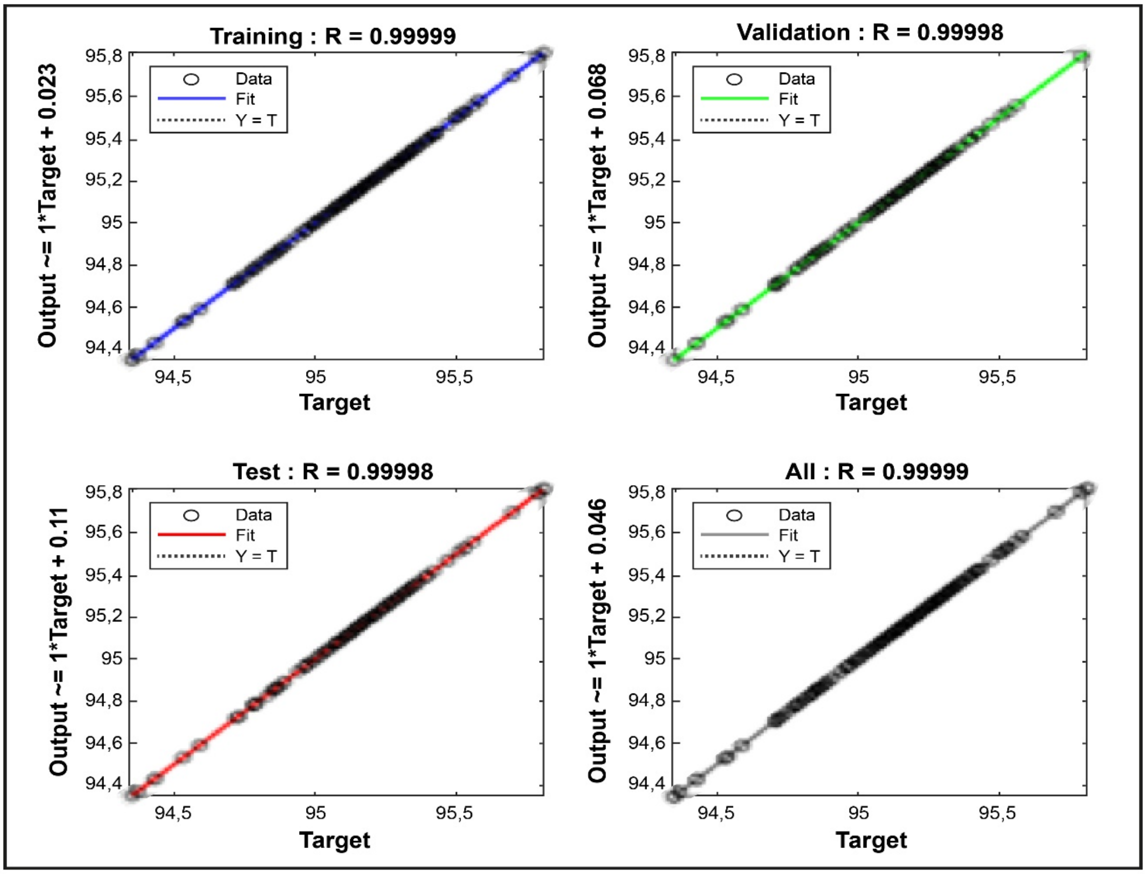

Fig. 3. Model regression analysis

The importance of each of the independent variables in artificial neural networks was also evaluated. Different methods can be used for this quantification. The connection weight approach was chosen because it provides the best overall methodology to accurately quantify the importance of the variables according to the comparison made by Olden et al. [10]. According to this algorithm the calculation is given by [11]:

$$RI_i = \left[\frac{S_i}{\sum_{i=1}^{n} S_i}\right] \times 100 \qquad \text{and} \qquad S_i = \sum_{j=1}^{m} \frac{|w_{ij} \times w_{jk}|}{\sum_{i=1}^{n} |w_{ij} \times w_{jk}|} \tag{3}$$

Where: RI_i, the relative importance of each variable; Si, the sum of the contributions of input neurons; i, j, and k represent the input, hidden and output layers of the network, respectively; W_{ij} represents the weight of the connection between the i^{th} input and the j^{th} neuron of the hidden layer; m, the total number of nodes in the hidden layer of the network and n, the total number of input variables in the neural network.

To determine the factors influencing the phosphoric acid yield, the resulting matrix containing the connection weights of the input-output neurons was used to measure the importance of the forty-one variables grouped in Table 1, using the equation (3). The results obtained are shown in Figure 4. These results show that the different parameters studied are significant. P_2O_5 (WS) Filter B, P_2O_5 (WS) Filter A and P_2O_5 (SYN) Filter B have the most significant effect on phosphoric acid yield, with RI equal to 6.38%, 6.08% and 6.05%, respectively. However, the SO_4^{2-} in the slurry has the least significant effect, with an RI value equal to 0.71%, followed by the Filter A gypsum flow rate with an RI of 0.86%.

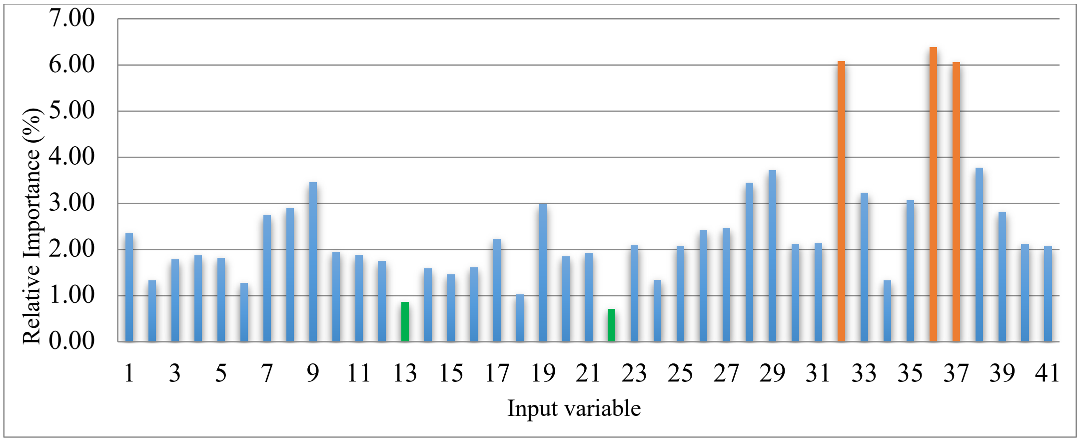

Fig. 4. Relative importance of input variables

5. Conclusion

In this work, we have successfully carried out a modeling using the ANN, relating to the industrial unit of phosphoric acid production. The results obtained indicated that the model can predict the phosphoric acid yield of this unit with a very satisfactory error, for this complex system. It was then used to analyze the importance of independent variables on the performance of the attack and filtration unit, which was evaluated based on the yield of phosphoric acid produced. Therefore, the developed model can be used for the simulation and prediction of the performance of the industrial unit. It is an efficient analysis and diagnostic tool that can be integrated into the control room, by plant operators and decision-makers, to understand and simulate its non-linear behavior.

Acknowledgements

The authors gratefully acknowledge support provided by OCP Group.

References

[1] Abu-Eishah, Samir I., and Nizar M. Abu-Jabal. (2001) “Parametric Study on the Production of Phosphoric Acid by the Dihydrate Process.” *Chemical Engineering Journal* 81 (1): 231–250.

[2] Papadopoulos, Athanasios I., and Panos Seferlis. (2009) “Generic Modelling, Design and Optimization of Industrial Phosphoric Acid Production Processes.” *Chemical Engineering and Processing: Process Intensification* 48 (1): 493–506.

[3] Messnaoui, Brahim, and Tijani Bounahmidi. (2006) “On the Modeling of Calcium Sulfate Solubility in Aqueous Solutions.” *Fluid Phase Equilibria* 244 (2): 117–127.

[4] Di Ciaccio, Agostino, Mauro Coli, and Jose Miguel Angulo Ibanez (2012) “Advanced statistical methods for the analysis of large data-sets”, (eds) *Springer Science & Business Media*.

[5] Philip Chen, C. L., and Chun-Yang Zhang (2014) “Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data.” *Information Sciences* 275: 314–347.

[6] Abiodun, Oludare Isaac, Aman Jantan, Abiodun Esther Omolara, et al. (2018) “State-of-the-Art in Artificial Neural Network Applications: A Survey.” *Heliyon* 4 (11): e00938.

[7] Shen, Xiaojun, Xuejiao Fu, and Chongcheng Zhou (2019A) “Combined Algorithm for Cleaning Abnormal Data of Wind Turbine Power Curve Based on Change Point Grouping Algorithm and Quartile Algorithm.” *IEEE Transactions on Sustainable Energy* 10 (1): 46–54.

[8] Khayet, M., and C. Cojocaru (2013) “Artificial Neural Network Model for Desalination by Sweeping Gas Membrane Distillation.” *Desalination. New Directions in Desalination* (308): 102–110.

[9] Godini, H. R., M. Ghadrdan, M. R. Omidkhah, and S. S. Madaeni (2011) “Part II: Prediction of the Dialysis Process Performance Using Artificial Neural Network (ANN).” *Desalination* 265 (1): 11–21.

[10] Olden, Julian D, and Donald A Jackson (2002) “Illuminating the “Black Box”: A Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks.” *Ecological Modelling* 154 (1): 135–150.

[11] Garson GD (1991) "Interpreting neural-network connection weights”, Artif. Intell. Expert 6:47–51.