25th ABCM International Congress of Mechanical Engineering
October 20-25, 2019, Uberlândia, MG, Brazil

COB-2019-0216
DESIGN OF EXPERIMENTS COMBINED WITH ARTIFICIAL NEURAL NETWORKS APPLIED ON THE CONTROL PARAMETERS OF A REAL STEAM GENERATOR

Lara Werncke Vieira
Bárbara Pacheco da Rocha
Paulo Smith Schneider
Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil
lara.vieira@ufrgs.br
pr_barbara@hotmail.com
pss@mecanica.ufrgs.br

Guilherme Lacerda Batista de Oliveira
Energy of Portugal - EDP
guilherme.oliveira@edpenergiapecem.com.br

Abstract. *The present paper aims to use statistical analysis through DoE methodology to select and rank the most influential parameters of a steam generator modeled by an ANN. The goal is to focus the efforts only on the significant operational input parameters according to their influence on the three responses: steam generator efficiency, electric power generation and flue gas outlet temperature. The main controllable parameters are: primary air flow, pulverized coal outlet temperature, velocity of the dynamic classifier, excess* O_2*, primary air's crossover duct pressure, secondary air's crossover duct pressure and coal flow. The ANN is trained with operational data provided by the PECEM power plant (2x360 MW), from Fortaleza, Brazil. The metrics that evaluate the ANNs performance are the mean absolute error, the mean square error and the mean percent error. According to the DoE results the primary air flow and the velocity of the dynamic classifier are not important to the flue gas outlet temperature output. A simplified ANN is developed with the selected parameters. For the two other outputs, all input parameters remained significant.*

Keywords: *Coal-fired Power Plant, Design of Experiments, Artificial Neural Networks, Steam Generator*

List of Symbols

C_O	Center points
k	Number of factors
N	Number of essays
$P1$	Primary air flow, *kg/s*
$P2$	Pulverized coal outlet temperature, $^\circ C$
$P3$	Velocity of the dynamic classifier, *rpm*
$P4$	Excess O_2, *%*
$P5$	Primary air collector pressure, *mbar*
$P6$	Secondary air collector pressure, *mbar*
$P7$	Coal flow, *ton/h*
$S1$	Flue gas outlet temperature , $^\circ C$
$S2$	Boiler efficiency, *%*
$S3$	Electric power generation, *MW*

1. INTRODUCTION

Coal-fired power plants account for 3.60% of the total electrical energy of Brazil's power grid and 38.25% of the world. Brazil's electrical energy matrix is an exception when compared with other countries predominant in renewable energies, especially due to the high insertion of hydroelectric power. Even so, the significance of coal power in global level is undoubted. Since 2000, coal-fired power generation has grown by nearly 900 GW worldwide (IEA, 2019; MME, 2018).

In coal-fired power plants, a large number of operational data is captured continuously. In order to fully understand the system's operation, these data must be carefully looked at. Pattern recognition and variables correlations are some of the methods usually explored. By aligning available data, efficient management and strategy it is possible to achieve the optimum point of the process. The constant monitoring allows noticing correlated events and operating configurations that leads to actual production parameters (GE, 2017; Smrekar *et al.*, 2009).

Developing a mathematical model for a steam generator can be very demanding. ANN models are capable of describing a system with lesser effort but with great utility. These models are trained on existing data by running large amounts of data until it finds enough patterns to be able to make accurate decisions about definite parameters (Hall *et al.*, 2015). Studies have already succeeded in modeling the steam generator of coal-fired power plants (Smrekar *et al.*, 2009; Strušnik *et al.*, 2015) as well as the whole system (Bekat *et al.*, 2012; Smrekar *et al.*, 2010; Tunckaya and Koklukaya, 2015).

Design of Experiments enables to investigate cause and effect relations and determine the influence that the input parameters have on the output response in a system. Analyzing the individual effects of each parameter and the interactions between them allows the development of a model that relates the response to the significative input parameters. This model can be used for improvements and support decision making (Montgomery, 2013).

The objective of this paper is to present a Design of Experiments methodology to select the most significant input parameters of an ANN that models a real 360 MW coal-fired power plant installed in Ceará, Brazil. Furthermore, the Response Surface Methodology will be used that approximates the model performance for different parameters configurations. Benefits of using DoE includes the necessity of running only a few experiments, efficient parameter selection based on statistical theory and definition of the most suited operating ranges (Lujan-Moreno *et al.*, 2018; Pirhadi *et al.*, 2018; Weissman and Anderson, 2015).

2. ARTIFICIAL NEURAL NETWORK

An artificial neural network consists of an information processing system created based on the functioning of biological neurons. Resembling the human brain, ANNs are composed of a large number of simple processing elements called neurons as first proposed by McCulloch and Pitts (1943) in their perceptron model. Neurons will gather information from the environment through a learning process and are connected with each other by targeted communication links reproducing a synapse, each with an associated weight (Haykin, 2014).

The perceptron network is a model that is bounded by one layer of input neurons and another of output neurons. Whenever intermediate layers are added, the model is called Multi-Layer Perceptron (MLP). The MLP architecture houses an input layer, an output layer, and intermediate layers called "hidden" layers. The inputs are associated with neurons in the left layer of the input, where external information feeds the network. As a next step, the information passes to the hidden layer to be processed. The processed information is then transferred to the output layer (Haykin, 2014).

The MLP model stands out for three characteristics: nonlinear activation function, hidden neurons and high degree of connectivity. The activation function should exhibit smooth nonlinearity for gradient variation and error to be reduced. Hidden neurons are responsible for the absorption of progressive knowledge, allowing the execution of more complex tasks. Finally, it is worth emphasizing that due to the network's high connectivity any modification requires it to be restructured (Haykin, 2014).

The metrics to evaluate the ANNs configuration perfomance are MAE (Mean Absolute Error), MPE (Mean Percent Error) and MSE (Mean Square Error), calculated as shown in the equations below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_{esp} - X_{obs}| \tag{1}$$

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{esp} - X_{obs}}{X_{esp}}\right| \tag{2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|X_{esp} - X_{obs}|^2 \tag{3}$$

where, in all equations presented above, X_{esp} represents the expected value and X_{obs} the value returned by the ANN.

3. DESIGN OF EXPERIMENTS

Design of Experiments (DoE) is a methodology for studying any response that varies as a function of one or more independent variables, called factors, based on ANOVA (analysis of variance). Statistical design of experiments refers to the process of planning experiments to collect appropriate data in order to study a process or system. The most significant advantage to using DoE is the ability to quickly detect how interactions between factors can affect the process and determine the significant parameters.

The significance of the parameters is determined through hypothesis testing. Hypothesis testing is the process of using statistics to determine the probability of whether the proposal hypothesis is true. In this case, the null hypothesis is that there is no significant correlation between the parameters and the alternative hypothesis is that there is. If the statistical test result is positive, this doesn't mean that the alternative hypothesis is true, but rather means that any evidence to disprove the alternative hypothesis was found. For example, using the hypothesis test considering a 95% confidence interval those with p-value greater than 0.05 are eliminated (Mathews, 2005; Montgomery, 2013; Pierson, 2015).

Factorial designs are a important class of experimental designs that are widely used in research works. Among other reasons, its importance is linked to the fact that they form the basis of other considerable designs. Full factorial designs consider all possible combinations of the level of the factors. In contrast, a Box-Behnken design has a lower necessity of experiments, estimation of the parameters of the quadratic model and build of sequential designs. Box-Behnken design is classified as a Response Surface Methodology (RSM) and has demonstrated to be more efficient than other methods as the Central Composite design. It is worth mentioning that factorial is not the most efficient way to model a quadratic relationship, and in this cases response surface designs are superior alternatives (Ferreira *et al.*, 2007; Montgomery, 2013).

The designed number of essays *N* of each methodology is shown by Eq.(4) for Box-Behnken design and by Eq.(5) for three full level factorial design (Ferreira *et al.*, 2007; Montgomery, 2013):

$$N = 2k(k-1) + Co \tag{4}$$

$$N = 3^k \tag{5}$$

where k is the number of factores and C_O are the center points. It is possible to notice that factorial designs include a lot of experiments comparing to Box-Behnken designs. As the number of factors increases this difference becomes more significant.

4. DESCRIPTION OF THE SYSTEM

Steam generators indicate machines equipped with super heaters, reheaters, economizers and air heaters. In coal-fired power plants or any other electric generation station load curves consider both power and steam requirements (Annaratone, 2008).

The system in analyses is the steam generator of PECEM power plant which is responsible for 50% of the energy generation complex in Ceará, Brazil. The power plant is composed of two independent groups. Each of the two superheated steam generators is a 360 MWe sub-critical, coal fired single furnace unit. The furnace operates under balanced draught conditions; with natural circulation and steam reheat. The boiler has a parallel back end forming two separate gas passes for the primary superheater and reheater banks (EDP, 2019).

Pulverized fuel is introduced to the furnace via twenty four Low NOx Axial Swirl Burners. Twelve after-air ports are provided for NOx reduction. The burners are arranged in two rows of six each on the furnace front and rear walls. The after air ports are arranged in two single rows of six each above the top rows of pulverized fuel burners. The pulverized fuel burners are each equipped with co-axial light fuel oil burners which provide for the boiler light up and flame stabilization. The oil burners are able to fire the boiler up to a load of 30% boiler maximum continuous rating (EDP, 2019). A schematic layout is presented in Fig.1.

5. METHODOLOGY

The methodology proposed in this paper focuses on applying DoE to an ANN model to analyze the system's behavior, select and rank the input parameters according to their order of significance. The methodology was constructed based on the structure presented in Fig. 2.

In the first step, data were collected to develop an ANN model. Data processing is an essential step for getting accurate results from the model. Data must be queried, summarized and visualized before and after training the models. Therefore, the data were plotted to search for the existence of any special pattern, as well as the presence of outliers, variation and

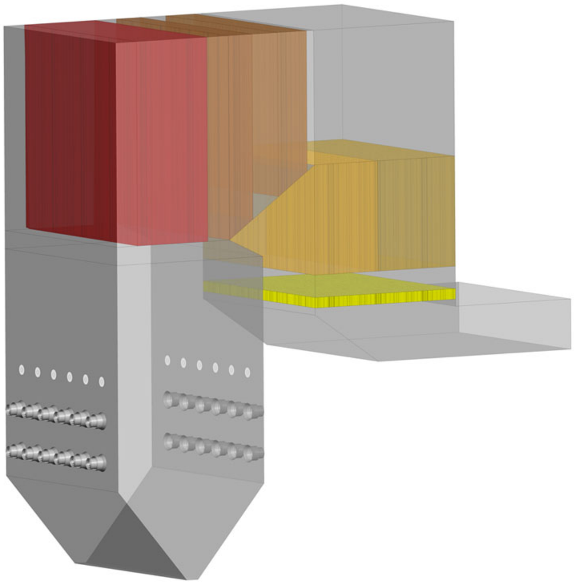

Figure 1. Steam generator schematic layout

Figure 2. Flowchart of the proposed method

distribution. The evaluation was made according to three characteristics, which are location (central tendency), variation (dispersion) and shape. The ANNs hyperparameters (number of hidden layers, number of hidden neurons per each hidden layer and activation functions) were defined through an interactive approach in search of best describing the problem at

Table 1. Model input parameters with their ranges selected for the Design of Experiments project

	Primary air flow* **(P1)**	Pulverized coal outlet temperature* **(P2)**	Velocity of the dynamic classifier* **(P3)**	Excess O2 **(P4)**	Primary air's crossover duct pressure **(P5)**	Secondary air's crossover duct pressure **(P6)**	Coal flow **(P7)**
High bound	24	65	80	2.00	10.0	51	27.0
Intermediate bound	26	75	95	2.75	18.5	62	38.5
Low bound	28	85	110	3.50	27.0	73	50.0
Unit	kg/s	°C	rpm	%	mbar	mbar	ton/h

* Parameter refers to the mills.

Table 2. Model output parameters

	Flue gas outlet temperature **(S1)**	Steam generator efficiency **(S2)**	Electric power generation **(S3)**
Unit	°C	%	MW

hand. Hyperparameters configurations were tested by a trial and error method guided by doubling the number of hidden neurons on each try (Mathews, 2005).

The input parameters were chosen from a large data set due to its controllability by the operators. Controllable parameters are those that can be directly manipulated by manual command and present an independent behavior among each other. The input parameters, also called factors by the DoE methodology, are described on Tab. 1. To facilitate further analysis visualization, the primary air flow, pulverized coal outlet temperature, velocity of the dynamic classifier, excess O_2, primary air's crossover duct pressure, secondary air's crossover duct pressure and coal flow will be called, respectively, P1, P2, P3, P4, P5, P6 and P7.

The outputs chosen in this paper are the flue gas outlet temperature, boiler efficiency and electric power generation. These outputs are not direct controllable parameters of the power plant, they are responses subjected to the different configurations of the input set of parameters. These outputs, or responses, were calculated by the ANN and are described in Tab. 2. These responses will be called S1, S2 and S3.

Once the ANN model is well established, a DoE is applied to show the correlation between the input and the output parameters considered. Two approaches for DoE were chosen, Box-Behnken and three full level factorial designs.

The parameters should be selected one step at a time, according to their statistical significance. The high order terms and the interactions between different input parameters are eliminated first. The significance considered a 95% confidence interval, then based on the hypothesis testing the terms with p-value greater than 0.05 must be eliminated. For hypothesis testing, the residuals are assumed to be normally and independently distributed random variables with mean and variance zero. If there are non-random patterns in the residuals, it means that probably the predictor are missing something. The simplest model that produces random residual its one of the assurances of a precise and unbiased model. Residual plots were used to check this assumptions.

After finalizing the parameter selection, these controllable parameters were ranked by order of importance according to their influence on each response. Thenceforth, a new model is proposed focusing on the selected parameters. Comparison between the models to check the error and prove the advantages of using a simplest model to predict the same responses. The best conditions to the responses are defined according to the new model.

6. RESULTS

Plant operating data is constantly collected over time through its data acquisition and supervision system. The supervisory system enables real-time visualization of the plant as well as the download of its history.

Initially, it was selected a data set consisting of the 10 parameters presented earlier on Tab. 1 and Tab. 2 stored every half hour within the period of January 2018 to May 2019. The only parameter in this group that is not directly measured on site is the steam generator efficiency that is calculated through other secondary parameter measures available.

Every measurement is subject to imperfections that reflect inadequate data. Therefore, the data set was analyzed and preprocessed to remove gross errors and outliers such as negative and null observations. The data was also filtered by electric power generation to reflect the 340 to 365 MW range.

The data set of 10 parameters and 6033 samples were randomized and divided into 70% training, 25% testing and validation, and 5% as a sample unseen by the ANN to be used for further analysis. The parameters were standardized by their standard deviation.

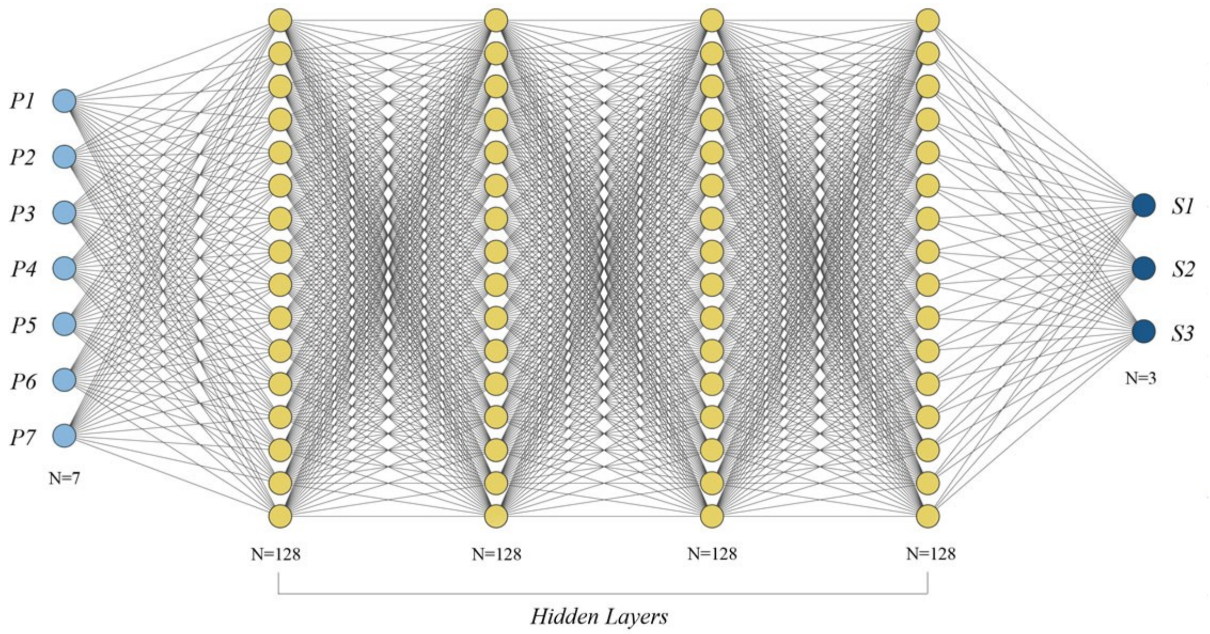

Figure 3. Chosen topology for ANN

The ANNs were developed using Python programming language through the Jupyter Notebook compiler. It was used for its construction the Keras programming interface provided by the Tensorflow machine learning library.

The topology of the ANN's hyperparameters tested followed the approach presented in the methodology, evaluating the performance of the combinations of 8, 16, 32, 64, 128 and 256 hidden neurons with 1 to 4 hidden layers. The number of neurons in each hidden layer was the same and the activation function used was the ReLU (Rectified Linear Unit).

Among the ANN's tested, the chosen one was built with one input layer, four hidden layers of 128 nodes each and one output layer. Its topology can be seen in Fig. 3. The standardized MSE and MAE of the test were respectively 0.2015 and 0.2741.

A DOE methodology was applied to analyze the control parameters of the steam generator modeled with the ANN, respecting the ranges listed on Tab. 1. The DoE choses a set of values as inputs and the ANN representing the steam generator's behavior offers the system response in order to feedback the DoE. Two DoE projects were chosen for a comparative and the respective operational details are shown in Tab. 3. The analyses were performed using the software Minitab®.

Table 3. Design of Experiments operational details

Box-Behnken			
Number of factors k	7	Replication	1
Number of essays	62	Total number of essays N	62
Number of blocks	1	Center points C_O	6
Three Level Full Factorial			
Number of factors k	7	Replication	1
Number of essays	2187	Total number of essays N	2187
Number of blocks	1	Center points C_O	0

Its possible to notice that for the same quantity of input parameters, a three level full factorial requires a larger amount of essays when with a Box-Behnken design. Even so, the ANN's fast response enables this comparative analyses. The system's responses are analyzed individually.

The first assessment was performed to identify the responses behavior in respect of the controlled parameters. Their main effects are evaluated separately for each response. The first results for S1 are shown in Fig. 4 for Box-Behnken and three level full factorial.

The difference between the two DoE results and be clearly visualized when looking at the graphics. The top set of results, (a), were generated by the Box-Behnken methodology and present the effect of curvature while for the three level full factorial shown in (b) the results are given considering only linear relations. The response S1 varies significantly with all the parameters according to three level full factorial, while Box-behnken metholody consider P1 and P3 as not

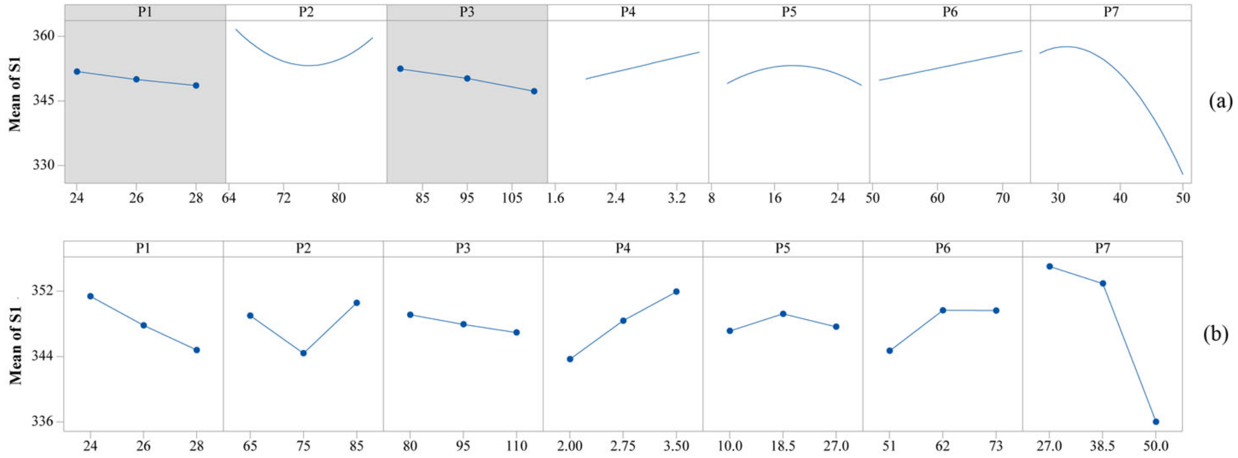

Figure 4. Main effects of the controlled parameters on the response S1 (a) Box-Behnken (b) three level full factorial

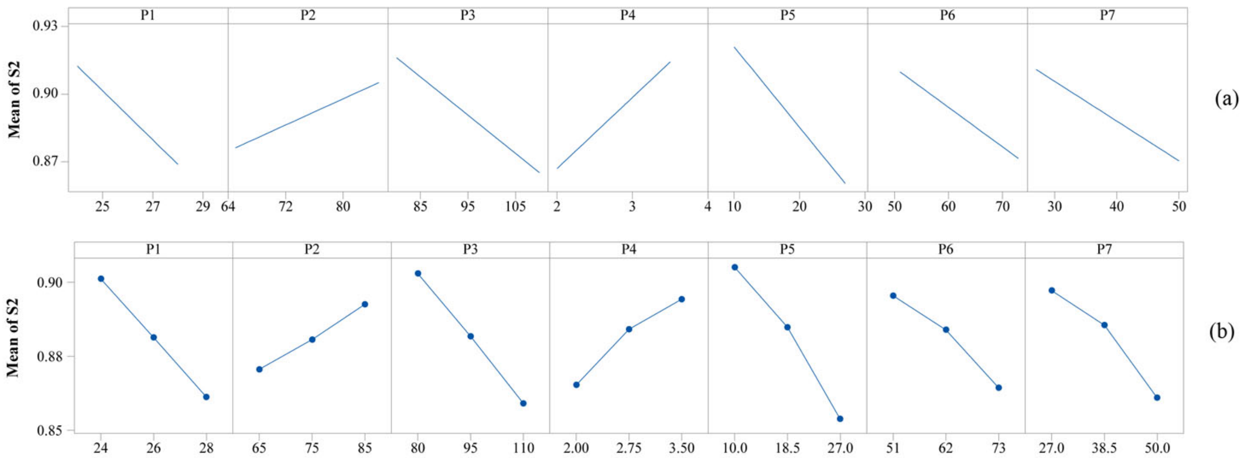

Figure 5. Main effects of the controlled parameters on the response S2 (a) Box-Behnken (b) three level full factorial

statistically significant (painted in gray at the figure). Even so, parameters behaviors and tendencies are the same when comparing the models. The increase of the parameters P4 e P6 leads to an increase on the response S1, showing a linear relationship. In contrast, the parameters P2, P5 and P7 have a non-linear relationship with the response.

The results found for S2 can be seen in Fig. 5. (a) shows the Box-Behnken analysis and (b) the three level full factorial.

In this case, both methods showed statistical significance and linear relationships between the parameters and the response S2. There is a positive correlation between the parameters P2 and P4 with S2. On the other side, a negative correlation its noticed among the other parameters and S2. Therefore, any increase on P1, P3, P5, P6 and P7 corresponds to a decrease on S2.

For the last response S3, the analysis results can be seen in Fig. 6. (a) again represents Box-Behnken and (b) the three level full factorial.

The difference between the two DoE models is emphasized due to the non-linearity behavior of the parameters with S3. Only the parameters P2 and P7 have positive linear relationships while P1 has a negative linear relationship. The parameter P5 has a huge influence on the response, noticeable on both approaches.

The next step consists of analyzing the combined effects among the parameters. A combined effect means that any adjustment in one parameter will affect the others. Combined effect analysis is essential because while sometimes a parameter is not statistically significant by itself, when analyzed together with the whole model it must be considered. Up to 6 way interactions in three level full factorial designs and 2 way interactions in Box-Behnken designs were analyzed in this study. Due to the massive amount of graphics generated by the interaction plots, these won't be presented in this paper. However, it is valid to stress that all the interactions were considered in both analysis.

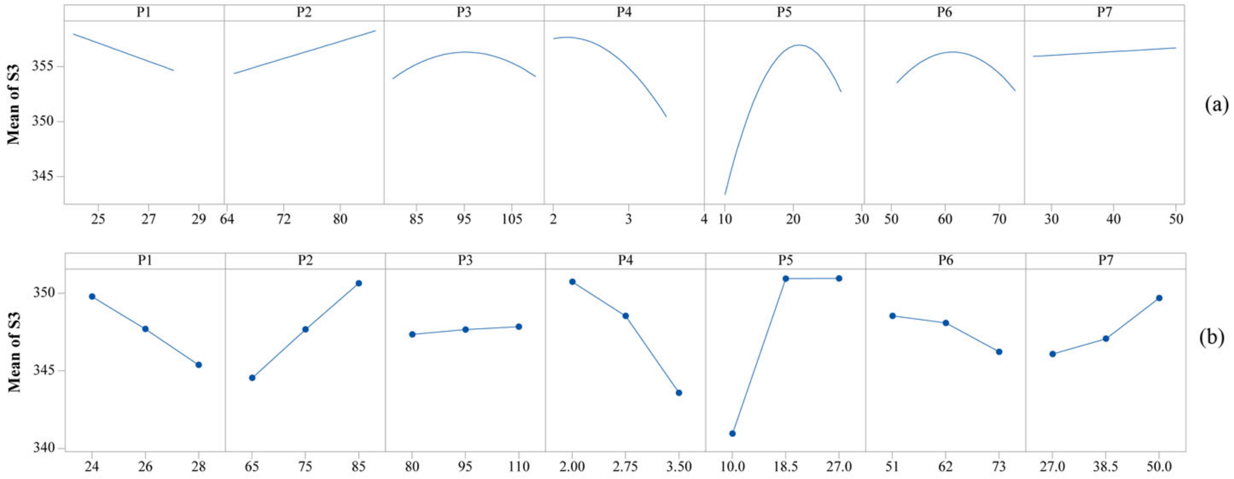

Figure 6. Main effects of the controlled parameters on the response S3 (a) Box-Behnken (b) three level full factorial

The residual plots were checked to guarantee normality, independency and constant variance. The squared correlation coefficient (R^2) of the ANN model represents the assurance of the equation developed through the DoE analysis. The R^2 adjust reveals the power of regression models that contain different numbers of control parameters and closing its value to the R^2 indicates high accuracy of the equation, in other words, how well the independent variables describe the dependent variable . Predicted R^2 measures the predict quality of the model. Tab. 4 presents the results.

Table 4. Summary of the squared correlation coefficients

	Box-Behnken			**Three level full factorial**		
	S1	S2	S3	S1	S2	S3
R^2	79.46%	81.66%	91.51%	99.79%	99.93%	99.85%
R^2 adjust	75.43%	77.63%	87.67%	99.26%	98.79%	99.32%
R^2 predictive	65.42%	72.20%	78.44%	97.32%	79.33%	96.88%

It is worth mentioning that both methodologies presented good results. The three level full factorial is the one with better and more precise results according to Tab. 4 However, by being a more thorough methodology, it required 35 more essays when compared to box-behnken. When dealing with an experimental approach, the number of essays to be considered is a crucial element to implement the study or not.

By applying DoE to the ANN model, it was revealed the significance of each control parameter through hypothesis testing. Therefore, it was perceived the response S1 was not affected by the parameters P1 and P3, even though the responses S2 and S3 were found to be affected by all parameters. Once the statistically significant parameters were determined, the next step was to rank them by order of importance. The ranking is presented if Fig. 7.

	Flue gas outlet temperature (S1)	**Steam generator efficiency (S2)**	**Electric Power Generation (S3)**
1	Coal flow (P7)	Primary air's crossover duct pressure (P5)	Primary air's crossover duct pressure (P5)
2	Secondary air's crossover duct pressure (P6)	Velocity of the dynamic classifier (P3)	Excess O2 (P4)
3	Excess O2 (P4)	Excess O2 (P4)	Pulverized coal outlet temperature (P2)
4	Pulverized coal outlet temperature (P2)	Primary air flow (P1)	Primary air flow (P1)
5	Primary air's crossover duct pressure (P5)	Coal flow (P7)	Coal flow (P7)
6	*N.A.**	Secondary air's crossover duct pressure (P6)	Secondary air's crossover duct pressure (P6)
7	*N.A.**	Pulverized coal outlet temperature (P2)	Velocity of the dynamic classifier (P3)

Figure 7. Parameters Ranking

The scale shows the level of importance of the parameters, representing the most important ones at the green upper side of the arrow and the least important ones at the red bottom of the arrow. Among the set of parameters studied, the coal flow (P7) presented itself as the most influential parameter for the flue gas outlet temperature (S1) response. As already discussed before, the primary air flow (P1) and velocity of the dynamic classifier (P3) were not statistically significant and therefore were not presented in the ranking. For both the steam generator efficiency (S2) and electric power generation

(S3), the primary air's crossover duct pressure (P5) was found to be the most important parameter. It can be noted that the ranking order was quite variable among the responses.

A new ANN model was then developed to estimate only the response S1. The same development method used for the first ANN was applied for the one and the same hyperparameters topology was tested. The second ANN for S1 prediction, was built with one input layer, four hidden layers of 64 nodes each and one output layer. The standardized MSE and MAE of the test were respectively 0.3759 and 0.4318.

After both ANNs were trained and validated, the sample data that was separated earlier from the rest of the dataset is now entered in both ANNs developed in order to analyze the models performances when dealing with new data. The comparison between the two ANNs studied can be seen in Tab. 5.

Table 5. ANNs comparison

	Original ANN		**Remodeled ANN**	
	MPE	MSE	MPE	MSE
Flue Gas Outlet Tempetature **(S1)**	0.320%	2.350	0.492%	5.548
Boiler Efficiency **(S2)**	0.452%	0.267	*N.A.**	*N.A.**
Electric Power Generation **(S3)**	0.208%	1.115	*N.A.**	*N.A.**

* Not applicable.

The MPE and MSE for S1 of the second ANN developed is larger when compared to the first ANN presented in this paper, but still low and extremely adequate to the problem addressed. Moreover, it should be noted that the second network topology is less complex with half of the first ANN's neurons per hidden layer. Since this is a problem applied to a real steam generator, the difference between a field operator needing to control seven parameters or five is crucial. The best conditions given by different configurations seek to achieve a minimum value for S1 and a maximum value for S2 and S3.

7. CONCLUSION

The main focus of this paper was to apply statistical analysis through DoE methodology in order to select and rank the most influential parameters of a steam generator modeled by an ANN. The relevance of this study is that by having identified the importance of each controllable parameter, it enables the operator on the power plant to understand and accurately manipulate the right parameters in order to achieve a new, safe, stable and more efficient condition.

The analysis includes as inputs the controllable parameters: primary air flow, pulverized coal outlet temperature, velocity of the dynamic classifier, excess O2, primary air's crossover duct pressure, secondary air's crossover duct pressure and coal flow. The responses, or outputs, were the steam generator efficiency, electric power generation and flue gas outlet temperature.

Two approaches to the DoE analysis were made, through Box-Behnken and Three level full factorial methodologies. DoE was implemented at the development of the ANN, searching to define acceptable ranges and significant parameters. First exploration pointed out all the parameters interactions. For the response S1, the parameters P1 and P3 were found not to be relevant while the responses S2 and S3 are influenced by all parameters. Finally, the ranking of the parameters by order of importance was obtained. The most important parameter for the flue gas outlet temperatura (S1) is the coal flow. For steam generator efficiency (S2) and electric power generation (S3) the primary air's crossover duct pressure (P5) was the most important parameter. The rankings for each of the responses were quite variable analyzing the whole set of controllable parameters.

The original ANN developed to model the steam generator with the entire data set presented standardized testing MAE and MSE of 0.2015 and 0.2741 and MPE and MSE for the unseen sample of 0.32% and 2.350. The remodeled ANN built to predict S1 with the five controllable parameters found to be significant showed standardized testing MAE and MSE of 0.3759 and 0.4318 and MPE and MSE for the unseen sample of 0.492% and 5.548. A simplest ANN to predict the same response allows the operators to concentrate their efforts to control the really impacting parameters for the operation of the power plant.

8. ACKNOWLEDGMENTS

Authors acknowledge Energy of Portugal EDP for the financial and technical support to this project; Vieira acknowledges CNPq– Brazilian National Council for Scientific and Technological Development,for her MSc grant; Smith Schneider acknowledges CNPq for his research grant (CNPq-PQ 305357/2013-1).

9. REFERENCES

Annaratone, D., 2008. "Steam Generators: Description and Design".

Bekat, T., Erdogan, M., Inal, F. and Genc, A., 2012. "Prediction of the bottom ash formed in a coal-fired power plant using artificial neural networks". *Energy*, Vol. 45, No. 1, pp. 882–887. doi:10.1016/J.ENERGY.2012.06.075.

EDP, 2019. "UTE PECÉM".

Ferreira, S., Bruns, R., Ferreira, H., Matos, G., David, J., Brandão, G., da Silva, E., Portugal, L., dos Reis, P., Souza, A. and dos Santos, W., 2007. "Box-Behnken design: An alternative for the optimization of analytical methods". doi:10.1016/j.aca.2007.07.011.

GE, 2017. "Data Science Services from GE Digital".

Hall, P., Phan, W. and Whitson, K., 2015. *The Evolution of the Technologies of Analytics*. Media, O'Reilly, 1st edition. ISBN 9781491954713. doi:10.1002/9781119198284.ch6.

Haykin, S., 2014. *Neural Networks and Learning machines*. ISBN 9780131471399. doi:978-0131471399.

IEA, 2019. "Statistics & Data". URL `https://www.iea.org/`.

Lujan-Moreno, G.A., Howard, P.R., Rojas, O.G. and Montgomery, D.C., 2018. "Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study". doi: 10.1016/j.eswa.2018.05.024.

Mathews, P.G., 2005. *Design of Experiments with Minitab*. American Society for Quality, Milwaukee. ISBN 0-87389-637-8.

McCulloch, W.S. and Pitts, W., 1943. "A logical calculus of the ideas immanent in nervous activity". *The bulletin of mathematical biophysics*, Vol. 5, No. 4, pp. 115–133. ISSN 1522-9602. doi:10.1007/BF02478259. URL `https://doi.org/10.1007/BF02478259`.

MME, 2018. "Balanço Energético Nacional".

Montgomery, D., 2013. *Design and Analysis of Experiments*. John Wiley & Sons, 8th edition.

Pierson, L., 2015. *Data Science for dummies*. John Wiley & Sons, Hoboken.

Pirhadi, N., Tang, X., Yang, Q. and Kang, F., 2018. "A new equation to evaluate liquefaction triggering using the response surface method and parametric sensitivity analysis". doi:10.3390/su11010112.

Smrekar, J., Assadi, M., Fast, M., Kuštrin, I. and De, S., 2009. "Development of artificial neural network model for a coal-fired boiler using real plant data". *Energy*, Vol. 34, No. 2, pp. 144–152. ISSN 03605442. doi: 10.1016/j.energy.2008.10.010.

Smrekar, J., Pandit, D., Fast, M., Assadi, M. and De, S., 2010. "Prediction of power output of a coal-fired power plant by artificial neural network". doi:10.1007/s00521-009-0331-6.

Strušnik, D., Golob, M. and Avsec, J., 2015. *Artificial neural networking model for the prediction of high efficiency boiler steam generation and distribution*, Vol. 57. doi:10.1016/j.simpat.2015.06.003.

Tunckaya, Y. and Koklukaya, E., 2015. "Comparative prediction analysis of 600 MWe coal-fired power plant production rate using statistical and neural-based models". *Journal of the Energy Institute*, Vol. 88, No. 1, pp. 11–18. doi: 10.1016/J.JOEI.2014.06.007.

Weissman, S.A. and Anderson, N.G., 2015. "Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications". doi:10.1021/op500169m.

10. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.