Energy and AI 3 (2021) 100040

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.elsevier.com/locate/egyai

Methodology for ranking controllable parameters to enhance operation of a steam generator with a combined Artificial Neural Network and Design of Experiments approach

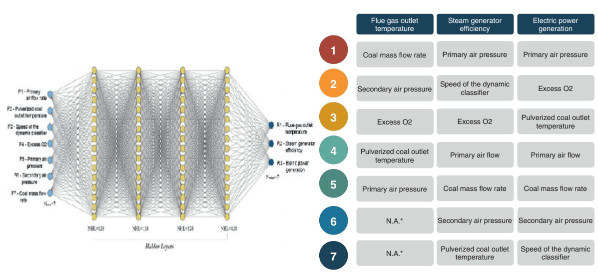
Lara Werncke Vieira[a,*], Augusto Delavald Marques[a], Paulo Smith Schneider[a], Antônio José da Silva Neto[b], Felipe Antonio Chegury Viana[c], Madhat Abdel-jawad[d], Julian David Hunt[a,e], Julio Cezar Mairesse Siluk[f]

[a] Federal University of Rio Grande do Sul - UFRGS, Porto Alegre, Brazil
[b] State University of Rio de Janeiro - UERJ, Nova Friburgo, Brazil
[c] University of Central Florida, Orlando, UCF, United States
[d] Advanced Analysis Australia - AAA, Perth, Australia
[e] International Institute for Applied Systems Analysis - IIASA, Vienna, Austria
[f] Federal University of Santa Maria - UFSM, Santa Maria, Brazil

HIGHLIGHTS

- Modeling of an existing coal-fired power plant with 360 MW in Brazil using real data.
- A combined approach of power plant design with artificial neural networks (ANN).
- Identification of the most relevant process parameters of the steam generator.
- Two Design of Experiment models are applied to compare the performance.
- Definition of the best operating ranges using Response Surface Methodology (RSM).

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 24 September 2020
Received in revised form 3 December 2020
Accepted 12 December 2020
Available online 16 December 2020

Keywords:
Coal-fired power plant
Artificial Neural Network
Design of Experiments
Response surface methodology
Steam generator

ABSTRACT

The operation of complex systems can drift away from the initial design conditions, due to environmental conditions, equipment wear or specific restrictions. Steam generators are complex equipment and their proper operation relies on the identification of their most relevant parameters. An approach to rank the operational parameters of a subcritical steam generator of an actual 360 MW power plant is presented. An Artificial Neural Network - ANN delivers a model to estimate the steam generator efficiency, electric power generation and flue gas outlet temperature as a function of seven input parameters. The ANN is trained with a two-year long database, with training errors of 0.2015 and 0.2741 (mean absolute and square error) and validation errors of 0.32% and 2.350 (mean percent and square error). That ANN model is explored by means of a combination of situations proposed by a Design of Experiment - DoE approach. All seven controlled parameters showed to be relevant to express both steam generator efficiency and electric power generation, while primary air flow rate and speed of the dynamic classifier can be neglected to calculate flue gas temperature as they are not statistically significant. DoE also shows the prominence of the primary air pressure in respect to the steam generator efficiency, electric power generation and the coal mass flow rate for the calculation of the flue gas outlet temperature. The ANN and DoE combined methodology shows to be promising to enhance complex system efficiency and helpful whenever a biased behavior must be brought back to stable operation.

* Corresponding author.

https://doi.org/10.1016/j.egyai.2020.100040

Nomenclature

DoE	Design of Experiments
ANN	Artificial Neural Network
Exp	Expected (actual value)
Obs	Observed (calculated with the ANN)
MAE	Mean absolute error
MPE	Mean percentual error
MSE	Mean square error
k	Number of factors
N	Total number of essays
C_O	Center points
LHV	Lower heating value
KKS	equipment identification codes
F1	Primary air flow rate, kg/s
F2	Pulverized coal outlet temperature, $°C$
F3	Speed of the dynamic classifier, rpm
F4	Excess O_2, %
F5	Primary air pressure, mbar
F6	Secondary air pressure, mbar
F7	Coal mass flow rate, ton/h
R1	Flue gas outlet temperature, $°C$
R2	Steam generator efficiency, %
R3	Electric power generation, MW
DCS	Distributed Control System
N HL	Number of neurons in the hidden layer
N input	Number of neurons in the first layer
N output	Number of neurons in the last layer

1. Introduction

Coal fuels approximately 40% of the world's electric supply, which has been growing by nearly 900 GW since 2000 [1,2]. The superheated water steam cycle is the most common technical solution for solid fuels like coal, nuclear and as well as renewable sources, such as sugar cane and solid waste, which increase the interest on enhancing plant performance and safety operation.

Operational data from coal-fired power plants are usually continuously acquired and available, allowing to better understand the system behaviour. Approaches based on pattern recognition and parametric correlation can allow for process optimization by aligning available data, efficient management and strategy, based on constant monitoring [3,4].

Different levels of modelling steam generators have been developed based on physical phenomena, but data based algorithms showed to be an attractive option as they are capable of modelling sophisticated systems with lesser effort but keeping their complexity representation. These models are trained with large amounts of actual data to find sufficient patterns that enable accurate decisions about the system parameters [5]. Studies have already succeeded in modeling steam generators by machine learning techniques. Romeo and Gareta [6] applied Artificial Neural Networks (ANN) to develop a methodology for a biomass boiler monitoring, concluding that the ANN can predict the operational parameters, as well as the fouling state of the boiler. Rusinowski and Stanek [7] used two ANN to calculate the flue gas and unburned losses. A model to predict a soot-blowing routine by ANN was presented by Shi et al. [8]. Also other authors used it to precidct boiler emissions like NOx [9–11].

ANN has been used to the integration of steam power plant components aiming to improve the overall performance of power plants [12,13]. ANNs were applied to entropy generation minimization of a combined heat and power system [14]. Also, the power production of a power plant was predicted using ANN considering as input the ambient temperature [13]. The real data on the amount of the generated steam in the existing system boilers was compared to the results of the model and results were used to analyze coal consumption savings and their impact on the environment. Navarkar et al. [15] studied the relationship between load cycling and the variations of the superheater outlet pressure, reheater inlet temperature, and flue gas temperature at the air heater inlet. An ANN trained with the data of the previous 10 years was able to predict these values for the next 10 h.

The studies found that apply ANN to steam generators focus on obtaining an architecture that provides a certain output with low value for the loss function, but there is little concern about how to implement the results in an operation.In this context, an ANN model linked with the control system of a power plant can guide the operator's decision making which will ensure an increase in efficiency along with the plant's stability.To enable the application of the model that aims to improve the operation or efficiency of a steam generator, it is necessary to study the controllability and impact of the parameters used as input of the model.

As an auxiliary tool for assessing any system behavior, the statistical methodology known as Design of Experiments – DoE enables to investigate cause and effect relations and to identify the influence of the input parameters on the system responses. Parameters can be individually analyzed and also their crossed interactions, allowing to propose models that can be used for improvements and support decision making [16,17]. The DoE can be applied in a wide range of processes. Kanimozhi et al [18] applied DoE and ANN to model and validate a thermal energy storage system, achieving the ranking factor for the charging process. Choi et al. [19] used DoE to identify and study the effect from controlling variables on thermal deformation in automative body parts.

The literature on power plants shows that it is possible to identify and model their behavior of these systems, but their operation in practice remains a field of development. The operation is subject to environmental factors, sensitivity to input variations, unexpected events and human aspects, which generate the need to propose coordinated and standardized actions. Based on this observation, this article proposes a methodology for ranking operating parameters that indicates ordered actions to maintain systems performance and to assure operational stability. The methodology is based on statistical analysis by applying a DoE approach to a system model built by neural networks. The case study presented is an actual 360 MW coal-fired power plant, but it can be extended to systems with identified control parameters.

2. Artificial Neural Network – ANN

The ANN gathers information from the environment through data. The Multi-Layer Perceptron (MLP) architecture houses an input layer, an output layer, and intermediate layers called "hidden" layers. The MLP model stands out for three main characteristics: nonlinear activation function, hidden neurons, and high degree of connectivity. Hidden neurons are responsible for the absorption of progressive knowledge, allowing the execution of more complex tasks [20–22].

The metrics to evaluate the ANNs configuration performance are the mean absolute error MAE, the mean percentual error MPE, and the mean square error MSE, as used by [13], are presented in Eq. (1–3):

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

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

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

with X_{exp} the output expected or actual value and X_{obs} its value calculated with the ANN.

3. Design of Experiments - DoE

DoE is a statistical methodology for studying any kind of system whose responses varies as a function of one or more independent parameters, called controllable factors, based on analysis of variance (ANOVA). The methodology allows planning experiments to collect appropriate data out of actual or modeled processes and systems. Changes in the average response due to factor swiping within a defined range or level is defined as an effect. Factors vary within ranges according to a defined number of levels which includes at least the level high and low. An interaction among factors is identified when the effect of one factor on the response depends on the level of some other factor. Interactions can occur between two, three, or more factors but three-factor interactions and beyond are usually assumed to be insignificant. The parameter significance is determined through hypothesis testing [16,17,23].

The three principles of experimental design, namely randomization, replication and blocking, can be utilized to improve the efficiency of experimentation, applied to reduce or even remove experimental bias [17]. The purpose of randomization is to remove all sources of extraneous variation which are not controllable in real-life settings. Replication means repetitions of an entire experiment or a portion of it, under more than one condition. Blocking is a method of eliminating the effects of extraneous variation due to noise factors and thereby improving the efficiency of experimental design. The idea is to arrange similar or homogeneous experimental runs into groups, called blocks [16,23].

Full factorial design is an important class of assessment procedure, which enables to evaluate individual effects and possible interactions of several factors, instead of the one-factor-at-a-time method. Its high number of combinations can lead to expensive and time consuming experiments, that can be reduced by choosing a Box-Behnken design, as one possible option. The designed number of essays N for each methodology, considering k factors, and C_O center points, is shown in Eq. (4) for a full three level factorial design, and in Eq. (5) for a Box-Behnken design [17,24]:

$$N = 3^k \tag{4}$$

$$N = 2k(k-1) + Co \tag{5}$$

4. System description

The PECEM coal-fired power plant was chosen to perform an assessment whose goal was to select and rank system parameters in order to better operate the plant. The power plant is located near the ocean coast of the State of Ceará, Brazil, composed of three identical and independent power groups. Each group is designed to produce 360 MW out of Colombian coal with a lower heating value (LHV) about 25,750 kJ/kg, burned on a sub-critical steam generator. The furnace operates under balanced drought conditions; with natural circulation and steam reheat. A parallel back end splits flue gas flows through the primary superheater and the reheater exchangers [25,26]. A schematic layout of the steam generator and its coupled coal mills is presented in Fig. 1.

Preheated air stream coming from an external heat recovery device at approximately 300°C is split into two feeding paths, the primary and

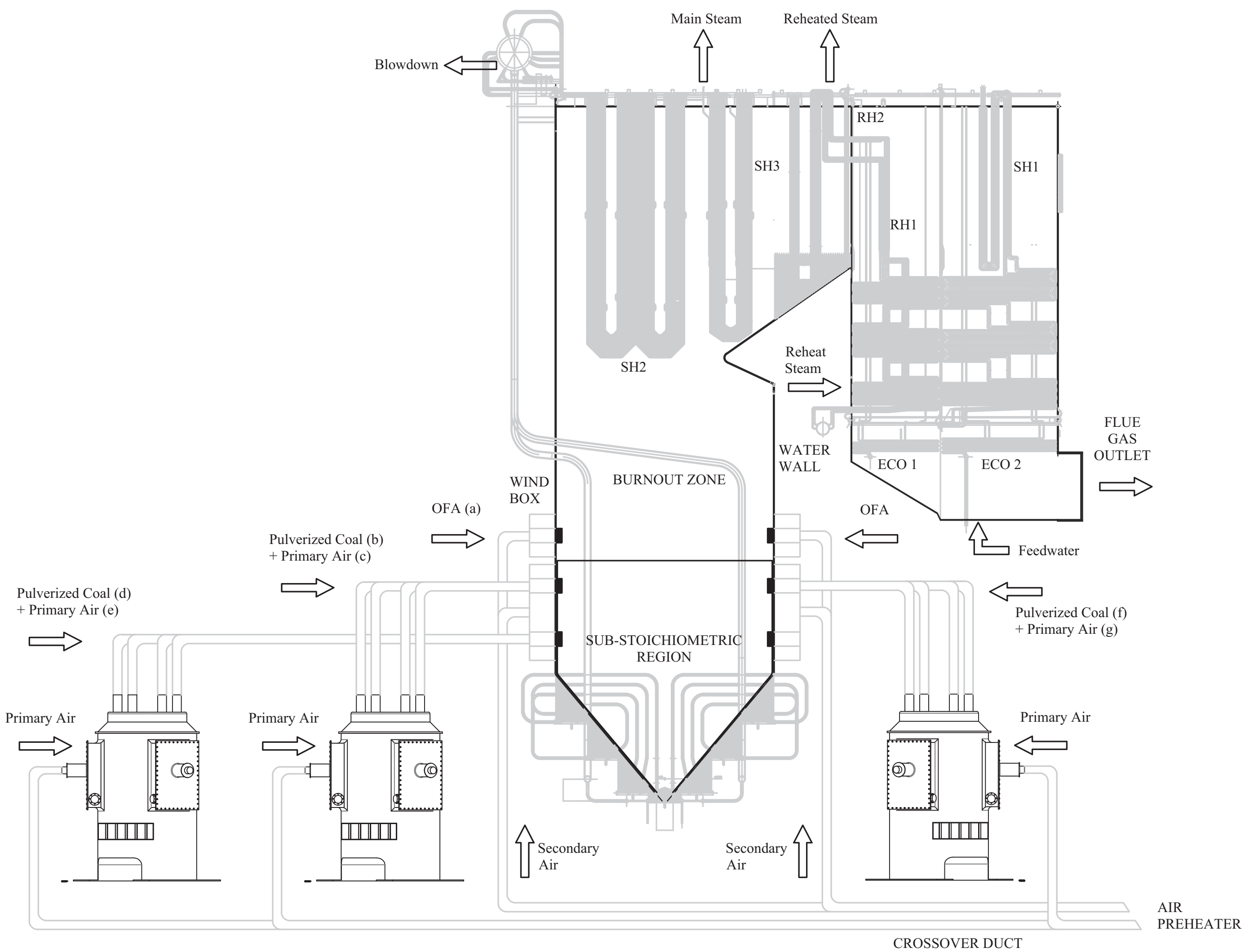


Fig. 1. Steam generator schematic layout (UTE PECÉM, Brazil).

secondary air flows. Primary air is admitted in the mill to both perform coal drying and transport it to the steam generator burners. Each mill feeds a burner line of six pulverized coal combustors or burners, placed in independent wind boxes. The pulverized fuel and the primary air are introduced into the furnace via a combination of twenty four Low NOx Axial Swirl Burners (letters b to g in Fig. 1) according to the load level, under sub-stoichiometric conditions. Combustion is completed on the furnace upper zone by twelve over fire air ports (OFAs, ports a in Fig. 1).The feedwater arrives at 276 æC and 168 bara, the output superheated steam at 538 æC drives the vapour cycle.

5. Methodology

The methodology strategy to select and rank the input parameters according to their order of significance is presented in Fig. 2.

Data processing is priorly performed in the first step to search for and identify the existence of special patterns, outliers, variation, and distribution [23]. An statistical test is performed to analyze the parameters and their respective ranges of operation. The input parameters are selected based on their controllability, which means, they can be directly impacted by the actions of the unit control operator.

The second step is dedicated to system modeling through ANNs. ANNs hyperparameters (number of hidden layers, number of hidden neurons per each hidden layer, and activation functions) are defined through an iterative approach that is intended to best describe the problem at hand. Hyperparameter configurations are tested by a trial and error method guided by doubling the number of neurons in the hidden layers on each try. The first ANN was developed with the simplest configuration, a single hidden layer. New networks were further on tested by doubling both the number of hidden layers and the number of neurons per layer. The simplest ANN with the best results is selected. The errors for the test and validation datasets are compared, in order to achieve the lowest error values for both datasets and ensure that there is no overfitting.

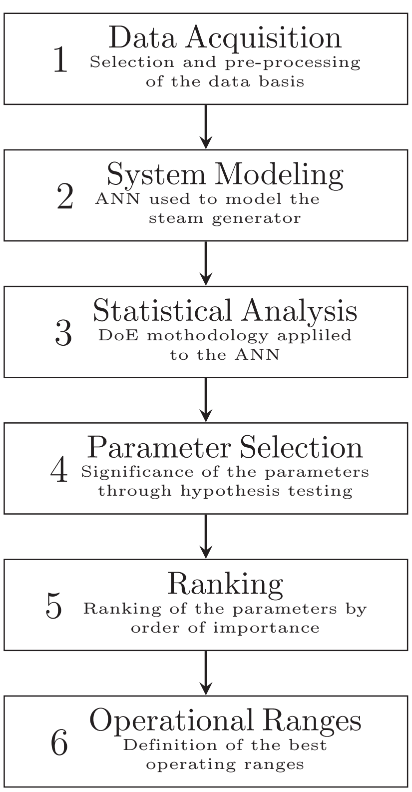


Fig. 2. Methodology strategy to select and rank the steam generator operational ranges.

The selected ANN algorithm is employed in the third step to evaluate the steam generator behavior by applying the DoE methodology. In the present work, both the three full level factorial and the Box-Behnken designs were tested. Parameter selection in the fourth step can be performed out of the results obtained in the prior step by hypothesis testing using ANOVA. The residual plots were checked to guarantee the ANOVA assumptions of a normal distribution, independence, and constant variance.

In step 5, the mathematical model produced by the DoE method was used to rank the parameters by order of importance according to each model response. Predicted coefficient of determination (R^2) was used to evaluate the prediction quality of the DoE mathematical model. Finally, the last step identifies the operating ranges in which the factors lead to the best possible system response.

6. Results and discussions

The controlled parameters were identified by means of three parallel and complementary sources: actual data and from the power station labeling system (KKS), list of parameters considered as significant to controllable losses on textbooks and technical standards, and advising from the PECEM in site technical staff. The list with 7 relevant controllable parameters and 3 system responses is presented in Tab. 1.

The primary air flow rate (F1) performs two prior functions, namely to dry the raw coal and convey it to the burners, already pulverized, whose amount is controlled by (F7), the coal mass flow rate. The speed of the dynamic classifier (F3) allows to select the fuel granulometry or pulverization level. Pulverized coal outlet temperature (F2) is measured at the mill outlet and it is related to the coal drying process. The steam generator is divided into two burner volumes, the sub-stoichiometric region with 4 rows of 6 burners each and the burnout zone, as showed in Fig. 1. The secondary air flow rate guaranties sub-stoichiometric combustion conditions, but it is not directly manipulated by the operator, which explains its exclusion as an ANN input.

The combustion total air is the summation of the primary, secondary, and over-firing air flows, and its global stoichiometry is kept approximately constant about 1.2. The excess of O2 (F4) is measured at the burnout zone and it indicates the global stoichiometry of the combustion process. Hot air flow from the air preheater serves both the primary and secondary streams via two independent systems, called the crossover ducts, in which we have as the input of the ANN the primary and secondary air pressure (F5 and F6). The output parameters flue gas outlet temperature (R1), steam generator efficiency (R2), and electric power generation (R3) were chosen for the system behavior representation.

The power plant Distributed Control System (DCS) continuously acquired the half-hour mean values of the parameters data during operation. The survey of equipment uncertainty data, measurement interval and calibration documents were carried out for all parameters. The DCS records only a variation above 0.5% of the previous value.

The complete dataset runs from January 2018 up to May 2019 in this work. Negative and null values were removed and then filtered with respect to the 340–365 MW range of electric power generation. This filter resulted in a set of 6033 records, which represents approximately 20% of the original dataset. The dataset was randomized and divided into 70% training, 25% testing, and 5% for validation [20]. Parameters were standardized with respect to their correspondent standard deviation.

ANNs were developed (step 2) using the Keras [27] programming interface running on top of the Tensorflow machine learning library [28].

The topology of the ANN hyperparameters was evaluated by performing combinations of 8, 16, 32, 64, 128, and 256 hidden neurons applied to each of the 4 hidden layers. The tested activation functions included ReLU (Rectified Linear Unit) and Tanh (hyperbolic tangent). ReLu is a typical activation function for MLP, especially to guarantee that the output will always be positive [21]. The investigation process

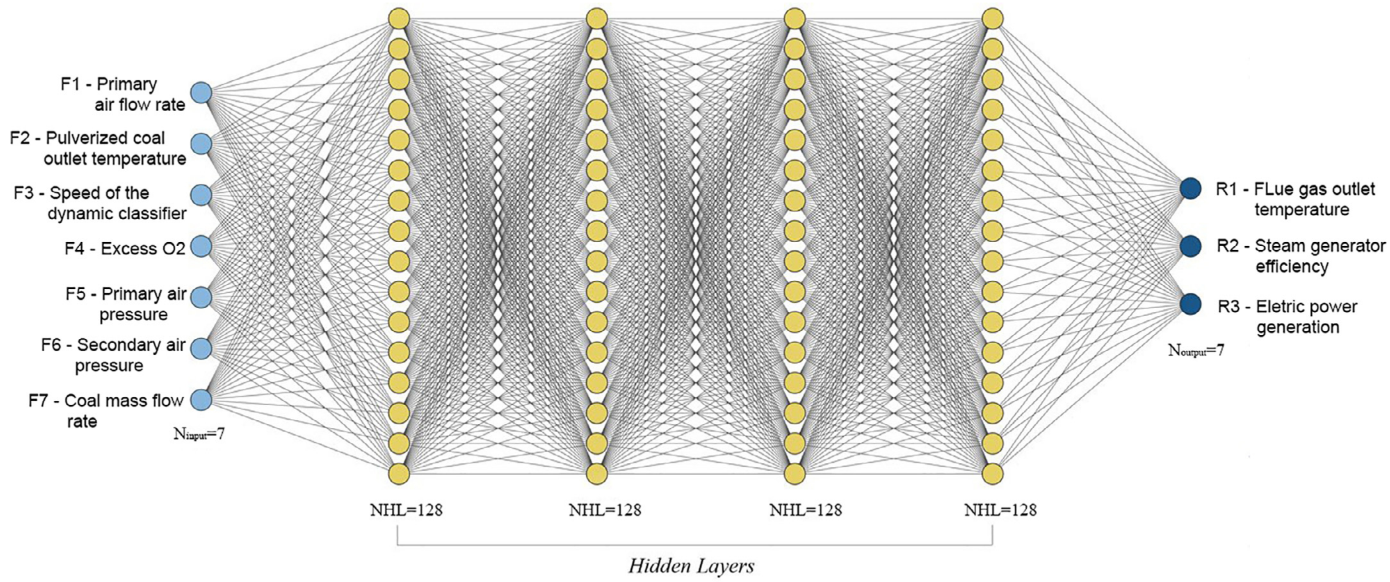


Fig. 3. Chosen topology for the ANN - the parameters details are presented in Table 1.

Table 1
Input and output parameters for the ANN model.

Input (controllable parameters)		Unit
Primary air flow rate	F1	kg/s
Pulverized coal outlet temperature	F2	°C
Speed of the dynamic classifier	F3	rpm
Excess O_2	F4	%
Primary air pressure	F5	mbar
Secondary air pressure	F6	mbar
Coal mass flow rate	F7	ton/h
Outputs (system responses)		**Unit**
Flue gas outlet temperature	R1	°C
Steam generator efficiency	R2	%
Electric power generation	R3	MW

started with the simplest ANN with 8 hidden neurons and one hidden layer. After that, the number of neurons was doubled as well as the hidden including a set of different combinations until 256 hidden neurons and 4 hidden layers. The main idea is to achieve the simplest ANN capable to represent our problem in analysis. Table 2 presents some of the tested ANNs.

The selected ANN was built with one input layer, with $N_{input} = 7$, corresponding to F1–F7, as shown in Table 1, four hidden layers of $N_{HL} = 128$ neurons each, and one output layer, with $N_{output} = 3$, corresponding to outputs (system responses). The ANN architecture is presented in Fig. 3.

Step 3 concerns the statistical analysis of the steam generator behavior simulated with the aid of the ANN algorithm. The ANN statistical metrics MAE and MSE were 0.2015 and 0.2741 with respect to the test data set, respectively. DoE was applied to the ANN according to the operational ranges of the selected input parameter as described in Tab. 3.

The operating ranges were determined according to the plant history and with the assistance of the PECEM technical team to provide safe and stable conditions. Simple data analysis did not allow to indicate if the power plant was running under expected conditions. Variability on coal moisture due to the rain, or unusual equipment behavior, for instance, cannot be observed with this approach. Thus, experimental investigation through DoE becomes essential because it performs a comprehensive analysis on the coupling of the operational parameters. Parameter values were kept within the range limits of regular operation. The plant ANN algorithm was tested by both the Box-Behnken and the three level Full Factorial designs, and details are shown in Table 4.

The three-level full factorial approach required a larger amount of essays when compared with the Box-Behnken design. Even so, the ANN fast response enabled to perform both approaches, presented hereafter to clarify their individual advantages. The first assessment was performed to identify the effect of each input parameter on the system responses, displayed separately.

Results for the flue gas outlet temperature R1 are shown in Fig. 4 for both the Box-Behnken and three-level full factorial approaches.

Parameter behavior and tendencies were quite the same when comparing the models. Relations were found to be close to linear for F4 and F6, and non-linear for F2, F5, and F7. Inputs F1 and F3 showed to be statistically not significant (gray boxes) with respect to the flue gas outlet temperature, according to the Box-Behnken model (a), whereas all parameters are relevant to the three-level full factorial model (b). This evaluation was made using hypothesis tests with a 95% confidence level. Results out of the Box-Behnken model are displayed with smooth curves while the three-level full factorial shown can only linearly link dots. Significant factors and interactions were selected by searching terms with p-value< α=0.05 according to the ANOVA. The high order terms and the interactions between different input parameters were eliminated first and the final model is a result of several model reduction iterations. The Table 6 in the Appendix presents the Analysis of variance (ANOVA) for the complete model with all linear, square, and interaction terms.

A similar assessment was performed for the steam generator efficiency R2 whose results are presented in Fig. 5.

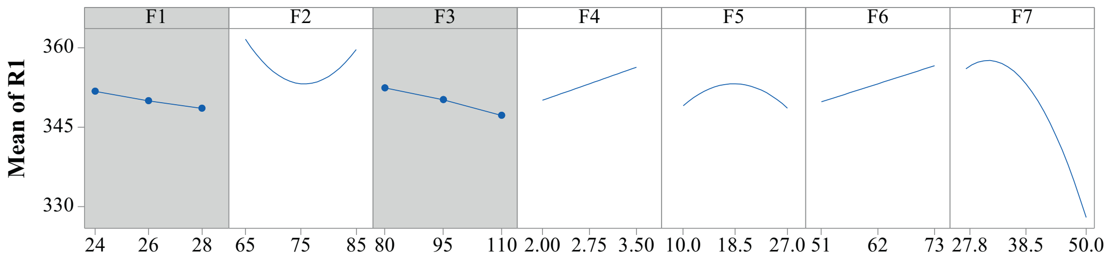
Both methods showed statistical significance and linear relationships between the parameters with respect to the steam generator efficiency R2. Direct correlations were found for parameters F2 and F4 and inverse ones for all others in respect to R2. The assessment of the electric power generation R3 is presented in Fig. 6.

The difference between the two DoE designs is emphasized due to the non-linearity behavior of the parameters with respect to R3. F2 and F7 displayed a positive relationship with the response while F1 displayed a negative relationship. F5 presented the highest influence on the response, noticeable on both approaches due to its span.

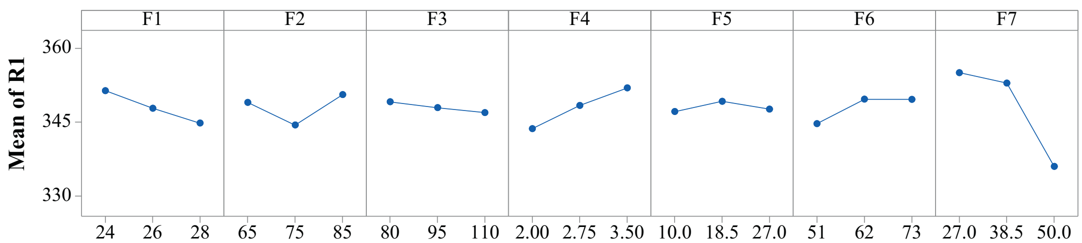
The next analysis of the fourth step (Fig. 2) consists of analyzing the interactions among factors, identified when the effect of one factor on the response depends on the level of some other factor. The present study focused on the analysis of 6-way interactions for the three-level full factorial design and 2-way interactions for the Box-Behnken design. All the 2-way interactions are presented in Figs. 7–9.

Table 2
Subset of the tested ANNs - Backpropagation learning algorithm and Multi-Layer Perceptron network type for 200 epochs with a batch size of 256.

ANN model	1	2	3	4
Hidden neurons	64 - 64 -64	64 - 64 -64	128 - 128 - 128 - 128	16 - 32 - 32 - 32
Hidden layers	3	3	4	4
Activation function	ReLU	Tanh	ReLU	Tanh - RelU
Training dataset size	4223	4223	4223	4223
Testing and validation dataset size	1810	1810	1810	1810
MAE train	0.2804	0.2505	0.1263	0.3447
MAE test	0.4287	0.3077	0.2741	0.388
MSE test	0.3537	0.2174	0.2015	0.4343

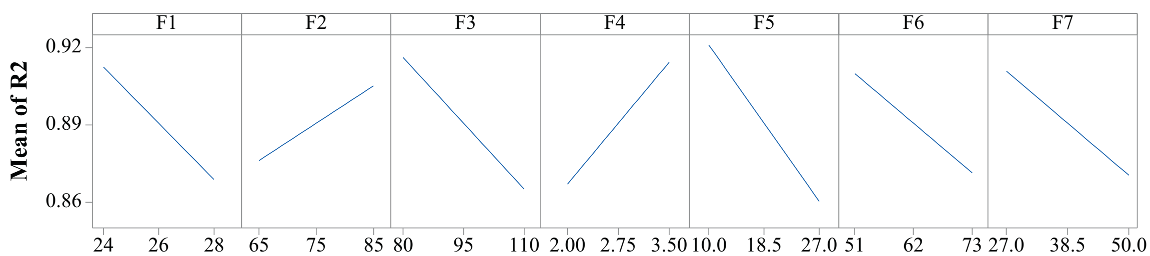


(a) Box-Behnken

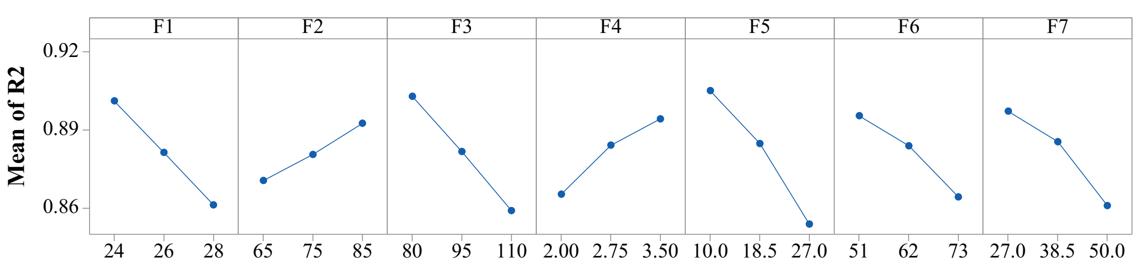


(b) Three level full factorial

Fig. 4. Main effects of the controlled parameters on the flue gas outlet temperature R1 with (a) Box-Behnken and (b) Three level full factorial.

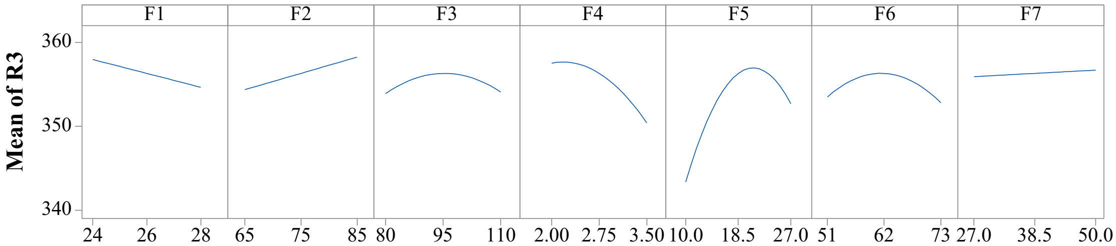


(a) Box-Behnken

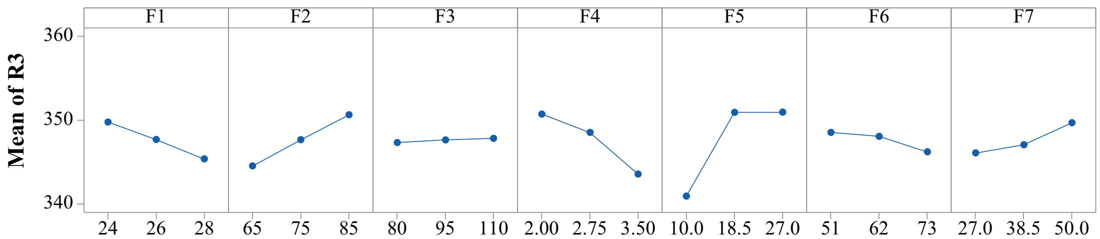


(b) Three level full factorial

Fig. 5. Main effects of the controlled parameters on the steam generator efficiency R2 with (a) Box-Behnken and (b) Three level full factorial.



(a) Box-Behnken



(b) Three level full factorial

Fig. 6. Main effects of the controlled parameters on the electric power output R3 with (a) Box-Behnken and (b) Three level full factorial.

Table 3
Model input parameters with their ranges selected for the Design of Experiments (DoE) project.

	F1*	F2*	F3*	F4	F5	F6	F7
Low level	24	65	80	2.00	10.0	51	27.0
Intermediate Level	26	75	95	2.75	18.5	62	38.5
High level	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 4
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

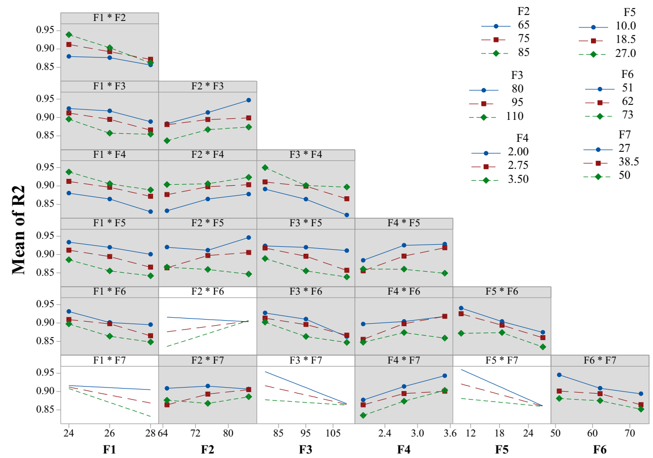


Fig. 8. Interaction plot for the response steam generator efficiency (R2).

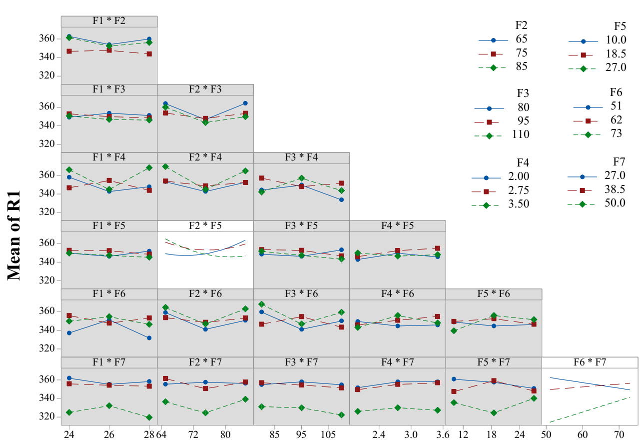


Fig. 7. Interaction plot for the response flue gas outlet temperature (R1).

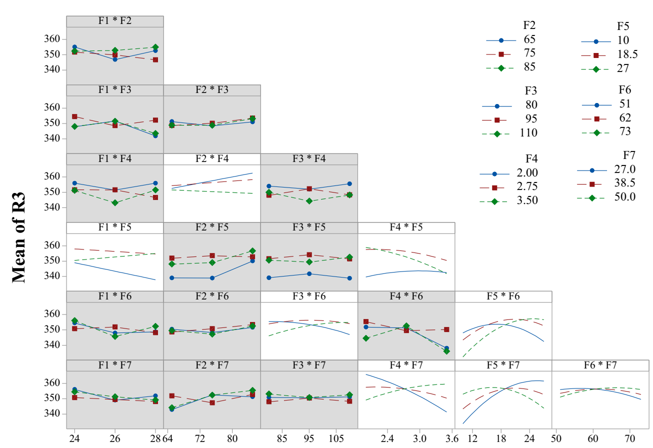


Fig. 9. Interaction plot for the response electric power output (R3).

Table 5
Summary of the coefficient of determination R^2.

	Box-Behnken			Three level full factorial		
	R1	R2	R3	R1	R2	R3
R^2	79.46%	81.66%	91.51%	99.79%	99.93%	99.85%
R^2 adjusted	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%

The crossing of the lines indicates that the interaction is significant, since the change in the level of the factor caused a change in the behavior of the other factor, altering its impact on the output. The levels are represented by the colors blue (low level), red (intermediate level), and green (high level). The behavior of the pulverized coal outlet temperature (F2) changes according to the three levels of the primary air pressure (F5). Based on the graph of F2xF5 (Fig. 7), if F5 = 10mbar, when F2 increases the output flue gas outlet temperature (R1) also increases. On the other hand, if F5 = 18.5mbar or F5= 27.0mbar, if F2 increases the output R1 decreases. The primary air pressure is directly related to the entry of primary air into the mill, which performs the drying of the coal and increases its temperature. The same occurs for the interaction between secondary air pressure (F6) and coal mass flow rate (F7). If F6 = 51mbar, as F7 increases the response R1 decreases.

The coal mass flow rate (F7) presents significant interactions with three other factors, namely the primary air flow rate (F1), speed of the dynamic classifier (F3), and secondary air pressure (F6). The impact on efficiency is proportional to the amount of coal the primary air needs to drag to the burners. It is possible to notice that the efficiency and performance of the steam generator are directly related to the performance of the mills.

The electric power output is the response with the greatest influence of cross-terms of parameters interaction. This response varies according to the whole power plant performance and for this reason, interactions are more significant.

The Tab. 5 presents the results of the coefficient of determination (R^2) as the prediction quality of the model considering Box-Behnken and three-level full factorial design, regarding each of the three responses: flue gas outlet temperature (R1), steam generator efficiency (R2), and electric power generation (R3).

The adjusted R-squared takes into account the number of predictors (factors) in the model, and it is lower than the R-squared. The predictive R-squared indicates how the model predicts the response for new observations. According to Tab. 5, the three-level full factorial displayed the highest values for the squared correlation coefficients. This result was expected due to the robustness of this design, which required 35 times more essays when compared to Box-Behnken (see Table 4). Dealing with an experimental approach, the number of essays to be considered can be a crucial element to implement the study or not. For this reason, the comparative analysis was carried out, in order to check the capability of Box-Behnken design to represent model tendency despite the huge difference in the required number of essays.

Hypothesis testing revealed the significance of each control parameter, which showed that the response of the flue gas outlet temperature R1 was not affected by the parameters F1 and F3, even though responses R2 and R3 were found to be affected by all parameters. The next step of the methodology concerned the parameter ranking by order of importance, as presented in Fig. 10.

The scale from 1 to 7 classifies the parameters in order of decreasing importance. The ranking order was quite variable as the positions of the parameters vary according to the response. Among the set of studied parameters, the coal mass flow rate (F7) presented itself as the most influential parameter for the flue gas outlet temperature (R1) response. In contrast, the primary air pressure (F5) was found to be the most important parameter for both the steam generator efficiency (R2) and electric

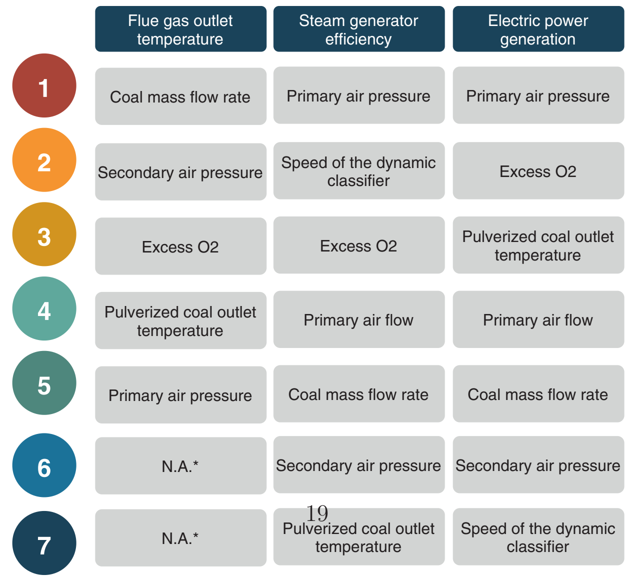


Fig. 10. Parameter ranking according to their impact on the flue gas outlet temperature (R1), steam generator efficiency (R2), and electric power generation (R3) responses.

power generation (R3). The primary air flow rate (F1) and speed of the dynamic classifier (F3) were not statistically significant for the flue gas outlet temperature (R2), and, therefore, were not presented in the ranking.

Since this is a problem applied to a real steam generator, make process controls adjustments, based on process history and parameter ranking, enables the right insight into all variability issues that interplay along the process. Such information provides guidance for engineers and operators to perform changes aiming at better operating conditions.

The last step of the proposed methodology consists on defining the operating ranges corresponding to the best response condition within the ranges defined in Table 3. That was performed using a Response Surface Methodology through Box-Behnken design since the previous analyses evidenced the same results tendency for Box-Behnken and three full factorial projects.

The contour plots presented in Fig. 11 represent the responses ranges based on the most impacting parameters. Two parameters for each response were selected while the others were kept constant. The graphics are represented by ranges of the response where the light green regions stand for the higher values achievable by each response considering the limits of the inputs.

The best conditions given by different configurations seek to achieve a minimum value for R1 and a maximum value for R2 and R3. The non-linear relationship of the parameters F2 and F7 with R1 reflects on its contour plot in Fig. 11 (a). For R2 and R3, the linear relationships are maintained as shown respectively in Fig. 11 (b) and (c). Each graphic contains the parameters ranges according to Tab. 3. It must be noted that for the linear relationships the increase of the input control parameters implicates the increase of the response. On the other hand, when dealing with a non-linear relationship as seen in Fig. 11 (a) there can be more than one region for the maximum response. In this case, the maximum possible can be achieved by the combination of low values for both F2 and F7 or low values of F7 and high values of F2. Clearly such results may be incorporate into the power plant control procedures.

The savings due to the increase in efficiency can be calculated through the efficiency equation by the direct method [29] for the steam generator. A 1.02 % efficiency gain leads to a saving up to 12,000 tons of coal per year and can reduce up to 3% of CO_2 emissions [30].

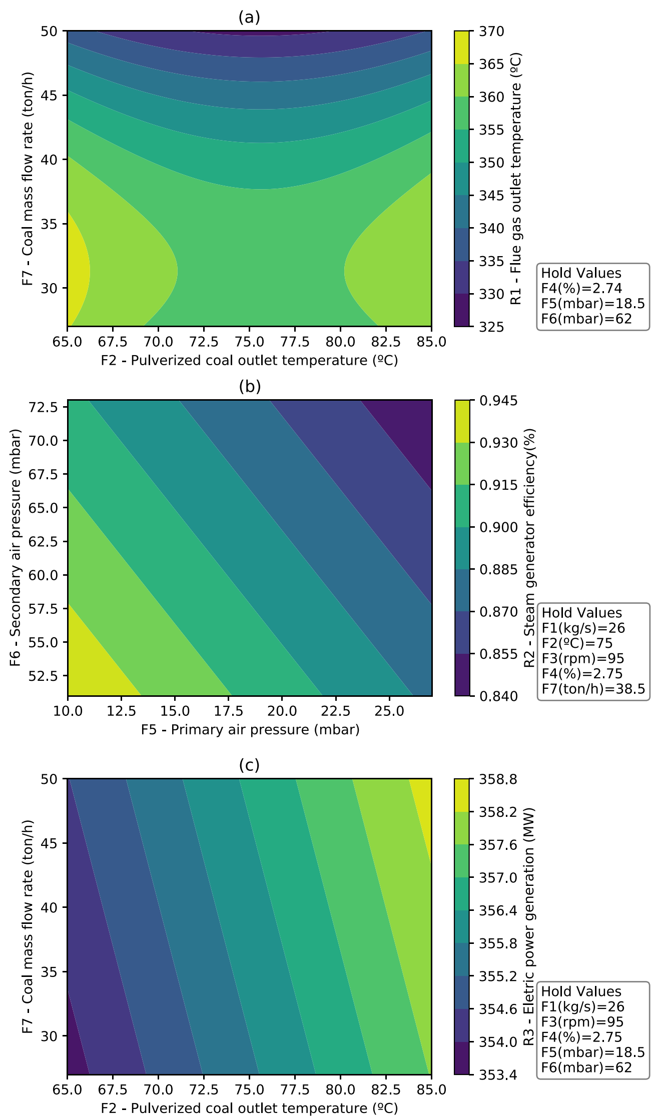


Fig. 11. Contour plots to the responses flue gas outlet temperature R1(a), steam generator efficiency R2 (b), and electric power generation R3 (c).

7. Conclusion

The main novelty brought in this work was the proposal of an approach to enhance the operational quality of a real complex system based on the identification of the distance from the actual operational conditions to the desired one, defined a priori by design. The Design of Experiments - DoE approach organized a set of maneuvers based on sweeping controllable operational parameters along their secure range of values. The system main responses were the flue gas outlet temperature, the steam generator efficiency, and the electric power generation.

In site experiments weren't available and the system was modeled with an Artificial Neural Network - ANN. The ANN model presented MAE and MSE of 0.2015 and 0.2741 for the test data set, and MPE and MSE of 0.32% and 2.350 for validation, respectively. That combined methodology allowed to rank the operational parameters of the steam generator and mills, and pointed out that the coal mass flow rate as the most relevant parameter with respect to the flue gas outlet temperature, while the primary air pressure was the most important parameter for both the steam generator efficiency and the electric power generation.

The present approach allows the identification of the controllable parameter's importance and its smooth-running range. It can also guide the power plant operator by helping him to understand and accurately manipulate the right parameters in real-time, in order to achieve a new, safe, stable, and more efficient condition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Authors acknowledge Energy of Portugal EDP for the financial and technical support to this project; Vieira acknowledges the INCT-GD and the financial support from CAPES 23038.000776/2017-54 for her PhD grant; Marques acknowledges the financial support from CNPq 132422/2020-4 for his MSc grant; Smith Schneider acknowledges CNPq for his research grant (PQ 305357/2013-1).

Appendix A

A1. Analysis of variance

In Table 6 DF, Adj SS, and Adj MS correspond to total degrees of freedom, adjusted sums of squares, adjusted mean squares respectively. The F-value is a test statistic while the p-value is a probability that measures the evidence against the null hypothesis.

Table 6
Analysis of variance (ANOVA) for the complete model with all linear, square and interactions terms for the response R1 through Box-Behnken Design.

Source	DF	Adj SS	Adj MS	*F*-Value	*P*-Value
Model	35	10935.6	312.45	5.98	0
Linear	7	5511.7	787.39	15.07	0
P1	1	62.8	62.83	1.2	0.283
P2	1	22.5	22.49	0.43	0.517
P3	1	162.5	162.47	3.11	0.090
P4	1	234	234.03	4.48	0.044
P5	1	1.20	1.16	0.02	0.883
P6	1	279.3	279.27	5.35	0.029
P7	1	4749.50	4749.5	90.92	0
Square	7	3370.8	481.54	9.22	0
P1*P1	1	30.8	30.82	0.59	0.449
P2*P2	1	556.3	556.3	10.65	0.003
P3*P3	1	55.8	55.78	1.07	0.311
P4*P4	1	123.7	123.68	2.37	0.136
P5*P5	1	395.4	395.43	7.57	0.011
P6*P6	1	131.9	131.95	2.53	0.124
P7*P7	1	2027.7	2027.74	38.82	0
2-Way Interaction	21	2053.1	97.77	1.87	0.065
P1*P2	1	2.8	2.77	0.05	0.82
P1*P3	1	19.7	19.70	0.38	0.544
P1*P4	1	78.6	78.65	1.51	0.231
P1*P5	1	21.9	21.87	0.42	0.523
P1*P6	1	2.2	2.21	0.04	0.839
P1*P7	1	1.5	1.50	0.03	0.867
P2*P3	1	57.0	57.00	1.09	0.306
P2*P4	1	8.0	8.01	0.15	0.699
P2*P5	1	552.3	552.29	10.57	0.003
P2*P6	1	24.0	23.97	0.46	0.504
P2*P7	1	1.7	1.70	0.03	0.858
P3*P4	1	73.6	73.55	1.41	0.246
P3*P5	1	87.3	87.34	1.67	0.207
P3*P6	1	0.4	0.42	0.01	0.929
P3*P7	1	38.9	38.90	0.74	0.396
P4*P5	1	10.7	10.72	0.21	0.654
P4*P6	1	38.8	38.80	0.74	0.397
P4*P7	1	13.9	13.89	0.27	0.61
P5*P6	1	107.9	107.93	2.07	0.163
P5*P7	1	107.5	107.48	2.06	0.163
P6*P7	1	804.5	804.45	15.4	0.001

A2. Contour plots

The contour plots in Figs 12–14 display response surfaces as a two-dimensional plane with response isolines. Graphs are assembled by pairs of factors, while all others parameters are hold at their average values.

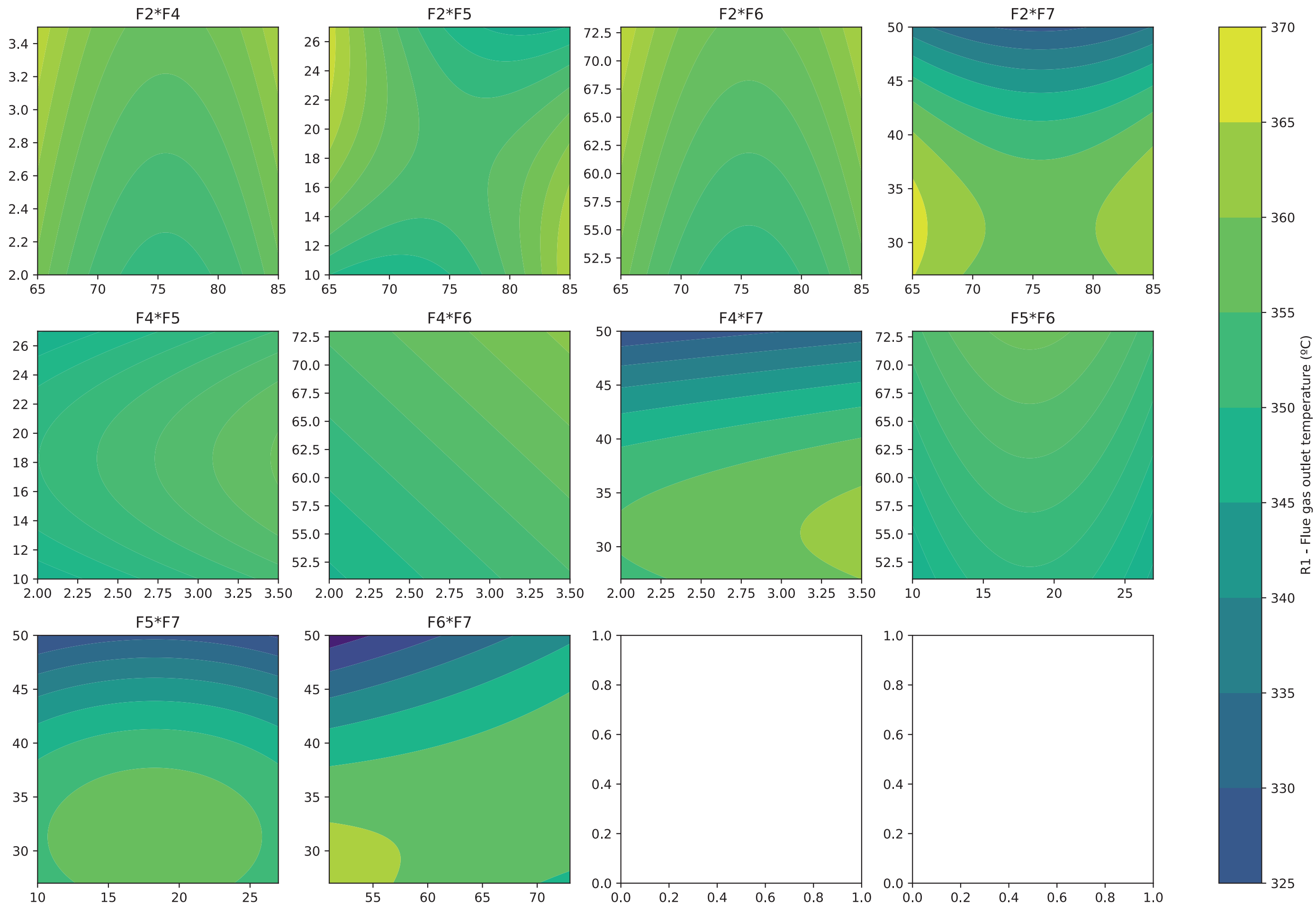


Fig. 12. Contour plots of the pairs of combined factors for the response flue gas outlet temperature (R1).

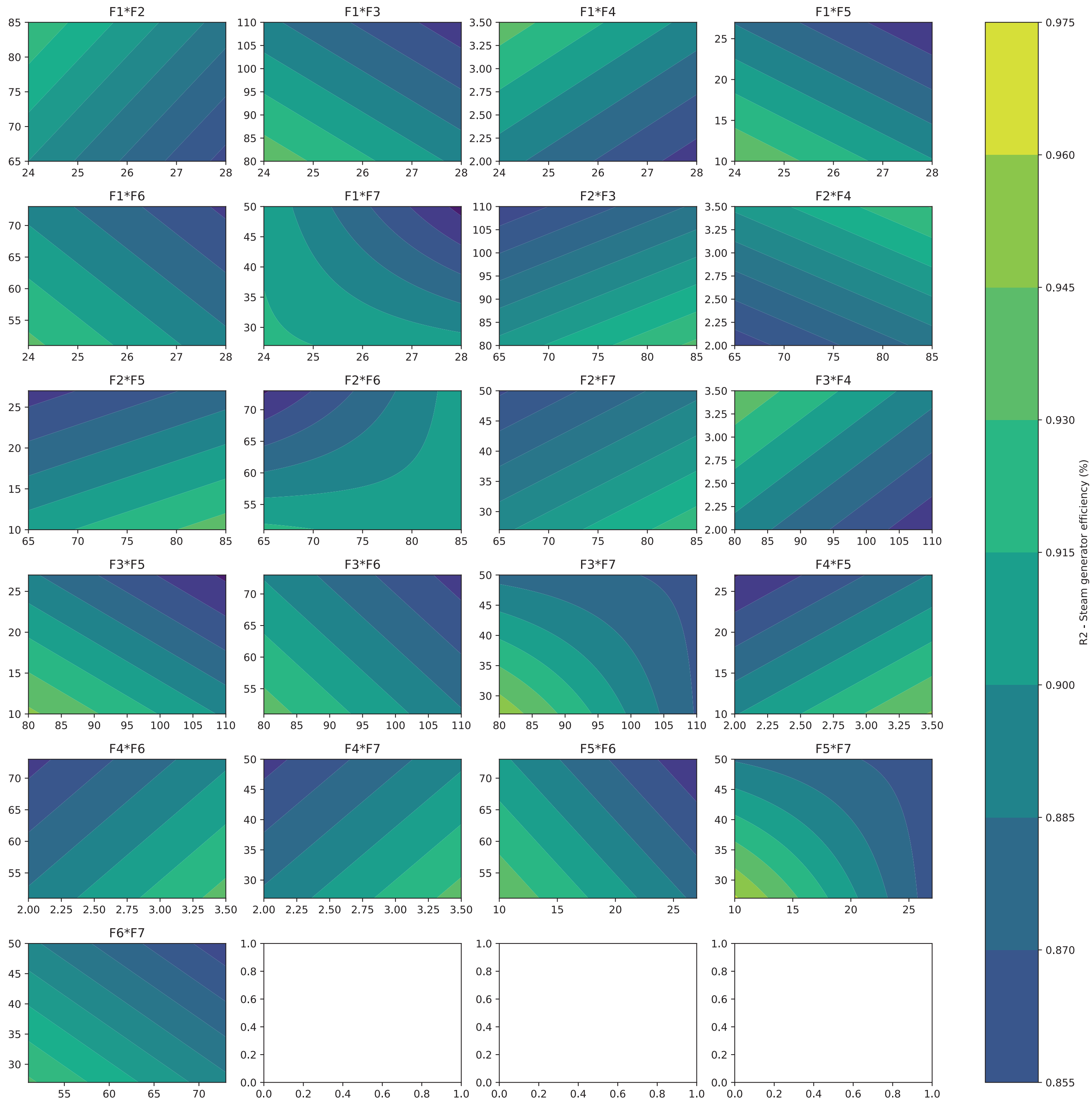


Fig. 13. Contour plots of the pairs of combined factors for the response steam generator efficiency (R2).

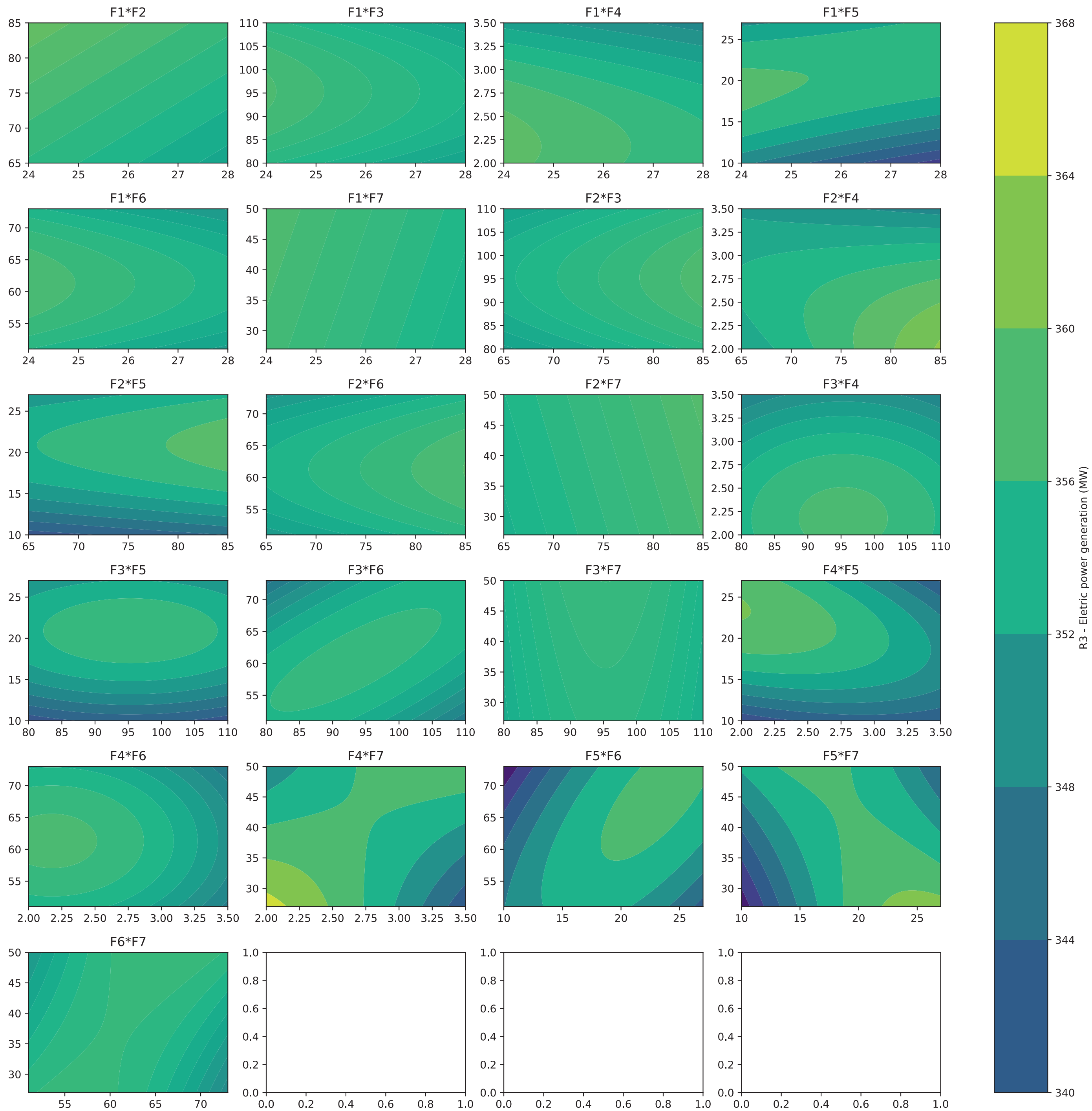


Fig. 14. Contour plots of the pairs of combined factors for the response electric power generation (R3).

References

[1] IEA. Statistics & Data. 2017. https://www.iea.org/data-and-statistics?country=WORLD&fuel=Energy supply&indicator=Coal production by type.

[2] MME. Brazilian energy balance (in Portuguese). 2018.

[3] GE. Data Science Services from GE Digital. 2017. https://www.ge.com/digital/services/data-science-services.

[4] Smrekar J, Assadi M, Fast M, Kuštrin I, De S. Development of artificial neural network model for a coal-fired boiler using real plant data. Energy 2009;34(2):144–52. doi:10.1016/j.energy.2008.10.010.

[5] Hall P, Phan W, Whitson K. The evolution of the technologies of analytics. 1st ed. Media, O'Reilly; 2015. doi:10.1002/9781119198284.ch6. ISBN 9781491954713

[6] Romeo LM, Gareta R. Neural network for evaluating boiler behaviour. Appl Therm Eng 2006;26(14–15):1530–6.

[7] Rusinowski H, Stanek W. Neural modelling of steam boilers. Energy Convers Manag 2007;48(11):2802–9.

[8] Shi Y, Wang J, Liu Z. On-line monitoring of ash fouling and soot-blowing optimization for convective heat exchanger in coal-fired power plant boiler. Appl Therm Eng 2015;78:39–50.

[9] Ilamathi P, Selladurai V, Balamurugan K, Sathyanathan V. Ann–ga approach for predictive modeling and optimization of NoX emission in a tangentially fired boiler. Clean Technol Environ Policy 2013;15(1):125–31.

[10] Zhou H, Cen K, Fan J. Modeling and optimization of the NoX emission characteristics of a tangentially fired boiler with artificial neural networks. Energy 2004;29(1):167–83.

[11] Shi Y, Zhong W, Chen X, Yu A, Li J. Combustion optimization of ultra supercritical boiler based on artificial intelligence. Energy 2019;170: 804–817.

[12] Kumar R, Nikam K, Jilte R. A simulation model to predict coal-fired power plant production rate using artificial neural network tool. In: Applied computer vision and image processing. Springer; 2020. p. 150–60.

[13] Strušnik D, Avsec J. Artificial neural networking and fuzzy logic exergy controlling model of combined heat and power system in thermal power plant. Energy 2015;80:318–30.

[14] Strušnik D, Golob M, Avsec J. Artificial neural networking model for the prediction of high efficiency boiler steam generation and distribution. Simul Model Pract Theory 2015;57:58–70.

[15] Navarkar A, Hasti VR, Deneke E, Gore JP. A data-driven model for thermodynamic properties of a steam generator under cycling operation. Energy 2020;211:118973.

[16] Antony J. Design of experiments for engineers and scientists. 2nd ed. Elsevier; 2014. doi:10.1016/C2012-0-03558-2. ISBN 9780080994178

[17] Montgomery DC. Design and analysis of experiments. 8th ed. John Wiley&Sons; 2013.

[18] Kanimozhi B, Bapu BR, Pranesh V. Thermal energy storage system operating with phase change materials for solar water heating applications: DoE modelling. Appl Therm Eng 2017;123:614–24.

[19] Choi B, Choi J, Heo S, Shin S. Parametric investigation of the effect from controlling variables on thermal deformation during stamping process in automotive body part. Appl Therm Eng 2018;140:373–85.

[20] Burkov A. The hundred-page machine learning book, 1. Andriy Burkov Quebec City, Can; 2019.

[21] Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media; 2019.

[22] Haykin S. Neural Networks and Learning machines; 2014. ISBN 9780131471399, DOI: 978-0131471399

[23] Mathews PG. Design of experiments with Minitab. Milwaukee: American Society for Quality; 2005.

[24] Ferreira S., Bruns R., Ferreira H., Matos G., David J., Brandão G., et al. Box-Behnken design: an alternative for the optimization of analytical methods. 2007. 10.1016/j.aca.2007.07.011

[25] The Babcock & Wilcox Company. Steam: its generation and use. 42n edd. Charlott: The Babcock & Wilcox Company; 2015. ISBN 978-0-9634570-2-8.

[26] EDP. UTE PECÉM. 2019. https://pecem.brasil.edp.com/pt-br.

[27] Chollet F., et al. Keras. https://keras.io; 2015.

[28] Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., et al. TensorFlow: large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org; https://www.tensorflow.org/.

[29] Shah S, Adhyaru DM. Boiler efficiency analysis using direct method. In: 2011 Nirma university international conference on engineering; 2011. p. 1–5. doi:10.1109/NUiConE.2011.6153313.

[30] Burnard K, Bhattacharya S. Power generation from coal; 2011. https://www.oecd-ilibrary.org/content/paper/5kg3n27ts06b-en. https://doi.org/https://doi.org/10.1787/5kg3n27ts06b-en