MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

STANDARDIZATION OF STEAM GENERATOR OPERATION IN ORDER TO INCREASE PERFORMANCE THROUGH PROCESS SURROGATE MODELS

por

Lara Werncke Vieira

Dissertação para obtenção do Título de
Mestre em Engenharia

Porto Alegre, Maio de 2020

STANDARDIZATION OF STEAM GENERATOR OPERATION IN ORDER TO INCREASE PERFORMANCE THROUGH PROCESS SURROGATE MODELS

por

Lara Werncke Vieira
Engenheira de Energia

Dissertação submetida ao Corpo Docente do Programa de Pós-Graduação em Engenharia Mecânica, PROMEC, da Escola de Engenharia da Universidade Federal do Rio Grande do Sul, como parte dos requisitos necessários para a obtenção do Título de

Mestre em Engenharia

Área de Concentração: Fenômenos de Transporte

Orientador: Prof. Dr. Paulo Smith Schneider

Aprovada por:

Prof. Dr. Antônio José da Silva Neto .... DEMEC / UERJ

Prof. Dr. Felipe Antonio Chegury Viana .... MAE /UCF

Prof. Dr. Felipe Roman Centeno .... PROMEC / UFRGS

Prof. Dr. Madhat Abdel-jawad .... Advanced Analysis Australia

Prof. Dr. Fernando Marcelo Pereira
Coordenador do PROMEC

Porto Alegre, 27 de Maio de 2020

ACKNOWLEDGMENTS

More important than words it is to be grateful in our own consciousness. All those around me know how grateful I am. Nonetheless, I would like to register my acknowledgment of those people who were present in my life in an exceptional manner.

I thank my mom for the joy of finding her smile. I thank my father for reminding me every day that I am not alone and for the comfort that his memory brings me. I thank my brother for understanding me, supporting me and being my daily motivation. I also thank the best cousin of all, Super D.

I would like to thank each professor for accepting the invitation and dedicate their time to be part of the examination board. It is an honor to have my work judged by your perspective.

To my advisor, the word is inspiration. Thanks for the support and friendship. There is no need to say that if you need a chimarrão, I am here to bring it to you.

To my lab colleagues, thank you very much. Lucky are the ones who go to work smiling and part of my smile I owe you.

To the friends of life know that you are in my heart.

I also would like to acknowledge EDP - Energia de Portugal, for the financial support that made possible the development of SMART-PECÉM R&D project. I would like to thank to the entire EDP technical team, that highly contributed to the development of this work, in special José Tarcísio Pimentel Neto and Guilherme Lacerda Batista de Oliveira.

Finally, I am very thankful for the financial support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), essential to keep cutting-edge science in Brazil.

RESUMO

Usinas termelétricas a carvão mineral são responsáveis por cerca de 40% da energia elétrica mundial e devem estar alinhadas a requisitos de controle ambiental. As ações para uma operação de alta qualidade podem ser apoiadas pela modelagem fina dos sistemas da planta, a fim de ajudar na operação de campo. O presente trabalho propõe a padronização da operação por meio de modelos substitutos para representar o gerador de vapor e seus moinhos, com base em duas abordagens: a simulação do sistema por um software comercial e, alternativamente, por uma Rede Neural Artificial treinada a partir de dados reais da planta. Uma revisão sistemática da literatura é conduzida para fornecer uma visão ampla da área de interesse e apontar a lacuna que justifica a necessidade de novas pesquisas na área. Uma metodologia para a construção de um modelo substituto para uma usina a carvão em operação é proposta e aplicada ao estudo de caso da usina PECEM. Ferramentas estatísticas como Projeto de Experimentos e Modelo de Superfície de Resposta (RSM) são usadas para identificar os principais parâmetros controláveis e interações do modelo e classificá-los por ordem de importância. O modelo substituto baseado no software comercial foi desenvolvido para simular a eficiência do sistema com sete parâmetros controláveis de entrada: vazão de ar primário, temperatura de saída do carvão pulverizado, velocidade do classificador dinâmico, estequiometria, excesso de O_2, pressão do coletor de ar secundário e pressão do coletor de ar primário. O desvio relativo máximo do modelo substituto em relação ao original é de 0,0172. O modelo substituto baseado em Redes Neurais Artificiais (RNA) também pode simular a eficiência do sistema, juntamente com a temperatura de saída dos gases de combustão e a geração de energia elétrica da planta. A vazão de carvão foi adicionada como um parâmetro de entrada controlável no lugar da estequiometria. Um modelo de RNA com sete entradas apresenta MAE e MSE de 0,2015 e 0,2741 para o conjunto de dados de treinamento e MPE e MSE de 0,32% e 2,350 para o conjunto de dados de validação. A operação padronizada do gerador de vapor visa que o operador respeite a cada alteração das condições de operação o ranqueamento dos parâmetros controláveis a partir da sua importância para o sistema. As faixas que garantem condições de maior desempenho do gerador de vapor devem ser seguidas, principalmente para os parâmetros com maior impacto na eficiência. O conhecimento do impacto de cada parâmetro controlável na operação, suas faixas permitidas de operação, bem como o seu comportamento e interações permitem a manipulação com precisão dos parâmetros corretos, a fim de alcançar uma condição nova, segura, estável e mais eficiente.

Palavras-chave: Box-Behnken Design; Modelo de Superfície de Resposta; Modelo Substituto; Projeto de Experimentos; Redes Neurais Artificais; Termelétrica a carvão

ABSTRACT

Coal-fired power plants (CPPs) provides about 40% of electricity worldwide and should be in line with the stringent environmental control requirements and continuous efficiency enhancement. Actions towards high-quality operation can be supported by fine modeling of plant systems in order to aid field operation. The present work proposes the standardization of operation through surrogate models to represent the assembly of the steam generator and its mills, based on two approaches: the system simulation by the EBSILON commercial software and alternatively by an Artificial Neural Network (ANN) trained with actual plant data. A systematic literature review is conducted to give the reader a broad vision of the area of interest and to point out the gap that justifies this master thesis. A methodology for the construction of a surrogate model to a coal-fired power plant in operation is proposed and applied to the case study of the PECEM power plant. Statistical tools like Design of Experiments (DoE) and Response Surface Methodology (RSM) are used to identify the model main controllable parameters and interactions to then rank them by order of importance. The surrogate model based on the commercial software is built to simulate the system efficiency with seven controllable input parameters: primary air flow, pulverized coal outlet temperature, speed of the dynamic classifier, stoichiometry, excess O_2, secondary and primary air crossover duct pressure, ranked by descendent significance. The maximum relative deviation of that surrogate model compared to the software simulation is 0.0172. The surrogate model based on Artificial Neural Networks can also simulate the system efficiency together with its flue gas outlet temperature and plant electric power generation with the addition of coal flow as a controllable input parameter. An ANN model with seven inputs presents mean absolute error (MAE) and mean square error (MSE) of 0.2015 and 0.2741 for the training data set and mean percentual error (MPE) and MSE of 0.32% and 2.350 for the validation data set. The standardized operation starts with the operator respecting the controllable parameters rank and initializing the alterations for a new condition always for the controllable parameters with a high effect on the steam generator efficiency. Their attention during operation must be kept on the most influential parameters. Finally, controllable parameters must attain the best operating ranges propose. The recommended operational ranges and order of operation by significance allows a precision action in order to achieve

a new, safe, stable, and more efficient condition.

Keywords: Artificial Neural Network; Box-Behnken Design; Coal-fired power plant; Design of Experiments; Response Surface Methodology; Surrogate Model

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF INITIALS AND ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
BBD	Box-Behnken Design
CCD	Central Composite Design
CPPs	Coal-Fired Power Plants
DoF	Degrees of Freedom
DoE	Design of Experiments
EDP	Energy of Portugal Company
GU	Generating Unit
HC	Heat Capacity
HHV	Higher Heating Value
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
MPE	Mean Percentage Error
MSE	Mean Square Error
NPP	Normal Probability Plot
OFA	Over Fire Air
PRS	Polynomial Response Surface
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RSM	Response Surface Methodology
SG	Steam Generator
SHC	Specific Heat Capacity
SSG	Surperheated Steam Generator
VIF	Variance Inflation Factor

LIST OF SYMBOLS

$\dot{m}$	Mass flow rate, kg/hr
$\dot{Q}$	Rate of heat transfer, W
C_O	Center points
$F1$	Primary air flow rate, kg/s
$F2$	Pulverized coal outlet temperature, $^{\circ}C$
$F3$	Speed of the dynamic classifier, rpm
$F4$	Excess O_2, %
$F5$	Secondary air collector pressure, $mbar$
$F6$	Primary air collector pressure, $mbar$
$F7$	Coal mass flow rate, ton/h
h	Specific enthalpy, kJ/kg
H_0	Null hypothesis
k	Number of factors
N	Number of essays
$P1$	Primary air mass flow rate, kg/s
$P2$	Pulverized coal outlet temperature, $^{\circ}C$
$P3$	Speed of the dynamic classifier, rpm
$P4$	Stoichiometry,
$P5$	Excess O_2, %
$P6$	Secondary air crossover duct pressure, $mbar$

$P7$ Primary air crossover duct pressure, $mbar$

$R1$ Flue gas outlet temperature , $^{\circ}C$

$R2$ Steam generator efficiency, %

$R3$ Electric power generation, MW

$S1$ Steam generator efficiency, %

x Design factor

z Standardized variable

Greek Symbols

α Significance level

β Unknown parameters

ϵ Random error

η Efficiency

μ Mean

σ Standard deviation

Subscripts

b Coal combustion

bd Blowdown

c Coal preheating

Exp Expected

f Fuel

fw Feedwater

ms Main steam

Obs Observed

pa	Primary air
rh	Reheat
sa	Secondary air
SG	Steam generator

1 THESIS CONTEXT

This work was developed to analyze the steam generator and its connected mills of a coal-fired power plant in order to increase performance through process surrogate models. The need to comply coal generation with increasingly stringent environmental control requirements pushed engineering to produce more electricity from less coal. Running plants efficiently, and consistently improving efficiency as they run, is the path to putting profits on the bottom line and to reduce environmental impacts. The steam generator is one of the components with the highest potential for efficiency improvements and for this reason will be the focus in this work. The application of surrogate modeling techniques plays an important role in assisting engineering decisions through a cheap and powerful tool for computational analysis of complex real-world systems.

1.1 Objectives

The present work is about the standardization of a coal-fired steam generator and its connected mills of an actual power plant through surrogate models, aiming at improving system efficiency.

The specific objectives of the study are meant to:

- Perform a systematic literature review to highlight areas where further original research is required;
- Provide a bibliometric analysis in surrogate modeling applied to coal-fired power plants;
- Propose surrogate models based on Design of Experiments and Response Surface Methodology as a tool for standardizing steam generator operation;
- Report an experimental investigation applied to a coal-fired power plant guided by Design of Experiments in a logbook format;
- Suggesting a sequence of operational maneuvers to standardize steam generation operation.

1.2 Thesis Outline

This work is composed of three independent chapters. Chapter 2 provides an overview of coal-fired power plant modeling based on surrogate models by means of a systematic literature review. The existing efforts on the field are displayed for further identifying how the current work fits into it, justifying the need for a new research. Chapter 3 proposes a methodology for the construction of a surrogate model of an actual coal-fired steam generator and its connected mills. Design of Experiments (DoE) techniques are applied to explored the controllable parameters of the power-plant, and major decisions are highlighted to guide further studies. Finally, the results to obtain a standardized operation consolidated by a suggestion of a sequence of maneuvers are presented. The last chapter proposes a similar methodology applied to trained Artificial Neural Networks. A higher number of parameters is explored due to a large amount of available plant data without the limitations of the real system.

2 COAL FIRED POWER PLANT MODELING BASED ON SURROGATE MODELS: A SYSTEMATIC LITERATURE REVIEW

2.1 Introduction

Coal-fired power plants (CPPs) play an important role in the energy supply, providing about 40% of electricity worldwide. Their significance is undoubted as they added nearly 900GW to the grid since 2000 . Nearly 4.3% of the Brazilian electric power supply is fuelled with coal, an exception when compared to the world average. Brazilian electric energy matrix is characterized by a strongly renewable share mainly based on hydropower [IEA, 2017].

The need to produce electricity from coal while respecting environmental restrictions pushed engineering to enhance efficiency. An unitary gain in efficiency on conventional pulverized coal power plants can lead to a 2-3% reduction in CO2 emissions [GP Strategies Corporation, 2013; IEA, 2017]. In this context, the steam generator is the component with the highest potential for improvements and therefore it was chosen as the focus of this review. Steam generation performance must be enhanced while use of auxiliary power and losses such as leaks and missing insulation should be reduced. Losses can be classified as a controllable parameter because it can be directly impacted by the actions of the unit control operator. Although the actual control is mostly automatic, some manned intervention can impact the magnitude of losses [GP Strategies Corporation, 2013].

In this context, performance can be improved by standardizing the steam generator operation. Tools must be identified to select significant parameters, estimate their impact on given outputs and correlate cause and consequence events. Surrogate models can aid to build tools focused on the enhancement of steam generator performance, and a systematic literature review on that matter is conducted to evaluate the most recent research efforts on surrogate models applied to CPPs. Some basic concept of Design of Experiments (DoE), Response Surface Methodology (RSM), and Surrogate Models are priorly presented to support the literature review.

2.2 Basic Concepts

2.2.1 Modeling overview

Design of Experiment - DoE

According to Mathews, 2005, DoE is a methodology for studying any response that varies as a function of one or more independent variables that refers to the process of planning, designing and analyzing the experiment. The purpose of a designed experiment is to understand the relationship between a set of input variables and an output.

Furthermore, a well-designed experiment indicates which parameters from a set inputs affect the process or system performance by ranking them by ordering of importance and also points out their interactions. A statistically valid mathematical model can be proposed by observing the response under a planned matrix. All DoE analysis are based on Analysis of Variance (ANOVA) [Antony, 2014; Montgomery, 2013].

Understanding cause-and-effect relationships in a system means to deliberately scan the input variables and observe system output responses. In order to properly assess a designed experiment, it is essential to have a good understanding of the process or system. Figure 2.1 illustrate a general process or system model [Antony, 2014; Montgomery, 2013].

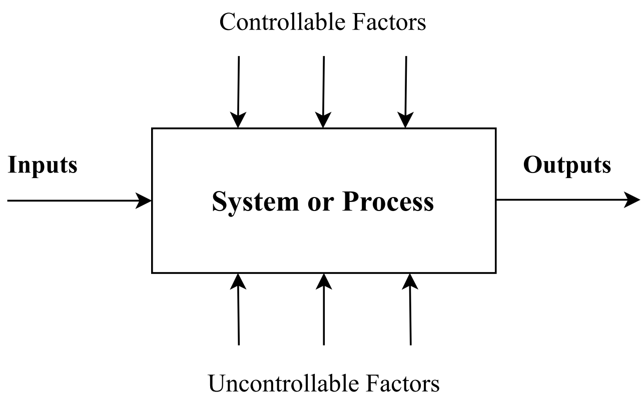


Figure 2.1 – General model of a process or system [Adapted from Montgomery, 2013].

A process is the transformation of inputs into outputs or responses. The factors which are intentionally changed in order to observe the process response are called control factors and must be independent. A carefully planned designed experiment is essential because the result and conclusion are highly dependent on the manner data were collected [Antony, 2014; Montgomery, 2013].

An experiment is a test or series of runs in which purposeful changes are imposed to the input variables to identify process or system behaviors and their causes. DoE aims to investigate a hypothesis by applying ANOVA. The three principles of experimental design are randomization, replication and blocking [Montgomery, 2013], whose sequence is organized in Figure 2.2.

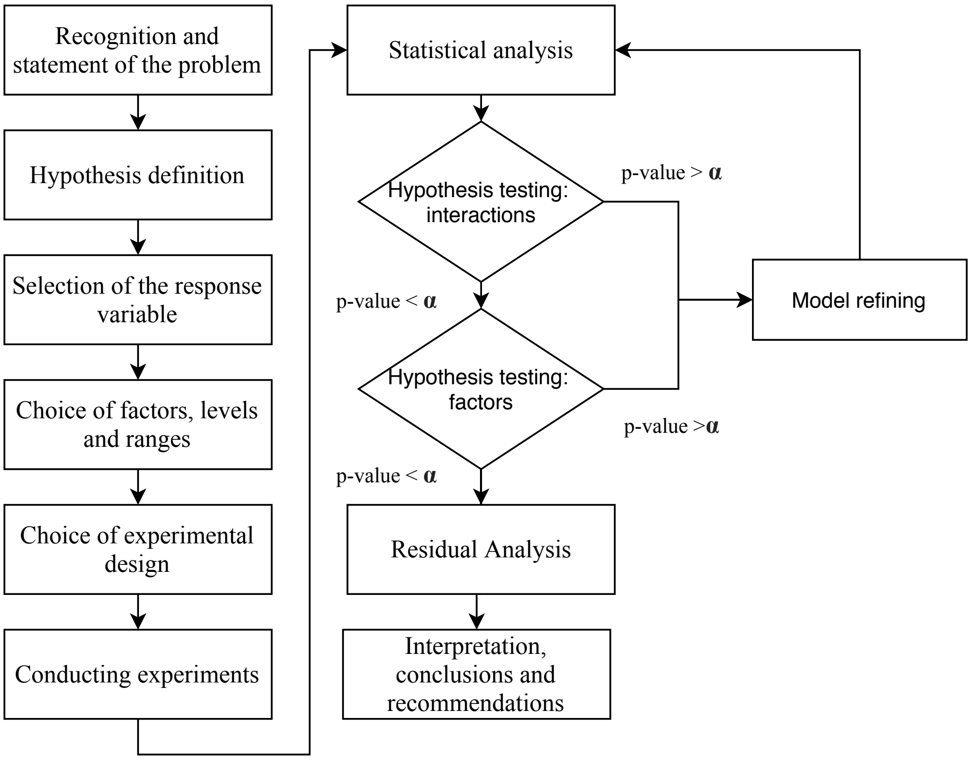


Figure 2.2 – Flowchart to conduct a Design of Experiment step by step

The first step concerns the problem statement and the definition of its control volume. This enables the hypothesis definition, that is the assumption that motivates the experiment. Next, it is necessary to select both the response variable and its corresponding input variables or control factors with their levels and ranges. A successful design depends on the process knowledge to define factor ranges, their appropriate number of levels and measurement unit.

The next step is related to the selection of the experimental design, which can be full factorial design, Box-Behnken design, central composite design, among others. The selected methodology will define the design matrix and its number of experiments. The type of variables, replication and blocking also have an influence.

Experiments can then be conducted after the planning stage is finished for either actual physical systems or their simulation models, when the former are a faithful representation of the real system. Collected data can now be statistically assessed. The hypothesis test is a statistical inference approach to determine the probability of a given hypothesis to be true, and it helps to understand factor interactions or correlations and their individual significance. The hypothesis reflects some conjecture about the problem situation as the significant correlation between parameters.

The null hypothesis is generally assumed to be true until evidence indicates otherwise and is the statement to be tested. The null hypothesis is usually a statement of no effect or no difference. In addition to it, there is an interval within which the value of a rated parameter would be expected to lie in, called confidence interval, defined by (1 - α), where α is the significance level of the test [Mathews, 2005; Montgomery, 2013]. The hypothesis test starts by defining the test significance level α and comparing it to the test p-values, as presented in Inequation 2.1.

$$\begin{cases} p - value < \alpha \quad reject \quad H_0 \\ p - value > \alpha \quad accept \quad H_0 \end{cases} \tag{2.1}$$

The p-value is the smallest level of significance that would lead to reject the null hypothesis H_0 or the smallest level at which the data are significant [Montgomery, 2013]. The p-value for a hypothesis test is calculated from the experimental statistic test under the assumption that the null hypothesis is true and provides a clear and concise summary of the significance of the experimental data. It should be pointed out that if the statistical test result is positive, p-value $<$ α leads to the rejection of H_0, this does not mean that the alternative hypothesis is true but rather means that some evidence to disprove the alternative hypothesis was found [Mathews, 2005].

It is possible to use either p-values or confidence intervals to determine whether the results are statistically significant. If the p-value is less than the determined significance level (α), the hypothesis test is statistically significant. If the confidence interval does not contain the null hypothesis value, the results are statistically significant. If the p-value is less than α, the confidence interval will not contain the null hypothesis value [Montgomery, 2013].

By the time the model is refined, the residual analysis is performed to validate results and prove normality. Interpretations, conclusions and recommendations on the model are presented. DoE indicates the correct approach to deal with several factors by conducting an experimental strategy in which factors are varied together, instead of one at a time. The use of DoE improves the parameter selection by applying a systematic method rather than a trial and error approach.

Response Surface Methodology - RSM

According to Montgomery, 2013, RSM is a collection of mathematical and statistical techniques useful for the modeling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response. Model parameters can be estimated most effectively if proper experimental designs are used to collect data, also called response surface designs. The field of RSM consists of the experimental strategies for exploring the space of the process [Myers et al., 2016].

RSM allows obtaining an approximate function between factors and responses through special experiments and statistical analysis, which makes it closely related to DoE. Fitting and analysing the results is greatly facilitated by the proper choice of an experimental design, specially if it is an appropriate design for fitting response surfaces [Lujan-Moreno et al., 2018; Myers et al., 2016; Montgomery, 2013].

Usually a low-order polynomial in some region of the independent variables is employed to build the approximate function. First-order models are used whenever linear functions can represent experiment (Equation 2.2), otherwise polynomial of higher degree functions are recommended (Equation 2.3) [Montgomery, 2013].

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon \tag{2.2}$$

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j} \sum + \beta_{ij} x_i x_j + \epsilon \tag{2.3}$$

These equations are able to solve most RSM modeling, conducted in accordance to DoE methodology. First-order model can be performed by 2^k factorials, whereas the most popular class of designs for second-order models is Central Composite Design (CCD). CCD consists of a 2^k factorial or fractional factorials with n_F factorial runs, 2^k axial or star runs, and n_C center runs. CCD arises through sequential experimentation of a 2^k in which axial runs are added to allow the quadratic terms to be incorporated into the model. Another important design is the Box–Behnken Design (BBD), with a set of three level designs for fitting response surfaces. These designs are formed by combining 2^k factorials with incomplete block designs. The resulting designs are usually very efficient in terms of the number of required runs. Also, they do not contain vertices points created by the upper and lower limits for each variable. This could be advantageous when points on the corners represent factor-level combinations that are prohibitively expensive or impossible to test because of physical process constraints [Myers et al., 2016; Montgomery, 2013].

Response surface designs are most often used to build predictions models. Commonly RSM is used to determine the optimum operating conditions for a process or system or to identify a region of the factor space in which operating requirements are satisfied. Three options are accepted to the response: maximize, minimize or a target. Similarly to DoE, RSM can be applied to simulation models or actual physical systems [Myers et al., 2016; Montgomery, 2013].

Surrogate Models

Surrogate models or metamodels are mathematical representations of actual or simulated system that can be seen as a model of a model, and enable the replacement of expensive procedures by approximating their inputs-outputs responses. The main objective is to predict performance of systems and facilitate the exploration of the design space to search for an optimal design [Cremanns et al., 2016; Jiang et al., 2020].

There are many commonly used surrogate modeling techniques such as Polynomial Response Surface (PRS), Kriging, Artificial Neural Network (ANN) and Radial Basis Function (RBF). The RSM correspond to the PRS methodology. DoE is often used to build surrogate models through experimental design to determine the coefficients of a polynomial. From this perspective, the surrogate approximation of the objective function is called response surface. The equation coefficient magnitudes can be used as a basis to judge the role of each parameter on the entire system response [Jiang et al., 2020; Montgomery, 2013; Antony, 2014].

Surrogate models have been used successfully in several fields where computational simulations or experimentation are time expensive or hard to perform. The technique serves as a powerful alternative for performance improvement of complex real-world system or other decision-making.

2.3 Systematic Literature Review

2.3.1 Literature research

A systematic literature review is a fundamental step in conducting a scientific research and contrast to the traditional narrative reviews by adopting a replicable, scientific and transparent process. If the methodology is clear, the review can be easily updated in the future with new research findings [Dresch et al., 2015].

The systematic review presented in this chapter is theme-centric and presents prior publications that have contributed to the development and understanding of themes and phenomena of interest. The step by step procedure to conduct the systematic review is presented in Figure 2.3.

Question Definition and Conceptual Framework

The first step in conducting systematic reviews is to define the central topic through the identification of a research question to define a conceptual framework [Dresch et al., 2015].

The conceptual framework serves as the basis for carrying out the systematic review. It maps out the actions required in the course of the study given his previous knowledge of other researchers' points of view and his observations on the subject of research. The definition of the research scope is the starting point that allows for understanding the review and its context. It can be developed, refined, or confirmed during the course of the research. In this research, the adopted conceptual framework concerned coal-fired power plants, DoE, RSM, and surrogate models.

From the knowledge of the methods presented previously and what they are capable of answering, it was arrived at the why and how of this work, synthesized in this research question: how surrogate modeling techniques supported by DoE and RSM can help enhancing coal-fired power plant efficiency?

Research strategy

One way to successfully perform a systematic review is to stablish a research strategy, that aims to define what and where will be searched, how to mitigate bias, what studies to consider and what will be the extent of the search, including the selection and combination of keyword(s) and database(s). These steps are subjective [Dresch et al., 2015] and all decisions made during this research strategy are found in the protocol avail-

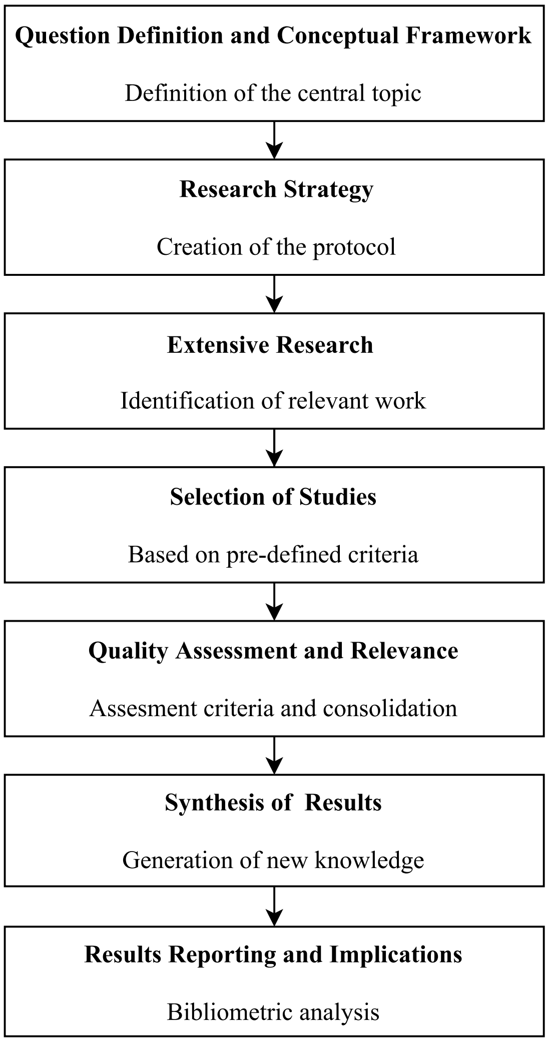


Figure 2.3 – Systematic literature review method of this survey [Based on Dresch et al., 2015; Linnenluecke et al., 2019; Khan et al., 2003].

able in the Appendix A (Table A.1).

The search terms are defined in accordance with the conceptual framework, followed by the respective search sources and finally the criteria for inclusion and exclusion. In the present work, the literature associated with the search terms "coal-fired power plant", "surrogate model", "Design of Experiments", and "Response Surface Methodology", were collected from academic databases including Scopus and Web of Science.

The search criteria were separated into inclusion (I) and exclusion (E) criteria, in accordance with: (I) applies DoE in coal-fired power plants; (I) applies RSM in coal-fired power plants; (I) develops a surrogate model based on significant parameters in thermal power plants; (E) article related to other areas, such as chemistry; (E) does not apply DoE, RSM or surrogate modeling methods in thermal power plants.

Extensive Research

Table 2.1 presents number of papers that were identified after the set of search expressions for Scopus Database.

Table 2.1 – Search expressions with the selected keywords, boolean operators and truncated terms to research at Scopus database with 250 articles as the total final result

Search	Expression	Total results
1	TITLE-ABS-KEY (("Design of Experiment*") AND (("power plant*") OR ("coal-fired") OR ("thermoelectric power")) AND (("boiler") OR ("steam generator")))	11
2	TITLE-ABS-KEY ((("Design of Experiment*") OR ("statistic* model*") OR ("response surface methodology")) AND (("power plant*") OR ("coal-fired") OR ("thermoelectric power")) AND (("decision making") OR ("parameter* selection*") or ("parameter* rank*")))	16
3	TITLE-ABS-KEY ((("Design of Experiment*") OR ("statistic* model*") OR ("response surface methodology")) AND (("power plant*") OR ("coal-fired") OR ("thermoelectric power")) AND (("boiler") OR ("steam generator*")))	40
4	TITLE-ABS-KEY ((("Design of Experiment*") OR ("statistic* model*") OR ("response surface methodology")) AND (("power plant*") OR ("coal-fired") OR ("thermoelectric power")) AND (("decision making") OR ("parameter*selection*") or ("parameter* rank*") or ("performance management")))	192
5	TITLE-ABS-KEY ((("Design of Experiment*") OR ("response surface methodology") or ("surrogate model*")) AND (("power plant*") OR ("coal-fired") OR ("thermoelectric power")))	250

The highest number of papers was found with the expressions of Search 5, whose operators were less restrictive. The same set of keywords, booleans operators and truncated terms were applied to search on the Web of Science database. The two database resumed 434 relevant works.

Selection of studies

The selection of relevant studies followed the process presented in Figure 2.4. Duplicate works were excluded and the remaining 423 out of 434 ones were filtered by reading all titles and abstracts. Only 13 studies were selected as relevant to answer the proposed review question, after the exclusion of 371 works based on title and abstract reading and 39 based on full-text analysis.

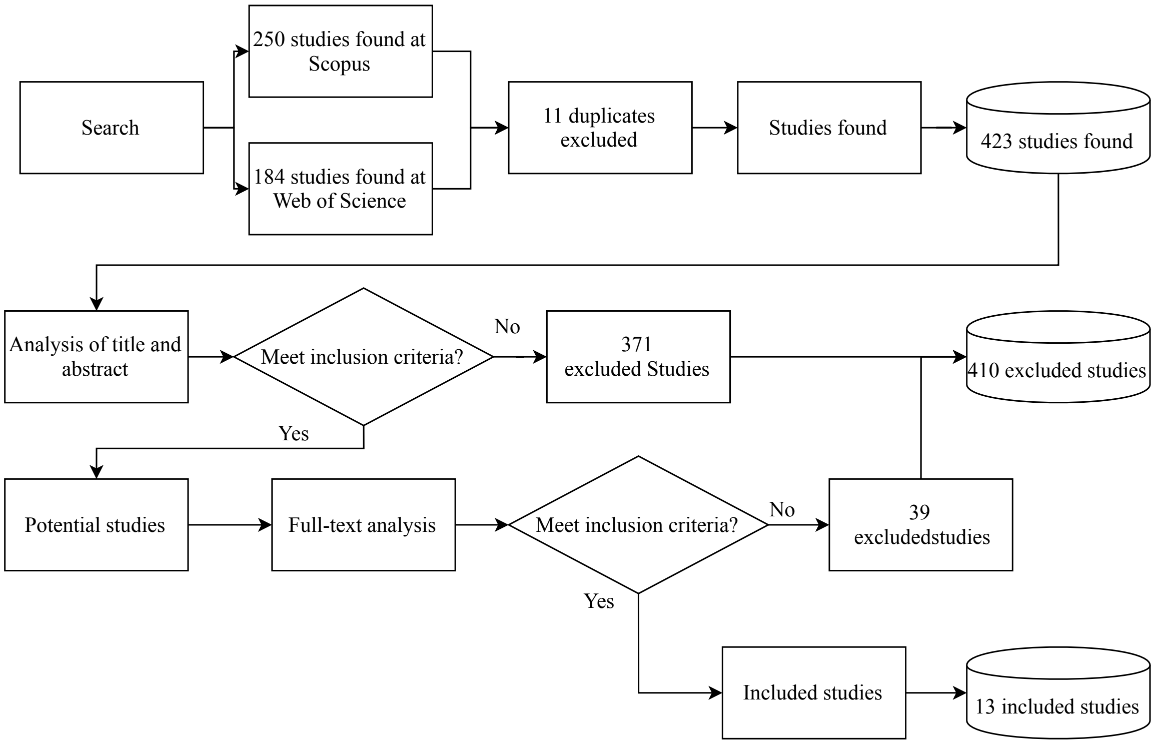


Figure 2.4 – Search process and eligibility [Adapted from Dresch et al., 2015].

Quality assessment and relevance

Quality criteria were proposed as shown in Table 2.2 to help classifying the selected studies into three levels, according to their fit when answering the proposed research question.

Table 2.2 – Assessment criteria of the dimensions of quality of primary studies [Adapted from Dresch et al., 2015].

	Dimension		
	1. Quality of the study performance	**2. Relevance to the review question**	**3. Relevance to the review focus**
High	The proposed method meets the standards required for the subject under study; the study strictly followed the proposed method; and the results are supported by facts and data.	The study precisely addresses the target subject of the systematic review. Meaning the standardization of the steam generator operation in a coal-fired power plant based on surrogate models of the process.	The study was conducted in the same context to the one defined for the review. Meaning surrogate models to the steam generator of a coal-fired power plant.
Medium	The proposed method has gaps regarding the standards required for the topic under study; or the study fails to show that it followed the proposed method in its entirety; or the results are not fully based on facts and data.	The study partially addresses the subject matter of the systematic review. Meaning it follows the first steps to build a surrogate model, but not finish the whole process.	The study was conducted in a similar context to the one defined for the review. Meaning surrogate models to a subsystem of a coal-fired power plant, not necessarily the steam generator.
Low	The proposed method does not comply with the standards required for the subject under study; or the study fails to show that it followed the proposed method; or the results are not based on facts and data.	The study only slightly addresses the systematic review subject. Meaning it stops at the first at of building a surrogate model.	The study was conducted in a different context from the one defined for the review. Meaning surrogate models of a process using a subproduct of a coal-fired power plant.

These criteria were applied to the set of 13 selected studies and delivered the final rank presented in Table 2.3.

Inclusion rules may change according to the evaluators based on three options: (i) include all studies, despite their grades, assigning smaller weights to low-quality studies in case of quantitative results; (ii) include all studies, describing their quality and relevance,

Table 2.3 – Assessment consolidation by using Table 2.2

	Assessment of Dimensions			**Study Assessment**
	1	**2**	**3**	Final
Amiri et al. [2017]	Medium	Medium	Low	Low
Chandane et al. [2017]	Medium	Medium	Low	Low
Chandane et al. [2018]	Medium	Medium	Low	Low
Chandrasekharan et al. [2017a]	Medium	Medium	High	Medium
Chandrasekharan et al. [2017b]	Medium	Medium	High	Medium
Cremanns et al. [2016]	Medium	Medium	Medium	Medium
Kumar et al. [2019]	Low	Low	Low	Low
Mahanta et al. [2019]	Medium	Medium	Low	Low
Remenárová et al. [2014]	Medium	Medium	Low	Low
Seetharama-Yadiyal et al. [2018]	Medium	Medium	Medium	Medium
Verma et al. [2006]	Low	Low	High	Low
Wakiru et al. [2019]	Low	Low	High	Low
Xu et al. [2019]	Medium	Medium	Low	Low

and the readers can have their own conclusions; (iii) perform sensitive analysis to verify the effects of including or not a study [Dresch et al., 2015]. All papers were included in the present review despite of their grade, due to the low number findings.

It is worth mentioning that the quality of the review depends on the whole process, and not only the quality of the selected studies. In this particular research, none of the selected studies displayed a high grade probably because the criteria were very demanding. The criteria were defined at the beginning of the assessment and none of them was changed.

Synthesis of results

As stated by Dresch et al., 2015, the synthesis process involves combining the results in a connected way to generate new knowledge that did not exist in the original primary studies. The synthesis techniques are dependent on both the type of question and the type of the review being conducted. The challenge is to transform the collected data into answers to the review question, organizing data available and identifying patterns and similarity among them.

As showed in Section 2.3 there is no study that achieved the highest grade when answering the review question. Nevertheless, the present review was capable of showing how each of the described methods were used in a similar context, how they complemented each other and how they were applied. Putting all these studies together it was possible to justify the presented research question and the connection between the research gap and current goal of the present work.

A brief description of each selected study is presented, ranked from medium to low quality. Their similarities were analysed and their most relevant aspects were highlighted.

Chandrasekharan et al., 2017b, and Chandrasekharan et al., 2017a, used RSM supported by DoE to optimize the operating parameters of an integrated boiler unit for a coal fired power-plant. These papers proposed the optimization of operational parameters considering as outputs the pressure and temperatures at the economizer, drum, superheater and the integrated boiler unit. The input parameters considered by Chandrasekharan et al., 2017b, were coal feed, feed water and air, while Chandrasekharan et al., 2017a, considered specific heat transfer rate of flue gas, flow rate of feed water and enthalpy of the feedwater.

Cremanns et al., 2016, proposed DoE combined to surrogate modeling to obtain an optimized design of labyrinth seal leakage in steam turbines through multi-dimensional target functions. It is a well-developed work, which presents all the steps of creating a CFD-based surrogate model.

Seetharama-Yadiyal et al., 2018, focused on the development and application of RSM to capture the performance of a complex power system through a surrogate model. The generated surrogate model become part of a wider computational platform and enabled to system optimization.

Amiri et al., 2017, applied DoE and RSM on a CO2 capture process from the flue gases of fossil fuel power plants. DoE and RSM were also applied by Chandane et al., 2017, that used a byproduct of coal-fired thermal power plants, the cenospheres, to develop a heterogeneous acid catalyst. RSM was used to optimize the various process parameters for the synthesis through Box-Behnken design. Polynomial model equations were developed to predict the esterification conversion and yield. Chandane et al., 2018, continued the previous work.

Kumar et al., 2019, performed experiments in a thermosyphon integrated thermoelectric generator using coal-fired fly ash (CFFA) collected from a local thermal power plant. The experiments were conducted through DoE and RSM to select process parameters for optimization, but there was no development of a surrogate model to represent the system.

Mahanta et al., 2019, used boron carbide along with fly ash from a thermal power plant as reinforcement particles in the aluminium matrix to fabricate a new class of composite. DoE and RSM were employed for investigation of the response variables. Multiple linear regression models were obtained to establish the functional relationship between response variables and process parameters.

Remenárová et al., 2014, verified the applicability of fly ash from the combustion of brown-coal in the ENO Novaky power plant (Slovak Republic) for the synthesis of zeolitic materials. A RSM using Box–Behnken design was applied for investigation of interaction and competitive effects in a binary metal system. Second-order polynomial models were obtained. In addition, Pareto graphs were used to present the effects of observed factors and their combined impacts.

Verma et al., 2006, approached the development and demonstration of a technology for ultra clean 21st century energy plants that could effectively remove environmental concerns associated with the use of fossil fuels for producing electricity. The DoE methodology was applied to identify the significant factors that affected the system performance, ranked them to further on propose system design modifications. The authors did not apply RSM nor developed surrogate models of the system.

Wakiru et al., 2019, established and selected significant optimization parameters that affected equipment performance with the objective of offering maintenance decision support on a thermal power plant. DoE approach was utilized to quantify the effects and interactions of the variables on equipment availability and total repair time. The paper did not stressed DoE capabilities nor proposed any RSM or surrogate model, but it presented DoE as fundamental for identifying important parameters to the process and their interactions. Finally, the optimization stage was conducted through simulation models.

Xu et al., 2019, employed RSM to model and optimize the electrodialysis process for Reverse Osmosis Concentrate (ROC) reclamation in coal-fired power plants. They applied DoE combined with RSM to develop a surrogate model.

Chandrasekharan et al., 2017a,b, papers stood out for the development of individual regression models for efficient calculation of boiler performance using RSM supported by DoE with Box-Behnken design. The statistics part was well described, but none of them mentioned parameter selection, rank or stabilization. Remenárová et al., 2014, applied RSM through Box-Bhenken design using multiple regression analysis to develop second-order polynomial models to a chemical process. However, they did not explain the individual impact of each parameter, their interactions or ranking.

Chandane et al., 2017, 2018, Xu et al., 2019, Amiri et al., 2017 and Kumar et al., 2019 applied RSM supported by Box-Behnken design for parameter optimization. However, just Chandane et al., 2017, Chandane et al., 2018 and Xu et al., 2019 employed polynomial equations to describe the system of interest. Similarly, Mahanta et al., 2019 applied RSM supported by central composite design. The process parameter contribution and their interactions were studied to develop equations to describe the system.

Cremanns et al., 2016, developed a surrogate model based on DoE applied to simulation models, but they did not include polynomial equations, RSM, parameter raking or the discussion of operational parameters. Seetharama-Yadiyal et al., 2018, identified the key design parameters and their impact on the system performance using RSM to build a surrogate model. The main objective was optimization. Finally, Verma et al., 2006, and Wakiru et al., 2019, applied DoE to assist the determination of critical system parameters to be optimized. Wakiru et al., 2019, applied a full-factorial 2^k in a simulation model, but focused on maintenance. Verma et al., 2006, presented the ranking of the parameters as an advantage.

In summary, nine of the selected studies applied RSM and DoE, two applied RSM without addressing DoE and the last two applied DoE solely. In this regard, even the studies that covered RSM, DoE and developed a surrogate model did not explore the full potential of the tools in a single study. Each paper presented their tool exploring one of its advantages, such as the system optimization or the parameter selection, but none of them included the interaction between parameters, ranking, construction of a surrogate model and then the optimization of the process. It is worth mentioning that not only the full potential of the tools was not explored simultaneously, but studies were applied to systems different than the one proposed in the present work or at different conditions. Only the complete reading of those articles allowed to reach the presented conclusions.

2.3.2 Bibliometric analysis

The aim of this section is to share the results in an easy and accessible way, allowing to assist decision making. Bibliometric analysis was chosen in the present work, among different available formats, to organize the sources.

Results are presented hereafter based on the original 423 papers identified in the beginning of the present work, in order to provide the reader a broad view about the field, although Table 2.3 brings 13 selected works. These 423 papers were classified according to their year of publication and journal. Networks maps were developed based on citation and common keywords. Figure 2.5 presents the publications per year.

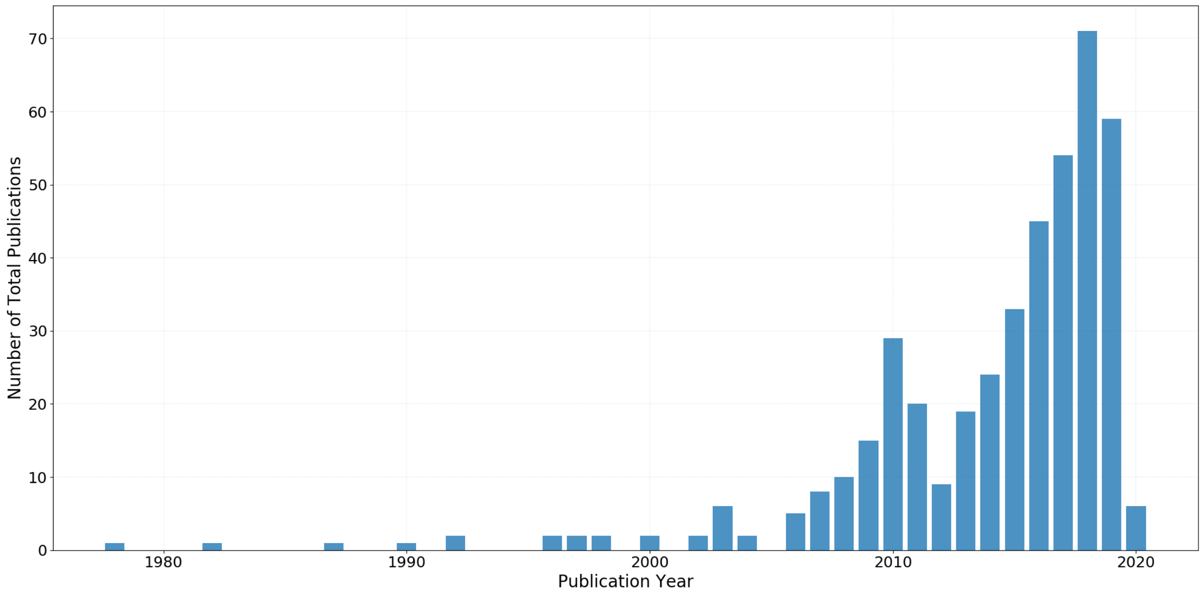


Figure 2.5 – Number of publications per year

The interesting part to be noted in the Figure 2.5 is the increase in the number of publications from 2005 on. A relevant number was noticed in 2018, reaching approximately 70 publications. Next analysis presented in Table 2.4 is about the number of publications per journal, accounting the 20 most relevant.

Table 2.4 – Number of published papers per relevant journals

Journal	**Number of Publications**
NUCLEAR ENGINEERING AND DESIGN	14
ENERGY	14
RENEWABLE ENERGY	9
NUCLEAR ENGINEERING AND TECHNOLOGY	8
APPLIED ENERGY	8
INTERNATIONAL JOURNAL OF GREENHOUSE GAS CONTROL	7
APPLIED THERMAL ENGINEERING	7
ENERGY CONVERSION AND MANAGEMENT	6
JOURNAL OF CLEANER PRODUCTION	5
ADVANCED MATERIALS RESEARCH	5
PROCEEDINGS OF THE ASME TURBO EXPO	5
AMERICAN SOCIETY OF MECHANICAL ENGINEERS, PRESSURE VESSELS AND PIPING DIVISION (PUBLICATION) PVP	5
INTERNATIONAL CONFERENCE ON NUCLEAR ENGINEERING, PROCEEDINGS, ICONE	4
INTERNATIONAL JOURNAL OF ENVIRONMENTAL SCIENCE AND TECHNOLOGY	4
MATERIALS TODAY: PROCEEDINGS	4
RSC ADVANCES	4
ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH	4
MATERIALS TODAY-PROCEEDINGS	4
HELIYON	4
PROCESSES	4

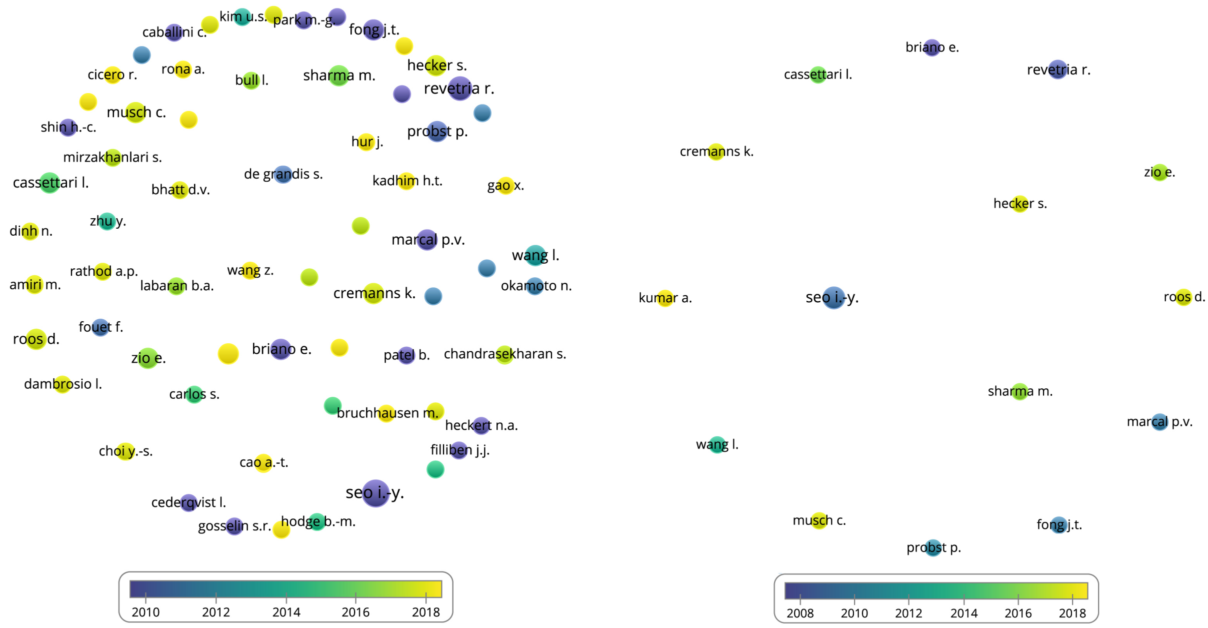
The journal with more publications is Nuclear Engineering and Design Energy, followed by Energy and Renewable Energy. Papers were mainly connected to the energy theme.

Next analysis was conducted by networks maps with the aid of the software VOSviewer [1] [van Eck and Waltman, 2010], a tool for constructing and visualizing bibliometric networks. The motivation was to look for trends in the topics of interest in the research literature. Color map indicates the publication year and circle size represents the number of citations. The citation network maps were developed based on the number of documents and citations per author.

Figure 2.6a displays authors with a minimum of two published papers with at least one citation per paper. Restricting the analysis, Figure 2.6b presents authors with at least

[1] https://www.vosviewer.com

three published papers and with at least one citation per paper.



(a) Minimum of two published papers per author with at least 1 citation.

(b) Minimum of three published papers per author with at least 1 citation.

Figure 2.6 – Citation network map. Circle sizes represent the number of citations.

From a universe of 780 authors, the first restriction returned 97 matches and the second one drop it down to 15 authors. It can be seen that the difference between the number of authors who met the threshold for a minimum of 2 or 3 documents was significant. The reduced number of publications indicates that there are no featured authors in this specific area of surrogate modeling applied on CPP.

The keyword network map is presented in Figure 2.7. The time scale for papers publication is given by colors. Purple represents papers until 2012 and yellow from 2018. Circle sizes represents the number of times the keyword was cited, the bigger the circle the more times the keyword was cited.

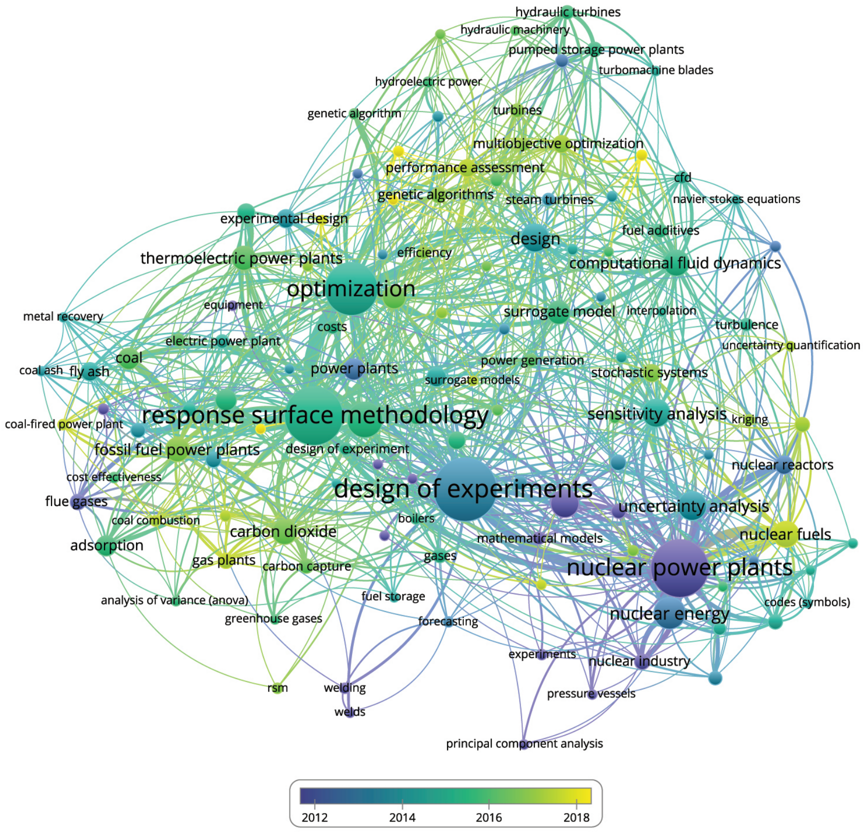


Figure 2.7 – Keyword network map. Circle sizes represent the number of times the keyword was cited.

The first feature that stands out is the strong connection between keywords, with an emphasis on the four most cited ones: *design of experiments*, *response surface methodology*, *optimization* and *nuclear power plants*. The keyword *nuclear power plant* was expected to be important as the journal with more publications (Table 2.4) is in the nuclear field. It is possible to notice that *response surface methodology* and *optimization* are more up to date keywords, while *design of experiments* and *nuclear power plants* are more frequently used in previous studies. The assessment was limited to publications with a minimum of 5 keywords, and 126 papers met the threshold out of 2948 keywords. The keyword network map (Figure 2.7) suggests that the DoE and RSM are widely used in different areas of research.

The same class of map was generated in Figure 2.8, limited to the most frequent keywords to facilitate visualization.

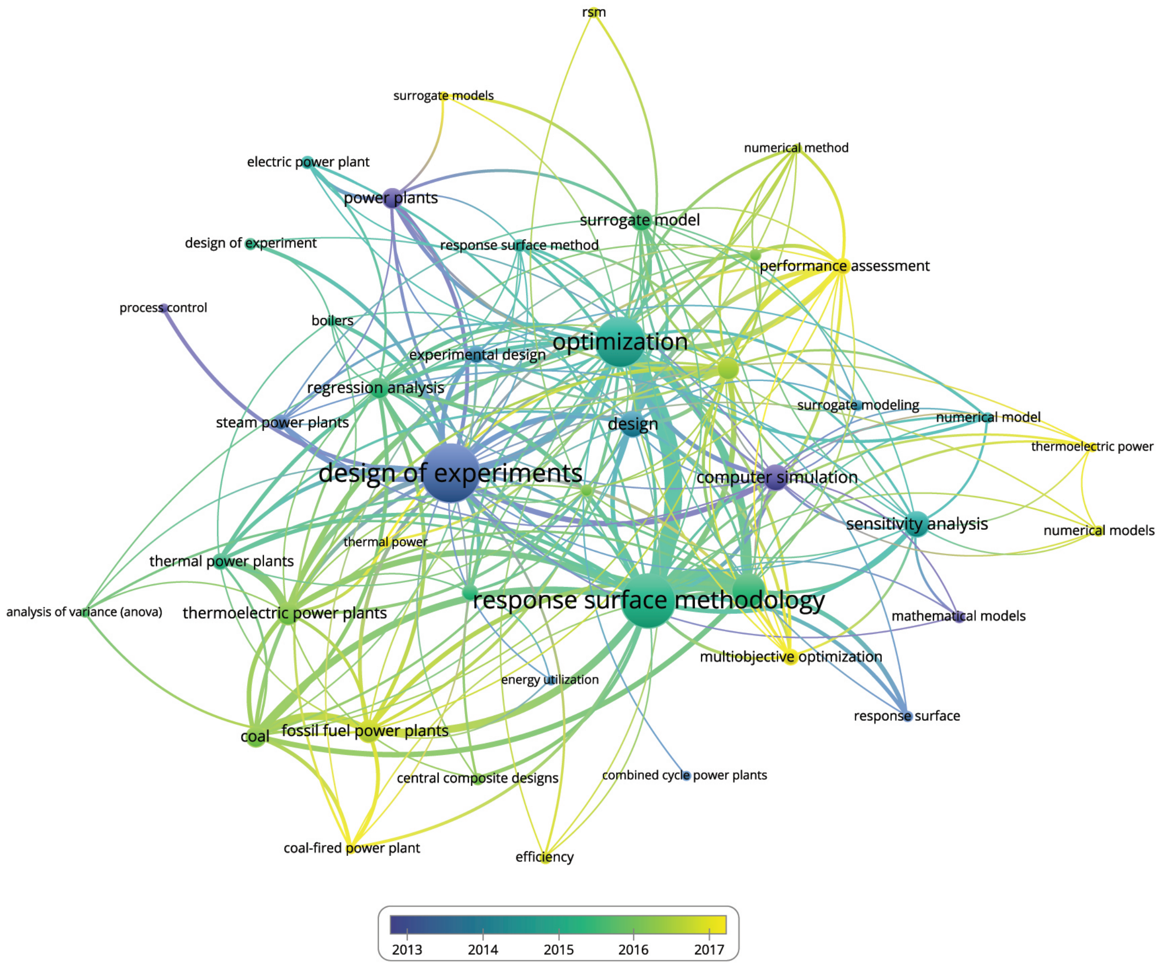


Figure 2.8 – Keyword map considering only the relevant ones. Circle sizes represent the number of times the keyword was cited.

Keywords related to *coal-fired power plants* and *surrogate models* are connected to more recent studies, and it can be identified keywords related to *performance assessment* and *process control*. It is worth mentioning the strong presence of the keyword *optimization*, which was never mentioned during the search.

The bibliometric analysis ends with the keyword network map (Figure 2.9) with the systematic review presented in Table 2.2.

No restrictions concerning the number of occurrences were considered, which highlighted the prevalence of *response surface methodology* rather then *design of experiments* and *surrogate model*. Besides that, draw the attention the keyword Box-Behnken design, a method of DoE for fitting response surfaces. The keyword Box-Behnken design was never used during the search.

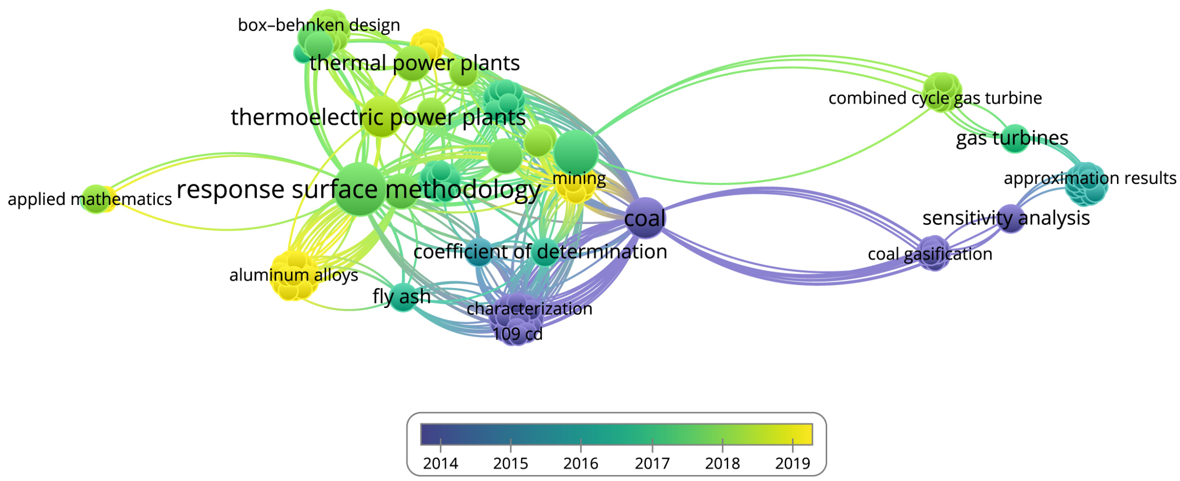


Figure 2.9 – Keyword map with the selected articles presented in Table 2.2. Circle sizes represent the number of times the keyword was cited.

2.4 Conclusions

The proposed systematic literature review aimed to answer the following question: how surrogate modeling techniques supported by DoE and RSM can help enhancing coal-fired power plant efficiency?

Thirteen studies were selected out from the initial list of four hundred and twenty-three, based on quality and relevance criteria. Four out of thirteen were classified as medium according to the quality assessment criteria. The rest of them were classified as low quality.

Of the selected studies, nine were classified as low because they were not conducted in the same context to the review or did not meet the standards required for the subject under study or did not target the subject. The other four were classified as medium based on the same assessment criteria. Nine of the selected studies applied RSM and DoE, two just RSM without addressing DoE and the last two just applied DoE. The studies did not explore the full potential of the methods. Each article presented the method exploring one of its advantages, however, none of the studies include the parameter selection, interaction between parameters, ranking according to their importance at the response, construction of a surrogate model and then the optimization of the process.

The systematic literature review was able to give the reader a broad vision of the area of interest and pointed out the gap that justifies the research question. The application of surrogate modeling techniques in coal-fired power plants supported by DoE and RSM play an important role in assisting engineering decisions through a cheap-to-run surrogate model and makes it easier to identify interesting regions to be explored and analyzed.

3 DESIGN OF EXPERIMENTS APPLIED TO THE STEAM GENERATOR OF PECEM POWER PLANT: SURROGATE MODELING APPROACH

3.1 Introduction

A coal-fired power plant is a complex system of interconnected processes that converts chemical energy into electricity. Its core is the steam generator, where heat released from the combustion process is transferred to the working fluid. Plant efficiency, fuel consumption, and capital cost are critically related [Annaratone, 2008; The Babcock & Wilcox Company, 2015; GP Strategies Corporation, 2013].

Power plant operation effectively takes place at the steam generator, as no other action on the remaining equipment can impact the overall performance to the same level [Annaratone, 2008; The Babcock & Wilcox Company, 2015]. The control system handles plant stability, leaving the operator to manage controllable losses [GP Strategies Corporation, 2013].

An experienced operator knows the plant characteristics and develops its particular way of command that guarantees the system integrity and performance. Although effective, there is room for reducing variability and improving the system performance process standardization, by means of decision support tools. These tools may be based on computational representations able to simulate the system behaviour in a broad range of conditions, also called surrogate models. The one developed in the present work was based on the Design of Experiments (DoE) and Response Surface (RSM) methodologies to standardize the steam generator and its mills operation to suggest operating conditions to the operator. A sequence of maneuvers are provided indicating the controllable parameter that the operator must act and the respective value of operation. The procedure to conduct DoE is applied to the PECEM power plant, located in São Gonçalo do Amarante, Ceará. The proposed methodology can be applied to other generation plants.

This chapter is organized as follow. Sections 3.2, 3.3 and 3.4 presents the basics concepts about the system and the surrogate modeling tools. Section 3.5 exposes the methodology for the construction of a surrogate model to a coal-fired power plant in operation. Section 3.6 discusses the application of the proposed methodology to the case study of PECEM power plant and highlights the major decisions to conduct the analysis during the operation. Section 3.7 introduces a simulation model to conduct experiments that could not be performed at the power plant and proceeds with the analysis. Finally, Section 3.8 the conclusions of the present chapter.

3.2 System Description

PECEM I [1] is composed by two independent sub-critical coal-fired power units of 360MW electric power output each. The identical steam generators are equipped with heat exchangers such as superheaters, reheaters, economizers and air heaters, arranged to efficiently absorb heat released by fuel combustion and deliver steam at rated temperature, pressure and capacity. These last parameters determine the steam generator configuration [Annaratone, 2008; The Babcock & Wilcox Company, 2015]. Three independent mills feed one steam generator with dry pulverised coal as shown in Figure 3.1. In fact, there are four mills available but one of them serves as a backup.

Air stream coming from a common heating device at approximately 300°C (air preheater) is split into two feeding paths, the primary and 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. Primary air and pulverized coal stream temperature at the combustor input ranges around 80°C. The secondary air stream is directly connected to the wind box, pressure-balanced, and admitted in the combustor to be mixed with the primary air and pulverized coal stream. The burners are arranged in four rows of six each on the furnace front and rear walls (letters (b), (c), (d), (e), (f), and (g)). Coal and air are rapidly mixed and burned in the furnace under sub-stoichiometric conditions, to be completed with extra oxygen from the over fire air (OFA) ports (a) at the burnout zone. OFAs are fed with secondary air from the preheater, arranged in two rows with six injectors each above the top rows of the pulverized fuel burners. Feedwater is admitted in countercurrent to the flue gases at the economizers (ECO1 and ECO2), to evaporate at the furnace water walls and superheated on three superheaters (SH1, SH2, and SH3). Both the main steam stream and the reheated stream feed the cycle turbine at 540°C, 180 bara, and 36 bara, respectively.

[1] https://pecem.brasil.edp.com/en/power-plant.

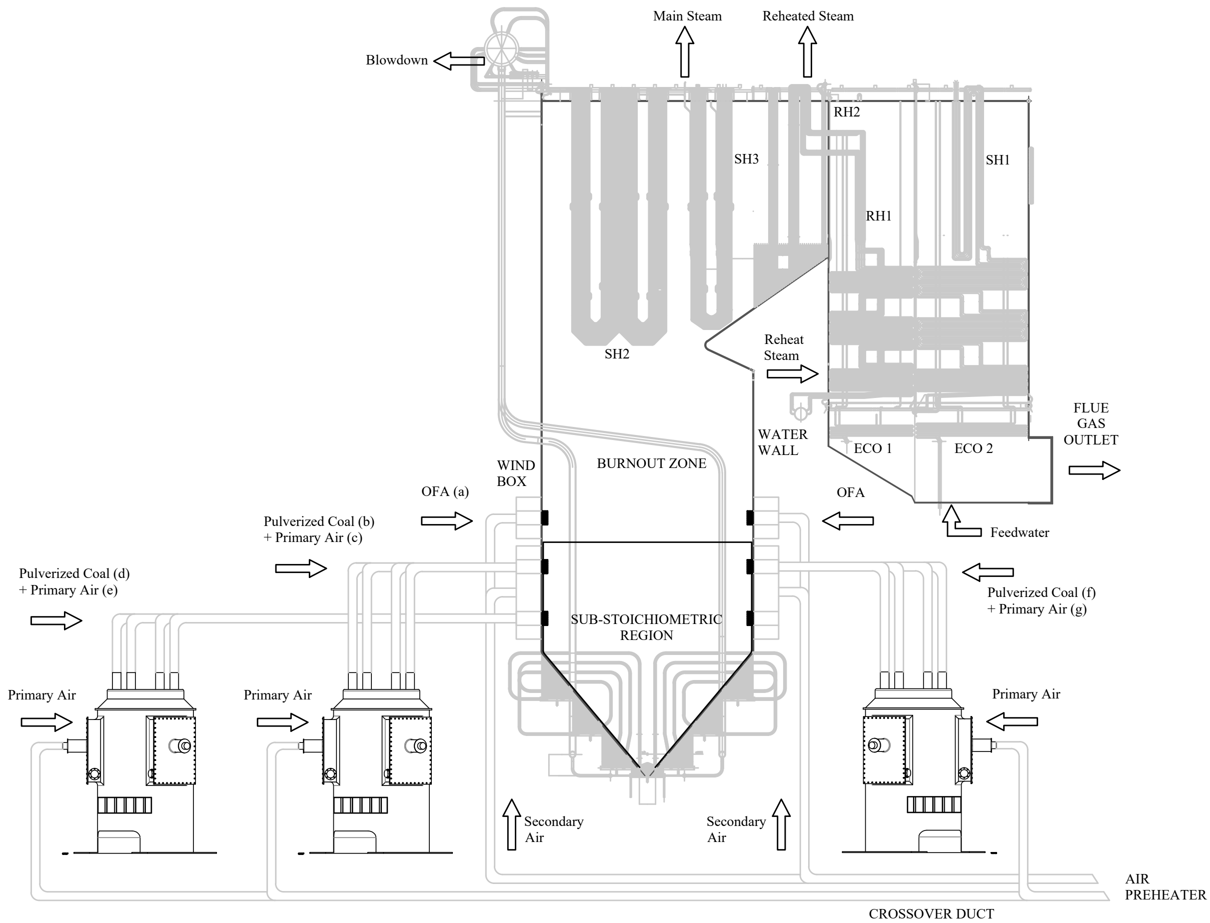


Figure 3.1 – Schematic layout of the steam generator and mills of PECEM power plant

3.2.1 Steam Generator Efficiency

Efficiency was calculated following the Direct Method (DM), which is basically the ratio of the output to the input heat streams [Chetan and Bhavesh, 2013], as presented in Equation 3.1,

$$\eta = \frac{\dot{Q}_{out}}{\dot{Q}_{in}} = \frac{\dot{m}_{ms}(h_{ms} - h_{fw})}{\dot{m}_f HHV} \tag{3.1}$$

with $\dot{m}_{ms}$ the generated steam mass flow rate (kg/hr), h_{ms} the steam specific enthalpy (kJ/kg), h_{fw} the feedwater specific enthalpy (kJ/kg), $\dot{m}_f$ the fuel mass flow rate (kg/hr) and HHV the fuel Higher Heating Value (kJ/kg). $\dot{m}_{ms}$ accounts for both the main and reheated generated steam outputs.

A more detailed ratio is proposed by DIN 1942, 1994, as presented in Equation 3.2,

$$\eta_{DIN} = \frac{\dot{Q}_{ms} + \dot{Q}_{rh} + \dot{Q}_{bd}}{\dot{Q}_{b} + \dot{Q}_{f} + \dot{Q}_{pa} + \dot{Q}_{sa}} \tag{3.2}$$

with heat transferred from output streams such as $\dot{Q}_{ms}$ (main steam), $\dot{Q}_{rh}$ (reheated steam), and $\dot{Q}_{bd}$ (blowdown steam), and the input streams $\dot{Q}_{b}$ (coal combustion), $\dot{Q}_{f}$ (coal preheating), $\dot{Q}_{pa}$ (primary air), and $\dot{Q}_{sa}$ (secondary air). Equations 3.3 to 3.9 detail their calculation.

$$\dot{Q}_{ms} = \dot{m}_{ms}(h_{outputsteam} - h_{fw}) \tag{3.3}$$

$$\dot{Q}_{rh} = \dot{m}_{rh}(h_{rh_{out}} - h_{ph_{in}}) \tag{3.4}$$

$$\dot{Q}_{bd} = 0.015\dot{m}_{ms}h_{bd} \tag{3.5}$$

$$\dot{Q}_{b} = \dot{m}_{coal}HHV \tag{3.6}$$

$$\dot{Q}_{f} = \dot{m}_{fuel}SHC_{coal}(T_{mill} - T_{reference}) \tag{3.7}$$

$$\dot{Q}_{pa} = \dot{m}_{pa}hc_{air}(T_{mill} - T_{reference}) \tag{3.8}$$

$$\dot{Q}_{sa} = \dot{m}_{sa}hc_{air}(T_{sa} - T_{reference}) \tag{3.9}$$

The specific heat capacity (SHC) of the coal is calculated from the thermal capacity of composition of the coal. Both secondary and primary air come from the air pre heater, presenting a considerable higher temperature than the reference temperature.

3.2.2 Mills

The mills process raw coal into a dry pulverized stream to feed the furnace throughout the burners (Figure 3.2).

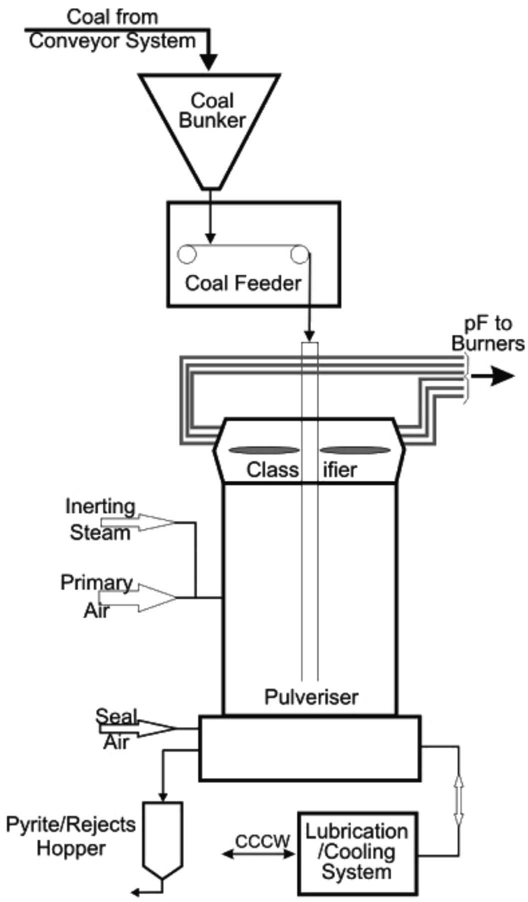


Figure 3.2 – Simplified pulverized fuel system [Doosan Babcock Energy, 2011]

Raw coal is delivered as required from the stockpile to the coal bunker via a conveyor system installed above the bunkers. Coal in the bunker flows by gravity to the coal feeder below. The coal feeder is a variable speed horizontal weighing conveyor. Coal discharged from the feeder flows down to the pulverizers grinding bowls. The pulverized coal overflows the grinding bowl and is carried upward by a flow of hot primary air to the burners. Larger coal particles fall back into the grinding zone while the finer material flows upward to the classifier located at the top of the mill. Each steam generator is served by four mills which ensure coal granulometry bellow 75 μm that are arranged alongside each other along the boiler front wall. Three mills are permanently in operation while one serves as a backup.

3.2.3 Operating Modes

The power plant can operate according to three modes, namely: (i) boiler-following control; (ii) turbine-following control; and (iii) coordinated boiler turbine control.

In operation mode (i) the steam generator follows the turbine operation. The load response is rapid because the stored energy in the boiler provides the initial change in load [The Babcock & Wilcox Company, 2015].

In operation mode (ii) the steam generator is fixed to a specific thermal load condition. The turbine is the one that reacts to that imposition, generating the electric power output in a variable manner. The load response is rather slow because the turbine-generator must wait for the boiler to change its energy output before repositioning the turbine control valves to change the load. Although the power generation varies, it remains close to the project level [The Babcock & Wilcox Company, 2015].

In operation mode (iii) the steam generator and turbine controllers work together to keep the generation stable and controlled. It is a more complex operation mode, but once achieved it guarantees stability and greater reliability of the system. This operation mode explores the advantages of the two previous ones [The Babcock & Wilcox Company, 2015].

3.3 Surrogate modeling tools

Statistical methods do not allow anything to be proved experimentally, but they do allow to measure the likely error in a conclusion or to attach a level of confidence to a statement [Montgomery, 2013]. Design of Experiments (DoE) and Response Surface Method (RSM) are employed in the present work to build a surrogate model based in statistics, whose highlights are presented hereon.

3.3.1 Basic Statistics

Hypothesis testing is the process of using statistics to determine the probability of whether the proposal hypothesis is true. The null hypothesis H_0 states that there is no significant correlation between the parameters. The hypothesis test starts by defining the test significance level α and comparing it to the test p-values. The test result is positive if p-value $< \alpha$, meaning that the alternative hypothesis is true and rejecting H_0.

The p-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 or the smallest level at which the data are significant [Montgomery, 2013].

Testing data for normality is a common and necessary step in the analysis of DoE problems as well as the examination of the residuals should be an automatic part of any analysis of variance [Mathews, 2005]. As DoE is based in analysis of variance (ANOVA), the assumptions include normality, constant variance and independence. These assumptions must be checked to validate the model using residuals plots, including: normal probability plot, histogram of residuals, residual versus fitted values and residual versus observation order. Once these assumptions are satisfied, then ordinary least squares regression produces unbiased coefficient estimates with the minimum variance [Montgomery, 2013].

The Normal Probability Plot (NPP) of residuals versus their expected values should approximately follow a straight line for normal distributions, and the residual histogram should look symmetric with about the same shape on each side for all observations. If residuals do not follow a normal distribution, the confidence intervals and p-values can be inaccurate [Montgomery, 2013]. The variance of residual terms must be constant with a mean of zero, otherwise the model may not be valid.

Residual plots also help in the identification of outliers. If one residual is larger than any of its neighbours, it can either be a mistaken result included in the analysis or brought from some ignored external influence, and both can mask the effect of the significant factors and results can be compromised. Although outliers require special attention, it is not recommended to reject or discard an outlying observation unless it has a justification [Mathews, 2005; Montgomery, 2013].

3.4 Design of Experiments

Design of Experiments (DoE) refers to the process of experiment planning, designing and analysis so that valid and objective conclusions can be drawn effectively and efficiently [Antony, 2014]. The set of experiments to be performed is expressed in the form of a design matrix, according to a chosen experimental design.

Understanding cause-and-effect relationships in a process or system include changing their input variables. Each set of input conditions is an experiment. An experiment includes uncontrollable and controllable parameters, called factors. The experimental design objective is to minimize the effects of the uncontrollable parameters and determine the effect of the controllable parameters, their interactions, the most influential, and their order of importance [Antony, 2014; Montgomery, 2013].

Changes in the average response due to factor swiping within a defined range or level is defined as an effect. 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. DoE allows recognizing and quantifying variable interactions so that they can be used to understand and better manage the response, in opposition to the One-Variable-at-A-Time (OVAT) or One-Factor-A-Time (OFAT) methods [Antony, 2014; Mathews, 2005].

DoE execution demands process knowledge and careful planning, including the determination of what to be measured in the experiment, the capability of the measurement system in place, which factors can be controlled for the experiment and the number of levels of each factor and its range. There are at least two levels: high and low. The most important issue is the choice of the highest and lowest levels, as they define the range of the factor [Antony, 2014; Mathews, 2005] to avoid risk and to guarantee safety to the operation.

Close levels may prevent to observe differences on system responses and important information about the process can be missed. There are several ways to choose the level spacing whenever three or more factor levels are considered, and the most common choice for a three-level factor is to use equally spaced intervals. Levels with constant increments are called a linear scale factor [Mathews, 2005].

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 [Antony, 2014; Montgomery, 2013]. The purpose of randomization is to remove all sources of extraneous variation which are not controllable in real-life settings. In other words, randomization can ensure that all levels of a factor have an equal chance of being affected by noise factors [Antony, 2014]. Replication means repetitions of an entire experiment or a portion of it, under more than one condition. The replication can decrease the experimental error and thereby increase precision. It allows the experimenter to obtain a more accurate estimate of the experimental error, a term which represented the differences that would be observed if the same experimental settings were applied several times to the same experimental units, as the operator, machine, material, etc. It also permits the experimenter to obtain a more precise estimate of the factor and interaction effect. However, replication can result in a substantial increase in the time needed to conduct an experiment. Their use in real life must be justified in terms of time and cost [Antony, 2014]. Blocking is a method of eliminating the effects of extraneous variation due to noise factors and thereby improving the efficiency of experimental design. If a factor must be included in an experiment but it is not the objective to make claims about differences between its levels then the levels of the factor are used to define blocks of experimental run. The idea is to arrange similar or homogenous experimental runs into groups, called blocks. Generally, a block is a set of relatively homogeneous experimental conditions. Variability between blocks must be eliminated from the experimental error, which leads to an increase in the precision of the experiment [Antony, 2014; Mathews, 2005].

3.5 Modeling approach

The methodology proposed in the present work follows three steps: planning and execution of the experiments with DoE, model fitting through RSM, and result analysis to build a surrogate model representing the system. Particular attention was given to DoE, justified by the importance of the planning phase. The proposal methodology is described in Figure 3.3.

The planning and execution phase following the DoE methodology is highlighted in yellow, and according to Antony, 2014, is crucial to the success of experiments. The design matrix with the necessary experiments to be carried out at the power plant is defined, based on six steps. The control volume (step 1) defines the scope of the study and its boundaries, by selecting the whole plant or some sub-system, such as the steam generator. Steps 2 to 4 follow the well known DoE procedure, and allows to chose the experimental design method in step 5. It must balance the amount of experiments with the available time and resources to conduct them, by taken into account the factor types and nature, replication, and blocking. The resulting design matrix contains the controlled factors, their levels and the experiment running order. The sixth step deals with procedures for conducting the experiments at the power plant.

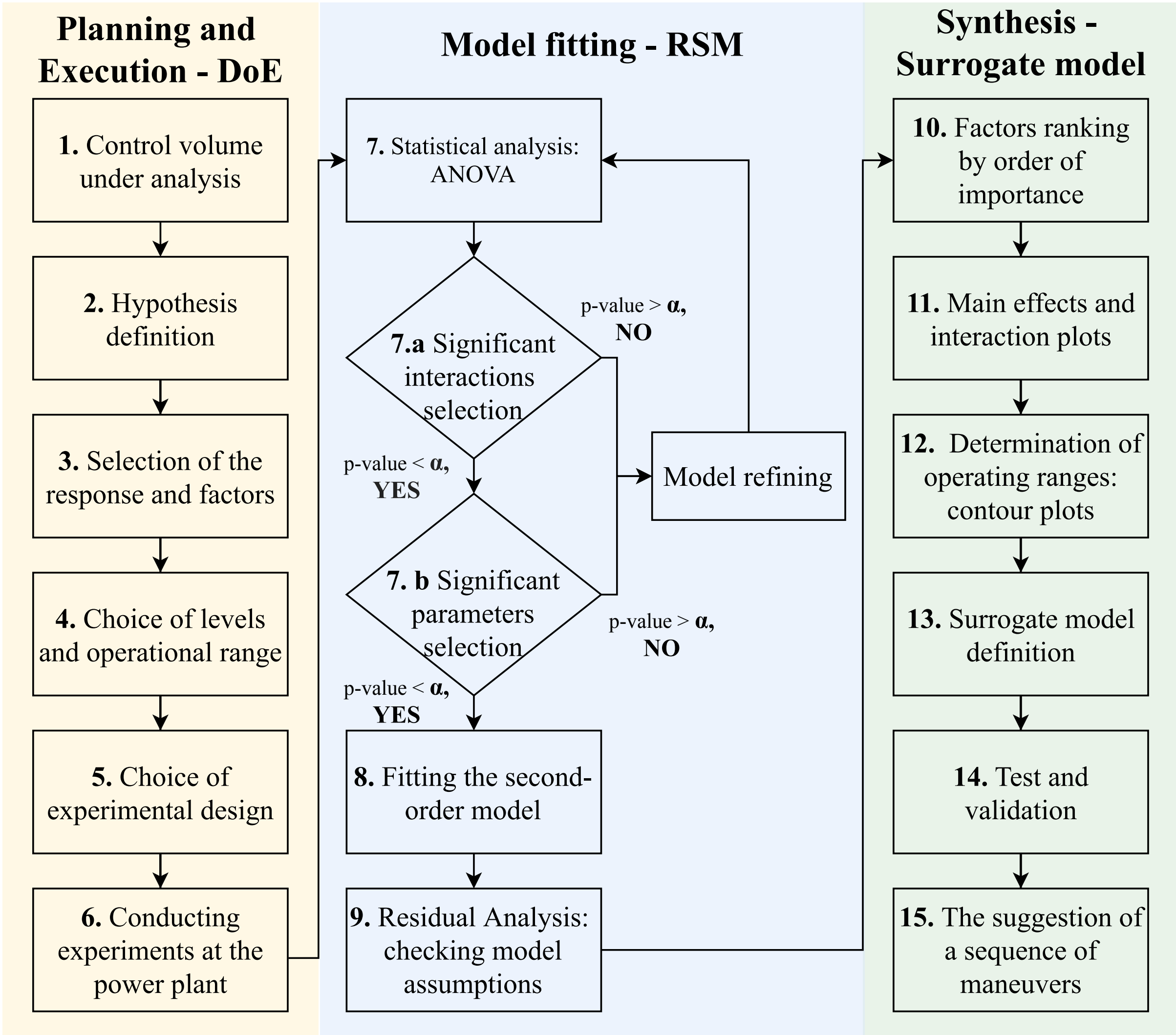


Figure 3.3 – Step by step to the construction of a surrogate model to a power plant

The whole execution process must be commented, as well as the occurrence of external variations that may generate interferences. In the case of system destabilization, the operator must return to regular operating conditions, for safety reasons. Any experiment that may cause an operating problem must be immediately suspended as safety is paramount. Results collection must be done at stable regime, i.e., when measurements do not change over time and not be adversely influenced by the operator and environmental changes [Antony, 2014].

The model fitting phase, highlighted in blue, builds a response surface model (RSM) out from the collected data. The seventh step employs Analysis of Variance (ANOVA) with the aid of MINITAB® to test the hypothesis defined at the beginning of the study, based on the definition of a confidence interval, and its complementary significance level (α). The interactions between factors are tested in step 7.a, starting with the higher-order interactions. The null hypothesis H_0 is rejected for p-value$<\alpha$, meaning that the interaction is significant, otherwise (p-value$>\alpha$) the interaction is removed from the model and the process restarted. This step is repeated until all remaining interactions in the model are considered as significant. Step 7.b tests the significance of individual factors. The null hypothesis H_0 is rejected for p-value$<\alpha$, which indicates that the effect of a given factor is significant. At the end of the seventh step, only the significant terms according to the response remain in the model. It is worth mentioning that if an interaction of a factor is significant than automatically the factor remains in the model (even if p-value $> \alpha$).

Step nine contains the residual analysis to check the model assumptions of normality, constant variance, and independence. Four residual plots are made in this step, namely normal probability plot, a histogram of residuals, residual versus fitted values and residual versus observation order. The simplest model that produces random residuals is a good candidate for a relatively precise and unbiased model. If some of the model assumptions could not be verified, the conduction of new analysis would become necessary. The possibilities include a missing variable, a missing higher-order term of a variable in the model to explain the curvature or a missing interaction between terms already in the model [Mathews, 2005; Montgomery, 2013].

The last phase, highlighted in green, builds a surrogate model to standardize the operation based on the analysis of the results of the previous steps. The first action concerns ranking the factors (controlled parameters) in descending order of importance in the response, in order to determine the optimal settings that minimizes variability. Key parameters are identified and ranked on a Pareto plot at the tenth step.

The 11^{th} step is the construction of the main and interaction effects plot to analyze factors behavior. This is necessary to determine the settings that yield the best performance to improve steam generator efficiency. Step 12^{th} settle operating ranges to divide the regions in which the important factors lead to the best possible response. The lines of constant yield are connected to form response contours using contour plots. These contours are projections on the interest regions [Montgomery, 2013].

The 13^{th} step defines the surrogate model as the final equation of the previous steps, which assures that only the significant terms are present. The 14^{th} step tests the proposed surrogate model to it is validation. New predictions are made at certain positions within the design space where no data points existed previously. At this moment it is essential to look at the results critically and use the subject knowledge area to evaluate if the results make sense.

In closing, the surrogate model is used in 15^{th} step to provide a sequence of maneuvers to the operator considering only the significant factors (controllable parameters). The operator order of action is defined according to the importance of the factor and the best operational ranges by factor are settled ensuring a standardized operation.

3.6 Pecem power plant: a logbook to build a surrogate model

The modeling approach presented in section 3.5 is applied to the case study of the PECEM power plant. Boxes depicted at Figure 3.3 flowchart are detailed and the new challenges that emerge during the process are discussed. The power plant assessment was carried out for the 360 MW eletrical output base load, as factor levels can display different ranges according to the plant load. It is worth recalling that the research goal is to standardize the steam generator operation in order to improve its performance and thereafter impact the plant overall behaviour.

3.6.1 Step 1 - Control volume

The natural choice of control volume (CV) is around the steam generator, but mills were included as they are directly related to the system performance. Coal consumption, air flow, granulometry, flame stability, among others, are all influenced by the mill activity. Four independent mills are connected to each plant steam generator, but the 360 MW base load operation is usually performed with a combination of three devices to rationalize costs and maintenance.

Mill operation can be assessed individually or by their average. In the first case, controllable factors may respond differently according to the equipment condition, which leads to the choice to adopt in the present study the individual analysis.

3.6.2 Step 2 - Hypothesis definition

Considering the objective of standardizing operation to reduce process variability related to the operator action, it is mandatory to identify which parameters are significant to the response. The significance of the controllable parameters and their interactions are the hypothesis to be tested.

3.6.3 Step 3 - Selection of the response and factors

The selection began with the identification of critical process parameters. In this study, the interest is in characterizing the efficiency of the steam generator as the response. The efficiency is capable of representing the performance of the steam generator in a single parameter and for this reason was selected. The factors (controllable parameters) selection requires the assurance of controllability and independence between them. The whole process of parameter selection is described in Figure 3.4.

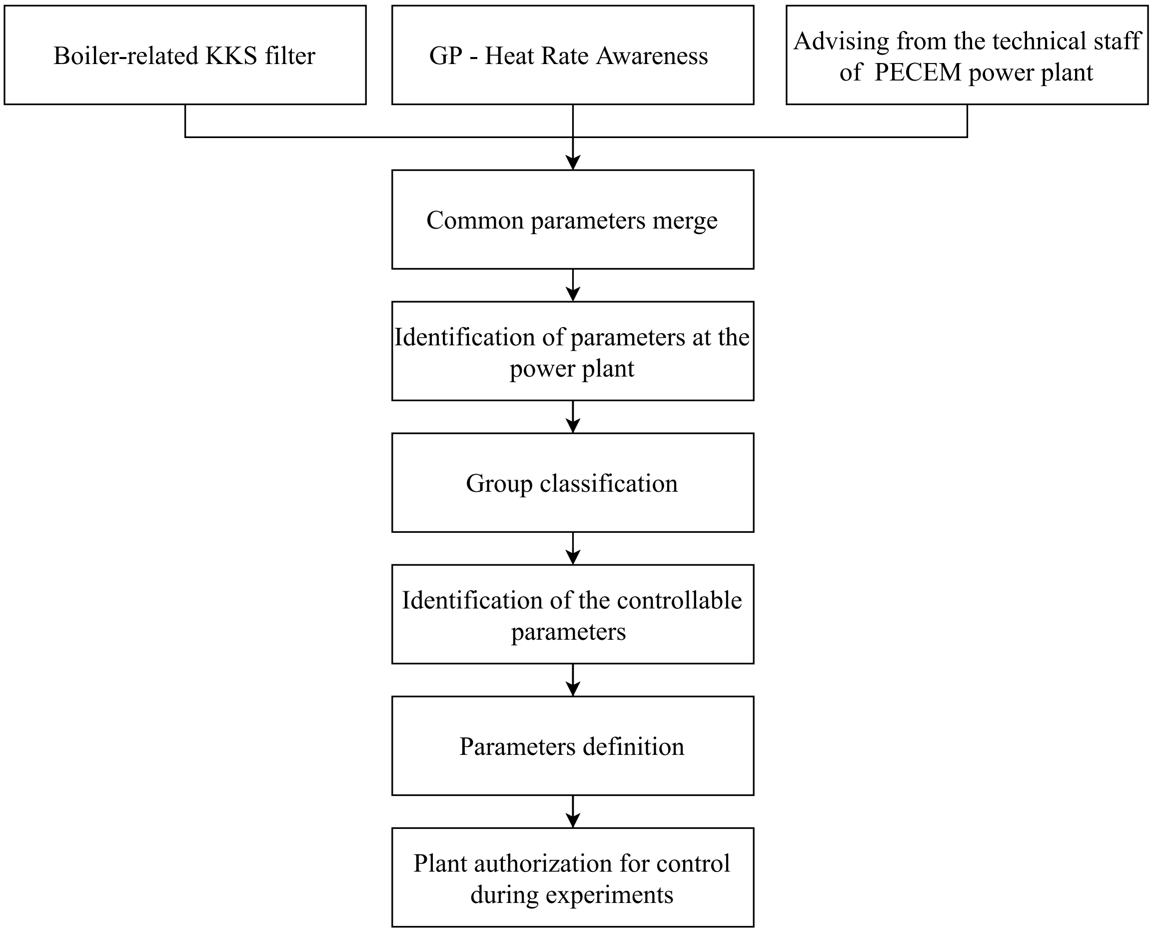


Figure 3.4 – Parameters selection flowchart

System parameters are identified using equipment identification codes (KKS). The initial list of parameters considered three sources: boiler-related KKS filter, GP Strategies Corporation, 2013, and advising from PECEM technical staff. The boiler-related KKS filter only concerned thermodynamic properties, such as temperatures, pressures, and flows. This choice was made based on system knowledge and focused on heat transfer processes. Boiler-KKS related filter resulted in a total of 29 parameters.

GP Strategies Corporation, 2013, was the second source, it is a textbook designed for operators, supervisors, engineers, and managers who are directly involved in the daily operation of power plants. Its content includes heat rate concepts, controllable and non-controllable losses, and the effects of components performance on operating costs. The textbook lists parameters considered as significant in line with controllable losses, and 52 were selected. Finally, new parameters were added based on the advising of PECEM power plant technical staff.

Parameters were grouped and the common ones were merged, identified by their respective KKS. Of the 52 parameters from GP Strategies Corporation, 2013, 11 were related to the boiler. The addition of new parameters related to the initials was performed by PECEM team, because not all the selected parameters had direct measurement or control, being thus observed through other parameters.

In summary, the list from GP Strategies Corporation, 2013, had 11 parameters related to the boiler and mills, boiler-related KKS accounted 29 and technical staff of PECEM power plant accounted more 23. The total number os parameters considering the three sources were 63.

The last classification was based on controllable and uncontrollable parameters. The controllable parameters by definition can be directly impacted by the actions of the unit control operator [GP Strategies Corporation, 2013]. Uncontrollable parameters or noise variables are those which are difficult or expensive to control or that cannot be controlled by the experimenter but can be monitored and included in the statistical model. The effect of such nuisance variables can be understated by the effective application of DoE principles such as blocking, randomization, and replication [Antony, 2014; Lujan-Moreno et al., 2018]. The controllable parameters account for a total of 11 and are presented in Table 3.1.

Table 3.1 – Selection of controllable parameters according to Figure 3.4

Group	Parameter	Manipulation
Mill	Primary air flow (kg/s)	By mill
	Pulverized coal outlet temperature (°C)	By mill
	Speed of the dynamic classifier (rpm)	By mill
	Coal mass flow rate (t/h)	By mill
Burner row	Secondary air flow (kg/s)	By burner row
	Stoichiometry* (dimensionless)	Sub-stoichiometric region
Air/fuel control	Excess O_2 (%)	Burnout zone
Boiler	Secondary air crossover duct pressure (mbar)	-
	Primary air crossover duct pressure (mbar)	-
	Power generation (MW)	Operation mode
	Differential pressure of the boiler feed valve (bar)	-
	Superheated spray flow (t/h)	-

* The stoichiometry and secondary air flow have alternative controllability

The first column at Table 3.1 refers to the group inside the control volume, second the name of the parameter and third refers to the intervention model of the operator in that specific parameter. The parameters related to the mill are controlled and can have different operational conditions by mill as well as the parameters related to the burner row, controlled by burner row.

The parameters of secondary air flow and stoichiometry are directly related and have alternative controllability. The second air flow is controllable, but if the operator changes the stoichiometry then the secondary airflow will change to respect that condition. The primary air flow influences the stoichiometry too, but the former is set at the mill and is more related to coal transport and moisture control (drying). The control system ensures that the flow and temperature of the primary air are sufficient for the current coal demand, while the secondary air flow is set at the burner row and it is more related to the direct control of boiler stoichiometry.

Once defined the parameters, the implementation of the actual DoE maneuvers or tests in the field demanded operational adequacy and therefore the agreement from the power plant management. Parameters shown in Table 3.1 were assessed by the PECEM technical team. It was pointed out that the boiler differential pressure and superheated spray flow should be removed as they can not actually be controlled. Coal flow and power generation were also removed due to authorization issues, related to the connection with the national electric grid although they are controllable parameters. It is worth mentioning that the air differential pressure at the mill is an important operational parameter, but it was removed because it cannot be controlled by the operator.

The steam generator efficiency was chosen as the unique response to be observed in the present work, among many other possible ones, because it can adequately resume the equipment performance. Each selected factors for the conduction of DoE at PECEM power plant were renamed from P1 to P7 and the response as S1, as depicted in Figure 3.5.

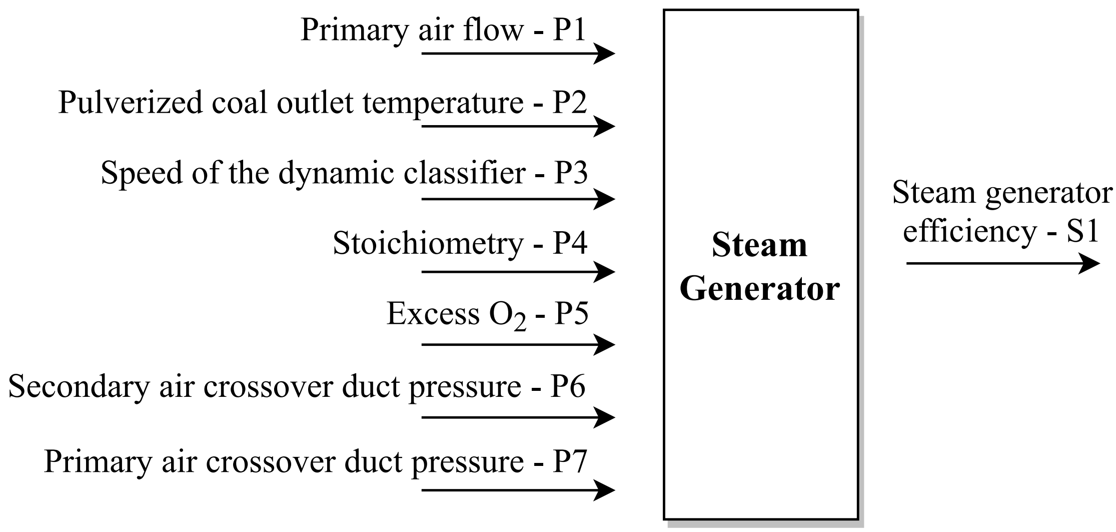


Figure 3.5 – Representation of the steam generator process using DoE on PECEM power plant.

Of the seven factors, three concerns the mills (P1 to P3) and the remaining ones are related to the steam generator. The controllable parameters P1 to P7 were situated in the schematic layout of steam generator and mills presented in Figure 3.6 according to the previous one Figure 3.1.

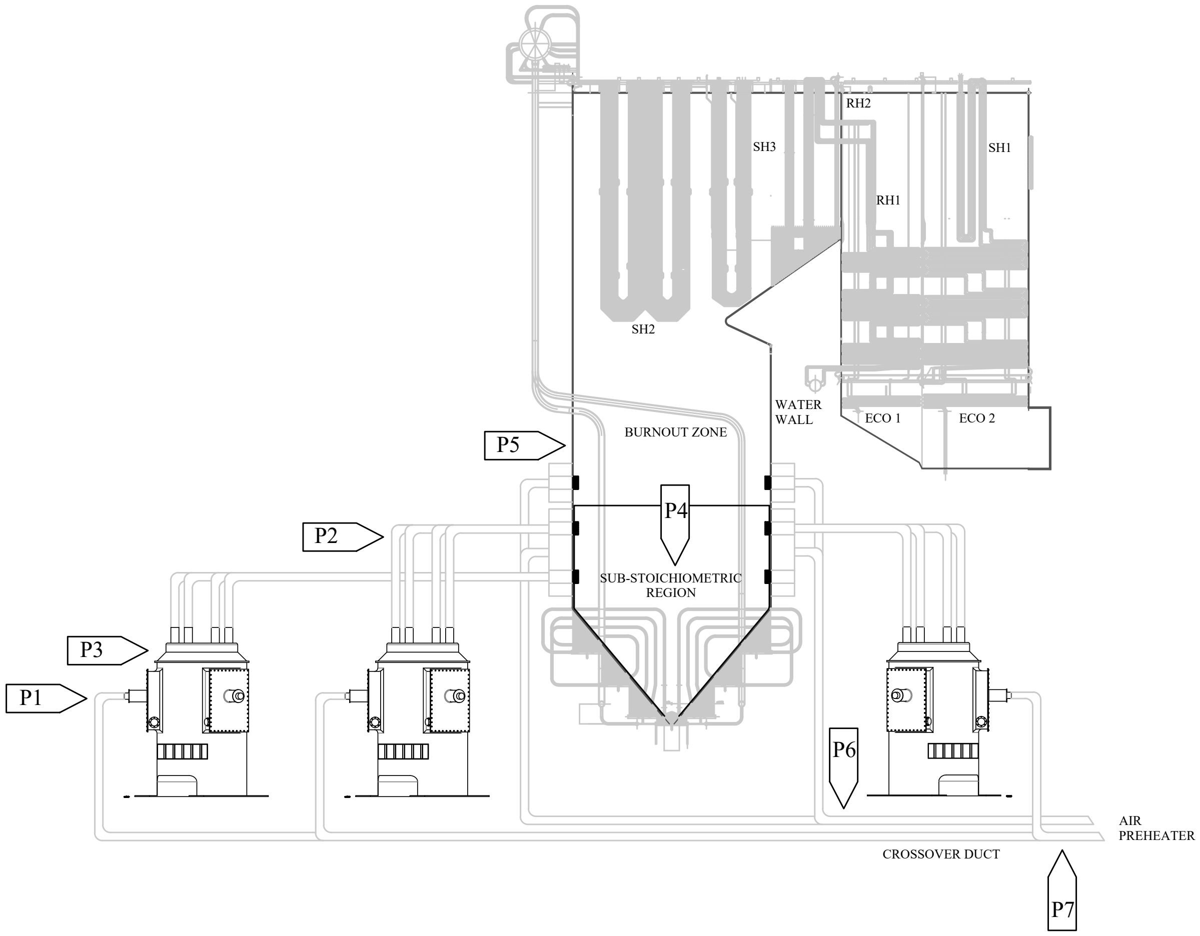


Figure 3.6 – Controllable parameters location according to the schematic layout presented in Figure 3.1.

Hot air flow from the air preheater serves both the primary and secondary streams via two independent systems, called the crossover ducts. The primary crossover duct supplies hot primary air to each of the three mills, who receive raw coal to be pulverized. The primary air flow (P1) has two prior functions that are to perform coal drying and then convey it to the burners, already pulverized. P1 is directly related to coal granulometry. Pulverized coal outlet temperature (P2) is measured at the mill outlet and is related to the coal drying process. A lower value must be guaranteed to reach drying requirements, but at the same time, it should not cause coal auto-ignition [Doosan Babcock Energy, 2011]. The speed of the dynamic classifier (P3) is the last parameter related to the mill, directly related to the coal granulometry, as well as P1. A schematic layout of the PECEM dynamic classifier is presented in Figure 3.7.

The dynamic classifier performs the second stage of pulverized coal classification

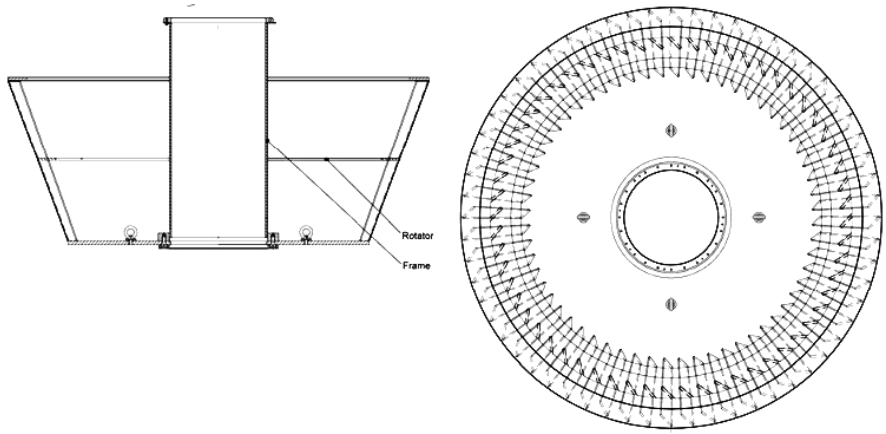


Figure 3.7 – Schematic layout of the dynamic classifier [Doosan Babcock Energy, 2011].

according to the speed of the classifier rotor. The rotor imparts a centrifugal force on the coal particles which, if greater than the force of the rising primary air stream, causes the coal to be returned to the pulveriser grinding zone. Finer material remains in the primary air stream and is conveyed to the burner [Doosan Babcock Energy, 2011].

The combustion air flow demand is determined by the steam generator flow and its fuel flow rate. Stoichiometry (P4) corresponds to the sub-stoichiometric region. The steam generator is divided in two burner volumes sub-stoichiometric region and burnout zone, as showed by Figure 3.6. The sub-stoichiometric region is set below 1.0 and ends at the fired air input. The excess of O_2 (P5) refers to the burnout zone and it defines the global stoichiometry of the combustion process and commands the OFA operation. 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 last two controllable parameters are the primary and secondary air crossover duct pressure (P6 and P7). The primary crossover duct supplies hot primary air to each of the three mills via control dampers while the secondary crossover duct delivers hot air to the burner windboxes [Doosan Babcock Energy, 2011]. The secondary crossover duct is lateral to the windbox and feeds both secondary air and OFA. This arrangement maintains the correct secondary air pressure for all firing conditions. The crossover duct pressure interferes with the flame stability and the speed at the burners enter of the secondary air flow, as well as its distribution. A detailed scheme of the duct arrangement is presented in Figures 3.8 and 3.9.

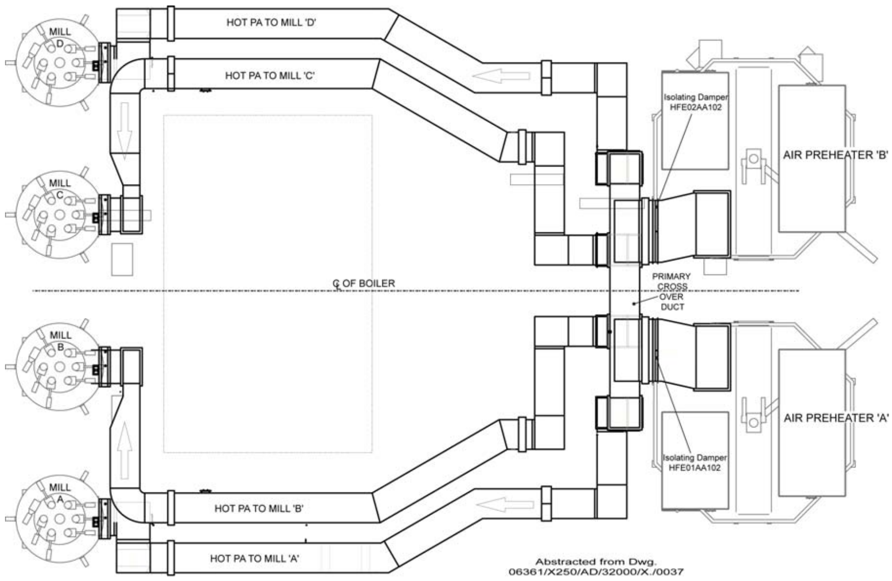


Figure 3.8 – Primary air ducting arrangement [Doosan Babcock Energy, 2011].

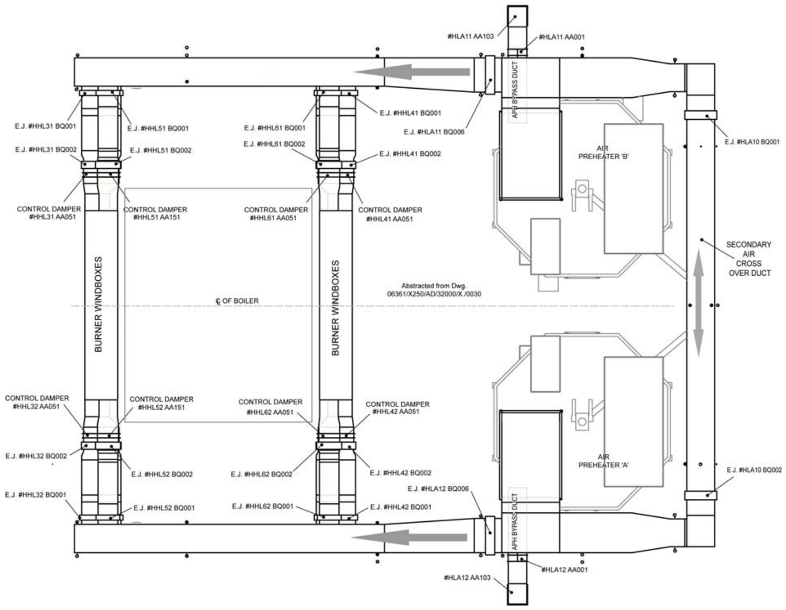


Figure 3.9 – Secondary air ducting arrangement [Doosan Babcock Energy, 2011].

3.6.4 Step 4 - Choice of levels and operational ranges

The operating range of the selected factors (controllable parameters) are determined according to the plant history to provide safe and stable conditions. Experiments must not cause additional stresses to the power plant, but to standardize operation ensuring safety.

Acquired data from Unit 2 were gathered for the electric power output within the range of 340 to 360 MW from January 2018 to May 2019, which led to 4738 registers with the seven selected factors. Mill factors (P1, P2, and P3) refer exclusively to mill A. A first assessment was carried out using behavior graphs to identify factors ranges and variability. The graphs are presented in Appendix B (B.1). The only factors with a clear behavior change was the speed of the dynamic classifier (P3) and the stoichiometry (P4) which are presented in Figure 3.10 and 3.11.

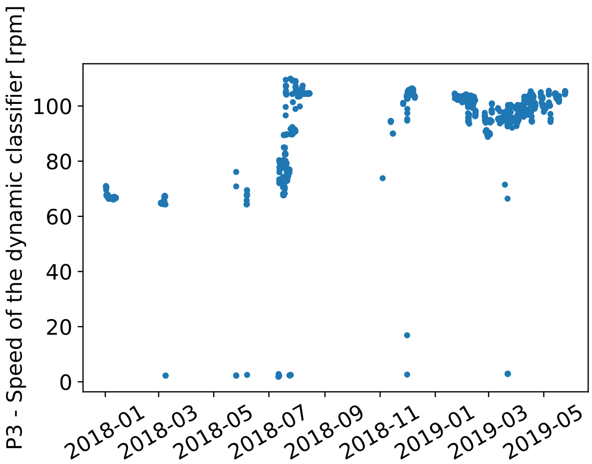


Figure 3.10 – Speed of the dynamic classifier versus time for group 2 @ 360 MW baseline

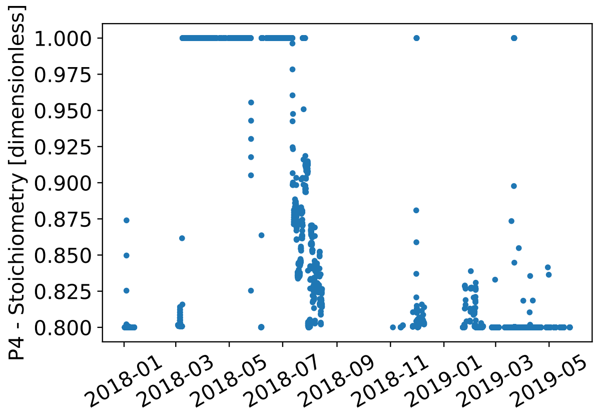


Figure 3.11 – Stoichiometry versus time for group 2 @ 360 MW baseline

Dynamic classifier speed (P3) was found to work around 70 rpm till June 2018, to rise to a reference baseline of 100 rpm from that moment on, due to a mill retrofitting. Stoichiometry (P4) found mostly to range from 0.8 to 1.0, with some intermediate points. For this reason, the data was filtered from June 2018 for this factors.

Periods with null data were common to all parameters and correspond to power plant shut down or power output leveled to 240 MW baseline.Results are summarized and detailed in Table 3.2.

The removed outliers were defined as the factor mean value minus 1.5 times the standard deviation. That criterion allowed to remove all values on the outside of the baseline range of 340 to 360 MW, which lead to 12 % lost of raw data. The first reasonable range for the parameters can be given by quartiles Q1 to Q3.

Table 3.2 – Statistics of the seven controllable parameters of group 2 from January 2018 to May 2019 @360 MW power output baseline

Factors (controllable parameters)		**Mean**	**Q1 (25%)**	**Q2 (50%)**	**Q3 (75%)**	**Max**	**Standard deviation**
Primary air flow (kg/s)	P1	24.91	22.90	25.37	26.58	28.83	1.91
Pulverized coal outlet temperature (°C)	P2	75.42	72.88	75.59	78.37	83.90	3.32
Speed of the dynamic classifier (rpm)	P3	100.20	96.01	102.25	104.47	109.92	5.03
Stoichiometry	P4	0.81	0.80	0.80	0.81	0.92	0.03
Excess O_2 (%)	P5	2.91	2.52	2.88	3.27	4.46	0.43
Secondary air crossover duct pressure(mbar)	P6	18.02	16.49	17.67	18.78	23.97	2.16
Primary air crossover duct pressure (mbar)	P7	86.19	78.93	87.70	93.07	94.35	6.68

As a complementary approach, each factor was plotted with respect to the steam generator efficiency followed by its respective heatmap. The proposal was to perform a pre-evaluation of the relevant parameters to the output. However, there is no clear and precise relationship between the factors and the efficiency of the steam generator (S1). The graphs are available in Appendix B.1 It is worth remembering that the graphs presented are not capable of representing the interaction between the parameters, but they serve only as an indicator of their individual effect.

The historical data does not enable the knowledge of the real conditions of the power plant at that time. Simple data analysis does not allow to conclude if the power plant is under normal operation. Coal moisture due to the rain, or unusual equipment behavior, for instance, cannot be observed with this approach. Thus, the performance of experiments through DoE is essential because it performs an analysis accompanied by the parameters and focused on maintaining the conditions of the other parameters stable. It is not possible to ignore the operation history, but it is also not possible to evaluate only by history.

Table 3.3 summarizes the main values collected for the controllable parameters for group 2 operating on the 340 to 360 MW range.

Table 3.3 – Summary of factors (controllable parameters) operation range and respective levels

Factor	Lower Level	Medium Level	Upper Level	Description
P1 (kg/s)	24.0	26.0	28.0	Primary air flow
P2 (°C)	65	75	85	Pulverized coal outlet temperature
P3 (rpm)	90	100	110	Speed of the dynamic classifier
P4 (dimensionless)	0.80	0.88	0.95	Stoichiometry
P5 (%)	1.5	2.3	3.0	Excess O_2
P6 (mbar)	18	21	23	Secondary air crossover duct pressure
P7 (mbar)	70	78	85	Primary air crossover duct pressure

The operational range of each parameter was defined with the assistance of the PECEM technical team, and limits were changed according to their experience and recommendations. It can be noticed that ranges are somehow limited but it always tried to reach the compromise of improving efficiency by respecting plant safety.

3.6.5 Step 5 - Choice of Experimental Design

The choice of the experimental design is directly associated with the costs and the available time for carrying on the experiments. This step corresponds to the number five of the flowchart (Figure 3.3) and involves the sample size, number of replicates, selection of the randomized order of experimentation, necessity of blocking and analysis of any restriction involved. Fitting and analyzing response surfaces is greatly facilitated by the proper choice of experimental design.

The number of required experiments by the full factorial 3^k is expressively bigger than the BBD and CCD. The full factorial 2^k is presented just as a comparative, because it is not able to consider second-order terms. The advantages of BBD become more representative as the number of factors increases. BBD stands out with only 62 experiments for 7 parameters whereas 3^k proposes 2187 experiments. On the top of it, BBD does not need to perform experiments at the range limits or extremes. This could be advantageous when the corner points could be prohibitively expensive or impossible to test because of physical process constraints [Myers et al., 2016; Montgomery, 2013].

Table 3.4 – Number of experiments according to the number of factors (controllable parameters) and the experimental design

		BBD	CCD based on half factorial	CCD based on full factorial	Full Factorial 2k	Full Factorial 3k
Number of factors	7	62	88	152	128	2187
	6	54	53	90	64	729
	5	46	32	52	32	243

* The experiments do not consider replication and blocking.

Finally, blocking and replication were not considered in the study. Although replication reflects sources of variability both between runs and (potentially) within runs there was a limitation imposed by the technical team of PECEM power plant.

3.6.6 Step 6 - Real Life Experiments

Execution of the experiments is prescribed in the sixth and last step of the planning and execution phase (Figure 3.3). The major decisions to conduct DoE during the operation are summarized in Figure 3.12. They are divided in the five big areas: coal, instrumentation, sampling, execution schedule and mills. Each of them is divided into subareas, according to their segmentation during the process. The decisions assumed during the process are practical recommendations that go beyond traditional analysis. Although taken for the PECEM case study, they can serve as a model for other generating plants.

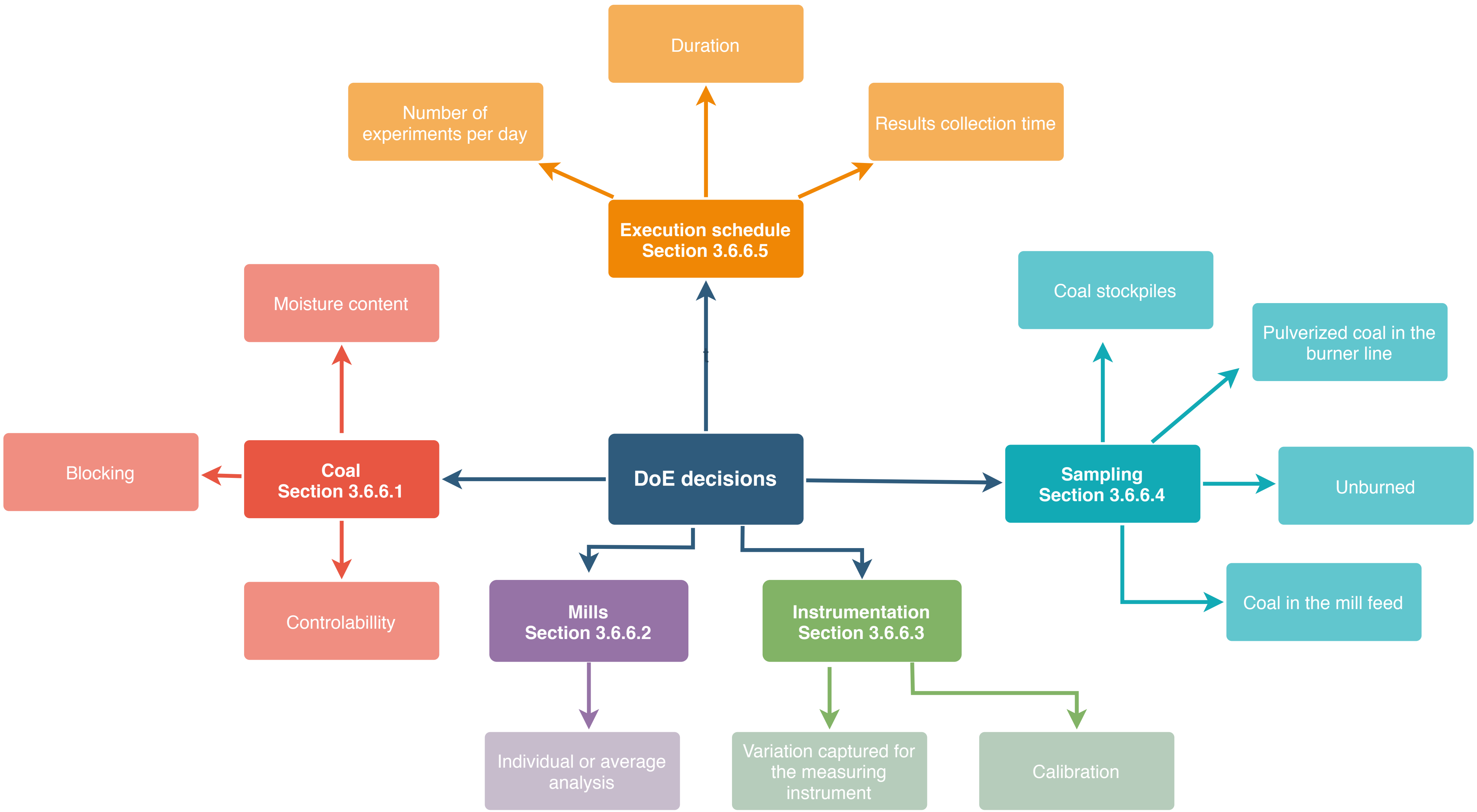


Figure 3.12 – Major decisions to conduct DoE in an operating coal-fired power plant

3.6.6.1 Coal

Blocking

The first raised question concerned the impact of coal variation from different ships. Coal is a controllable parameter as one can choose its type and coal flow rate. However, there is no possibility of returning to the same coal in case an experiment needs to be redone. Besides that, there is no guarantee that all experiments are able to be carried out within the same coal stockpile. An alternative to that situation is the blocking of coal by lots, which would decrease the number of experiments per type of coal, ensuring that they could be carried out within the same pile and reduce the impact of the different stockpiles if any. In the other hand, the plant coal supplier is the same, so there should be no difference between coal lots. Laboratory tests of the coal stockpiles also indicated that there was no variation between them (see section 3.6.6.4). Therefore, it was decided not to block the coal lots in the present study to reduce its complexity.

Moisture and water content

Coal moisture was a problem faced during the operation. Pulverized coal outlet temperature (P2) hardly exceeded 65°C on rainy days. Even so, moisture is not a controllable parameter, and the only action to work around the situation was to stop experiments during that period. Since the power plant is located in a region with long periods of drought. The dry season runs from August to December, the period in which the experiments were conducted. During this time rains were almost non-existent.

There is a correlation that helps identifying the moisture content, based on the confrontation of the mill performance test to the ΔT between the coal inlet and outlet temperatures. If that ΔT decreases it means that the moisture content is sinking energy. For instance, the operator expects for a coal inlet temperature of 300°C and outlet temperature of 85°C, and that output temperature does not exceed 75°C when the moisture content is high.

Controllability

Coal flow rate was not allowed to be used as a DoE factor by the PECEM, although it was assigned as a controllable parameters. The total amount of coal flow is the result of a combined operation of the three mills, which prevented to perform that experiment. A fixed coal flow of 45 ton/h was determined for mill A.

3.6.6.2 Mills

The decision was to perform the experiments by mill. Factors P1 to P3 were tested for mill A without replicates, due to their complexity. The conduction of that same experiments on the remaining mills should be presented to the technical team of the PECEM power plant for further conduction.

In the burners handling occurs by burners row, corresponding to the mill in operation. Subsequently, the same sequence of experiments must be performed for the other mills. The obtained results from the experiments in the mill A should be presented to the technical team of the PECEM power plant in order to justify the conduction for the remaining mills.

3.6.6.3 Instrumentation

The measurement system is an essential part of any experiment and must be capable, stable, robust and insensitive to environmental changes. It is necessary to ensure that the equipment exists and is accessible and calibrated. As stated by Antony, 2014, the quality of a measurement system is usually determined by the statistical properties of the data it generates over a period of time which captures both long and short-term variations.

The survey of equipment uncertainty data, measurement interval and calibration documents were carried out for all KKS directly related to the experiment parameters or that were included in the calculation of non measured parameters. At the power plant, the instruments capture a variation above 0.5% of the value of the measurement range.

3.6.6.4 Sampling

Four different coal samplings were planned at the PECEM site in order to characterize fuel conditions.

Coal stockpiles

That sampling aimed to verify if the incoming raw coal composition was in agreement with the one declared by its supplier, and therefore exclude the possibility of including coal type as a DoE factor.

Samples of five types of coal delivered to PECEM were collected in September

2018 following ASTM International, 2019, methodology. Results discarded the eventual blending of the delivered coal shipments. In parallel to that procedure, it is worth noticing that coal can change along operation due to its turnover, and further investigation about the eventual impacts were performed.

Pulverized coal in the burner row and unburned fuel

An experimental investigation was performed to investigate the impacts of the primary air flow (P1) and the dynamic classifier speed (P3) on coal granulometry and ash carbon content. Table 3.5 planning composes a set of samples to be performed with the DoE experiments.

Table 3.5 – Sampling planning to pulverized coal in the burner row and unburned

P1	P3	Sample
24 kg/s	90 rpm	1
	100 rpm	2
	110 rpm	3
26 kg/s	90 rpm	4
	100 rpm	5
	110 rpm	6
28 kg/s	90 rpm	7
	100 rpm	8
	110 rpm	9

During the experiments, when they are in the primary air conditions and dynamic classifier speed described in Table 3.5, it should be carried out the pulverized coal collection in the burner row and unburned. It is not necessary to follow the collection order and only one collection per condition needs to be performed.

Coal in the mill feed

Moisture is a non-controllable parameter that can impose penalties on the overall operation. Fuel sampling at the mill inlet is a difficult task and offers risk to the operation, and it should be carried out before the storage silo. Although it is not so effective, it was preferred due to safety reasons.

It is expected to perform the analysis of the moisture present in the coal at least twice a day. Programming two collections per day to measure the surface moisture in the coal. Planning was to use the sampler on the conveyor that leads to the silos.

3.6.6.5 Execution schedule

The design matrix (Table 3.6) was proposed to the team in charge of conducting the experiments prior to its execution. It gathers all factor settings at different levels and their running order [Antony, 2014]. It is important to assure a second approval from that team due to the technical and cost aspects.

Table 3.6 – Execution schedule of the experiments

Date	Experiment number	Responsible operator	Adjustments start time	Adjustments end time	Experiment end time	Coal Stockpile	Factors (controllable parameters) P1 (KKS)	P2 (KKS)	P3 (KSS)	P4 (KKS)	P5 (KSS)	P6 (KSS)	P7 (KSS)	Sampling	Sootblowing*	S1 (KKS)
	1						26.0	65	90	0.88	2.3	23	78			
	2						24.0	75	100	0.88	2.3	18	70	proceed sampling		
	3						26.0	75	100	0.80	3.0	23	78			
	4						26.0	75	100	0.80	3.0	18	78			
	5						26.0	75	100	0.88	2.3	21	78			
	6						24.0	75	110	0.88	3.0	21	78	proceed sampling		
	7						26.0	65	100	0.88	3.0	21	85			
	8						28.0	75	100	0.88	2.3	23	85			
	9						26.0	85	100	0.88	3.0	21	85			
	10						24.0	75	100	0.88	2.3	18	85			
	11						24.0	85	100	0.95	2.3	21	78			
	12						28.0	75	110	0.88	3.0	21	78			
	13						24.0	75	90	0.88	3.0	21	78	proceed sampling		
	14						26.0	65	110	0.88	2.3	23	78			
	15						26.0	85	90	0.88	2.3	18	78			
	16						28.0	75	100	0.88	2.3	23	70			
	17						24.0	85	100	0.80	2.3	21	78			
	18						26.0	85	100	0.88	3.0	21	70	proceed sampling		
	19						26.0	65	110	0.88	2.3	18	78			
	20						26.0	75	100	0.88	2.3	21	78			
	21						28.0	85	100	0.80	2.3	21	78			
	22						28.0	65	100	0.95	2.3	21	78			
	23						26.0	75	100	0.95	3.0	18	78			
	24						24.0	65	100	0.80	2.3	21	78			
	25						26.0	75	110	0.80	2.3	21	70			
	26						28.0	75	100	0.88	2.3	18	85			
	27						26.0	75	100	0.95	3.0	23	78			
	28						26.0	75	100	0.88	2.3	21	78			
	29						28.0	75	100	0.88	2.3	18	70			
	30						26.0	75	100	0.88	2.3	21	78			

Continued on next page

Table 3.6 – *Continued from previous page*

							Factors (controllable parameters)									
Date	Experiment number	Responsible operator	Adjustments start time	Adjustments end time	Experiment end time	Coal Stockpile	P1 (KKS)	P2 (KKS)	P3 (KSS)	P4 (KKS)	P5 (KSS)	P6 (KSS)	P7 (KSS)	Sampling	Sootblowing*	S1 (KKS)
	31						26.0	75	90	0.80	2.3	21	85	proceed sampling		
	32						24.0	75	110	0.88	1.5	21	78			
	33						26.0	75	100	0.95	1.5	18	78			
	34						26.0	75	90	0.80	2.3	21	70			
	35						26.0	75	100	0.95	1.5	23	78			
	36						26.0	85	110	0.88	2.3	18	78	proceed sampling		
	37						26.0	75	110	0.95	2.3	21	70			
	38						24.0	75	90	0.88	1.5	21	78			
	39						26.0	75	90	0.95	2.3	21	85			
	40						26.0	65	100	0.88	1.5	21	85			
	41						26.0	65	100	0.88	1.5	21	70			
	42						28.0	75	90	0.88	1.5	21	78	proceed sampling		
	43						24.0	65	100	0.95	2.3	21	78			
	44						26.0	85	110	0.88	2.3	23	78			
	45						24.0	75	100	0.88	2.3	23	85			
	46						26.0	75	90	0.95	2.3	21	70			
	47						28.0	85	100	0.95	2.3	21	78	proceed sampling		
	48						26.0	65	90	0.88	2.3	18	78			
	49						26.0	75	100	0.80	1.5	23	78			
	50						26.0	65	100	0.88	3.0	21	70			
	51						26.0	85	100	0.88	1.5	21	70			
	52						26.0	75	100	0.88	2.3	21	78			
	53						26.0	85	90	0.88	2.3	23	78			
	54						26.0	85	100	0.88	1.5	21	85			
	55						28.0	75	110	0.88	1.5	21	78	proceed sampling		
	56						26.0	75	110	0.80	2.3	21	85			
	57						26.0	75	100	0.88	2.3	21	78			
	58						28.0	65	100	0.80	2.3	21	78			
	59						26.0	75	110	0.95	2.3	21	85			
	60						28.0	75	90	0.88	3.0	21	78			
	61						24.0	75	100	0.88	2.3	23	70			
	62						26.0	75	100	0.80	1.5	18	78			

Adjustment start time refers to the launching of the first factor, whose calling order is not relevant. The only required condition is that each and all factors must achieve the prescribed values. The prescribed values for factors P1 to P7 are hardly reached, but they must be in accordance with their corresponding uncertainties. The adjustment end-time refers to the moment when factors effectively reached their prescribed values, under stable condition. The moment the output value is observed determines the experiment end-time. External and non-controllable factors must be monitored in order to avoid interferences in the steam generator operation, like changes in the condensing system or soot blowers effects. Attention must be paid whenever a different stockpile is selected to feed the combustion system since experimental data is acquired for a given type of coal. Sootblowing was also stopped for around 30 min. Actual manoeuvrers were slowly conducted in order to avoid plant destabilization.

Logical control options during the experiments

There are three types of logical control related to operator intervention: (i)external; (ii)internal; (iii) manual. The two first modes are automatic and differs by the way the set point is defined. The manual set point is a local command in which the operator set the value for a parameter and assumes the risk. That option can be performed as a remote manual control or as a local manual control, when a field technician acts directly on a specific equipment. Manual control was chosen to operate the experiments as it allows for faster responses than the two other modes, assuming a higher risk to the plant.

Response observation

Steam generator efficiency S1 is released from the supervisory chart after reaching a steady state regime. A copy of the command screen was captured and saved with the experiment number, together with two 30 minutes long trend graphs. Only the average value is to be retained.

Partial results

The controllable parameters were set one at a time, allowing to observe the development of the operation and to ensure safe control. Results of the performed experiments are presented in Table 3.7.

Table 3.7 – Execution of the experiments at the PECEM power plant in accordance with Table 3.6

Date	Experiment number	Responsible operator	Adjustments start time	Adjustments end time	Experiment end time	Coal Stockpile	**Factors (controllable parameters)** P1 (KKS)	P2 (KKS)	P3 (KSS)	P4 (KKS)	P5 (KSS)	P6 (KSS)	P7 (KSS)	Sampling	Sootblowing*	S1 (KKS)
8/12/2019	1	Operator A	11:36	13:05	14:20	2D	26.0	65	90	0.88	2.3	23	78		Sootblowing SH; start time 10:30 end time 13:40	91.20
8/12/2019	2	Operator A	14:20	15:23	16:24	2D	24.0	75	100	0.88	2.3	18	70	proceed sampling	Sootblowing primary SH ; start time 14:49	91.10
8/12/2019	3	Operator B	16:30	17:46	18:43	3A	26.0	75	100	0.80	3.0	23	78		Finalyzing sootblowing primary SH	90.90
8/12/2019	4	Operator B	18:47	20:05	21:05	3A	26.0	75	100	0.80	3.0	18	78		Sootblowing furnace; start time 20:03	90.30
8/12/2019	5	Operator B	22:25	23:10	0:10	3A	26.0	75	100	0.88	2.3	21	78		Finalyzing sootblowing SH	90.30
8/13/2019	6	Operator C	9:28	10:44	11:15	3A	24.0	75	110	0.88	3.0	21	78	proceed sampling	Sootblowing SH final; start time 8:42 end time 10:20	91.00
8/13/2019	7	Operator C	11:26	12:34	13:04	3A	26.0	65	100	0.88	3.0	21	85		Sootblowing stopped	90.30
8/13/2019	8	Operator C	13:58	15:22	15:38	3A	28.0	75	100	0.88	2.3	23	85		Sootblowing final; start time 13:26 end time 15:00	90.40
8/13/2019	9	Operator C	15:51	17:08	17:14	3A	26.0	85	100	0.88	3.0	21	85		Sootblowing stopped	90.10
8/13/2019	10	Operator C	17:31	18:31	18:48	3A	24.0	75	100	0.88	2.3	18	85		Sootblowing primary SH start time 16:33	90.00
8/13/2019	11	Operator A	20:22	21:50	22:10	3A	24.0	85	100	0.95	2.3	21	78		Sootblowing stopped	90.40
8/13/2019	12	Operator A	22:10				28.0	75	110	0.88	3.0	21	78			

Only eleven experiments out of the 62 planned ones were performed, without sampling. It was noticed a 1.2% of variation in the steam generator efficiency - S1. Although this variation may seem small, an increase in efficiency of 1.2% can represent a reduction of 10707 tons on fuel consumption (coal) per year [2].

These results are not conclusive and should be followed by a complete set of experiments, repetitions, inclusion of mills B, C and D, evaluation of external factors, among other measures. It is worth highlighting that performing experiments on a plant the size of PECEM is a hard task to accomplish, despite all the strictness and attention given by the operational team.

Some relevant observations are reported as follows:

Experiment 1

Conducted during 30 minutes after a two hours sootblowing.

Experiment 2

Primary air crossover duct (P7) changing was too slow, be careful because of the risk in the seal air fans and other equipment.

Experiment 3

The experiments took more time than expected due to system inertial response time. New changes were only performed after system stabilization and respected the sootblowing routine.

Experiment 6

The most powerful equipment is the feed pump and for this reason, a change in this equipment can cause oscillations in many others. A failure happened in the execution of this experiment due to this problem and not because of the experiment itself.

Experiment 12 - Failure

A failure on the mill inverter during the execution of the experiment, caused by a relay overload that led the mills to stop. Generation output dropped from 360 to 190 MW. A failure of this magnitude is the worst consequence that can occur during the execution of an experiment. The primary air flow (P1) and the speed of the dynamic classifier (P3) at its upper limits caused an increase in the current of the mill inverter, which exceeded the equipment allowed limit.This experiment brought extremely relevant information to a new limitation that was not within the system alarms, and must be included. Historical

[2]The calculation considered 2018 as the base year. The power plant operated 40,6% of the time in baseload.

data showed that this current value had only been reached once before, which also caused failure and consequently stopped the mill. An investigative process was carried out, and it was found a 20 A deviation from the supervisory reading (72 A) and the actual value (92 A) acquired on the field, close to its limit value of 100 A. The inverter was replaced but still mills presented an extra current failure, causing the experiment to be aborted.

3.7 Simulation Model

The proposed experiments at Table 3.6 could not be concluded at the PECEM power plant and were substituted by a simulation routine assemble with the aid of the EBSILON® professional software, which allowed to fulfil step number six of Figure 3.3.

EBSILON (Energy Balance and Simulation of the Load response of power generating or process controlling network structures) solves non-linear mass and energy balances to simulate a variety of thermodynamic cycles. Its disposes of extensive component, fluid and fuel libraries [STEAG, 2020]. Figure 3.13 represents the EBSILON implementation of PECEM steam generator and mills.

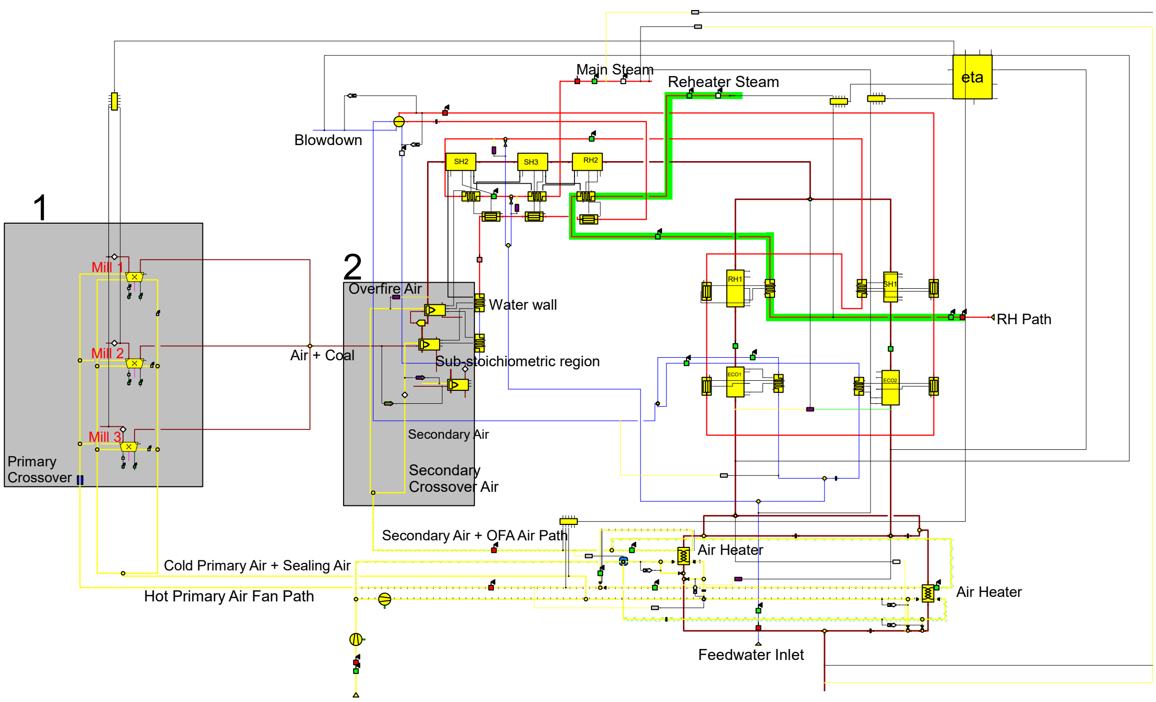


Figure 3.13 – Representation of the PECEM steam generator simulation in the EBSILON® simulation program.

The complete system is composed by the steam generator, three mills and auxiliary equipment as heat exchangers, pumps and tanks, connected by working fluid and fuel streams, modeled by 149 components. Two special subsystems are highlighted in the Figure 3.13 for the mills (1) and the steam generator furnace (2). Subsystem 1 input parameters were defined as hot and cold primary air, coal flow and sealing air, coal and air mixture outlet temperature.The outputs are the fuel air mixture and it is moisture.

Subsystem 2 segregates two volumes with different models to calculate the sub-stoichiometric combustion and a complementary one, with excess air, in the burnout zone. Inputs are the secondary air and over fire air (OFA) flow rates and outputs are the flue gases sent to the heat exchangers and the steam generation rate. EBSILON model is also capable of calculating the system efficiency based on the direct method by DIN 1942, 1994 (Section 3.2.1), for design and off-design conditions.

3.7.1 Model assessment

Simulated results from the EBSILON model were compared to the ones displayed in Table 3.7, and the obtained efficiencies are presented in Table 3.8. The steam generator efficiency of the PECEM power plant was recalculated according to the direct method by DIN 1942, 1994 (Section 3.2.1) in order to be directly compared to the simulation model results.

Table 3.8 – Relative deviation of real experiments at the PECEM power plant and the simulation model

Steam Generator Efficiency (S1)			
Experiment number	PECEM power plant	Simulation model	Relative deviation
1	84.19%	83.52%	0.80
2	84.37%	83.53%	1.00
3	84.02%	83.00%	1.21
4	82.89%	83.00%	-0.13
5	83.90%	83.52%	0.45
6	83.61%	82.81%	0.96
7	83.19%	82.80%	0.47
8	83.76%	83.52%	0.29
9	82.92%	82.79%	0.16
10	82.82%	83.49%	-0.82
11	83.71%	83.46%	0.30

The relative deviation was calculated by the ratio between the efficiency difference of the PECEM plant and the simulation model in relation to the PECEM plant efficiency.

Efficiency displayed a 1.55% variation for the actual case and no more then 0.53% for the simulation model, which is a controlled and conservative environment. Efficiency deviations also came from their calculation approach, and the maximum relative values for each given case reached 1.21. Besides the deviations, simulation model showed to be capable of represent the system trend. For this reason, data presented in Table 3.8 was standardized. The procedure to standardize the efficiencies is in accordance with Equation 3.10.

$$z = \frac{\eta_{SG} - \mu}{\sigma} \tag{3.10}$$

The results are exhibited in Figure 3.14.

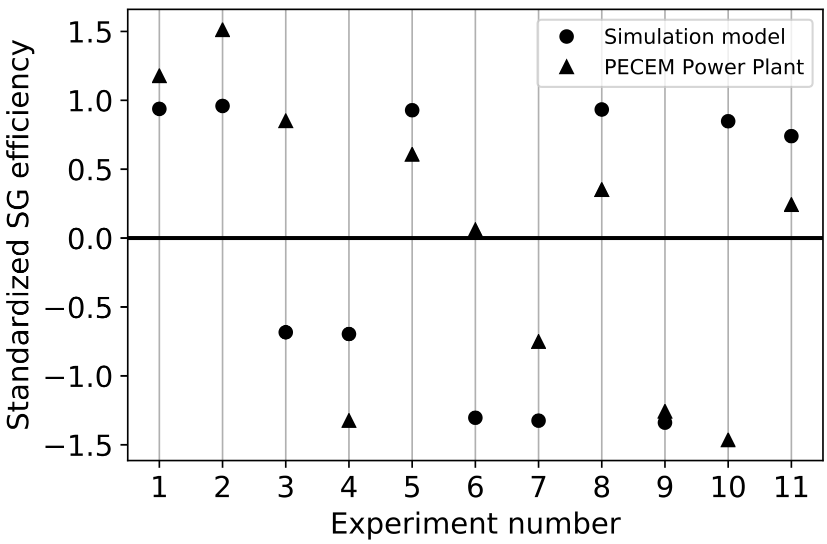


Figure 3.14 – Comparison between the standardized steam generator efficiencies of the PECEM power plant and the simulation model

It is possible to notice that the circular and triangular points in the graph show the same behavior trend, with the exception of experiments 3, 6 and 10. The execution of the experiments in the power plant in operation suffers a much greater impact from external variables than the experiments in the controlled environment in the simulation model, which may have caused this divergence.

3.7.2 DoE applied on the simulation model

The simulation model was able to take into account all controllable parameters defined in Figure 3.5 but the speed of the dynamic classifier (P3). The design matrix for the BBD method was downsized to 54 experiments keeping the same operational range (see Table 3.4 and 3.3). Results for steam generator efficiency (S1) are presented in Table 3.9.

Table 3.9 – Steam generator efficiency (S1) calculated with the simulation model according to a DoE planning

	Factors (controllable parameters)						**Response**
Experiment number	P1	P2	P4.B	P5	P6	P7	S1
1	28.0	75	0.80	2.3	21	85	83.42
2	28.0	85	0.88	3.0	21	78	82.65
3	28.0	65	0.88	3.0	21	78	80.97
4	28.0	65	0.88	1.5	21	78	82.38
5	26.0	75	0.88	2.3	21	78	83.33
6	26.0	85	0.88	2.3	23	85	83.42
7	26.0	75	0.80	1.5	21	85	84.15
8	26.0	85	0.80	2.3	23	78	83.56
9	26.0	75	0.95	1.5	21	85	83.94
10	26.0	65	0.95	2.3	18	78	81.71
11	28.0	75	0.88	3.0	18	78	82.52
12	24.0	75	0.80	2.3	21	70	83.55
13	26.0	65	0.88	2.3	23	85	81.77
14	28.0	75	0.95	2.3	21	85	83.18
15	24.0	75	0.95	2.3	21	70	83.34
16	26.0	85	0.80	2.3	18	78	83.55
17	26.0	75	0.95	3.0	21	85	82.55
18	26.0	75	0.95	1.5	21	70	83.94
19	28.0	75	0.95	2.3	21	70	83.18
20	26.0	75	0.88	2.3	21	78	83.33
21	24.0	75	0.88	1.5	18	78	84.06
22	24.0	75	0.88	3.0	18	78	82.71
23	24.0	85	0.88	1.5	21	78	84.06
24	26.0	75	0.80	3.0	21	85	82.78
25	24.0	75	0.88	3.0	23	78	82.71
26	26.0	85	0.95	2.3	18	78	83.35
27	24.0	85	0.88	3.0	21	78	82.71
28	26.0	75	0.88	2.3	21	78	83.33
29	26.0	65	0.80	2.3	18	78	81.91
30	26.0	75	0.88	2.3	21	78	83.33
31	26.0	85	0.88	2.3	18	70	83.41
32	26.0	85	0.88	2.3	18	85	83.41
33	24.0	65	0.88	1.5	21	78	83.98
34	28.0	75	0.88	1.5	23	78	83.95
35	24.0	65	0.88	3.0	21	78	82.62
36	24.0	75	0.80	2.3	21	85	83.55
37	26.0	65	0.88	2.3	18	70	81.77
38	26.0	85	0.88	2.3	23	70	83.42
39	26.0	65	0.88	2.3	23	70	81.77
40	26.0	75	0.80	3.0	21	70	82.78
41	26.0	65	0.95	2.3	23	78	81.71
42	26.0	65	0.88	2.3	18	85	81.77
43	28.0	75	0.88	1.5	18	78	83.95

Continued on next page

Table 3.9 – *Continued from previous page*

	Factors (controllable parameters)						Response
Experiment number	P1	P2	P4.B	P5	P6	P7	S1
44	26.0	75	0.88	2.3	21	78	83.33
45	26.0	65	0.80	2.3	23	78	81.92
46	28.0	75	0.88	3.0	23	78	82.52
47	24.0	75	0.88	1.5	23	78	84.07
48	28.0	75	0.80	2.3	21	70	83.42
49	28.0	85	0.88	1.5	21	78	84.06
50	26.0	75	0.80	1.5	21	70	84.15
51	26.0	85	0.95	2.3	23	78	83.35
52	26.0	75	0.88	2.3	21	78	83.33
53	24.0	75	0.95	2.3	21	85	83.34
54	26.0	75	0.95	3.0	21	70	82.55

Efficiency ranged 3.19% in absolute values, which is quite sensitive for that response factor. The BBD details are presented in Table 3.10.

Table 3.10 – Box-Behnken Design (BBD) details to perform DoE on the simulation model

Number of factors	6	Replication	1
Number of essays	54	Total number of essays	54
Number of blocks	1	Center points	6

Next steps on the process was to model the response surface RSM, corresponding to the second phase highlighted in blue in Figure 3.3.

3.7.3 Step 7 - Statistical analysis

The seventh step is the statistical analysis with Analysis of Variance (ANOVA).

As strange as it can look, outliers can occur in simulated processes as the result of the combination of extreme situations. Four occurrences were identified in experiments 2, 33, 35, and 49, all of them related to the combination of P1, P2, and P5. These results were removed from the set of experiments.

Steps 7.a and 7.b were performed based on ANOVA results, presented in Table 3.11. Significant factors and interactions were selected by searching terms with p-value$<\alpha$=0.05, which reject the null hypothesis and corresponds to a minimum confidence level of 95%.

The ANOVA presented in Table 3.11 referred to the complete model with all terms and the final model, as a result of several model reduction iterations. Non-significant 2 way terms were removed one by one, followed by the square and linear ones. ANOVA was recalculated after each term removal as a result of an iterative calculation loop to update values. It is possible to notice that in the final model the not statistically significant terms were removed. The complete table for ANOVA is available in Appendix B (Section B.2).

Table 3.11 – Analysis of variance (ANOVA) for the complete and the final model with all linear, square and interactions terms

	Complete model	**Final model**
Source	**P-value**	**P-value**
Model	0.000	0.000
Linear	0.000	0.000
P1	0.000	0.000
P2	0.000	0.000
P4	0.000	0.000
P5	0.000	0.000
P6	0.466	0.380
P7	0.955	0.946
Square	0.000	0.000
P1*P1	0.072	0.030
P2*P2	0.000	0.000
P4*P4	0.000	0.000
P5*P5	0.000	0.000
P6*P6	0.001	0.000
P7*P7	0.082	0.036
2 - Way interaction	0.154	0.000
P1*P4	0.082	0.036
P1*P5	0.001	0.000
P1*P6	0.979	-
P1*P7	0.924	-
P2*P4	0.897	-
P2*P5	0.603	-
P2*P6	0.995	-
P2*P7	0.969	-
P4*P5	0.320	-
P4*P6	0.981	-
P4*P7	0.999	-
P5*P6	0.988	-
P5*P7	0.964	-
P6*P7	1.000	-

Linear and square terms with statistic significance were kept in the final model.

The importance of each factor and their effects were not a priori known, although inferred due to the process knowledge. The interactions P1 with P4 and P1 with P5 were the only kept in the model after performing the interactions.

3.7.4 Step 8 - Fitting the second-order model

The eighth step is dedicated to the fitting of the steam generator efficiency (S1) second-order model. The first equation contains all the terms and is presented in Equation 3.11.

$$\begin{aligned}
S1 = 42.99 + &\ 0.1658P1 + \ 1.2255\ P2 - \ 14.28\ P4 - \ 0.260\ P5 - \ 0.1280\ P6 \\
- &\ 0.0222\ P7 - \ 0.00219P1^2 - \ 0.007614\ P2^2 + \ 8.394\ P4^2 - \ 0.04651\ P5^2 \\
+ &\ 0.003144\ P6^2 + \ 0.000147\ P7^2 - \ 0.0620\ P1P4 - \ 0.01275\ P1P5 \\
- &\ 0.00003\ P1P6 - \ 0.000033\ P1P7 - \ 0.00089\ P2P4 - \ 0.00062\ P2P5 \\
+ &\ 0.000001\ P2P6 + \ 0.000003\ P2P7 - \ 0.0923\ P4P5 + \ 0.0007\ P4P6 \\
+ &\ 0.00001\ P4P7 + \ 0.00004\ P5P6 + \ 0.000041\ P5P7 + \ 0.000000\ P6P7
\end{aligned} \tag{3.11}$$

Next, equation 3.12 presents the final model containing only the terms statistically significant.

$$\begin{aligned}
S1 = &\ 43.44 + 0.1604P1 + 1.22355P2 - 14.54P4 - 0.4101P5 \\
&- 0.128P6 - 0.0228P7 - 0.002193P1^2 - 0.007614P2^2 + 8.394P4^2 \\
&- 0.04651P5^2 + 0.003144P6^2 + 0.000147P7^2 - 0.062P1P4 - 0.01172P1P5
\end{aligned} \tag{3.12}$$

The final model displayed an adjusted R^2 of 99.98% and a predicted R^2 of 99.95% which can be considered suitable to calculate S1 [Salkind, 2017].

3.7.5 Step 9 - Residual analysis

The objective of the ninth step is to check the model assumptions of normality, constant variance, and independence through residual plots. Figure 3.15 presents results with the aid of a Normal Probability Plot (NPP) and a correspondent histogram, residual versus fitted values, and residual versus observation order.

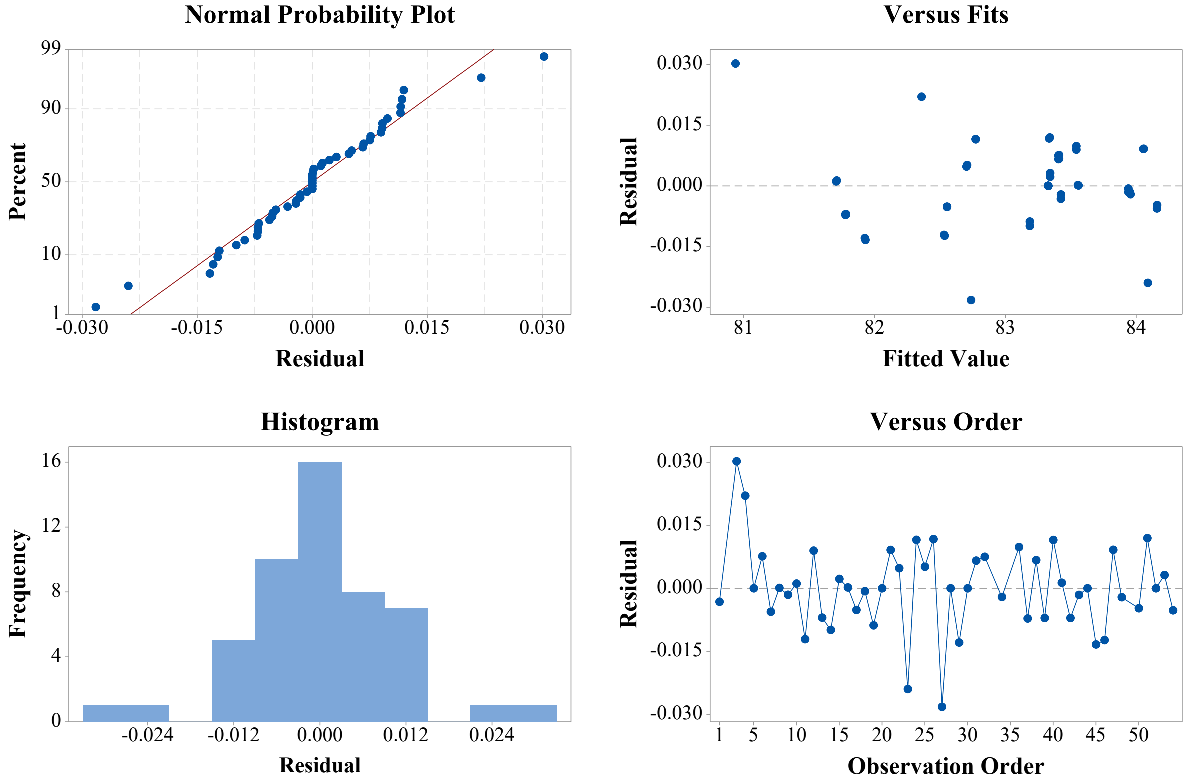


Figure 3.15 – Residual plots for the response steam generator efficiency (S1)

NPP shows that the data approximately follow a straight line and the histogram displays a symmetrical distribution, similar to a Gaussian distribution. The residual versus fitted values show a random distributed around the zero line with constant variance. The residual versus observation order plot shows no recognizable patterns or trends, and both NPP and histogram plots indicate that data come from a normal population. The present step concludes the second phase model fitting - RSM.

3.7.6 Step 10 - Factors ranking by order of importance

Ranking of the factors that influence variability in the response S1 is one of the major goals of the methodology. Table 3.12 presents these factors based on the evaluation of Equation 3.12 coefficients.

Table 3.12 – Regression coefficients of the second-order models in terms of coded and uncoded coefficients

Term	**Coded Coefficient**	**VIF**	**Uncoded Coefficient**
Constant	83.3262		43.4400
P1	-0.0684	1.04	0.1604
P2	0.8141	1.04	1.2235
P4	-0.1095	1.00	-14.5400
P5	-0.6930	1.00	-0.4101
P6	0.0022	1.00	-0.1280
P7	-0.0002	1.00	-0.0228
P1*P1	-0.0088	1.24	-0.0022
P2*P2	-0.7614	1.63	-0.0076
P4*P4	0.0472	1.23	8.3940
P5*P5	-0.0262	1.24	-0.0465
P6*P6	0.0197	1.52	0.0031
P7*P7	0.0083	1.23	0.0001
P1*P4	-0.0093	1.00	-0.0612
P1*P5	-0.0176	1.00	-0.0117

The coded coefficients are calculated as if all the factors were to be varied in the same range. Therefore, different orders of magnitude of the factors won't impact results and for this reason, they are used to rank the factors. The third column is the Variance Inflation Factor (VIF). A VIF of 5 or greater indicates multicollinearity, which is not the case in the present model. Last column brings the original or uncoded coefficients. In either cases, the coefficient signal indicates direct or indirect proportionality in respect to the equation response. Uncoded coefficients allows to express the equation response in a more meaningful way as they express the original operational scales and ranges. Similar results to the coded coefficients can be observed graphically through the Pareto chart presented in Figure 3.16.

The Pareto chart presents the factors ranked by order of importance, and by adding linear and quadratic terms of the coefficients presented in Table 3.12. Factors by order of importance in respect to the system efficiency (S1) were the pulverized coal outlet temperature (P2), excess O_2 (P5), stoichiometry (P4), primary air flow (P1), secondary air crossover duct pressure (P6), and primary air crossover duct pressure (P7).

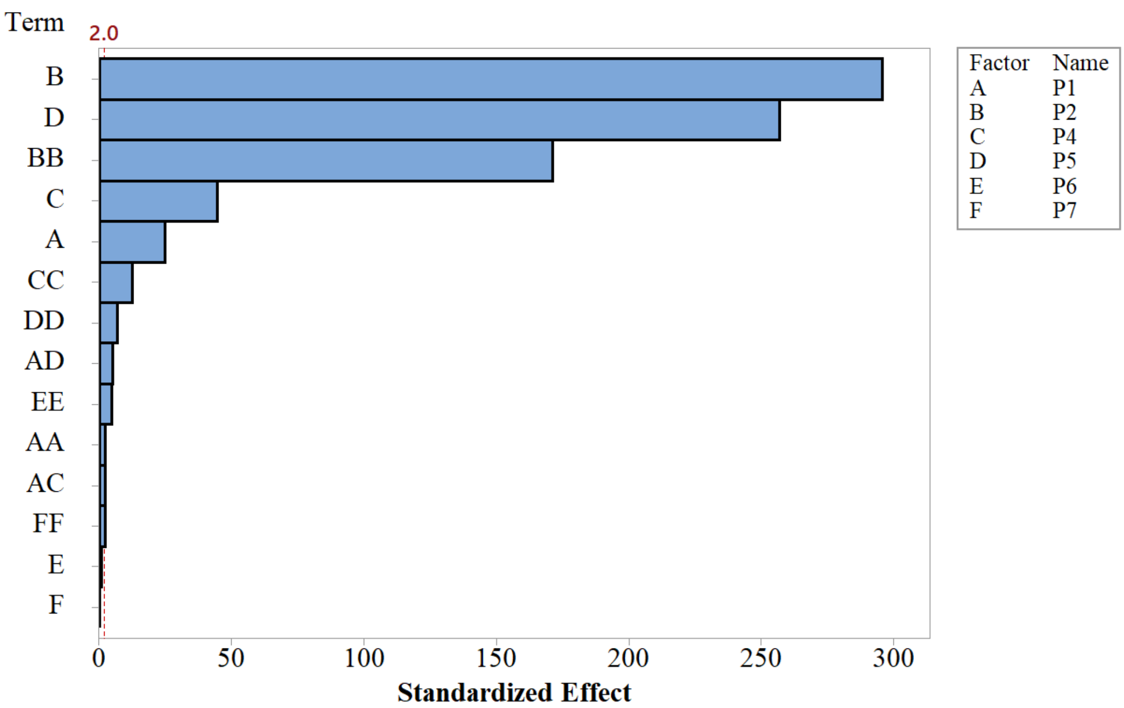


Figure 3.16 – Pareto chart of the standardized effects (response S1, α=0.05)

3.7.7 Step 11- Main effects and interaction plots

Single effect on the steam generator efficiency (S1) in respect to each factor are displayed in Figure 3.17. Both a main effects plot and a Pareto plot are used to identify the key process parameters or factors which have an impact on variability.

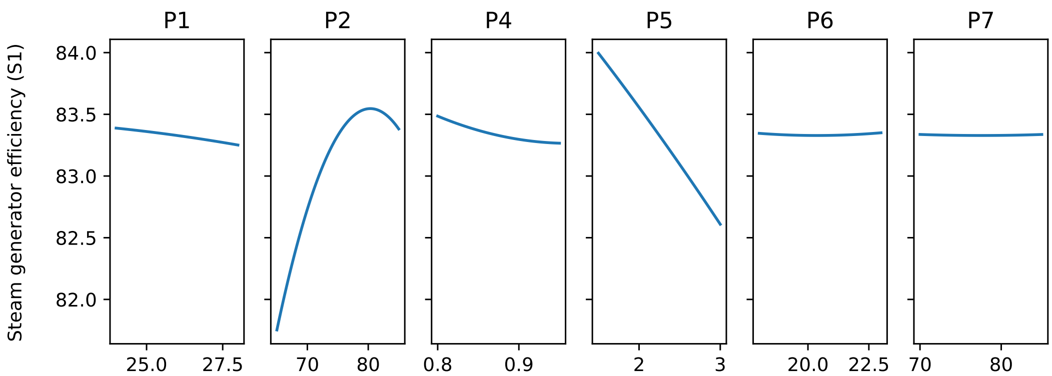


Figure 3.17 – Main effects plot for the response steam generator efficiency (S1)

The slope is proportional to the effect. Factors P2 and P5 displayed a significant impact on S1 variation, confirmed on the former Pareto chart (Figure 3.16) which ranked the factors P2 and P5 in the first and second position, respectively.

The steam generator efficiency (S1) increases as the primary air flow (P1), stoichiometry (P4) and excess O_2 (P5) decrease. In the burning process, the more air is presented the greater the energy is used to promote the combustion. It is worth remembering that the stoichiometry (P4) is related to the sub stoichiometry region while the Excess O_2 (P5) is related to the burnout zone (Figure 3.6). Regarding the pulverized coal outlet temperature (P2) the higher the temperature of the pulverized coal the better for the burning process. This temperature must be high enough to remove coal moisture, however, it cannot be so high as to cause the auto-ignition process. The condition of higher efficiency is around 80°C, corresponding to the nominal operating point of the mills.

The parameters with the least impact on the steam generator efficiency (S1) are the secondary and primary air's crossover duct pressure (P6 and P7). Changing the pressure in the crossover duct changes the air enthalpy and the amount of energy. Despite being statistically significant parameters and remained in the model, the effect of P6 and P7 on the steam generator efficiency (S1) is much less than those of the other factors.

The interaction plots presented in Figure 3.18 allow to identify if the effect of one given factor depends on some other one by searching for line crossings.

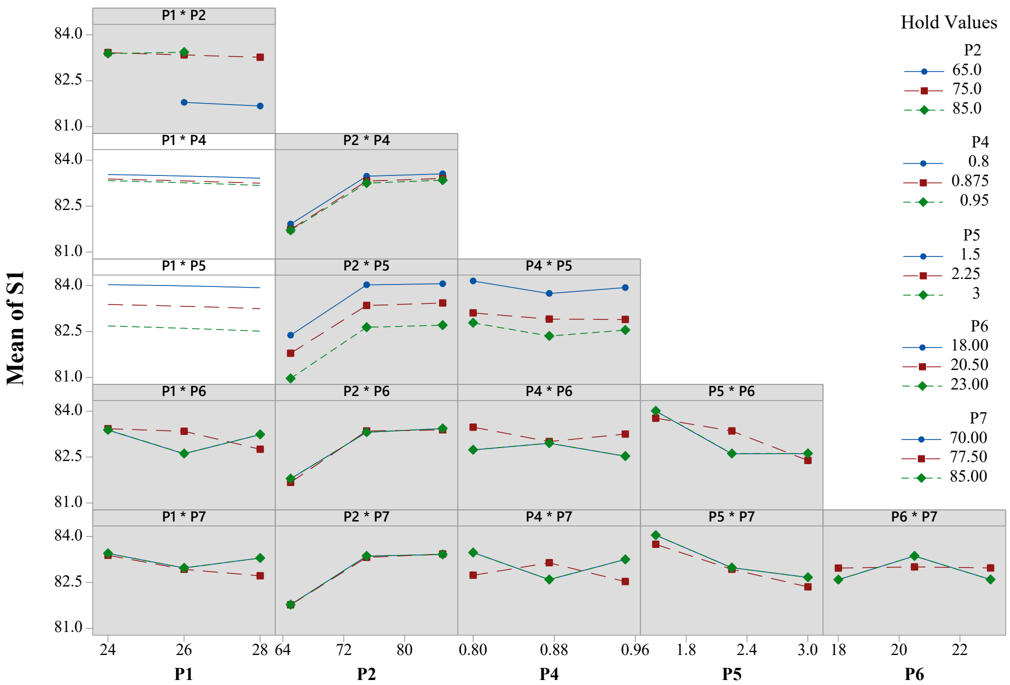


Figure 3.18 – Interaction plot for the response steam generator efficiency (S1)

P1 with P4 and P1 with P5 were the only pair of factors that displayed significant interactions according to the hypothesis test. This interaction is in conformity with the physical process because the three factors are related to the total air flow in the boiler.

3.7.8 Step 12 - Surface and contour plots

The contour plots 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. Factors P2 and P5 showed to be the more relevant in respect to the system efficiency S1 and its contour plot is presented in Figure 3.19.

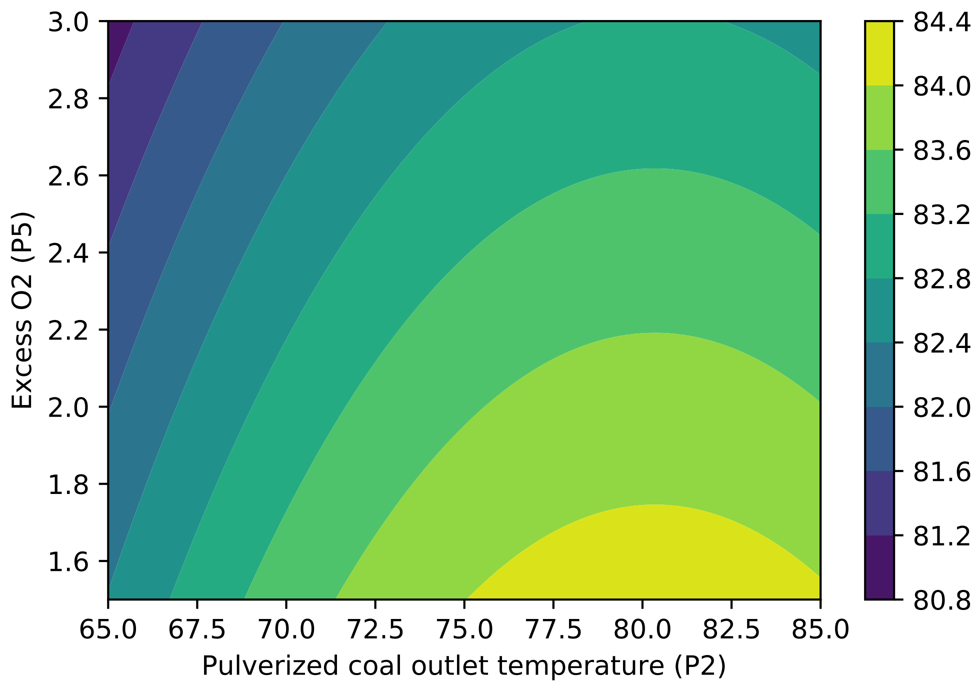


Figure 3.19 – Contour Plot of P2 x P5

The highest steam generator efficiency (S1) ranges are related to high pulverized coal outlet temperatures (P2), around 80°C, and low values for the excess O_2, up to around 1.6%. The variation in S1 for changing the levels of these factors is 3.6%. The higher the pulverized coal outlet temperature the better the process, because the energy required to burn coal will be less. On the other hand, the higher the Excess O_2 more energy will be required to burn the air, decreasing the efficiency of the process. Whole set of contour plots are presented in Figure 3.20.

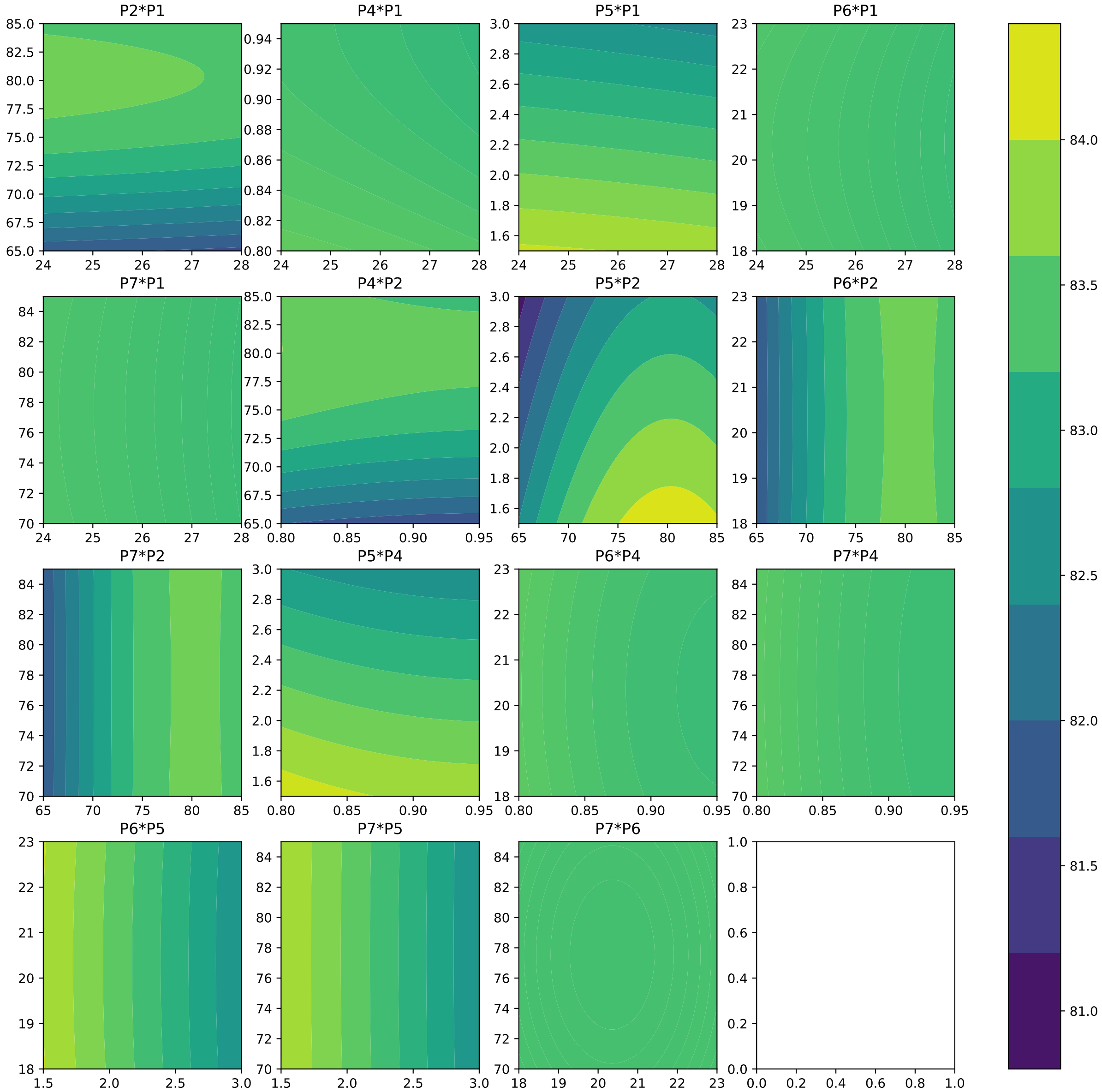


Figure 3.20 – Contour plots of the pairs of combined factors

The pairs of combined factors with the greatest range of variation in steam generator efficiency (S1) were P2 and P5, P2 and P4, and P5 and P4, in accordance with the ranking of the factors (Figure 3.16).

3.7.9 Step 13 - Surrogate model definition

The surrogate modeling technique used in the present work was the Polynomial Response Surface (PRS) based on DoE and RSM. The second-order model was chosen to predict the steam generator efficiency S1, presented in Equation 3.12.

The surrogate model substitutes more sophisticated models and its validity is constrained to the range of the selected factors presented in Table 3.3.

3.7.10 Step 14 - Test and validation

The generated surrogate model was tested for different situations and results were compared to the ones forming the BBD 54 experiments. In that context, the highest relative deviation was 0.0183 (Appendix B - Table B.2).

The generated response surface produces 20 predictions for new and untested operational conditions, willing to assess the model accuracy for unknown states, and results are shown in Table 3.13.

Table 3.13 – Relative deviation of the simulation model and the surrogate model for 20 new operating conditions

Test	P1	P2	P4	P5	P6	P7	Simulation Model	Surrogate Model	Relative Deviation
1	26.0	75	0.80	2.3	21	85	83.48	83.49	-0.0002
2	26.5	76	0.82	2.4	21	83	83.33	83.40	-0.0008
3	27.0	77	0.84	2.5	22	81	83.19	83.30	-0.0013
4	27.5	78	0.86	2.6	22	79	83.05	83.20	-0.0017
5	24.5	67	0.90	1.7	19	75	83.77	82.70	0.0128
6	25.0	68	0.92	1.8	20	77	83.66	82.77	0.0105
7	25.5	69	0.94	1.9	20	79	83.54	82.84	0.0083
8	26.0	70	0.95	2.0	21	81	83.42	82.90	0.0063
9	26.5	71	0.90	2.1	21	83	83.36	82.98	0.0046
10	24.5	65	0.80	1.6	18	70	83.99	82.54	0.0172
11	25.0	66	0.82	1.7	19	72	83.85	82.62	0.0147
12	25.5	67	0.84	1.8	19	74	83.71	82.68	0.0124
13	26.0	68	0.86	1.9	20	76	83.58	82.73	0.0102
14	26.5	69	0.88	2.0	20	78	83.45	82.77	0.0081
15	27.0	70	0.90	2.1	21	80	83.33	82.80	0.0063
16	24.5	66	0.81	1.6	18	72	84.08	82.78	0.0154
17	25.0	69	0.82	1.6	19	74	83.96	83.31	0.0078
18	25.5	72	0.83	1.6	19	76	83.95	83.70	0.0029
19	26.0	75	0.84	1.6	20	78	83.93	83.96	-0.0004
20	26.5	78	0.85	1.6	20	80	83.92	84.09	-0.0020

Factors were randomly varied to represent twenty new operational conditions. The maximum relative deviation was found to be 0.0172. Results show a slightly larger relative deviation for these new operating conditions.

3.7.11 Step 15 - Recommendation of a sequence of maneuvers

For the suggestion of a sequence of maneuvers there are no further constraints except the factors limits. The proposal of this section is to define the factors values of operation to improve steam generator performance. The desired operational conditions to operate the steam generator efficiency (S1) are presented in Table 3.14.

Table 3.14 – Optimum operational condition to maximize the response S1 - steam generator efficiency

S1 = 84.43%					
P1 (kg/s)	P2 (°C)	P4 (dimensionless)	P5 (%)	P6 (mbar)	P7 (mbar)
24	80	0.8	1.5	23	70

The steam generator efficiency varies from 80.80 to 84.43%. If the best operating conditions are defined as those with steam generator efficiency above 84% a set of input conditions can be chosen. Table 3.15 presents the possible operating conditions to guarantee steam generator efficiencies above 84% which is only possible for P2 avoce 75°C and P5 below 2.0%.

Table 3.15 – Operation maneuvers to assure best-operating conditions

	P2=85°C and P5=1.5%			**P2=75°C and P5=1.5%**	
	Lower limit	Upper limit		Lower limit	Upper limit
P1	24	26	P1	24	25
P4	0.8	0.95	P4	0.8	0.9
P6	18	23	P6	18	23
P7	70	85	P7	70	85
	P2=80°C and P5=1.5%			**P2=80°C and P5=2%**	
	Lower limit	Upper limit		Lower limit	Upper limit
P1	24	28	P1	24	
P4	0.8	0.95	P4	0.8	
P6	18	23	P6	23	
P7	70	85	P7	70	85

The most important factors according to the rank presented in Section 3.7.6 were the pulverized coal outlet temperature (P2) and the excess O_2 (P5). If the levels of these factors are kept constant at their optimum, the other factors may vary throughout their operating range but the steam generator efficiency will always remain above 84%. If P5 is set in 2% the factors P1, P4 and P6 must be kept at their optimum conditions to assure efficiencies above 84%. On the other hand, if the temperature drops to 70°C, regardless of the operating range of the other factors the steam generator efficiency will not reach values greater than 83.62%. This could indicate high moisture content which proves the difficulty of maintaining stable and high steam generator performance on rainy days.

3.7.12 Summary and Results Discussions

The steps to build a surrogate model using RSM and DoE were applied to the case study of the PECEM power plant. The challenges encountered during the execution of the study can serve as a basis for new applications or as a reference model to the application of the methodology in other power plants, considering the specificity of each case.

The study considered primary air flow, pulverized coal outlet temperature, speed of the dynamic classifier, stoichiometry, excess O_2, secondary air crossover duct pressure, and primary air crossover duct pressure as the input of the model to build a second-order polynomial. The speed of the dynamic classifier was removed from the model due to the impossibility of consideration of the simulation model. The analysis commenced with data collection in agreement with the DoE design chosen, model-fitting in a response surface design using the results in which only the significant factors (controllable parameters) and interactions remain in the model. The whole set of factors was significant which is in agreement with process knowledge but only the interactions between primary air flow with stoichiometry and primary air flow with excess O_2 were statistically significant.

The obtained algebraic expression (Equation 3.12) is capable of representing the steam generator behavior and suit as a surrogate model of the original system. The equation displayed an adjusted R^2 of 99.98% and a predicted R^2 of 99.95%. The model validation varied the factors randomly to represent twenty new operational conditions besides the reproduction of the 54 initial experiments. The maximum relative deviation was found to be 0.0172.

The factors were ranked based on their order of importance, where the first one has the higher impact. The most important factors were the pulverized coal outlet temperature (P2) and the excess O_2 (P5), followed by the stoichiometry (P4), primary air flow (P1), secondary air crossover duct pressure (P6), and primary air crossover duct pressure (P7). The effects of the factors and their interactions were analyzed according to their impact on the steam generator efficiency variability using main effects plots. Pulverized coal outlet temperature, primary and secondary air crossover duct showed non-linear behaviors. The best-operating factor ranges were defined using contour plots.

The standardized operation of the steam generator starts with the operator respecting the controllable parameters rank and initializing the alterations for a new condition always for the controllable parameters with a high effect on the steam generator efficiency. Their attention during operation must be kept on the most influential parameters, consequently, the primary and secondary crossover duct pressures are the last to concern with. Finally, controllable parameters must attain the best operating ranges according to Table 3.15.

3.8 Conclusions

The current work aimed to improve steam generator performance through process standardization. The proposed methodology applied RSM based on the DoE approach to build a surrogate model capable of capture the behavior of a coal-fired power plant system focused on the steam generator and its mills across a defined design space.

The DoE was applied in the case study of the PECEM power plant, however, it could not be finalized. For this reason, a simulation model based on mass and energy balances were developed to perform the proposed experiments. The steam generator efficiency calculated from the simulation model showed a maximum relative deviation of 1.21 when compared with the steam generator efficiency of the PECEM power plant. The initial input parameters, factors on DoE methodology, were the primary air flow (P1), pulverized coal outlet temperature (P2), speed of the dynamic classifier (P3), stoichiometry (P4), excess O_2 (P5), secondary air crossover duct pressure (P6), and primary air crossover duct pressure (P7). The speed of the dynamic classifier was removed due to a limitation on the simulation model. All the remained factors were analyzed and classified as statistically significant on the response steam generator efficiency (S1), as well as the interactions between the primary air flow (P1) and stoichiometry (P4) and between the primary air flow (P1) and excess O_2. The most important factors were the pulverized coal outlet temperature (P2) and the excess O_2 (P5), followed by the stoichiometry (P4), primary air flow (P1), secondary air crossover duct pressure (P6), and primary air crossover duct pressure (P7).

The experiments performed through the DoE were used to to build a second-order polynomial model according to the RSM. The result is an algebraic expression (Equation 3.12) capable of representing the steam generator behavior that suit as a surrogate model of the original system. The equation displayed an adjusted R^2 of 99.98% and a predicted R^2 of 99.95%. The model validation varied the factors randomly to represent twenty new operational conditions besides the reproduction of the 54 initial experiments. The maximum relative deviation was found to be 0.0172. The optimum operational condition to maximize the steam generator efficiency (S1 = 84.43%) corresponds to maximum value of P6 = 23 mbar, minimum values of P1 = 24 kg/s, P4 = 0.8, P5 = 1.5 %, and P7 = 70 mbar, and an intermediate value for P2 = 80°C. The standardized operation is a guidance for the operator to follows the controllable parameters rank, focus on the main controllable parameters and interactions, and attain the best operating ranges according to the response.

The use of surrogate models helps in drastically reducing the modeling time or experimentation hard to perform. The significant variables become decision variables to the operator. The surrogate model defined using RSM and DoE set the best-operating conditions and propose operation maneuvers to improve performance. During operation, this standard order of the factors must be followed by the operator when intervention occurs. The results add objectivity to the decision-making process during operation, reduces variability and improves quality assuring a standardized operation.

4 DESIGN OF EXPERIMENTS COMBINED WITH ARTIFICIAL NEURAL NETWORKS APPLIED ON THE CONTROL PARAMETERS OF A REAL STEAM GENERATOR [1]

4.1 Introduction

Coal-fired power plants account for 4.30% of the total electrical energy of Brazil's power grid and around 40% of the world. The Brazilian 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, 2017; 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].

[1] Article published in the proceedings of the 25th International Congress of Mechanical Engineering

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 to approximate 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].

4.2 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 layers 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.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |X_{exp} - X_{obs}| \tag{4.1}$$

$$MPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{X_{exp} - X_{obs}}{X_{exp}}\right| \tag{4.2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} |X_{exp} - X_{obs}|^2 \tag{4.3}$$

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

4.3 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 in 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 Equation (4.4) for Box-Behnken design and by Equation (4.5) for three full level factorial design [Ferreira et al., 2007; Montgomery, 2013]:

$$N = 2k(k-1) + Co \tag{4.4}$$

$$N = 3^k \tag{4.5}$$

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.

4.4 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 Figure4.1.

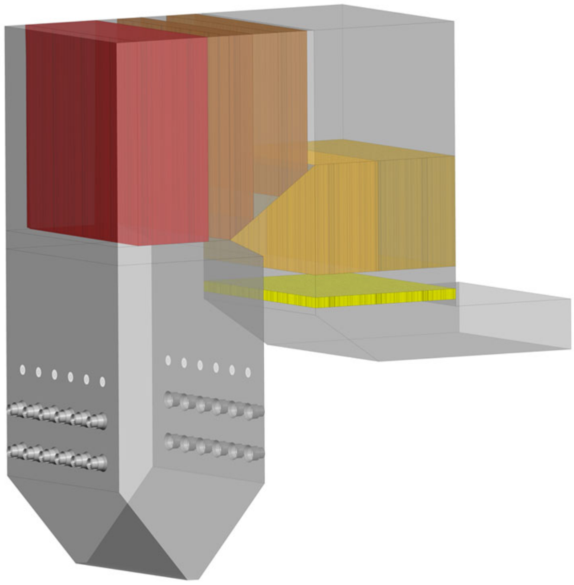


Figure 4.1 – Steam generator schematic layout

4.5 Methodology

The methodology proposed in this chapter 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 Figure 4.2.

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 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 hid-

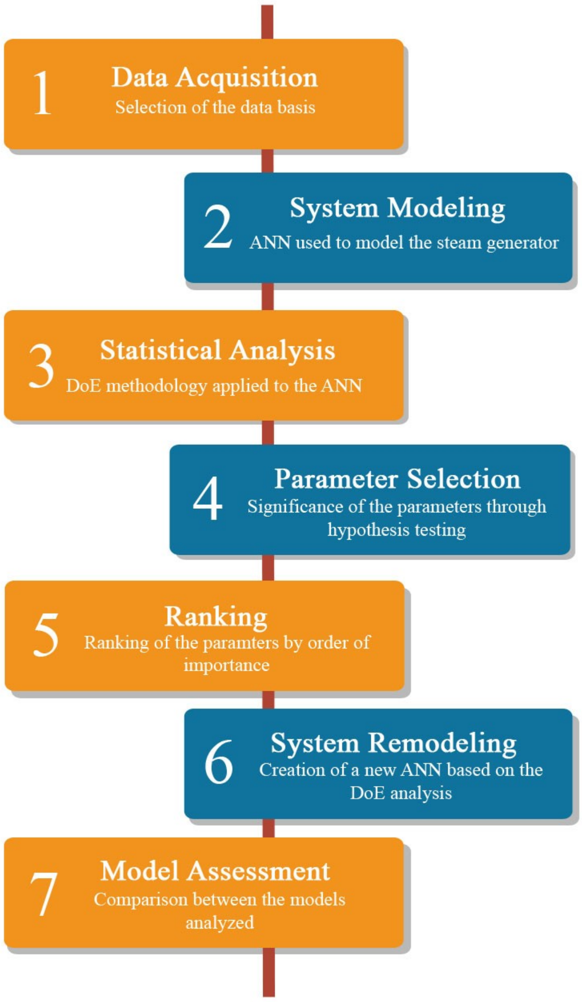


Figure 4.2 – Flowchart of the proposed method

den layer and activation functions) were defined through an iterative approach in search of best describing the problem at 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 Table 4.1. To facilitate further analysis visualization, the primary air flow, pulverized coal outlet temperature, speed of the dynamic classifier, excess O_2, primary air crossover duct pressure, secondary air crossover duct pressure and coal flow rate will be called, respectively, F1, F2, F3, F4, F5, F6 and F7.

Table 4.1 – Model input parameters with their ranges selected for the Design of Experiments 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 4.2 – Model response parameters

	Flue gas outlet temperature **(R1)**	Steam generator efficiency **(R2)**	Electric power generation **(R3)**
Unit	°C	%	MW

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 Table 4.2. These responses will be called R1, R2 and R3.

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 predictors are missing something. The simplest model that produces random residual is 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. Thereafter, 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.

4.6 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 Table 4.1 and Table 4.2 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.

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 Figure 4.3. The standardized MAE and MSE of the test were respectively 0.2015 and 0.2741.

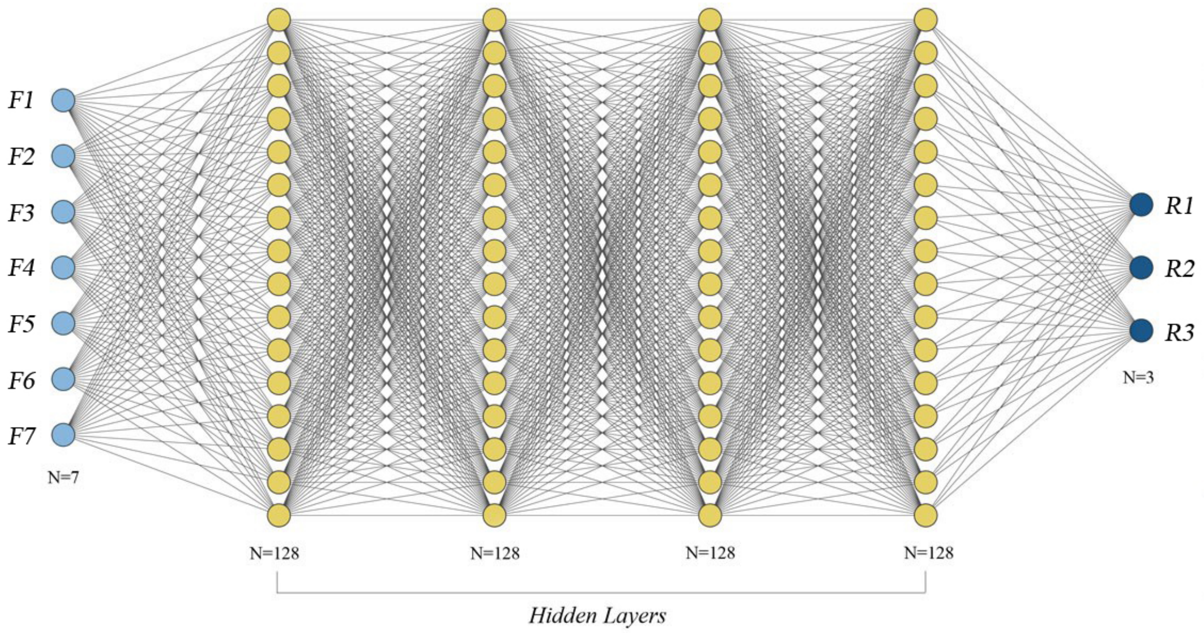


Figure 4.3 – Chosen topology for ANN

A DoE methodology was applied to analyze the control parameters of the steam generator modeled with the ANN, respecting the ranges listed in Table 4.1. 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 Table 4.3. The analyses were performed using the software Minitab®.

Table 4.3 – 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 fast response enables this comparative analyses. The system responses are

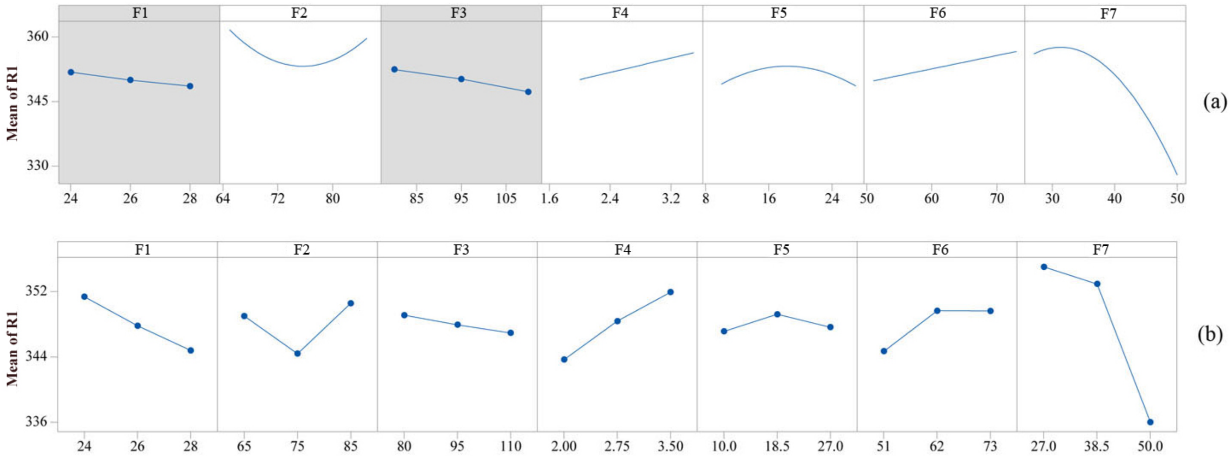


Figure 4.4 – Main effects of the controlled parameters on the response R1 (a) Box-Behnken (b) three level full factorial

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 R1 are shown in Figure 4.4 for Box-Behnken and three level full factorial.

The difference between the two DoE results are 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 R1 varies significantly with all the parameters according to three level full factorial, while Box-behnken design consider F1 and F3 as not 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 F4 and F6 leads to an increase on the response R1, showing a linear relationship. In contrast, the parameters F2, F5 and F7 have a non-linear relationship with the response.

The results found for R2 can be seen in Figure 4.5. (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 R2. There is a positive correlation between the parameters F2 and F4 with R2. On the other side, a negative correlation its noticed

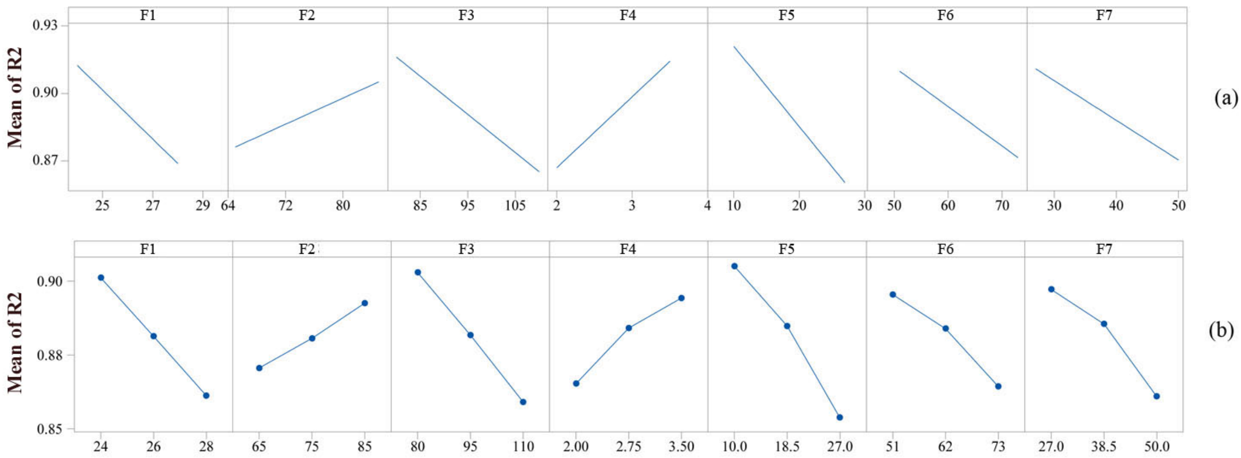


Figure 4.5 – Main effects of the controlled parameters on the response R2 (a) Box-Behnken (b) three level full factorial

among the other parameters and R2. Therefore, any increase on F1, F3, F5, F6 and F7 corresponds to a decrease on R2.

For the last response R3, the analysis results can be seen in Figure 4.6. (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 R3. Only the parameters F2 and F7 have positive linear relationships while F1 has a negative linear relationship. The parameter F5 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 work. However, it is valid to stress that all the interactions were considered in both analysis.

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

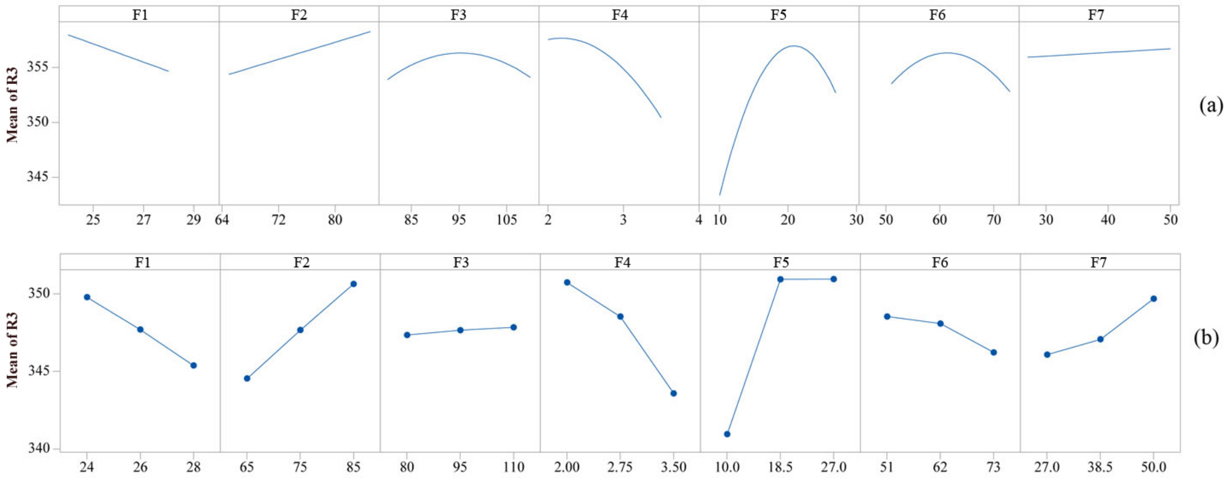


Figure 4.6 – Main effects of the controlled parameters on the response R3 (a) Box-Behnken (b) three level full factorial

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 prediction quality of the model. Table 4.4 presents the results.

Table 4.4 – Summary of the squared correlation coefficients

	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 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 Table 4.4. 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 R1 was not affected by the parameters F1 and F3, even though the responses R2 and R3 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 Figure 4.7.

	Flue gas outlet temperature (R1)	**Steam generator efficiency (R2)**	**Electric Power Generation (R3)**
1	Coal flow (F7)	Primary air crossover duct pressure (F5)	Primary air crossover duct pressure (F5)
2	Secondary air crossover duct pressure (F6)	Velocity of the dynamic classifier (F3)	Excess O2 (F4)
3	Excess O2 (F4)	Excess O2 (F4)	Pulverized coal outlet temperature (F2)
4	Pulverized coal outlet temperature (F2)	Primary air flow (F1)	Primary air flow (F1)
5	Primary air crossover duct pressure (F5)	Coal flow (F7)	Coal flow (F7)
6	*N.A.**	Secondary air crossover duct pressure (F6)	Secondary air crossover duct pressure (F6)
7	*N.A.**	Pulverized coal outlet temperature (F2)	Velocity of the dynamic classifier (F3)

Figure 4.7 – 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 (F7) presented itself as the most influential parameter for the flue gas outlet temperature (F1) response. As already discussed before, the primary air flow (R1) and speed of the dynamic classifier (F3) were not statistically significant and therefore were not presented in the ranking. For both the steam generator efficiency (R2) and electric power generation (R3), the primary air crossover duct pressure (F5) 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 R1. The same development method used for the first ANN was applied for the second one and the same hyperparameters topology was tested. The second ANN for R1 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 Table4.5.

Table 4.5 – ANNs comparison

	Original ANN		**Remodeled ANN**	
	MPE	MSE	MPE	MSE
Flue Gas Outlet Tempetature **(R1)**	0.320%	2.350	0.492%	5.548
Boiler Efficiency **(R2)**	0.452%	0.267	*N.A.**	*N.A.**
Electric Power Generation **(R3)**	0.208%	1.115	*N.A.**	*N.A.**

* Not applicable.

The MPE and MSE for R1 of the second ANN developed is larger when compared to the first ANN presented in this chapter, but it is 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 R1 and a maximum value for R2 and R3.

4.7 Conclusion

The main focus of this chapter 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, speed of the dynamic classifier, excess O_2, primary air crossover duct pressure, secondary air 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 R1, the parameters F1 and F3 were found not to be relevant while the responses R2 and R3 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 temperature (R1) is the coal flow rate. For steam generator efficiency (R2) and electric power generation (R3) the primary air crossover duct pressure (F5) 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.

5 CONCLUSIONS

The present work proposes alternatives for standardization of operation through surrogate models to represent the assembly of the steam generator and its mills, based on two approaches: the system simulation by a commercial software and alternatively by an Artificial Neural Network trained with actual plant data.

A systematic literature review was carried out to answer the following research question: how surrogate modeling techniques supported by DoE and RSM can help enhancing coal-fired power plant efficiency? The systematic literature review was able to give the reader a broad vision of the area of interest and pointed out the gap that justifies the research question.

A methodology for the construction of a surrogate model to a coal-fired power plant in operation is proposed based on DoE and RSM. The steps to conduct the methodology were described and applied to the case study of the PECEM power plant. Statistical tools like Design of Experiments (DoE) and Response Surface Methodology (RSM) were used to identify the model main controllable parameters and interactions to then rank them by order of importance. The experiments were conducted in the PECEM power plant but they could not be finalized. For this a reason, a simulation model with a commercial software was built to simulate the system efficiency. The surrogate model was built with six controllable input parameters: primary air flow, pulverized coal outlet temperature, stoichiometry, excess O_2, secondary and primary air crossover duct pressure. The maximum relative deviation of that surrogate model compared to the software simulation is 0.0172.

The surrogate model based on Artificial Neural Networks (ANN) can also simulate the system efficiency together with its flue gas outlet temperature and plant electric power generation with the addition of coal flow as a controllable input parameter. An ANN model with seven inputs presents MAE and MSE of 0.2015 and 0.2741 for the training data set and MPE and MSE of 0.32% and 2.350 for the validation data set.

The standardized operation starts with the operator respecting the controllable parameters rank and initializing the alterations for a new condition always for the controllable parameters with a high effect on the steam generator efficiency. Their attention during operation must be kept on the most influential parameters. Finally, controllable parameters must attain the best operating ranges propose. The recommended operational ranges and order of operation by significance allows a precision action in order to achieve a new, safe, stable, and more efficient condition.

The positive results using surrogate models supported by DoE, RSM, and ANN showed the application potential of this methodology. The challenges encountered during the execution of this research applied on the PECEM power plant can serve as a basis for new applications. The use of surrogate models helps in drastically reducing the modeling time or experimentation, assisting engineering decisions through a cheap-to-run surrogate model, and makes it easier to identify interesting regions to be explored and analyzed. The results add objectivity to the decision-making process during operation, reduces variability, and improves quality assuring a standardized operation.

5.1 Future work

New researches could be conducted to:

- Explore the significance of new parameters, including other systems and responses;
- Finalize the experiments conducted through the Design of Experiments methodology in the PECEM power plant;
- Use the Design of Experiments results as input to develop an ANN of the system;
- Explore different surrogate modeling techniques;
- Compare the results using Artificial Neural Networks (ANN) and the simulation model based on mass and energy balances to improve the analysis;
- Compare the surrogate models based on RSM and ANN to analyze precision, complexity, cost, among other factors;
- Apply the surrogate models for other power plants.

REFERENCES

Amiri, M., Shahhosseini, S., and Ghaemi, A. Optimization of CO2 Capture Process from Simulated Flue Gas by Dry Regenerable Alkali Metal Carbonate Based Adsorbent Using Response Surface Methodology, **Energy and Fuels**, vol. 31(5), p. 5286–5296, 2017.

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

Antony, J. **Design of Experiments for Engineers and Scientists**. Elsevier, 2 edition, 2014.

ASTM International. **ASTM D2234 / D2234M-19 - Standard Practice for Collection of a Gross Sample of Coal**, 2019.

Bekat, T., Erdogan, M., Inal, F., and Genc, A. Prediction of the bottom ash formed in a coal-fired power plant using artificial neural networks, **Energy**, vol. 45(1), p. 882–887, 2012.

Chandane, V. S., Rathod, A. P., Wasewar, K. L., and Sonawane, S. S. Efficient cenosphere supported catalyst for the esterification of n-octanol with acetic acid, **Comptes Rendus Chimie**, vol. 20(8), p. 818–826, 2017.

Chandane, V. S., Rathod, A. P., Wasewar, K. L., and Sonawane, S. S. Synthesis of cenosphere supported heterogeneous catalyst and its performance in esterification reaction, **Chemical Engineering Communications**, vol. 205(2), p. 238–248, 2018.

Chandrasekharan, S., Panda, R. C., and Swaminathan, B. N. Predictive Modeling of Integrated Boiler Unit for a Coal Fired Thermal Power Plant, **International Journal of Emerging Electric Power Systems**, vol. 18(3), 2017a.

Chandrasekharan, S., Panda, R. C., and Swaminathan, B. N. Statistical modeling of an integrated boiler for coal fired thermal power plant, **Heliyon**, vol. 3(6), p. e00322, 2017b.

Chetan, T. and Bhavesh, P. K. Efficiency with different GCV of coal and efficiency improvement, **International Journal of Innovative Research in Science, Engineering and Technology**, vol. 2(5), p. 1518–1527, 2013.

Cremanns, D. R. K., Hecker, S., Dumstorff, P., Almstedt, H., and Musch, C. **Efficient multi-objective optimization of labyrinth seal leakage in steam turbines based on hybrid surrogate models**. In Proceedings of ASME Turbo Expo 2016: Turbomachinery Technical Conference and Exposition, p. 1–11, Seoul. Siemens Energy, 2016.

DIN 1942. **Acceptance Testing of Steam Generators**, 1994.

Doosan Babcock Energy. **Pecem Units 1 & 2 Boiler Operating & Maintenance Manual**. Doosan Babcock Energy, 2011.

Dresch, A., Lacerda, D. P., and Jr, o. A. V. A. **Design Science Research: Método de pesquisa para avanço da tecnologia e ciência**. Bookman, 2015.

EDP. **UTE PECÉM**, 2019.

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. **Box-Behnken design: An alternative for the optimization of analytical methods**, 2007.

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

GP Strategies Corporation. **Heat Rate Awareness**, 2013.

Hall, P., Phan, W., and Whitson, K. **The Evolution of the Technologies of Analytics**. Media, O'Reilly, 1 edition, 2015.

Haykin, S. **Neural Networks and Learning machines**, 2014.

IEA. **Statistics & Data**, 2017.

Jiang, P., Zhou, Q., and Shao, X. **Surrogate Model-Based Engineering Design and Optimization**. Springer, 1 edition, 2020.

Khan, K. S., Kunz, R., Kleijnen, J., and Antes, G. Five steps to conducting a systematic review, **Journal of the Royal Society of Medicine**, vol. 96(3), p. 118–121, 2003.

Kumar, M., Goswami, L., Singh, A. K., and Sikandar, M. Valorization of coal fired-fly ash for potential heavy metal removal from the single and multi-contaminated system, **Heliyon**, vol. 5(10), p. e02562, 2019.

Linnenluecke, M. K., Marrone, M., and Singh, A. K. Conducting systematic literature reviews and bibliometric analyses, **Australian Journal of Management**, vol. 5(October), 2019.

Lujan-Moreno, G. A., Howard, P. R., Rojas, O. G., and Montgomery, D. C. Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study, **Expert Systems with Applications**, vol. 109, p. 195–205, 2018.

Mahanta, S., Chandrasekaran, M., Samanta, S., and Arunachalam, R. Multi-response ANN modelling and analysis on sliding wear behavior of Al7075/B4C/fly ash hybrid nanocomposites, **Materials Research Express**, vol. 6(8), 2019.

Mathews, P. G. **Design of Experiments with Minitab**. American Society for Quality, Milwaukee, 2005.

McCulloch, W. S. and Pitts, W. A logical calculus of the ideas immanent in nervous activity, **The bulletin of mathematical biophysics**, vol. 5(4), p. 115–133, 1943.

MME. **Balanço Energético Nacional**, 2018.

Montgomery, D. C. **D esign and Analysis Ninth Edition**. John Wiley & Sons, Inc., 8 edition, 2013.

Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. **Response Surface Methodology: process and product optimization using designed experiments**. John Wiley & Sons, Inc., Hoboken, 2016.

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

Pirhadi, N., Tang, X., Yang, Q., and Kang, F. **A new equation to evaluate liquefaction triggering using the response surface method and parametric sensitivity analysis**, 2018.

Remenárová, L., Pipíška, M., Florková, E., Horník, M., Rozložník, M., and Augustín, J. Zeolites from coal fly ash as efficient sorbents for cadmium ions, **Clean Technologies and Environmental Policy**, vol. 16(8), p. 1551–1564, 2014.

Salkind, N. J. **Statistics for people who (think they) hate statistics**. SAGE, London, 6 edition, 2017.

Seetharama-Yadiyal, V., Brighenti, G. D., and Zachos, P. K. Advanced gas turbine performance modelling using response surface methods, **Aeronautical Journal**, vol. 122(1258), p. 1871–1883, 2018.

Smrekar, J., Assadi, M., Fast, M., Kuštrin, I., and De, S. Development of artificial neural network model for a coal-fired boiler using real plant data, **Energy**, vol. 34(2), p. 144–152, 2009.

Smrekar, J., Pandit, D., Fast, M., Assadi, M., and De, S. **Prediction of power output of a coal-fired power plant by artificial neural network**, 2010.

STEAG. **Design and engineer your plants with EBSILON®Professional – STEAG's innovative, world-leading planning tool**, 2020.

Strušnik, D., Golob, M., and Avsec, J. **Artificial neural networking model for the prediction of high efficiency boiler steam generation and distribution**. vol. 57, 2015.

The Babcock & Wilcox Company. **Steam: its generation and use**. The Babcock & Wilcox Company, Charlott, 42 edition, 2015.

Tunckaya, Y. and Koklukaya, E. 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(1), p. 11–18, 2015.

van Eck, N. J. and Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping, **Scientometrics**, vol. 84(2), p. 523–538, 2010.

Verma, A., Rao, A. D., and Samuelsen, G. S. Sensitivity analysis of a Vision 21 coal based zero emission power plant, **Journal of Power Sources**, vol. 158(1), p. 417–427, 2006.

Wakiru, J. M., Pintelon, L., Muchiri, P. N., and Chemweno, P. K. A simulation-based optimization approach evaluating maintenance and spare parts demand interaction effects, **International Journal of Production Economics**, vol. 208(October 2018), p. 329–342, 2019.

Weissman, S. A. and Anderson, N. G. **Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications**, 2015.

Xu, Y., Sun, Y., Ma, Z., Wang, R., Wang, X., Wang, J., Wang, L., Gao, X., and Gao, J. Response surface modeling and optimization of electrodialysis for reclamation of RO concentrates in coal-fired power plants, **Separation Science and Technology (Philadelphia)**, vol. 0(00), p. 1–11, 2019.

APPENDIX A – Systematic Literature Review

A.1 Research Base

Table A.1 – Search strategy protocol for the conduction of the systematic literature review [Adapted from Dresch et al., 2015].

Research question How to standardize the operation of the steam generator of a coal fired power plant in order to increase its performance based on surrogate models of the process? Are there any DoE and RSM studies for building surrogate models of coal-fired power plants?
Conceptual framework Coal-fired power plants; Surrogate Models; Design of Experiments; RSM
Horizon Limitless
Theoretical currents There is no theoretical current to be followed.
Languages English and Portuguese
Review Strategy Aggregative
Search criteria: (I) Inclusion (E) Exclusion (I) applies DoE in coal-fired power plants; (I) applies RSM in coal-fired power plants; (I) develops a surrogate model based on significant parameters in thermal power plants; (E) article related to other areas, such as chemistry; (E) does not apply DoE, RSM or surrogate modeling methods in thermal power plants
Search terms TITLE-ABS-KEY ((("Design of Experiment*") OR ("response surface methodology") or ("surrogate model*")) AND (("power plant*") OR ("coal-fired") OR ("thermoelectric power"))
Search sources Scopus; Web of science.

APPENDIX B – Design of Experiments Applied to the Steam Generator of PECEM power plant

B.1 Behavior Graphs

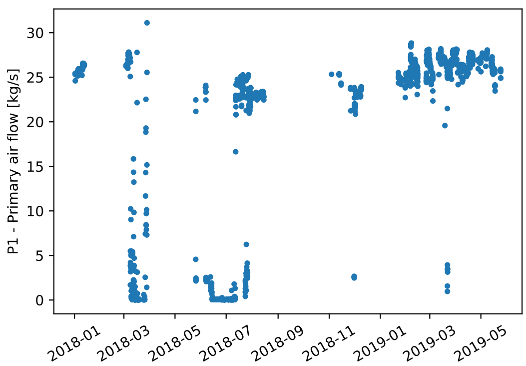


Figure B.1 – Primary air flow versus time for group 2 @ 360 MW baseline

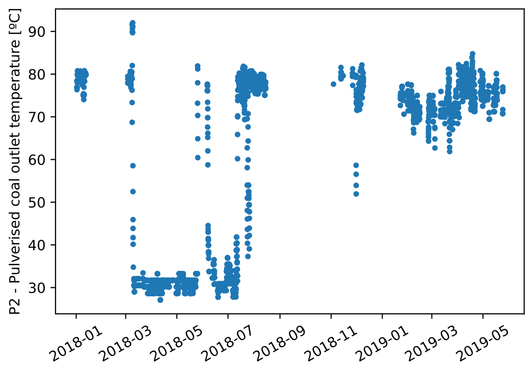


Figure B.2 – Pulverized coal outlet temperature versus time for group 2 @ 360 MW baseline

Primary air flow (P1) was mostly found to operate around 25kg/s during the assessed period, without important variation around that value, but with some periods off duty (flow rate close to zero). Pulverized coal outlet temperature (P2) ranged from 65 to 85°C, but with some atypical values around 30°C along March to June 2018 P2, as mill A was undergoing maintenance job.

Excess O_2 (P5) is presented in Figure B.3 ranged from 1.5 to 3.5%. The values start to decrease in the beginning of 2019 until the end of May 2019, but remain in this range.

Secondary and primary air crossover duct pressure P6 and P7 displayed an important variation for the same power output baseline (Figures B.4 and B.5). They seem to have a change in behavior from November 2018 but it is not so evident. P6 has a larger data cluster around 16mbar from this date and P7 around 90mbar

As a complementary analysis, each factor was plotted in respect to the steam generator efficiency followed by its respective heatmap. Figure B.6 presents the primary air flow (P1) versus the steam generator efficiency (S1).

As the data set is large, the graphs present in many cases a point cloud that makes

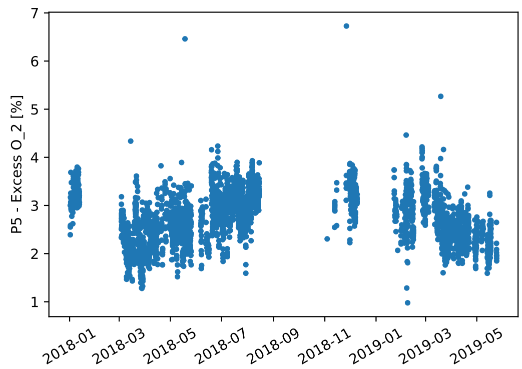


Figure B.3 – Excess O_2 versus time for group 2 @ 360 MW baseline

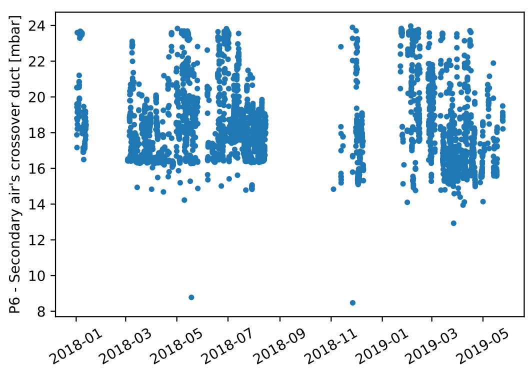


Figure B.4 – Secondary air crossover duct pressure versus time for group 2 @ 360 MW baseline

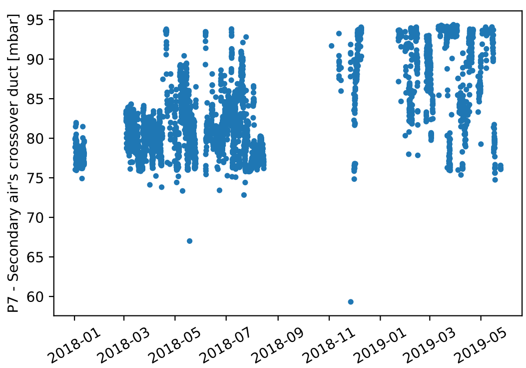
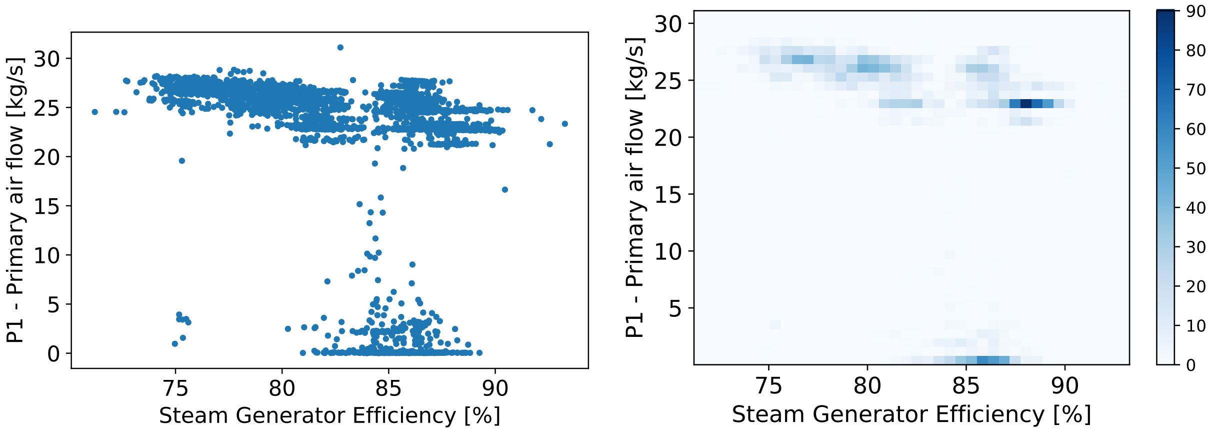


Figure B.5 – Primary air crossover duct pressure versus time for group 2 @ 360 MW baseline

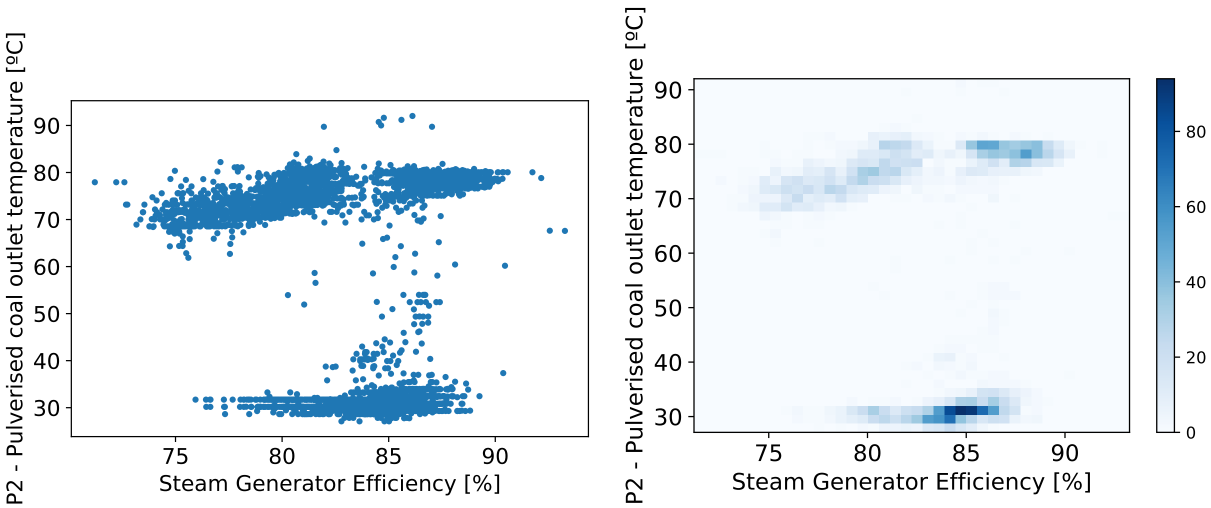
it difficult to realize which condition has the greatest number of repetitions. Analyzing Figure B.6a S1 values by 80% seem to be the condition with the greatest number of points, but it is hard to be sure. The heatmap enables a clear analysis, where the regions with darker colors are the regions with most operating points, while lighter regions aren't as operated. Now it is possible to notice that S1 values around 86% is the most common and the equivalent P1 is around 24kg/s.

The Figure B.7a looks like a point cloud. However, it is possible to notice from Figure B.7b that the predominate condition around 80°C to the steam generator efficiency achieving values close to 90%. The pulverized coal outlet temperature oscillates between



(a) Plot graph (b) Heatmap

Figure B.6 – Primary air flow (P1) by steam generator efficiency (S1) from January 2018 to May 2019 - GU2 of PECEM power plant



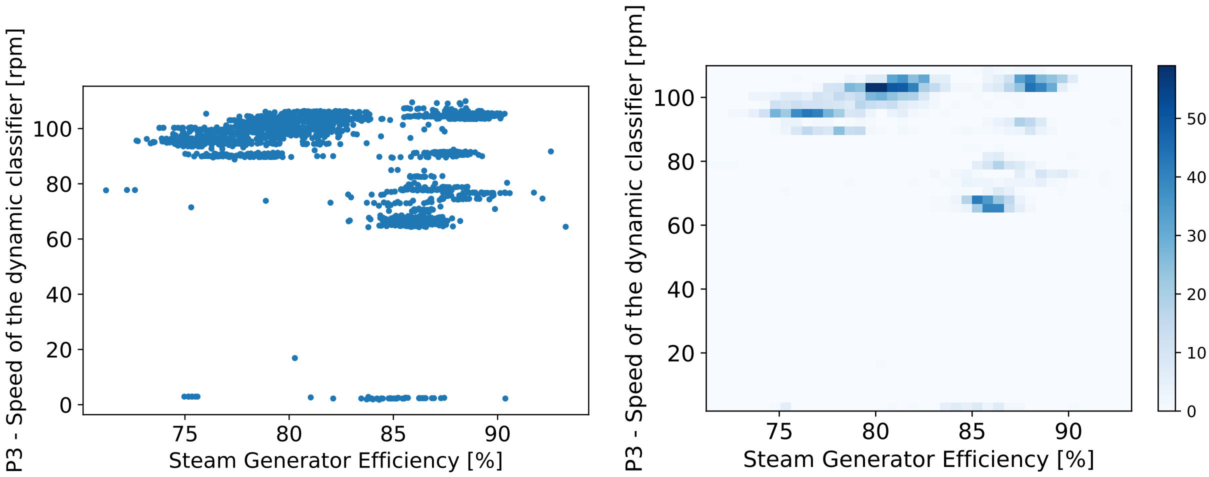
(a) Plot graph (b) Heatmap

Figure B.7 – Pulverized coal outlet temperature (P2) by steam generator efficiency (S1) from January 2018 to May 2019 - GU2 of PECEM power plant

70 and 80°C. The lower values of P2 are associated with lower values of S1.

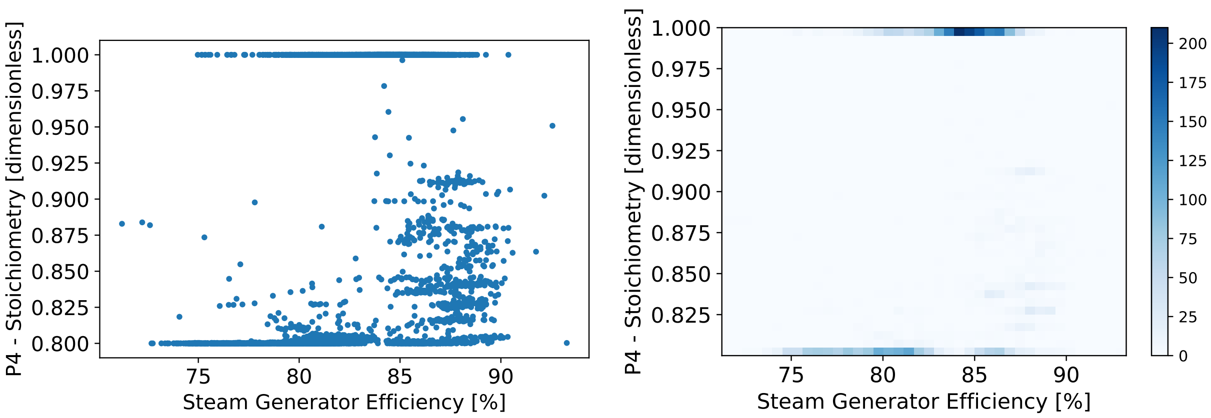
Analyzing Figures B.8a and B.8b it is not clear the relation between the parameters. The operation range of P3 keeps around 100 rpm even with the variation of S1.

The stoichiometry (P4) has a well-defined behavior and operate or in 1.0 either in 0.8. The correspondent S1 presents bigger values for P4 equal to 1.0. In the scale



(a) Plot graph (b) Heatmap

Figure B.8 – Speed of the dynamic classifier (P3) by steam generator efficiency (S1) from January 2018 to May 2019 - GU2 of PECEM power plant

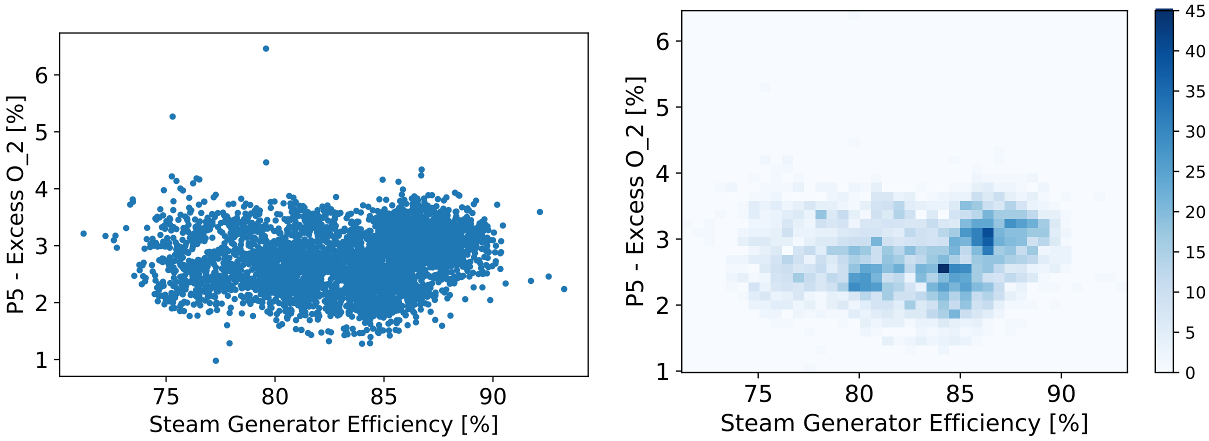


(a) Plot graph (b) Heatmap

Figure B.9 – Stoichiometry (P4) by steam generator efficiency (S1) from January 2018 to May 2019 - GU2 of PECEM power plant

of the colors presented in Figure B.9b the dark blue is equal or higher than 300 points. Moreover, it was possible to notice in Figure 3.11 that the higher values of stoichiometry happened before July 2018. Even if the higher values of S1 are related to higher values of P4 cannot ignore the timing of these events and how the rest of the plant was operating.

There are no preferential condition looking at the graphs of Excess O_2 (P5) versus steam generator efficiency (S1) presented in B.10 which is confirmed by the heatmap.

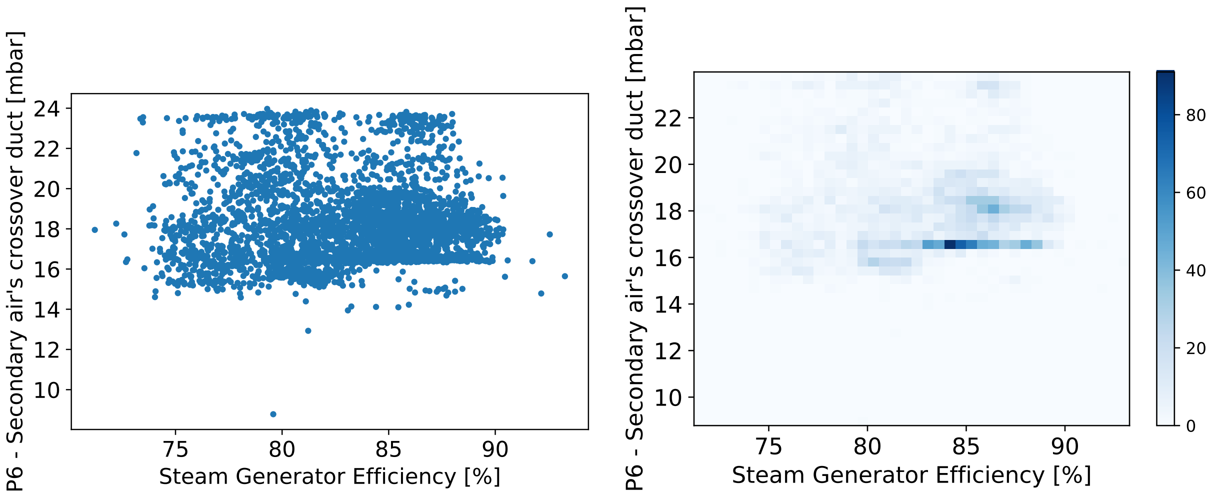


(a) Plot graph

(b) Heatmap

Figure B.10 – Excess O_2 (P5) by steam generator efficiency (S1) from January 2018 to May 2019 - GU2 of PECEM power plant

The dark blue represents only 50 points, much less when compared with the previous one (Figure B.9) The variation occurs around 1.5 and 3.5% for P5.

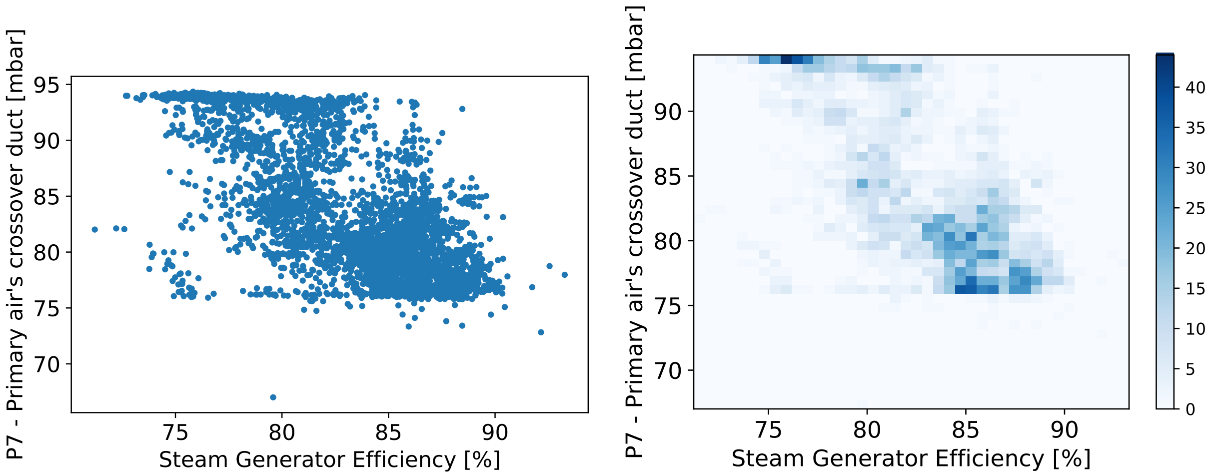


(a) Plot graph

(b) Heatmap

Figure B.11 – Secondary air's crossover duct (P6) by steam generator efficiency (S1) from January 2018 to May 2019 - GU2 of PECEM power plant

The data is relatively scattered for secondary air crossover duct pressure (P6) versus steam generator efficiency (S1). The higher values of S1 could be related to the range around 16 and 18 mbar.



(a) Plot graph (b) Heatmap

Figure B.12 – Primary air's crossover duct (P7) by steam generator efficiency (S1) from January 2018 to May 2019 - GU2 of PECEM power plant

According to Figure B.12 the secondary air crossover duct pressure works in two main operational conditions of 75 and 93 mbar, however, there is a lot of data between them. The higher values of S1 seems to be related to the lower values of P7.

B.2 ANOVA

Table B.1 – Analysis of variance (ANOVA) for the complete model with all linear, square and interactions terms

Source	DF	Adj Sum of Squares	Adj Mean Squares	F-Value	P-Value
Model	26	30.4291	1.1703	5602.98	0.000
Linear	6	23.6877	3.9479	18900.55	0.000
P1	1	0.0898	0.0898	429.75	0.000
P2	1	12.7262	12.7262	60926.19	0.000
P4	1	0.2879	0.2879	1378.27	0.000
P5	1	9.6043	9.6043	45979.99	0.000
P6	1	0.0001	0.0001	0.55	0.466
P7	1	0.0000	0.0000	0.00	0.955
Square	6	6.7367	1.1228	5375.29	0.000
P1*P1	1	0.0007	0.0007	3.56	0.072
P2*P2	1	4.2692	4.2692	20438.42	0.000
P4*P4	1	0.0226	0.0226	108.07	0.000
P5*P5	1	0.0066	0.0066	31.70	0.000
P6*P6	1	0.0032	0.0032	15.16	0.001
P7*P7	1	0.0007	0.0007	3.31	0.082
2 - Way interaction	14	0.0047	0.0003	1.60	0.154
P1*P4.B	1	0.0007	0.0007	3.31	0.082
P1*P5	1	0.0029	0.0029	14.02	0.001
P1*P6	1	0.0000	0.0000	0.00	0.979
P1*P7	1	0.0000	0.0000	0.01	0.924
P2*P4	1	0.0000	0.0000	0.02	0.897
P2*P5	1	0.0001	0.0001	0.28	0.603
P2*P6	1	0.0000	0.0000	0.00	0.995
P2*P7	1	0.0000	0.0000	0.00	0.969
P4*P5	1	0.0002	0.0002	1.03	0.320
P4*P6	1	0.0000	0.0000	0.00	0.981
P4*P7	1	0.0000	0.0000	0.00	0.999
P5*P6	1	0.0000	0.0000	0.00	0.988
P5*P7	1	0.0000	0.0000	0.00	0.964
P6*P7	1	0.0000	0.0000	0.00	1.000

B.3 Surrogate model validation

Table B.2 – Execution of the experiments through the simulation model

	Factors (controllable parameters)						Response S1	
Experiment number	P1	P2	P4.B	P5	P6	P7	Simulation model	Surrogate model
1	28.0	75	0.80	2.3	21	85	83.42	83.43
2	28.0	85	0.88	3.0	21	78	82.65	82.57
3	28.0	65	0.88	3.0	21	78	80.97	80.94
4	28.0	65	0.88	1.5	21	78	82.38	82.36
5	26.0	75	0.88	2.3	21	78	83.33	83.33
6	26.0	85	0.88	2.3	23	85	83.42	83.41
7	26.0	75	0.80	1.5	21	85	84.15	84.16
8	26.0	85	0.80	2.3	23	78	83.56	83.56
9	26.0	75	0.95	1.5	21	85	83.94	83.94
10	26.0	65	0.95	2.3	18	78	81.71	81.71
11	28.0	75	0.88	3.0	18	78	82.52	82.53
12	24.0	75	0.80	2.3	21	70	83.55	83.54
13	26.0	65	0.88	2.3	23	85	81.77	81.78
14	28.0	75	0.95	2.3	21	85	83.18	83.19
15	24.0	75	0.95	2.3	21	70	83.34	83.34
16	26.0	85	0.80	2.3	18	78	83.55	83.56
17	26.0	75	0.95	3.0	21	85	82.55	82.55
18	26.0	75	0.95	1.5	21	70	83.94	83.94
19	28.0	75	0.95	2.3	21	70	83.18	83.19
20	26.0	75	0.88	2.3	21	78	83.33	83.33
21	24.0	75	0.88	1.5	18	78	84.06	84.05
22	24.0	75	0.88	3.0	18	78	82.71	82.70
23	24.0	85	0.88	1.5	21	78	84.06	84.09
24	26.0	75	0.80	3.0	21	85	82.78	82.77
25	24.0	75	0.88	3.0	23	78	82.71	82.71
26	26.0	85	0.95	2.3	18	78	83.35	83.34
27	24.0	85	0.88	3.0	21	78	82.71	82.74
28	26.0	75	0.88	2.3	21	78	83.33	83.33
29	26.0	65	0.80	2.3	18	78	81.91	81.93
30	26.0	75	0.88	2.3	21	78	83.33	83.33
31	26.0	85	0.88	2.3	18	70	83.41	83.41
32	26.0	85	0.88	2.3	18	85	83.41	83.41
33	24.0	65	0.88	1.5	21	78	83.98	82.46
34	28.0	75	0.88	1.5	23	78	83.95	83.96
35	24.0	65	0.88	3.0	21	78	82.62	81.11
36	24.0	75	0.80	2.3	21	85	83.55	83.54
37	26.0	65	0.88	2.3	18	70	81.77	81.78
38	26.0	85	0.88	2.3	23	70	83.42	83.41
39	26.0	65	0.88	2.3	23	70	81.77	81.78
40	26.0	75	0.80	3.0	21	70	82.78	82.77
41	26.0	65	0.95	2.3	23	78	81.71	81.71
42	26.0	65	0.88	2.3	18	85	81.77	81.78
43	28.0	75	0.88	1.5	18	78	83.95	83.95
44	26.0	75	0.88	2.3	21	78	83.33	83.33
45	26.0	65	0.80	2.3	23	78	81.92	81.93
46	28.0	75	0.88	3.0	23	78	82.52	82.54
47	24.0	75	0.88	1.5	23	78	84.07	84.06
48	28.0	75	0.80	2.3	21	70	83.42	83.43
49	28.0	85	0.88	1.5	21	78	84.06	83.99
50	26.0	75	0.80	1.5	21	70	84.15	84.16
51	26.0	85	0.95	2.3	23	78	83.35	83.34
52	26.0	75	0.88	2.3	21	78	83.33	83.33
53	24.0	75	0.95	2.3	21	85	83.34	83.34

Continued on next page

Table B.2 – *Continued from previous page*

	Factors (controllable parameters)						**Response S1**	
Experiment number	P1	P2	P4.B	P5	P6	P7	Simulation model	Surrogate model
54	26.0	75	0.95	3.0	21	70	82.55	82.56