**Abstract:** Some (non)manufacturing industries are becoming more energy efficient, but many of them are losing cost-effective energy-savings opportunities, namely, by lack of knowledge or underestimation of good engineering and management practices as well as guidance on techniques or tools for that purpose. This study points out that Design of Experiments is a tool that cannot be ignored by managers and other technical staff, namely, by those who have the responsibility to eliminate energy waste and promote energy-efficiency improvement in industry, mainly in energy-intensive manufacturing industries. A review on Design of Experiments for physical and simulation experiments, supported on carefully selected references, is provided, since process and product improvement at the design and manufacturing stages increasingly rely on virtual tests and digital simulations. However, the expense of running experiments in complex computer models is still a relevant issue, despite advances in computer hardware and software capabilities. Here, experiments were statistically designed, and several easy-to-implement yet effective data analysis methods were employed for identifying the variables that must be measured with more accurate devices and methods to better estimate the energy efficiency or improve it in a billets reheating furnace. A simulation model of this type of furnace was used to run the experiments and the results analysis shows that variables with practical effect on the furnace’s energy efficiency are the percentage of oxygen in the combustion gases, the fuel flow in the burners, and the combustion air temperature.

**Keywords:** computer experiments; sustainability; furnace; simulation; screening; factorial design

**1. Introduction**

Energy efficiency is one of the most cost-effective ways to mitigate climate change, improve energy security, and grow economies while delivering environmental and social benefits [1]. To boost the progress on energy efficiency is imperative to achieve these purposes and, more than ever, a strong push is needed, since the global energy-efficiency progress has been slowing in the last five years. The reasons for this slowdown include [2]: (i) energy-intensive industries’ demand for all primary energy fuels has increased; (ii) weather—a cooler winter and a warmer summer in the United States and a milder one in Europe drove up energy use for both heating and cooling; (iii) longer-term structural factors—technologies and processes are becoming more efficient, but transport modes and more building floor area per person are dampening the impact of these technical efficiency gains on energy demand; (iv) policy progress and investment—the coverage and strength of energy-efficiency obligation programmes as well as investment targeting efficiency have remained largely unchanged.

In Europe, energy efficiency has gained in significance and is one of the five dimensions of the European Energy Union [3]. In December 2018, a revised Energy Efficiency Directive [4] entered into force, updating some specific provisions and introducing some new elements to European Union (EU) Directive 2012/27 [5]. The key element of the amended directive is a headline EU energy-efficiency...
target for 2030 of at least 32.5% (compared to projections) with a clause for a possible upwards revision by 2023. As an integral part of the European Commission’s strategy to implement the United Nation’s 2030 Agenda and other priorities to sustainable development of the EU’s economy, also recently launched was the European Green Deal [6], whose major purpose is to make the EU a climate-neutral region by 2050. The International Energy Agency (IEA) is also putting emphasis on the need for urgent global action on energy efficiency and established, in the summer of 2019, an independent high-level global commission to examine how progress on energy efficiency can be rapidly accelerated through new and stronger policy action. At the United Nations Climate Change Summit 2019, fifteen leading countries also committed to accelerate progress on energy efficiency as well [2]. In addition to twelve strategies to set up global energy efficiency submitted to IEA high-level commission by three expert non-governmental organizations [7], a call to action for energy efficiency was also launched by the American Council for an Energy-Efficient Economy [8]. Recommendations on how governments can accelerate energy-efficiency progress were already reported by the Global Commission for Urgent Action on Energy Efficiency [9].

2. Methodology and Objective

Despite energy efficiency’s tremendous potential, the world is struggling to capture its full benefits [8]. Nevertheless, progress with regards to implementation of various provisions included in the Energy Efficiency Directive is apparent [10]. Organizations are increasingly integrating energy efficiency into their business strategy, because a (small or moderate) financial burden for energy-efficiency programs cannot realistically be considered a strong determinant to reduce their competitiveness [11]. Energy-efficiency programs significantly contribute to business sustainability and improvement, compliance with tightening regulatory requirements, substantial direct and indirect costs reduction, and the image of an industrial or consumer goods manufacturing company [12]. There are many energy-efficient technologies, methods, engineering and management practices that can be implemented for energy-efficiency improvement and energy savings in a (non)manufacturing industry, and considerable efforts have been exerted to achieve green and environmental benign processes [13]. Unfortunately, besides the existing examples of good engineering and management practices and guidance on techniques or tools for these purposes published by scientific journals and non-governmental organizations like the European Council for an Energy-Efficient Economy, European Union, European Commission, and IEA (see as examples, [14–19]), much more must be done for boosting those practices and tools [20,21]. Efforts have been put into information and knowledge transfer, and learning energy-efficiency networks exist in some countries/regions, but one cannot ignore that many organizations, namely, the small and medium-sized ones, do not automatically translate information into useable knowledge [22]. In fact, policy experience in energy efficiency is rich and extensive, but policy makers have much to learn from each other in policy selection, design and implementation [9].

Energy-efficiency improvement of energy-intensive equipment is, surely, a priority decision or immediate action that organizations can (should) take as a simple strategic objective, a result of energy management system implementation, due to governmental stimulus programmes, and either simultaneously or before other practices and actions, namely, equipment update or replacement for energy savings, energy performance improvement, and costs reduction. Billets reheating furnaces are examples of energy-intensive equipment, and their modelling and control have been a research topic since the 1970s [23]. However, to the best of our knowledge, no study was found where Design of Experiments (DoE-A tool through which data is collected and analysed by statistical methods) was employed for assessing its efficiency or improving its performance, namely, for identifying variables with practical effect or statistically significance in a furnace’s energy efficiency. This is not surprising, because many practitioners, namely, those who are not experts in statistics, do not feel comfortable in using statistical tools [24,25]. When statistically designed experiments are planned and run appropriately, researchers and practitioners can expect to take better-informed decisions as well as
improve and optimize processes’ efficiency and products’ quality (reliability, durability, performance, robustness, etc.) in a structured, faster, and cheaper way at the design and manufacturing stages. Unfortunately, the trial-and-error, one-factor-at-time, and brute force are common practices in industry to understand process or equipment and product behaviour or to improve its performance. As examples, Saxena et al. [26] stated that earlier studies generally investigated the effect of One-Factor-At-Time (OFAT) to provide information on combustion behaviour, along with performance and emission characteristics of a diesel engine, and Cheng and Liu [27] adopted the OFAT procedure to optimize an injection moulding process and maximize the related energy savings. DoE is a widely tested and more efficient tool that has been successfully used in chemical, aerospace, automotive, and electronics industries, to cite only a few. Examples in systems (process/equipment) energy-efficiency improvement and energy saving were reported in Ref. [28–34]. The objective of most academic and industrial research works is response optimization, but researchers and practitioners must be aware that screening variables (separating dominant input variables from the non-dominant ones) is critical for that purpose in terms of efficiency and efficacy. Variables screening cannot be ignored or undervalued, and to employ more than one analysis method for the same data set is a recommended approach for taking a more informed decision about dominant input variables. This is true for physical and for simulation experiments as well, though data analysis methods for variables screening, and DoE in general, are not enough disseminated among researchers and practitioners, namely, among energy managers and auditors. As an example, to know the variables with significant effect or practical influence on the energy efficiency of an equipment is very useful for saving time and to get more accurate results when energy audits are performed, because energy auditors can select more accurate devices and methods for measuring those variables. Thus, this paper has a threefold purpose: (1) present a review on DoE for physical and simulation experiments; (2) review and illustrate easy-to-implement data analysis methods for separating dominant input variables from the non-dominant ones; (3) identify the variables with practical effect or influence on the energy efficiency of a billets reheating furnace so that they are measured with more accurate measuring devices and methods to better estimate the furnace’s energy efficiency.

The remainder of this paper is organized as follows: Section 3 reviews the DoE for physical experiments; Section 4 introduces the DoE for simulation experiments; Section 5 describes the case study, analysis methods, and presents the results discussion. Conclusions are in Section 6.

3. Design of Experiments—An Overview

The Design of Experiments (hereafter termed DoE) has been a fundamental tool in various activity areas, namely, in the aeronautic, chemistry, pharmaceutical, automotive, mechanical, electronic, and biomedical sectors. It can be adjusted to the characteristics of the phenomena and variables under study, as well as to the type of information that is intended to be obtained from an experimental study. DoE is a flexible and cost-effective tool for developing and optimizing a process’s efficiency and product’s quality, as well as services, where the benefits and challenges of DoE usage are also identified [35–37].

Comprehensive guidelines to help researchers and practitioners in planning, conducting, and analyzing experiments were reported in Ref. [38–41]. However, unfortunately, many researchers and practitioners are not familiar with the DoE or use it sparingly at best [42]. Such as Bergquist [24] stated, the reasons why DoE is not often used by those working on research and development of processes and products, including those who knew this tool, lie in the technical domain and in how statistical methods are viewed by decision-makers (managers, engineers, and other technical staff). Costa [43] revisited and discussed the hindrances to DoE usage as well as bad practices in using this tool and argued that, to succeed with DoE, a solid background on nonstatistical issues (characterize the problem; define objective, response, variables and test levels; select, plan, conduct, and control people, materials and devices) is required from the users, in addition to statistical ones (design the experiments, collect, analyze, and interpret data). Without such a background, trying to implement
DoE can become a frustrating task, because unexpected barriers may arise and bad practices are not avoided, resulting in a waste of resources, misconceptions about DoE usefulness, and doubts about DoE user technical competences.

3.1. DoE’s Principles and Experimentation Strategy

Today, there are many off-the-shelf software packages that allow for designing and analysing experiments in a friendly manner. However, software is not a magic wand that conveys answers to the research questions. To understand DoE’s principles is fundamental to succeed with this tool. Blocking, Randomization, and Replication are the three DoE principles. These are not just theoretical (statistical) concepts. They impact significantly on the experimental results and conclusions. Blocking is a useful procedure to reduce or eliminate the nuisance variables contribution to experimental error and provides a more accurate estimate of variables effect on response, while Replication allows getting a more precise estimate of pure error (the representative value of experimental variability) and response mean. Randomization is a recommended practice for safeguarding the experimental data from an imbalanced effect of unknown sources of variation (lurking variables or noise) in some treatments and from any type of systematic bias, which improves the quality of collected data, makes the data analysis less sophisticated, and provides more confidence in results interpretation.

If any principle is violated, the validity or reliability of the conclusions will be compromised. This is true at any phase of an investigation guided by the so-called scientific method, which involves making conjectures (hypotheses), deriving predictions from them as logical consequences, and then carrying out experiments or empirical observations based on those predictions to determine whether the original conjecture was correct.

It is widely accepted that an efficient and recommended way to build knowledge about process and/or product is to adopt a sequential experimentation strategy which may consist of three experimental phases (Screening, Characterization, and Optimization) whose objective is the following:

a) Identifying the so-called active variables (input variables with practical or significant effect on response)—Screening.

b) Understanding the relationship between the selected variables and the response by guiding these variables, and their interactions, to the region where the response yields the most favourable results—Characterization.

c) Modelling one or more responses, usually by second order models, and identifying optimal settings for the influential (significant) variables—Optimization.

This sequential learning process is inherent to the called Response Surface Methodology (RSM). RSM is not a single experiment (series of runs where purposeful changes are made to the input variable values of a system (process/equipment or product). Even in each phase, it may consist of more than one set of runs, each one building upon what the team learns from the previous experiment.

Curiously, most case studies reported on scientific journals and textbooks on DoE do not illustrate the full sequential nature of an investigation. In many problems, the authors justify their approach to solve the problem, and use two out of three experimental phases, namely, screening/characterization [44] and screening/optimization [45]. A comprehensive example of the sequential learning process in the response surface methodology framework, including the optimization of multiple responses simultaneously, is presented by Lv et al. [46].

3.2. DoE–Experimental Design Selection and Results Analysis Methods

Developments in computer hardware and software definitively expanded the DoE usage, avoiding complex and tedious hand calculations. Today, many off-the-shelf software packages have good capability to generate a great variety of experimental designs and tools to analyse the collected data. However, they do not come with the ability to choose neither the correct design nor tool for data analysis nor to interpret the results.
Experimental design selection depends on the problem characterization, study objective, number and type of variables, and resources available. The design choice has relied heavily on (classical) two-level factorial designs, which have good properties for a wide range of applications. They can be considered first when planning an experiment; however, to deal with resource or budget constraints, design region constraints, and when a nonstandard model is expected to be required to adequately explain the response, other design options are available. With the advent of readily available computer power, there has been a strong movement toward (A-, D-, G-, I-, . . . ) optimal and, more recently, definitive screening designs [47]. Nevertheless, notice that no design can provide in one shot what is expected to be done through a sequential (screening–characterization–optimization) experimentation approach. For further guidelines and discussion on experimental design selection at various experimental stages in the response surface methodology framework, namely, when constraints exist, the reader is also referred to [39,48–55]. To deal with multiobjective when constructing a design, the reader is referred to Lu et al. [56].

Conclusions drawn from an experiment (series of runs in which purposeful changes are made to the input variable values of a system (process/equipment or product) so that we may observe and identify the reasons for changes in the response) depend to a large extent on the manner in which and what data were collected. This means that no results analysis method can neutralize a badly designed experiment. In fact, experimental design and results analysis method selection are tasks that are closely intertwined, though methods performance and the analyst’s preference can be also considered in the selection of a results analysis method.

In the early phase of an investigation, it is not expected that an experimenter has a complete and profound knowledge about new processes and/or products. Thus, the screening phase plays a very important role in the RSM, because the monetary and time costs of experimentation grow exponentially with the number of variables or factors in the subsequent phases of experimentation. How to identify the location effects (those that have effect on response mean) from unreplicated designs has been a widely researched topic, and three general guidelines (see Ockuly et al. [57] for an empirical quantification of these guidelines in the RSM framework based on a meta-analysis) must be considered in the results interpretation:

a) Only a small fraction of tested factors will be active (influence or have a statistically significant effect on response). This is called the effects sparsity principle.

b) Factors main effect is generally larger than the effect of interactions between two factors, and the latter is larger than that of three-factor interactions. This is called the effects hierarchy principle.

A two-factor interaction has more chance to influence the response (is active) when, at least, one of the individual factors is active. This is called the effects inheritance principle.

Worthy additions to the ongoing efforts to help researchers and practitioners in selecting a method for analysing unreplicated experiments from physical experiments are the works by Hamada and Balakrisnan [58] and Chen and Kunert [59], who compared the performance of a wide variety of methods to identify the location effects on an equitable basis without destroying their essence. Hamada and Balakrisnan [58] adjusted the methods so that their Individual Error Rate, that is, the probability they identify as active, an inactive effect under the null hypothesis that no contrast is active, was as close to 5% as possible when active contrasts are of the same size. Chen and Kunert [59] adjusted the methods so that its Experimentwise Error Rate, that is, the probability of declaring at least one inactive effect as active under the null hypothesis that no contrast is active, was as close to 5% as possible when active contrasts with the same and different size exist. Doing so, these authors tried to minimize the Type II error, i.e., to misidentify an active effect as inactive, which is more severe than to misidentify an inactive effect as active (Type I error).

Unfortunately, one method that performs better than all the others in the screening phase was not found so far [60]. The methods performance depends on the number and size of the active effects as well as on the existence of abnormalities in the data. This is a serious limitation, because the number
and size of active effects can only be estimated. Thus, Costa et al. [61] tested various methods from a non-expert DoE user’s perspective (no adjustment in the methods was made) and argued that several methods must be used simultaneously to avoid Type I and Type II errors, and to contemplate the possibility of outliers in the data. To identify or select a method(s) for factor screening in nonregular factorial designs and when response is non-normal, the readers are referred to [62–64].

Researchers and practitioners sometimes aim at achieving a practical solution for a problem rather than the optimal solution, namely, when it is imperative to get out of an embarrassing situation with a client(s) and to produce a commercially valid product as fast as possible [65]. For this purpose, guiding selected variables, and their interactions, to the region where the responses value yield, the most favourable results, and fulfil the specification is the appropriate or recommended procedure. It may include dropping and adding factors from an initial (fractional) factorial design, rescaling some factors, and replicate runs at the design center point to account for response curvature. In this experimentation phase, the analysis of variance (ANOVA) has been often used to test for the significance of main effects and interactions. However, only after investigating the assumptions for using ANOVA, the experimenter is ready to draw practical conclusions from the data analysis. In fact, it is unwise to rely on the ANOVA results until the assumptions validity (errors in the model fitted to response are normally, independently, and identically distributed with zero mean and a constant variance) has been checked. Violations of the assumptions, and model adequacy, can be easily investigated by residual analysis and other diagnostic checking procedures (for details, see any classic textbook on DoE such as that by Montgomery [66]. Unfortunately, such as Sheil and Hale [25] also stated in medical device manufacturing, there are evidences that a significant number of researchers and practitioners use statistical techniques in the belief that knowledge of the ‘theory’ (assumptions, limitations, etc.) is not required, since the validation of ANOVA results is not illustrated or commented on in many works published in journals. These computations are easily performed with any software package, so the burden of data analysis is no longer an acceptable excuse to not make appropriate use of statistical tools. This is also true in the optimization phase of experimentation, where the exploration of the response surface is carried out. In practice, this consists in fitting a second-order model to response(s), and then identifying the optimal settings for influential or significant variables. OLS is often used for modelling uncorrelated responses whose variance is homogeneous. When responses’ variance is non-homogeneous, Generalized Least Squares must be employed [67,68]. In the case of correlated responses and homogeneous variance, Seemingly Unrelated Regression technique has proved to perform well (is recommended), such as Ref. [69,70] show.

To aggregate the multiple response models into a composite function and then proceed to its optimization, it is a current strategy in the RSM framework. The two most popular composite functions (optimization criteria) are built on desirability and loss functions. An extensive review on desirability-based criteria and loss function-based criteria are available in Ref. [71,72]. Sophisticated desirability-based and loss function-based criteria are available in Ref. [73,74]. The applicability and computational aspects of various criteria in different decision-making contexts were discussed by Ardakani and Wulff [75], who also categorize and integrate the foremost approaches. Costa and Lourenço [76] evaluated and compared the working ability of several easy-to-use criteria with that of a theoretically sound method, and concluded that easy-to-implement criteria, namely, the mathematical programming-based and the compromise-based criteria reported in Ref. [77,78], were able to generate solutions similar to those generated by methods mathematically and computationally much more sophisticated, even when the objective is to depict the Pareto frontier (a set of nondominated solutions where any improvement in one response cannot be achieved without degrading, at least, another response). To help the decision-maker in selecting a nondominated solution, Costa and Lourenço [79] proposed two metrics to assess the predicted variability of nondominated solutions: (1) the predicted standard error, which quantifies the uncertainty in the estimated value for each response; (2) the quality of predictions, which quantifies the uncertainty associated to each generated solution. Another research direction that has been explored is focused on “how to optimize”
(modify or develop a search algorithm). For a review on the state of the art, special features, trends on the development of search algorithms, and systematic comparison of some local and global algorithms, the reader is referred to Ref. [80,81], as examples. Notice that the mathematical programming-based and the compromise-based criteria can be easily implemented in Microsoft Excel, and the solver tool used to optimize them.

4. DoE and Simulation Experiments

Computer codes (or simulators) are increasingly used to accelerate learning at the design and manufacturing stages of processes and products. They allow for conducting experiments with the purpose of understanding the behaviour and evaluating the impacts produced by changes in the system, which could hardly be possible otherwise, namely, when the system does not physically exist, the experimentation is time-consuming, too expensive, very difficult, or risky to be performed. The spectrum of applications ranges from the nano (molecular simulations) to mega scale (structural analysis of buildings and bridges) in various science fields, namely, in chemistry, mechanics, toxicology and pharmaceutics, materials, electronics and communication, biotechnology, aeronautics, and so on [82].

Despite advances in computer hardware and software capabilities, the expense of running experiments in simulators is still a relevant issue. Single evaluations of stress, thermal, and impact/crash analyses can take hours to days, if not longer. To minimize this inconvenient or hindrance, an alternative has been running experiments in meta-models or “models of the models.” This has spawned a great deal of research in the last three decades and sparked the development of new statistical methods for designing and analysing the called simulation or computer experiments. Simulators are becoming increasingly prevalent surrogates for physical experiments, but practitioners often treat their experiments with simulators (or simulation models) very unprofessionally [83,84]. Trial-and-error and brute force (massive computation) practices are not recommended and, in most situations, must be not used, because they may neither uncover all the input variables with practical effect on the response nor the best functional relationship between input variables and response, and therefore, the most favourable settings for input variables.

Selection of Experimental Design and Analysis Method

To design simulation experiments and analyse the data are more of a science than an art, and it is not a one-time task. It requires critical thinking and not simply some clicks from a mouse (supply a vector of design/input variables to the computer code and obtain a vector of responses/outputs). Many tools for designing and analysing computer experiments have been put forward in the literature, but there is a strong need for making them more accessible to the practitioners [82,85]. It is known that many practitioners, namely, those who have a low background on statistics, as well as on computation, do not feel comfortable in using (complex or sophisticated) statistical tools [25,42]. Thus, choosing an experimental design, defining the number of experimental runs, selecting and implementing methods (tools) for data analysis in each experimental phase, and interpreting the respective results can be daunting (very difficult or even impossible to perform) tasks for those practitioners. Nevertheless, the systematic literature review applied to discrete simulation-based optimization undertaken by Junior et al. [86] shows that, in the last 25 years, many researchers and practitioners did use the design and analysis of computer experiments for developing and improving industrial processes and products.

A feature that set physical experiments off from computer experiments is the variability in the response values due to known (controllable and uncontrollable) and unknown sources of variation. Variability is intrinsic to physical experiments and several procedures are defined in the RSM framework for minimizing it, because the reliability of conclusions drawn from data analysis (strongly) depends on how the sources of variability are managed. In simulation experiments, it is possible to evaluate stochastic and deterministic responses. Deterministic responses are output variables whose values have no variability, whereas stochastic responses are random variables whose values have variability artificially generated. When deterministic computer codes are used, what it is still too often in practice,
the output response is a not random variable so it makes no sense to repeat runs at the same factor settings and to randomly run the designed experiments for estimating the true experimental error and safeguarding the experimental data from the effect of unknown sources of variation (lurking variables or noise). To repeat runs at the same factor settings yield always the same result, and unknown sources of variation are not possible to include in the simulator.

The variety of experimental designs and modelling techniques for computer experiments is abundant and, just like in physical experiments, the selection of experimental design and data analysis technique depend on experimentation purpose (screening or modelling). Garud et al. [82] and Kleijnen [84] provided general reviews on experimental designs and modelling techniques for computer experiments. The former summarizes classic linear regression metamodels, including polynomials, and their designs, and explains how sequential bifurcation can screen hundreds of variables. He also summarizes Kriging and its designs and explains simulation optimization that uses either low-order polynomials or Kriging, including robust optimization. The later authors reviewed metrics to quantify space-filling, presented a detailed classification and chronological evolution of DoE, a comprehensive overview of research on DoE, discussing the static and adaptive DoE techniques as well as numerical and visual analyses of the prominent DoE techniques. They concluded with future directions followed by a list of possible opportunities and unexplored fields in DoE for computer experiments.

Strategies for screening, ranging from designs originally developed for physical experiments to those especially tailored to experiments on numerical models (group screening experiments, including factorial group screening and sequential bifurcation), were explored by Woods and Lewis [87]. A proposal to deal with contaminated data in screening problems (Robust Sequential Bifurcation) was proposed by Liu et al. [88]. Draguljic et al. [89] also provided a comprehensive assessment and comparison of screening strategies for interactions using two-level supersaturated designs, group screening, and a variety of data analysis methods including shrinkage regression and Bayesian methods. Georgiou [90] reviewed several methods for constructing and analysing two-, multi-, or mixed-level supersaturated designs. Joseph et al. [91] proposed a new version of space-filling designs for robustness experiments (more capable of accurately estimating the control-by-noise interactions, which is the key to make the system robust or insensitive to the noise factors). Kong et al. [92] presented a general method for designing sequential follow-up experiments in computer experiments. Bhosekar and Ierapetritou [85] reviewed recent advances in the area of surrogate models, and tested a variety of problems of two frequently used surrogates, the Radial Basis Functions and Kriging. They also provided an extensive review on model selection strategies, including model fitness and validation, as well as guidelines for the choice of surrogate model for modelling, feasibility analysis, and optimization purposes. For new statistical tests for leave-one-out cross-validation of Kriging models, the reader is referred to Kleijnen and van Beers [93].

Software and computational power advances enable researchers to develop and use a large variety of model-building techniques and data collection (sampling) strategies for better understanding and improving systems, as well as to solve new and increasingly complex real-life problems. However, the gap between academic research and practitioner needs is increasing. There is evidence that academia is focusing on ever-more complicated methodologies and, unfortunately, many are of no practical value, they are purely for academic consumption [94]. Academia must push the knowledge frontiers in developing new techniques and methodologies to handle more complex, large, and unstructured problems [95]; however, practitioners’ difficulties for implementing those techniques and methodologies and solving real problems cannot be ignored.

There are no universally best techniques and methodologies for solving problems, as each has its unique aspects and challenges. Regarding the choice of a surrogate model, this is not straight-forward because of trade-offs associated with each surrogate. There are many parametric, semiparametric, and non-parametric model-building techniques, but no single type of surrogates outperforms all other types for all types of problems [85]. Moreover, the performance of a surrogate model depends strongly on the quality as well as the number of samples. Thus, sampling strategies
(what data points are the best for building useful surrogates) are also critical in problem solving. The popular space-filling designs aim at filling the experimental region evenly with as few gaps as possible and are robust to modelling choices, so they are widely used designs for computer experiments. The behaviour of several Latin Hypercube Designs in the context of the Gaussian process model was examined by Pistone and Vicario [96], and Joseph [97] presented a review of space-filling designs, including the recently proposed maximum projection design. Space-filling designs with nominal factors for nonrectangular regions were investigated by Lekivetz and Jones [98].

A considerable body of literature on design and analysis of computer experiments exists, and one can state that Gaussian Process and the popular space-filling Latin Hypercube Design have been usual choices in a large number of theoretical and practical applications with modelling and optimization purposes. The Gaussian Process model is flexible and can fit complex surfaces with data from deterministic computer experiments. It provides an exact fit to that data, but this is no assurance that it will interpolate well at locations in the region of interest where there is no data, and no one seriously believes that the Gaussian Process model is the correct model for the relationship between the response and the design variables. Many engineering systems exhibit non-linearity, but in the early stage of experimentation, it is technically and economically unnecessary about any worry about non-linearity. In this stage, the so-called screening, the goal is to reduce the relatively large list of input variables or factors to a manageable few, identifying those (individual factors and two-factor interactions) with the largest effect on response, what can be done with the Sequential Bifurcation method. This is considered the most efficient and effective method for screening in simulation experiments [88]. This is particularly true when hundreds of factors are involved, the direction of influence of each factor is known and nonnegative, and high experimental costs are expected if RSM designs (classical designs for physical experiments) are adopted [84,99]. As these tools mature, the need for practitioner-oriented guidance and easy-to-use free or commercial software is clear [83,94], because there is no evidence that computer experiments analysis is included in the curricular program of most engineering courses as a curricular unit, such as classical experimental design, which is not considered yet [43].

5. Case Study

Physics-based computer simulation tools, whether for mechanical stress, thermal, fluid dynamics, or electromagnetics, among other examples, are becoming essential in the researchers’ and practitioners’ toolbox. Professionals from different fields are also realizing the benefits of simulation, and this case study is another contribution.

Integrated in a more extensive project supported by the European Regional Development Fund (AuditF project Lisboa-01-0247-FEDER-017980), the case study presented here was undertaken with the objective of supporting energy auditors in identifying the variables that must be measured with more accurate devices and methods to estimate the energy efficiency of a billets reheating furnace when energy audits are performed, though the approach presented below can be employed to any type of equipment or industrial process.

To build a billets reheating furnace in a real or reduced scale for experimental purposes was unrealistic, and running experiments in an industrial environment was impossible, namely due to technical and economic reasons. The alternative was to model the furnace, which is currently used to heat cylindrical billets with a length between 1.5 and 1.6 m and a diameter between 0.2 and 0.3 m. The furnace is 22 m long, can heat 10 tons/h of billets (copper metal alloys) up to 770 °C, and has a total of 312 burners, whose propane consumption is of 40 kg/h and the thermal power production is approximately 500 kW. The furnace modelling was performed with a commercial software and, due to the complexity of thermodynamic phenomena inside the furnace and the size of this equipment, the furnace simulator was built by zones. The results presented here were collected from zone 8, the penultimate furnace zone with 16 burners with a thermal power of 25 kW. This zone represents appropriately the furnace operation conditions, was validated by values collected in industrial environment, and allows running the planned experiments in the planned (imposed) time to do it.
5.1. Design and Analysis of Experiments: Theoretical Framework and Results

The study presented here focused on the so-called screening phase of RSM, since the objective was to identify the variables that must be measured with more accurate devices and methods to better estimate the furnace’s energy efficiency. The variables and their test values (current, low, and high levels) considered in this study are listed in Table 1, and they were defined by senior energy auditors and academic experts in thermodynamics, combustion technologies and processes, and energy management.

Table 1. Input Variables and Settings.

| Variable                        | Value          |
|---------------------------------|----------------|
| v1 Fuel Flow in the Burner (kg/s) | 0.0011 0.0009 0.0013 |
| v2 O₂ in Combustion Gases (% vol dry) | 11 9 13       |
| v3 Combustion Air Temperature (°C) | 275 260 290   |
| v4 Billet Input Temperature (°C) | 730 725 735   |
| v5 Billet Output Temperature (°C) | 745 740 750   |
| v6 Billet Emissivity            | 0.80 0.72 0.88 |

With this (small) set of variables, a two-level fractional factorial design (2ᵏ⁻ᵖ design with k variables or factors and p independent generators) was selected. This design type is often used for screening in physical and simulation experiments when the number of variables is small (say up to about 15 to 20 variables, as it is common in physical experiments) due to their efficiency, effectiveness, and versatility for sequential experimentation [57,84]. In 2ᵏ⁻ᵖ designs, each factor is tested at two levels and only a fraction (1/2ᵖ) of all factor-level combinations are run. The degree of acceptable confounding (or aliasing) among estimated effects determines the resolution of the design. The lower the resolution (denoted by the roman numerals III, IV, V, …), the more the aliasing will be. When some interactions are expected to have effect on response value, resolution IV designs are appropriate.

In the case study presented here, a Minimum Aberration Design (a fractional factorial of resolution IV with sixteen experiments (2⁶⁻²) and a minimum number of confounded effects) was adopted to establish the experimental runs. Variables v5 and v6 were used as generators, assuming v5 = v1 × v2 × v3 and v6 = v1 × v2 × v4, because, according to the experts, it was not expected that variables v5 and v6 were the most determinant for the study purpose. The aliased effects structure is presented in Table 2, and the designed experiments are listed in Table 3. Two-factor interactions that cannot be estimated separately due to the reduced number of experiments (16 out of 64 experiments), though they may have effect on response (energy efficiency), are called redundant effects and are listed in the second row of Table 2. The energy-efficiency values listed in Table 3, calculated by the ratio between the heat transferred to the billet and the heat released by the fuel’s combustion and expressed in percentage (%), correspond to the results of the sixteen performed experiments.

Table 2. Aliased Effects Structure.

| Aliased Effects         |
|-------------------------|
| Effect                  |
| v1 × v2                 |
| v1 × v3                 |
| v1 × v4                 |
| v1 × v5                 |
| v1 × v6                 |
| v3 × v4                 |
| v3 × v6                 |
| Alias                   |
| v3 × v5, v4 × v6        |
| v2 × v5                 |
| v2 × v6                 |
| v2 × v3                 |
| v2 × v4                 |
| v5 × v6                 |
| v4 × v5                 |
Table 3. Experimental Design and Results.

| Exp. | v1   | v2  | v3  | v4  | v5  | E  | Efficiency (%) |
|------|------|-----|-----|-----|-----|----|----------------|
| 1    | 0.0009 | 9   | 260 | 725 | 740 | 0.72 | 56.1          |
| 2    | 0.0013 | 9   | 260 | 725 | 750 | 0.88 | 51.9          |
| 3    | 0.0009 | 13  | 260 | 725 | 750 | 0.72 | 45.7          |
| 4    | 0.0013 | 13  | 260 | 725 | 740 | 0.72 | 55.8          |
| 5    | 0.0009 | 9   | 290 | 725 | 750 | 0.72 | 57.6          |
| 6    | 0.0013 | 9   | 290 | 725 | 740 | 0.72 | 53.7          |
| 7    | 0.0009 | 13  | 290 | 725 | 740 | 0.72 | 46.5          |
| 8    | 0.0013 | 13  | 290 | 725 | 750 | 0.72 | 41.2          |
| 9    | 0.0009 | 9   | 260 | 735 | 740 | 0.88 | 56.7          |
| 10   | 0.0013 | 9   | 260 | 735 | 750 | 0.72 | 50.6          |
| 11   | 0.0009 | 13  | 260 | 735 | 750 | 0.72 | 42.6          |
| 12   | 0.0013 | 13  | 260 | 735 | 740 | 0.72 | 39.7          |
| 13   | 0.0009 | 9   | 290 | 735 | 750 | 0.72 | 58.2          |
| 14   | 0.0013 | 9   | 290 | 735 | 740 | 0.72 | 52.5          |
| 15   | 0.0009 | 13  | 290 | 735 | 740 | 0.72 | 45.4          |
| 16   | 0.0013 | 13  | 290 | 735 | 750 | 0.88 | 41.7          |

The analysis of non-replicated fractional factorial designs has been considered an adventure, because it is not possible to estimate the true experimental error and, consequently, the reliability of the conclusions resulting from the data analysis may not be the most desired. This is particularly true when the principle of effects scarcity is not verified, that is, when there are many factors and two-factor interactions with significant effect on response.

A very common practice in the analysis of non-replicated fractional factorials is to employ a Normal or “Half-normal” probability plot. In this type of plot, the values of the effects corresponding to the factors (variables) with low contribution to the (average) value of the response tend to be disposed of along a straight line. The effects of variables with greater influence on the response, called active or significant effects, are away from that straight line. This practice was recently criticized by Lenth [100], who argued that, in some cases, the plot analysis is excessively dependent on the sensitivity and knowledge of the process or product by the analyst so bias in the conclusions taken from the data can be introduced. Many other methods have been proposed and their performance (usefulness) reported in various published works. See Costa et al. [61] and references herein as examples.

If researchers and practitioners decide or have the chance to use software for data analysis, it is important that they have an extreme caution with software features. Some software vendors may claim that software users do not need to worry about the experimental protocol, because they understand the proper experimental protocol; software will generate an appropriate optimal design if the factors, their levels, and the experimental constraints are inputted. They may also claim that software can do all the required data analysis. Accepting in full these claims is risky. Data analysis software is an extremely important tool; however, it requires intelligent use. Fontdecaba et al. [101] studied and evaluated five well-known statistical software packages and stated that all packages use different methods and criteria that deliver different results in the analysis of unreplicated factorial designs. Moreover, they show that some of those methods are clearly incorrect and deliver biased results, so methods selection must be carefully made. This is one of the major difficulties felt by practitioners and, unfortunately, no existing method performs better than all the others. The number and size of variable effects, in addition to
abnormalities in the data, have influence on methods performance. Another difficulty is to identify the journals and access the papers where the methods are reported. In fact, some effective methods are not available in commercial statistical packages and are published in distinct journals, which requires time to look for and financial resources to access the papers.

In the case study presented here, and based on results reported in Ref. [58,59,61] and discussion on Lenth’s paper [100], the methods reported in Ref. [102–104] are reviewed and employed in addition to the popular half-normal probability plot for identifying the factors (location effects—those that influence the mean response) with practical or statistical significance effect on the furnace’s energy efficiency.

The multistage procedure published by Al-Shiha and Yang [102] makes use of the test statistic $L_{m,r}$ (1),

$$L_{m,r} = \frac{\sum_{i=m-r+1}^{m} |c_i|^2 / r}{\sum_{i=1}^{m-r} |c_i|^2 / (m-r)}$$  \hspace{1cm} (1)

where $m$ is the number of contrasts, $r$ is the number of potentially active contrasts, and $c_i$ represents the estimated contrasts obtained from the experimental results. When $L_{m,r}$ is larger than a critical value, denoted by $L_{m,r,a}$, the null hypothesis ($H_0$: No active contrasts exist) is rejected, and one can accept that $r$ contrasts are active at the significance level $\alpha$. The $L_{m,r,a}$ values are available at Al-Shiha and Yang [105].

Benski [104] proposed a normality test coupled with an outlier test to identify the active contrasts. To implement this method, it is necessary to perform a test of normality, assess its significance, and then identify the active contrasts as follows [106]:

1) Perform the $W'$ test of Normality,

$$W' = \left( \frac{\sum_{i=1}^{m} z_i^2}{\sum_{i=1}^{m} z_i^2 + m} \right)$$  \hspace{1cm} (2)

where $c$ is the average of the ordered effects or contrasts, $c_i$ and $z_i$ are the expected standard normal order statistics for a sample of size $m$. The $z_i$ can be approximated by

$$z_i = \phi^{-1}(p_i)$$  \hspace{1cm} (3)

where $\phi^{-1}$ is the inverse normal distribution, and $p_i = (i - a) / (m - 2a + 1)$ with

$$a = \begin{cases} 0.275499 + 0.072884(ln(m))^{0.41148}, & 1 < i < m \\ 0.205146 + 0.1314965(ln(m))^{0.226701}, & i = 1, m \end{cases}$$  \hspace{1cm} (4)

2) Calculate the significance level ($P_1$) of $W'$ test

$$P_1 = \exp(C)$$  \hspace{1cm} (5)

where

$$C = \frac{(W' - A) / B + 0.0486128}{0.0276039} - \ln(100)$$  \hspace{1cm} (6)

The parameter $A = 1.031918 - 0.183573(0.1m)^{-0.5447402}$ and $B = -0.5084706 + 2.076782(0.1m)^{-0.4005993}$. If $P_1$ is not small, go to point four of this procedure. If $P_1$ is small ($P_1 < 0.05$), calculate the significance level $P_2$ of the outlier test ($d_F$) for any data point outside the interval $[-2d_F, +2d_F]$, where $d_F = F_U - F_L$, is the interquartile range, and $F_L$ and $F_U$ are the first and the third quartiles of the ($c_i$) contrasts. Under normality, $P_2$ can be estimated as

$$P_2 \approx 0.00698 + (0.4 / m)$$  \hspace{1cm} (7)
Calculate $P_C = 2x\ln(1/(P_1 x P_2))$ and respective significance level, assuming $P_C$ follows a chi-square distribution with four degrees of freedom. If the combined test is rejected at the significance level associated to $P_C$, the largest contrast (in absolute value) can be declared active, removed from the list of contrasts and steps 1–4 repeated with the remaining contrasts.

Stop; consider active the contrasts removed in point four (if applicable). Notice that the confidence in claiming that significant effects exist is enhanced when (1–Pc) is closer to 1 than (1–P1).

When the Dong’s method [101] is employed, a contrast $c_i$ is declared active if

$$|c_i| > t_{\gamma,m_{inactive}} \times S_{Dong}$$

where \( \gamma = \left(1 + 0.98^{1/m}\right)/2, S_{Dong} \) is an estimate of the standard error defined by

$$S_{Dong} = \sqrt{\frac{1}{m_{inactive}} \sum_{|c_i|<2.5S_0} c_i^2}$$

and \( m_{inactive} \) is the number of inactive contrasts, among the \( m \) contrasts, characterized by \( |c_i| \leq 2.5S_0 \).

5.2. Results Analysis and Discussion

The graphical representation in a half-normal probability plot of the experimental results listed in Table 3, as shown in Figure 1, allows to assume that the variables with the greatest effect on the response are the percentage of oxygen in the combustion gases (v2), the fuel flow in the burners (v1), and the combustion air temperature (v3). The billet emissivity (v6) is not considered a variable with relevant effect on the response, such as Figure 1 suggests, because its effect is aliased with (or is achieved from the combination of) the effects of three variables ($v1 \times v2 \times v4$), namely with the effect of $v1$ and $v2$, which are the most influential variables in the furnace’s efficiency.

![Half-Normal Probability Plot](image)

Figure 1. Half-Normal Probability Plot. Factors and Interactions effects.

The results from Al-Shiha and Yang’s method corroborate the interpretation made to Figure 1. Only the variables $v2$, $v1$, and $v3$ are statistically significant (their effects are active) at a significance level $\alpha = 1\%$, with $L_{15,3} = 3095.3 > L_{15,3,1\%} = 19.8$; $L_{14,2} = 119.9 > L_{14,2,1\%} = 19.2$; $L_{13,1} = 40.5 > L_{13,1,1\%} = 19.9$. Dong’s method, with $S_{Dong} = 0.0021$, $m = 15$, and $m_{inactive} = 12$, confirms that $v2$, $v1$, and $v3$ are statistically significant. Benskt’s method also identifies $v2$, $v1$, and $v3$ as statistically significant variables. Thus, one can feel quite confident about the validity of this solution, and to assume that the remaining variables are not influential or statistically significant (see Table 4).
Table 4. Benski’s method results.

| Variable | W'   | P_i  | ±2d_F | P_c  |
|----------|------|------|-------|------|
| v2       | 0.5177 | 8.5 × 10⁻⁷ | 0.0077 | 34.75 |
| v1       | 0.6087 | 1.7 × 10⁻⁶ | 0.0073 | 33.27 |
| v3       | 0.6961 | 3.3 × 10⁻⁶ | 0.0081 | 31.77 |
| Others   | 0.8445 | 6.7 × 10⁻⁶ | 0.0036 | 30.24 |

To obtain indication of the best level for these three variables, a classical procedure was used [107]: the mean values of each variable were calculated at the high and low levels separately, and the highest value of this mean was selected so that the energy efficiency of the furnace is the highest possible. In this case, the efficiency will be higher when v2 is at low level, v1 is at low level, v3 is at high level, and all other variables are kept at low level for technical–economic reasons.

To validate the data analysis, a confirmatory experience was performed using the computer model, and an efficiency value equal to 57.8% was obtained. This result suggests a rigorous measurement of the three aforementioned variables to adequately quantify the energy efficiency of the furnace under study.

6. Conclusions

Energy efficiency is still today an option for most industrial and service organizations, so more intensified policy efforts and regulations are crucial for ensuring that sufficient progress is made in the coming years at (non)industrial settings. There are still barriers that hinder the adoption of energy-efficiency policies and programs, but recognized and generally accepted engineering and energy management practices, a broad range of energy-efficient technologies, methodologies, and many types of mathematical/statistical tools can be implemented to increase energy savings and optimize energy efficiency at low cost. As an example, energy auditors and those who have the responsibility to eliminate energy waste in the manufacturing sector cannot ignore tools that help them in identifying critical process variables and estimating the energy efficiency more accurately by using more accurate devices and methods to collect data, namely, the Design of Experiments. Trustworthy energy consumption calculations give to facility owners and managers valuable feedback on their energy performance status, helping them to adjust energy conservation measures design or operations to improve savings and achieve greater persistence of savings over time. This is also critical for convincing investors in energy-efficiency projects of the benefit and cost-effectiveness of such investments and to replace or defer supply-side capital investments.

This study points out that DoE is an appropriate tool to identify the critical process or equipment variables and assess the energy consumption and efficiency of large and complex equipment, but it is not limited thereto. To try out many settings of an input variable in a “what if” scenario or testing many variables simultaneously in an unstructured way to achieve energy savings are counterproductive practices. In fact, the operation, energy-saving, and energy-efficiency maximization of systems supported on users’ empirical knowledge is an unsustainable practice. To appropriately plan and run statistically designed experiments is a more structured, faster, cheaper, and reliable practice for that purpose at the systems design and manufacturing stages. Even when it is not possible to perform experiments in an industrial environment, due to technical, economic, lack of resources, or other acceptable reasons, running statistically designed experiments in computer models and analyzing the results with tested and validated methods is an efficient approach for maximizing the systems performance.

The trend in technology favors a shorter product development cycle and a quicker reaction to market opportunities, so shorter-duration non-sequential experiments will become more popular in engineering. However, practitioners cannot ignore that valid and practical conclusions drawn from experimental studies depend to a large extent on which data were collected and the way how the data
were collected. This study shows that fractional factorial designs and easy-to-implement methods for screening are alternatives to more sophisticated approaches for designing and analyzing results from computer models when the number of input variables is small. Al-Shiha and Yang’s, Dong’s, and Benski’s methods are structurally different, efficient, and easy-to-implement in Microsoft Excel. Moreover, they can supplement or replace the popular half-normal probability plot. In this case study, all the analysis methods lead to the same conclusion so one can assume their result confidently.

In this study, a computer model of a billets reheating furnace was used, and it was possible to conclude, only with 16 out of 64 experiments, that the percentage of O₂ in the combustion gases (v2) is the variable with the greatest influence on the furnace’s energy efficiency. Its effect is 2.5 times higher than that of fuel flow in the burners (v1) and 6 times higher than that of combustion air temperature (v3). The monitoring and measurement as accurate as possible of the v2 value is determinant in the furnace’s energy-efficiency evaluation, but a similar approach is required to variables v1 and v3. Results analysis provides evidence that v2 and v1 should be set at the low level, while v3 should be set at the high level. All the other variables must be set at the low level. For future work, it is suggested to validate the presented results in other furnace zones or, preferably, in a complete furnace model. In addition to process parameter optimizations, implementing an energy management system, as well as other tools to reduce energy consumption, are also recommended since they have the potential for improvements in the medium–long term.

Author Contributions: Both authors participated in all the development phases of this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work and APC were funding by the AuditF project Lisboa-01-0247-FEDER-017980.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. IEA. Energy Efficiency 2018—Analysis and Outlooks to 2040; IEA Publications: Paris, France, 2018; Available online: https://www.iea.org/efficiency2018/ (accessed on 12 June 2019).
2. IEA. Energy Efficiency 2019; IEA Publications: Paris, France, 2019; Available online: https://www.iea.org/efficiency2019/ (accessed on 12 June 2019).
3. European Union. Regulation (EU) 2018/1999. Official Journal of the European Union, 21/12/2018, L 328/1-328/77. Available online: https://eur-lex.europa.eu/oj-direct-access.html (accessed on 6 June 2019).
4. European Union. Directive (EU) 2018/2002. Official Journal of the European Union, 21/12/2018, L 328/210-328/230. Available online: https://eur-lex.europa.eu/oj-direct-access.html (accessed on 6 June 2019).
5. European Union. Directive (EU) 2012/27. Official Journal of the European Union, 14.11.2012, L 315/1-315/56. Available online: https://eur-lex.europa.eu/oj-direct-access.html (accessed on 6 June 2019).
6. European Commission. The European Green Deal. COM (2019) 640 final, Brussels, 11.12.2019. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52019DC0640 (accessed on 31 March 2020).
7. ECEEE. Twelve Strategies to Step up Global Energy Efficiency. 2019. Available online: https://www.eceee.org/static/media/uploads/site-2/policy-areas/globalaction/12-eestrategiesfor-ieahighlevel-enrev.pdf (accessed on 15 September 2020).
8. ACEEE. Call to Action for Energy Efficiency. 2020. Available online: https://www.aceee.org/sites/default/files/pdfs/call-to-action.pdf (accessed on 15 September 2020).
9. IEA. Recommendations of the Global Commission for Urgent Action on Energy Efficiency; IEA: Paris, France, 2020; Available online: https://www.iea.org/reports/recommendations-of-the-global-commission-for-urgent-action-on-energy-efficiency (accessed on 22 June 2020).
10. Tsemekidi-Tzeiranaki, S.; Economidou, M.; Cuniberti, B.; Bertoldi, P. Analysis of the Annual Reports 2019 under the Energy Efficiency Directive; EUR 30160 EN; Publications Office of the European Union: Luxembourg, 2020; Available online: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120194/synthesis_report_final.pdf (accessed on 16 March 2020).
11. Fawcett, T.; Rosenow, J.; Bertoldi, P. Energy efficiency obligation schemes: Their future in the EU. Energy Eff. 2019, 12, 57–71. [CrossRef]
12. Cornelis, E. History and prospect of voluntary agreements on industrial energy efficiency in Europe. *Energy Policy* **2019**, *132*, 567–582. [CrossRef]

13. Cai, W.; Liu, C.; Lai, K.; Li, L.; Cunha, J.; Hu, L. Energy performance certification in mechanical manufacturing industry: A review and analysis. *Energy Convers. Manag.* **2019**, *186*, 415–432. [CrossRef]

14. European Union. Guide on Good Practice in Energy Efficiency for Central and South Eastern Europe. Available online: https://op.europa.eu/en/publication-detail/-/publication/cf3b622e-22ae-11e9-8d04-01aa75ed71a1/language-en/format-PDF/source-search (accessed on 3 May 2020).

15. European Union. Good Practice in Energy Efficiency—For a Sustainable, Safer and more Competitive Europe. Available online: https://op.europa.eu/en/publication-detail/-/publication/54b16aa-2982-11e7-ab65-01aa75ed71a1/language-en/format-PDF (accessed on 3 May 2020).

16. EU-MERCI. European Industrial Energy Efficiency Good Practices Platform. 2017. Available online: http://www.eumerci-portal.eu/web/guest/home (accessed on 3 August 2020).

17. IEA. *Capturing the Multiple Benefits of Energy Efficiency*; IEA Publications: Paris, France, 2015; Available online: https://webstore.iea.org/capturing-the-multiple-benefits-of-energy-efficiency (accessed on 12 June 2019).

18. US-Department of Energy. *Barriers to Industrial Energy Efficiency—Report to Congress*; United States Department of Energy: Washington, DC, USA, 2015.

19. European Commission. Reference Document on Best Available Techniques for Energy Efficiency. 2009. Available online: https://eippcb.jrc.ec.europa.eu/reference/BREF/ENE_Adopted_02-2009.pdf (accessed on 15 July 2020).

20. European Commission. 2018 Assessment of the Progress Made by Member States towards the National Energy Efficiency Targets for 2020 and towards the Implementation of the Energy Efficiency Directive as Required by Article 24 of the Energy Efficiency Directive 2012/27/EU. COM (2019) 224 Final, Brussels, 9.4.2019. Available online: https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-224-F1-EN-MAIN-PART-1.PDF (accessed on 28 July 2020).

21. European Union. EU Energy in Figures—Statistical Pocketbook. EU Publications. ISSN 263-247X. Available online: https://ec.europa.eu/energy/en/data-analysis/energy-statistical-pocketbook (accessed on 28 July 2020). [CrossRef]

22. Palm, J.; Backman, F. Energy efficiency in SMEs: Overcoming the communication barrier. *Energy Effic.* **2020**, *13*, 809–821. [CrossRef]

23. Santos, H.; Almeida, P.; Cardoso, R. Fuel Costs Minimization on a Steel Billet Reheating Furnace Using Genetic Algorithms. Available online: https://www.hindawi.com/journals/mse/2017/2731902 (accessed on 25 April 2020). [CrossRef]

24. Bergquist, B. Some ideas on why factorial designs are seldom used for full-scale experiments in continuous production processes. *TQM J.* **2015**, *26*, 1242–1254. [CrossRef]

25. Sheil, J.; Hale, D. The ‘Democratisation’ of Statistics: Parastatistics in Manufacturing. *Qual. Reliab. Eng. Int.* **2012**, *28*, 524–534. [CrossRef]

26. Saxena, V.; Kumar, N.; Saxena, V.K. Multi-objective optimization of modified nanofluid fuel blends at different TiO2 nanoparticle concentration in diesel engine: Experimental assessment and modelling. *Appl. Energy* **2019**, *248*, 330–353. [CrossRef]

27. Cheng, C.-C.; Liu, K.-W. Optimizing energy savings of the injection molding process by using a cloud energy management system. *Energy Effic.* **2018**, *11*, 415–426. [CrossRef]

28. Zin, A.; Moradi, M. An experimental investigation of price elasticity in electricity markets using a response surface methodology. *Energy Effic.* **2019**, *12*, 667–680. [CrossRef]

29. Bellotti, D.; Cassettari, L.; Mosca, M.; Magistri, L. RSM approach for stochastic sensitivity analysis of the economic sustainability of a methanol production plant using renewable energy sources. *J. Clean. Prod.* **2019**, *240*, 117947. [CrossRef]

30. Liang, X.; Zhao, B.; Zhang, F.; Liu, Q. Compact research for maritime selective catalytic reduction reactor based on response surface methodology. *Appl. Energy* **2019**, *254*, 113702. [CrossRef]

31. Khalid, A.; Yaakob, Z.; Abdullah, S.; Takriff, M. Analysis of the elemental composition and uptake mechanism of Chlorella sorokiniana for nutrient removal in agricultural wastewater under optimized response surface methodology conditions. *J. Clean. Prod.* **2019**, *210*, 673–686. [CrossRef]

32. Ji, Q.; Li, C.; Zhu, D.; Jin, Y.; He, J. Structural design optimization of moving component in CNC machine tool for energy saving. *J. Clean. Prod.* **2020**, *246*, 118976. [CrossRef]
33. Rahimipetroudi, I.; Rashid, K.; Yang, J.; Dong, S. Use of response surface methodology to optimize NOx emissions and efficiency of W-type regenerative radiant tube burner under plasma-assisted combustion. J. Clean. Prod. 2020, 244, 118626. [CrossRef]

34. Tang, S.-Z.; Wang, F.-L.; He, Y.-L.; Yu, Y.; Tong, Z.-X. Parametric optimization of H-type finned tube with longitudinal vortex generators by response surface model and genetic algorithm. Appl. Energy 2019, 239, 908–918. [CrossRef]

35. Anthony, J.; Coleman, S.; Montgomery, D.; Anderson, M.; Silvestrini, R. Design of experiments for non-manufacturing processes: Benefits, challenges and some examples. Proc. Inst. Mech. Eng. Part B J. Eng. Manuf. 2011, 225, 2078–2087. [CrossRef]

36. Ree, S.; Park, Y.; Yoo, H. A study on Education Quality using the Taguchi Method. Total Qual. Manag. Bus. Excell. 2014, 25, 935–943. [CrossRef]

37. Anthony, S.; Anthony, J. Academic leadership and Lean Six Sigma: A novel approach to systematic literature review using design of experiments. Int. J. Qual. Reliab. Manag. 2016, 33, 1002–1018. [CrossRef]

38. Costa, N.; Pires, A.; Ribeiro, C. Guidelines to help practitioners of design of experiments. TQM Mag. 2006, 18, 386–399. [CrossRef]

39. Simpson, J.; Listak, C.; Hutto, G. Guidelines for planning and evidence for assessing a well-designed experiment. Qual. Eng. 2013, 25, 333–355. [CrossRef]

40. Tanco, M.; Viles, E.; Álvarez, M.; Ilzarbe, L. Implementation of Design of Experiments projects in industry. Appl. Stoch. Models Bus. Ind. 2009, 25, 478–505. [CrossRef]

41. Freeman, L.; Ryan, A.; Kensler, J.; Dickinson, R.; Vining, G. A tutorial on the planning of experiments. Qual. Eng. 2013, 25, 315–332. [CrossRef]

42. Lundkvist, P.; Bergquist, B.; Vanhatalo, E. Statistical methods—Still ignored? The testimony of Swedish alumni. Total Qual. Manag. Bus. Excell. 2018, 31, 245–262. [CrossRef]

43. Costa, N. Design of Experiments—Overcome hindrances and bad practices. TQM J. 2019, 31, 772–789. [CrossRef]

44. Johnson, R.; Hutto, G.; Simpson, J.; Montgomery, D. Designed Experiments for the Defense Community. Qual. Eng. 2012, 24, 60–79. [CrossRef]

45. Zhang, J.; Liang, Y.; Harpalani, S. Optimization of methane production from bituminous coal through biogasification. Appl. Energy 2016, 183, 31–42. [CrossRef]

46. Lv, S.; He, Z.; Valeria, A.; Mirabile, Y.; Vining, G. Process optimization using sequential design of experiment: A case study. Qual. Eng. 2019, 31, 473–483. [CrossRef]

47. Jones, B.; Nachtsheim, C. Effective Design Based Model Selection for Definitive Screening Designs. Technometrics 2017, 59, 319–329. [CrossRef]

48. Anderson-Cook, C.; Borror, C.; Montgomery, D. Response surface design evaluation and comparison (with discussion). J. Stat. Plan. Inference 2009, 139, 629–674. [CrossRef]

49. Anderson-Cook, C.; Lu, L. Best Bang for Your Buck—Part 1. Qual. Prog. 2016, 49, 45–47.

50. Anderson-Cook, C.; Lu, L. Best Bang for Your Buck—Part 2. Qual. Prog. 2016, 49, 50–52.

51. Tyssedal, J.; Chaudhry, M. The choice of screening design. Appl. Stoch. Models Bus. Ind. 2017, 33, 662–673. [CrossRef]

52. Jensen, W. Open Problems and Issues in Optimal Design. Qual. Eng. 2018, 30, 583–593. [CrossRef]

53. Santos, C.; Rato, T.; Reis, M. Design of Experiments: A comparison study from the non-expert user’s perspective. J. Chemom. 2019, 33, e3087. [CrossRef]

54. Mylona, K.; Gilmour, S.; Goos, P. Optimal Blocked and Split-Plot Designs Ensuring Precise Pure-Error Estimation of the Variance Components. Technometrics 2019, 62, 57–70. [CrossRef]

55. Shi, C.; Tang, B. Supersaturated designs robust to two-factor interactions. J. Stat. Plan. Inference 2019, 200, 119–128. [CrossRef]

56. Lu, L.; Anderson-Cook, C.; Lin, D. Optimal designed experiments using a Pareto front search for focused preference of multiple objectives. Comput. Stat. Data Anal. 2014, 71, 1178–1192. [CrossRef]

57. Ockuly, R.; Weese, M.; Smucker, B.; Edwards, D.; Chang, L. Response surface experiments: A meta-analysis. Chemom. Intell. Lab. Syst. 2017, 164, 64–75. [CrossRef]

58. Hamada, M.; Balakrishnan, N. Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals. Stat. Sin. 1998, 8, 1–41.
59. Chen, Y.; Kunert, J. A new quantitative method for analysing unreplicated factorial designs. *Biom. J.* 2004, 46, 125–140. [CrossRef]

60. Lawson, J. Comparison of conditional main effects analysis to the analysis of follow-up experiments for separating confounded two-factor interaction effects in $2^{k-p}$ fractional factorial experiments. *Qual. Reliab. Eng. Int.* 2020, 36, 1454–1472. [CrossRef]

61. Costa, N.; Pires, R.; Fontes, P. On the selection of significant variables from unreplicated factorial designs. *Int. J. Product. Qual. Manag.* 2013, 12, 161–180. [CrossRef]

62. Tyssedal, J.; Hussain, S. Factor screening in nonregular two-level designs based on projection-based variable selection. *J. Appl. Stat.* 2016, 43, 490–508. [CrossRef]

63. Aguirre, V. Bayesian analysis of definitive screening designs when the response is nonnormal. *Appl. Stoch. Models Bus. Ind.* 2016, 32, 440–452. [CrossRef]

64. Drosou, K.; Koukouvinos, C.; Lappa, A. Screening Active Effects in Supersaturated Designs with Binary Response. *Qual. Reliab. Eng. Int.* 2017, 33, 1475–1483. [CrossRef]

65. Hibbert, D. Experimental design in chromatography: A tutorial. *J. Chromatogr. B* 2012, 9, 2–13. [CrossRef]

66. Montgomery, D. *Design and Analysis of Experiments*, 8th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013.

67. Kleijnen, J. Regression and Kriging metamodels with their experimental designs in simulation: A review. *Eur. J. Oper. Res.* 2017, 256, 1–16. [CrossRef]

68. Mukhopadhyay, S.; Khuri, A. Optimization in a multivariate generalized linear model situation. *Comput. Stat. Data Anal.* 2008, 52, 4625–4634. [CrossRef]

69. Peterson, J.; Miró-Quesada, G.; Del Castillo, E. A bayesian reliability approach to multiple response optimization with seemingly unrelated regression models. *Qual. Technol. Quant. Manag.* 2009, 6, 353–369. [CrossRef]

70. Costa, N.; Lourenço, J.; Pereira, Z. Responses modeling and optimization criteria impact on the optimization of multiple quality characteristics. *Comput. Ind. Eng.* 2012, 62, 927–935. [CrossRef]

71. Costa, N.; Lourenço, J.; Pereira, Z. Desirability function approach: A review and performance evaluation in adverse conditions. *Chemom. Intell. Lab. Syst.* 2011, 107, 234–244. [CrossRef]

72. Murphy, T.; Tsui, K.; Allen, J. A review of robust design methods for multiple responses. *Res. Eng. Des.* 2005, 15, 201–215. [CrossRef]

73. Salmasnia, A.; Bashiri, M. A new desirability function-based method for correlated multiple response optimization. *Int. J. Adv. Manuf. Technol.* 2015, 76, 1047–1062. [CrossRef]

74. Soh, W.; Kim, H.; Yum, B.-J. A multivariate loss function approach to robust design of systems with multiple performance characteristics. *Qual. Reliab. Eng. Int.* 2016, 32, 2685–2700. [CrossRef]

75. Ardakani, M.; Wulff, S. An Overview of Optimization Formulations for Multiresponse Surface Problems. *Qual. Reliab. Eng. Int.* 2013, 29, 3–16. [CrossRef]

76. Costa, N.; Lourenço, J. Multiresponse problems: Desirability and other optimization approaches. *J. Chemom.* 2016, 30, 702–714. [CrossRef]

77. Kim, K.; Lin, D. Dual response surface optimization: A fuzzy modelling approach. *J. Qual. Technol.* 1998, 30, 1–10. [CrossRef]

78. Costa, N.; Pereira, Z. Multiple response optimization: A global criterion–based method. *J. Chemom.* 2010, 24, 333–342. [CrossRef]

79. Costa, N.; Lourenço, J. Reproducibility of nondominated solutions. *Chemom. Intell. Lab. Syst.* 2017, 168, 1–9. [CrossRef]

80. Amaran, S.; Sahinidis, N.; Sharda, B.; Bury, S. Simulation optimization: A review of algorithms and applications. *Ann. Oper. Res.* 2016, 240, 351–380. [CrossRef]

81. Cui, Y.; Geng, Z.; Zhu, Q.; Han, Y. Review: Multi-objective optimization methods and application in energy saving. *Energy* 2017, 125, 681–704. [CrossRef]

82. Garud, S.; Karimi, I.; Kraft, M. Design of computer experiments: A review. *Comput. Chem. Eng.* 2017, 106, 71–95. [CrossRef]

83. Parker, P. Discussion of “Space-filling designs for computer experiments: A review.” *Qual. Eng.* 2016, 28, 39–41. [CrossRef]

84. Kleijnen, J. Design and Analysis of Simulation Experiments: Tutorial. In *Advances in Modeling and Simulation. Simulation Foundations, Methods and Applications*; Tolk, A., Fowler, J., Shao, G., Yücesan, E., Eds.; Springer: Cham, Switzerland, 2017; Chapter 8.
85. Bhosekar, A.; Ierapetritou, M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comput. Chem. Eng.* 2018, 108, 250–267. [CrossRef]
86. Júnior, W.; Montevechi, J.; Miranda, R.; Campos, A. Discrete simulation-based optimization methods for industrial engineering problems: A systematic literature review. *Comput. Ind. Eng.* 2019, 128, 526–540. [CrossRef]
87. Woods, D.; Lewis, S. Design of Experiments for Screening. In *Handbook of Uncertainty Quantification*; Ghanem, R., Higdon, D., Owhadi, H., Eds.; Springer: Cham, Switzerland, 2017.
88. Liu, L.; Ma, Y.; Byun, J.-H. Robust sequential bifurcation for simulation factor screening under data contamination. *Comput. Ind. Eng.* 2019, 129, 102–112. [CrossRef]
89. Draguljić, D.; Woods, D.; Dean, A.; Lewis, S.; Vine, A. Screening Strategies in the Presence of Interactions. *Technometrics* 2014, 56, 1–10. [CrossRef]
90. Georgiou, S. Supersaturated designs: A review of their construction and analysis. *J. Stat. Plan. Inference* 2014, 144, 92–109. [CrossRef]
91. Joseph, V.; Gu, L.; Shan, B.; Myers, W. Space-Filling Designs for Robustness Experiments. *Technometrics* 2019, 61, 24–37. [CrossRef]
92. Kong, X.; Ai, M.; Tsui, K. Design for Sequential Follow-Up Experiments in Computer Emulations. *Technometrics* 2018, 60, 61–69. [CrossRef]
93. Kleijnen, J.; van Beers, W. *Statistical Tests for Cross-Validation of Kriging Models*; CentER Discussion Paper; CentER, Center for Economic Research: Tilburg, The Netherlands, 2019; Volume 2019–2022.
94. Vining, G.; Kulaahi, M.; Pedersen, S. Recent Advances and Future Directions for Quality Engineering. *Qual. Reliab. Eng. Int.* 2016, 32, 863–875. [CrossRef]
95. Hoerl, R.; Snee, R. Statistical Engineering: An Idea Whose Time Has Come? *Am. Stat.* 2017, 71, 209–219. [CrossRef]
96. Pistone, G.; Vicario, G. Comparing and generating Latin Hypercube designs in Kriging models. *ASIA Adv. Stat. Anal.* 2010, 94, 353–366. [CrossRef]
97. Joseph, V. Space-filling designs for computer experiments: A review. *Qual. Eng.* 2016, 28, 28–35. [CrossRef]
98. Lekivetz, R.; Jones, B. Fast flexible space-filling designs with nominal factors for nonrectangular regions. *Qual. Reliab. Eng. Int.* 2019, 35, 677–684. [CrossRef]
99. Ankenman, B.; Cheng, R.; Lewis, S. Screening for Dispersion Effects by Sequential Bifurcation. *ACM Trans. Modeling Comput. Simul.* 2014, 25, 2–27. [CrossRef]
100. Lenth, R. The Case Against Normal Plots of Effects. *J. Qual. Technol.* 2015, 47, 91–97. [CrossRef]
101. Fontdecaba, S.; Grima, P.; Tort-Martorell, X. Analyzing DOE With Statistical Software Packages: Controversies and Proposals. *Am. Stat.* 2014, 68, 205–211. [CrossRef]
102. Al-Shiha, A.; Yang, S. Critical Values and Some Properties of a New Test Statistic. *Biom. J.* 2000, 42, 605–616. [CrossRef]
103. Dong, F. On the identification of active contrasts in unreplicated fractional factorials. *Stat. Sin.* 1993, 3, 209–217.
104. Benski, C. Use of a Normality Test to Identify Significant Effects in Factorial Designs. *J. Qual. Technol.* 1989, 21, 174–178. [CrossRef]
105. Al-Shiha, A.; Yang, S. A multistage procedure for analyzing unreplicated factorial experiments. *Biom. J.* 1999, 41, 659–670. [CrossRef]
106. Costa, N.; Pereira, Z. Decision-Making in the Analysis of Unreplicated Factorial Designs. *Qual. Eng.* 2007, 19, 215–225. [CrossRef]
107. Myers, R.; Montgomery, D.; Anderson-Cook, C. *Response Surface Methodology*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NY, USA, 2009; p. 282.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).