Energy & Buildings 250 (2021) 111298

Contents lists available at ScienceDirect

Energy & Buildings

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

Designing the design of experiments (DOE) – An investigation on the influence of different factorial designs on the characterization of complex systems

Aleksandar Jankovic, Gaurav Chaudhary, Francesco Goia *

Department of Architecture and Technology, Norwegian University of Science and Technology, NTNU, Trondheim, Norway

ARTICLE INFO

Article history:
Received 16 February 2021
Revised 25 June 2021
Accepted 20 July 2021
Available online 22 July 2021

Keywords:
Design of experiments (DOE)
Characterization
Double-skin facades
Taguchi design
Response surface methodology (RSM)
Definitive screening design (DSD)
Full factorial design (FFD)

ABSTRACT

Although a general set of guidelines and procedures for performing the design of experiments (DOE) exists, the literature lacks a recommended course of action for finding and selecting the optimal design of experiments among a large range of possible designs. This research tries to fill this gap by comprehensively testing more than thirty different DOEs through nearly half a million simulated experimental runs. The performance of various DOEs in the characterization of the thermal behaviour of a double skin façade (DSF) is assessed by comparing the outcomes of the different designs and using the full factorial design (FFD) as the *ground truth*. Besides the finding for the specific case study used in this investigation, this research allowed us to obtain some broad conclusions on the behaviour of different DOEs, which are summarized and translated into recommendations and a general decision tree chart for selecting the suitable DOE(s). The outcomes of this study help researchers and designers to apply DOEs that consider the extent of nonlinearity and interaction of factors in the investigated process in order to select the most successful and the most efficient designs for the specific process characterization.

1. Introduction

Although developed primarily for agricultural purposes by British statistician Sir Ronald Fischer in the 1920s [1], the *design of experiment* (DOE) as a statistical method has been widely applied in different fields of science and industry, especially to support the design, development, and optimization of products and processes [2]. The *design of experiments* includes a series of applied statistics tools used to systematically classify and quantify cause-and-effect relations between variables and outputs in the studied process or phenomenon, which may result (if that is the aim) in finding the settings and conditions under which the process becomes optimized.

Well-established, general guidelines and procedures are available to support the implementation of DOE methods [3]. These steps include defining the objectives and response variables, determining factors, levels, experimental design type and experiment execution. Variables in the DOE such as the number of factors, levels, and the logic to select them usually depend on the type of investigation (screening, characterization, or optimization), the process type and the available resources. However, there is a multitude of different DOEs that can theoretically match the type of investigation, and therefore it is not straightforward to identify which design provides the best possible insight using the least resources. In general terms, a good experimental design ensures the validity of the given insight. However, good and excellent DOEs differ in efficiency, i.e., the ratio between extracted information from the examined process and the invested resources. Unfortunately, there is very little information in the literature that investigates and explains what types of procedures and steps need to be taken to find the optimal DOE among all the possible alternative options which have been developed and proposed.

The knowledge gap on how to select optimal DOEs became evident in our planning of the experiments in a controlled environment on a mock-up of an advanced fenestration system based on a double-skin façade (DSF) concept. This lack of recommended procedures in the literature, not only for the specific case of building envelope systems or even buildings but also in more general terms, motivated us to plunge into the search for answers to the following research questions: *to what extent and why do different design of experiments give different results?* And, further: *what are the recommended steps to be followed to find optimal design(s) of experiments for given research*?

We tackled this problem through a case study, where an advanced façade system was examined. The aim was to identify

* Corresponding author.
E-mail address: francesco.goia@ntnu.no (F. Goia).

https://doi.org/10.1016/j.enbuild.2021.111298
0378-7788/

general guidelines that can facilitate finding one or more optimal designs of an experiment for different types of problems or processes, thus going beyond the specific case used in our investigation. Through the use of building performance simulation, we compared and analyzed a wide variety of fractional factorial designs to find one or more optimal methods for our specific case. More than 400 000 simulation runs were performed to assess 30 different DOEs. The use of simulation tools to support the design of experiments is not a novel concept [5]. However, in this study, we used the possibility given by simulating a huge number of cases as a strategy to explore how it is possible to find optimal DOEs. As a secondary effect of this research strategy, we also showed how simulation tools could support the selection of the best DOEs and, at the same time, serve as good preparation for physical experiments.

Nowadays, researchers primarily select the DOE based on the assumed importance of the factors and the desired number of experimental runs [4–6]. If the aim is in-depth characterization, not knowing the nature of the complex process can lead to the wrong selection of experimental design and false conclusions about the importance of different factors, the extent and the type of nonlinearity within the process. A better understanding of what is the optimal or the best DOE(s) is important to assure that such a powerful investigation technique is properly used. Therefore, the results of this study can help researchers who need to find the optimal design for an experiment, using as few resources as possible and discovering as many details as possible about the process.

The remainder of this paper is organized as follows. In the following Section 2, we provide the reader with a background on DOEs and linked statistical tools. In the last part of this section, we also present a brief review of studies where simulation tools were employed to support experimental design. In Section 3 (Methodology), we give an overview of the general flow of the research, with details on: the specific case study, the numerical model implemented in software for building performance simulation, the description of the selected factors, the response variables, tested experimental designs, analysis of variance (ANOVA) and data analysis methodology for comparing various designs. In the fourth section, Results and Discussion, we present the outcome of our investigation of the case study based on the statistical analysis of variance performed on full factorial design (FFD) and its comparison with other DOEs. There, we summarize the performance of the different DOEs from the case study and provide a generalized flow-chart to facilitate selecting the appropriate DOE. Finally, in the fifth section, Conclusions, we recall the main lessons learned from the paper.

2. Background

2.1. Overview of the main DOEs

Each DOE can be seen as being composed of a series of steps: the planning, the execution of the experiment, and the analysis of the collected experimental data using various statistical methods in order to draw valid and objective conclusions [7]. Each DOE starts with selecting the system/process and recognizing the investigation problem. The problem statement leads to establishing the objectives based on which the performance indicator (response variable) needs to be defined. The response variable should represent a quantitative measure of system behavior. As an essential step in the whole process, the factors affecting the performance indicator and how they are discretized, the number of experimental runs, and a suitable array need to be defined in the second stage [8]. The third stage covers the performance of the experiment according to the designed array and collection of data. The last step includes data analysis using statistical tools (ANOVA and associated statistical methods) and interpretation of results, leading to a better understanding of system behavior or its optimization.

In order to examine the impact of several factors and interactions among them on the response quantity [9], experiments need to be performed systematically using factorial experiments (so-called. factorial experimental designs/arrays), where several factors are altered during each experimental run. A factorial experiment whose design consists of all possible combinations of the chosen factors and levels is called **full factorial design (FFD).** Effects of all factors (main effects) and interactions among them are considered in this design [8], making it a potent tool that, compared to other experimental designs, provides the most comprehensive insight into the system's behavior. If all factors k has the same number of levels n, the total number of runs is equal to n^k. By increasing the number of factors and levels, the number of experimental runs grows hugely. For classical experiments, this brings high costs and time consumption. However, due to enormous diversity in combinations, the response quantity variance can be explained, decomposed and attributed to all possible causes, thereby providing in that way an almost-realistic depiction of the process. The nature of the FFD means that its results can be considered good references to discuss other designs' performance in the characterization.

Besides FFD, there is a wide variety of factorial designs, and they differ in the insight they offer. The depth of the characterization depends on the resolution level of the design, which identifies the order of confounding the main effects and their interactions [10]. Resolution designs below III levels are not helpful, because, by definition, I level design consists of only one experimental run, while in II level, main effects are mutually confounded [11]. The most common types are III, IV, and V level designs [12]. Third-level resolution designs assess only the impact of factors, but these main effects are confounded with two-factor interactions. Fourth-level resolution designs consider main effects, and they are not aliased with two-factor interactions, but two-factor interactions are confounded with each other. In fifth-level resolution designs, the main effects are not aliased with each other or two-factor interactions, and two-factor interactions are not aliased with each other. However, higher-order interactions may originate a background noise in lower-order terms.

Selecting the "right" design means identifying the best way to sample the domain of possibilities. There is a wide variety of factorial designs. Some are used to screen out important variables (III resolution), others to characterize processes (IV-V resolution) and a third type to optimize them (≥IV resolution). Some designs, such as definitive screening or designs associated with response surface methodology (RSM), have been derived from factorial designs and can be considered partial factorial designs that include points (runs) that are not covered by standard factorial designs. In the following text, the main characteristics of experimental designs that have been studied in this research are presented. In our study we selected designs that are most often employed for characterization, both in science and industry. For the sake of completeness, we need also to recall that other types of designs, such as reliability, optimal custom, mixture, and split-plot designs can also be adopted [13–16]. However, they are employed either for purpose other than characterization or for the experiments that require special conditions, which is not of particular interest to our research.

The difference between various designs can be visualized, for example, assuming a problem with three factors (A, B, C), each discretized in five levels (a, b, c, d, e), as in Fig. 1.

Definitive screening design (DSD), unlike other screening designs, introduces the third (middle) level for continuous fac-

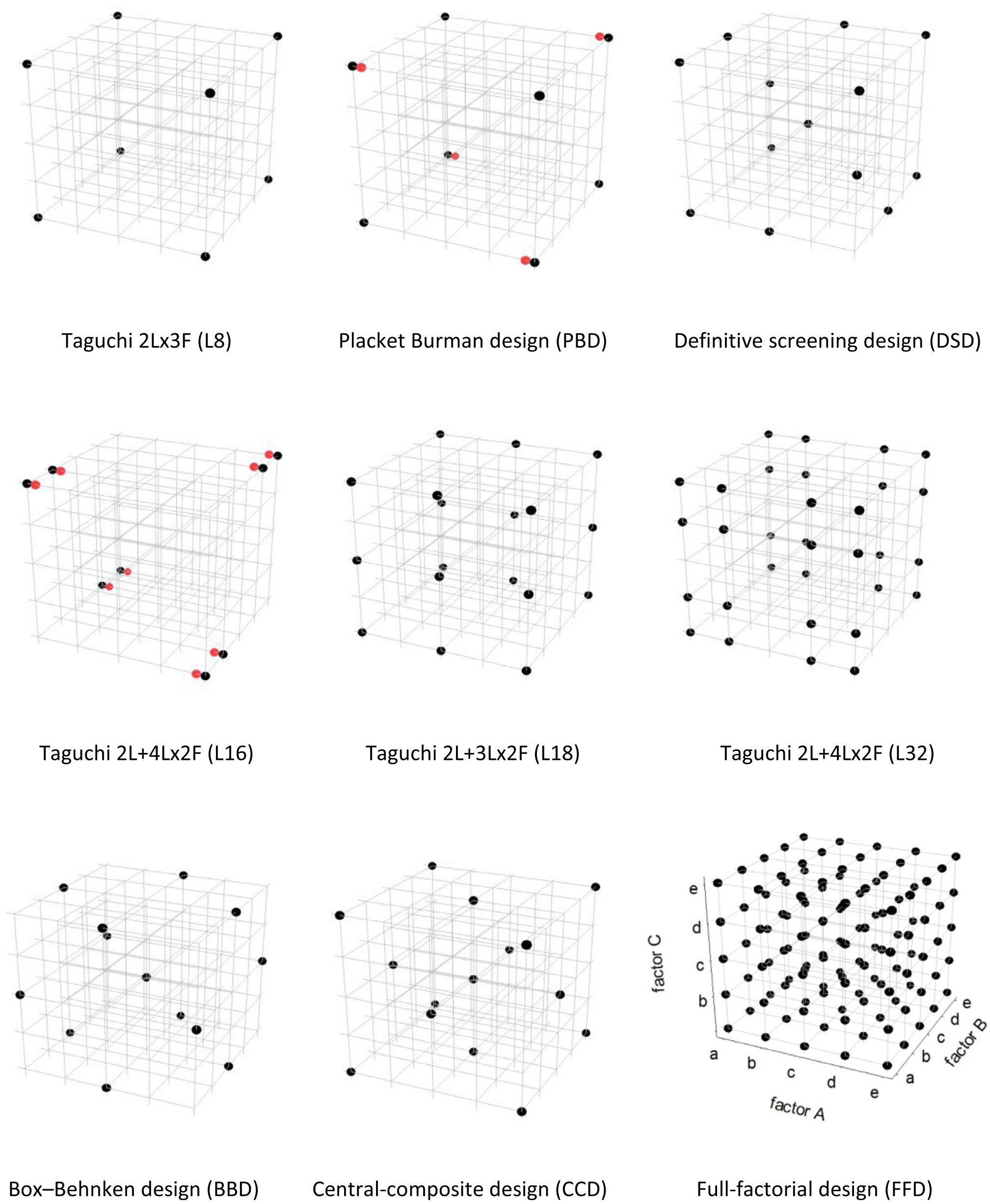

Fig. 1. Various designs for the three-factor model. The intersections of the lines represent the possible combinations of the three factors, while the black dots on some of these intersections represent the combinations to be investigated with that design. Red dots are tests repeated on some combinations to reach the minimum number of tests according to the specific design. In full-factorial design, all the possible combinations are explored. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tors. At the same time, it lowers the number of required experimental runs to *2 k + 1* (k – number of factors) [17], which can be very suitable when some process is affected by a large number of factors [18]. This design can estimate two-factor interactions using a low number of experimental runs, which is impossible in other screening methods. For the characterization of complex processes affected by a high number of factors and interactions, DSD could be a suitable choice for screening and even characterization.

R.L. Plackett and J.P. Burman developed **Placket-Burmann design (PBD)** in 1946. Up to now, it represents one of the most applied screening methods [19,20] for recognition of the most significant factors among a large number of variables. Classic PBD is a third-level resolution design that offers insight into only the main (first-order) effects [21]. However, folding this design increases the number of experimental runs and resolution levels (to 4th), thus enabling the determination of two-factor interactions [22]. Both classic and folded PBD are studied in this research.

Taguchi design (TD). Despite its divided reception in the scientific community [23], Taguchi design, due to its practicality, became the most commonly applied experimental design, both in industry and science [18]. Its strength lies in effective orthogonal matrices where factor levels are distributed in a balanced way, reducing the required number of experimental runs [24]. Most orthogonal arrays are focused only on the main effects, but some designs allow the estimation of specific interactions. Given the practicality of TD, most of the experimental designs examined in this research are Taguchi's. Some of the designs are suitable for screening (two-level TD), while others are more suitable for more in-depth characterization and optimization.

Designs associated with Response Surface Methodology (RSM). These class of designs were developed in 1951 by Box and Wilson [25]. Over time, two main groups of designs have evolved: central-composite and Box-Behnken designs (CCD and BBD). They offer an understanding of the system behavior (reveal a type of connection between factor and response) and its optimization at the same time. CCD is usually applied after narrowing down important factors by some screening methods. It consists of central and axial points beside cube points, allowing insight into a curvature of response and estimation of higher-order effects [26]. BBD is similar to CCD but requires fewer experimental runs and does not contain points at the vertices of the cube (low/high points), which can be very useful for physical experimentation since extreme points are sometimes expensive or difficult to test [27]. For this reason, compared to standard CCD, BBD design contains regions of lower prediction quality [28]. Although RSMs are primarily oriented toward optimizing the system, we decided to include RSMs in the study since they offer an assessment of higher-order terms (quadratic or cubic), which is not possible with the other methods.

2.2. Statistical tools

Parallel to selecting representative runs that can sample the domain of exploration effectively, dedicated methods need to be employed to post-process the results of the experiment. In this way, qualitative and quantitative information on the impact of the different independent variables on the dependent variable is obtained. The analysis of variance (ANOVA) is an analytical and statistical procedure that determines if there are differences between group means in a sample and whether these differences exist only due to randomness or can be attributed to a specific cause. When it comes to DOE, the sample represents a series of experimental runs determined by design (i.e., L32, the sample of 32 experimental runs). In contrast, groups within a sample are a set of runs associated with a particular level/factor/interaction. By comparing group means and variation among them, ANOVA decomposes the total variance and attributes it to all the different causes, thus quantifying the effects of various predictors (factors and interactions) on the dependent variable.

While ANOVA evaluates the impact of various factors, interactions and randomness on the response quantity [29], regression analysis, on the other hand, builds a quantitative relationship between them in the form of a regression equation (model), most often using the least-squares method [30]. The coefficient of determination indicates how well an estimated regression equation explains the variance of the response variable. The RSM allows this regression model to be expanded even further to fit a polynomial function that includes cross-product terms that may be raised up to any power [31]. However, since the structure of the RSM model is only adapted to fit the low-degree Taylor series so it can perform well for the localized region, the class of lower-degree polynomials (up to three) is most often used. Sometimes, built models contain too many predictors, which may be impractical for predictive purposes, if that is the aim. In that case, a delicate balance between the complexity of the model (number of factors) and how well it predicts the response can be found through factor selection procedures. The most commonly adopted strategies to do so are forward entry, backward elimination, and stepwise selection procedures [35]. In an effort to keep this article concise, we cannot report here all the details about the calculation procedures of the ANOVA and the associated statistical methods, but the readers who want to obtain more information can find a synthetic descriptions of these aspects in Appendix A.

2.3. Applications of DOE methods in simulation studies for building physics problems

The use of simulations is becoming a commonly exploited strategy to support the optimization of the experimental runs in various fields of engineering and technology. At the same time, the use of methods for experimental design in computer simulations is an increasingly popular approach. These trends can be considered two sides of the same coin: in both cases the goal is to minimize the resources necessary to understand a certain phenomenon – whether through experiments or through simulations.

Accurate numerical simulations (such as CFD simulations) can be costly, time-consuming and require expensive software and computing equipment, and are therefore limited in their applicability. The utilization of DOE investigation methods provides a rationale to limit the number of accurate numerical simulations without losing the reliability of the overall picture obtained by the simulation runs. Furthermore, applying DOEs to simulations allows one to realize surrogate models that show an acceptable prediction accuracy and can be used in a fast and effortless way to explore large domains. DOE-based high-fidelity simulations can provide researchers with general trends and a high-level understanding of the relationships between variables used to optimize experimental runs. While a relatively large number of factors, levels and "experimental" runs can be oftentimes explored through DOEs implemented in simulations, high-quality experiments can be used to investigate in more depth particular regions or subdomains for a given problem that is identified after a simulation-based pre-screening.

In some cases, simulations could also be the only possibility when physical experimentation is not available or where material saving in terms of labor costs and time is significant [32]. For the purposes of robust parameter design, noise factors can be more easily controlled in simulations compared to experiments, while also experimental variation (noise, error) is absent due to deterministic nature of simulations [33]. In a simulation-based DOE there is no need for randomization, replication of experimental runs, and division of runs into experimental blocks. This is due to the deterministic nature of simulations [34], consistency of input quantities and the ability to control noise (uncontrollable) factors. However, replication can be useful in folding designs where it is needed to increase the design's resolution. However, simulation-based DOE cannot be considered the perfect solution and should be used with caution. In addition to the common problems of physical experiments (selection of factors, levels, and optimal design), the final result of simulation-based DOE also depends on the simulations' accuracy, i.e., the physics captured by the numerical representation of the mathematical model. If the simulation quality is not sufficient and therefore not reliable, then any DOE application makes no sense. Alternatively, using a perfect model/simulation that exactly replicates reality (if it exists) is useless as, in that case, everything is already known about the phenomenon/process. Yet, adopting the DOE approach in building performance simulation can be considered an effective strategy to combine detailed, computationally extensive simulations with the exploration of a large domain as an alternative approach to using other methods like parametric analysis and optimization.

There are, however, relatively few applications of simulation-based DOE in building energy performance research. Sadeghifam et al. [35] examined the influence of various building components and interior temperature on cooling energy loads of buildings in tropical areas using *EnergyPlus* simulations. Full factorial design with four factors and two-level resolution was applied and replicated three times to account for higher-order interactions. ANOVA analysis showed that the main effects were dominant (82.7 %), among which the ceiling had the most substantial influence. Jaffal [36] developed a simple polynomial function using DOE and regression analysis, which estimated the annual energy demand of a low energy building based on its envelope parameters. Simulations were done in *TRNSYS*. A total of 11 parameters with the two-level resolution were used: U-values of vertical walls, floor, windows, roof, the linear transmission coefficient of the thermal bridges, solar heat gains through north, east, south, and west windows, infiltration, and ventilation rates. Several different experimental designs were applied for three different climates (continental, oceanic and Mediterranean) to find optimal function: Taguchi's L12 and L20, a face-centered composite design with 35 runs, and four D-optimum designs (L68, L80, L136, and L160). Overall, the full quadratic model (D160) showed the best performance and lowest error.

Delgram [37] used *EnergyPlus* as a simulation tool and OFAT (one-factor-at-time) as an experimental design to assess the impact of building orientation, optical characteristics and size of windows, overhang system, and envelope thermal characteristics on building energy demand and annual lightning of a typical room for four different climate types in Iran. Variance-based sensitivity analysis showed window size as a prevailing parameter for building energy demand and glazing visible transmittance for annual lightning [37]. The Box–Behnken experimental design with 28 simulations performed in *EnergyPlus* was used to optimize an integrated daylighting and HVAC system [38]. Shen [27] compared full factorial (FFD), central composite rotation (CCRD), optimal (OPD), Box–Behnken (BBD) and space-filling design (SFD) to find the optimal design in the sense of a balance between accuracy and number of experimental runs. The aim was to find a regression model to assess ventilation rate from three factors with two-level resolution: sidewall opening size, outdoor wind speed and direction. For simulations, the CFD numerical model was used, and the SFD proved to be the most accurate, while the BBD was the most efficient.

In conclusion, it is possible to see that DOE methods have been used in combination with building performance simulation to investigate an array of different problems. However, in this area, as in many other areas of engineering and technology, there is a lack of guidelines and suggestions about how different designs can influence the results of an analysis based on the application of these methods. This knowledge-gap is addressed in this study through using simulation as a platform to investigate the implications of employing different DOEs (either in pure experiments or in simulation-based analyses) for a specific case study (a double-skin façade), which we believe can well represent the complexity of many building science problems.

3. Methodology

The research methodology in this study is based on the objectives as listed below.

I) To select a problem and identify the independent variables and the dependent variable(s) which will be the target of the study.
II) To develop a numerical representation of the problem to be able to carry out simulations.
III) To identify a number of DOEs that can be utilized to study the selected problem.
IV) To apply a series of DOEs, and of the full factorial case, by using numerical simulations.
V) To post-process simulation data and execute the data analysis according to the relevant methods and tools on both the DOEs and on the full factorial case.
VI) To assess through comparison what information is obtained and what is the quality of such information from the different DOEs, using as the "ground truth" the results that are obtained by the full factorial design.
VII) To develop a series of guidelines for making the selection of the DOE a more grounded choice, supported by evidence collected both though the review of the literature in the field and the lessons-learned of this study.

The central assumption behind the overall methodological approach of this investigation is that it is possible to know the "true" behavior of a system when such a system is investigated through the full factorial design. In other words, the execution of the full factorial design allows us to "fully" know the impact of every single permutation (i.e. combining all the different levels and all the different independent variables) on the response variable. In this way, we obtain the representation of the entire complexity of the system, and by comparing this with the representations produced by another design (with a lower number of experiments compared to the full factorial design), we can assess how good the latter factorial design is in returning the "real" behavior of the system.

3.1. Case study

The case study selected for this research is a mechanically ventilated double-skin façade (DSF), where both constructional and operational features are selected as independent variables, and the global thermal performance corresponds to the response quantity, as explained more in detail at the end of this section. DSFs are advanced envelope systems where thermophysical, fluid mechanical and optical phenomena that regulate the overall performance are highly dynamic and in constant interaction with different structural elements. Methods based on DOEs seem to be suitable tools to unveil the complex and intertwined interactions between different driving forces in DSFs. The impact of the different construction and operational features on how DSFs behave is still, in large aspects, incompletely evaluated and represents a current knowledge gap when it comes to assessing the performance of these envelope systems [51]. In the context of this study on DOE methods, the intrinsically articulated and multi-domain characteristics of a DSF make it suitable for use as a comprehensive yet relatively simple case technology to investigate the impact of the DOE formulation on the characterization of a certain performance.

The physical–mathematical modeling of such a building system is not a straightforward task [39], and for the sake of this investigation, a model of the DSF was implemented in the simulation environment *EnergyPlus* by making use of the in-build function "airflow window" [40]. This routine allows the users to specify different features of a DSF, together with the usual construction characteristics, and to model the five possible ventilation modes the airflow window can take: indoor air curtain (I-I), outdoor air curtain (O-O), air supply (O-I), exhaust (I-O), and air buffer mode (AB).

The DSF is incorporated in a virtual cubicle where only one surface (where the DSF is modeled) is exposed to the outdoor conditions, while all other surfaces are set as adiabatic and with a fixed temperature equal to indoor air temperature. The nature of this study required systematic experimental procedures in controlled conditions (i.e., fixed values of indoor and outdoor air temperature and solar irradiance). Therefore, dedicated settings were implemented to ensure the right conditions to replicate steady-state simulations.

The model of DSF simulated in this study has a transparent frontal area of dimensions 1.4 m (W) $\times$ 2.8 m (H), with a cavity depth that can range from 20 to 60 cm. A venetian blinds system with 50 mm blinds is positioned at the center of the ventilated cav-

Table 1
Thermal and optical properties of glazing and venetian blinds implemented in EnergyPlus.

	Inner/outer glazing			The front side of the slat			The back side of the slat		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
Total thickness [cm]	1	1	2.3	0.23	0.23	0.23	0.23	0.23	0.23
Solar transmittance [–]	0.83	0.59	0.43	0.03	0.03	0.03	0.03	0.03	0.03
Front side solar reflectance [–]	0.08	0.27	0.26	0.36	0.59	0.83	–	–	–
Back side solar reflectance [–]	0	0	0	–	–	–	0.36	0.59	0.83
Front side IR emissivity [–]	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Back side IR emissivity [–]	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Conductivity [$Wm^{-1}K^{-1}$]	0.3	0.3	0.3	205	205	205	205	205	205

ity between the inner and outer skin. The thermal and optical properties of glazing and venetian blinds of the case study DSF are summarised in Table 1.

When it comes to the independent variables that can influence the global thermal performance of the facade, the following factors were chosen:

- the temperature difference between indoor and outdoor air [°C] (discretized in 5 levels)
- incident solar irradiance on the vertical plan [Wm^{-2}] (discretized in 5 levels)
- slat angle of venetian blinds [°] (discretized in 5 levels)
- airflow rate [m^2s^{-1}] (discretized in 3 levels)
- cavity depth [cm] (discretized in 3 levels)
- optical and thermal properties of the inner and outer glazing [-] (discretized in 3 levels/types), and
- optical properties of front and back surface of blinds [-] (discretized in 3 levels each).

In the simulation runs, the indoor temperature was kept constant at 20 °C while wind influence was excluded from this research. Three types of inner/outer glazing and front/back surface of venetian blinds were taken into consideration, and their characteristics are given, together with information on the different levels for the independent variables, in Table 2.

The dependent variable, i.e., the response quantity in this study, was the global thermal performance of the DSF, which is constituted by the total heat gain density associated with examined DSF element according to the general equation:

$$q_{net} = q_{sol,SW} + q_{sol,LW} + q_{conv} + q_{vent} + q_{ent} \quad (1)$$

where the intensity of transmitted solar irradiance through the element is given by $q_{sol,SW}$, longwave irradiance exchanged between the inner glazing and the interior environment by $q_{sol,LW}$, and the heat flux density transferred by convection between the interior surface of the glazing and indoor environment by q_{conv}; q_{vent} indicates the convective gain or loss due to the airflow that passes through the cavity, and q_{ent} gives the contribution necessary to compensate the convective heat gain or loss due to infiltration to assure the air mass balance in those configurations where mass exchange occurs between inside and outside.

All quantities are normalized per unit area, and they are given with the next equations:

$$q_{sol,SW} = \tau \cdot I \quad (2)$$

$$q_{sol,LW} + q_{conv} = (h_{conv} + h_{rad}) \cdot (t_i - t_{is}) \quad (3)$$

Transmitted solar irradiance depends on the optical properties of glazing and shading (τ – solar transmissivity of the DSF system). Convective and infrared heat exchanged between surface and indoor air is dependent on its temperature difference (t_i-t_{is}) and emissivity of glazing that faces an indoor environment. There are various empirical formulas for convective and radiative heat transfer coefficients (h_c and h_r) implemented in *EnergyPlus*, and the following algorithms were adopted in this study: TARP and DOE-2. Convective heat exchange between indoor air and airflow that passes the cavity and enters the indoor environment is given with q_{vent}, in equation (1):

$$q_{vent} = \dot{m}\, c_p (t_{gap} - t_i) \quad (4)$$

Where $\dot{m}$ represents air mass flow rate, c_p specific heat capacity at constant pressure, and t_{gap} temperature of air that enters the interior from the gap. Convective heat exchange is equal to zero due to the absence of interaction with indoor air in O-O, AB, and I-O ventilation mode. The last term in Eq. (1) represents energy that needs to be added or subtracted from the air infiltrated from the outside in order to bring its temperature to the interior one:

$$q_{ent} = \dot{m}\, c_p (t_o - t_i) \quad (5)$$

The infiltrated air from the outside replaces the air ventilated through the DSF cavity to the outside environment. In this way, the air balance is maintained. As might be expected, this quantity is different from zero only for exhaust ventilation mode.

Table 2
Factors and corresponding levels for FFD.

	Level I	Level II	Level III	Level IV	Level V
Temperature difference [^{0}C] (DBT)	−20	−10	0	10	20
Incident solar irradiance [Wm^{-2}] (ISR)	0	200	400	600	800
Slat angle of blinds [^{0}] (SA)	OFF	90	60	30	0
Airflow rate [m^2s^{-1}] (AR)	0	0.04	0.08		
Cavity depth [cm] (CD)	20	40	60		
Inner glazing [see Table 1] (IG)Table 2.	LOW	MED	HIGH		
Outer glazing [see Table 1] (OG)	LOW	MED	HIGH		
The front side of the slat [see Table 1] (FSR)	LOW	MED	HIGH		
The back side of the slat [see Table 1] (BSR)	LOW	MED	HIGH		

3.2. Tested DOEs

Full factorial design (FFD) considers the highest number of factors and levels, which results in by far the largest number of experimental runs. Such diversity in combinations means that ANOVA results for this design can be considered a benchmark for other designs' performance in process characterization. The following factors and corresponding levels were chosen for four ventilation modes (O-O, I-I, O-I, I-O):

The total number of simulations was 91,125 simulations for each ventilation mode, except AB, where the number of factors was reduced to 8 (no airflow rate), leading to a lower number of simulations (30,375). Considering all five ventilation modes, the total number of simulations was nearly 400,000.

Screening designs - *Definitive screening* (L17/L21), *Plackett-Burman* (L12), *folded Plackett-Burman* (L48) and *two two-level Taguchi designs* (L12 and L32) were tested as screening designs. Levels in the screening designs were chosen to correspond to high and low levels from FFD, while the same number of factors was considered. DSD introduces an additional intermediate level between high and low for continuous factors, making the characterization deeper. In order to suit the structure of this design, five factors whose physical properties are continuous were arbitrarily discretized. Therefore, level II (MED) was defined as the intermediate level for the factors: inner and outer glazing type and the front and backside of the slat. For the slat angle of the blinds, open blinds (90°, level II) were taken to represent the intermediate level that is found between low (0°) and high (OFF) levels. Here it is useful to highlight that although some of the factors in the problem we analyze may appear to be categorical (like the glazing type, or the shading position), they are simply technological implementations of continuous factors, as the fundamental equations that describe the relations between independent and dependent variable make use of continuous factors (which are the thermal and radiative properties of the different layers in the DSF). For example, in this case the slat angle of venetian blinds can be described with the direct solar transmissivity of a layer placed in the middle of the DSF (the shading system), whose value goes between zero or close to zero (fully closed) and one (no shading system). In DSD, AB ventilation mode has a lower number of experimental runs (L17) than other ventilation modes (L21) because it contains fewer factors (8 compared to 9).

Taguchi multilevel designs - *Twenty-three different Taguchi multilevel designs* were tested in five different ventilation modes, resulting in 115 designs and 2480 experimental runs. Arrays differ in the number of experimental runs (8–54) and considered factors (2–6), leading to some being able to assess the main effects (F), while others can evaluate both main effects and interaction (F&I). The values for the levels in Taguchi designs are chosen to correspond to those of FFD. Unlike screening designs, Taguchi multilevel and RS designs do not consider all factors, so it was necessary to define base levels for factors not included in designs. Base levels were set to 'mid' levels or in a state where the related element is not present and cannot influence heat transfer (Table 3). Factors included in Taguchi multilevel designs were selected based on the magnitude of their contribution seen by the ANOVA applied on FFD.

Designs associated with RSM - *Central composite* and *Box-Behnken designs* (CCD and BBD) were chosen as representative cases of designs associated with RSM. The number of experimental runs in these cases grows significantly if categorical runs exist. Though some of the factors in this investigation may seem to be categorical at first sight (e.g. the glazing type), because the underlying physics is based on continuous physical properties and functions, there was no need to treat such factors as categorical. This makes it possible to limit the numbers of runs to 25–45 for CCD, depending on the number of factors included in the design (four to six). For Box-Behnken design, the number of experimental runs goes from 24 to 48, depending on the number of factors considered (four to six factors). The RSDs include only those factors that ANOVA and FFD see with a percentage of contribution greater than 1% (including interactions). Factors not included in RSDs are tuned to their base levels, just like for the Taguchi multilevel designs. The face-centered type of CCD was selected with an alpha value of one so that the axial points fall into the interest range and correspond to low and high levels of FFD. The type of glazing, blind's radiative properties and shading system state were chosen to suit the corresponding axial and center points required by the CCD and BBD. Due to the deterministic nature of the simulations, one center point, along with no replication and randomization of runs, was selected as the preferred settings for both designs

Since a large number of simulation runs were carried out in this study (i.e., nearly 400,000 for the FFD and around 3500 for the whole set of investigated DOEs) through the use of EnergyPlus, the simulation workflow, including inputting data, running simulations, reading output and classifying data, had to be automatized. A template input file for *EnergyPlus* was created as the first step in the workflow, and through a dedicated custom-made Python script, individual input data files were then authomatically created for each simulation (hence changing the independent variables). Another Python script was then used to run all the group *EnergyPlus* simulations and post-process the simulation's output data. The output data for all simulations were then collected in a single CSV file, later used for further analysis.

3.3. Data analysis

Simulation outputs from both the FFD and the different factorial designs of the investigated DOEs were processed by applying ANOVA to identify the nature of the process and the cause-and-effect relationships between the variables. ANOVA quantified the impact of factors, and factors' interactions, on the response variables and estimated the amount of the variance that cannot be explained and attributed to the factors and their interactions. However, that means that ANOVA of the FFD also contained a certain amount of unexplainable variance. To consider full factorial analysis highly successful, the amount of unattributable part of variance needs to be negligibly small, or in other words, the coefficient of determination needs to be very high ($R^2 \approx 0.95$–1.00). An additional condition that ensures that variance is explained only by significant factors is that the values of the adjusted coefficient of determination and of the predicted coefficient of determination are similar to the value of the standard coefficient of determination R^2.

Table 3
Base levels of factors that were not included in Taguchi's multilevel designs.

Factors	Temperature [°C]	Incident solar irradiance [Wm^{-2}]	Slat angle [°/−]	Airflow rate [m^2s^{-1}]	Cavity depth [cm]	Inner glazing [−]	Outer glazing [−]	Inner blinds [−]	Outer blinds [−]
Base level	20	0	OFF	0	40	MED	MED	MED	MED

To assess how well each (simpler) factorial design matched the information extracted from the full factorial design for a given factor and interaction, and to have such an assessment carried out in a quantitative way, we introduced in this study a new metric called fitting coefficient ([−]). This indicator was conceived as a one-value number that provides a comprehensive assessment of the "distance" between the output information of a certain design and the correspondent information in the full factorial design, as well as the assessment of the match between the unexplained variance in the FFD and the certain design. This coefficient is calculated for each specific factor and interaction and can assume values between *0* and *1*, where *0* means that the particular design does not detect any statistically significant dependence of the response variable on factors, while the FFD explains all the variance (tested model fails, FFD succeeds). Some designs are able to estimate the contributions of each individual factor, but if they do not leave any degrees of freedom for the calculation of the error, they are not able to assess whether these contributions are statistically significant. Therefore, these designs are considered unsuccessful in characterization. A value equal to *1* means that the compared design provides an identical picture as FFD and that at the same time, FFD can explain all variance (both tested model and FFD succeed) in a statistically significant way. The mathematical formulation of the fitting coefficient is given in Equation 6, where $SS_{F\&I}/SS_{T,FFD}$ is the contribution of the specific factor/interaction in a full factorial design, and $SS_{F\&I}/SS_{T,D}$ is the corresponding contribution in tested design. Furthermore, $SS_{E,FFD}/SS_{T,FFD}$ represents the extent of randomness (unpredictability) in the process seen by ANOVA in the full factorial design, while $SS_{E,D}/SS_{T,D}$ represent same for the tested design. A total number of factors and interactions is given with max, and depends on the FFD. For the considered case, this number is equal to 45.

$$f = 1 - \frac{\sum_{F\&I=1}^{\max} \left| \frac{SS_{F\&I,FFD}}{SS_{T,FFD}} - \frac{SS_{F\&I,D}}{SS_{T,D}} \right| + \left| \frac{SS_{E,FFD}}{SS_{T,FFD}} - \frac{SS_{E,D}}{SS_{T,D}} \right|}{2} \quad (6)$$

For each design applied to a particular ventilation mode, the fitting coefficient *f* was systematically calculated. Since there are five possible ventilation modes, the range of variation of the fitting coefficient was identified for each design and the average value of the fitting coefficient *f* calculated using all possible modes. These two quantities (the range of *f* and the average *f*) were used to classify the performance of the different DOEs against the FFD.

In FFD, a second order fixed-effects model in FFD was adopted, and this showed excellent coefficients of determination. We did not consider it necessary to employ higher-order fixed-effect models because this would have been harder to physically interpret (if physical interpretations were even possible) and, in the end, to compare with lower resolution designs. The existence of statistically significant higher interactions in the fixed-effect model may indeed not have an obvious physical interpretation, but can only mean that any optimization must simultaneously take into account pairs, three or more n-tuples of factors. In CCD and BBD we adopted a full quadratic polynomial model and both the contribution of first-order and the contribution second-order cross-product were considered when assessing the contribution of each factor

4. Results and Discussion

4.1. ANOVA for full factorial design and factors' influence

The analysis of variance for the FFD revealed the functional dependence of total thermal gain. It showed that the DSF's behaviour can be represented in a very satisfying way with a model that includes only the main effects and interactions (Table 4). The addition of higher-order terms (quadratic, triple interactions, or cubic) complicates the model and reduces the coefficients of determination (adjusted and predicted). Each ventilation mode was assessed independently, and separate ANOVAs were carried out for each of the five ventilation modes. It was impossible to produce one FFD that encompassed all the ventilation modes since the air buffer (AB) mode does not have the same number of factors. Furthermore, when the four ventilation modes were analyzed together, a less satisfactory result (R^2 = 86.6 %, Error = 13.4 %) was achieved compared to FFDs for separate ventilation modes and corresponding, independent ANOVA.

The ANOVA analysis showed that 2-way interactions contributed considerably to a global heat transfer for outdoor air curtain (16.90 %), exhaust air (34.72 %) and supply air (31.10 %) ventilation modes. In contrast, the influence of interaction was much less relevant for the other two modes. Among the main factors, the cavity depth was the least influential variable, and for the air buffer mode, it was not even significant in the control of the heat transfer. Regarding the indoor air curtain mode, the incident solar irradiance had by far the greatest impact on the global heat transfer (79.21 %), followed by the slat angle (4.06 %) and the airflow rate (3.48 %). The low impact of temperature difference on global heat transfer performance can be explained by the lack of interaction between indoor and outdoor air. Regarding exhaust air mode, the temperature difference emerged, instead, as the most critical factor (52.89 %), followed by the 2-way interaction between temperature difference and airflow rate (32.22 %) and incident solar irradiance (7.69 %). The importance of the first two factors originated from the enthalpy flow rate, which was directly dependent on temperature difference and the rate of air extraction through the cavity (equation 5).

The general picture for supply ventilation mode is similar to exhaust mode; the highest impact was caused by the temperature difference (48.12 %), followed by interaction between temperature difference and airflow rate (29.48 %) and incident solar irradiance (18.03 %). The influence of solar radiation is amplified here, possibly due to the shading heated by the radiation, which in turn warms up air delivered from the outside through the cavity. The outdoor air curtain mode showed the most diverse situation, where five factors and interactions contributed more than 3 %: incident solar irradiance (51.60 %), slat angle (16.38 %), the interaction between incident solar irradiance and slat angle (8.27 %), airflow rate (5.27 %), and type of outdoor glazing (3.65 %). In addition to this, there was an influence of six other factors and interactions larger than 1 %. This order in contributions of factors probably comes from the absence of direct interaction between inner and outer air. Outer glazing and airflow through ventilation effects additionally control the heat transfer. The analysis of variance for the air buffer mode did not recognize cavity depth as a significant variable. Here, incident solar irradiance played the most crucial role by far (75.73 %), followed by the slat angle (9.17 %) and interaction between them (4.58 %). Air acts as an insulator, decreases the heat flow due to transmission, and reduces the impact of temperature difference in this way.

The graphical representation of ANOVA results is given in Fig. 2, where the effects of the six most influential factors are shown. The average values of global heat gain density for each ventilation mode, i.e., the grand mean of all runs associated with a particular mode, are presented with horizontal dashed lines. This quantity can be interpreted as the overall (bulk) efficiency or capability of each ventilation mode in damping the net heat transfer across the whole domain of boundary conditions utilized as independent variables in the factorial designs. The highest value (169.5 Wm^{-2}) characterizes indoor air curtain, while the lowest (95.8 Wm^{-2}) characterizes exhaust air mode. The average values of the levels for certain factors are denoted with circles. For example, in supply ventilation mode, an average value for level 5 (20 °C) of tempera-

Table 4
Contribution of all factors and interactions for five ventilation modes.

Ventilation modes	Contribution [%]				
	I-I	I-O	O-I	O-O	AB
Model R^2	98.46	99.73	99.59	97.90	99.02
Model R^2 (predicted)	98.45	99.73	99.59	97.89	99.01
Model R^2 (adjusted)	98.46	99.73	99.59	97.90	99.01
Linear	91.23	65.01	68.49	81.00	90.34
2-Way Interactions	7.23	34.72	31.10	16.90	8.68
Error	1.54	0.27	0.41	2.10	0.98
DBT	1.36	52.89	48.12	2.03	0.81
ISR	79.21	7.69	18.03	51.60	75.73
SA	4.06	2.47	1.04	16.38	9.17
AR	3.48	1.10	0.54	5.27	–
CD	0.01	0.00	0.00	0.00	Not sign.
IG	0.57	0.30	0.14	2.03	2.17
OG	2.04	0.55	0.50	3.65	2.22
FSR	0.47	0.00	0.10	0.03	0.23
BSR	0.04	0.00	0.01	0.02	0.00
DBT*ISR	Not sign.	0.00	0.01	0.01	0.00
DBT*SA	Not sign.	0.00	0.00	0.00	Not sign.
DBT*AR	0.11	32.22	29.48	0.11	Not sign.
DBT*CD	0.00	Not sign.	0.00	Not sign.	Not sign.
DBT*IG	0.00	0.00	0.00	0.08	0.02
DBT*OG	0.05	0.00	0.00	0.01	0.01
DBT*FSR	Not sign.	Not sign.	0.00	Not sign.	Not sign.
DBT*BSR	Not sign.	Not sign.	0.00	Not sign.	Not sign.
ISR*SA	2.05	1.24	0.53	8.27	4.58
ISR*AR	1.71	0.41	0.30	2.79	–
ISR*CD	0.00	0.00	0.00	0.00	Not sign.
ISR*IG	0.29	0.15	0.08	1.02	1.12
ISR*OG	1.04	0.28	0.26	1.84	1.10
ISR*FSR	0.24	0.00	0.05	0.01	0.11
ISR*BSR	0.02	0.00	0.00	0.01	0.00
SA*AR	0.19	0.05	0.04	0.28	–
SA*CD	Not sign.	Not sign.	0.00	Not sign.	Not sign.
SA*IG	0.14	0.06	0.03	0.43	0.22
SA*OG	0.59	0.21	0.15	1.41	1.11
SA*FSR	0.43	0.01	0.09	0.10	0.28
SA*BSR	0.05	0.00	0.01	0.01	0.01
AR*CD	0.00	0.00	0.00	0.00	–
AR*IG	0.12	0.01	0.03	0.04	–
AR*OG	0.03	0.04	0.01	0.25	–
AR*FSR	0.05	0.01	0.01	0.06	–
AR*BSR	0.01	0.00	0.00	0.02	–
CD*IG	0.00	Not sign.	Not sign.	Not sign.	Not sign.
CD*OG	0.00	Not sign.	0.00	Not sign.	Not sign.
CD*FSR	Not sign.	Not sign.	Not sign.	Not sign.	Not sign.
CD*BSR	Not sign.	Not sign.	Not sign.	Not sign.	Not sign.
IG*OG	0.04	0.02	0.01	0.13	0.09
IG*FSR	0.00	0.00	0.00	0.00	0.01
IG*BSR	0.00	0.00	0.00	0.00	0.00
OG*FSR	0.06	0.00	0.01	0.00	0.02
OG*BSR	0.01	0.00	0.00	0.00	0.00
FSR*BSR	0.00	0.00	0.00	0.00	Not sign.
Total	100.00	100.00	100.00	100.00	100.00

ture difference is 429.0 Wm^{-2}, and this is denoted with the highest blue circle on the first graph. A greater range of average values of levels for some particular factor (vertical extent of solid lines on the figures below) means at the same time a greater impact of that quantity on the thermal metric [41]. For example, based on the graphs below, temperature difference and incident solar irradiance were the dominant factors in exhaust and supply ventilation modes Fig. 3.

However, some other interesting conclusions can be drawn from the presented graphs; for example, overall net heat transfer increased by increasing the airflow rate for indoor air curtain and supply ventilation mode. For outdoor air curtain and exhaust mode, the situation is the opposite. Consequently, it seems that airflow in some modes makes DSF more efficient and, in others, less efficient. It is also visible that increasing the angle of venetian blinds from 0° to 60° generally promoted net heat transfer. However, the maximum of transfer occured for the angle between 60° and 90°, due to the fact that the simulations were carried out assuming fully diffuse radiation, which is conventionally modeled in many software tools for building performance simulations as corresponding to direct radiation with an impinging angle on the surface between 60° and 70°. Interestingly, the preferred type of inner glazing for reducing overall heat transfer was the one with high transmittance for all ventilation modes. However, when it comes to the outer glazing, the type with medium transmittance was preferred for all ventilation modes, except for outdoor air curtain and supply mode.

4.2. Screening designs

The detailed information on the results of the ANOVA of screening designs is given in Table 5. The fourth column indicates

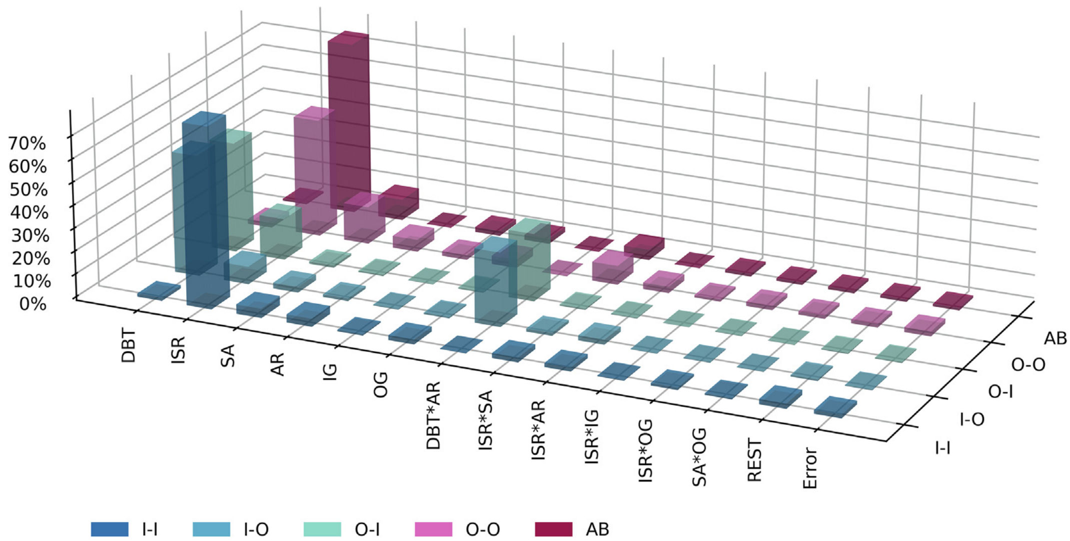

Fig. 2. Contribution of the most relevant factors and interactions for different ventilation modes.

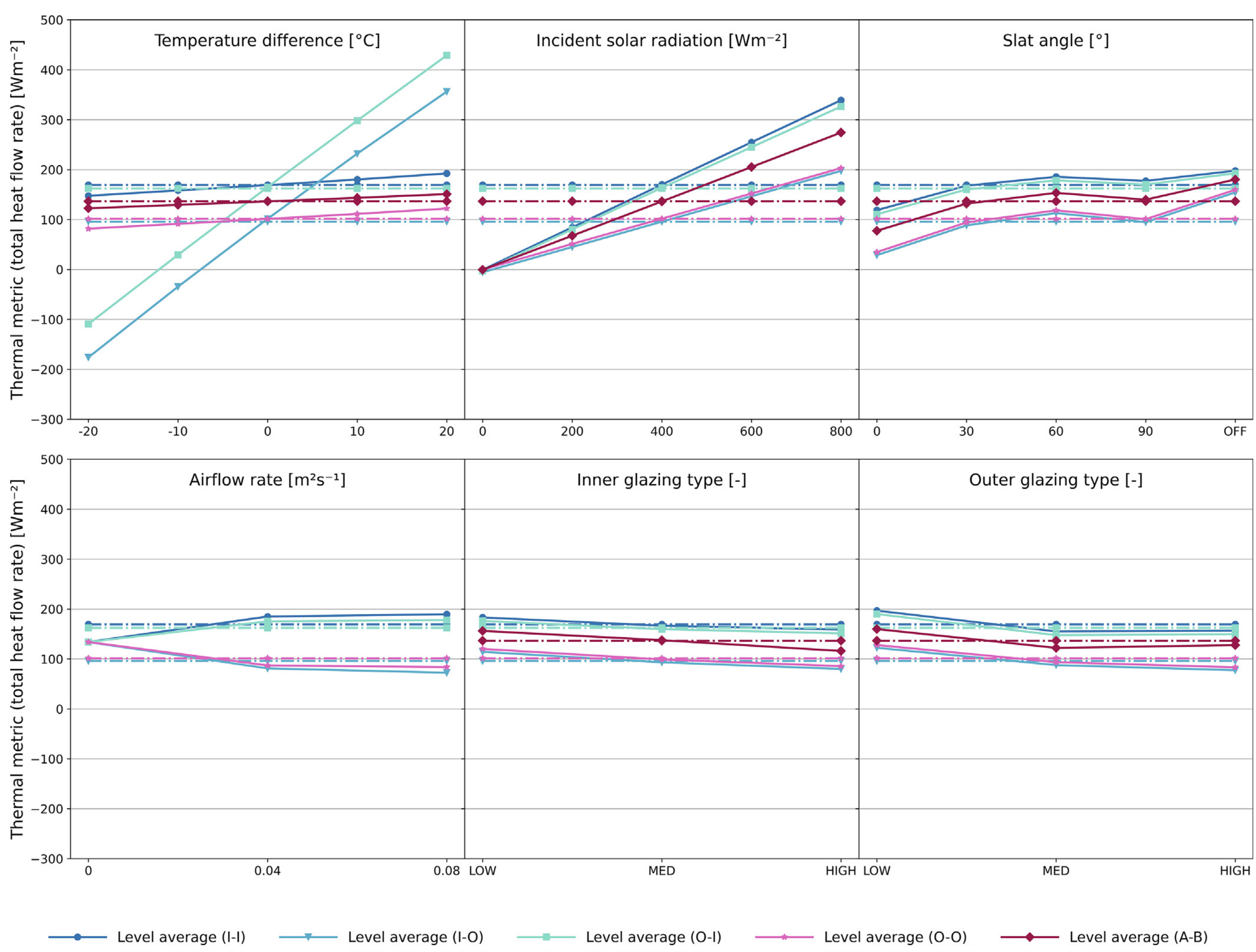

Fig. 3. Results of analysis of variance ANOVA.

Table 5
Values of fitting coefficients for various definitive screening designs.

Designs	Number of runs	Number of factors	Factors/Factors & Interactions	Recommended selection procedure	Fitting coefficient (range)	Fitting coefficient (average)
Taguchi (L32)	32	8/9	F&I	Backward	0.88–0.90	0.89
Definitive screening	17/21	8/9	F&I	Backward mainly	0.82–0.92	0.86
Placket-Burmann 48	48	8/9	F&I	Forward mainly	0.68–0.90	0.82
Taguchi (L12)	12	8/9	F	Forward/backward	0.49–0.84	0.66
Placket-Burmann 12	12	8/9	F	Forward/backward	0.39–0.77	0.58

whether a design can assess only the main effects (F), or if it is capable of evaluating both the influence of factors and interactions (F&I). In the fifth column, the recommended selection procedure is indicated, and when both methods are specified (forward / backward), the final result is the same regardless of choice. When "NO" is displayed, it is not recommended to apply any of the selection procedures, while "-" means that the ANOVA model fails to recognize any statistically significant dependence on the factors (both with the application and no application of selection procedures). The alpha value (critical p-value) for both forward and backward selection procedures is set to 0.05, indicating a high probability that the considered variable is significant. The sixth column shows the range of values for the fitting coefficient from the lowest to the highest. The average fitting coefficient is obtained as the mean value of fitting coefficients for five ventilation modes.

The analysis shows that Taguchi's (L32) design gave the best results overall among all tested screening designs. The definitive screening design can be regarded as the most efficient since it had roughly 30% fewer experiments than Taguchi's L32 array. It is not recommended to use screening designs with fewer experimental runs (Placket Burman and Taguchi L12) since doing so means that only the main effects can be assessed. However, adding interactions does not necessarily mean greater accuracy, i.e., the folded Plackett-Burmann design did not show considerably higher accuracy than III level resolution designs, despite a higher number of experimental runs (48).

The performance of the two best designs in screening out important variables is worth analyzing even more deeply. In the Table 6 we can see a number of factors and interactions whose contribution is higher than a certain percentage (1, 5, or 10 %) in the FFD. It is also shown how many factors and interactions were recognized in DSD and TD L32 designs with an appropriate success rate. If we consider factors whose contribution is higher than 10 %, then these three designs filter out the same factors. The same stands for factors whose contribution is higher than 5 %, except in outdoor air curtain ventilation mode, where three out four factors are recognized. If we consider how good screening designs recognize factors and interactions with a higher share than 1%, then DSD has an average success rate of 57 %, while TD L32 has 78 %. Despite this, DSD and TD L32 showed very good performance in filtering out important variables and consequently can facilitate finding the optimal design. In contrast, screening designs that do not include interactions should not be used to filter out important variables as they show poor performance even in recognizing factors that have a contribution of more than 10%.

4.3. Taguchi multilevel designs

As shown in Table 7, Taguchi's multilevel designs 2L + 3Lx3F (L54), 2L + 4Lx2F (L32), 2L + 3Lx2F (L18), and 3Lx3F (L27) showed the highest accuracy with an average value of the fitting coefficient equal to or higher than 0.88. The second and third designs had a low range of fitting coefficient values, which means that they performed very well for all ventilation modes. However, 3Lx3F (L27) design had a broader range of values, and for some ventilation modes, such as outdoor air curtain and air buffer, the value of the fitting coefficient was around 0.75. The most efficient design is 2L + 3Lx2F (L18), which used only 18 experimental runs and allows the input of three factors. The design 2L + 3Lx3F (L54) is very accurate but cannot be considered among the most efficient ones, as it used a large number of experimental runs. The analysis of the results for different Taguchi multilevel arrays shows how, in general terms, it is not recommendable to use designs with too low a number of experimental runs. It is desirable that a total number of runs is higher than the sum of degrees of freedom for factors and first-order interactions (Appendix A, Tables A.1 and A2) so that design can recognize the basic extent of nonlinearity. However, having many experimental runs is, in itself, not an assurance of good performance. For example, the designs 5Lx2F (L27) and 3Lx2F (L27) were inaccurate in characterizing the role of the factors in the system compared to the full factorial design, although they had fewer experimental runs than some other more successful DOE arrays. Here, the number of the factors that were taken into consideration plays a crucial role in why these designs fail. The average value of the fitting coefficient for designs that could only assess the main factors' influence was 0.55. Those arrays that can evaluate both factors and interactions instead showed a fitting coefficient of 0.78. For non-linear processes, such as the heat transfer phenomena in DSF, it is essential to consider designs that will assess both factors and interactions.

Table 6
Performance of definitive screening and Taguchi L32 designs in filtering out important variables.

	>1%					>5%					>10%				
	I-I	O-O	I-O	O-I	AB	I-I	O-O	I-O	O-I	AB	I-I	O-O	I-O	O-I	AB
	Number of factors (–)														
Full factorial design	8	11	6	4	8	1	4	3	3	3	1	2	2	3	1
Definitive screening	5	5	4	3	3	1	3	3	3	3	1	2	2	3	1
Taguchi (L32)	5	8	5	4	6	1	3	3	3	3	1	2	2	3	1
	Success rate (%)														
Definitive screening	62.5	45.5	66.7	75.0	37.5	100	75.0	100	100	100	100	100	100	100	100
Taguchi (L32)	62.5	72.7	83.3	100	75.0	100	75.0	100	100	100	100	100	100	100	100

Table 7
Fitting coefficient values for various Taguchi's multilevel designs.

Designs	Number of runs	Number of factors	F/F&I	Recommended selection procedure	Fitting coefficient (range)	Fitting coefficient (average)
2L + 3Lx3F (L54)	54	4	F + I	No	0.85–0.95	0.90
2L + 3Lx2F (L18)	18	3	F&I	Backward	0.85–0.91	0.88
2L + 4Lx2F (L32)	32	3	F&I	Backward	0.85–0.91	0.88
3Lx3F (L27)	27	3	F&I	Backward	0.75–0.96	0.88
4L + 2Lx2F (L16)	16	3	F&I	Backward	0.70–0.88	0.81
2L + 4Lx3F (L32)	32	4	F&I	Backward mainly	0.53–0.90	0.79
5Lx6F (L25)	25	6	F	Backward mainly	0.64–0.88	0.74
3Lx4F (L27)	27	4	F + I	Forward mainly	0.55–0.88	0.73
5Lx5F (L25)	25	5	F	Backward mainly	0.64–0.85	0.73
5Lx3F (L25)	25	3	F	Backward mainly	0.53–0.86	0.72
3Lx5F (L27)	27	5	F&I	Forward	0.50–0.85	0.71
4Lx3F (L16)	16	3	F	Backward	0.49–0.88	0.69
4Lx5F (L16)	16	5	F	Backward mainly	0.49–0.83	0.67
5Lx4F (L25)	25	4	F	Forward / backward	0.49–0.83	0.67
3Lx2F (L27)	27	2	F + I	No	0.09–0.90	0.54
2L + 3L (L18)	18	2	F + I	No	0.08–0.89	0.54
3Lx3F (L9)	9	3	F	Backward mainly	0.02–0.87	0.53
3Lx4F (L9)	9	4	F	Forward	0.02–0.81	0.53
4Lx2F (L16)	16	2	F	Forward	0.09–0.86	0.53
5Lx2F (L25)	25	2	F	Forward	0.09–0.86	0.52
4L + 2Lx2F (L8)	8	3	F	Backward mainly	0.02–0.80	0.51
3Lx2F (L9)	9	2	F	Forward	0.09–0.90	0.5
2L + 3Lx3F (L18)	18	4	F + I	Forward	0.11–0.81	0.49
4L + 2L (L8)	8	2	F	–	0.01–0.19	0.06

Table 8
Fitting coefficient values for RSM designs.

Designs	Number of runs	Number of factors	F/F&I	Selection procedure	Fitting coefficient (range)	Fitting coefficient (average)
Central composite	25–45	4–6	F&I	Backward	0.89–0.92	0.91
Box-Behnken design	24–48	4–6	F&I	Backward	0.84–0.93	0.88

4.4. Designs associated with RSM

The CCD (Table 8) shows excellent results with the highest average fitting coefficient of 0.91. However, the number of experimental runs was relatively high for some ventilation modes (45 for outdoor air curtain). When there were five or fewer factors, the total number of experimental runs was lowered to a value considered more acceptable (<27). A similar picture is seen for Box-Behnken design, which has the advantage of not using too many extreme levels simultaneously. However, it has a slightly higher number of experimental runs and a broader range of fitting coefficient values with a lower average value.

4.5. Summary of the DOEs performance assessment

Fig. 4 represents a graphical summary of the DOEs performance assessment, where the values of fitting coefficients for each ventilation mode and tested design are given. The ventilation modes are indicated with different colors, while the tested designs are determined by the corresponding radial directions on which the values of the fitting coefficient for five different ventilation modes lie. Among the screening designs, the Taguchi (L32) and the definitive screening designs proved to be the best ones. The former showed very good performance for all ventilation modes (low fitting coefficient range) and the slightly higher average value of the fitting coefficient. Screening designs are recommended as the initial step when the nature of the process is unknown or where the number of possible factors that may affect the response quantity is high. In this way, by performing screening designs, the important parameters can be filtered out. From the previous analysis, both DSD and Taguchi L32 have proved to be reliable in filtering out the factors that most contribute to the variation of response quantity while pointing to the possible existence of higher-order terms.

Unlike screening, Taguchi multilevel designs tend to use fewer experimental runs, but they do not allow one to include a high number of factors. In this analysis, three designs showed very good performance: 2L + 4Lx2F (L32), 3Lx3F (L27), and 2L + 3Lx2F (L18), where the last one had the highest efficiency in comparison to all other tested designs. Taguchi multilevel designs are recommended when one is sure that a complex process is affected by few(er) factors. At the same time, the possibility to carry out experimental runs is very limited. This type of design allows one to have higher discretization for factors that are assumed to be more important than others, not only when it comes to their direct impact but also their interactions with other factors. In this way, a better insight into the nature of interaction can be obtained compared to what screening designs offer with their two-level approach. However, one should be careful not to choose the design that is overloaded with levels as a sufficient number of degrees of freedom may not be secured for assessment of interactions.

Finally, CCD showed the best performance among all tested designs, while BBD appeared somewhat less consistent with FFD than CCD. However, BBD could be a suitable choice if it is hard or expensive to replicate conditions where several factors are set at extreme levels. If the experiment is limited with runs, the highest number of considered factors is five for CCD and four for BBD.

4.6. General guidelines for the selection of optimal DOEs

Based on the information available in the literature and the results of the investigation presented in this study on the specific case of a DSF, we tried to define some general guidelines to help researchers and designers select optimal DOEs that go beyond the considered case. Since every investigation is different, it is impossible and meaningless to define a ranking for more or less efficient DOEs in general terms. Instead, the recommendations

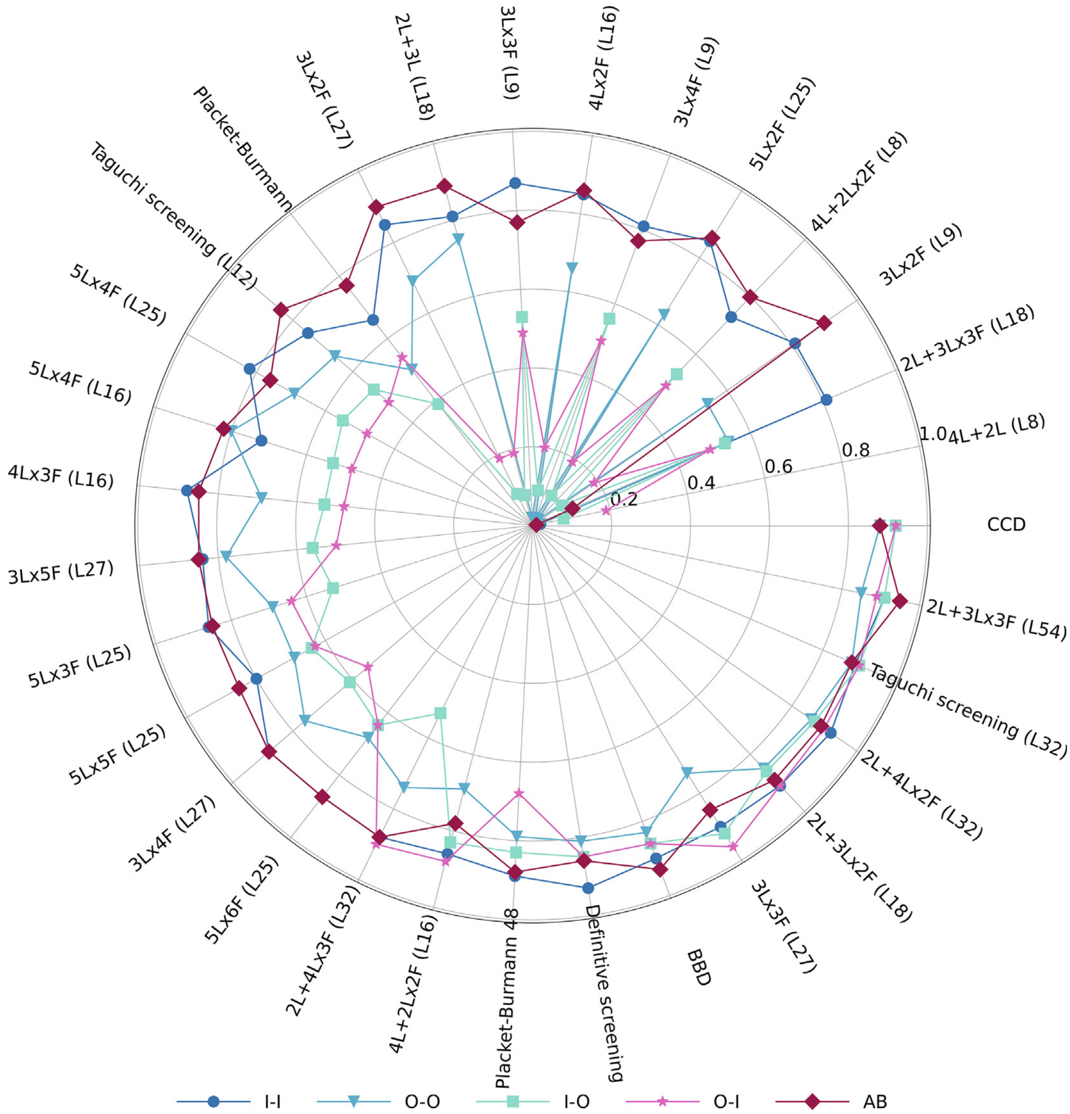

Fig. 4. Fitting coefficient values of all tested designs, where for each one, five different values are indicated in the radial direction (representing five different ventilation modes). Satisfactory designs, such as CCD, Taguchi L32, 2L + 3Lx2F (L18), or BBD, have high fitting coefficient values for all the ventilation modes, i.e., the points representing these values are close to the circumference, and they are characterized by a small extent in the radial direction. In such designs, there is a balance between the number of experimental runs (moderate), factors (moderate or high), and levels (low or moderate), and ANOVA can unveil the strong statistical significance of both factors and interactions.

we report here provide a general approach to have a more thoughtful and grounded selection process when adopting a DOE-based approach. Furthermore, this strategy refers to resource-limited experiments that aim to investigate complex processes/systems characterized by a certain amount of nonlinearity, which we usually encounter in many processes in nature and building physics.

We summarize and visualize the different steps and checks that we recommend in order to carry out to select suitable options for DOEs in Fig. 5.

The essential step in the recommended approach is the proper preparation of input data. In order to make the results of an experiment more general and applicable, it is recommended to identify as many factors as possible that can influence response quantity. They must be mutually independent so that a change in one factor does not induce a change in another factor (not to be confused with interactions). The input data preparation also includes the assignment of low- and high-level values to each factor. These values should be selected based on the range of interest. However, the physical experiment's limitations should be considered since the extreme values may sometimes be complicated to replicate, primarily when several factors with extreme values are run simultaneously.

Care should be taken in analyzing the problem to properly assess whether truly categorical factors exist, and this can sometime be tricky to assess in building science problems at first sight. Technology in the building industry usually promotes discrete product classification, such as components with sets or combinations of predetermined mechanical, thermal, or optical characteristics. These elements may seem to be only described in the form of categorical factors, which can increase the numbers of simulations/tests in some DOEs. While some properties/products can indeed only be described in the form of categories in opposition to continuous factors (such as the ventilation path in the case presented in this article), many variables that can appear at first glance as categorical are, from a physics perspective, continuous. Here, the researcher's experience with the underlying physics and phenomena described in the problem is of fundamental importance to understand the "true" nature of the factor's physical properties behind the technological implementations.

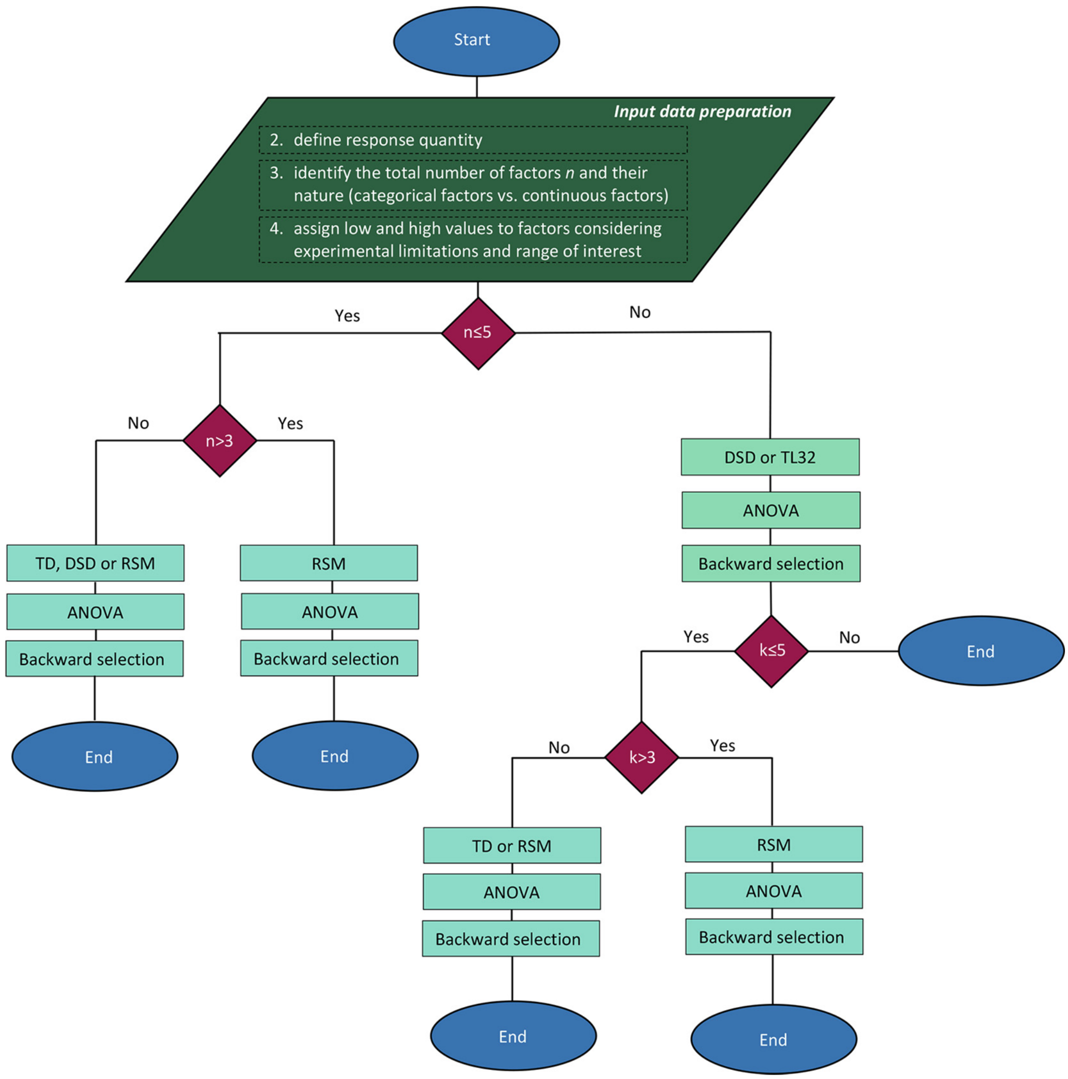

Fig. 5. Recommended decision tree to support the selection of DOEs to investigate a given process. The total number of factors (n) and the number of important factors (k) are used to determine the suitable DOEs for different cases.

In the next step (Fig. 5), the total number of factors n determines whether it is necessary to filter important factors. If the total number of considered factors is six or more, then filtering important factors k is recommended (Fig. 5). DSD or Taguchi L32 designs are favored designs for screening, where the factors with a contribution higher than 1–2 % (both through individual action or through interaction) should be considered important factors. The CCD design is recommended when the number of considered factors has dropped to five or less. If it is difficult to perform an experiment when several factors simultaneously have extreme values, BBD could be a more suitable design than CCD. In the RSDs, the extreme levels (one that corresponds to the axial points) should be chosen to conform to the limits of the interest range so that the "classic" low / high levels fall inside the interest domain if alpha is higher than one.

However, besides RSDs, Taguchi multilevel or DSD can be optimal if only two or three factors are considered. Here, special care should be taken to leave a sufficient number of degrees of freedom for the error, especially with the Taguchi multilevel designs. If the array is designed so that it leaves no degrees of freedom for the error, then it will not be able to evaluate statistical significance of individual contributions. Attention should also be paid to the magnitude of the error. A very large error (>20%) means that ANOVA fails to explain a large part of variance and that either an experiment has not been performed well, or that inappropriate design, factors or even processes are chosen for the analysis. In general, errors greater than 5% for a limited number of experimental runs make it very difficult to argue for a strong statistical significance ($p < 0.05$) of the factors and their interactions. To retain a sufficiently high resolution and enable the assessment of the influence of higher-order terms (interactions), it is not recommended to overload design with too many levels. Taguchi multilevel designs containing at most three or four levels have shown remarkable performances. A backward selection procedure is recommended if the aim is to include only significant terms and reduce the model's complexity so it can be used for prediction. If there are enough resources, it is always wise to compare the analysis of variance obtained by two different designs to ascertain the validity of the obtained characterization picture. This is not so far-fetched considering that the individual design points are common to different types of DOE arrays. DOE is applicable to experiments in a controlled environment, where the factors can be systematically manipulated. However, DOE can be applied to natural experiments, such as time-series data, as long as the response quantity is not inert to factor values changes.

5. Conclusions

Through the results of an extensive simulation study, we have investigated how different design of experiments (DOEs) can lead to the different characterization of the same phenomenon and how the optimal design(s) can be selected to obtain the best possible characterization of the process using the fewest possible experimental runs. In the specific case study used in this research (investigating the thermal behaviour of a double skin façade as a function of its constructional and operational features), the RSM (Response Surface Methodology) with central composite design showed the best performance in the characterization with the average value of the fitting coefficient of 0.90. The number of experimental runs differed for this design, based on the total number of factors that influence more considerable (>1%) response quantity. This number went from 25 experiments for exhaust air, supply air, and air buffer, through 27 for indoor air curtain, to 45 for outdoor air curtain ventilation mode. However, the most efficient design that best balanced the number of experimental runs and accuracy is the Taguchi L18 array 2L + 3Lx2F. This array considered only two factors with three levels and one factor with two levels using 18 experimental runs, but it could explain almost 80 % of the total variance. Some of the Taguchi designs surprisingly failed in characterization, so one should be very careful when choosing the appropriate design.

Based on the central study results, general guidelines that go beyond the considered case are established. These guidelines recommend procedures for preparing input data for various types of experimental designs. They encompass the definition and interpretation of factors along with assigning level values and ranges. Depending on the initial number of factors, screening procedures can be used to filter out the most significant factors. The extent of nonlinearity in the process determines the resolution of optimal design. If the higher-order terms are significant, some of the RSM designs are advisable. On the other side, if only the main effects and interactions influence response quantity, then some Taguchi design of lower resolution is sufficient (resolution IV). However, during the selection of optimal design, one must carefully consider the physical experiment's limitations, such as time and material resources and the ability to perform experiments under extreme conditions. The selected design should secure a comprehensive picture of interactions using as few resources as possible during the physical experiment.

Declaration of Competing Interest

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

Acknowledgments

The activities presented in this paper were carried out within the research project "REsponsive, INtegrated, VENTilated - REINVENT – windows", supported by the Research Council of Norway through the research grant 262198, and the partners SINTEF, Hydro Extruded Solutions, Politecnico di Torino, and Aalto University. We thank the anonymous reviewers whose comments helped improve this article and Duncan Harkness for proofreading the manuscript.

Appendix A

Let's consider a general case where the impact of three factors A, B, and C, containing a certain number of levels (a, b, c, respectively), needs to be assessed. The response variable may not be dependent only on factors in a linear way, as interactions between them can also have an effect. For example, one factor's influence can be affected by the other factor's level. The mathematical procedure starts with the calculation of the total sum of squares SS_T, which equals a sum of squared differences between each run x_{ijkl} and grand mean $\bar{x}$. This quantity can be partitioned into several components: the sum of squares for the factors (SS_A, SS_B, *and* SS_C), the sum of squares for the interactions (SS_{AB}, SS_{AC}, SS_{BC}, and SS_{ABC}), and the error sum of squares (SS_E). Squared difference between group ($\bar{x}_i, \bar{x}_j, \bar{x}_k$) and grand mean, multiplied by the number of the runs within the group, represent the sum of squares for that factor [42]. A similar definition can be derived for the interaction sum of squares. On the other side, the error sum of squares (SS_E) is the sum of squared differences between individual runs and group means. This term indicates the extent of randomness, and if it is large, then there is less probability that factors or interactions influence response quantity in a statistically significant way.

To conclude whether the main effects or interactions exist, the F-ratio for these elements needs to be calculated. This number represents the ratio between the group and the error variance, where variance can be defined as a particular sum of squares divided by its degree of freedom (DOF), where DOF is dependent on the number of levels (*a,b* and *c*). The calculated value of *F-ratio* is compared with the critical *F value* determined from statistical tables. If it is higher than critical, then there is evidence that at least two levels of a factor differently affect the response variable (i.e., factor significantly affects response variable) or that interaction between factors exists [43]. The following Table A1 summarizes the concepts, procedures and equations adopting the terminology mentioned above.

A general three-factor model that accounts for interactions can be written through regression analysis in the following form:

$$x_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + e_{ijkl} \quad (1)$$

where μ represents the overall mean response (or intercept grand mean), a number that is constant regardless of the level settings, e_{ijkl} is the residual term (error), which represents the effect of all other factors that are not considered in the model [44]. The main effects (i.e., factor *A* at level *i*, factor *B* at level *j* and factor *C* at level *k)* are given in Table A2 with terms α_i, β_j and γ_k, while interactions (between *A* and *B*, *A* and *C*, *B* and *C* and A, B, and C) are given with $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$ and $(\alpha\beta\gamma)_{ijk}$, respectively:

The selection of the factors to be included in the regression model as predictors of the response quantity is done via the variable selection procedures (e.g. forward entry, backward elimination, and stepwise selection procedures [45]). Forward entry starts with the null model and gradually adds one factor at a time from the most to the least significant until the previously specified criterion (critical p-value, F-ratio) is not met or until all factors are included in the model. This method is recommended when the number of factors under consideration is larger than the number of experimental runs. The backward elimination criterion starts the full model. It gradually removes one factor at a time, from the least to the most significant, until the previously specified criterion (critical p-value, F-ratio) is not met or until all factors are excluded from the model. Generally, backward elimination is preferred over forward entry because it is less negatively affected by the collinearity of the model's factors, except when the number of experimental runs is low [46]. The stepwise selection criterion combines forward and backward, so it adds one factor at a time and recalculates the significance of all the factors considered in the model up to this step [47]. If a nonsignificant factor is found, then it is eliminated from the model. This selection procedure requires two specified criteria, one for the entry of the factor and

Table A1
Terminology in the three-factor ANOVA.

Source of variation	Degrees of freedom	Sum of squares	Mean square	F-ratio
Factor A	$df_A = a - 1$	$SS_A = nbc\sum_{i=1}^{a}\left(\bar{x}_i - \bar{x}\right)^2$	$MS_A = \frac{SS_A}{df_A}$	$F_A = \frac{MS_A}{MS_E}$
Factor B	$df_B = b - 1$	$SS_B = nac\sum_{j=1}^{b}\left(\bar{x}_j - \bar{x}\right)^2$	$MS_B = \frac{SS_B}{df_B}$	$F_B = \frac{MS_B}{MS_E}$
Factor C	$df_C = c - 1$	$SS_C = nab\sum_{k=1}^{c}\left(\bar{x}_k - \bar{x}\right)^2$	$MS_C = \frac{SS_C}{df_C}$	$F_C = \frac{MS_C}{MS_E}$
Interaction AB	$df_{AB} = (a-1)(b-1)$	$SS_{AB} = nc\sum_{j=1}^{b}\sum_{i=1}^{a}\left(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}\right)^2$	$MS_{AB} = \frac{SS_{AB}}{df_{AB}}$	$F_{AB} = \frac{MS_{AB}}{MS_E}$
Interaction AC	$df_{AC} = (a-1)(c-1)$	$SS_{AC} = nb\sum_{k=1}^{c}\sum_{i=1}^{a}\left(\bar{x}_{ik} - \bar{x}_i - \bar{x}_k + \bar{x}\right)^2$	$MS_{AC} = \frac{SS_{AC}}{df_{AC}}$	$F_{AC} = \frac{MS_{AC}}{MS_E}$
Interaction BC	$df_{BC} = (b-1)(c-1)$	$SS_{BC} = na\sum_{k=1}^{c}\sum_{j=1}^{b}\left(\bar{x}_{jk} - \bar{x}_j - \bar{x}_k + \bar{x}\right)^2$	$MS_{BC} = \frac{SS_{BC}}{df_{BC}}$	$F_{BC} = \frac{MS_{BC}}{MS_E}$
Interaction ABC	$df_{ABC} = (a-1)(b-1)(c-1)$	$SS_{ABC} = n\sum_{k=1}^{c}\sum_{j=1}^{b}\sum_{i=1}^{a}(\bar{x}_{ijk} - \bar{x}_{ij} - \bar{x}_{ik} - \bar{x}_{jk} + \bar{x}_i + \bar{x}_j + \bar{x}_k - \bar{x})^2$	$MS_{ABC} = \frac{SS_{ABC}}{df_{ABC}}$	$F_{ABC} = \frac{MS_{ABC}}{MS_E}$
Error	$df_E = n - abc$	$SS_E = \sum_{l=1}^{n}\sum_{k=1}^{c}\sum_{j=1}^{b}\sum_{i=1}^{a}\left(x_{ijkl} - \bar{x}_{ijk}\right)$	$MS_E = \frac{SS_E}{df_E}$	
Total	$d_T = n - 1$	$SS_T = \sum_{l=1}^{n}\sum_{k=1}^{c}\sum_{j=1}^{b}\sum_{i=1}^{a}\left(x_{ijkl} - \bar{x}\right)$	$MS_T = \frac{SS_T}{df_T}$	

Table A2
Three-factor model and its terms.

Three-factor model terms	
Main effects	$\mu = \bar{x}\alpha_i = \bar{x}_i - \bar{x}\beta_j = \bar{x}_j - \bar{x}\gamma_k = \bar{x}_k - \bar{x}$
Two-way Interactions	$(\alpha\beta)_{ij} = \bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}(\alpha\gamma)_{ik} = \bar{x}_{ik} - \bar{x}_i - \bar{x}_k + \bar{x}(\beta\gamma)_{jk} = \bar{x}_{jk} - \bar{x}_j - \bar{x}_k + \bar{x}$
Three-way interaction and the error term	$(\alpha\beta\gamma)_{ijk} = \bar{x}_{ijk} - \bar{x}_{ij} - \bar{x}_{ik} - \bar{x}_{jk} + \bar{x}_i + \bar{x}_j + \bar{x}_k - \bar{x}e_{ijkl} = x_{ijkl} - \bar{x}_{ijk}$

the other for its elimination, where the first one needs to be greater than the second.

There is a close connection between RSM and regression analysis [48]. While regression analysis seeks an empirical relationship between the response variable and its affecting factors, RSM represents supplementary techniques including planning, model testing procedures and optimization employed ahead of, during and after regression analysis [49]. Response surface modeling is based on the assumption that response function (surface) can be approximated by a Taylor series expansion and that the surface is curved around the optimum. To describe such response adequately, cross product terms need to be incorporated [50]. The response function can be approximated by polynomials of order higher than three, but if the experimental region is not too broad, lower-degree polynomials (at most three) can successfully approximate the response function.

References

[1] J. Antony, "1 - Introduction to Industrial Experimentation," J. B. T.-D. of E. for E. and S. (Second E. Antony, Ed. Oxford: Elsevier, 2014, pp. 1–6.
[2] B. Durakovic, Design of experiments application, concepts, examples: State of the art, Period. Eng. Nat. Sci. 5 (3) (Dec. 2017) 421–439.
[3] D.C. Montgomery, Design and analysis of experiments, John Wiley & Sons, New York, 2001.
[4] A. Alafaghani, A. Qattawi, Investigating the effect of fused deposition modeling processing parameters using Taguchi design of experiment method, J. Manuf. Process. 36 (2018) 164–174.
[5] F. Kolivand, R. Rahmannejad, Estimation of geotechnical parameters using Taguchi's design of experiment (DOE) and back analysis methods based on field measurement data, Bull. Eng. Geol. Environ. 77 (4) (2018) 1763–1779.
[6] C. Shen, L. Wang, W. Cao, L. Qian, Investigation of the effect of molding variables on sink marks of plastic injection molded parts using taguchi DOE Technique, Polym. Plast. Technol. Eng. 46 (3) (Mar. 2007) 219–225.
[7] S.M. Zahraee, M. Hatami, N. Mohd Yusof, J. Mohd Rohani, F. Ziaei, Combined use of design of experiment and computer simulation for resources level determination in concrete pouring process, J. Teknol. 64 (1) (2013) Nov.
[8] M.A. Farooq, H. Nóvoa, A. Araújo, S.M.O. Tavares, An innovative approach for planning and execution of pre-experimental runs for Design of Experiments, Eur. Res. Manag. Bus. Econ. 22 (3) (2016) 155–161.
[9] N.R. Smalheiser, in: Chapter 5 - Experimental Design: Design Strategies and Controls, Academic Press, 2017, pp. 65–85.
[10] W.J. Kolarik, Creating quality: concepts, systems, strategies, and tools, McGraw-Hill New York, 1995.
[11] G. Box, J. Hunter, W. Hunter, "Chapter 6 Fractional Factorial Designs", in *Statistics for experimenters design, innovation, and discovery*, 2nd ed., Wiley-Interscience, Hoboken, New Jersey, 2005, pp. 235–271.
[12] J. Antony, in: 2 - Fundamentals of Design of Experiments, Elsevier, Oxford, 2014, pp. 7–17.
[13] B. Jones, K. Allen-Moyer, and P. Goos, "A-optimal versus D-optimal design of screening experiments," J. Qual. Technol., pp. 1–14, May 2020.
[14] L. Eriksson, E. Johansson, C. Wikström, Mixture design—design generation, PLS analysis, and model usage, Chemom. Intell. Lab. Syst. 43 (1) (1998) 1–24.
[15] K.K. Choi, B.D. Youn, R.-J. Yang, Moving least square method for reliability-based design optimization, Proc. 4th World Cong. Struct. Multidiscip. Optim. (2001).
[16] B. Jones, C.J. Nachtsheim, Split-Plot Designs: What, Why, and How, J. Qual. Technol. 41 (4) (Oct. 2009) 340–361.
[17] W. Libbrecht et al., Optimization of soft templated mesoporous carbon synthesis using Definitive Screening Design, Chem. Eng. J. 259 (2015) 126–134.
[18] L. Ilzarbe, M.J. Álvarez, E. Viles, M. Tanco, Practical applications of design of experiments in the field of engineering: a bibliographical review, Qual. Reliab. Eng. Int. 24 (4) (2008) 417–428.
[19] S.K. Ahuja, G.M. Ferreira, A.R. Moreira, Application of Plackett-Burman design and response surface methodology to achieve exponential growth for aggregated shipworm bacterium, Biotechnol. Bioeng. 85 (6) (2004) 666–675.
[20] J. Zhou et al., Optimization of phenol degradation by Candida tropicalis Z-04 using Plackett-Burman design and response surface methodology, J. Environ. Sci. 23 (1) (2011) 22–30.
[21] K. Vanaja, R.H. Shobha Rani, Design of Experiments: Concept and applications of plackett burman design, Clin. Res. Regul. Aff., Jan. 24 (1) (2007) 1–23.
[22] A. Miller, R.R. Sitter, Using the Folded-Over 12-Run Plackett—Burman Design to Consider Interactions, Technometrics 43 (1) (2001) 44–55.
[23] G. Box, S. Bisgaard, C. Fung, An explanation and critique of taguchi's contributions to quality engineering, Qual. Reliab. Eng. Int. 4 (2) (1988) 123–131.
[24] K.N. Ballantyne, R.A. van Oorschot, R.J. Mitchell, Reduce optimisation time and effort: Taguchi experimental design methods, Forensic Sci. Int. Genet. Suppl. Ser. 1 (1) (2008) 7–8.
[25] G.E.P. Box, K.B. Wilson, On the experimental attainment of optimum conditions, J. R. Stat. Soc. Ser. B 13 (1) (Sep. 1951) 1–45.
[26] S. Karimifard, M.R. Alavi Moghaddam, Application of response surface methodology in physicochemical removal of dyes from wastewater: A critical review, Sci. Total Environ. 640–641 (2018) 772–797.
[27] X. Shen, G. Zhang, B. Bjerg, Assessments of experimental designs in response surface modelling process: Estimating ventilation rate in naturally ventilated livestock buildings, Energy Build. 62 (2013) 570–580.
[28] W.F. Guthrie, NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151), Natl. Inst. Stand. Technol. (2020).
[29] H.C.J. Hoefsloot, D.J. Vis, J.A. Westerhuis, A.K. Smilde, J.J. Jansen, in: 2.23 - Multiset Data Analysis: ANOVA Simultaneous Component Analysis and Related Methods, Elsevier, Oxford, 2009, pp. 453–472.

[30] S. Weisberg, "Linear Hypothesis: Regression (Basics)," N. J. Smelser and P. B. B. T.-I. E. of the S. & B. S. Baltes, Eds. Oxford: Pergamon, 2001, pp. 8884–8888.
[31] K. M. Carley, N. Y. Kamneva, and R. J, "Response surface methodology," in CASOS Techn. Rep. CMU ISRI-04-136, Pittsburgh: Carnegie Mellon University, 2004.
[32] D.R. Gunasegaram, D.J. Farnsworth, T.T. Nguyen, Identification of critical factors affecting shrinkage porosity in permanent mold casting using numerical simulations based on design of experiments, J. Mater. Process. Technol. 209 (3) (2009) 1209–1219.
[33] I. Lorscheid, B.-O. Heine, M. Meyer, Opening the 'black Box' of Simulations: Increased Transparency and Effective Communication through the Systematic Design of Experiments, Comput. Math. Organ. Theory 18 (1) (2012) 22–62.
[34] A. Giunta, S. Wojtkiewicz, M. Eldred, Overview of Modern Design of Experiments Methods for Computational Simulations (Invited), American Institute of Aeronautics and Astronautics, 2003.
[35] A.N. Sadeghifam, S.M. Zahraee, M.M. Meynagh, I. Kiani, Combined use of design of experiment and dynamic building simulation in assessment of energy efficiency in tropical residential buildings, Energy Build. 86 (2015) 525–533.
[36] I. Jaffal, C. Inard, C. Ghiaus, Fast method to predict building heating demand based on the design of experiments, Energy Build. 41 (6) (2009) 669–677.
[37] N. Delgarm, B. Sajadi, K. Azarbad, S. Delgarm, Sensitivity analysis of building energy performance: A simulation-based approach using OFAT and variance-based sensitivity analysis methods, J. Build. Eng. 15 (2018) 181–193.
[38] W. Kim, Y. Jeon, Y. Kim, Simulation-based optimization of an integrated daylighting and HVAC system using the design of experiments method, Appl. Energy 162 (2016) 666–674.
[39] E. Catto Lucchino, F. Goia, G. Lobaccaro, G. Chaudhary, Modelling of double skin facades in whole-building energy simulation tools: A review of current practices and possibilities for future developments, Build. Simul. 12 (1) (2019) 3–27.
[40] A. Gelesz, E. Catto Lucchino, F. Goia, V. Serra, A. Reith, Characteristics that matter in a climate façade: A sensitivity analysis with building energy simulation tools, Energy Build. 229 (2020) 110467.
[41] A. Jankovic, F. Goia, A simulation study on the performance of double skin façade through experimental design methods and analysis of variance, IOP Conf. Ser. Mater. Sci. Eng. 609 (2019) 62003.
[42] J. Berk and S. Berk, "Chapter 11 - ANOVA, Taguchi, and Other Design of Experiments Techniques: Finding needles in haystacks...," J. Berk and S. B. T.-Q. M. for the T. S. Berk, Eds. Woburn: Butterworth-Heinemann, 2000, pp. 106–123.
[43] B. Bowerman et al., Experimental Design and Analysis of Variance, in: In Business Statistics and Analytics in Practice, 3rd ed.,., McGraw Hill, 2003, pp. 399–443.
[44] B. R. Clarke, "Helmert Matrices and Orthogonal Relationships," Linear Models: The Theory and Application of Analysis of Variance. pp. 77–112, 18-Jul-2008.
[45] A. M. McIntosh, M. Sharpe, and S. M. Lawrie, "9 - Research methods, statistics and evidence-based practice," E. C. Johnstone, D. C. Owens, S. M. Lawrie, A. M. McIntosh, and M. B. T.-C. to P. S. (Eighth E. Sharpe, Eds. St. Louis: Churchill Livingstone, 2010, pp. 157–198.
[46] D.C. Montgomery, E.A. Peck, G.G. Vinning, Variable Selection and Model Building, in: Introduction to Linear Regression Analysi, 5th ed.,., Wiley, 2012, pp. 327–372.
[47] L. J. Hintze, "Stepwise Regression," in NCSS User's Guide III: Regression and Curve Fitting, Kaysville, Utah, pp. 311-1,311-10.
[48] R.H. Myers, D.C. Montgomery, C.M. Anderson-Cook, Response surface methodology : process and product optimization using designed experiments, Wiley, Hoboken, 2019.
[49] A.I. Khuri, J.A. Cornell, Response Surfaces : Second Edition, Routledge, Boca Raton, 2018.
[50] R. B. T.-D. H. in S. and T. Carlson, Ed., "Chapter 12 Response surface methods," in Design and Optimization in Organic Synthesis, vol. 8, Elsevier, 1992, pp. 249–324
[51] Aleksandar Jankovic, Francesco Goia, Impact of double skin facade constructional features on heat transfer and fluid dynamic behaviour, Building and Environment 196 (2021), https://doi.org/10.1016/j.buildenv.2021.107796 107796.