processes

Article

Using Lean-and-Green Supersaturated Poly-Factorial Mini Datasets to Profile Energy Consumption Performance for an Apartment Unit

Spyridon Zarkadas [1,2] and George Besseris [1,2,*]

1 Mechanical Engineering Department, The University of West Attica, 12241 Egaleo, Greece; spyzas93@hotmail.com
2 Advanced Industrial & Manufacturing Systems Graduate Program, Kingston University, London KT1 2EE, UK
* Correspondence: besseris@uniwa.gr

Abstract: The *Renovation Wave for Europe* initiative aspires to materialize the progressive greening of 85–95% of the continental older building stock as part of the *European Green Deal* objectives to reduce emissions and energy use. To realistically predict the energy performance even for a single apartment building is a difficult problem. This is because an apartment unit is inherently a customized construction which is subject to year-round occupant use. We use a standardized energy consumption response approach to accelerate the setting-up of the problem in pertinent energy engineering terms. Nationally instituted Energy Performance Certification databases provide validated energy consumption information by taking into account an apartment unit's specific shell characteristics along with its installed electromechanical system configuration. Such a pre-engineered framework facilitates the effect evaluation of any proposed modifications on the energy performance of a building. Treating a vast building stock requires a mass-customization approach. Therefore, a lean-and-green, industrial-level problem-solving strategy is pursued. The TEE-KENAK Energy Certification database platform is used to parametrize a real standalone apartment. A supersaturated mini dataset was planned and collected to screen as many as 24 controlling factors, which included apartment shell layout details in association with the electromechanical systems arrangements. Main effects plots, best-subsets partial least squares, and entropic (Shannon) mutual information predictions—supplemented with optimal shrinkage estimations—formed the recommended profiler toolset. Four leading modifications were found to be statistically significant: (1) the thermal insulation of the roof, (2) the gas-sourced heating systems, (3) the automatic control category type 'A', and (4) the thermal insulation of the walls. The optimal profiling delivered an energy consumption projection of 110.4 kWh/m^2 (energy status 'B') for the apartment—an almost 20% reduction in energy consumption while also achieving upgrading from the original 'C' energy status. The proposed approach may aid energy engineers to make general empirical screening predictions in an expedient manner by simultaneously considering the apartment unit's structural configuration as well as its installed electromechanical systems arrangement.

Keywords: energy consumption; apartment unit energy screening; supersaturated datasets; performance improvement; main effect plot; partial least squares; entropic mutual information

check for updates

Citation: Zarkadas, S.; Besseris, G. Using Lean-and-Green Supersaturated Poly-Factorial Mini Datasets to Profile Energy Consumption Performance for an Apartment Unit. *Processes* **2023**, *11*, 1825. https://doi.org/10.3390/pr11061825

Academic Editor: Kian Jon Chua

Received: 30 April 2023
Revised: 22 May 2023
Accepted: 13 June 2023
Published: 15 June 2023

1. Introduction

Improving the energy efficiency of the existing building stock has been given high priority in the *European Green Deal* [1]. On track to a climate neutral Europe by 2050, the '*Renovation Wave for Europe*' initiative [2,3] calls for the progressive greening of 85–95% of the older building stock by reducing emissions and energy use. The imposing of such an emphasis on construction processes has transpired because 40% of the energy consumption and 36% of the energy-related greenhouse gas emissions in Europe are attributed to buildings [4–6]. Consequently, the '*Renovation Wave for Europe*''s aim is to campaign for action now in order to improve the energy consumption for a forecasted number of 35 million inefficient buildings by 2030; the initiative anticipates adherence to the target of reducing emissions by at least 55%. Meanwhile, the overall decarbonization of the heating and cooling processes are, of course, in direct alignment to the broader global sustainability goals as they are promulgated through the *United Nations 17 Sustainable Development Goals* (SDGs) [7]. Specifically, the impact of attaining several interlacing SDGs, such as Goal #7 (affordable and clean energy), Goal #11 (sustainable cities and communities), Goal #12 (responsible consumption and production), and Goal #13 (climate action), is envisaged through new-age, digitalized, and smarter home management, which is prognosticated to enhance the overall quality of life of the population.

Incentivizing green construction through effective government policies has shown promise through the concept of the Leadership in Energy and Environmental Design (LEED) certifications [8–14]. The Minimum Energy Performance (MEP) standards for existing buildings are formalized through the Energy Performance of Building Directive (EPBD). The Energy Performance Certificates (EPCs) are instituted on a national level; in essence, they encourage "progressive renovations over the lifetime of a building" through the establishment of the Building Renovation Passports (BRPs). The EPC is the rating output from a green-building certification system that assesses an apartment unit's performance by employing a lifecycle method that considers specific design, construction, and operation details, while upholding the emphasis of quality on the environmental and sustainability perspectives [15].

To tackle the many barriers that hinder a realistic and optimized prediction of the year-round energy consumption of older-building apartment units, innovative solutions may be facilitated by the introduction of data-centric engineering know-how. Old-building stock renovation projects rely on suitable retrofitting modeling instruments that usually address the whole building as a problem [16–24]. From an energy engineer's point of view, retrofitting is essentially a screening-and-optimization energy-performance exercise which is characterized by high complexity. To be pragmatic, the retrofitting optimization problem's complexity should be simplified, as it is known to implicate contributions of uncertainty from the structural details of the examined apartment unit, its electromechanical systems configuration, the unit's actual demands on annual local weather conditions, and indigenous occupant behavior tendencies [25–33]. A sophisticated multicriteria software tool is often necessitated in order to simulate and analyze improvement opportunities for various types of energy leaks, such as to prioritize the retrofitting tasks while paying heed to green considerations [34–43]. Of course, it is anticipated that the energy professional in charge of the retrofitting project should be skillful and knowledgeable in several areas of expertise, including designing and computing, to handle empirical modelling and complicated forecasts. Conducting optimization simulations for large buildings is becoming more common owing to simplification of the apartment unit problem, which allows for the relaxing of some technical assumptions. On the other hand, the energy-consumption optimization of micro- and small-scale systems is more perplexing to carry out than larger ones, because of constraints arising from unit-to-unit differentiation within a building, thus allowing the technical variability of these systems to be deemed substantial [44]. Moreover, an apartment renovation task is a more stringent problem to simulate because the energy efficiency optimization procedure may interrupt the current unit's energy distribution practices. Nevertheless, the model calibration may be conducted on a standardized framework to be convincing and meaningful [45,46]. A major issue that arises from evaluating sustainable retrofits—key to successfully advancing the renovation process—is related to the optimal profiling of alternative retrofit scenarios [47–51]. The synchronous simplification of the complexity in the considered energy system is congruent to the extent of customization at the apartment unit level.

The novelty of this work rests on the notion that screening and optimizing the energy performance for an apartment unit may be facilitated after adopting the Lean Six Sigma initiative, a contemporary manufacturing philosophy which is applied to improving products, processes, and services by instituting a holistic and datacentric engineering strategy [52,53]. While the green Lean Six Sigma initiatives have been launched in the construction industry, instructive paradigms for the screening and optimization of energy performance studies are lacking [54,55]. The examined problem is an excellent paradigm of greening the energy consumption of a building by customizing the optimization solution at the point of interest, which is the apartment unit. The proposed tactic fosters a duality that merges lean thinking and quality improvement [56]. The lean and green aspects of a study are aligned to effectuate the sustainability aspect in a circular process [57–64]. The lean and green objectives are accomplished by: (1) screening and optimizing the reduction in an apartment unit's energy consumption, and (2) introducing lean thinking into the problem-solving process itself. The benefits of the former objective are immediately evident: (1) lower overall energy demand, (2) lower gas emissions, and (3) lower bottom-line energy costs per apartment unit [65–67]. The latter objective may be viewed as more subtle because it regards the energy and resource commitment in order to research the problem. The lean thinking approach is applied in such a fashion as to greatly reduce the amount of work that is necessary to carry out this project. To be a practical and expedient project, it is necessary to save time and resources, thus shortening the data collection cycle. This effort appears innately time-consuming, because year-round data would be necessitated for a building energy-performance study to be rendered pertinent. Quite possibly, a simulation study would be valuable in understanding the mechanisms that regulate the energy distribution within an apartment [68–76]. The complete description of the inherent energy physics that govern the heat flow and the electricity distribution, across the installed electromechanical systems in a real apartment unit, would absorb a lot of resources and knowledge because the multifarious sources of uncertainties should also be accounted. Even so, it would be meaningful only for that specific single customized case. A more general method that relies on standardized data is greatly desirable to heed the practical aspect of the application; method agility and reliability are construed to be attractive features in product/process improvement studies.

To counter the perceived complexity surrounding the data collection part, the proposed methodology orients toward a strategy that draws valid data from a national Energy Performance Certification database [77,78]; it is a standardization output that has been established to track down the decarbonizing progress of existing building stocks. In this manner, the year-round energy parametrization of residential buildings is accomplished by considering the customized structural and electromechanical systems requirements, which are matched to their performances according to standardized specifications. Since the national Energy Performance Certification database holds an immense amount of customized and validated operational coefficients and parameters, any ensuing energy performance analysis is certainly amenable to any residential apartment unit. In Greece, an EU member state, this is transacted by the mandatory government-instituted Energy Certification program that permits, through a software platform, the TEE KENAK (1.31.1.19), the standardized parametrization of existing residential buildings [79,80]. The TEE KENAK software allows energy engineers to extract a vast amount of information in order to conduct energy studies which can lead to awarding an official certification of the energy consumption status to any residential apartment unit. It is on the TEE KENAK database that the energy consumption reduction will be carried out in this investigation.

An energy performance improvement study is a difficult task to undertake because there are, intrinsically, many controlling factors that should be considered. This is easily comprehended by merely interacting with the Energy Performance Certification software, which expects many building layout and operation parameters to be fed into the program [14]. Consequently, a design of experiments (DOE) plan [81,82] should involve a large number of controlling factors that are usually not convenient to program. Further, there is an enormous number of parameter setting combinations that should be organized which, in turn, is followed by the serial execution of the resulting factorial recipes on the Energy Performance Certification software. If this software is to be a realistically useful tool to an energy engineer, who regularly embarks on such renovation improvement studies, a more sophisticated data planning approach should be deployed. Industrial-level trial scheduling which allows for the implementation of an efficient DOE sampling scheme becomes imperative in such a situation [83,84]. Supersaturated factorial samplers, which are often used in large industrial studies, are very appealing because they allow for the disproportional screening of many controlling factors by extracting information from a modest number of data points which is much less than the number of screened factors [85–90].

Interestingly, it was suggested that it may be practical to pace the retrofitting process of older residential buildings by examining them on an apartment unit basis [91]. By taking advantage of the construction-design modularity and the comprehensive information on the electromechanical systems configuration, which are stored in the national building certification register platforms, EPC-generating software packages may be utilized to conveniently screen and optimize the energy-consumption performance of any residential apartment unit. Besides granting a green energy-efficiency rating to a residential structure, an EPC authentication is also accompanied with a MEP standard, which is a customized estimation for the particular apartment unit. Therefore, MEPs may be taken as consistent improvement yardsticks that could guide the apartment unit's future renovation actions. MEPs summarize the yearly-weighted energy performance of a residential structure in a single estimation. A recently published case study on the retrofitting of an apartment unit in a residential building has demonstrated the usefulness on implementing EPC-created information to advance design of experiments (DOE) in order to screen a sizable number of exclusively electromechanical-system controlling factors [91]. The intention was to research a single apartment in a residential building, in which case each floor corresponded to a single apartment unit. The main objective was the minimization of the energy consumption of the single apartment unit which was situated at the highest floor. It became evident that it is a difficult task to successfully simulate an approximate distribution of the energy demands of an apartment unit under year-round realistic operating conditions. By considering the factual structural layouts and electromechanical systems operating information, simulations may be additionally hampered by the inherent high complexity of the problem which relies heavily on the interplay of the specifics of the installed electromechanical systems against the apartment unit's real architectural details. To circumvent the exigent task of screening as many as 24 electromechanical controlling factors, Rousali and Besseris [91] developed an empirical tool that facilitated the estimation of the contributions of the screened strong factors from their regression analysis after pre-treating all of the factors according to their location and dispersion tendencies, using a modified Matrix Data Analysis Chart (MDAC) tool; the MDAC shortened the initial list of factors to a mere few that statistically outperformed the rest.

The challenge for this work is three-fold in terms of the newly introduced case study features: (1) attempt a screening study on a real single-family home, (2) introduce building shell components in the controlling factor list, (3) use tools that are directly accessible from mainstream statistical software packages, (4) extend the data analysis to implicate more advanced statistics. In comparison to the study by Rousali and Besseris [91], this proposal has several novelties. Firstly, it utilizes the latest version of the TEE KENAK software, since the previous publication was based on a software version which is now a decade old. This means that the updated version, which is utilized for this work, is commensurate to the current reality of using and operating a more modern apartment unit. Thus, there will be access to a more mature EPC platform in synergy with a more comprehensive and accurate database. Second, the study by Rousali and Besseris [91] ignored contributions from the unit's shell properties. This weakness is remedied in this new attempt. Third, the overall type of building structure that will be studied here is significantly different from the apartment building studied in reference [91]. This means that a new and challenging case study is added in a very critical research area that now has commenced to investigate how to improve building's energy performance—an objective that is highly prioritized by the EU. Finally, a new analysis methodology is suggested to alleviate the customized approach that was presented by Rousali and Besseris [91]. This means that only simple techniques are used which are immediately available to commercial and freeware statistical analysis software.

The developments of this paper are organized as follows. Next, the Materials and Methods section elucidates the technical description of the shell structural details and the installed electromechanical systems of the modeled apartment unit, along with the supersaturated data planning, collection, and analysis steps that also describe the computational support toolbox. The Results section provides the collected supersaturated dataset which was obtained by manipulating the apartment unit parameters on the TEE-KENAK platform. Subsequently, the factorial screening and optimization solutions are presented in a cooperative application of response graphs/tables and linear regression techniques. In the Discussion section, a peripheral analysis is appended to explore the validity of the predictions using partial least squares, analysis of variance, entropic representations, and hierarchical clustering treatments. Descriptive statistics, correlation analysis, and nonparametric comparisons complete the prediction validation of the proposed methodology. In the Conclusion section, the key findings are summarized and recommendations for further research are provided.

2. Materials and Methods

2.1. Technical Description of the Studied Building Apartment Unit

2.1.1. Basic Location and Energy Consumption Status Information for the Building Apartment Unit

The residential building that is modelled for this study is located in a suburban area in proximity to the city of Athens (Greece). The construction was erected in the year 2004. Regionally, the district is situated within Climate Zone "B" in Greece. For the particular apartment unit, the thermal space that will be examined has been projected to be a surface area of 174.90 m^2. Based on the national cumulative statistics from the estimations on the issued Energy Performance Certificates, 83.82% of the residential buildings which have been constructed before the year 1980 (55% of the total available building stock) have been awarded an energy consumption status of 'H'. Even for more contemporary buildings such as the one that will be analyzed in this work, the majority of the energy performance ratings have been categorized as either a 'C' or 'D' class. The particular apartment unit was certified to require a year-round primary energy consumption of 133.8 kWh/m^2. This was compared against a (theoretical) reference-building energy demand of 129.2 kWh/m^2. The ratio of the two energy consumption estimations (the former over the latter) provides a measure for the energy efficiency. Since the estimated energy efficiency corresponds to the standardized interval (1.00, 1.41), it was awarded a rating status of 'C'. This study is meaningful because, to reach the desired transitory 'yellow zone,' the energy efficiency score should lie within the standardized interval (0.75, 1.00), which corresponds to an energy consumption grade of 'B'. The top rating (class 'A+') is awarded for energy efficiency scores lower than 0.33. However, the 'green' status for a residential building is awarded upon certification after attaining at least a 'B+' rating, which corresponds to an energy efficiency standardized interval of (0.50, 0.75). Therefore, any recommended improvement interventions should lead, at least, to climbing up to the 'yellow zone' scale, before any apartment renovation gains become substantial enough to contribute to the 'greener outlook' of the building's energy performance. Ostensibly, the respective intervention costs, the financial status of the apartment owners/occupants, and any potential government incentives may affect the pace of progress toward reaching an enhanced green building status.

2.1.2. Apartment Unit Structural Details

The complete layout for the apartment unit structure has been drawn in Figure 1 (AutoCAD, Autodesk, San Francisco, CA, USA). The building shell details have been tabulated in Table 1 [92]. They are required inputs for the software package TEE-KENAK (1.31.1.19) [79,80], which evaluates the energy consumption compared to a reference building, such that an energy performance certification can be issued. The case study is typical of a single-family (raised) residential unit with six rooms; two rooms (#1 and #2) are assessed as a single compartment. There are available windows in all of rooms; four out of five have exactly the same window surface area. Similarly, there are ventilation ductways for all rooms, with middling variations among them. For energy consumption estimates to be realistic, even moderate differentiations in the dimensions of the structural elements in the four sides of the construction (Figure 2) may become critical in rendering valid evaluations. The dimensional parameters (column, window, and brick wall surfaces) of all four apartment sides are also entered (Table 2) in the TEE-KENAK software package.

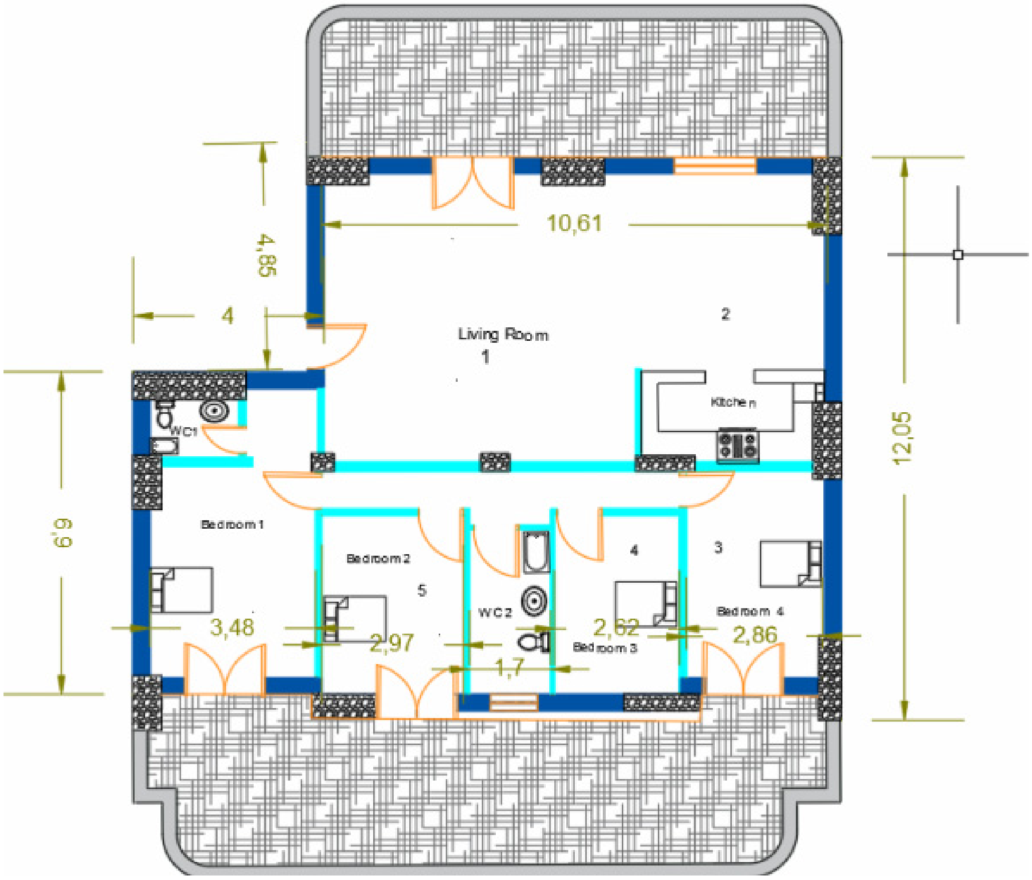


Figure 1. Complete floor layout for the modeled apartment unit.

Table 1. Apartment unit shell structural surfaces.

Room No.	Floor Surface Area (m^2)	Windows Surface Area (m^2)	Ventilation Surface Area (m^2)
1,2	72.6	7.6	5.08
3	20.2	2.99	1.42
4	18	2.99	1.26
5	18	2.99	1.26
6	24.3	2.99	1.7

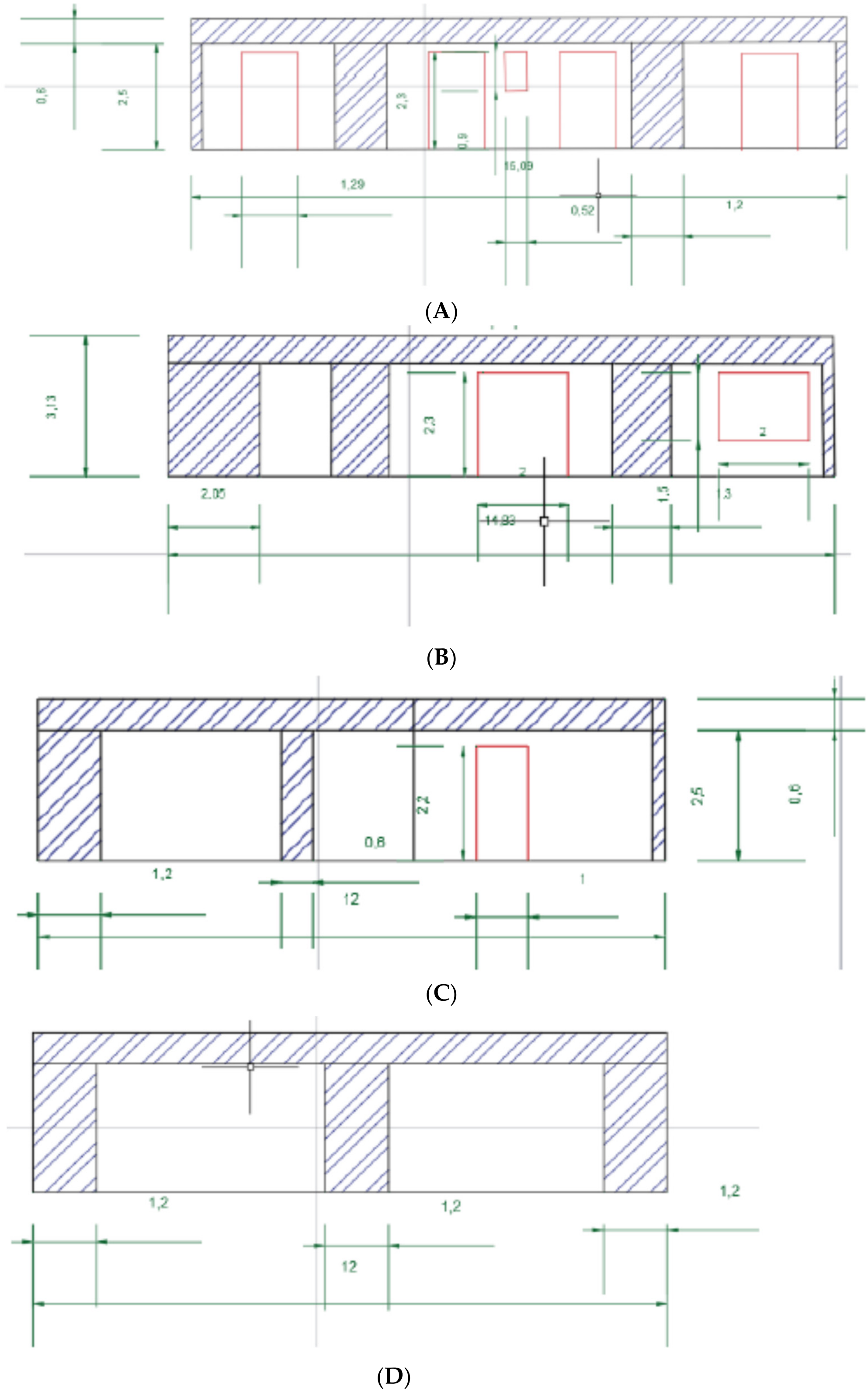


Figure 2. Outer wall layout for each of the four sides: (**A**–**D**).

Table 2. Apartment unit shell structural elements details.

Structural Elements	Side A Surface Area (m^2)	Side B Surface Area (m^2)	Side C Surface Area (m^2)	Side D Surface Area (m^2)
Columns	16.3	21.1	12.3	16.2
Windows	11.5	7.6	2.2	0
Brick wall	19	17.2	22.7	21

In Table 3, shell elements are parametrized in terms of wall and window inputs, which are appropriately coded for orientation, surface area (F), and thermal permeability (k) properties. In Table 4, there is indicative information regarding the input unit block data, given the fact that the structure is raised such that the ground floor level can be used as a parking space as well as to support other auxiliary building facilities.

Table 3. Apartment unit shell energy transfer data.

Shell Element	Element Coding	Orientation (°)	Surface F (m^2)	k (kcal/m^2)
Walls	W1	346	35.3	0.61
	W2	166	38.3	0.61
	W3	76	35	0.58
	W4	256	37.2	0.59
Windows	F1	346	11.5	2.6
	F2	166	7.6	2.6
	F3	76	2.2	3
	F4	256		
	S		167.1	

Table 4. Apartment unit block data.

Surface Area (m^2)	174.9
Volume (m^3)	570.25
Concrete Height Level (m)	3.25
Final Height Level (m)	3.3

2.1.3. Electromechanical and Renewable Energy Systems

In Table 5, the basic characteristics of the electromechanical systems that were in operation during this study are listed. Natural gas was used as a heating source—the boiler power capacity was set at 25 kW. Refrigerant heat transfer was regulated by heat pumps with a total power capacity of 9 kW. Domestic hot water generation was attained by utilizing both solar and electricity resources. Renewable energy systems were solely based on solar panels that stored water in a building roof tank. The respective details for the renewable energy system characteristics are listed in Table 6.

Table 5. Electromechanical systems data for the apartment unit.

System	Source	Distribution Network of Thermal Medium	Season	Power (kW)
Heating	Natural Gas	Yes	Winter	25
Cooling	Electricity	No	Summer	9
Hot Water	Solar/Electricity	No	Year-round	5

Table 6. Renewable energy systems data for the apartment unit.

Panel Angle (°)	Panel Surface Area (m^2)	Shade Coefficient	Orientation (°)
45	4	0.8	180

2.1.4. The Energy Efficiency Certification Software Package TEE-KENAK

The TEE-KENAK (1.31.1.19) software was developed by the Energy Saving Team of the Institute for Environmental and Sustainable Development Research of the National

Observatory of Athens in cooperation with the Technical Chamber of Greece [79,80]. This software applies appropriate algorithms for the valid calculation of the energy efficiency performance of buildings in Greece. It is based on a methodology that is outlined in the European-adopted energy performance standards (EN ISO 13790), as well as the relevant national standards for heating and cooling spaces, in conjunction with the Technical Guides of the Technical Chamber of Greece. According to the assessed output report from the TEE-KENAK software, a building apartment is officially certified to an energy efficiency class. The TEE-KENAK software requires entering basic factual information about the building characteristics. Formal input information includes the owner identification, the ownership status, the building address, and the year of issuance of the building permit. Technical details regarding the surfaces and volumes of the building (total surface area, heated surface area, cooled surface area, etc.) as well as information about the energy sources of the building are also keyed-in online. Opaque elements of the building are described by providing information on the orientation, the surface area, the coefficient of thermal permeability, and the pertinent shadowed area pattern. Dimensioning of the opaque surfaces with respect to the building foundations also involves lower and upper height details for each relevant structural element. Data for the transparent building surfaces additionally include glazing conditions, thermal break options, pane gap details, etc. The TEE-KENAK software package is compartmentalized to receive factual information with regards to the installed electromechanical systems in the apartment unit by considering energy demands for heating, cooling, and domestic hot water generation equipments, while also opting for renewable energy sources such as solar energy collection systems. The input data indicate the electromechanical system power demands/consumption, which are supplemented with their associated efficiency performance ratings. To estimate the solar collector contribution, specific details such as the usage rate, the effective surface area, and the optimal tilt and orientation angles are also submitted.

2.2. The Statistical Analysis Approach

The TEE-KENAK software platform permits a large number of relevant parameters to be synchronously investigated, whenever the scope of a study is to undertake energy-consumption reduction recommendations. This provision makes it impractical to test the energy consumption response by either a trial-and-error or a full factorial method. In fact, it was remarked [91] that succeeding in modeling the energy performance of an apartment unit is not adequate, if the modelling aspect is restricted solely to the installed electromechanical systems. It would not be convenient to carry out the resulting simulated estimations without resolving the drawback of voluminous (poly-factorial) combinatorial computations. Simulated predictions ought to be practical to allow for obtaining them on a per project basis and as part of an engineer's daily routine if they are to be deemed helpful. It should be reiterated that this difficulty stemmed from the fact that, in attempting to simultaneously screen the tendencies of as many as 24 electromechanical controlling factors [91]—each factor minimally adjusted to two settings—the trial volume demand blew up (2^{24} = 16,777,216 trials). The experimental tactic of resorting to supersaturated designs befits the condition to drastically compress the trial schedule. Therefore, the 14-run, 24-parameter supersaturated design of Williams [87], with its versatile parameter screening acceleration properties [89], which was implemented for the lean experimental data collection of Rousali and Besseris [91], is determined to be an attractive sampling planner for this work, as well. The adopted supersaturated design class is modified factorial half-fractions [88], which may also include the special case of half-split Plackett–Burman [93] design matrices. The 24 controlling factors which will be accommodated in the 14-run, 24-parameter supersaturated screening design have been tabulated in Table 7.

Table 7. Controlling factors and their settings for influencing the energy consumption of the residential unit.

Coded	Factors	Natural Gas Boiler(−)	Petroleum Boiler(−)	Natural Gas Boiler(+)	Petroleum Boiler(+)
F1	Automation for hot water	no		yes	
F2	Category of automatic control	A		D	
F3	Number of ceiling fans	0		5	
F4	Energy source for heating systems	gas			petroleum
F5	Efficiency of power generation of heating systems	0.977	0.9	0.955	0.84
F6	Passage of distribution network of heating systems	externally		internally	
F7	Efficiency of terminal units of heating systems	0.89		0.93	
F8	Type of cooling systems	air cooled		water cooled	
F9	Power of cooling systems (KW)	6		9	
F10	Efficiency of power generation of cooling systems (EER)	2.5		5.3	
F11	Efficiency of terminal units of cooling systems	0.9		0.96	
F12	Recirculation of distribution network (yes or no)	yes		no	
F13	Efficiency of domestic hot water storage system	1	0.98	0.98	0.93
F14	Type of solar panels	Simple		Vacum	
F15	Surface area of solar panels	2		4	
F16	Utilization rate of solar radiation for domestic hot water	0.344		0.38	
F17	Thermal insulation of walls	yes		no	
F18	Installation of awnings	yes		no	
F19	Presence of shutters	yes		no	
F20	Thermal insulation of roof	yes		no	
F21	Type of window's frame	wooden		metallic	
F22	Air gap between glasses	6 mm		12 mm	
F23	Percentage of window frame	30%		20%	
F24	Type of exit door	Thermal Insulation		No thermal insulation	

For the stepwise regression analysis procedure [94], the basic model is defined by the predictor matrix, **X** = $\{X_{ij}\}$, of size $n \times m$, where the number of the supersaturated design explanatory variables is m and the number of supersaturated design recipes is n. Then, the response matrix **Y**, of size $n \times r$, where the number of responses is r = 1 for this work (energy consumption), and in conjunction with the prescribed condition for design supersaturation in m regressors, i.e., $m > n + 1$, is written as:

$$Y_i = \beta_o + \sum_{j=1}^{m} \beta_j X_{ij} + \epsilon_i$$

where β_o and β_j, with $1 \leq j \leq m$, symbolize the coefficients of regression and ϵ_i is denoted as the error term for $1 \leq i \leq n$, which is assumed to be an independent and identically distributed random normal variable. The stepwise regression method uses forward sequences of F-test applications, but the model selection technique alternatives will include assessments which consider: (1) the adjusted coefficient of determination (adj R^2), (2) the Bayesian information criterion [95], and (3) the Mallow's C_p metric [96] for best subsets regression.

To use the latent variable approach in the partial least squares (PLS) model [97,98], the number of the supersaturated design explanatory variables remains in the formalism m, the number of responses is r (r = 1), and the number of supersaturated design recipes is n. Then, the predictor matrix **X**, of size $n \times m$, and the response matrix **Y**, of size $n \times r$, are prescribed for $m > n + 1$:

$$\mathbf{X} = \mathbf{C}\mathbf{L}_X{}^T + \mathbf{E}_X$$

$$\mathbf{Y} = \mathbf{D}\mathbf{L}_Y{}^T + \mathbf{E}_Y$$

The projection matrices of **X** and **Y** are defined as **C** and **D**, respectively, and they are both of size $n \times p$. The orthogonal loading matrices $\mathbf{L}_X$ and $\mathbf{L}_Y$ correspond to the matrices **X** and **Y**, with dimensions $m \times p$ *and* $r \times p$, respectively. The error terms, $\mathbf{E}_X$ and $\mathbf{E}_Y$, corresponding to the respective **X** and **Y** matrix models, are assumed to be independent and identically distributed random normal variables. The subsequent maximization of the covariance of the matrices **C** and **D** permits the decomposition of the matrices **X** and **Y**.

2.3. The Computational Aids

Descriptive statistics (median, interquartile range, skewness, and kurtosis) were computed per a factorial-setting basis using the 'Frequencies' selection from the IBM SPSS (v.29) software package. The 'stepwise' regression analysis was used to create a model summary with the statistically strong controlling factors using the probability of the F-distribution to sieve through the effects while adjusting the stepping method criteria for entry and removal at the α levels of 0.05 and 0.10, respectively. The model coefficients were supplemented with estimates of the coefficients of determination, collinearity diagnostics, Durbin–Watson residuals testing [99], and P-P plot residuals assessment.

To obtain the main effects plot for the 24-factor, 14-run supersaturated energy-consumption response dataset, the function 'maineffectsplot()' (MATLAB (R2022b) software package) was utilized after individually entering the poly-factorial combinations through the function 'table2array()' of the statistical freeware platform R [100].

The data analysis included the factorial coefficient predictions that were obtained from the partial least squares treatment (IBM SPSS v.29), which offer additional information such as the proportion of the explained variance by the investigated regressors, while testing the variable importance in the latent factorial projection. The selection cut of the strong predictors was also assessed using the Mallow's C_p metric. Therefore, the best-subsetting factorial combinations of the supersaturated energy consumption dataset were determined using the linear regression outcomes from two R-packages (v.4.1), 'leaps()' (v.3.1) and 'StepReg()' (v.1.4.4).

The clustering analysis (IBM SPSS v.29) was implemented to refine the grouping tendency of the four summary estimators (median, interquartile range, skewness, and kurtosis) of the supersaturated dataset using an unsupervised classifier [101]. Correlation results (Spearman's ρ coefficient [102]) among the four measures were pre-screened (IBM SPSS v.29). The goodness of clustering was obtained by estimating the gap statistic [103], using the function 'clusGap()' from the R-package 'cluster()' (v.2.1.2). Furthermore, to affirm the optimal cluster size, the auto-clustering method of the Schwarz's Bayesian Criterion (BIC) was also employed (IBM SPSS v.29). The BIC change, the ratio of the BIC changes, and the ratio of the distance measures were computed. Hierarchical cluster sequence identifications were provided through a dendrogram. The statistical significance of the clustered groups of the four summary estimators was computed using the Mann–Whitney test [104] (IBM SPSS v.29).

The entropic screening of the predominant effects was carried out using the bin-discretization of the supersaturated energy-consumption dataset per a factorial setting basis. After partitioning in bins using the function 'discretize2d()' from the R-package 'entropy()' (v.1.3.1), the empirical (Shannon) mutual information [105] between same factor setting pairs was computed using the function 'mi.empirical()' from the R-package 'entropy()' (v.1.3.1). Supplementary information corrections, via the optimal shrinkage intensity estimations, were completed by implementing the function 'entropy.shrink()' from the R-package 'entropy' (v.1.3.1)).

2.4. The Methodological Outline

The methodology may be recapitulated in brief in the following steps:

(1) Gather the required building apartment unit structural layout designs, along with the information for the installed electromechanical and renewable energy equipment information.
(2) Determine which featured characteristics will be investigated for the selected apartment unit.
(3) Determine the range values for the featured apartment unit characteristics and code them into controlling factor levels.
(4) Select an appropriate supersaturated screening design to accommodate the large number of controlling factors from steps 2 and 3.
(5) Execute the supersaturated plan runs by inputting each time trial recipe information (from step 4) into the TEE-KENAK software package.
(6) Record the energy consumption (real and reference) estimates from each supersaturated trial run.
(7) Prepare the response table and response graph for the energy consumption estimates.
(8) Conduct stepwise regression analysis and evaluate the model summary results.
(9) Determine the active controlling factors from step 8 and suggest a possible solution for the factorial settings.
(10) Confirm the energy consumption performance improvement by inputting the optimal solution into the TEE-KENAK software package.

(11) Assess and discuss the overall solution using other known methods such as PLS, entropic, and hierarchical clustering comparisons on key descriptive estimators of the energy consumption response.

3. Results

The results from executing the 14 poly-factorial recipes are tabulated in Table 8 in terms of the energy consumption response, along with its respective energy class. The energy consumption values ranged from 105.9 (trial #12) to 321.7 (trial #13) kWh/m^2. Meanwhile, the corresponding energy classification ratings varied from status categories of 'B' (0.75–1.0 ratio to the reference building's energy consumption) to 'Z' (2.27–2.73 times the reference building's energy consumption). In Table 9, the energy demands and consumption details are indicatively tabulated for the first trial of the supersaturated trial-design schedule; they result from loading the software platform TEE KENAK 1.31.1.19 with the input from the prescribed recipe. From the ensuing response table (Table 10), it is observed that the factorial variability of the energy consumption performance declines from 102.8 (F4) to 2.11 (F16) kWh/m^2. The two leading factors which contributed to the magnitude of the variability are: (1) the energy source for heating system's power generation (F4), and (2) the thermal insulation of the roof (F20). This behavior becomes more transparent in the response graph (Figure 3 (MATLAB R2022b)), where the optimal settings are identified at the lower levels of both factors. Factor F4, adjusted at the 'gas' setting, reduces the energy consumption estimation down to as low as 171.07 kWh/m^2 (an intra-factorial difference of 102.8 kWh/m^2). Similarly, the 'insulated roof' option of factor F20 reduces the energy consumption to 179.87 kWh/m^2 (an intra-factorial difference of 85.2 kWh/m^2). From the response graph, it is apparent that factors such as F2, F3, F5, F17, F21, and F22 might also be statistically assessed for their contributing effects to the overall improved performance of the energy consumption response.

Table 8. The response output for energy consumption (EC) in kWh/m^2 and its energy classification status.

Run #	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Energy Consumption Response	252.1	115.1	257.1	292	299.8	268.6	227.3	144.9	157.6	247.8	230.1	105.9	321.7	194
Class	E	B	E	E	Z	E	D	C	C	E	D	B	Z	D

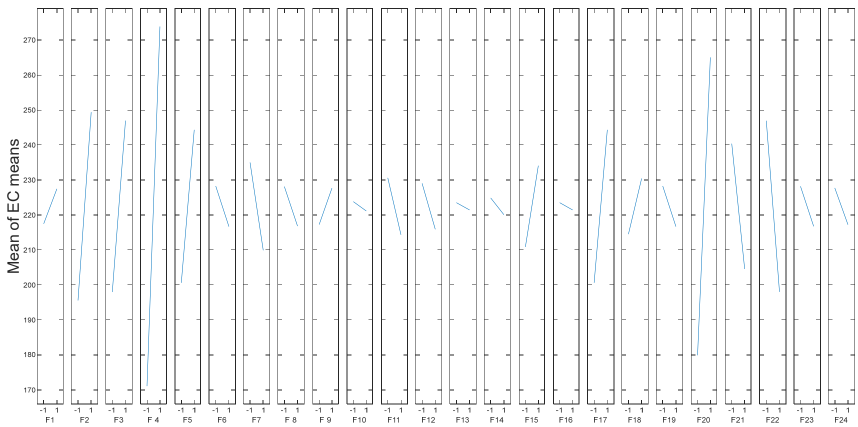


Figure 3. Response graph (MATLAB R2022b) for the poly-factorial profiling of the energy consumption (EC) in kWh/m^2.

Table 9. Energy demands and consumption results for the first simulation trial according to the supersaturated design schedule.

Energy Demand (kWh/m^2)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Heating	44.1	35.3	26.3	4.6	0	0	0	0	0	0	17.8	37.2	165.3
Cooling	0	0	0	0	1	11.7	24.3	20.4	1.5	0	0	0	58.9
Hot Water	2.1	1.9	2.1	1.8	1.6	1.3	1.2	1.2	1.3	1.6	1.8	2	19.9
Energy Consumption (kWh/m^2)	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Total**
Heating	55.6	44.5	33.1	5.8	0	0	0	0	0	0	22.5	46.8	208.3
Cooling	0	0	0	0	0.1	1.3	2.6	2.2	0.2	0	0	0	6.4
Hot Water	1.7	1.5	1.5	1.2	1	0.7	0.5	0.5	0.7	1	1.3	1.6	13.2
Hot Water (from solar)	0.4	0.4	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.5	0.4	0.3	6
Total	57.7	46.4	35.1	7.5	1.7	2.6	3.7	3.3	1.5	1.5	24.2	48.7	233.9

Table 10. Response table for the factorial profiling of the energy consumption (EC) in kWh/m^2.

Factor/ Setting	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
1	217.47	195.56	197.95	171.07	200.58	228.28	235.03	228.13	217.24	223.81	230.61	229.03
2	227.47	249.39	246.9	273.87	244.36	216.66	209.91	216.81	227.7	221.13	214.32	215.91
Range	10	53.83	48.95	102.8	43.78	11.62	25.12	11.32	10.46	2.68	16.29	13.12
Rank	20	3	5	1	6	15	9	17	19	22	11	13
Factor/ Setting	**F13**	**F14**	**F15**	**F16**	**F17**	**F18**	**F19**	**F20**	**F21**	**F22**	**F23**	**F24**
1	223.53	224.85	210.86	223.52	200.61	214.38	228.29	179.87	240.39	246.9	228.21	227.7
2	221.41	220.1	234.1	221.41	244.3	230.43	216.66	265.07	204.56	198	216.73	217.23
Range	2.12	4.75	23.24	2.11	43.69	16.05	11.63	85.2	35.83	48.9	11.48	10.47
Rank	23	21	10	24	7	12	14	2	8	4	16	18

A typical linear regression analysis for the supersaturated dataset is presented in Table 11 (IBM SPSS v.29). The 'stepwise' method was selected with entry and removal probabilities—to pace the stepping criteria—for the F-test values, at cutoff points of 0.05 and 0.10, respectively. The model summary indicates that four factors should be retained in the active factor group (F2, F4, F17, and F20); these adequately account for 96.2% of the total variation, according to the adjusted coefficient of determination (adj R^2). Adding the last influence, F17, to the model corrected the prediction by only 0.041%; the factor F17 was found to be statistically significant at a level of 0.05. Overall, the four-factor prognostication is also statistically significant at a Bonferroni-corrected familywise error rate of 0.05. The two statistically stronger factors, F4 and F20, contributed 60% and 29% to the total variation, respectively. The Durbin–Watson statistic was estimated at a value of d = 3.17 (>dU = 2.296); it does not provide any hint that the successive error terms might be positively autocorrelated. Moreover, the alternative estimation, 4-d (= 0.833), is within the critical value interval (0.505, 2.296), as computed for the test parameters n = 14, k′ = 5, and α = 0.05. Hence, the test for the presence of a negative autocorrelation is inconclusive. The normal P–P plot (IBM SPSS v.29) of the regression-analysis standardized residuals (Figure 4) does not reveal any detectable abnormalities. The model coefficient and collinearity statistics (IBM SPSS v.29) are listed in Table 12. The unstandardized/standardized coefficients of the four active factors are statistically significant to at least an error rate of 0.01. Moreover, the variance inflation factor (VIF) has been estimated to a maximum value of 1.35. Thus, there seem to be no apparent multicollinearity tendencies across effects, with respect to the two leading factors (F4 and F20) in particular.

Table 11. Stepwise-regression model summary (IBM SPSS v.29) for selecting statistically strong controlling factors.

Model [e]	R	R^2	Adjusted R^2	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R^2 Change	F Change	df1	df2	Sig. F Change	
1	0.775 [a]	0.600	0.567	45.30	0.600	18.024	1	12	0.001	
2	0.943 [b]	0.889	0.868	24.97	0.288	28.485	1	11	<0.001	
3	0.966 [c]	0.933	0.912	20.38	0.044	6.516	1	10	0.029	
4	0.987 [d]	0.974	0.962	13.41	0.041	14.100	1	9	0.005	3.167

[a] Predictors: (Constant), F4; [b] Predictors: (Constant), F4, F20; [c] Predictors: (Constant), F4, F20, F2; [d] Predictors: (Constant), F4, F20, F2, F17; [e] Dependent Variable: EC.

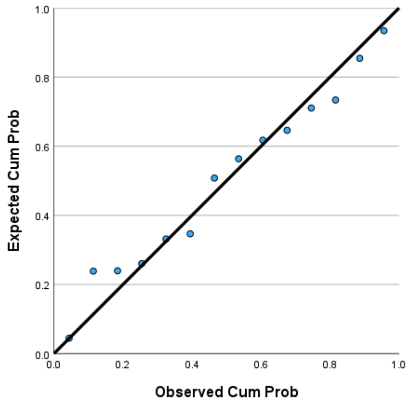


Figure 4. Normal P–P plot of the regression standardized residuals for the dependent variable EC (IBM SPSS v0.29).

Table 12. Stepwise-regression model summary coefficients and collinearity statistics (IBM SPSS v.29) for the statistically active controlling factors.

Model [a]		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	222.471	12.107		18.375	<0.001	196.092	248.851		
	F4	51.400	12.107	0.775	4.245	0.001	25.021	77.779	1.000	1.000
2	(Constant)	222.471	6.674		33.332	<0.001	207.781	237.162		
	F4	46.258	6.744	0.697	6.860	<0.001	31.416	61.101	0.980	1.021
	F20	35.992	6.744	0.543	5.337	<0.001	21.149	50.834	0.980	1.021
3	(Constant)	222.471	5.447		40.842	<0.001	210.335	234.608		
	F4	39.014	6.192	0.588	6.300	<0.001	25.217	52.811	0.774	1.292
	F20	39.285	5.653	0.592	6.950	<0.001	26.690	51.880	0.929	1.077
	F2	15.806	6.192	0.238	2.553	0.029	2.009	29.603	0.774	1.292
4	(Constant)	222.471	3.584		62.074	<0.001	214.364	230.579		
	F4	35.781	4.164	0.539	8.593	<0.001	26.361	45.201	0.741	1.350
	F20	38.207	3.730	0.576	10.242	<0.001	29.769	46.646	0.923	1.083
	F2	19.039	4.164	0.287	4.572	0.001	9.619	28.459	0.741	1.350
	F17	14.007	3.730	0.211	3.755	0.005	5.569	22.446	0.923	1.083

[a] Dependent Variable: EC.

In Table 13, the optimal setting recommendations—in conjunction with the results of Figure 3—are summarized in terms of the four leading recommended modifications: (1) the thermal insulation of the roof, (2) the gas-sourced heating systems, (3) the automatic control category type 'A', and (4) the thermal insulation of the walls. The remaining 20 weaker factors may be adjusted by also considering practical/economic implications, or left at their original conditions. A complete final profiled-factor solution is shown in Table 13. At this point, it is worthwhile to assess the results by comparing the 'before and after' benefits that may be realized from this rudimentary study. The original (certifiable) energy consumption performance, as it was computed by the TEE KENAK 1.31.1.19 software program, was found to be 133.8 kWh/m^2 (energy status 'C'), and was to be contrasted against a reference building estimation of 129.2 kWh/m^2. After completing the screening/optimization work, the recommended factorial settings from Table 13 were input to the TEE KENAK 1.31.1.19 software program to confirm any accruing energy savings. The improved solution delivered an energy consumption projection of 110.4 kWh/m^2 (energy status 'B') for the apartment, which was to be contrasted against a reference building estimation of 125.9 kWh/m^2. This is an almost 18% reduction in energy consumption, which may be considered satisfactory given the fact that only a subset of the total available variables in the TEE KENAK 1.31.1.19 software program was actually studied in this paradigm.

Table 13. Optimal settings for several key controlling factors (Key Settings). Combination solution for all controlling factors (Full Settings).

Key Settings					
	Factors	**Natural Gas Boiler(−)**	**Petroleum Boiler(−)**	**Natural Gas Boiler(+)**	**Petroleum Boiler(+)**
1	**Automation for hot water**	no		yes	
2	**Category of automatic control**	A		D	
3	**Number of ceiling fans**	0		5	
4	**Energy source for heating systems**	gas			petroleum
5	**Efficiency of power generation of heating systems**	0.977	0.9	0.955	0.84
6	**Passage of distribution network of heating systems**	externally		internally	
7	**Efficiency of terminal units of heating systems**	0.89		0.93	
8	**Type of cooling systems**	Air-cooled		Water-cooled	
9	**Power of cooling systems (KW)**	6		9	
10	**Efficiency of power generation of cooling systems (EER)**	2.5		5.3	
11	**Efficiency of terminal units of cooling systems**	0.9		0.96	
12	**Recirculation of distribution network (yes or no)**	yes		no	
13	**Efficiency of domestic hot water storage system**	1	0.98	0.98	0.93
14	**Type of solar panels**	Simple		Vacuum	
15	**Surface area of solar panels**	2		4	
16	**Utilization rate of solar radiation for domestic hot water**	0.344		0.38	
17	**Thermal insulation of walls**	yes		no	
18	**Installation of awnings**	yes		no	
19	**Presence of shutters**	yes		no	
20	**Thermal insulation of roof**	yes		no	
21	**Type of window's frame**	wooden		metallic	
22	**Air gap between glasses**	6 mm		12 mm	
23	**Percentage of window frame**	30%		20%	
24	**Type of exit door**	Thermal Insulation		No insulation	

Full Settings											
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
-	-	-	-	-	+	+	+	-	+	+	+
13	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
+	+	-	+	-	-	+	-	+	+	+	+

4. Discussion

Supersaturated datasecorrelation coefficients and their respective ts require probing by a multitude of statistical techniques. The partial least squares (PLS) method is an alternative approach that might offer additional information about the validity of the factorial screening results, which were obtained in the preceding section. The proportion of the explained variance, using a maximum of five latent variables, is shown in Table 14 (IBM SPSS v.29). The output solution in Table 15 (IBM SPSS v.29) lists the factorial coefficients and their corresponding variable importance in the projection using the five latent factors. A practical roundup cut includes the candidate controlling factors F2, F3, F4, F5, F17, F20, and F22. In this profiling, only the solution factor members that were suggested in the Results section are considered.

Table 14. The proportion of the explained variance for the supersaturated dataset using the PLS method.

	X Variance	Cumulative X Variance	Y Variance	Cumulative Y Variance (R^2)	Adjusted R^2
1	0.081	0.081	0.984	0.984	0.983
2	0.063	0.144	0.014	0.998	0.998
3	0.083	0.227	0.001	1.000	0.999
4	0.074	0.301	0.000	1.000	1.000
5	0.033	0.333	4.570×10^{-5}	1.000	1.000

Using the stepwise elimination option of PLS, the ANOVA treatment diagnostics (IBM SPSS v.29) reaffirm the high confidence to the previously recommended four-factor model (Table 16). To further study the possible inclusion of additional contributions, the Mallows's C_p metric is utilized to reassess the linear regression results using the best subsets approach (R-packages 'leaps()' (v.3.1) and 'StepReg()' (v.1.4.4)). The suggested solution (F2, F4, F17, F22) achieves an adjusted R^2 value of 96.2% and a corresponding C_p value of 28.1. Adding as many as five extra regressors (F5, F7, F11, F21, and F22) via the best subsetting approach increases the adjusted R^2 to a value of 98.9% and, thus, substantially reduces the corresponding C_p value to 10.3. However, a 99% confidence interval estimation for the adjusted R^2 in the original solution also includes the latter prediction. It is inferred that the small supersaturated dataset may not allow for discerning the need for additional predictors by relying only on the C_p criterion.

Table 15. The factorial coefficients and the variable importance (latent factors) in the projection for the supersaturated dataset using the PLS method.

Controlling Factors	PLS Coefficients	Latent Factors *				
		1	2	3	4	5
F1	3.208	0.268	0.270	0.270	0.271	0.271
F2	12.227	1.443	1.436	1.436	1.436	1.436
F3	13.633	1.315	1.308	1.307	1.307	1.307
F4	26.549	2.756	2.738	2.736	2.736	2.735
F5	12.178	1.174	1.170	1.169	1.169	1.169
F6	−1.606	0.312	0.313	0.317	0.318	0.318
F7	−7.509	0.673	0.669	0.669	0.670	0.670
F8	−4.134	0.303	0.324	0.324	0.324	0.324
F9	0.856	0.280	0.312	0.313	0.313	0.313
F10	−1.585	0.072	0.102	0.137	0.145	0.145
F11	−6.211	0.437	0.452	0.454	0.454	0.454
F12	0.563	0.352	0.469	0.470	0.470	0.470
F13	0.491	0.057	0.095	0.098	0.107	0.107
F14	−3.333	0.128	0.188	0.192	0.193	0.193
F15	4.502	0.623	0.636	0.636	0.636	0.636
F16	0.491	0.057	0.095	0.098	0.107	0.107
F17	10.085	1.172	1.164	1.166	1.167	1.167
F18	3.653	0.427	0.427	0.428	0.428	0.428
F19	−2.059	0.312	0.313	0.313	0.313	0.313
F20	25.524	2.284	2.281	2.280	2.280	2.280
F21	−7.199	0.961	0.968	0.967	0.967	0.967
F22	−13.994	1.312	1.306	1.306	1.305	1.305
F23	−3.650	0.308	0.308	0.307	0.308	0.308
F24	−1.843	0.281	0.284	0.284	0.284	0.284

* Cumulative Variable Importance.

Table 16. ANOVA results for the stepwise PLS treatment of the supersaturated dataset.

Model [a]		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	36,987.440	1	36,987.440	18.024	0.001 [b]
	Residual	24,625.969	12	2052.164		
	Total	61,613.409	13			
2	Regression	54,752.927	2	27,376.463	43.895	<0.001 [c]
	Residual	6860.482	11	623.680		
	Total	61,613.409	13			
3	Regression	57,459.467	3	19,153.156	46.108	<0.001 [d]
	Residual	41,53.941	10	415.394		
	Total	61,613.409	13			
4	Regression	59,994.975	4	14,998.744	83.407	<0.001 [e]
	Residual	1618.433	9	179.826		
	Total	61,613.409	13			

[a] Dependent Variable: EC; [b] Predictors: (Constant), F4; [c] Predictors: (Constant), F4, F20; [d] Predictors: (Constant), F4, F20, F2; [e] Predictors: (Constant), F4, F20, F2, F17.

A convenient way to filter out the weak effects is to exploit the limiting dichotomous nature of the supersaturated dataset and the advantageously large number of the studied predictor variables. Consequently, essential information about the behavior of the energy consumption response may be synopsized by at least four statistical descriptive measures including the data location, dispersion, skewness, and flatness. The four respective statistical estimators, the median (M), the interquartile range (I), the skewness (S), and the kurtosis (K), re-organized the collected supersaturated dataset in terms of 48 individual factor levels. The summarized supersaturated EC dataset is listed in Table 17 (IBM SPSS v.29) in a data-reduced form and tabulated per each individual factor level. Before advancing the information generation process, the extent of potential correlations among the four summary estimators should be assessed. From Table 18, it can be observed that the two-variable correlation estimations, according to the Spearman's ρ coefficient, returned three statistically significant outcomes ($p < 0.001$), in three out of the six possible pairings. The three detected estimator relationships were between: (1) the median and the skewness, (2) the interquartile range and the kurtosis, and (3) the skewness and the kurtosis. However, the magnitudes of the coefficients for the median–skewness and skewness–kurtosis pairs may not be considered strong. It is noteworthy that the resulting correlation between the median–kurtosis estimator pair does not show up as significant. For data processing purposes, it is assumed that there exists a mediocre correlation between the median–skewness and the skewness–kurtosis estimator relationships. Thus, both are retained in the succeeding data analysis steps. The interquartile range–kurtosis estimator pair may be rated close to strong, whereas the lower magnitude of its correlation coefficient does not also exclude a mediocre rating ($|\rho| = 0.549$). Therefore, the interquartile range is maintained in the analysis as well.

At this stage, an unsupervised screening approach is employed to cluster all individual factor levels by their grouping tendency to enter different memberships. To initiate the clustering process, the optimal cluster number is sought by computing the goodness of the clustering measure, the gap statistic (function 'clusGap()' in R-package 'cluster()' v.2.1.2); the 'firstSEmax' method is selected to evaluate the partitioning around the medoids ('pam') by generating 1000 simulated reference sets. From Figure 5, the optimal cluster number is found to be two. This outcome is also affirmed by the direct auto-clustering result (IBM SPSS v.29), which utilizes the Schwarz's Bayesian Criterion (BIC) to locate the optimal change point of the measure (Table 19). From Figure 6, the cluster quality due to cohesion and separation is rated as 'Fair', according to the Silhouette measure evaluation (IBM SPSS v.29). Using hierarchical cluster analysis, the obtained identification cluster membership predictions are now tabulated in the right-hand side of Table 17.

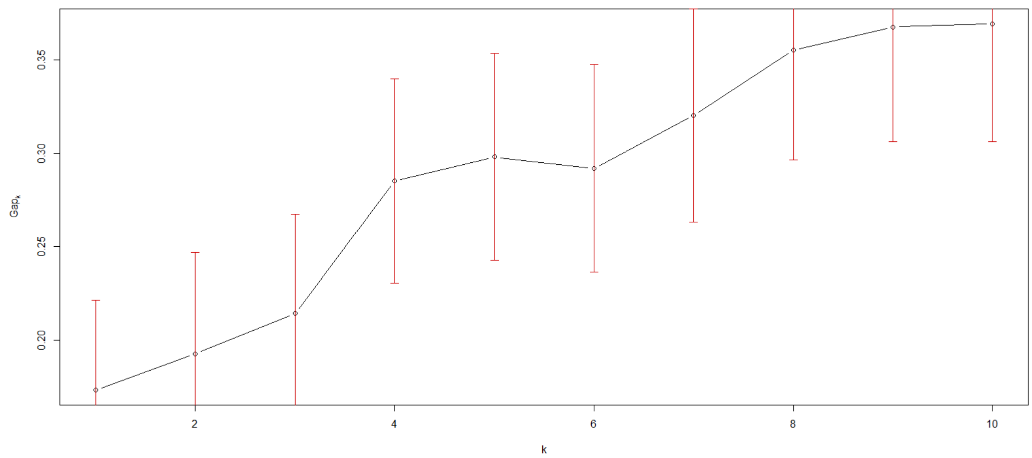


Figure 5. Gap statistic performance for profiling optimal clustering size for the summarized supersaturated dataset of Table 17.

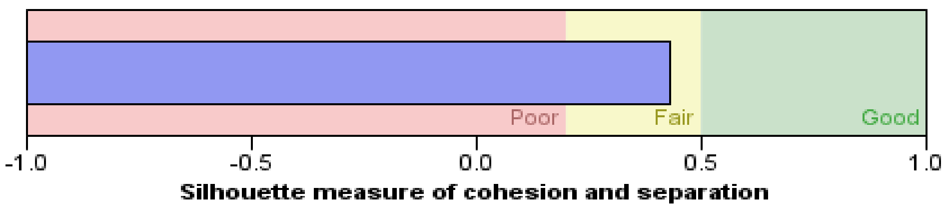


Figure 6. Cluster quality rating using the silhouette measure of cohesion and separation (IBM SPSS v.29).

Table 17. Summary statistics of the supersaturated dataset (median (M), interquartile range (I), skewness (S), kurtosis (K)) per factorial setting, and their hierarchical cluster identification.

Factor	Level	M	I	S	K	Cluster ID
F1	1	227.3	111.0	0.11	−1.06	1
	2	252.1	176.3	−0.81	−0.95	1
F2	1	194.0	152.9	0.15	−1.48	1
	2	252.1	61.9	−0.95	2.04	2
F3	1	227.3	141.4	−0.19	−1.71	1
	2	252.1	105.8	−0.67	−0.04	1
F4	1	157.6	111.6	0.35	−1.38	1
	2	268.6	52.0	0.19	−1.02	2
F5	1	194.0	107.2	0.12	−1.59	1
	2	257.1	72.5	−1.36	2.59	2
F6	1	230.1	63.1	−0.57	1.59	2
	2	247.8	147.1	−0.36	−2.03	1
F7	1	230.1	98.0	−0.48	−0.26	1
	2	247.8	152.9	−0.15	−1.74	1
F8	1	257.1	134.4	−0.52	−1.01	1
	2	230.1	107.2	−0.67	−0.44	1
F9	1	227.3	99.5	−0.14	−1.31	1
	2	252.1	184.1	−0.72	−1.12	1
F10	1	230.1	142.2	−0.11	−1.03	1
	2	252.1	123.7	−1.01	−0.43	1
F11	1	247.8	154.9	−0.55	−0.91	1
	2	230.1	111.0	−0.64	−0.55	1
F12	1	252.1	147.1	−0.86	−0.88	1
	2	227.3	99.5	−0.17	−0.04	1
F13	1	230.1	111.0	0.24	−0.76	1
	2	252.1	176.3	−0.84	−1.09	1
F14	1	252.1	154.9	−0.30	−2.11	1
	2	230.1	63.1	−1.61	2.72	2
F15	1	227.3	107.2	0.00	0.00	1
	2	257.1	134.4	−1.07	−0.32	1
F16	1	230.1	111.0	0.24	−0.76	1
	2	252.1	176.3	−0.84	−1.09	1
F17	1	194.0	176.3	0.35	−1.55	1
	2	252.1	41.3	−1.27	2.74	2
F18	1	247.8	112.2	−0.59	−0.86	1
	2	230.1	134.4	−0.51	−0.70	1
F19	1	252.1	142.2	−0.33	−1.20	1
	2	230.1	123.7	−0.87	−0.41	1
F20	1	157.6	132.1	0.16	−2.25	1
	2	268.6	72.5	−0.46	−0.59	2
F21	1	252.1	98.0	−0.55	−0.56	1
	2	227.3	152.9	0.04	−1.69	1
F22	1	247.8	64.7	−0.36	0.88	2
	2	194.0	152.9	0.06	−2.07	1
F23	1	247.8	74.6	−1.31	1.86	2
	2	230.1	154.9	−0.06	−1.73	1
F24	1	257.1	147.1	−0.41	−1.46	1
	2	230.1	94.5	−0.82	0.30	1

Table 18. Spearman’s ρ correlation coefficients and their respective 95% confidence intervals for median(M), interquartile range (I), skewness (S), and kurtosis (K) of the supersaturated dataset.

	Spearman’s ρ	**Significance(2-tailed)**	**95% Confidence Intervals (2-tailed) [a,b]**	
			Lower	**Upper**
M–I	−0.054	0.716	−0.341	0.242
M–S	−0.579	<0.001	−0.745	−0.346
M–K	0.280	0.054	−0.013	0.529
I–S	0.190	0.195	−0.108	0.457
I–K	−0.723	<0.001	−0.838	−0.546
S–K	−0.627	<0.001	−0.777	−0.410

[a] Estimation is based on Fisher’s r-to-z transformation. [b] Estimation of standard error is based on the formula proposed by Fieller, Hartley, and Pearson.

Table 19. Auto-clustering of the summarized supersaturated datasets (Table 17) using the Schwarz's Bayesian Criterion (BIC) (IBM SPSS v.29).

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change [a]	Ratio of BIC Changes [b]	Ratio of Distance Measures [c]
1	162.043			
2	154.178	−7.865	1.000	1.378
3	156.959	2.781	−0.354	1.864
4	172.807	15.849	−2.015	1.365
5	192.696	19.889	−2.529	1.456
6	216.057	23.361	−2.970	2.013
7	243.246	27.189	−3.457	1.072
8	270.690	27.444	−3.489	1.090
9	298.424	27.734	−3.526	1.082
10	326.404	27.980	−3.558	1.355
11	355.168	28.764	−3.657	1.551
12	384.715	29.547	−3.757	1.022
13	414.294	29.578	−3.761	1.074
14	443.968	29.674	−3.773	1.126
15	473.787	29.819	−3.791	1.214

[a] The changes are from the previous number of clusters in the table. [b] The ratios of changes are relative to the change for the two-cluster solution. [c] The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

The rationale is that, as long as the two settings for the same controlling factor are found to belong to the same cluster, then the regressor cannot be further considered as a viable predictor of the total solution. In other words, an alternative method is attempted to reduce the initial regressor list of the supersaturated EC dataset. The hierarchical cluster distribution for the combined four statistical-estimator groupings is shown in Table 20 (IBM SPSS v.29). A total of 81.3% of the members are identified as belonging to cluster '1′; hence, there is a clear asymmetry in the factor-level distribution. Such asymmetry has been elicited owing to the cluster variability in the mean estimation for the interquartile range, and also for both skewness and kurtosis, based on their mean and their standard error for their respective mean estimations (Table 20). In Figure 7, the dendrogram for the hierarchical clustering solution (using median linkage) demonstrates the inherently complicated configuration display of all 48 factor-levelled EC-response datasets. To ensure that the dichotomizing of all four summarizing estimators is meaningful, in Figure 8, the distribution of the datapoints for the median, interquartile range, skewness, and kurtosis is contrasted with their clustered memberships. Further, the comparison application outcomes of the Mann–Whitney test aid in inferring that the cluster separations, for all four estimators, are statistically significant at least on the typical 0.05 level.

Table 20. Hierarchical clustering and combined statistics for the four individual summarizing estimators (IBM SPSS v.29) from Table 17.

HIERARCHICAL		M	I	S	K
1	N	39	39	39	39
	Mean	231.782	133.336	−0.3441	−1.0326
	Std. Error of Mean	4.0194	4.1793	0.06563	0.10261
2	N	9	9	9	9
	Mean	250.478	62.856	−0.8556	1.4233
	Std. Error of Mean	4.6410	3.5568	0.19677	0.46687
Total	N	48	48	48	48
	Mean	235.288	120.121	−0.4400	−0.5721
	Std. Error of Mean	3.5261	5.2895	0.07009	0.18282

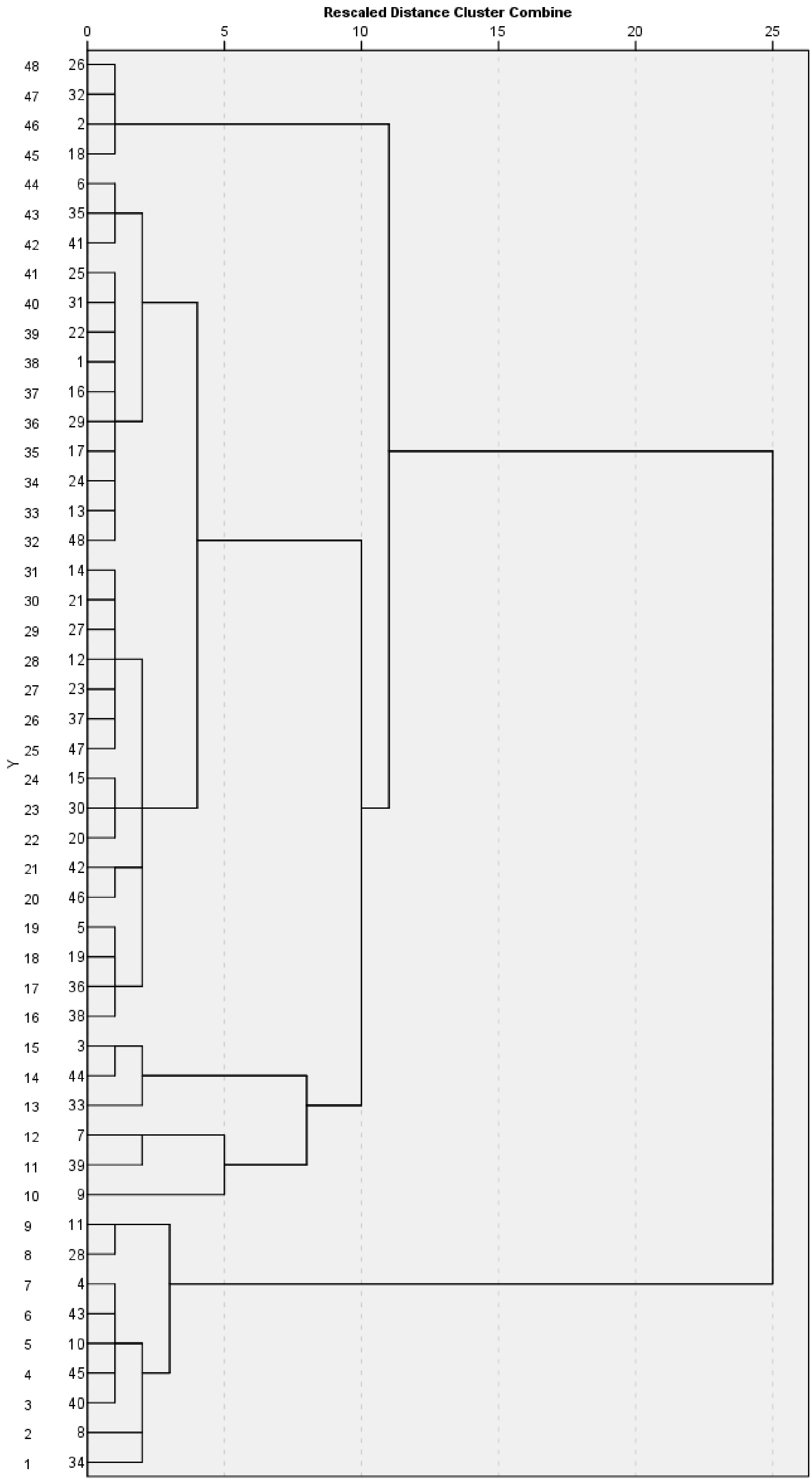
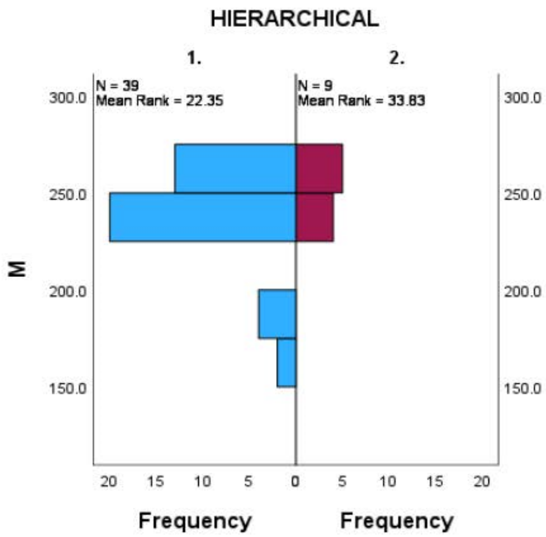


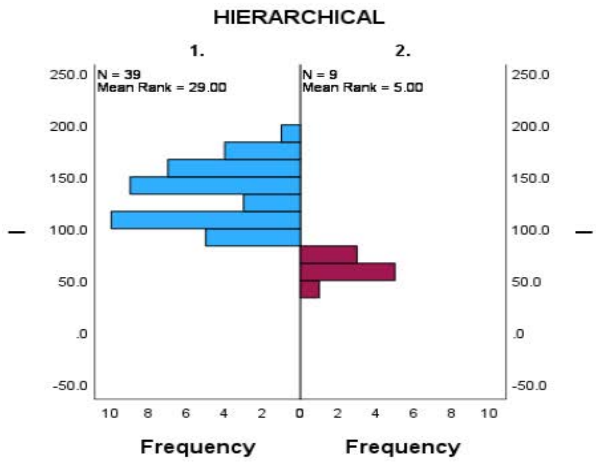
Figure 7. Dendrogram using median linkage for the 48 factor levels resulting from the summarized four-estimator supersaturated dataset (IBM SPSS v.29).



Independent-Samples Mann-Whitney U Test Summary

Total N	48
Mann-Whitney U	259.500
Wilcoxon W	304.500
Test Statistic	259.500
Standard Error	37.170
Standardized Test Statistic	2.260
Asymptotic Sig.(2-sided test)	0.024
Exact Sig.(2-sided test)	0.025

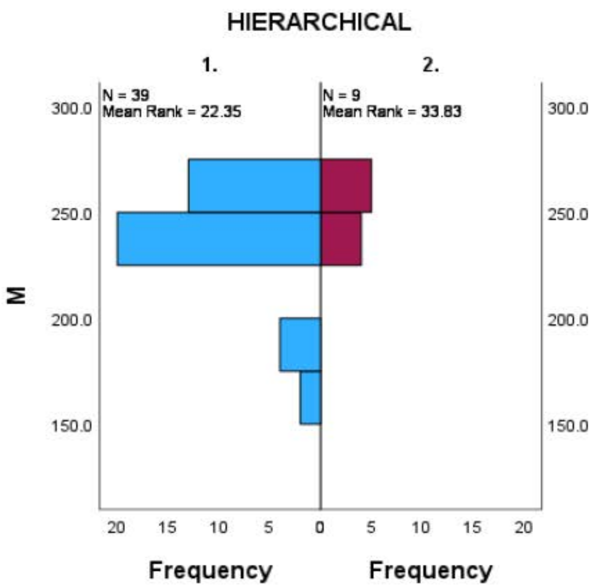
(A)



Independent-Samples Mann-Whitney U Test Summary

Total N	48
Mann-Whitney U	0.000
Wilcoxon W	45.000
Test Statistic	0.000
Standard Error	37.805
Standardized Test Statistic	−4.642
Asymptotic Sig.(2-sided test)	<0.001
Exact Sig.(2-sided test)	0.000

(B)

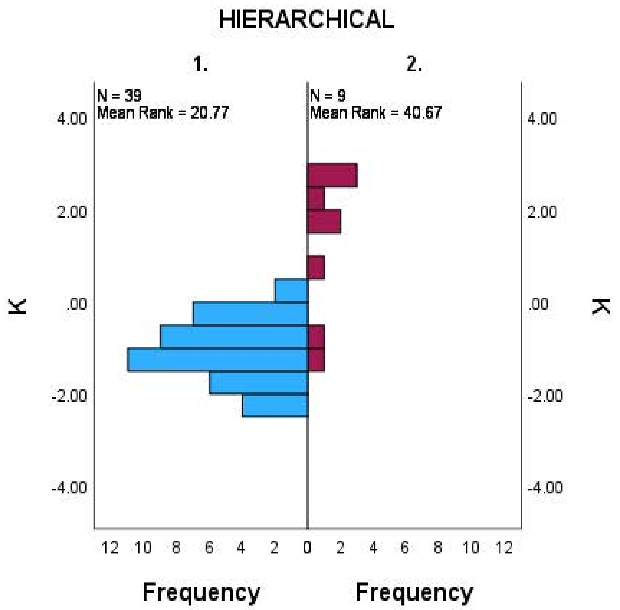


Independent-Samples Mann-Whitney U Test Summary

Total N	48
Mann-Whitney U	259.500
Wilcoxon W	304.500
Test Statistic	259.500
Standard Error	37.170
Standardized Test Statistic	2.260
Asymptotic Sig.(2-sided test)	0.024
Exact Sig.(2-sided test)	0.025

(C)

Figure 8. *Cont*.



Independent-Samples Mann-Whitney U Test Summary

Total N	48
Mann-Whitney U	321.000
Wilcoxon W	366.000
Test Statistic	321.000
Standard Error	37.855
Standardized Test Statistic	3.844
Asymptotic Sig.(2-sided test)	<0.001
Exact Sig.(2-sided test)	0.000

(D)

Figure 8. Individually contrasting the clustered supersaturated datasets for their four summarizing estimators: (**A**) median (M), (**B**) interquartile range (I), (**C**) skewness (S), and (**D**) kurtosis (K).

Accordingly, returning to Table 17, the reduced list of the nominated controlling factors, in this instance, includes F2, F4, F5, F6, F14, F17, F20, F22, and F23. Repeating the stepwise selection process on this group of regressors, the resultant factorial profile appears identical to the regression solution which was obtained in the previous section, i.e., F2, F4, F17, and F20.

Finally, to examine the preponderance of the leading controlling factors from a non-parametric viewpoint, an entropic approach is implemented. A convenient way to achieve this is to evaluate, for all individual controlling factors, the joint distribution of their paired settings. Due to the small number of data points that needs to be accommodated by a larger factorial base, the continuous EC dataset was discretized each time to reflect the response data that corresponded to each factorial setting. To render a common bin number for all computations, the Freedman–Diaconis formula [106] was used, i.e., # of bins = range_{EC}/h with $h = 2 \cdot \text{IQR}/\sqrt[3]{n}$ (IQR = interquartile range, n = number of EC response entries). Inputting the values of range_{EC} = 215.8 kWh/m^2, IQR = 99.03 kWh/m^2, and n = 14, the number of common bins was computed to be approximately three. Next, the synchronous two-setting discretization was conducted using the function 'discretize2d()' (R-package 'entropy()' (v.1.3.1)). Then, the empirical (Shannon) mutual information of the setting pairs was computed using the function 'mi.empirical()' (R-package 'entropy' (v.1.3.1)). The two controlling factors with the two lower mutual information estimations between settings, along with their lower-setting optimal shrinkage intensity estimations (function 'entropy.shrink()' from the R-package 'entropy' (v.1.3.1)), were found to be: (1) F20 (0.08 nats) with optimal shrinkage intensity lowered at 0.263 at the second level, and (2) F4 (0.202 nats) with optimal shrinkage intensity lowered at 0.263 at the second level.

5. Conclusions

The *Renovation Wave for Europe* is a great campaign under the *European Green Deal* which aims to effectuate the energy consumption and gas emission reduction of millions of older residential buildings in Europe. However, the energy performance improvement effort would be attained one apartment unit at a time. This complication might be remedied by adopting the mass customization philosophy that has been successfully espoused for several decades in industrial engineering. Consequently, a lean and green datacentric approach was attempted to carry out such a gradual improvement process, borrowing ideas and techniques from the Green Lean Six Sigma toolbox that is implemented to model and solve manufacturing problems. Therefore, the problem of examining the energy consumption reduction due to a candidate group of controlling factors was easily organized into a mini dataset with the aid of a supersaturated trial planner. The resulting empirical model is conveniently formulated, and the ensuing data analysis becomes manageable to expediently complete. The practicality of reaching a quick result is advantageous to an energy engineer. The proposed methodology demonstrated its responsiveness aspect toward attaining the goal of obtaining a solution quickly. The case study was a real-life standalone apartment unit that actually underwent a green certification process. By programming only 14 specific supersaturated recipe combinations for as many as 24 controlling factors, each factor was adjusted at two selected settings. The profiling relied on the variability potential that a factor carried to influence the energy consumption of the apartment unit. The novelty of this work is owed to the fact that it was the first time that shell properties and electromechanical system modifications were allowed to enter the empirical modelling effort. It is noted that it is a great advantage to prioritize the influence of the examined controlling factors on the standardized energy consumption ratings for two reasons. First, the generated dataset is trustworthy because it is based on derived estimations from a parametrized Energy Performance Certification database which is considered officially validated on a national level. Secondly, The Energy Performance Certification software platform (TEE-KENAK) generates actual and reference energy consumption ratings on the year-round performances of: (1) the original apartment unit and (2) on the apartment unit's modifications.

An assortment of main effects plots, best-subsets partial least squares, and entropic (Shannon) mutual information predictions formed the profiler engine of the proposed methodology. The statistical filtering of the proposed apartment shell properties and its electromechanical system modifications identified four strong effects: (1) the thermal insulation of the roof, (2) the gas-sourced heating systems, (3) the automatic control category type 'A', and (4) the thermal insulation of the walls. The prediction for the optimal energy consumption corresponds to 110.4 kWh/m^2 (energy status 'B') for the apartment. It accounts for an almost 20% reduction in energy consumption. Moreover, the 'greener' status rating has improved from the original 'C' status. Future work could involve forecasting the costs of apartment unit renovations and optimized predictions that combine economical and technical parameters, as well as occupant usage trends.

Author Contributions: Conceptualization, S.Z. and G.B.; methodology, S.Z. and G.B.; validation, S.Z. and G.B.; formal analysis, S.Z.; investigation, S.Z. and G.B.; resources, S.Z. and G.B.; writing—original draft preparation, G.B.; writing—review and editing, S.Z. and G.B.; supervision, G.B.; project administration, G.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data are available through Mr. Zarkadas' thesis as submitted to the Kingston University depository per ref. [92].

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

References

1. EC-COM(2019) 640 Final: The European Green Deal. Available online: https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF (accessed on 30 March 2021).
2. EC-COM(2020) 562 Final: Stepping Up Europe's 2030 Climate Ambition. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0562&from=EN (accessed on 3 April 2021).
3. EC-COM(2020) 662 Final: A Renovation Wave for Europe—Greening Our Buildings, Creating Jobs, Improving Lives. Available online: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:081:0018:0036:EN:PDF (accessed on 30 March 2021).
4. IEA. *Change in Final Energy Consumption by Sector, 2000–2018, and by Scenario to 2040*; IEA: Paris, France, 2019.

5. IEA. *Tracking Buildings*; IEA: Paris, France, 2019.
6. IEA; IRENA; UNSD; WB; WHO. *Tracking SDG 7: The Energy Progress Report*; WB: Washington, DC, USA, 2019.
7. United Nations Foundation. Sustainable Development Goals. 2023. Available online: https://unfoundation.org/what-we-do/issues/sustainable-development-goals/ (accessed on 30 January 2023).
8. Bond, S.; Devine, A. Incentivizing green single-family construction: Identifying effective government policies and their features. *J. Real Estate Financ. Econ.* **2016**, *52*, 383–407. [CrossRef]
9. Newsham, G.R.; Mancini, S.; Birt, B.J. Do LEED-Certified buildings save energy? Yes, but . . . . *Energy Build.* **2009**, *41*, 897–905. [CrossRef]
10. Adekanye, O.G.; Davis, A.; Azevedo, I.L. Federal policy, local policy and green building certifications in the U.S. *Energy Build.* **2020**, *209*, 109700. [CrossRef]
11. Porumb, V.A.; Maier, G.; Anghel, I. The impact of building location on green certification price premiums: Evidence from three European countries. *J. Clean. Prod.* **2020**, *272*, 122080. [CrossRef]
12. Sanderford, A.R.; McCoy, A.P.; Keefe, M.J. Adoption of energy star certifications: Theory and evidence compared. *Build. Res. Inf.* **2018**, *46*, 207–219. [CrossRef]
13. Kok, N.; McGraw, M.; Quigley, J.M. The diffusion of energy efficiency in building. *Am. Econ. Rev.* **2011**, *101*, 77–82. [CrossRef]
14. Zhang, L.; Li, R. Impacts of Green Certification Programs on Energy Consumption and GHG Emissions in Buildings: A Spatial Regression Approach. *Energy Build.* **2022**, *256*, 111677. [CrossRef]
15. Sołoducho-Pelc, L.; Sulich, A. Natural Environment Protection Strategies and Green Management Style: Literature Review. *Sustainability* **2022**, *14*, 10595. [CrossRef]
16. He, Q.; Hossain, M.U.; Ng, S.T.; Skitmore, M.; Augenbroe, G. A cost-effective building retrofit decision-making model-Example of China's temperate and mixed climate zones. *J. Clean. Prod.* **2021**, *280*, 124370. [CrossRef]
17. Wilson, C.; Pettifor, H.; Chryssochoidis, G. Quantitative modelling of why and how homeowners decide to renovate energy efficiently. *Appl. Energy* **2018**, *212*, 1333–1344. [CrossRef]
18. Jafari, A.; Valentin, V. An optimization framework for building energy retrofits decision-making. *Build. Environ.* **2017**, *115*, 118–129. [CrossRef]
19. Chantrelle, F.P.; Lahmidi, H.; Keilholz, W.; El Mankibi, M.; Michel, P. Development of a multicriteria tool for optimizing the renovation of buildings. *Appl. Energy* **2011**, *88*, 1386–1394. [CrossRef]
20. Jafari, A.; Valentin, V. Sustainable impact of building energy retrofit measures. *J. Green Build.* **2017**, *12*, 69–84. [CrossRef]
21. Pardo-Bosch, F.; Cervera, C.; Ysa, T. Key aspects of building retrofitting: Strategizing sustainable cities. *J. Environ. Manag.* **2019**, *248*, 109247. [CrossRef]
22. Ma, Z.; Cooper, P.; Daly, D.; Ledo, L. Existing building retrofits: Methodology and state-of-the-art. *Energy Build.* **2012**, *55*, 889–902. [CrossRef]
23. Fan, Y.; Xia, X. Building retrofit optimization models using notch test data considering energy performance certificate compliance. *Appl. Energy* **2018**, *228*, 2140–2152. [CrossRef]
24. Fan, Y.; Xia, X. Energy-efficiency building retrofit planning for green building compliance. *Build. Environ.* **2018**, *136*, 312–321. [CrossRef]
25. Foucquier, A.; Robert, S.; Suard, F.; Stephan, L.; Jay, A. State of the art in building modelling and energy performances prediction: A review. *Renew. Sustain. Energy Rev.* **2013**, *23*, 272–288. [CrossRef]
26. Asadi, E.; da Silva, M.G.; Henggeler Antunes, C.; Dias, L. Multi-objective optimization for building retrofit strategies: A model and an application. *Energy Build.* **2012**, *44*, 81–87. [CrossRef]
27. Castro, S.S.; Lopez, M.J.S.; Menendez, D.G.; Marigota, E.B. Decision matrix methodology for retrofitting techniques of existing buildings. *J. Clean. Prod.* **2019**, *240*, 118153. [CrossRef]
28. Andrade-Cabrera, C.; O'Dwyer, C.; Finn, D.P. Integrated cost-optimal residential envelope retrofit decision-making and power systems optimisation using Ensemble models. *Energy Build.* **2020**, *214*, 109833. [CrossRef]
29. Wu, R.; Mavromatidis, G.; Orehounig, K.; Carmeliet, J. Multi-objective optimisation of energy systems and building envelope retrofit in a residential community. *Appl. Energy* **2017**, *190*, 634–649. [CrossRef]
30. Schütz, T.; Schiffer, L.; Harb, H.; Fuchs, M.; Müller, D. Optimal design of energy conversion units and envelopes for residential building retrofits using a comprehensive MILP model. *Appl. Energy* **2017**, *185*, 1–15. [CrossRef]
31. Costa, A.; Keane, M.M.; Torrens, J.I.; Corry, E. Building operation and energy performance: Monitoring, analysis and optimization toolkit. *Appl. Energy* **2013**, *101*, 310–316. [CrossRef]
32. Kheiri, F. A review on optimization methods applied in energy-efficient building geometry and envelope design. *Renew. Sustain. Energy Rev.* **2018**, *92*, 897–920. [CrossRef]
33. Waibela, C.; Evins, R.; Carmeliet, J. Co-simulation and optimization of building geometry and multi-energy systems: Interdependencies in energy supply, energy demand and solar potentials. *Appl. Energy* **2019**, *242*, 1661–1682. [CrossRef]
34. Barber, K.A.; Krarti, M. A review of optimization based tools for design and control of building energy systems. *Renew. Sustain. Energy Rev.* **2022**, *160*, 112359. [CrossRef]
35. Nguyen, A.-T.; Reiter, S.; Rigo, P. A review on simulation-based optimization methods applied to building performance analysis. *Appl. Energy* **2014**, *113*, 1043–1058. [CrossRef]

36. Evins, R. A review of computational optimization methods applied to sustainable building design. *Renew. Sustain. Energy Rev.* **2022**, *160*, 112359.
37. Shi, X.; Tian, Z.; Chen, W.; Si, B.; Jin, X. A review on building energy efficient design optimization from the perspective of architects. *Renew. Sustain. Energy Rev.* **2016**, *65*, 872–884. [CrossRef]
38. Balali, A.; Yunusa-Kaltungo, A.; Edwards, R. A systematic review of passive energy consumption optimization strategy selection for building through multiple criteria decision-making techniques. *Renew. Sustain. Energy Rev.* **2023**, *171*, 113013. [CrossRef]
39. Machairas, V.; Tsangrassoulis, A.; Axarli, K. Algorithms for optimization of building design: A review. *Renew. Sustain. Energy Rev.* **2014**, *31*, 101–112. [CrossRef]
40. Kistelegdi, I.; Horvath, K.R.; Storcz, T.; Ercsey, Z. Building geometry as a variable in energy, comfort, and environmnetal design optimization-A review from the perspective of architects. *Buildings* **2022**, *12*, 69. [CrossRef]
41. Hashempour, N.; Taherkhani, R.; Mahdikhani, M. Energy performance optimization of existing buildings: A literature review. *Sustain. Cities Soc.* **2020**, *54*, 101967. [CrossRef]
42. Tian, Z.; Zhang, X.; Jin, X.; Zhou, X.; Si, B.; Shi, X. Towards adoption of building energy simulation and optimization for passive building design: A survey and a review. *Energy Build.* **2018**, *158*, 1306–1316. [CrossRef]
43. Abbas, S.; Saleem, O.; Rizvi, M.A.; Kazmi, S.M.S.; Munir, M.J.; Ali, S. Investigating the Energy-Efficient Structures Using Building Energy Performance Simulations: A Case Study. *Appl. Sci.* **2022**, *12*, 9386. [CrossRef]
44. Rong, A.; Su, Y. Polygeneration systems in buildings: A survey on optimization approaches. *Energy Build.* **2017**, *151*, 439–454. [CrossRef]
45. Gassar, A.A.A.; Koo, C.; Kim, T.W.; Cha, S.H. Performance optimization studies on heating, cooling and lighting energy systems of buildings during the design stage: A review. *Sustainability* **2021**, *13*, 9815. [CrossRef]
46. Azevedo, L.; Gomes, R.; Silva, C. Influence of model calibration and optimization techniques on the evaluation of thermal comfort and retrofit measures of a Lisbon household using building energy simulation. *Adv. Build. Energy Res.* **2021**, *15*, 630–661. [CrossRef]
47. Bale, C.S.S.; Varga, L.; Foxon, T.J. Energy and complexity: New ways forward. *Appl. Energy* **2015**, *138*, 150–159. [CrossRef]
48. Manson, S.M. Simplifying complexity: A review of complexity theory. *Geoforum* **2001**, *32*, 405–414. [CrossRef]
49. Menassa, C.C. Evaluating sustainable retrofits in existing buildings under uncertainty. *Energy Build.* **2011**, *43*, 3576–3583. [CrossRef]
50. Lopion, P.; Markewitz, P.; Robinius, M.; Stolten, D. A review of current challenges and trends in energy systems modeling. *Renew. Sustain. Energy Rev.* **2018**, *96*, 156–166. [CrossRef]
51. Priesmann, J.; Nolting, L.; Praktiknjo, A. Are complex energy system models more accurate? An intra-model comparison of power system optimization models. *Appl. Energy* **2019**, *255*, 113783. [CrossRef]
52. Pyzdek, T.; Keller, P. *The Six Sigma Handbook*; McGraw-Hill: New York, NY, USA, 2018.
53. George, M.; Rowlands, D.; Price, M.; Maxey, J. *The Lean Six Sigma Pocket Toolbook*; McGraw-Hill: New York, NY, USA, 2005.
54. Hussain, K.; He, Z.; Ahmad, N.; Iqbal, M.; Nazneen, S. Mapping green, lean, Six Sigma enablers through the lens of a construction sector: An emerging economy's perspective. *J. Environ. Plan. Manag.* **2023**, *66*, 779–812. [CrossRef]
55. Hussain, K.; He, Z.; Ahmad, N.; Iqbal, M.; Saeed, M.Z. Establishing a Green, Lean and Six Sigma implementation model for sustainable construction industry: An analysis of driving forces through ISM-MICMAC approach. *Environ. Sci. Pol. Res.* **2023**, *30*, 30462–30492. [CrossRef]
56. Womack, J.P.; Jones, D.T. *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*, 2nd ed.; Revised and Updated; Free Press: New York, NY, USA, 2003.
57. Dhingra, R.; Kress, R.; Upreti, G. Does lean mean green? *J. Clean. Prod.* **2014**, *85*, 1–7. [CrossRef]
58. Johansson, G.; Sundin, E. Lean and green product development: Two sides of the same coin? *J. Clean. Prod.* **2014**, *85*, 104–121. [CrossRef]
59. Garza-Reyes, J.A. Lean and green—A systematic review of the state of the art literature. *J. Clean. Prod.* **2015**, *102*, 18–29. [CrossRef]
60. Fercoq, A.; Lamouri, S.; Carbone, V. Lean/Green integration focused on waste reduction techniques. *J. Clean. Prod.* **2016**, *137*, 567–578. [CrossRef]
61. Rathi, R.; Kaswan, M.S.; Garza-Reyes, J.A.; Antony, J.; Cross, J. Green Lean Six Sigma for improving manufacturing sustainability: Framework development and validation. *J. Clean. Prod.* **2022**, *345*, 131130. [CrossRef]
62. Jum'a, L.; Zimon, D.; Ikram, M.; Madzik, P. Towards a sustainability paradigm; the nexus between lean green practices, sustainability-oriented innovation and Triple Bottom Line. *Int. J. Prod. Econ.* **2022**, *245*, 108393. [CrossRef]
63. Mishra, M.N. Identify critical success factors to implement integrated green and Lean Six Sigma. *Int. J. Lean Six Sigma* **2022**, *13*, 765–777. [CrossRef]
64. Kalemkerian, F.; Santos, J.; Tanco, M.; Garza-Reyes, J.A.; Viles, E. Analysing the alignment between the Green Lean and Circular strategies: Towards a Circular Lean approach. *J. Manuf. Technol. Manag.* **2022**, *33*, 1059–1079. [CrossRef]
65. Allouhi, A.; El Fouih, Y.; Kouskou, T.; Jamil, A.; Zeraouli, Y.; Mourad, Y. Energy consumption and efficiency in buildings: Current status and future trends. *J. Clean. Prod.* **2015**, *109*, 118–130. [CrossRef]
66. Gabbar, H.A. *Energy Conservation in Residential, Commercial, and Industrial Facilities*, 1st ed.; Wiley-IEEE Press: Piscataway, NJ, USA, 2018.

67. Harvey, L.D.D. Reducing energy use in the building sector: Measures, costs, and examples. *Energy Effic.* **2009**, *2*, 139–163. [CrossRef]
68. Chen, X.; Yang, H. Integrated energy performance optimization of a passively designed high-rise residential building in different climatic zones of China. *Appl. Energy* **2018**, *215*, 145–158. [CrossRef]
69. Zhang, Y.; He, C.-Q.; Tang, B.-J.; Wei, Y.-M. China's energy consumption in the building sector: A life cycle approach. *Energy Build.* **2015**, *94*, 240–251. [CrossRef]
70. Ascione, F.; Bianco, N.; Mauro, G.M.; Vanoli, G.P. A new comprehensive framework for the multi-objective optimization of building energy design: Harlequin. *Appl. Energy* **2019**, *241*, 331–361. [CrossRef]
71. Sorrell, S. Reducing energy demand: A review of issues, challenges and approaches. *Renew. Sustain. Energy Rev.* **2015**, *47*, 74–82. [CrossRef]
72. Yao, J. Energy optimization of building design for different housing units in apartment buildings. *Appl. Energy* **2012**, *94*, 330–337. [CrossRef]
73. Merkel, E.; McKenna, R.; Fehrenbach, D.; Fichtner, W. A model-based assessment of climate and energy targets for the German residential heat system. *J. Clean. Prod.* **2017**, *142*, 3151–3173. [CrossRef]
74. Benndorf, G.A.; Wystrcil, D.; Rehault, N. Energy performance optimization in buildings: A review on semantic interoperability, fault detection, and predictive control. *Appl. Phys. Rev.* **2018**, *5*, 041501. [CrossRef]
75. Harish, V.S.K.V.; Kumar, A. Reduced order modeling and parameter identification of a building energy system model through an optimization routine. *Appl. Energy* **2016**, *162*, 1010–1023. [CrossRef]
76. Xia, X.; Zhang, J. Mathematical description for the measurement and verification of energy efficiency improvement. *Appl. Energy* **2013**, *111*, 247–256. [CrossRef]
77. Dascalaki, E.; Droutsa, K.G.; Balaras, C.A.; Kontoyiannidis, S. Building typologies as a tool for assessing the energy performance of residential buildings—A case study for the Hellenic building stock. *Energy Build.* **2011**, *43*, 3400–3409. [CrossRef]
78. Dascalaki, E.; Kontoyiannidis, S.; Balaras, C.A.; Droutsa, K.G. Energy certification of Hellenic buildings: First findings. *Energy Build.* **2013**, *65*, 429–437. [CrossRef]
79. *Technical Chamber of Greece TEE KENAK 1.31.1.19*; Software Inspection & Certification Of Energy Buildings, Study of Energy Efficiency Boiler Inspection/Heating Installations and Facilities Air Conditioning-Manual. Technical Chamber of Greece: Athens, Greece, 2018.
80. *Technical Chamber of Greece TOTEE 20701-1/2017*; Detailed National Specifications of Parameters for Calculating the Energy Efficiency of Buildings and the Issue of the Energy Performance Certificate. Technical Chamber of Greece: Athens, Greece, 2017.
81. Fisher, R.A. *The Design of Experiments*; Macmillan: New York, NY, USA, 1971.
82. Fisher, R.A. *Statistical Methods, Experimental Design, and Scientific Inference*; Oxford University Press: Oxford, UK, 1990.
83. Tanco, M.; Viles, E.; Ilzarbe, L.; Alvarez, M.J. Implementation of Design of Experiments projects in industry. *Qual. Reliab. Eng. Int.* **2009**, *25*, 478–505. [CrossRef]
84. Ilzarbe, L.; Alvarez, M.J.; Viles, E.; Tanco, M. Practical applications of design of experiments in the field of engineering: A bibliographical review. *Qual. Reliab. Eng. Int.* **2008**, *24*, 417–428. [CrossRef]
85. Satterthwaite, F. Random balance experimentation. *Technometrics* **1959**, *1*, 111–137. [CrossRef]
86. Booth, K.H.V.; Cox, D.R. Some Systematic Supersaturated Designs. *Technometrics* **1962**, *4*, 489–495. [CrossRef]
87. Williams, K.R. Designed Experiments. *Rubber Age* **1968**, *100*, 67–71.
88. Lin, D.K.J. A New Class of Supersaturated Designs. *Technometrics* **1993**, *35*, 28–31. [CrossRef]
89. Allen, T.T.; Bernshteyn, M. Supersaturated Designs That Maximize the Probability of Identifying Active Factors. *Technometrics* **2003**, *45*, 90–97. [CrossRef]
90. Weese, M.L.; Smucker, B.J.; Edwards, D.J. Searching for powerful supersaturated designs. *J. Qual. Technol.* **2015**, *47*, 66–84. [CrossRef]
91. Rousali, C.; Besseris, G. Lean Screening for Greener Energy Consumption in Retrofitting a Residential Apartment Unit. *Appl. Sci.* **2022**, *12*, 6631. [CrossRef]
92. Zarkadas, S. Quality Optimization of Electromechanical Systems and Building Shell Insulation. Master's Thesis, Advanced and Industrial Manufacturing Systems, Kingston University, London, UK, 2022.
93. Plackett, R.L.; Burman, J.P. The design of optimum multifactorial experiments. *Biometrika* **1946**, *33*, 303–325. [CrossRef]
94. Draper, N.; Smith, H. *Applied Regression Analysis*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1981.
95. Schwarz, G.E. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
96. Mallows, C.L. Some comment on Cp. *Technometrics* **1973**, *8*, 661–675.
97. Hair, J.F.; Sarstedt, M.; Hopkins, L.; Kuppelwieser, V.G. Partial least squares structural equation modeling (PLS_SEM). *Eur. Bus. Rev.* **2014**, *26*, 106–121. [CrossRef]
98. Dijkstra, T.K.; Henseler, J. Consistent partial least squares path modeling. *MIS Quart.* **2015**, *39*, 297–316. [CrossRef]
99. Durbin, J.; Watson, G.S. Testing for serial correlation in least squares regression. *Biometrika* **1950**, *37*, 409–428. [PubMed]
100. R Core Team. *R: A Language and Environment for Statistical Computing, Version 4.2.2.*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: https://www.R-project.org/ (accessed on 31 October 2022).
101. Rousseuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Comp. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]

102. Spearman, C. The proof and measurement of association between two things. Am. J. *Psychol.* **1904**, *15*, 72–101.
103. Tibshirani, R.; Walther, G.; Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B* **2001**, *63*, 411–423. [CrossRef]
104. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [CrossRef]
105. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; The University of Illinois Press: Urbana, IL, USA, 1949.
106. Freedman, D.; Diaconis, P. On the histogram as a density estimator L_2 theory. *Prob. Theory Rel. Fields* **1981**, *57*, 453–476.