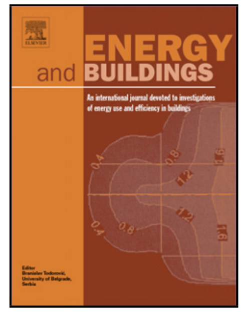

Title: The effect of spatial and temporal randomness of stochastically generated occupancy schedules on the energy performance of a multiresidential building

Author: Salvatore Carlucci Gabriele Lobaccaro Yong Li Elena Catto Lucchino Roberta Ramaci

PII: S0378-7788(16)30384-X
DOI: http://dx.doi.org/doi:10.1016/j.enbuild.2016.05.023
Reference: ENB 6658

To appear in: *ENB*

Received date: 11-12-2015
Revised date: 8-5-2016
Accepted date: 9-5-2016

The effect of spatial and temporal randomness of stochastically generated occupancy schedules on the energy performance of a multiresidential building

Salvatore Carlucci[1,*], Gabriele Lobaccaro[2], Yong Li[3], Elena Catto Lucchino[1], Roberta Ramaci[2]

[1] *NTNU Norwegian University of Science and Technology, Department of Civil and Transport Engineering, Trondheim, Norway*

[2] *NTNU Norwegian University of Science and Technology, Department of Architectural Design, History and Technology, Trondheim, Norway*

[3] *Shanghai Jiao Tong University, Institute of Refrigeration and Cryogenics, Shanghai, China*

[*] *Corresponding author. Tel.: +47 735 94634. E-mail address: salvatore.carlucci@ntnu.no (S. Carlucci).*

Highlights

- The study has its origins in the *Sustainable Energy in Cities* summer school held in Shanghai, China, in July 2015
- It statistically explores the effect of temporal and spatial randomness of stochastically generated occupancy schedules on a building's energy performance
- It adopts a scalarized single-objective optimization to minimize heating and cooling energy needs
- It presents a quality assurance procedure for numerical models of buildings that cannot be calibrated using measured data
- Modeling of high-performance buildings requires a (spatially) detailed and (timely) precise description of occupancy and occupant-dependent input variables

Table of contents

Keywords

Mathematical optimization; multiresidential building; occupancy models; quality assurance; spatial randomness; temporal randomness, Shanghai.

Abbreviations

ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers
BAS	Building Automation System
BMS	Building Management Systems
CO_2	Carbon dioxide
$C_V(RMSE)$	Coefficient of variation of the Root Mean Square Error
df	Degree of freedom
EBC	Energy in Buildings and Communities
EIA	Energy International Administration
HSCW	Hot Summer and Cold Winter
HVAC	Heating Ventilation and Air Conditioning
IEA	International Energy Agency
IWEC	International Weather for Energy Calculations
MBE	Mean Bias Error
Md	Median
Mtce	Million tons of coal equivalent
NTNU	Norwegian University of Science and Technology
PSO	Particle Swarm Optimization
PSOIW	Particle Swarm Optimization with Inertia Weight
SEniC	Sustainable Energy in Cities
SHC	Solar Heating and Cooling
SJTU	Shanghai Jiao Tong University

Abstract

Building performance simulation is frequently used to support building design, renovation, and operation. However, modelers are traditionally concerned with accurately describing technical input data, and have only limited interest in investigating the influence of occupant behavior on buildings' energy performance.

To fill this gap, this article examines the effects of stochastically generated occupancy schedules on the energy performance of a multiresidential high-rise building located in Shanghai, China. The building's energy performance is analyzed under two design proposals: a law-compliant proposal developed by the designers, and a second proposal conceived through an automatized optimization process. A statistical analysis quantifies the energy implications of adopting different degrees of randomness when creating occupancy and occupancy-dependent schedules.

Simulation outcomes show that temporal and spatial randomness of occupancy and occupancy-dependent schedules have a statistically significant influence on the building's energy performance, with an estimated uncertainty of up to 10%. At least in Shanghai, occupant behavior affects cooling more than heating, and its influence on the energy performance is stronger in high-performance buildings than in poorly insulated ones.

Finally, accurate modeling of high-performance buildings would require a detailed and precise description of occupancy and occupant-dependent input variables even if this increases the modeling effort and costs.

1 Introduction

This article investigates the influence of occupant behavior on the energy performance of a multiresidential high-rise building located in Shanghai, China, and has its origins in the *Sustainable Energy in Cities* (SEniC) summer school organized by the Norwegian University of Science and Technology (NTNU) in collaboration with the Shanghai Jiao Tog University (SJTU) and held in Shanghai in July 2015.

The object of this study was chosen because, over recent decades, there has been a growing interest in reducing the environmental impact of the building sector, which is believed to be responsible for more than two-thirds of the world's primary energy usage and more than one-third of the world's greenhouse gas emissions [1].

Focusing on China, the International Energy Agency (IEA) and the World Bank state that China is currently the biggest greenhouse gas emitter in the world [2-4]. Furthermore, the US Energy Information Administration (EIA) claims that the Chinese building sector was responsible for up to 18% of the overall Chinese greenhouse gas emissions in 2009 [5]. Regarding energy demand, the EIA estimated that China is the largest energy user worldwide, with a rate of 18% in 2010 [6], and that one-fifth of China's total primary energy is attributable to the building sector. The IEA has also pointed out that China's residential and commercial energy usage were ranked respectively first and third among those of all the world's countries [2]. Additionally, it shall be recalled that, after the *Reform and Opening-up* policy launched in 1978, China entered into a frenetic period of rapid urbanization [7, 8], with a level of urbanization that rose from 19.4% in 1980 to 53.7% in 2013 [9] and is expected to achieve saturation by 2030 [10]. This aspect is closely linked with the phenomenon of the rapid growth of the Chinese population, which increases the demand for residential buildings.

China's census indicates that energy usage has increased by more than 3.5 times from 1990 to 2013 (i.e., from 987 million tons of coal equivalent, Mtce to 3750 Mtce), while the amount of carbon dioxide (CO_2) emissions increased almost four times in the same period (from 2269 Mtce to 8106 Mtce) [11]. Berardi [5], elaborating data from IEA [12] and World Bank [4], pointed out that, in China, the energy requirement in 2050 will be 15 times higher than the level in 1970. In this scenario, some tremendous challenges related to environmental pressures, including energy usage and CO_2 emissions, have become more and more urgent in China [13, 14].

Furthermore, as urbanization continues to increase rapidly, much still needs to be done to achieve energy security and environmental sustainability and to sensitize users' awareness of energy utilization. A study conducted by Murata, Hailin and Weisheng [15] showed that, in China, at least 21% of residential energy (up to 50% in some regions) may be saved by using household appliances with a higher efficiency. However, since the building standards' requirements are getting stricter, the relative impact of occupants on a building's energy usage is going to increase, and "better models of occupation presence and interaction are necessary" [16]. Indeed, the presence of people in a building affects its thermal and energy performance not only through the production of sensible and latent heat, but also, and to a large extent, through their activities and interaction with the building's systems, devices, and appliances. However, improper use of electric devices and appliances is a key factor that makes occupancy one of the weakest points of the energy balances of a building [17]. Wu, Zhu and Zhou [18] compared electricity consumption, measured in 1999, among 410 apartments in Beijing and provided evidence that variations in household electricity use is mainly due to occupant behavior. In another study carried out in Beijing in 2006, Li, Jiang and Wei [19] monitored the summer use of air-conditioning in 25 identical apartments in a low-rise building. Outcomes showed relevant discrepancies in energy usage among these apartments, with a maximum value of 991 kWh per year and a minimum value of 170 kWh per year due to different occupant behaviors. Jian, Li, Wei, Zhang and Bai [20] found a significant impact of occupant behavior on the whole electricity use in 44 individual apartments in Beijing. Regarding heating, Guo, Yan, Peng, Cui, Zhou and Hu [21], during the winter of 2013, monitored energy use for heating, indoor temperatures, and CO_2 emissions in 48 dwellings located along the Yangtze River (China) and belonging to the so-called hot-summer-and-cold-winter (HSCW) climate zone. Unlike Northern China, this region is not provided with district heating, and heating needs are managed at the building or even individual apartment level. Consequently, the operation time of each heating device varies significantly among different units and families' patterns of use. Measurements showed that the heating consumption was quite low, due to the fact that heating devices were used just for 30% of the entire heating season. Furthermore, indoor air temperatures, which reached on average 16 °C, were largely below any thermal comfort zone. In relation to all of the above-mentioned matters, one of the research questions investigated in this work is to estimate to what extent occupant behavior influences the energy performance of a multiresidential building while ensuring comfortable conditions.

It can be inferred from the aforementioned examples that, to fully characterize the performance of a building, several occupancy patterns seem necessary, shifting building modeling from a deterministic to a stochastic approach. Unfortunately, modeling occupant behavior is complex. Six typologies of models are available to describe occupant behavior[1]: psychological models, average value models, deterministic models, probabilistic models, agent-based models, and action-based models [22]. Focusing only on probabilistic models, different techniques can be used for developing these models, such as logistic regressions, state-transition analyses using

[1] In general, these typologies of models allow a description of the occupant's status and action or reaction in response to external or internal stimuli in order to adapt ambient environmental conditions that can affect the energy performance of a building. However, in this article, only occupancy and occupancy-related energy usages are addressed.

Markov chains, Monte Carlo modeling, and artificial neural networks. Several reviews are available in the scientific literature that discuss the strengths and weaknesses of these models and techniques, for example Reff. [23, 24]. Therefore, another research question consists of evaluating to what extent the (temporal and spatial) implementation of a probabilistic occupancy model in a simulation tool influences the energy performance of a multiresidential building.

Moreover, social, cultural, and economic factors provide a further significant contribution to defining occupants' attitudes towards energy usage in buildings. However, no occupancy model specifically built for Chinese society that was suitable to be used in the summer school was identified in the literature; hence, an occupancy probabilistic model developed for Japanese society [25] was implemented after a qualitative check was made regarding the reliability of its extension to the presented case study by the Chinese students and professors participating in the summer school.

In summary, this work aims at providing further insights into the influence of occupant behavior on energy uses for heating and cooling, electric lighting, and appliances by modeling stochastic schedules for occupancy and occupant-dependent input data in both a current law-compliant and an optimized design proposal for a high-rise residential building in Shanghai. However, in this article, occupant behavior is used exclusively to describe occupancy and occupant-dependent energy uses for electric lighting and appliances. It does not refer to the effect of those actions taken by occupants to manipulate the built environment to create more comfortable and pleasant indoor conditions, such as the operation of windows, solar shading devices, and thermostats, etc.

2 Methodology

During the SEniC summer school, three main learning outcomes were pursued to provide students with information and knowledge in the following areas: (i) numerical modeling and dynamic energy simulation of buildings, (ii) mathematical optimization techniques that are useful for supporting the design of high-performance buildings, and (iii) statistical analyses that are useful for interpreting the sets of data populated with the simulation outcomes due to the implementation of several stochastically generated occupancy schedules. During the first week of the summer school, 15 students were split into two groups, which worked in parallel creating (i) the numerical model of the Base case, according to the law-compliant proposal developed by the designers and (ii) several stochastically generated occupancy schedules. Starting from the same blueprints, two subgroups, autonomously and in isolation, created two numerical models of the Base case, which were eventually compared and refined against input errors. During the second week, most of the students were involved in the simulation of the Base case using the stochastically generated occupancy schedules. Another smaller group set up and carried out the mathematical optimization of the Base case, obtaining the so-called Optimized case. All the students then simulated the Optimized case with the same stochastically generated occupancy schedules previously integrated into the Base case. Finally, they collaborated to summarize the work activities in a report and a poster. The work done during the summer school was the basis for the present article, which has recently been refined by the authors.

2.1 Description of the case study

The object on which students worked during the summer school was the Zhoukanghang (周康航) resettlement residential project located in Shanghai's southeastern area. This resettlement residential project acts on a neighborhood scale and represents a typical Chinese housing development project constructed after the introduction of the *Reform and Opening-up* policy [7, 8]. In this section, the main features of the local climate are briefly introduced and a description of the study area is presented.

2.1.1 Climate of the Shanghai region

China is characterized by large regional differences in climate. According to the Chinese *Code for Thermal Design of Civil Buildings* (GB50176-93) [26], China is divided into five climate zones: a severe-cold region, a cold region, a hot-summer-and-cold-winter region, a hot-summer-and-warm-winter region, and a mild region [27]. The city of Shanghai (latitude 31.20° N, longitude 121.50° E) is classified as being in the hot-summer-and-cold-winter (HSCW) region and is characterized by being in a humid, subtropical climate zone with hot, moist, and rainy summers and overcast, cold winters. It is indicated with the code Cfa in the climate classification of Köppen and Geiger [28]. The annual weather data are presented in Figure 1. They were elaborated from the .epw weather file made available by the US Department of Energy [29] and represent the monthly mean daily values of the principal meteorological quantities, that is, an average for each of the 24 hours of the day is calculated and represented per month.

Regarding the air temperature, the winter outdoor temperature can be below 0 °C and can reach −5 °C in December, while in summer, it usually reaches above 30 °C with peaks of over 35 °C in July and August. In relation to the annual solar radiation, the highest contribution of the direct component is registered in winter, thanks to a lower level of cloudiness[2]. However, due to the combined contribution of direct and diffuse radiation, in summer the global radiation reaches the highest values. Finally, the level of relative humidity can achieve values that are higher than 85% during summertime.

[2] Air pollution seems, hence, not to be considered by the weather file.

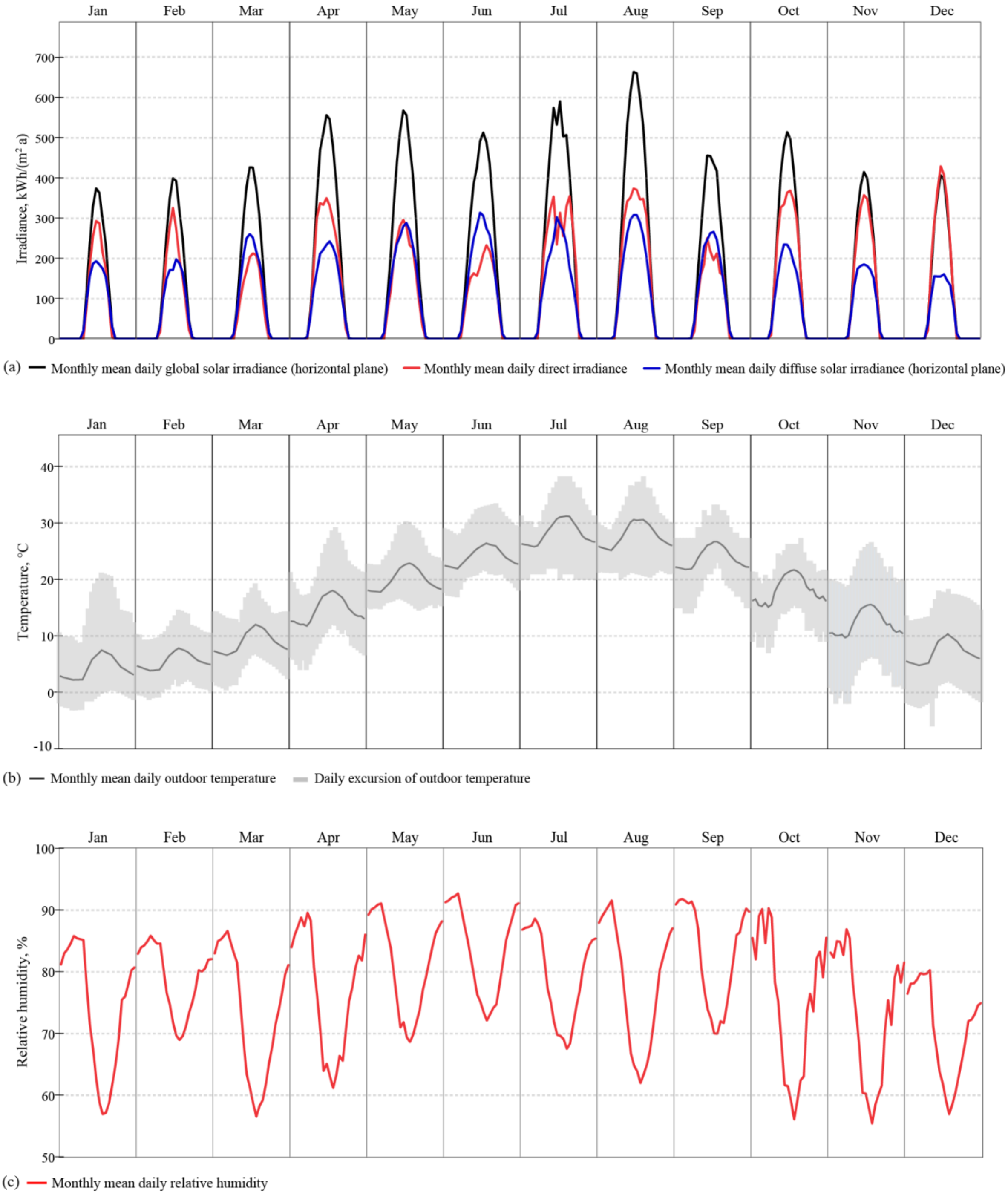

Figure 1: Trends of (a) global solar radiation on a horizontal plane, and direct and diffuse solar radiation, (b) outdoor temperature and its daily excursion and (c) relative humidity in Shanghai.

2.1.2 The study area

The Shanghai urban region covered an area of about 176 km^2 in 1984, while, during the period of rapid urbanization, it grew until it covered 412 km^2 in 1996 and 886 km^2 in 2008 [30]. The chosen case study is included in one of the new urban densification interventions: the Zhoukanghang (周康航) resettlement residential project. It is located in the Pudong District, in the southeastern area of Shanghai, and is intended for the relocation of people formerly living downtown. The project started in 2013 and is still under construction. It is composed of a cluster of nine high-rise multiresidential building blocks and a community center. The entire intervention covers an area of 2.45 hectares and has a total net floor area of 59 672 m^2, which allows for the possibility of hosting 588 households. Each block has 18 conditioned storeys plus a free-running basement. Each storey includes six apartments: four designed for three-person families and two for couples. The layout of the apartments is symmetrical and is represented in Figure 2.

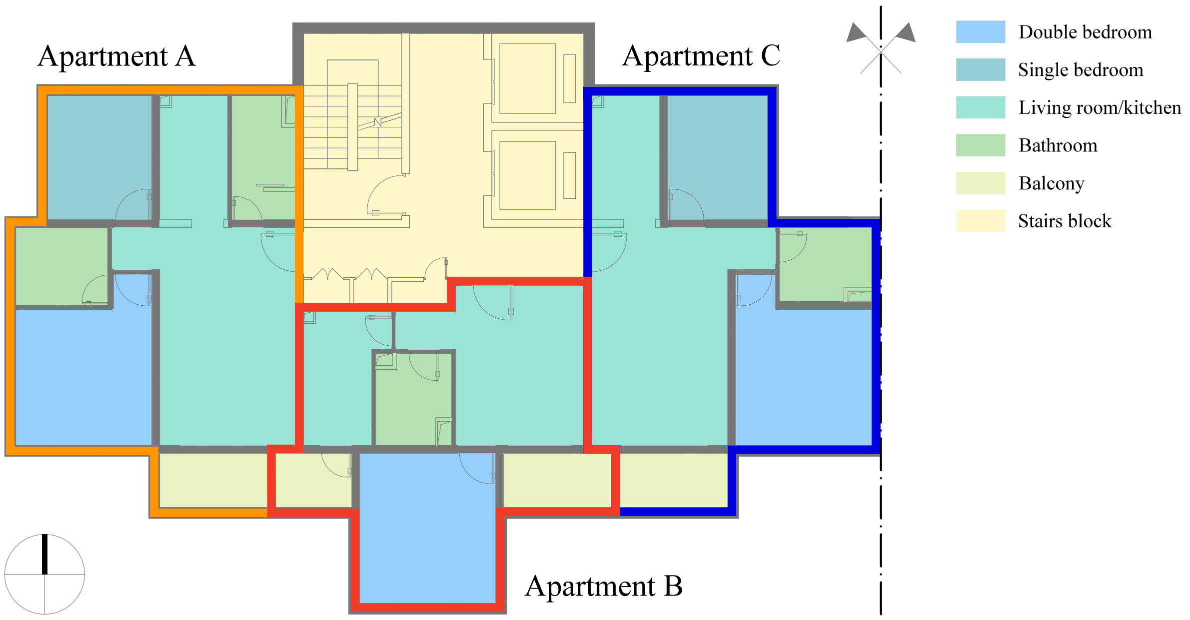

Figure 2: Layout of the typical floor unit of the Zhoukanghang (周康航) resettlement residential project.

The apartments A, B, and C have a total net floor area of about 66, 50, and 60 m^2 respectively. According to the current design proposal, the external walls, the roof, and the floor are made of concrete and have a thin insulating layer. The windows are made of a clear, air-filled double glazing unit mounted into an aluminum frame with a thermal break. Furthermore, no external solar shading devices or overhangs are foreseen to control the incoming solar radiation. The characteristics of the building envelope components according to the current design proposal are summarized in Table 1.

Table 1: Characteristics of the building components in the Base case.

Building component	**Steady-state transmittance, U, W/(m^{2}K)**	**Periodic transmittance, Y_{IO}, W/(m^{2}K)**
External walls	1.46	0.23
Flat roof	0.71	0.17
Floor against basement	1.80	0.03
Basement walls	0.48	0.08
Internal partitions between rooms	2.63	2.06
Internal partitions between apartments	1.40	0.34
Internal floor between apartments	1.43	0.32
Glazing unit	2.80	Not applicable
Frame with thermal break	3.45	Not applicable

2.2 Description of the numerical model

In this section, a description of the process followed by the students to create the numerical model of the building in the Base case is presented, as well as the necessary information to comprehensively understand the feature of the model and all the assumptions made.

2.2.1 Modeling simplifications and thermal zoning

Building performance simulation (BPS) is a computer-based process for assessing through a numerical model some aspects of a building performance based on fundamental physical principles and engineering models. As with all other models used in science, its purpose is to represent as closely as possible a given object or real phenomenon, but it always remains distinct from the object or phenomenon itself. Therefore, the purpose of BPS is to create an abstracted building representation that reproduces a selected behavior with a controlled deviation. Moreover, for large buildings, a trade-off has to be found between available computational capacity, the running period of each individual simulation, and the number of the exported outputs.

For this purpose, only one high-rise multiresidential building block in the Zhoukanghang (周康航) resettlement residential project was modeled in DesignBuilder. The building block was chosen to be representative of the average energy performance of the nine building blocks, taking into account the mutual shadows projected by the designed blocks and the surrounding buildings (Figure 3, left). Given the symmetry of the original project (Figure 2), only half of the building block was modeled (Figure 3, right), and the partition normally facing the other half of the building was assumed to be adiabatic.

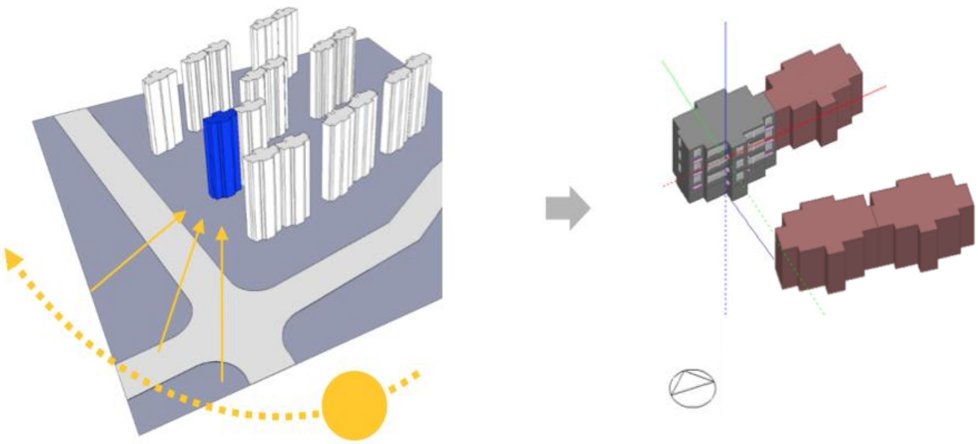

Figure 3: Graphical representation of the Zhoukanghang (周康航) resettlement residential project and of the simplified geometry of the numerical model.

This choice halved the number of apartments without affecting significantly the simulation outcome since energy benchmarks are typically normalized by the conditioned net floor area.

Since each building block has 18 storeys, the total number of apartments to model would have been 54, and, zoning for each room, the total number of thermal zone would have been 325; these are numbers outside the simulation capabilities of most BPS software. Therefore, the model was reduced in size in order to take into account only three storeys and the free-running basement for a total of nine apartments and 50 thermal zones.

Basement aside, the three storeys are: (i) the top floor, mostly affected by solar radiation incident on the flat roof, (ii) an intermediate storey representing all the storeys from the second to the seventeenth floor, and (iii) the bottom storey, which is in thermal contact with the free-running basement. Obviously, this simplification affects the energy balance of the whole building block since the top and the bottom storeys have a more significant impact on the overall energy bill. Looking next at the whole building, the reduced-size model is characterized by a lower thermal inertia and a higher shape ratio (S/V). Furthermore, the lower height of the reduced-size building model implies a lower over-shading effect from surrounding buildings, which were therefore moved closer to the analyzed building block than set out in the original project (Figure 3).

Since one of the purposes of this article is to assess the effect of the stochastically generated occupancy schedules on the energy performance of the building, every room was modeled as an individual thermal zone (Figure 2). In addition to the thermal zones of the apartments, that is, bedrooms, living rooms, kitchens, and bathrooms, some buffer zones were also modeled to better account for the heat exchanges between the indoor and outdoor environments. Hence, the basement, entrance, stair block, and sheltered balconies were modeled as free-running zones. The surrounding building blocks were modeled in all simulations without any thermal zones, but with surfaces that can shade or reflect direct solar radiation.

The values for maximum occupancy and the installed power for electric lighting and appliance set for each room of the model are reported in

Table 2.

Table 2: Definition of internal gains for each type of room of the building model.

	Maximum number of people per room	**Installed power for lighting (W/m^2)**	**Installed power for appliances (W/m^2)**
Double bedroom	2	5	10
Single bedroom	1	10	10
Living room of Apartments A and C	3	12	10
Living room of Apartment B	2	12	10
Bathrooms	-	-	-

2.2.2 Setup of the numerical models

The weather file used in all simulations was the *International Weather for Energy Calculations* (IWEC) file of Shanghai made available by the US Department of Energy [29].

The case study was modeled in DesignBuilder, version 4.2.0.054, which was used as the interface for the whole-building dynamic simulation engine EnergyPlus [31], version 8.1.0.009. Each released version of EnergyPlus undergoes two major types of validation tests [32]: analytical tests, according to ASHRAE Research Projects 865 and 1052, and comparative tests, according to the ANSI/ASHRAE 140 [33] and IEA SHC Task 34/Annex 43 BESTest method. Within the capability of EnergyPlus, the building models were set up with a priority of reproducing in significant detail the geometrical features of the building and the physical phenomena that determine the thermal behavior of the building, although this was at the expense of a rather large computational time. The update frequency for calculating sun paths was set to 7 days, rather than the default 20 days, to better estimate solar gains entering the model. The heat conduction through the opaque envelope was calculated via the finite difference method using a 3-minute time-step, rather than the default transfer function method that has a

15-minute time-step, in order to improve calculation accuracy in the presence of components with high thermal inertia, such as the thick concrete layers [34]. The natural convection heat exchange near external and internal surfaces was calculated via the adaptive convection algorithm [35] to better meet the local conditions of each surface of the model. The initialization period of simulation was set at 25 days to reduce the uncertainties connected to the thermal initialization of the numerical model. The voluntary ventilation and involuntary air infiltration were calculated using the *AirflowNetwork* module to better calculate the contribution of natural ventilation and infiltration.

2.2.3 Additional assumptions

Besides the selection of physical models and numerical schemes, it is necessary to make a number of assumptions and choices to comprehensively describe the numerical model used for the simulations. Ideal systems were integrated into the model to provide heating and cooling. Therefore, hereafter we will refer only to the *energy need for heating and cooling*[3] without assessing the efficiencies of any building systems. Furthermore, since the building is considered conditioned throughout the year, the Fanger thermal comfort model [37] is adopted to set suitable set-point operative temperatures for the heating and cooling periods [38, 39]. For both periods, it is assumed that (i) the indoor operative temperature is calculated as the mean of the air and the mean radiant temperature, (ii) the external work of the occupants is zero, (iii) the occupants are in a sedentary activity[4], so their metabolic rate is 1.2 met, (iv) the relative humidity is fixed at 50%, and (v) air speed is fixed at 0.1 m/s. Specifically, the occupants are supposed to wear typical summer clothing with a clothing resistance, I_{clo}, of 0.5 clo during the cooling period, while the clothing resistance is set at 1.0 clo during the heating period. Accordingly, the cooling set-point operative temperature is 24.7 ºC and the heating set-point operative temperature is 21.5 ºC. Such assumptions are implemented in all the models simulated in this work: the Base case, the Optimized case, and in all building variants simulated in the optimization run. Furthermore, operable solar shading devices and a mechanical ventilation system equipped with a high efficient heat recovery unit (Eff = 80%) have been introduced in the reference building model used in the optimization run and, hence, in the Optimized case, which is the optimal building variant identified by the optimization run (Section 3.2).

2.3 Quality assurance of a numerical model

Since every model is a simplified representation of a real-world problem, it is necessary to be confident that a building models provides an accurate representation of how a building and its systems would behave in reality. Quality assurance is a process that aims to develop confidence in the predictions of a simulation tool [40]. This is of fundamental importance because designers base design decisions on the results of simulations. The main strategies used to enhance the quality of a BPS are undertaking rigorous *validation* and *calibration* of a building model.

[3] The *energy need for heating or cooling* is defined as “heat to be delivered to, or extracted from, a conditioned space to maintain the intended temperature conditions during a given period of time” [36] CEN, Energy performance of buildings - Overall energy-use and definition of energy ratings, in, European Committee for Standardization, Brussels, Belgium, 2008.
[4] The metabolic rate was set according to table A.3 of the EN 15251 for “Residential buildings, living spaces (bed room’s living rooms etc.) Sedentary activity ~ 1,2 met”.

In general, *validation* is "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model" [41]. Its purpose is therefore to assess the *physical fidelity* of a model for a specific predictive application [42]. Operationally, it is a procedure that uses statistical metrics to evaluate the deviation in the prediction of a model built on a data sample commonly called a *training set*, with respect to the actual data of the sample used to carry out validation, commonly called a *test set*. If a training set and a test set belong to the same population of data, this assessment process is called *internal validation*; if they belong to a different population of data, it is called *external validation*. *Internal validation* results in an evaluation of the reproducibility of a model on a different data set belonging to the same population of data, whereas e*xternal validation* evaluates the generalizability, or transportability, of a model to a related, but different, population from that used for developing the model itself. In the specific case of BPS, the model is not built on data collected from the field using, for example, regression techniques, but is a mathematical system of partial differential equations that represent physical phenomena and is solved by approximation using adequate numerical methods. In order to assess the accuracy of a model in representing the behavior of an actual building, *aleatory uncertainties* (i.e., "the inherent variation associated with the physical system or the environment under consideration" [43]) have to be minimized, and, hence, a building model should be simulated using the most accurate boundary conditions, for example, weather conditions and occupancy profiles, etc., which represent as much as possible the real conditions to which the actual building is exposed. Therefore, only external validation can be employed to evaluate the quality of a BPS model.

Calibration "is the process of improving the agreement of a code calculation or set of code calculations with respect to a chosen set of benchmarks through the adjustment of parameters implemented in the code" [42]. Its purpose is therefore to help the analyst to choose those values of the design variables that improve the agreement of a simulation model with a defined set of physical benchmarks, increasing the *credibility* of the model. Operationally, it is a process that starts with the choosing of a physical benchmark (e.g., delivered energy, indoor air temperature, etc.) with which to calibrate a model. At the same time, the *epistemic uncertainty* (i.e., "a *potential* inaccuracy in any phase or activity of the modeling process that is due to *lack of knowledge*" [43]) of a set of design variables (also called independent variables or input variables) has to be quantified. Several versions of the model are then generated, setting different values for each design variable, which are compatible with the already identified epistemic uncertainties. Finally, all models are simulated, and the individual simulation outcomes are collected and compared with the measured values of the same benchmark. The agreement between simulation outcomes and measurements is assessed via statistical metrics. ASHRAE Guideline 14 [44] suggests the use of the *mean bias error*, *MBE*, and the *coefficient of variation of the root mean square error*, $C_V(RMSE)$. *MBE* is a non-dimensional measure of the overall bias error between the measurements and the simulation outcomes in a known time resolution, and it is usually expressed as a percentage:

$$MBE = \frac{\sum_{i=1}^{N_p}(m_i - s_i)}{\sum_{i=1}^{N_p} m_i} \quad [\%] \tag{1}$$

where m_i ($i = 1, 2, \ldots, N_p$) are the measured data, s_i ($i = 1, 2, \ldots, N_p$) are the simulated data at time interval i, and N_p is the entire number of data values. Positive values indicate that regression underpredicts values; negative values indicate that the model predicts values for the benchmark that are higher than the actual values, whereas $C_V(RMSE)$ indicates overall uncertainty in a model. The lower $C_V(RMSE)$ is, the smaller the residuals between the measurements and the simulation outcomes are. This is defined as:

$$C_V\left(RMSE\right)=\frac{1}{\overline{m}}\sqrt{\frac{\sum_{i=1}^{N_p}\left(m_i-s_i\right)^2}{N_p}} \quad [\%] \tag{2}$$

where, besides the quantities already introduced in Eq.(1), $\overline{m}$ is the average of measured data values. ASHRAE Guideline 14 [44] also provides useful criteria that can be used to declare a model calibrated (Table 3).

Table 3: Acceptable calibration tolerances according to ASHRAE Guideline 14.

Calibration type	**Acceptable value of *MBE*** *	**Acceptable value of C_V(*RMSE*)** *
Monthly	±5%	15%
Hourly	±10%	30%

* Lower values indicate better calibration.

The building that was studied in this research is still under design, and therefore its numerical model cannot be validated against monitored data, nor can it be calibrated to identify those values of the design variables that increase model credibility. Therefore, a dedicated quality assurance procedure was implemented in order to minimize the following:

- *specification uncertainty* due to interpretation of the design specification
- *modeling uncertainty* due to the construction of the geometrical model and the selection of the physical models.

However, any quality assurance procedure cannot prevent:

- *numerical uncertainty* that might be due to the settings of the solver and that therefore mostly depends on the modeler's knowledge and expertise
- *scenario uncertainty*, which refers to uncertainty regarding all types of external factors to which the building and its technical systems are exposed, such as weather conditions and occupant behavior, and which is inherently present in any virtual experiments.

In addition, a seven-step procedure was designed and adopted:

1. All of the following modeling specifications were assembled in a design package: blueprints; all technical reports about the project developed by the designers; the typical weather file of the site to be used in all simulations; standard profiles for occupancy, as recommended by the UK National Calculation Methodology, to be used in the optimization; and electric lighting and appliances usages to be used in optimization.
2. The design package was delivered to two independent groups of students, who developed two building models in absolute autonomy and avoided reciprocal interference.
3. An indicator suitable for characterizing the thermal behavior of the building zones was identified. The purpose of the research is to assess the influence of different occupancy profiles in the several zones of a complex building model; hence, the criteria used to guide the selection of an indicator were: (i) that it should be comprehensive and capable of describing the entire energy performance of a thermal zone, (ii) that it should be simple to extract from the model and easy to statistically elaborate it, (iii) that a time serial metric rather than a cumulative metric should be used so that the dynamic behavior of the

building could be represented. The hourly operative temperature reached in the selected thermal zones of the models (living room plus kitchen) was finally chosen.

4. A comparison of the quality benchmarks of the two models was carried out graphically using a scatterplot.
5. The bias between the two models was quantified using *MBE*, and the overall uncertainty was evaluated using $C_V(RMSE)$.
6. An independent comparison of the input parameters of the two numerical models was run to find the source of discrepancy that caused the deviation in the quality benchmark.
7. Finally, the errors were fixed and the reference model was finalized for the subsequent analyses.

The above procedure cannot guarantee that the simulation outcome reflects the actual performance of the building, but it can help to minimize both specification and modeling uncertainties.

2.4 Building energy optimization

The energy design of a building is a multivariable problem that, leading to a large number of alternative solutions that cannot all be simulated in a time span compatible with the design phase of a building [45, 46]. In order to explore a very large number of building variants in a relatively short time, mathematical optimization is used to reduce the energy need for space conditioning of the building while guaranteeing high indoor thermal comfort conditions [47-50].

From among several options, scalarized optimization was considered a suitable trade-off between serving a pedagogical purpose and searching for a solution to the given design problem. It is a process whereby an optimization algorithm assesses an *a priori* combination of two or more *objective functions* that is called a *utility function*. Based on the values of the utility function, the optimization algorithm guides a simulation engine to simulate several building variants of the entire *design space* that makes up the optimization problem.

In our optimization problem, the two objective functions are the energy need for heating and cooling. They are combined linearly, through a simple summation, into a new variable that represents the energy need for space conditioning. The target of the optimization process is to find the building variant that minimizes the energy need for space conditioning of the entire building model by varying the values of six design variables: the constructions of exterior walls, the roof, the floor, the glazing units of windows, the control strategy of solar shading devices, and the set-point value of the control strategy of the solar shading devices.

First of all, the students built the numerical model of the reference building in DesignBuilder and exported it in the .idf format, which is the source file of the simulation engine EnergyPlus. Next, an automated iterative simulation process was built by coupling an optimization engine, GenOpt version 3.1.0 [51], with EnergyPlus. Finally, GenOpt was configured by customizing four files: the *Model template file*, the *Command file*, the *Initialization file*, and the *Configuration file*. The *Model template file* is a copy of the *Building model file* in .idf, which incorporates special codes substituting the names of the design variables. In the *Command file*, these special codes indicating the design variables are defined using all the values that can be taken by each design variable, and the optimization algorithm and the stopping criterion are tuned. In the *Initialization file*, the objective functions are recalled from EnergyPlus's .eso file, and the utility function is defined as an algebraic equation of the objective functions. Finally, the *Configuration file* manages the optimization process and is available for several operating systems; in general, it is not necessary to modify it. A flowchart of the entire optimization process is depicted in Figure 4.

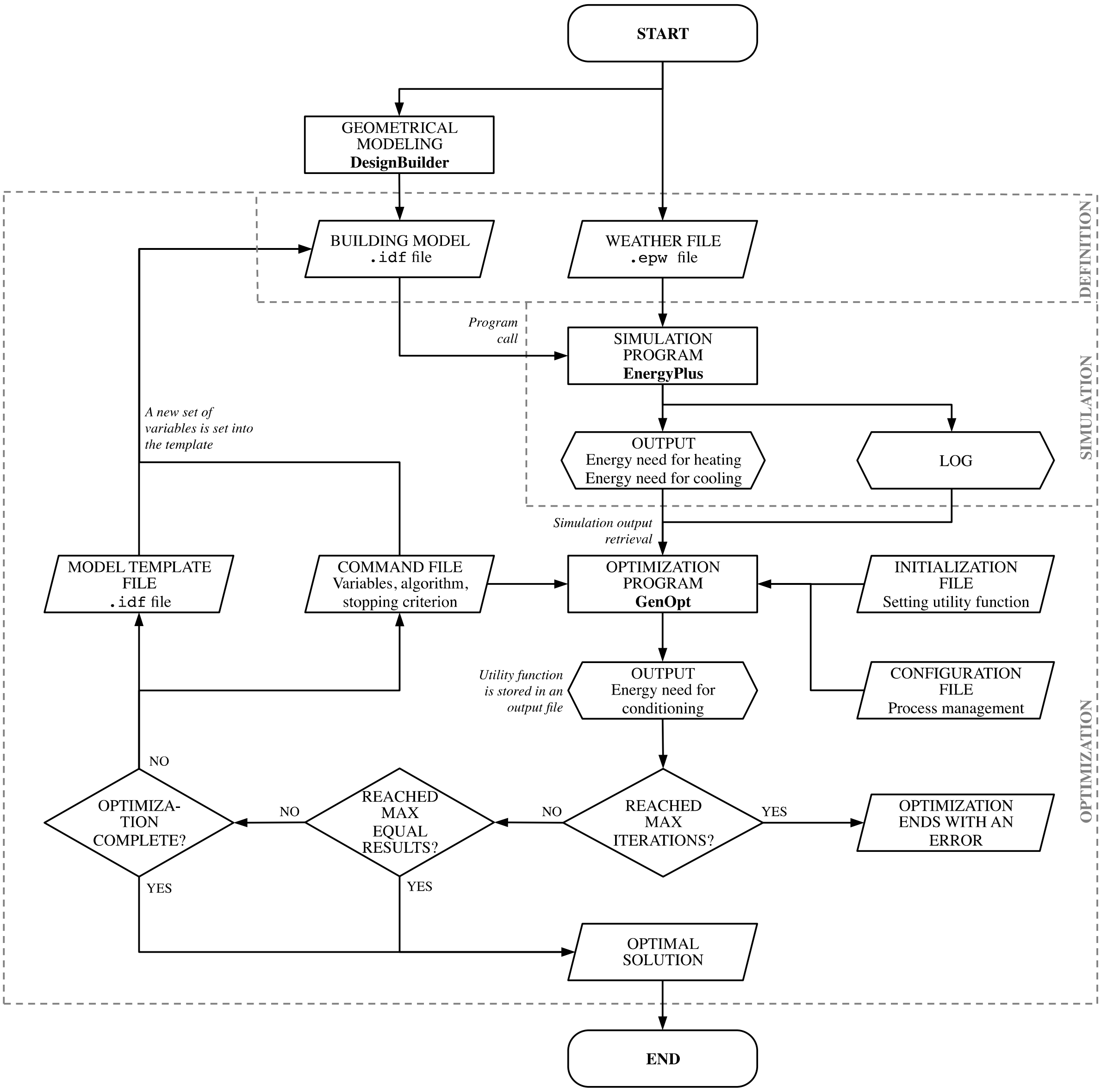

Figure 4: Flowchart of the optimization process.

2.4.1 Scalarized optimization

In mathematics, optimization is a process that aims to select the best element or set of elements, according to given criteria, that minimizes, or maximizes, one or more objective functions. More generally, optimization means finding the *best available* values of some objective functions given a defined domain, called the problem space, and also subject to a given set of equality or inequality constraints. A general optimization problem can be formulated as [52, 53]

$$\min_{z \in Z} F(\mathbf{x}) \quad \text{subject to} \begin{cases} G_i(\mathbf{x}) = 0 \\ H_j(\mathbf{x}) \geq C \\ \{i, j\} \in \frown \end{cases} \tag{3}$$

where $F(\mathbf{x})$ is the set of objective functions, $\mathbf{x}$ is the vector of design variables, z are the simulations analyzed, Z is the problem space, i equality and j inequality constraints are given for a general optimization problem in terms of the functions $G(\mathbf{x})$ and $H(\mathbf{x})$, which are dependent on the $\mathbf{x}$ design variables, and C is a constant.

On the basis of the number of the objective functions, an optimization is called a single- or a multi-objective optimization. In the case of single-objective optimization, a single optimum solution of the optimization problem exists and it is either its global maximum or minimum, depending on the purpose, whereas for a multi-objective optimization, a set of non-dominated variants belonging to the so-called Pareto front are the solution to the optimization problem. Finally, two or more objective functions, $F(\mathbf{x})$, can also be analytically combined into a utility function $U(\mathbf{x})$. This is called scalarized optimization and *a priori* condenses a multi-objective optimization problem into a single-objective one.

To solve the optimization problem related to this case study, a scalarized optimization combining both the energy need for heating and the energy need for cooling into a utility function built using the weighted sum method [54] and with unitary weighting factors was set as follows:

$$\begin{cases} U(\mathbf{x}) = \sum_{i=1}^{k} w_i F_i(\mathbf{x}) \\ \text{subject to: } w_i = 1 \;\; \forall \;\; i \in \mathbb{N} \end{cases} \rightarrow E_{\text{cond}}(\mathbf{x}) = E_{\text{heat}}(\mathbf{x}) + E_{\text{cool}}(\mathbf{x}) \tag{4}$$

where $E_{\text{cond}}(\mathbf{x})$ is the total energy need for space conditioning, $E_{\text{heat}}(\mathbf{x})$ is the energy need for heating, $E_{\text{cool}}(\mathbf{x})$ is the energy need for cooling (sensible plus latent), i is the counter of the objective functions, and k is the total number of the objective functions of the optimization problem.

2.4.2 Design variables and available design options

The building is assumed to be fully conditioned by an ideal system. Therefore, the design variables were selected from among those that only influence the building envelope and the solar shading control strategies, such as external-wall construction, roof construction, floor construction, glazing system, control strategy of solar shading devices, and set-point value of the solar shading control strategy. For the constructions, nine options were created in order to modulate the steady-state thermal transmittance, U, and the periodic transmittance, Y_{IO}, over three values labeled with '-' for a low performance, 'o' for a medium performance, and '+' for a high performance (Figure 5).

Six options, however, were available for the glazing unit and were created by modulating the steady-state thermal transmittance, U, and the solar factor, g, over the same aforementioned three values (Figure 5).

Solar shading devices and natural ventilation strategies were introduced in the numerical model, and the control strategies of solar shading devices can take three options in the optimization run:

- The solar shading devices installed on the sun-facing windows of the entire building close if the outdoor air temperature exceeds a set-point value in a given time-step (*OutTemp*).

- The solar shading devices installed on the sun-facing windows of a given thermal zone close if the indoor air temperature of the zone exceeds a set-point value in a given time-step (*InTemp*).
- The solar shading devices installed on the sun-facing window close if the irradiance incident on the window exceeds a set-point value in a given time-step (*WinTemp*).

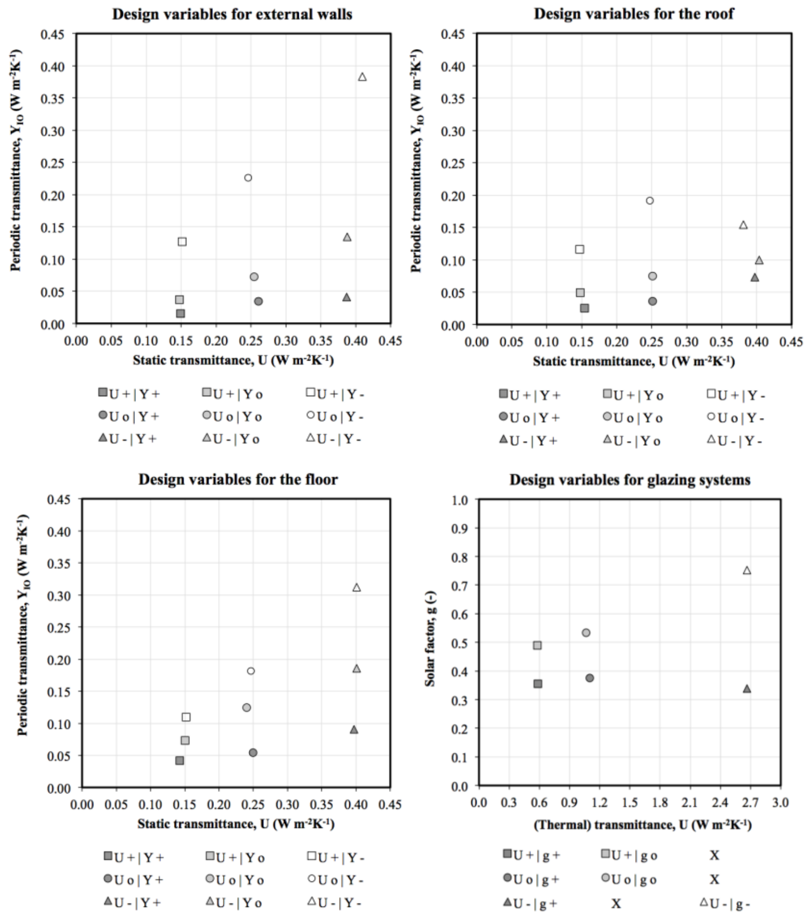

Figure 5: Available options for the design variables of the building envelope.

Finally, for each control strategy of the solar shading devices, four available set-point values are available:

- Set-point value of a zone's indoor air temperature: [23, 24, 25, 26] ºC
- Set-point value of the outdoor air temperature: [23, 24, 25, 26] ºC
- Set-point value of irradiance incident on a window: [80, 100, 120, 150] W/m^2.

The problem space of this optimization consists of 183 708 possible building variants, which clearly highlights the need to use an appropriate optimization technique that can guide the simulation engine towards the optimal solution without exploring explicitly all the building variants that form the entire problem space.

2.4.3 Optimization algorithm

Particle swarm optimization (PSO) is the optimization algorithm chosen to solve this scalarized single-objective optimization because of (i) its robustness in controlling parameters and (ii) its computational efficiency compared with other existing heuristic algorithms and (iii) the fact that it can be applied to nondifferentiable, nonlinear, and huge search problem spaces [55]. It is a heuristic population-based optimization algorithm developed by Kennedy and Eberhart [56], which is inspired by the social behavior of fish and birds in search of food. The searching units exploring the problem space that are looking for the optimal solution are called particles in the semantics of PSO. The swarm of particles aims at those areas characterized by lower values of the objective function where there is a minimization problem. The speed with which each particle moves depends on its best value compared with the overall best value found by the particles in its neighborhood. In order to improve the search-and-balance exploitation and exploration process of this swarm, particle velocity is controlled by the concept of *inertia weight*, developed by Shi and Eberhart [57]. Therefore, a particle swarm optimization with inertia weight (PSOIW) was implemented in GenOpt using the code reported in Code 1.

```
Algorithm{
      Main = PSOIW;
      NeighborhoodTopology = vonNeumann;
      NeighborhoodSize = 5;
      NumberOfParticle = 32;
      NumberOfGeneration = 40;
      Seed = 0;
      CognitiveAcceleration = 2.8;
      SocialAcceleration = 1.3;
      MaxVelocityGainContinuous = 0.5;
      MaxVelocityDiscrete = 4;
      InitialInertiaWeight = 1.2;
      FinalInertiaWeight = 0;
}
OptimizationSettings{
      MaxIte = 2000;
      MaxEqualResults = 100;
      WriteStepNumber = false;
      UnitsOfExecution = 0;
}
```

Code 1: PSOIW algorithm and the optimization settings implemented in the Command file GenOpt.

As recommended by van den Bergh and Engelbrecht [58], the number of particles and generations was set in order to obtain accurate results; specifically, 32 particles per generation was set for a total number of 40 generations. Moreover, a stopping criterion was also implemented in order to stop the optimization run if the algorithm should find 100 of the same instance of the best solution.

2.5 Stochastically generated occupancy and occupancy-dependent schedules

In our study, a probabilistic model that uses parameters and equations to evaluate the occupants' state was used to generate the probability distributions for a number of family-component types living in an apartment. Typical days were then stochastically extracted from the probability distributions for each family-component type and assembled to create family profiles corresponding to typical Chinese family types. Next, the daily family-component schedules assembled per family type were translated into room occupancy schedules suitable for the apartment types in the case study. Afterwards, the stochastically generated occupancy schedules for apartment types were applied to all the apartments in the model of the multiresidential building, with the aim of taking into account the proportions of the Chinese family types that accord with those detailed in the national census. In the following subsections, all the mentioned steps are described in detail.

While most of the techniques used to develop probabilistic models, such as logistic regressions, state-transition analyses using Markov chains, Monte Carlo modeling, and artificial neural networks, allow an automatic generation of occupancy schedules to be implemented in a whole-building simulation, these automatized processes cannot control family composition and the proportion of family types in the multiresidential building. To overcome these limits, a manual procedure was implemented by the students during the summer school to create *ad hoc* family types representing Chinese society (see Section 2.6.1). This allowed to take into account for the proportion of family types in the model in each occupancy schedule (for each room of each apartment) and occupant-dependent schedules (electric lighting and appliance usages and availability of heating and cooling) in each simulation. The adopted methodology was very time-consuming, but it permitted only reliable simulation scenarios, for example, scenarios with just children (i.e., without parents) or with only two nonworking people were not admitted.

2.5.1 Development of the schedules

In the Zhoukanghang buildings, there are three apartments on each storey: one small and two large (Figure 2). For each apartment, different types of family compositions were hypothesized.

In order to create a distribution of families that could represent a typical Chinese scenario, data were gathered from the Sixth National Chinese Census [59] and used to define the family types that were used in the building model. This census identifies the compositions for the main types of Chinese families: *nuclear families* (60.9%), *linear families* (23.0%), and *singles* (13.7%). A *nuclear family* is a traditional family composed of two parents and one child. A *linear family* is made up of two or more generations. Finally, *single* is a family type made up of a single working person living on their own. On the basis of these family types, several family-component types were identified, and the probability method proposed by Yamaguchi and Shimoda [25] was used to stochastically generate probability distributions of the occupant's activity for these family-component types. A trade-off between the accuracy of the characterization of family types and the total number of the analyzed alternatives was made, and only four family-component types were defined and used in this analysis. These four family-component types were defined on the basis of demographic attributes such as gender, age, and working activity and were used to create the four Chinese family types mentioned above:

- *Working person* (female or male). This family-component type is based on the data referring to a male aged from 20 to 65, who is a full-time worker who works both in the morning and the afternoon.
- *Nonworking person* (female or male). This family-component type is based on the data referring to a female without a job aged from 20 to 65 who takes care of a student who attends school (primary to high school).
- *Student* (female or male). This family-component type is based on the data referring to a male student attending junior high school.
- *Retired person* or *weekend* (female or male). This family-component type is based on the data referring to a female older than 65, who lives with a family. It is also used for the weekends of the *Working person*, *Nonworking person*, and *Student* schedules.

Required data about (i) the probability distribution of the duration of routine behaviors, (ii) the probability distribution of the beginning or ending times for routine behaviors, (iii) the percentage of people who adopt the behavior at time t (*PB*), and (iii) the probability distribution of duration for non-routine behaviors (*CF*) were taken from the outcomes of the national survey administered by Statistics Japan in 2006 [60]. Unfortunately, such data are not available for China, and a cross-cultural analysis between Japan and China was not found in the literature. However, even if typical daily activities carried out by Japanese and Chinese people might be slightly different, their energy usage in dwellings is more similar than to that of Western ones (specifically American and Canadian ones) [61]. For the purpose of this research, it is assumed that the behavior of typical Japanese family-components may be transferred to the Chinese context.

Moreover, according to the design instructions, the small apartment is designed for a couple whereas the two bigger apartments can host no more than three people. Nine types of family compositions were developed: four for the small apartment (indicated by S1, S2, S3, and S4) and five for the big ones (indicated by B1, B2, B3, B4, and B5). They are represented in Table 4.

Table 4: Types of family compositions for small and big apartments.

Family-component type	**Number of people per family-component type**								
	Small (Apartment B)				**Big (Apartments A and C)**				
	S1	**S2**	**S3**	**S4**	**B1**	**B2**	**B3**	**B4**	**B5**
Working person	1	2	1	0	3	2	1	2	1
Nonworking person	0	0	1	0	0	0	1	0	1
Child	0	0	0	0	0	1	1	0	0
Retired person	0	0	0	2	0	0	0	1	1
TOTAL	**1**	**2**	**2**	**2**	**3**	**3**	**3**	**3**	**3**

So, typical days were randomly extracted from the probability distributions of each family-component type and assembled to create specified family types and assigned to and merged together in each room (living room plus kitchen, double bedroom, and single bedroom)[5] of each apartment of the multiresidential building model. Therefore, on the base of the occupant's activity (sleeping, working/studying, not in the apartment, eating, housework, watching TV or engaging with other media, e.g., social media), hourly occupancy of each room was defined for each apartment in the models. During the composition of the room schedules, a control quality procedure was implemented to check that the number of occupants never exceeded the maximum allowed by the room type, for example, no more than two people in the double bedroom and no more than one person in the single bedroom, and by the apartment size. In this way, the randomly assigned and stochastically generated occupancy schedules should reflect a reliable daily occupancy time for each apartment.

One of the research questions investigated in this work is to what extent the creation of occupancy schedules and how they are linked to model thermal zones can influence the energy performance of a multiresidential building. Thus, randomized distributions of occupancy schedules were created to test (i) the effect of the method used to generate yearly schedules with an hourly resolution repeating a reference to a stochastically generated period (temporal randomness), such as a day, a week, or a month and (ii) the effect of spatial randomness of the occupancy schedules when they are linked to thermal zones.

[5] It is assumed that the bathrooms were unoccupied since the schedule time-step is 1 hour, and the constant presence of a person for that period or more was unlikely. The balconies, the basement, the entrance, and the stair block were also modeled as unoccupied.

Regarding temporal randomness, starting from the daily schedule of each family type (Table 4), yearly schedules were created using four temporal randomness approaches to allocate and combine the stochastically generated daily schedules:

- *Day-repeated schedule*: one random weekday and one weekend day were opportunely repeated 257 and 108 times respectively.
- *Week-repeated schedule*: one week composed of five random weekdays and two random weekends was repeated 52 times.
- *Month-repeated schedule*: one month composed of four random weeks was repeated 12 times.
- *Whole-year randomized schedule*: an entire year was composed of 365 random days. Weekdays and weekend days were opportunely allocated.

Regarding spatial randomness, two approaches were implemented:

- *Apartment-repeated schedules*: the same set of occupancy schedules were repeated for the same apartment of all the storeys.
- *Whole-building randomized schedules*: the set of occupancy schedules are randomized in all the apartments.

Applying this randomization process, 25 different schedules were generated for each type of temporal randomness approach, which were applied in the two spatial randomness approaches for a total of 200 occupancy scenarios; these were then implemented for both the two design quality proposals, Base case and Optimized case. The same randomization process was applied to the schedule of the appliances, the electric lighting, and the power system. Due to the direct link between these schedules and the occupancy schedules, no further randomization is needed. The occupant-dependent schedules (lighting, appliances, heating, and cooling availability schedules) were, accordingly, randomized as well. Specifically, the schedules for the electric appliance usage were directly linked to the occupancy of each room. They can take the binary values *on*/*off*, and it is *on* when at least one person is inside a given room in a given hour. As with the electric appliances, the electric lighting was also linked to the presence of the occupants, but lighting is always *off* from 0:00 A.M. to 5:00 A.M., and from 7:00 A.M. to 5:00 P.M. due to the presence of enough daylight. The heating and cooling systems were assumed to be autonomous for each apartment, so that they are linked to apartment occupancy, and the heating and cooling systems are *on* in an apartment when at least one person is inside a given apartment. Next, the coefficient of variation (C_V) was calculated to provide an overall assessment of the dispersion of frequency distribution of the energy needs in the Base case and in the Optimized case. This can be formulated as

$$C_V = \frac{\sigma}{\mu} \tag{5}$$

where σ is the standard deviation of a frequency distribution and μ is its mean.

2.5.2 Statistical analysis

The following statistical analysis was carried out on the seven variables using the software package IBM® SPSS® Statistics version 21: design proposal, temporal randomness, spatial randomness, energy need for heating, energy need for cooling, energy use for electric lighting, and energy use for electric appliances. However, the influence of the design proposal was not analyzed in relation to the energy use for electric lighting or the energy use for electric appliances since there is no difference in this use between the Base case and the Optimized case according to the assumptions mentioned in Section 2.5. Specifically, *Design proposal* is a within-groups independent, binary, and balanced variable that takes the two values *Base case* and *Optimized case*. *Spatial randomness* is a between-groups independent, binary, and balanced variable that takes the two values *Apartment-repeated schedules* and *Whole-building randomized schedules*. *Temporal randomness* is a between-groups independent, categorical, and balanced variable that takes the four values *Day-repeated schedules*, *Week-repeated schedules*, *Month-repeated schedules*, and *Whole-year randomized schedules*. *Energy need for heating, energy need for cooling, energy use for electric lighting, and energy use for electric appliances* are dependent and continuous variables expressed in kWh/(m^2 a). Several statistical techniques were adopted to extract patterns and general findings from the dataset created by collecting all the data exported from the simulations. A few techniques aimed to explore relationships between groups of data, and others compared groups in the dataset. The continuous variables were tested for normality for each category of the independent variables using the Shapiro-Wilk's statistic [62], due to the dimension of the samples. Since this testing of the continuous variables produced results that were not normally distributed for $p \leq 0.05$ (see Table 6), Spearman's Rank Order Correlation was used to explore the strength of correlation among continuous variables, and non-parametric statistics were adopted to explore the differences among groups of data. Wilcoxon Signed Rank test was used to statistically characterize the energy performance of the building in the two proposed design qualities: *Base case* and *Optimized case*. The Mann-Whitney U test was used to assess the dependency of energy quantities on spatial randomness that takes two values: *Apartment-repeated schedules* and *Whole-building randomized schedules*. The Kruskal-Wallis H test tested for the dependency of energy quantities on temporal randomness that takes four values: *Day-repeated schedules*, *Week-repeated schedules*, *Month-repeated schedules*, and *Whole-year randomized schedules*. Finally, the Wilcoxon Signed Rank Test and the Mann-Whitney U test were used to estimate the magnitude of the difference between each pair of energy terms by computing the effect size defined by Cohen [63] as

$$r \equiv \frac{z}{\sqrt{N}} \tag{6}$$

where r is the effect size, z is the statistic's value and N is the sample size. According to Cohen [63], the effect size is large if the value of r varies more than 0.5, medium if it varies by around 0.3, and low if is around 0.1.

3 Results and discussion

The main outcomes of the article are presented and discussed in the following subsections. Specifically, the quality check of the numerical modeling is discussed in Section 3.1, its mathematical optimization is presented in Section 3.2, and the statistical analyses exploring the impact of stochastically generated schedules on the building's energy performance are reported in Section 3.3.

3.1 Quality check of the Base case model

The building has not been built yet and measured data are not available. Hence, a quality check procedure was implemented to minimize the errors that could have been caused because of the interpretation of design information, as well as the construction of the numerical model. From two modeling groups, called Group A and Group B, two building models were developed avoiding reciprocal interference, both starting from the same design package. The hourly indoor operative temperature in the living room and the kitchen was adopted as a benchmark, and the criteria for hourly data, recommended for calibration by the ASHRAE Guideline 14 [44], were used to check the models' quality. Figure 6 shows the comparison of the simulation outcomes produced by the two groups of students for the initial model and for the model refined after the tutor applied an independent control.

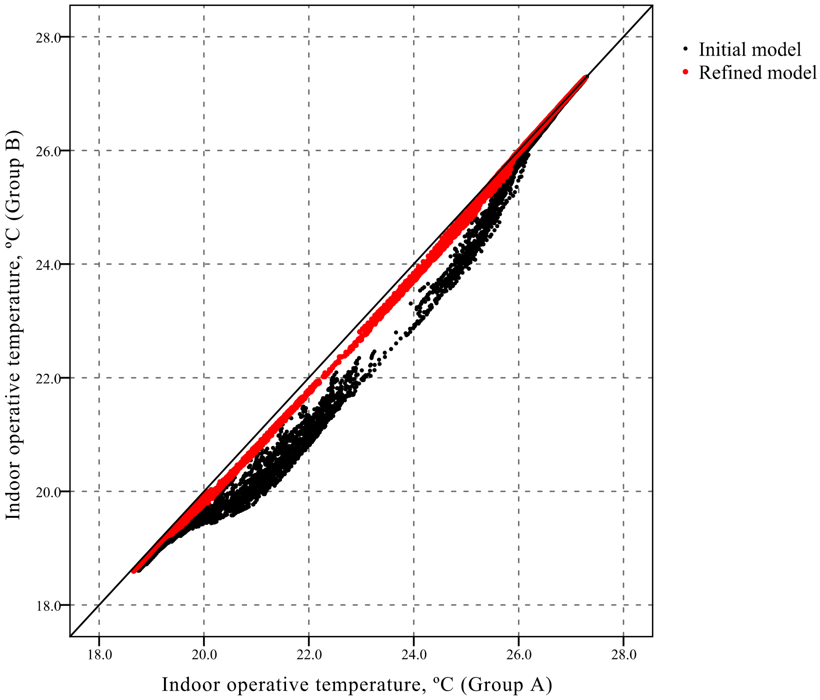

Figure 6: Quality check of the two numerical models of the Base case.

The initial comparison showed a substantial difference in the temperatures calculated by the two models: *MBE* = 33.4% and $C_V(RMSE)$ = 2.0%. The differences were not randomized but were affected by a strong bias (Figure 6). During the independent control carried out by comparing the two source files, several differences were identified: the type of indoor temperature used for the set-point, the height of the rooms, the degree of orientation of the building, the construction used for the floor of the first storey (i.e., against the basement), the material used for the window frames, the thermo-physical properties of a few materials, and the monthly average temperatures of the ground below the building. All of the identified differences were fixed, and the comparison of the two refined models showed *MBE* = 9.1% and $C_V(RMSE)$ = 0.6%, which met both of the ASHRAE Guidelines 14 criteria, but a few differences still affected the two models. Therefore, to find the sources of these differences, the two models were exported in the .idf format, and the two texts were automatically compared. The two models were affected by small geometrical differences regarding the entire building and some rooms: by a slightly different position of the windows on the façades, and by marginally different dimensions of balconies. Finally, the model that more accurately represented the architectural design concept was chosen to represent the Base case model.

3.2 Identification of the optimal building variant

The simulation engine executed 1440 simulations in 13 h 13′ 42″ on a workstation equipped with an Intel Xeon Processor E5-1660 v3 (8-Core, 20 MB Cache, 3.0 GHz Turbo), 16 GB RAM (2 133 MHz, DDR4) and an AMD FirePro W4100 graphic card (2 GB of dedicated memory). The optimization engine combined 727 different building variants throughout the entire optimization: the optimal building variant was identified four times as a possible solution. It was tested the first time in the 33rd generation after 1180 simulations and was the 702nd building variant combined by the algorithm. Figure 7 shows the evolution of the values taken by the objective function throughout the optimization process.

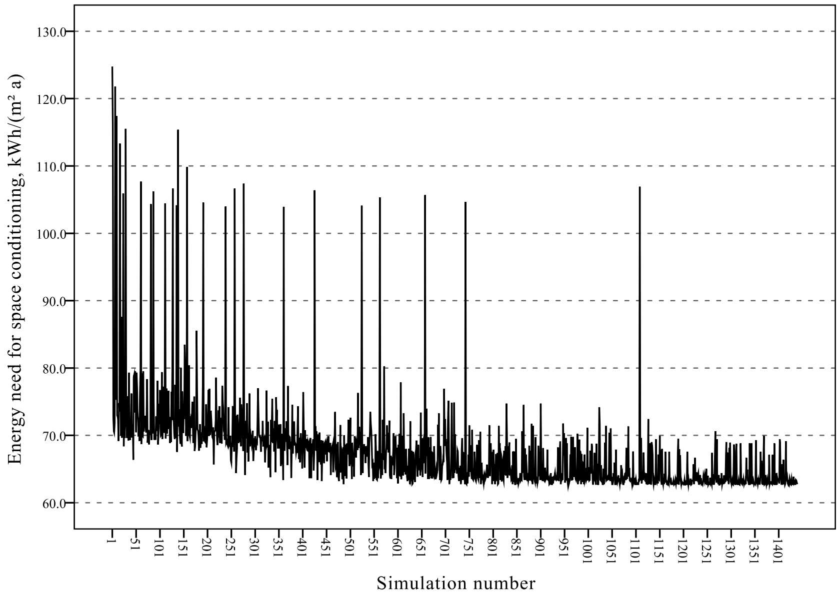

Figure 7: Minimization of the objective function throughout the optimization run.

With respect to the Base case, the optimal building variant reduces the total energy need for space conditioning from 124.8 to 62.7 kWh/(m^2 a) with a percentage saving (evaluated with respect to the Base case) of 78% and 26% respectively for heating and cooling. This optimal building variant is characterized by 16.2 and 46.5 kWh/(m^2 a) respectively for heating and cooling whereas the minimum values of energy needs for heating and cooling identified during the entire optimization are 12.9 and 43.2 kWh/(m^2 a) respectively. It should be recalled that the energy need for cooling considers both sensible and latent contributions.

Figure 8 shows that the two objective functions (energy need for heating and energy need for cooling) are antagonistic in the present case study, and a multi-objective optimization could provide a more detailed outcome in the form of a Pareto front [47].

Finally, the options for all design variables that characterize the optimal building variant are reported in Table 5. The optimal building variant is characterized by a highly insulated building envelope, high-performance glazing systems, and external solar shading that has to effectively prevent unwanted solar gains. The identified optimal building variant is called Optimized case outside this section.

Table 5: Technical features of the optimal building variant.

Design variable	**Code**	*** Steady-state transmittance, U, W/(m²K)** **◊ value**	**† Periodic transmittance, Y, W/(m²K)** **‡ Solar factor, g (%)**
External walls	*U+ | Yo*	0.15*	0.04†
Flat roof	*U+ | Yo*	0.15*	0.05†
Basement floor	*U - | Y+*	0.40*	0.30†
Glazing	*U+ | g+*	0.60*	35%‡
Solar shading control strategy	*InTemp*	Internal air temperature◊	
Set-point value for solar control strategy	*InTemp23*	23 ºC◊	

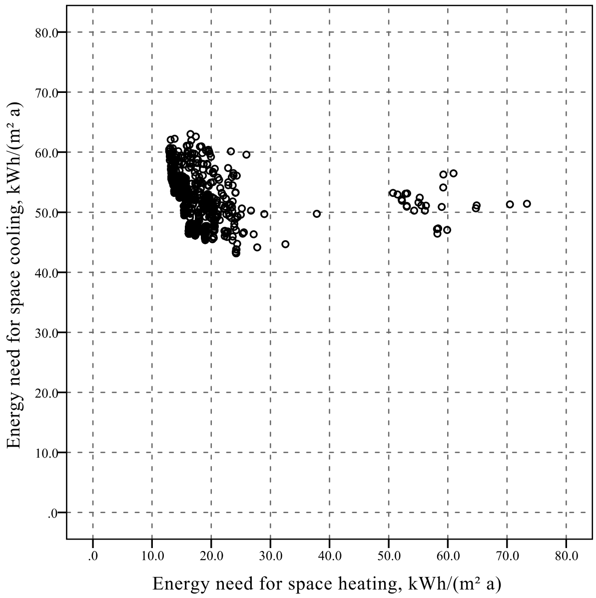

Figure 8: Scatterplot of the energy needs for heating and cooling of the building variants simulated in the optimization.

3.3 Impact of stochastic schedules on the building's energy performance

A total of 400 simulations were carried out to assess the impact of stochastically generated schedules for occupancy and occupancy-dependent input variables (electric lighting usage, electric appliances usage, and availability of active heating and cooling in each apartment) on the energy performance for heating and cooling in the Base case and in the Optimized case. Specifically, three dimensions of the problem were investigated: (i) the design proposals, (ii) the spatial randomness of occupancy schedules, and (iii) the temporal randomness of occupancy schedules. These dimensions are expressed with categorical data. In particular, the names used to describe the design proposals, that is, the different levels of the building envelope quality and of the integration of passive strategies are *Base case* (described in Section 2.1) and *Optimized case* (modeled in Section 2.2 and described in Section 3.2). The names used to describe spatial randomness of schedules are *Apartment-repeated schedules* and *Whole-building randomized schedules* (both described in Section 2.5). The names used to describe temporal randomness of schedules are *Day-repeated schedules, Week-repeated schedules, Month-repeated schedules*, and *Whole-year randomized schedules* (all described in Section 2.5). Figure 9 shows the outcomes of simulations as frequency distributions of the energy quantities for both design proposals, Base case and Optimized case. Table 6 summarizes the principal statistical descriptors of the energy distributions.

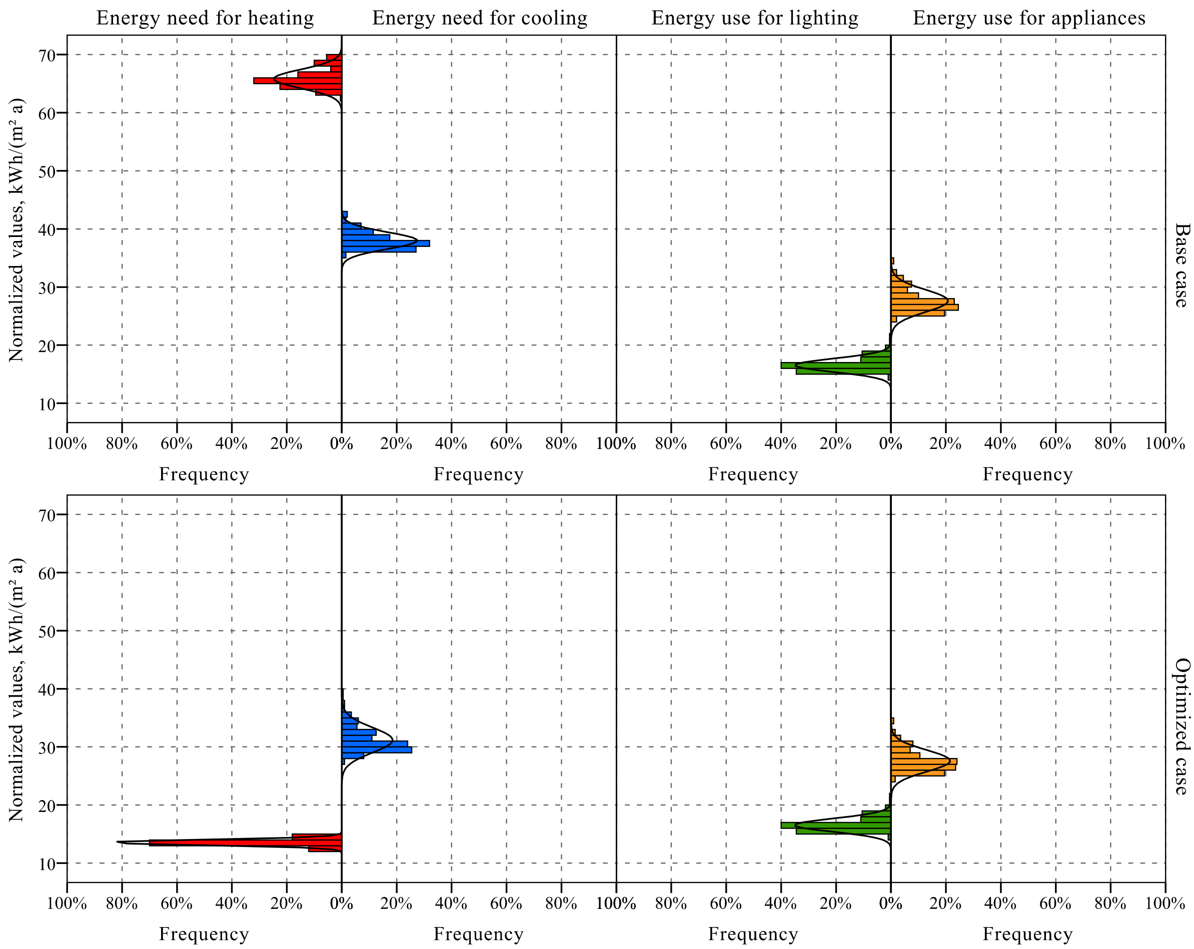

Figure 9: Frequency distributions of the energy quantities for each design quality category.

The energy needs for heating and cooling are much higher in the poorly insulated multiresidential buildings. A high-performance building envelope can drastically reduce the heating need of the building by about four times, while cooling need (sensible plus latent) can be reduced by about 25%. Furthermore, it emerges that the energy need for cooling is higher than the heating need for high-performance residential buildings in the HSCW climate zone of China, as already mentioned in Ref.[64]. Furthermore, simulation outcomes show that the spread of values of the energy need for heating is reduced in low-energy buildings due to several occupant patterns. This result agrees with the findings of the final report of the IEA-EBC Annex 53 [22]. The coefficient of the variation of the frequency distributions of the energy needs for heating and cooling shows that, even if the spread of energy needs is reduced in low-energy buildings due to several occupant patterns, the spread of the frequency distribution standardized by the distribution mean for the optimized building is from 42% to 82% higher than for the building in the Base case (Table 6). Therefore, there is a need for further investigation of the robustness of low-energy buildings against occupant behavior.

According to the assumptions mentioned in Section 2.5, energy uses for electric lighting and appliances only depend on room occupancy and are hence the same in both the Base case and the Optimized case.

Table 6: Principal descriptive statistics of the energy distributions grouped by design quality categories.

	N	Max.*	Min.*	Mean*	Median*	St.dev*,†	C_V**	Skewness	Kurtosis	Shapiro-Wilk *p*-value
					Base case					
Energy need for heating	200	69.85	62.88	65.81	65.51	1.61	2.4	0.69	-0.21	0.000
Energy need for cooling	200	42.31	35.80	38.00	37.77	1.45	3.8	0.87	0.27	0.000
Energy use for electric lighting	200	21.19	14.98	16.54	16.19	1.14	6.9	1.40	2.10	0.000
Energy use for electric appliances	200	34.32	24.79	27.62	27.08	1.92	6.7	1.10	0.79	0.000
					Optimized case					
Energy need for heating	200	14.48	12.11	13.57	13.68	0.47	3.4	-0.71	0.26	0.000
Energy need for cooling	200	39.01	27.93	31.18	30.59	2.16	6.9	1.11	0.96	0.000
Energy use for electric lighting	200	21.19	14.98	16.54	16.19	1.14	6.9	1.40	2.10	0.000
Energy use for electric appliances	200	34.32	24.79	27.62	27.08	1.92	6.7	1.10	0.79	0.000

* Values in kWh/(m^2 a).
† St.dev stands for standard deviation.
* C_V stands for coefficient of variation that is expressed in %.

The statistical analysis pursues two main purposes: (i) the statistical characterization of the energy performance of the building in the two design proposals and (ii) the estimation of the effects of spatial and temporal randomness on buildings' energy performance for heating and cooling.

First, a normality test was carried out to address the selection of suitable statistical methods and tests. All the dependent variables are affected by skewness and kurtosis[6]. A Shapiro-Wilk's test ($p > 0.05$) showed that all the depended variables reject the null hypothesis for each category of the independent variables and that they are non-normally distributed.

Regarding the first purpose of the statistical analysis, Figure 9 shows a clear reduction in the energy needs for heating and cooling and that the energy uses for electric lighting and appliances are identical by definition for the Base case and the Optimized case. The Wilcoxon Signed Rank Test quantified a statistically significant reduction in the energy needs for heating and cooling in the Optimized case, for both $z = -12.26$[7] and $p < 0.01$, with a large effect size ($r = 0.87$) according to Cohen [63]. The median score on the energy need for heating decreases from the Base case, $Md = 65.51$ kWh/(m^2 a), to the Optimized case, $Md = 13.68$ kWh/(m^2 a), and the energy need for cooling decreases from the Base case, $Md = 37.77$ kWh/(m^2 a), to the Optimized case, $Md = 30.59$ kWh/(m^2 a).

The relationships between all the energy quantities in the two design proposals are assessed using Spearman's Rank Order Correlation (Table 7).

In general, a Spearman's Rank Order Correlation shows that the strength of the linear relationship between all the pairs of energy quantities is stronger in the Optimized case than in the Base case. Therefore, the influence of occupant-dependent quantities, such as the availability schedules for electric lighting and appliances, on energy needs for cooling and heating is stronger in high-performance buildings than in poorly insulated buildings.

[6] Even if they are not the cause for rejecting the normality hypothesis since the *z*-values are both lower than 1.96 for all the continuous variables.
[7] We recall that a nonparametric statistic converts scores to ranks and compares them.

Table 7: Correlation analysis among the energy quantities. The values referring to the Base case are shaded in light gray, and those referring to the Optimized case are shaded in dark gray.

Spearman's Rank Order Correlation	Energy need for cooling	Energy need for heating	Energy use for lighting	Energy use for appliances
Energy need for cooling		0.361[*,**]	0.698[*,**]	0.929[*,**]
Energy need for heating	-0.651[*,**]		-0.056[*,**]	0.362[*,**]
Energy use for lighting	0.809[*,**]	-0.742[*,**]		0.719[*,**]
Energy use for appliances	0.985[*,**]	-0.583[*,**]	0.719[*,**]	

[*] Correlation is significant at the 0.01 level (2-tailed).
[**] N = 200.

The energy need for cooling in the Base and the Optimized case is strongly and positively related to energy use for both electric lighting and appliances, meaning that internal electric gains contribute to increasing the need for cooling in a statistically significant way in the HSCW climate. On the contrary, the energy need for heating is quite strongly and negatively correlated with energy uses for both electric lighting and appliances only in the Optimized case. In this sense, electric internal gains can effectively contribute to reduce heating need only in highly insulated, low-energy buildings. In other words, even if the absolute effect of electric internal gains is the same for both building types, their relative contribution in covering the total heating requirement is higher in the Optimized case due to its much lower heating demand. Finally, the energy needs for heating and cooling are characterized by a weak but positive correlation in the Base case, that is, when one increases, the other is also likely to increase, whereas the correlation is stronger and negative in the Optimized case, that is, when one increases, the other is quite likely to decrease. These last two behaviors are well represented in Figure 10. This means that, assuming occupants do not deliberately modify their seasonal behavior, some virtuous behaviors can reduce the overall energy performance of poorly insulated buildings, diminishing both heating and cooling needs. For highly insulated buildings, however, occupant behavior has an antagonistic effect on energy needs for heating or cooling, that is, if the former is reduced, the latter has a tendency to increase. Furthermore, Figure 10 shows better than Figure 9 that, given the same occupancy schedules, the spread of the energy need for cooling is much higher in the Optimized case than in the Base case, that is, the sensitivity to the impact of user behavior has a tendency to become greater with a better insulated building. On the contrary, given the same occupancy schedules, the spread of the energy need for heating is lower in the Optimized case compared to the Base case, confirming what has already been found by Polinder, Schweiker, van der Aa, Schakib-Ekbatan, Fabi, Andersen, Morishita, Wang, Corgnati, Heiselberg, Yan, Olesen, Bednar and Wagner [22]. This means that low-energy buildings are, from a heating perspective, more robust against occupant behavior than poorly insulated buildings.

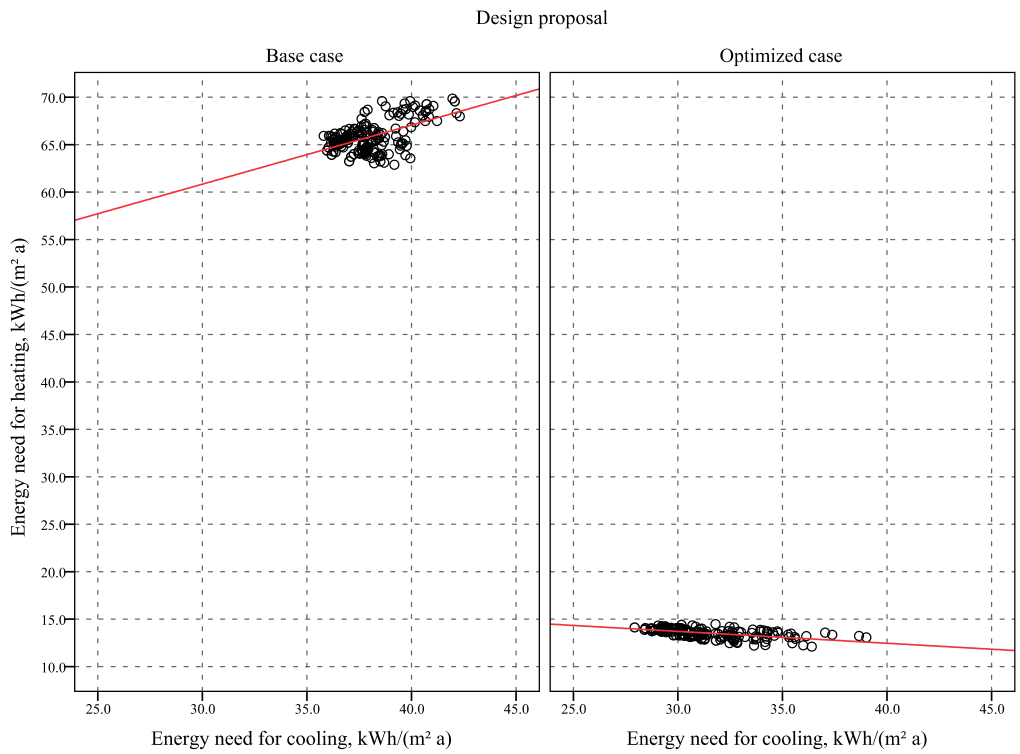

Figure 10: Scatterplots of the energy needs for heating and cooling in the Base case and the Optimized case. The linear regression lines for the two set of data are shown in red.

The second purpose of the statistical analysis aimed to identify possible patterns in the energy demands of the building caused by the different assumptions used to create the spatial and temporal randomized schedules. Regarding spatial randomness, a Mann-Whitney *U* test revealed:

- A statistically nonsignificant difference in the energy need for heating in the two spatial randomization scenarios, U = 17 905.5, z = –1.81, and p = 0.07, with a small effect size (r = 0.13) according to Cohen [63]. The median score for the energy need for heating remains the same for *Apartment-repeated schedules, Md* = 38.68 kWh/(m^2 a), and *Whole-building randomized schedules, Md* = 38.68 kWh/(m^2 a).
- A statistically nonsignificant difference in the energy use for electric lighting in the two spatial randomization scenarios, U = 19 538.0, z = –0.40, and p = 0.69, with a small effect size (r = 0.03) according to [63]. The median score for the energy use for electric lighting is slightly higher for *Apartment-repeated schedules*, *Md* = 16.30 kWh/(m^2 a), than for *Whole-building randomized schedules*, *Md* = 16.13 kWh/(m^2 a).
- A statistically significant difference in the energy need for cooling in the two spatial randomization scenarios, U = 17 037.5, z = –2.56, and p = 0.01, with a small effect size (r = 0.18) according to [63]. The median score for the energy need for cooling is slightly higher for *Apartment-repeated schedules, Md* = 36.16 kWh/(m^2 a), than for *Whole-building randomized schedules, Md* = 35.64 kWh/(m^2 a). It is relevant to recall that Shanghai is classified as being in a humid, subtropical climate zone, indicated with Cfa in the Köpper and Greiger's classification [28].
- A statistically significant difference in the energy use for electric appliances in the two spatial randomization scenarios, U = 14 016.0, z = –5.176, and p < 0.01, with a medium effect size (r = 0.37)

according to [63]. The median score for the energy need for heating remains almost the same for *Apartment-repeated schedules, Md* = 27.44 kWh/(m^2 a), and *Whole-building randomized schedules, Md* = 26.83 kWh/(m^2 a).

On the basis of the Mann-Whitney *U* test outcomes, it can be stated that a detailed definition of occupancy schedules for all the individual apartments is important to accurately estimate the energy need that is required to face the predominant local climate challenge, for example, cooling in summer-dominated climates. Moreover, a detailed definition of occupancy schedules for all the individual apartments is also important for estimating the occupancy-dependent energy use for electric appliances. However, it is not statistically relevant for estimating the occupancy-dependent energy use for electric lighting. This might be due to a stronger dependency on occupancy for energy use for electric appliances than for energy use for electric lighting, since, during the daytime in the simulations, lighting is assumed to be switched off whether or not the room is occupied.

Regarding temporal randomness, a Kruskal-Wallis *H* test revealed a statistically significant difference among the four approaches used to generate the sets of schedules for each apartment. A total of 100 schedules were created for each of the four approaches. Specifically, the outcomes of the test are:

- A statistically significant difference in the energy need for heating across the four temporal-randomness approaches, χ^2 (3 *df*) = 25.24 and $p < 0.01$. The lowest median score for the energy need for heating is achieved by *Day-repeated schedules*, *Md* = 38.51 kWh/(m^2 a), then by *Month-repeated schedules*, *Md* = 39.04 kWh/(m^2 a), *Week-repeated schedules*, *Md* = 39.10 kWh/(m^2 a), and eventually by *Whole-year randomized schedules, Md* = 39.35 kWh/(m^2 a).
- A statistically significant difference in the energy need for cooling across the four temporal-randomness approaches, χ^2 (3 *df*) = 36.91 and $p < 0.01$. The lowest median score for the energy need for heating is achieved by *Month-repeated schedules*, *Md* = 35.70 kWh/(m^2 a), then by *Whole-year randomized schedules*, *Md* = 35.79 kWh/(m^2 a), *Week-repeated schedules*, *Md* = 36.26 kWh/(m^2 a), and eventually by *Day-repeated schedules, Md* = 37.03 kWh/(m^2 a).
- A statistically significant difference in the energy use for electric lighting across the four temporal-randomness approaches, χ^2 (3 *df*) = 143.88 and $p < 0.01$. The lowest median score for the energy need for heating is achieved by *Whole-year randomized schedules, Md* = 15.95 kWh/(m^2 a), then by *Month-repeated schedules, Md* = 16.12 kWh/(m^2 a), *Week-repeated schedules*, *Md* = 39.10 kWh/(m^2 a), and eventually by *Day-repeated schedules*, *Md* = 38.51 kWh/(m^2 a).
- A statistically significant difference in the energy use for electric appliances across the four temporal-randomness approaches, χ^2 (3 *df*) = 117.59 and $p < 0.01$. The lowest median score of energy need for heating is achieved by *Whole-year randomized schedules, Md* = 25.98 kWh/(m^2 a), then by *Month-repeated schedules*, *Md* = 26.33 kWh/(m^2 a), by *Week-repeated schedules*, *Md* = 27.44 kWh/(m^2 a), and eventually by *Day-repeated schedules, Md* = 28.50 kWh/(m^2 a).

In order to assess to what extent the six available pairs of schedules differ for each energy quantity, some *post-hoc* Mann-Whitney *U* tests were carried out between pairs of groups. According to Cohen [63], the effect size is large if the value of *r* varies by more than 0.5, medium if it varies by around 0.3, and low if it varies by around 0.1. A summary of the analysis is reported in Table 8.

Table 8: Effect size according to Cohen [63] of the differences in the energy quantities for each of the four temporal randomness approaches used to create the occupancy schedules.

Paired simulation outcomes		**Energy use for electric lighting**	**Energy use for electric appliances**	**Energy need for cooling**	**Energy need for heating**
Day-repeated schedules	Month-repeated schedules	Large	Large	Medium	Small
Day-repeated schedules	Whole-year randomized schedules	Large	Large	Medium	Small
Week-repeated schedules	Whole-year randomized schedules	Medium	Large	Small	Small
Day-repeated schedules	Week-repeated schedules	Large	Small	Small	Small
Month-repeated schedules	Whole-year randomized schedules	Large	Small	Small	Small
Week-repeated schedules	Month-repeated schedules	Small	Medium	Small	Small

Therefore, at least for the climate of Shanghai and tanking into account the simulation assumptions mentioned in Section 2.2, according to a Kruskal-Wallis H test and six follow-up Mann-Whitney U tests, (i) the differences in the energy need for heating caused by the four temporal randomness approaches are always statistically significant but are always small; (ii) the differences in the energy use for electric lighting caused by the four temporal randomness approaches are always statistically significant and are quite large; (iii) the differences in the energy use for electric appliances caused by the four temporal randomness approaches are always statistically significant and the effect size is high when comparing approaches separated by at least one temporal category; (iv) the differences in the energy need for cooling caused by the four temporal randomness approaches are always statistically significant and the effect size is medium when comparing the most repeated approach (*Day-repeated schedules*) with the most randomized ones (*Month-repeated schedules* and *Whole-year randomized schedules*); (v) for all the energy quantities, the difference between the *Month-repeated schedules* and the *Whole-year randomized schedules* is negligible when compared with the *Day-repeated schedules*. However, it should be noted that such outcomes strictly depend on the assumptions made, and additional work would be needed to generalize them. For example, if individual differences in preferred set-point temperatures or diverse usage patterns were considered, some of the findings shown above might be characterized differently.

Finally, it is possible to summarize as follows: (i) the different approaches to creating temporal randomized schedules cause a statistically significant difference in all the analyzed energy quantities, which affects the internal electric gains more since they are directly dependent on the occupancy schedules rather the heating and cooling needs and (ii) intermediate temporal randomness approaches (i.e., *Week-repeated schedules* and *Month-repeated schedules*), do not appreciably change the assessment with respect to the closest extreme (i.e., *Day-repeated schedules* and *Whole-year randomized schedules*, respectively), and therefore it is suggested that *Whole-year randomized schedules* should be used generally (this is stated on the basis of previous findings too). However, (iii) for simplified simulations, *Day-repeated schedules* could be sufficient if the main interest is solely energy performance.

Finally, if the energy needs for heating and cooling are depicted with respect to spatial and temporal randomness for the Base case and the Optimized case (Figure 11 and Figure 12 respectively), it is possible to observe that (i) the spread is appreciably reduced in the space- and temporal-randomized schedules, (ii) apartment-repeated schedules are generally characterized by higher energy needs both for heating and for cooling, and (iii) a higher randomization of temporal schedules causes an increase in the energy need for heating and a reduction in the energy need for cooling.

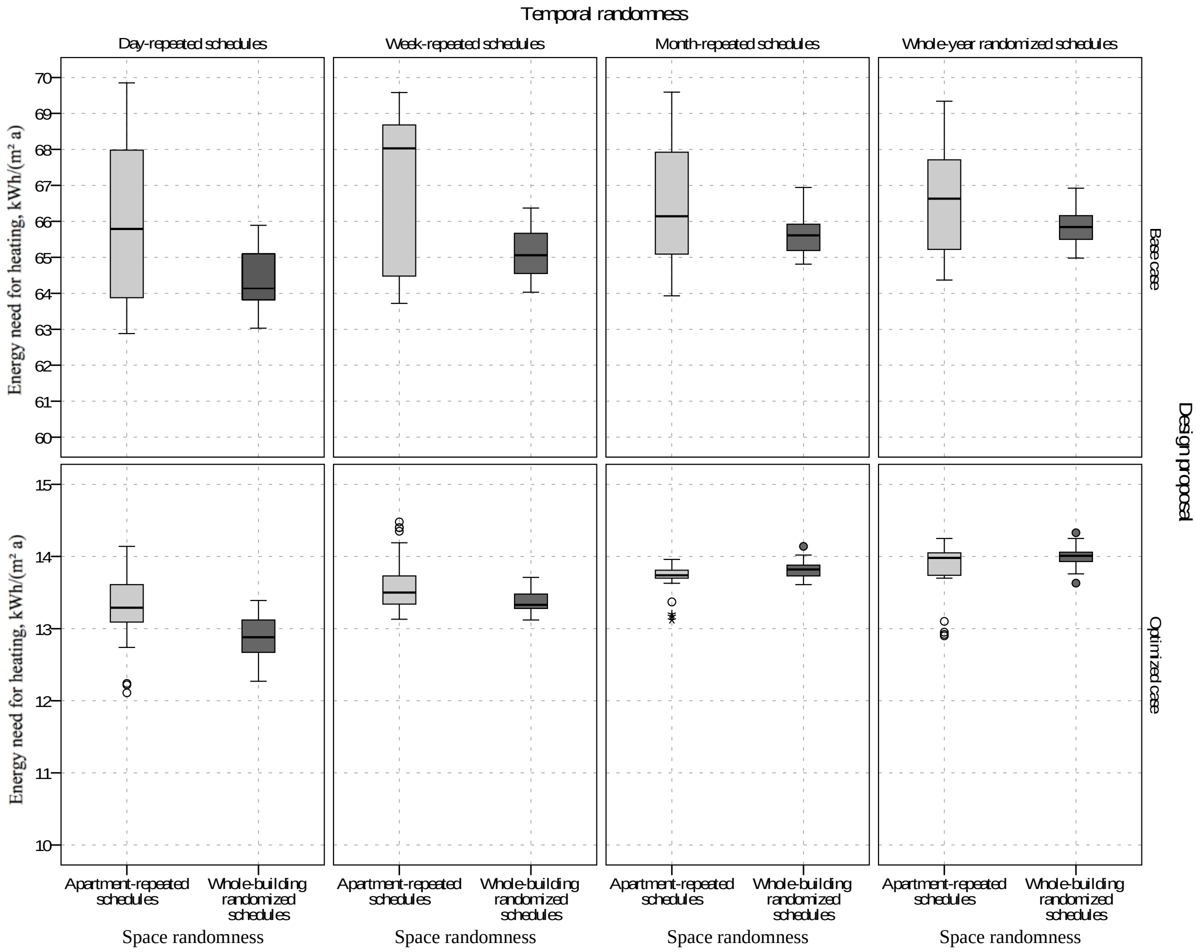

Figure 11: Energy need for heating depicted with respect to spatial and temporal randomness for the Base and Optimized cases.

Regarding the medians (*Md*) of the distributions, Table 9 shows the relative uncertainty (U_{Md}) due to spatial and temporal randomness for the energy need for heating and cooling.

Table 9: Relative uncertainty due to spatial and temporal randomness for the energy needs for heating and cooling.

Relative uncertainty	Energy need for heating		Energy need for cooling	
	Base case	Optimized case	Base case	Optimized case
Spatial randomness	4.4%	3.8%	4.0%	3.4%
Temporal randomness	4.7%	8.6%	4.6%	9.8%
Spatial and temporal randomness	5.8%	8.6%	6.4%	9.8%

Relative uncertainty (U_{Md}) is defined as

$$U_{Md} \equiv \frac{Md_{max} - Md_{min}}{Md_{max}} . \qquad (7)$$

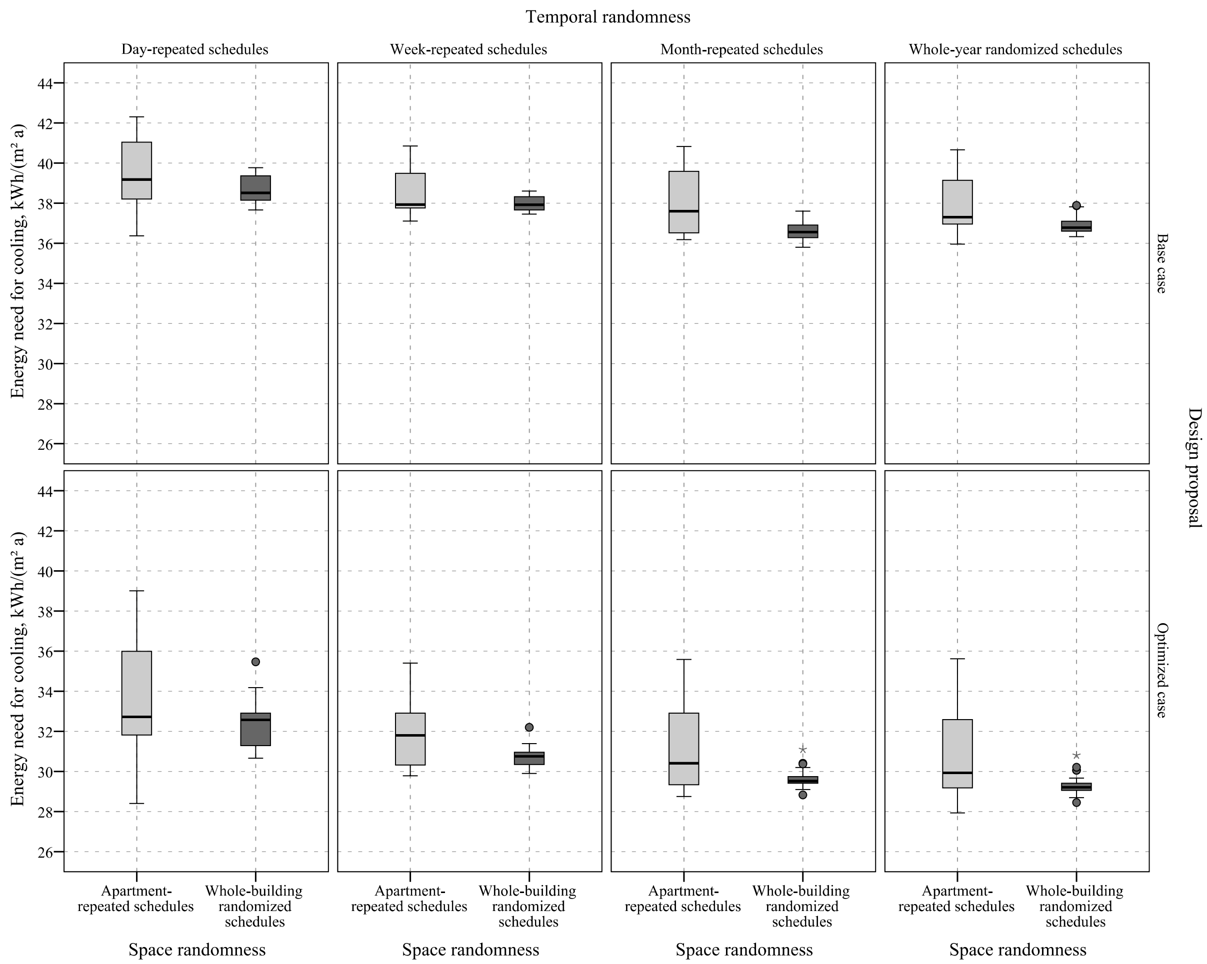

Figure 12: Energy need for cooling depicted with respect to spatial and temporal randomness for the Base and Optimized cases.

The principal findings of this analysis are:

- Uncertainty in the energy performance due to spatial randomness seems to have less weight than temporal randomness and is around 5%. This might be because of the similar structure of the apartments.
- Uncertainty in the energy performance due to temporal randomness is in the order of 10%.
- Uncertainty due to occupant modeling is higher in low-energy buildings than in poorly insulated ones.
- Uncertainty due to occupant modeling is almost the same for the energy needs for heating and cooling.

4 Conclusions

This study has its origins in the *Sustainable Energy in Cities* summer school hosted in Shanghai in July 2015 and jointly organized by the Norwegian University of Science and Technology (NTNU) and the Shanghai Jiao Tong University (SJTU).

Fifteen students from Europe and China with different academic backgrounds worked intensely together for two weeks using: (i) an optimization technique to improve the quality of and reduce the energy needs for heating and cooling of a new neighborhood of high-rise residential buildings located in Shanghai and (ii) computer-based, dynamic, whole-building simulations to estimate the impact of occupant behavior on the energy performance of such buildings. During the months following the summer school, the work was further developed, refined, and clarified by a more limited number of students and the tutors.

The article has three main purposes: (i) to carry out a quality check of a whole-building numerical model for which calibration cannot be executed, (ii) to adopt a scalarized single-objective optimization to minimize the energy need for space conditioning of a reference building, and (iii) to explore the implications of temporal and spatial randomness of stochastically generated occupancy schedules on the energy performance of the reference building through several statistical tools.

Regarding the first purpose, a seven-step quality assurance procedure is presented in this article and a method that can be used for the quality assessment of the whole-building model is suggested. However, the presented procedure cannot guarantee that the simulation outcome reflects the actual performance of the building, but it can help to minimize at least specification uncertainty (due to the interpretation of the design specification) and modeling uncertainty (due to the construction of the geometrical model and the selection of the physical models).

Regarding the optimization of the high-rise residential building block, a particle swarm optimization algorithm with inertia weight (PSOIW) guided a scalarized single-objective optimization that reduced the total energy need for space conditioning from 124.8 to 62.7 kWh/(m^2 a), that is, by 50.0%, with a percentage reduction in heating of 78% and cooling of 26%, evaluated with respect to the current design proposal. It should be recalled that the energy need for cooling considers both sensible and latent contributions. The optimal building variant for the discussed optimization problem is characterized by a highly insulated building envelope, high-performance glazing systems with low thermal transmittance and a low solar factor, and automated external solar shading that has to effectively prevent unwanted solar gains.

Regarding the third purpose, the outcomes of 400 simulations were statistically analyzed to assess the impact of stochastically generated schedules for occupancy and occupancy-dependent input variables on the energy performance for heating and cooling of the building model in the current design proposal and in the optimized design proposal. The following findings were found:

- The influence of occupant-dependent quantities, such as internal electric gains, on the energy needs for heating and cooling is statistically significant and is stronger in high-performance buildings than in poorly insulated buildings.
- Internal electric gains contribute to increasing the cooling need in all buildings located in the HSCW climate zone.
- Internal electric gains contribute in an effective manner to reducing the heating need only in highly insulated, low-energy buildings, where the relative contribution is higher due to a much lower heating demand.
- A detailed definition of occupancy schedules for all the individual apartments is important to accurately estimate the energy need that is required to face the predominant local climate challenge, for example, cooling in summer-dominated climates; however, further analysis for a winter-dominated climate is required to confirm this behavior and to allow a full generalization of this finding.
- The different approaches to creating temporal randomized schedules causes a statistically significant difference in all the energy quantities analyzed, which affects the internal electric gains more since they are directly dependent on the occupancy schedules rather than on the heating and cooling needs.

- Intermediate temporal randomness approaches, i.e. *Week-repeated schedules* and *Month-repeated schedules*, do not change the assessment appreciably with respect to the closest extreme; therefore, the following is suggested:
    - generally, *Whole-year randomized schedules* should be used, and
    - *Day-repeated schedules* could be sufficient for simplified assessment if the main interest is solely energy performance of the building.
- Uncertainty in the energy performance due to spatial randomness is estimated to be in the order of 5%, whereas uncertainty due to temporal randomness is in the order of 10%.
- Uncertainty due to occupant modeling is higher in low-energy buildings than in poorly insulated ones and is almost the same in magnitude for the energy needs for heating and cooling.

Finally, it is possible to state that an accurate modeling of high-performance buildings requires a spatially detailed and temporally-precise description of occupancy and occupant-dependent input variables for each thermal zone of the building even if this implies that more effort and increased costs might be needed to achieve such modeling.

5 Acknowledgments

This research was developed in the framework of the *Sustainable Energy in Cities* summer school that was partially supported by the Norwegian Centre for International Cooperation on Education [UTF-2014/10069]. The authors wish to thank all the students involved in the summer school for their great commitment and all the participants of the IEA-EBC Annex 66 entitled *Definition and Simulation of Occupant Behavior in Buildings* for the inspiring discussions.

6 References

[1] UNEP, SBCI, Buildings and climate change - Summary for Decision-Makers, United Nations Environment Programme and Sustainable Buildings & Climate Initiative, Paris, France, 2009.

[2] IEA, Energy balances of non-OECD Countries, in, International Energy Agency, Paris, France, 2007.

[3] CEG, China Energy Databook v7.0, China Energy Group, 2008.

[4] World Bank, Energy Use, (2016).

[5] U. Berardi, A cross-country comparison of the building energy consumptions and their trends, Resources, Conservation and Recycling, (2016) in press.

[6] EIA, International Energy Outlook, in, U.S. Energy Information Administration, 2014.

[7] Q. Wang, Effects of urbanisation on energy consumption in China, Energy Policy, 65 (2014) 332-339.

[8] X. Bai, S. Peijun, L. Yansui, Realizing China's urban dream, Nature, 509 (2014) 158-160.

[9] Q. Wang, S.D. Wu, Y.E. Zeng, B.W. Wu, Exploring the relationship between urbanization, energy consumption, and CO2 emissions in different provinces of China, Renewable and Sustainable Energy Reviews, 54 (2016) 1563-1579.

[10] UN, World Urbanization Prospects: the 2014 Revision, United Nations, 2014.

[11] NBSC, China energy statistical year book., in: Statistics Press, National Bureau of Statistics of China, Beijing, China, 2013.

[12] IEA, Energy Efficiency in the North American Existing building Stock, in, International Energy Agency, 2009, pp. 108.

[13] C. Zhang, Y. Lin, Panel estimation for urbanization, energy consumption and CO2 emissions: A regional analysis in China, Energy Policy, 49 (2012) 488-498.

[14] D.W. Jones, Urbanization and Energy Use In Economic Development, The Energy Journal, 10 (4) (1989) 29-44.

[15] A. Murata, M. Hailin, Z. Weisheng, Electricity demand in the Chinese urban household-sector, Applied Energy, 85 (12) (2008) 1113-1125.

[16] J. Virote, R. Neves-Silva, Stochastic models for building energy prediction based on occupant behavior assessment, Energy and Buildings, 53 (2012) 183-193.

[17] O.T. Masoso, L.J. Grobler, The dark side of occupants' behaviour on building energy use, Energy and Buildings, 42 (2) (2010) 173-177.

[18] H.B. Wu, Y.X. Zhu, P. Zhou, The investigation on electricity use amount of domestic air conditioners in urban residences in Beijing, New Technology in HVAC 2, (2000) 52-56.

[19] Z. Li, Y. Jiang, Q. Wei, Survey on energy consumption of air conditioning in summer in a residential building in Beijing, Journal of heating ventilation and air conditioning, 37 (4) (2007) 46-51.

[20] Y. Jian, Y. Li, S. Wei, Y. Zhang, Z. Bai, A Case Study on Household Electricity Uses and Their Variations Due to Occupant Behavior in Chinese Apartments in Beijing, Journal of Asian Architecture and Building Engineering, 14 (3) (2015) 679-686.

[21] S. Guo, D. Yan, C. Peng, Y. Cui, X. Zhou, S. Hu, Investigation and Analyses of Residential Heating in the HSCW Climate Zone of China: Status Quo and Key Features, Building and Environment, (2015).

[22] H. Polinder, M. Schweiker, A. van der Aa, K. Schakib-Ekbatan, V. Fabi, R. Andersen, N. Morishita, C. Wang, S. Corgnati, P. Heiselberg, D. Yan, B. Olesen, T. Bednar, A. Wagner, Occupant behavior and modeling, in: H. Polinder, M. Schweiker, A. van der Aa, K. Schakib-Ekbatan (Eds.) Total energy use in buildings - Analysis and evaluation methods, International Energy Agency (IEA) - Energy in buildings and Communities (EBC), 2013.

[23] H.X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews, 16 (6) (2012) 3586-3592.

[24] M. Schweiker, M. Shukuya, A Critical Review on Thermal Factors as Predictors for Occupant Behaviour: Towards a Purpose-Rank Based Model of reference levels, in: Windsor Conference 2010, Network for Comfort and Energy Use in Buildings (NCEUB), Cumberland Lodge, Windsor, UK, 2010.

[25] Y. Yamaguchi, Y. Shimoda, Behavior Model of Occupants in Home based on Japanese National Time Use Survey in: ASim 2014 - IBPSA Asia Conference, Nagoya, Japan, 2014, pp. 617-624.

[26] MOC, Code for thermal design of civil buildings, in, Beijing, China, 1993.

[27] P. Xu, Y. Shen, L. Chen, J. Mao, E. Chang, Y. Ji, Assessment of energy-saving technologies retrofitted to existing public buildings in China, Energy Efficiency, (2015) 1-28.

[28] W.P. Köppen, R. Geiger, Handbuch der klimatologie, Gebrüder Borntraeger, Berlin, 1930.

[29] US-DoE, Weather Data, in, US Department of Energy, 2016.

[30] W. Ma, L. Zhou, H. Zhang, Y. Zhang, X. Dai, Air temperature field distribution estimations over a Chinese mega-city using MODIS land surface temperature data: the case of Shanghai, Front. Earth Sci., (2015) 1-11.

[31] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, J. Glazer, EnergyPlus: Creating a new-generation building energy simulation program, Energy and Buildings, 33 (4) (2001) 319-331.

[32] US-DoE, Testing and Validation, in: U.S.D.o. Energy (Ed.) EnergyPlus Energy Simulation Software, 2012.

[33] ANSI/ASHRAE 140, Standard Method of Test for the Evaluation of Building Energy Analysis Computer Programs, in, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta (GA), USA, 2011, pp. 272.

[34] G. Beccali, M. Cellura, M. Lo Brano, A. Orioli, Is the transfer function method reliable in a EN building context? A theoretical analysis and a case study in the south of Italy, Applied Thermal Engineering, 25 (2005) 341-357.

[35] US-DoE, InputOutput Reference: The Encyclopedic Reference to EnergyPlus Input and Output, U.S. Department of Energy, 2010.

[36] CEN, Energy performance of buildings - Overall energy-use and definition of energy ratings, in, European Committee for Standardization, Brussels, Belgium, 2008.

[37] P.O. Fanger, Thermal comfort: Analysis and applications in environmental engineering, Danish Technical Press, 1970.

[38] S. Carlucci, Thermal Comfort Assessment of Buildings, Springer, London, 2013.

[39] S. Carlucci, L. Pagliano, W. O'Brien, K. Kapsis, Comfort considerations in Net ZEBs: theory and design, in: A. Athienitis, W. O'Brien (Eds.) Modeling, Design, and Optimization of Net-Zero Energy Buildings, Wilhelm Ernst & Sohn, Hoboken, New Jersey, USA, 2015, pp. 75-106.

[40] J.L.M. Hensen, R. Lamberts, Building performance simulation for design and operation, Spon Press, Oxon, UK, 2011.

[41] AIAA, Guide for the verification and validation of computational fluid dynamics simulations, in, American Institute of Aeronautics and Astronautics, Reston, VA, USA, 1998.

[42] T.G. Trucano, L.P. Swiler, T. Igusa, W.L. Oberkampf, M. Pilch, Calibration, validation, and sensitivity analysis: What's what, Reliability Engineering and System Safety, 91 (2006) 1331-1357.

[43] W.L. Oberkampf, S.M. DeLand, B.M. Rutherford, K.V. Diegert, K.F. Alvine, Error and uncertainty in modeling and simulation, Reliability Engineering and System Safety, 75 (2002) 333-357.

[44] ASHRAE, Measurement of Energy and Demand Savings, in, American Society of Heating, Refrigerating, and Air-Conditioning Engineers Atlanta (GA), USA, 2002.

[45] S. Attia, M. Hamdy, W. O'Brien, S. Carlucci, Computational optimisation for zero energy buildings design: Interviews results with twenty eight international experts, in: 13th IBPSA Conference, BS 2013, International Building Performance Simulation Association, Chambery, France, 2013, pp. 3698-3705.

[46] S. Attia, M. Hamdy, W. O'Brien, S. Carlucci, Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design, Energy and Buildings, 60 (2013) 110-124.

[47] S. Carlucci, G. Cattarin, F. Causone, L. Pagliano, Multi-objective optimization of a nearly zero-energy building based on thermal and visual discomfort minimization using a non-dominated sorting genetic algorithm (NSGA-II), Energy and Buildings, 104 (2015) 378-394.

[48] S. Carlucci, L. Pagliano, An optimization procedure based on thermal discomfort minimization to support the design of comfortable net zero energy buildings, in: 13th IBPSA Conference, BS 2013, International Building Performance Simulation Association, Chambery, France, 2013, pp. 3690-3697.

[49] S. Carlucci, L. Pagliano, P. Zangheri, Optimization by discomfort minimization for designing a comfortable net zero energy building in the mediterranean climate, in: Z. Chen, L. Guo, J. Wu (Eds.) Advanced Materials Research, Trans Tech Publications, Wuhan, China, 2013, pp. 44-48.

[50] F. Causone, S. Carlucci, L. Pagliano, M. Pietrobon, A zero energy concept building for the Mediterranean climate, Energy Procedia, 62 (2014) 280-288.

[51] M. Wetter, GenOpt - A Generic Optimization Program, in: Seventh International IBPSA Conference, Rio de Janeiro, 2001, pp. 601 - 608.

[52] L.L. Faulkner, Design and Optimization of Thermal Systems, 2nd ed., CRC Press, Boca Ranton, FL, USA, 2008.

[53] S. Attia, M. Hamdy, S. Carlucci, L. Pagliano, S. Bucking, A. Hasan, Building performance optimization of net zero-energy buildings, in: A. Athienitis, W. O'Brien (Eds.) Modeling, Design, and Optimization of Net-Zero Energy Buildings, Wilhelm Ernst & Sohn, Hoboken, New Jersey, USA, 2015, pp. 175-206.

[54] R.T. Marler, J.S. Arora, Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization, 26 (6) (2004) 369-395.

[55] J.C. Bansal, P.K. Singh, M. Saraswat, A. Verma, S.S. Jadon, A. Abraham, Inertia Weight Strategies in Particle Swarm Optimization, in: 2011 Third World Congress on Nature and Biologically Inspired Computing (NaBIC), IEEE, Salamanca, Spain, 2011, pp. 633-640.

[56] J. Kennedy, R.C. Eberhart, Particle swarm optimization, in: IEEE international conference on neural networks, Perth, Australia, 1995, pp. 1942-1948.

[57] Y. Shi, R. Eberhart, A modified particle swarm optimizer, in: Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on, IEEE, 1998, pp. 69-73.

[58] F. van den Bergh, A.P. Engelbrecht, Effects of swarm size on cooperative particle swarm optimisers, in: Genetic and Evoluationary Computation Conference 2001, San Francisco, USA, 2001.

[59] NBSC, Sixth National Census, in, National Bureau of Statistics of China.

[60] Statistics Bureau, Survey on Time Use and Leisure Activities in 2006, in, Ministry of Internal Affairs and Communications, 2014.

[61] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews, 13 (8) (2009) 1819-1835.

[62] S.S. Shapiro, M.B. Wilk, An Analysis of Variance Test for Normality (Complete Samples), Biometrika, 52 (3) (1965) 591-611.

[63] J.W. Cohen, Statistical power analisys for the behavioral science, 2nd edn ed., Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1988.

[64] J. Ouyang, J. Ge, K. Hokao, Economic analysis of energy-saving renovation measures for urban existing residential buildings in China based on thermal simulation and site investigation, Energy Policy, 37 (1) (2009) 140-149.

List of figure captions

Figure 13: Trends of (a) global solar radiation on a horizontal plane, and direct and diffuse solar radiation, (b) outdoor temperature and its daily excursion and (c) relative humidity in Shanghai.

Figure 14: Layout of the typical floor unit of the Zhoukanghang (周康航) resettlement residential project.

Figure 15: Graphical representation of the Zhoukanghang (周康航) resettlement residential project and of the simplified geometry of the numerical model.

Figure 16: Flowchart of the optimization process.

Figure 17: Available options for the design variables of the building envelope.

Figure 18: Quality check of the two numerical models of the Base case.

Figure 19: Minimization of the objective function throughout the optimization run.

Figure 20: Scatterplot of the energy needs for heating and cooling of the building variants simulated in the optimization.

Figure 21: Frequency distributions of the energy quantities for each design quality category.

Figure 22: Scatterplots of the energy needs for heating and cooling in the Base case and the Optimized case. The linear regression lines for the two set of data are shown in red.

Figure 23: Energy need for heating depicted with respect to spatial and temporal randomness for the Base and Optimized cases.

Figure 24: Energy need for cooling depicted with respect to spatial and temporal randomness for the Base and Optimized cases.