2014

Development Of Sets Of Simplified Building Models For Building Simulation

Giovanni Pernigotto
University of Padova, Italy, pernigotto@gest.unipd.it

Alessandro Prada
Free University of Bozen-Bolzano, Italy, alessandro.prada@unibz.it

Andrea Gasparella
Free University of Bozen-Bolzano, Italy, andrea.gasparella@unibz.it

Jan L. M. Hensen
Eindhoven University of Technology, The, j.hensen@tue.nl

Pernigotto, Giovanni; Prada, Alessandro; Gasparella, Andrea; and Hensen, Jan L. M., "Development Of Sets Of Simplified Building Models For Building Simulation" (2014). *International High Performance Buildings Conference.* Paper 128.
http://docs.lib.purdue.edu/ihpbc/128

Development of sets of simplified building models for building simulation

Giovanni PERNIGOTTO[1,*], Alessandro PRADA[2], Andrea GASPARELLA[3] and Jan L. M. HENSEN[4]

[1]University of Padova, Department of Management Engineering,
Vicenza, Italy
Tel: +39 0444 998746, Fax: +39 0444 998889, email: pernigotto@gest.unipd.it

[2]Free University of Bozen-Bolzano, Faculty of Science and Technology,
Bolzano, Italy
Tel: +39 0471 017204, Fax: +39 0471 017009, email: alessandro.prada@unibz.it

[3]Free University of Bozen-Bolzano, Faculty of Science and Technology,
Bolzano, Italy
Tel: +39 0471 017200, Fax: +39 0471 017009, email: andrea.gasparella@unibz.it

[4]Eindhoven University of Technology, Department of the Built Environment,
Eindhoven, The Netherlands
Tel: +31 (0)40 2472988, Fax: +31 (0) 40 2438595, email: j.hensen@tue.nl

* Corresponding Author

ABSTRACT

This work proposes a method to manage the complexity of variables involved in building simulation studies and to identify groups of simplified building models suitable to have statistically significant results. The method is described by means of an applicative example, whose aim is the definition of a set of configurations appropriate for the analysis of TRNSYS and EnergyPlus discrepancies in monthly energy needs, hourly peak loads and time of occurrences of hourly peak loads – for both heating and cooling. The proposed procedure for the definition of a reference set of building configurations moves on from the selection of a set of candidate variables describing the building envelope characteristics, paying attention to implications of each choice and to cross-correlations among variables. This is obtained by means of a screening analysis with a simple statistical index (Spearman's rank correlation coefficient). Two sample sizes are considered in order to evaluate the effects on the selection procedure. For each of the six considered outputs, the most significant group of configuration variables is identified, and the differences among those groups are described.

1. INTRODUCTION

The investigation of a physical phenomenon, such as a particular aspect in the performance of a building system or the contribution of a particular building component, requires some level of simplification with respect to the complexity of the reality, in order to limit the number or the detail of the evaluations that are needed to characterize the observed behavior. Many of the techniques that can be adopted to this purpose relate to the Design of Experiments, as is for instance with the use of factorial experiments. In these approaches, the complexity of reality can be simplified by an appropriate choice of a representative sample of configurations, which should preserve the original variance, and therefore the generality of the outcomes, and allow the study of its relation with the most relevant features of the investigated system.

In building simulation, either when using simulation to analyze the building performance (such as the optimization of the combination of refurbishment interventions, the comparison of the performance of alternative technologies, the sensitivity analysis, etc.), or when assessing building modelling itself, methods and assumptions (such as the weather inputs reliability, the accuracy of solar radiation and sky luminance models, the validation of simulation codes, etc.), a proper sample of building configurations is generally considered.

When the interest is in existing buildings, the sample has to be representative of the building stock (Ballarini *et al.*, 2014) – especially when reference buildings are not available or not completely suitable for inferential statistics. For new building solutions or technologies, instead, proper samples are needed for their evaluation. Moreover, the researchers' subjects can vary from HVAC load prediction to indoor environmental quality, as explained by Hensen and Lamberts (2011). In the second case, as well, proper building configurations have to be determined and those, for instance, for the assessment of energy balance models can be different from those for the evaluation of thermal-hygrometric or visual comfort models.

In a large number of studies in the literature, the sample of building configurations is more or less implicitly defined by considering sets of building characteristics and performing some analyses. The authors usually identify these sets by exploiting some experience or empirical knowledge about building physics aspects and building stock characteristics. Quite often, the set is assumed as an input in the research instead of the result of a pre-process within the research itself. In contrast, a sample of buildings should always be defined in consideration of the specific target of the analysis, while the use of the same set for targets different from the original one should be always avoided since it can produce relevant statistical errors, undermining the value of the findings and leading to misinterpretation of the results.

In this paper, following the Design of Experiment (DoE) approach, we define a general procedure – applicable both to building performance and to simulation modeling studies, to go beyond the trade-off between completeness and computational costs and be able to identify those building features that have the largest relevance. This methodology can be useful to researchers studying assumptions and model algorithms and to researchers developing new products or control strategies, who need sets of buildings able to emphasize a particular numerical or physical phenomenon in order to check the effectiveness of their solutions. To exemplify the approach, the proposed procedure has been used to determine which set of configuration variables should be considered when investigating TRNSYS and EnergyPlus discrepancies in monthly energy needs, hourly peak loads and time of occurrences of hourly peak loads – for both heating and cooling.

2. DESCRIPTION OF THE METHODOLOGY AND APPLICATION TO THE INTER-MODEL COMPARISON

The procedure presented in this work consists of three main parts: (a) definition of the targets of the analysis, (b) choice of the candidate variables and (c) screening analysis for the identification of the most relevant quantities for the analysis targets.

2.1 Definition of the targets of the analysis

The first step is the definition of the research target for which the sample of buildings is developed. Multi-objective sets (e.g., for both energy and lighting aims) can be developed but the specific targets need to be identified before the definition of the set.

This work illustrates the proposed procedure by using our previous research on the comparison between building energy simulation (*BES*) codes (Pernigotto and Gasparella, 2013) as an example. This example has been chosen since the research of very simple building configurations as reference for *BES* validation is a well-known theme and under development since the 80s in order to include the more and more advanced features of *BES* codes and new potential application fields of *BES*. As concerns our target definition, it discusses up to six distinct objectives, which are deviations between TRNSYS and EnergyPlus in the estimation of (1) heating needs, (2) cooling needs, (3) heating peak loads, (4) cooling peak loads, (5) time of occurrence of heating peak loads and (6) time of occurrence of cooling peak loads for residential buildings in Mediterranean regions. These are considered in the present research to evaluate to which extent a set of reference buildings developed for a given target can be suitable for another one and the possibility of combining the different characteristics in order to build a multi-objective set.

2.1.1 Analyzed outputs: The targets can be expressed by means of specific *BES* outputs, which are, in this case, deviations of energy needs, hourly peak loads and time of peak loads. As remarked also in (Pernigotto and Gasparella, 2013), these quantities are the same considered in the state of the art of *BES* validation (Judkoff and Neymark, 1995). Since the analysis involves two *BES* codes, one is selected as reference (TRNSYS) and the deviations of the other (EnergyPlus) are considered. Each quantity depends on the time discretization and the relevance of the differences, as well: hourly peak loads, for instance, can be defined daily, weekly, monthly or yearly. The time discretization results from a trade-off between the level of detail and the analysis cost: with shorter time of discretization, a larger and more time-consuming number of comparisons for each building configuration has to be considered. In (Judkoff and Neymark, 1995) the outputs time discretization is a year while in a previous research (Pernigotto and Gasparella, 2013) a monthly reference period was considered. Shorter discretization periods are not suitable for screening studies and should be considered only for more detailed analyses.

2.2 Variable choice

Characteristics and features of a building can be described in different ways and, furthermore, they are managed differently by each *BES* code. As a rule of thumb, an effective approach is to start with primitive quantities (e.g., wall sizes, layer thicknesses, etc.), which can be used directly as inputs in most of *BES* codes. For example, the thermal transmittance of a wall is not an input, since the transmitted heat is calculated by means of numerical methods – such as *CTF* (conduction transfer functions) or *FDM* (finite differences methods), which use thermal conductivity, heat capacity, density and thickness of each layer as inputs. Some quantities have identical or similar effects, as for instance, the specific heat or the density of a wall layer. Changing all possible inputs related to the targets stated before is highly inconvenient since it requires a very high number of simulations in the screening analysis. Therefore, the researcher has to (a) identify categories of quantities, (b) define parameters and variable quantities and (c) select ranges for the variables. Finally, the quantities (both parameters and variables) can be correlated and mutually conditioned (e.g., the size of windows and that of the corresponding façade). The user has to decide which quantity constrains and which is constrained in consideration of the target of the analysis.

2.2.1 Identification of the categories: In this example, only a set of variables and parameters can be selected for all targets since all the objectives (the various simulation outputs) require the same quantities as inputs. In case of miscellaneous - distinct or combined, research aims (e.g., energy and comfort), different sets have to be considered. Since the targets are related to the energy need, the description can focus on the following categories: (1) the building envelope geometry, (2) its thermal properties, (3) other heat balance contributions and (4) internal and external boundary conditions.

2.2.2 Definition of the parameters and variable quantities and selection of variables ranges: As concerns the building shape and size, we choose a parallelepiped shape of the thermal zone defined by means of the length (x), the depth (y) and the height (z). Together with the variable z, x determines the size of south/north façades while y that of east/west façades. We decide to fix one size (z) and to vary the remaining two: this implies that we are focusing on a specific kind of thermal zones – a room, an apartment with a single floor or a floor of a high-rise building, depending on x and y. We assume that x and y values can vary from 4 m to 12 m and, thus, the net floor area passes from 16 m^2 (a room) to 144 m^2 (an entire apartment) and the net volume from 48 m^3 to 432 m^3, with the largest zones approximately ten times the smallest ones. The chosen sizes encompass the majority of Italian dwellings: only 10 % of Italian residential buildings have a floor area larger than 150 m^2 according to ISTAT census data (ISTAT, 2011). Moreover, a portion of the envelope can be modelled as adiabatic (i.e., the boundary condition on the external side is the same found for the internal side), in order to simulate the adjacency to thermal zones with the same air conditions. The absolute amount of adiabatic envelope is clearly conditioned by the size of the walls so it has been decided to vary the adiabatic fraction of each wall, independently, by means of six variables ratios: Ad_1, Ad_2, Ad_3 and Ad_4 are the adiabatic fraction of the vertical façades (respectively, south, east, north and west façade), Ad_5 of the floor and Ad_6 of the ceiling. Each variable Ad varies between 0 (i.e., the wall is totally exposed to the external environment) and 1 (i.e., totally adiabatic wall). As regard the windows, their dimensions have to be compatible with the façades size and with the non-adiabatic fractions. Also in this case, the window size is expressed by means of a percentage of the available surface – independently for each vertical surface. Win_1, Win_2, Win_3 and Win_4 controls the window fraction of south, east, north and west dispersing walls, respectively. The largest window size is possible for the maximum façade size (i.e., 36 m^2) and Ad null.

The composition of the opaque envelope can be described by means of different variables: the number of layers n_l, their position p and their thermal properties. In order to simplify the problem, the same envelope definition is adopted for all the dispersing opaque elements (vertical walls, floors and ceilings). In a generic wall structure, we can distinguish those layers characterized by a high thermal resistance (i.e., insulating layers) and those with a relevant thermal mass (i.e., massive layers). In this example, we impose $n_l = 2$, with just one insulating and one massive layer for each structure – respectively a polystyrene and a concrete layer. Moreover, the insulating layer can be positioned on the external or on the internal side. For each layer, we keep constant thermal conductivity, specific heat capacity and density and consider the thickness s as variable (Table 1). All the adiabatic walls are non-insulated.

Table 1: Material layer properties

Layer	Thermal conductivity (W m^{-1} K^{-1})	Specific heat capacity (J kg^{-1} K^{-1})	Density (kg m^{-3})	Thickness (m)
Polystyrene	0.04	1470	40	s_{ins}: 0.00 – 0.20
Concrete	0.37	840	1190	s_{conc}: 0.05 - 0.30

In case of dissipating uninsulated wall (i.e., polystyrene thickness is zero), the worst performance is given by the concrete wall with the minimum thickness. In this case, the reference thermal transmittance for a vertical wall according to EN ISO 6946:2007 (CEN, 2007a) is 3.28 W m^{-2} K^{-1} and the periodic thermal transmittance and the time-shift according to EN ISO 13786:2007 (CEN, 2007b) are 3.23 W m^{-2} K^{-1} and 0.9 h, respectively. The wall configuration with the maximum thickness for the concrete and insulation layer has a thermal transmittance of 0.17 W m^{-2} K^{-1}, a periodic thermal transmittance slightly larger than 0.01 W m^{-2} K^{-1} and a time-shift around 15 h, both if the polystyrene is on the external and on the internal side. Similar ranges of values could have been obtained also assuming thermal properties as variables but with more difficulty due to the correlation among the three thermo-physical properties.

The surface thermal properties (i.e., solar absorptance α and infrared emissivity ε) are not accounted among the variables. The physical effects of the solar absorptance and the emissivity are, respectively, on solar radiation and radiative exchanges in the surface heat balance. Since these thermal flows are also largely affected by other elements (e.g., weather conditions) and the emissivity cannot be controlled in TRNSYS 16.1, solar absorptance and emissivity are kept constant. All the wall surfaces have emissivity equal to 0.9 for the external side and 1 for the internal one. Moreover, they are considered light colored (α = 0.3), with the exception of the internal side of the floor and the external side of the ceiling, which are assumed as dark-colored (α = 0.6).

The approach for the definition of the window is similar to that adopted for the opaque elements. To simplify the problem, window frames are neglected and the focus is only on the glazed part. This means that the fraction of window area that is opaque, fr_i, expressed as a percentage of each Win_i, is always 0. In the transparent part, we have two kinds of layers: the glass and the gaps. Among all possible types, we select two kinds of glass - (1) clear glass of 4 mm and (2) clear glass of 4 mm with low-e treatment, and two kinds of gaps – (A) gap with 12.7 mm thick and argon filling and (B) gap with 12.7 mm thick and krypton filling. In this context, we do not control the layer thicknesses but the number of gaps (variable n_{gap}, from 0 to 2), the kind of gas-filling (variable *gas*) and build all allowed combinations (Table 2). As a rule, the glass (2) is oriented in order to have the low-e side towards a gap and, when n_{gap} = 2, the glass between the gaps has no low-e treatment.

Table 2: Glazings

Glz_id	n_{gap}	gas	**Composition** (outside to inside)	U_{gl} (W m^{-2} K^{-1})	SHGC (-)
1	0	-	1	5.75	0.84
2	1	A	1-A-1	2.65	0.76
3	1	A	1-A-2	1.19	0.60
4	1	A	2-A-1	1.19	0.58
5	1	A	2-A-2	1.13	0.54
6	2	A	1-A-1-A-1	1.72	0.68
7	2	A	1-A-1-A-2	0.96	0.54
8	2	A	2-A-1-A-1	0.95	0.53
9	2	A	2-A-1-A-2	0.66	0.49
10	1	B	1-B-1	2.47	0.76
11	1	B	1-B-2	0.77	0.60
12	1	B	2-B-1	0.77	0.58
13	1	B	2-B-2	0.70	0.54
14	2	B	1-B-1-B-1	1.57	0.68
15	2	B	1-B-1-B-2	0.65	0.54
16	2	B	2-B-1-B-1	0.65	0.54
17	2	B	2-B-1-B-2	0.41	0.49

The thermal transmittances of the modelled glazing systems range from 0.41 W m^{-2} K^{-1} to 5.75 W m^{-2} K^{-1} while their solar heat gain coefficient (*SHGC*) from 0.49 to 0.84. Differently from the opaque components, the number of alternatives is discrete and finite and, thus, in the screening analysis, the variable n_{gap} and *gas* are combined into a single label indicator *Glz_id* to simplify the analysis. The presence of thermal bridges, shading systems, are not considered in this example.

As regards the other heat balance contributions and boundary conditions, even if many of these (e.g., ventilation rate, internal gains, setpoints) can be described by means of profiles rather than single values, time-constant values have been preferred for the screening analysis. Further and more detailed evaluations can be left for quantities that have appeared to be significant in this phase.

As concerns ventilation, accordingly to the analysis targets on energy needs and peak loads, only natural ventilation or infiltration is modelled. We decide to express the ventilation rate in terms of air change per hour (variable *ACH*), from a minimum of 0 ACH (i.e., absence of ventilation) to a maximum of 1 ACH. The absolute volumetric air flow rate depends on the actual size of the thermal zone. The internal gains are generally expressed per unit of surface (variable *G*) and, in this work, are comprised between 0 and 8 W m^{-2} (two times the average value of 4 W m^{-2} suggested for residential dwellings by Italian technical specification UNI/TS 11300-1:2008 (UNI, 2008)). In the current example, we choose to have a radiative quote G_{rad} constant to 0.5, as suggested by the European technical standard EN ISO 13790:2008 (CEN, 2008), leaving for further analysis the investigation with different values.

Both heating and cooling controls are assumed always active (i.e., dead-band regulation) with a fix heating setpoint (set_H) of 20 °C and different possible cooling setpoints (set_C) from a minimum of 20 °C (i.e., null regulation band) to a maximum of 30 °C. As concern the timestep *t*, a unique value complying with the recommendations of both *BES* codes (i.e., 10 minutes) is assumed. Sensitivity of the deviations to the timestep choice can be very time-consuming and unsuitable for a screening study and it can be left for those configurations characterized by larger deviations in a more detailed study.

The surface convective heat transfer coefficients (h_c) could be provided by the user or calculated automatically by the *BES* code according to the chosen implemented model. For the current work, the same values in TRNSYS 16.1 and EnergyPlus should be used. Since a detailed algorithm is available in TRNSYS only for the internal side and it is different from those of EnergyPlus, the use of constant values is preferred. According to the EN ISO 6946:2007 (CEN, 2007a), 20 W m^{-2} K^{-1} is assumed at the external side, while 0.7, 2.5 and 5 W m^{-2} K^{-1}, respectively for downward, horizontal and upward heat flows (i.e., for floors, vertical walls and ceilings) at the internal side.

Generally, internal walls and furniture are not modelled in detail in *BES*. Their contribution in terms of thermal inertia can be accounted as internal mass or modifying the thermal zone air capacitance. We decided to change parametrically the air zone thermal capacitance by means of a multiplier coefficient (*m*) both in TRNSYS and in EnergyPlus. This has been varied from 1 to 10, allowing the study of the influence of different internal capacitance magnitude on *BES* output deviation.

Finally, the external weather conditions and their variability depend on the choice (and number) of localities (indicated by the configuration variable *Loc*), which depend on the focus of the analysis. Since the interest is both on heating and cooling, the climates have to be selected consequently: in order to have cooling needs also for set_C higher than the classic 26 °C, the choice has fallen on a location with long heating season and short cooling season (Milan), and on one with short heating season and long cooling season (Messina). As some input quantities are not directly available from historical recordings, the same models are used to provide aligned boundary conditions to the simulation codes (Pernigotto and Gasparella, 2013).

The discussed configuration quantities are listed in Table 3. In Figure 1, correlations and interdependencies between the variables are described.

Table 3: Quantities for the screening analysis

Category	Parameters	Variables
Building envelope geometry	z	x, y, Ad_i, Win_i
Building envelope thermal properties	n_l, α, ε, fr_i	p, s_{ins}, s_{conc}, n_{gap}, gas
Other heat balance contributions	G_{rad}	ACH, G
Boundary conditions	set_H, t, h_c	set_C, m, Loc

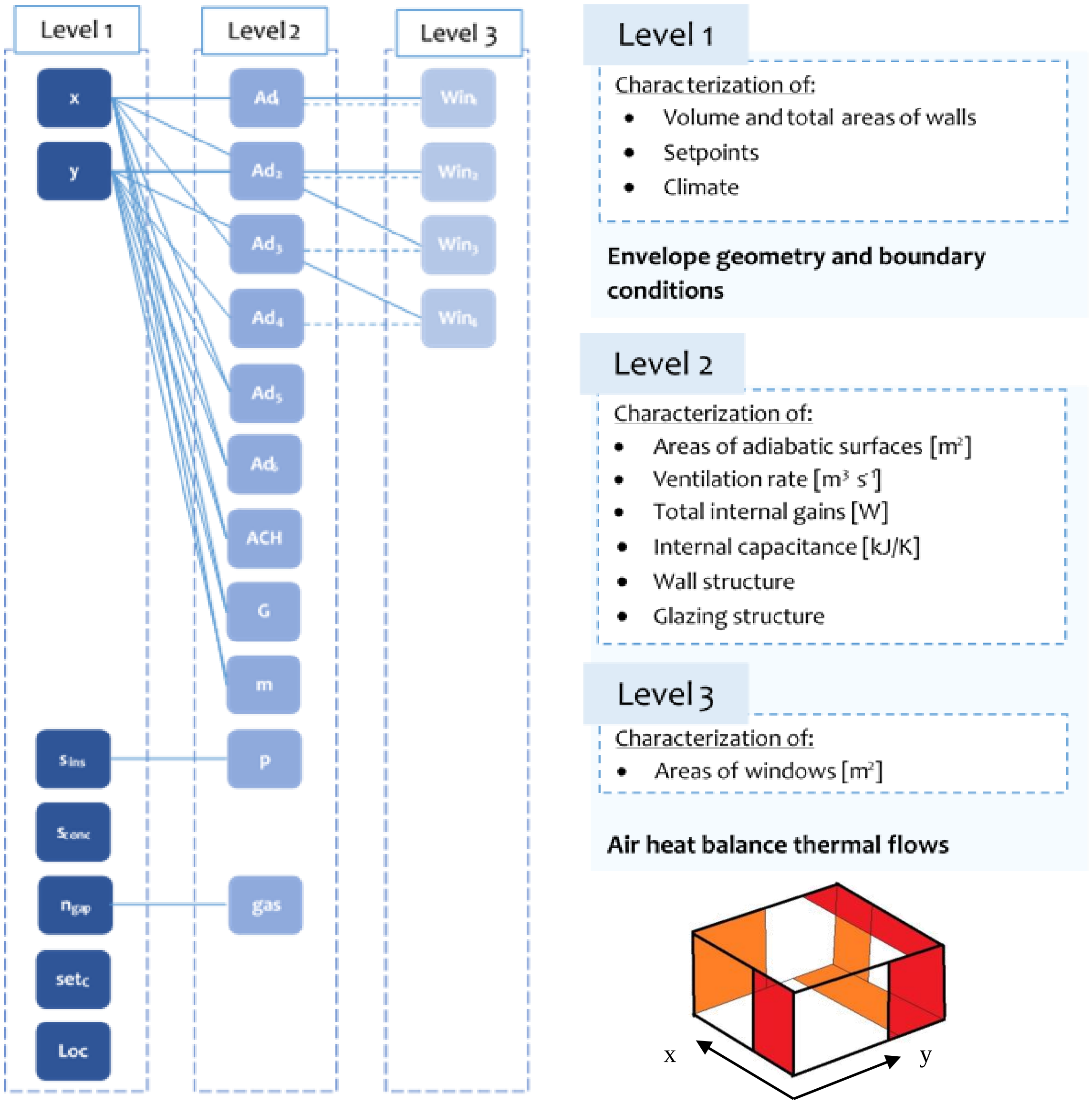

Figure 1: Correlations among the variables and simplified sketch of the building module with the adiabatic fraction of the walls in red

According to the method used to identify the quantities for the screening analysis, a particular structure of interconnections among the variables and a workflow for the definition of the inputs for each case have been developed. The first level represents all those quantities that can be used directly as inputs in *BES* (i.e., the thermal zone volume, façade areas and some boundary conditions) while, in order to allow a complete characterization of each air-heat balance component, further levels are required.

2.3 Screening analysis

Once variables and ranges are defined, it is necessary to adopt a criterion to extract the sample of configurations for the simulations in the screening analysis. Many sampling techniques and, consequently, many approaches to the sample size problem can be found in the statistical scientific literature (Saltelli *et al.*, 2008). The general objective is to balance the complexity of the analysis (i.e., number of quantities), the robustness of the sampling technique and the minimization of the sample size. Although the higher the number of inputs, the larger is the sample size to get a certain level of significance, some techniques are more robust than others and lead to a reduction of the minimum sample size. Some previous works by Burhenne *et al.* (2011) confirmed the superiority of the sampling according to Sobol's sequence (Saltelli *et al.*, 2008) for *BES*, with fastest convergence of mean estimates and stronger robustness. However, for a screening analysis, also simpler sampling techniques can be adequate, especially if the number of inputs to control is limited. No recommendations of general effectiveness are possible, since the sample size depends on both focus of the screening analysis and adopted technique. In the screening studies, interactions among the selected variables are generally not investigated, which allows keeping the number of configurations relatively low. As a rule, more sample sizes can be assessed in order to be sure that the convergence is reached. In this work, we have chosen a sampling technique based on Sobol's sequence and two sample sizes: (a) 512 and (b) 1024 (i.e., 2^9 and 2^{10}, respectively).

The statistical techniques adopted in the screening analysis are generally simple and aimed at detecting the presence of correlations between the considered variables and the outputs. Even if not quantified, the presence of interdependencies and correlations among the variables has to be taken into account since some quantities can be conditioned by other variables for physical or modelling reasons (e.g., percentage of transparent envelope and total envelope). If completely neglected, the results given by some statistical techniques can be seriously affected (e.g., the multi-collinearity in multiple regressions).

For the current screening analysis, the Spearman's rank correlation coefficient ρ has been adopted. This index assesses how well the relationship between two variables can be described by a monotonic function and it can be positive (i.e., $0 < \rho \leq 1$) or negative (i.e., $-1 \leq \rho < 0$). When $|\rho|$ is larger than 0.7 there is a strong correlation, between 0.3 and 0.7 a moderate correlation and for $|\rho|$ lower than 0.3 the correlation is weak or even null. The main limitation of Spearman's rank correlation coefficient is that it can be inappropriate for non-monotonic relationships, such as the parabolic or harmonic ones. At the very first step, however, the simplest relationships among the variables (e.g., linear or, generally, monotonic) are researched while more complex correlations are considered afterwards or at this stage only if known in advance.

As, at this stage, the interest is not in studying the correlation strength but only in detecting its presence, only those quantities whose relationships with the output are expressed by $|\rho| \geq 0.1$, statistically significant with respect to significance levels of 95 % and 99 % (i.e., with p-value lower than 0.05 or 0.01, respectively) have been considered. Since the Spearman's rank correlation coefficient works with two variables a time, we neutralize the interdependencies discussed before (Figure 1) and use directly the inputs for *BES* air heat balance instead of quantities in second or third levels (Table 4). Two variables (*Loc* and *Glz_id*) are categorical: to perform the correlation analysis, for each alternative a dummy variable is used (i.e., Loc_1 and Loc_2 for the variable location and from Glz_id_1 to Glz_id_{17} for the kind of glazing).

Table 4: Quantities for the statistical analysis

Description	Original quantity	Quantity for statistical analysis
Adiabatic surfaces	$Ad_1, Ad_2, Ad_3, Ad_4, Ad_5, Ad_6$	$x{\cdot}z{\cdot}Ad_1, y{\cdot}z{\cdot}Ad_2, x{\cdot}z{\cdot}Ad_3, y{\cdot}z{\cdot}Ad_4, x{\cdot}y{\cdot}Ad_5, x{\cdot}y{\cdot}Ad_6$
Glazing structure	n_{gap}, *gas*	*Glz_id*
Internal mass	m	$x{\cdot}y{\cdot}z{\cdot}m$
Total internal gains	G	$x{\cdot}y{\cdot}G$
Ventilation rate	ACH	$x{\cdot}y{\cdot}z{\cdot}ACH$
Wall structure (insulation position)	s_{ins}, p	$s_{ins.i}, s_{ins.e}$
Windows areas	$Win_1, Win_2, Win_3, Win_4$	$x{\cdot}z{\cdot}(1-Ad_1){\cdot}Win_1, y{\cdot}z{\cdot}(1-Ad_2){\cdot}Win_2, x{\cdot}z{\cdot}(1-Ad_3){\cdot}Win_3, y{\cdot}z{\cdot}(1-Ad_4){\cdot}Win_4$

3. RESULTS AND DISCUSSION

3.1 Overview

The deviations for each objective have similar ranges of those analyzed in (Pernigotto and Gasparella, 2013). This confirms that in our previous research no relevant variables have been overlooked. The deviations of monthly heating and cooling energy needs and peak loads are represented in Figure 2 for sample (b). As regards the differences of time of peak loads, they are larger than 1 h for around 15 % of heating peak occurrences for both samples, 25 % of cooling peak occurrences for the smallest sample and 17 % for the largest one.

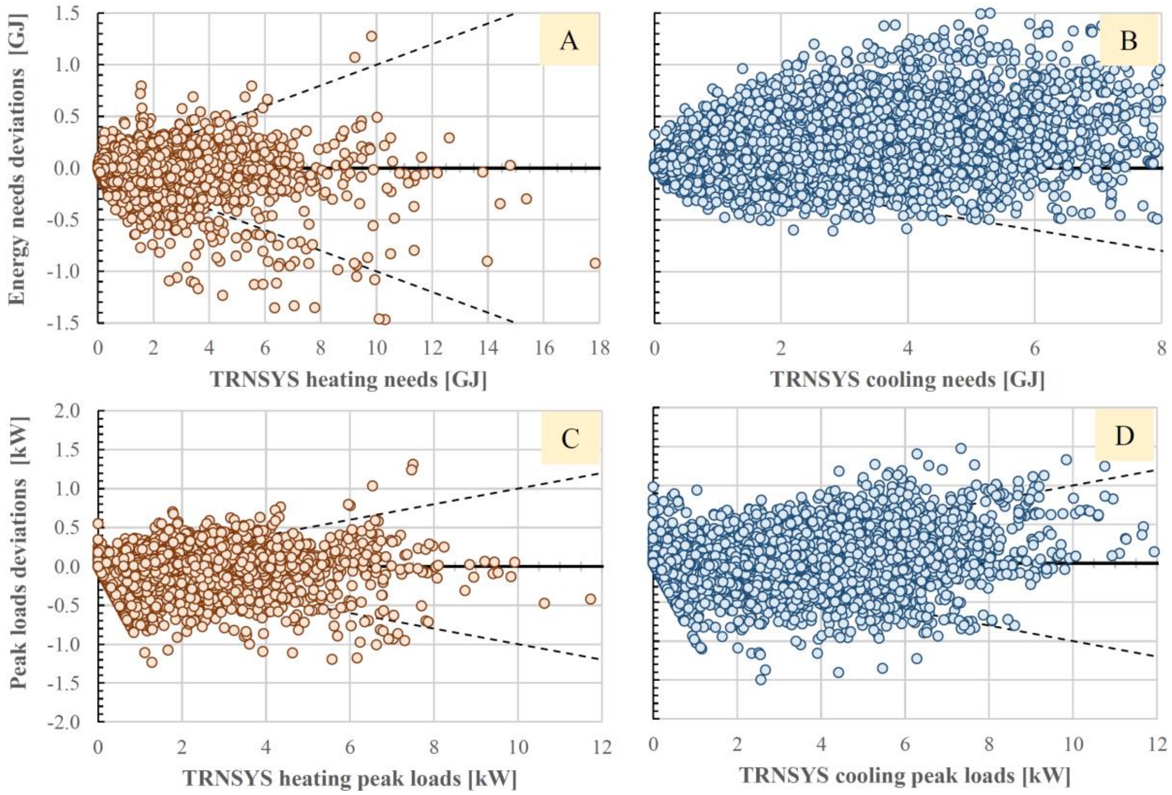

Figure 2: Deviations of EnergyPlus outputs with respect to TRNSYS in monthly (A) heating needs (B) cooling needs (C) heating peak loads (D) cooling peak loads. The dotted lines represent deviations of 10 %

3.2 Rank correlation coefficients

The statistically significant correlation coefficients for the deviations of monthly energy needs and peak loads for sample (b) cases can be found in Table 5. They resulted the same with both 1 % and 5 % statistical significance levels. The absolute correlations are generally weak (lower than 0.3). Some variables which appeared relevant with sample (a) have been rejected with sample (b), such as y, $x\cdot y\cdot Ad_6$, $y\cdot z\cdot(1\text{-}Ad_2)\cdot Win_2$ and $y\cdot z\cdot(1\text{-}Ad_4)\cdot Win_4$ for the heating needs, y, $x\cdot z\cdot(1\text{-}Ad_1)\cdot Win_1$, $x\cdot y\cdot z\cdot m$ and Glz_id_{16} for the cooling needs, $x\cdot y\cdot Ad_5$ and Glz_id_{14} for the heating peak loads, y, $x\cdot y\cdot Ad_6$, $x\cdot z\cdot(1\text{-}Ad_1)\cdot Win_1$, $x\cdot z\cdot(1\text{-}Ad_3)\cdot Win_3$, $y\cdot z\cdot(1\text{-}Ad_4)\cdot Win_4$, $x\cdot y\cdot G$, Loc_1 and Loc_2, for the cooling peak loads. Except Glz_id_{16}, which is no more statistically significant in sample (b), the rejection of these variables is due to refinement in ρ estimation in the largest sample leading to $|\rho| < 0.1$. In contrast, some quantities excluded with sample (a) are included with sample (b) (grey cells in Table 5). The adoption of a larger sample for the screening analysis allowed to reduce the number of candidate variables for further detailed analyses and to consider some quantities that would have been rejected with a smaller sample (e.g., the glazing type for cooling peak loads).

By means of this analysis, the number of variables to consider can be reduced. For the heating needs deviations, the largest $|\rho|$ are those of $x\cdot y\cdot Ad_5$, $x\cdot z\cdot(1\text{-}Ad_1)\cdot Win_1$, $s_{ins.e}$, s_{conc} and Glz_id (i.e., area of adiabatic floor, south window area, composition of the opaque walls and kind of glazing). As concerns the first two quantities, since length, depth and amount of adiabatic surface of the south façade do not result sufficiently correlated, we can keep them constant and vary just Ad_5 and Win_1. For the dummy variables for the glazing type, we can find different behaviours: some of them are negatively correlated with the deviations (e.g., single glass Glz_id_1) and some positively (e.g., double glazing Glz_id_{10}). Clustering the different kinds of glazing allows finding some relationships between their properties and the correlation with the deviations and, so, characterizing the most interesting for further analyses.

For the cooling needs deviations, the highest $|\rho|$ are encountered for $x\cdot y\cdot Ad_5$, $x\cdot y\cdot Ad_6$, $y\cdot z\cdot(1\text{-}Ad_2)\cdot Win_2$, $y\cdot z\cdot(1\text{-}Ad_4)\cdot Win_4$, $x\cdot y\cdot G$, $s_{ins.e}$, s_{conc} and Glz_id. (i.e., areas of adiabatic floor and adiabatic ceiling, east and west window area, total gains, composition of the opaque walls and kind of glazing). With respect to heating need deviations, the amount of adiabatic ceiling is added and the amount of transparent fractions on east and west façades are considered in place of that on south façade. For heating peaks loads, $x\cdot y\cdot z\cdot ACH$, s_{conc}, Glz_id and Loc are the variables most correlated while $x\cdot y\cdot Ad_5$, $y\cdot z\cdot(1\text{-}Ad_2)\cdot Win_2$, $s_{ins.i}$, $s_{ins.e}$, s_{conc} and Glz_id the most correlated with cooling peak loads.

Table 5: Spearman's coefficients for sample (a) and deviations with respect to sample (b). Coefficients in bold are relevant according to the exposed criteria; the grey cells identify the variables added in sample (b)

	Heating needs		**Cooling needs**		**Heating peak loads**		**Cooling peak loads**	
	ρ	$\Delta\rho$	ρ	$\Delta\rho$	ρ	$\Delta\rho$	P	$\Delta\rho$
x	-0.04	-	0.10	0.02	-	-	0.04	-0.03
y	-0.04	0.09	0.09	-0.09	-	-	0.04	-0.08
$x \cdot z \cdot Ad_1$	0.07	-	-	-	-	-	-0.05	0.04
$y \cdot z \cdot Ad_2$	-	-	-	-	-	-	-0.05	-
$x \cdot z \cdot Ad_3$	-	-	0.05	-	-	-	-	-
$y \cdot z \cdot Ad_4$,	-	-	-0.05	-	-	-	-0.08	-0.02
$x \cdot y \cdot Ad_5$	**-0.18**	0.06	**0.40**	-0.04	-0.09	0.08	**0.37**	-0.01
$x \cdot y \cdot Ad_6$	-0.05	0.06	**0.14**	-0.02	-	-	0.06	-0.04
$x \cdot z \cdot (1-Ad_1) \cdot Win_1$	**-0.12**	0.00	0.06	-0.13	-	-	0.06	-0.18
$y \cdot z \cdot (1-Ad_2) \cdot Win_2$	-0.07	0.05	**0.12**	-0.10	-	-	**0.13**	-0.07
$x \cdot z \cdot (1-Ad_3) \cdot Win_3$	-0.05	-	0.06	-0.01	-	-	0.07	-0.04
$y \cdot z \cdot (1-Ad_4) \cdot Win_4$	-0.07	0.04	**0.11**	-0.08	-	-	0.09	-0.09
$x \cdot y \cdot z \cdot ACH$	0.06	-	-0.04	-	**0.19**	-	-0.05	-
$x \cdot y \cdot G$	-0.07	-	**0.13**	-0.05	-	-	0.06	-0.05
$x \cdot y \cdot z \cdot m$	-	-	0.07	-0.05	-	-	-	-
$s_{ins.i}$	-	-	-0.09	-0.01	-	-	**-0.22**	-0.07
$s_{ins.e}$	**0.10**	0.01	**0.13**	0.02	0.07	-	**0.22**	0.10
s_{conc}	**0.23**	-0.03	**-0.14**	0.06	**0.14**	-0.07	**-0.15**	0.15
Glz_id_1	**-0.26**	0.01	**0.10**	-	**-0.36**	-0.07	0.09	0.15
Glz_id_2	**-0.21**	-0.04	**0.15**	-	**-0.19**	-0.04	**0.17**	0.12
Glz_id_3	-	-	0.09	0.04	-	-	**0.15**	-
Glz_id_4	-0.03	-	-	-	-	-	-	-
Glz_id_5	**0.15**	-	**-0.30**	-0.35	0.07	-	**-0.33**	-0.40
Glz_id_6	**-0.14**	-0.09	**0.11**	-	**-0.11**	-0.01	**0.13**	0.18
Glz_id_7	-	-	0.04	-	-	-	0.04	-
Glz_id_8	-	-	-	-	-	-	-	-
Glz_id_9	0.09	-	**-0.22**	-	-	-	**-0.24**	-
Glz_id_{10}	**-0.20**	-0.10	**0.14**	-	**-0.14**	-0.08	**0.15**	-
Glz_id_{11}	**0.14**	-0.07	0.10	-	**0.19**	-0.04	**0.14**	-
Glz_id_{12}	**0.19**	-0.01	-0.03	-	**0.22**	0.02	-0.03	0.02
Glz_id_{13}	**0.24**	0.09	**-0.30**	-	**0.24**	0.02	**-0.33**	-
Glz_id_{14}	**-0.11**	-0.04	**0.15**	-	-0.09	0.02	**0.19**	-
Glz_id_{15}	0.06	-	0.08	0.03	**0.10**	-	**0.11**	0.06
Glz_id_{16}	**0.16**	0.06	-	-	**0.17**	0.08	-	-
Glz_id_{17}	**0.11**	0.01	**-0.22**	-0.16	0.06	-0.02	**-0.23**	-0.17
set_C	-	-	-	-	-	-	0.07	-
Loc_1	0.07	-	-0.04	0.05	**0.27**	-0.05	-0.09	0.04
Loc_2	-0.07	-	0.04	-0.05	**-0.27**	0.05	0.09	-0.04

Some quantities are common to energy needs and peak loads deviations – both heating and cooling: the amount of adiabatic floor (for the heating peak loads deviation, $|\rho|$ is a little bit lower than 0.1), the kind of glazing and composition of the walls – in particular the massive layer.

As regards the deviations in time occurrences, in most of cases they are null. For the differences in time occurrence of heating peak loads, the only relevant quantity is the thickness of the concrete layer (ρ = 0.1), followed by the location (ρ = 0.09). There are no variables with $|\rho|$ larger than 0.1 for deviations in time occurrence of cooling peak loads and some weak connections with glazing area, location and wall composition.

4. CONCLUSION

In this work, we showed how it is possible to pre-process the variables describing building features characteristics and boundary conditions in order to reduce the number of quantities to analyze in detail, going beyond the trade-off between computational costs and analysis completeness.

As concerns the example on the comparison between EnergyPlus and TRNSYS, the number of variables has been reduced to, respectively, 6 for the study of heating needs deviations, 8 for cooling needs and peak loads deviations and 4 for heating peak loads deviations. The correlations of the processed quantities with the peak load time occurrences are very weak and the deviations are limited, making the last two objectives not worthy of further investigation. With the objectives limited to 4 and the variables to study reduced to 1/3 or 1/6, depending on the objective, it is possible to decide how to go in detail with the analysis. The alternatives are to (a) study the relevant variables, select their most interesting values (e.g., those which maximize the deviations) and define simplified buildings with full factorial approach as the shoeboxes in (Pernigotto and Gasparella, 2013) or (b) keep working on the *domain* of variables relevant for the current target with sampling techniques, without actual definitions of sets of building. The two choices have different outlooks on statistical generalizability but both can be effective for further investigations aimed at, for instance, (1) seeing the effect of increasing the time discretization (e.g, hourly or sub-hourly outputs) or (2) analyzing cross-correlations and interactions between the actually relevant variables.

Regarding the aims of the specific topic discussed with the example, i.e., the inter-code comparison and the evaluation of the relative accuracy of *BES* results, the proposed approach can be exploited in order to go in detail of *BES* validation process and set new challenging targets. The deviation of about 10 % on yearly results accepted according to the state of the art could be integrated with additional analyses on results with shorter time-discretization for the building configurations with largest deviations. Also, a more detailed comprehension of the building configurations whose performance estimation is failed by a given *BES* code can drive to its improvement and a continuous development for more accurate simulation tools.

REFERENCES

Ballarini, I., Corgnati, S. P., Corrado, V., 2014, Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project, *Energy Policy*, vol. 68, p. 273–284.

Burhenne S., Jacob D., Henze, G. P., 2011, Sampling based on Sobol' sequences for Monte Carlo techniques applied to building simulations, *Building Simulation 2011*, IBPSA, p. 1816-1823.

Ente Nazionale Italiano di Normazione (UNI), 2008. *UNI/TS 11300-1:2008 – Energy performance of buildings Part 1: Evaluation of energy need for space heating and cooling*, Milan, Italy.

European Committee for Standardization (CEN), 2007a, *EN ISO 6946:2007 – Building components and building elements - Thermal resistance and thermal transmittance - Calculation methods*, Brussels, Belgium.

European Committee for Standardization (CEN), 2007b. *EN ISO 13786:2007 – Thermal performance of building components - Dynamic thermal characteristics - Calculation methods,* Brussels, Belgium.

European Committee for Standardization (CEN), 2008. *EN ISO 13790:2008 – Energy performance of buildings - Calculation of energy use for space heating and cooling,* Brussels, Belgium.

Hensen, J. L. M., Lamberts, R., 2011, *Building performance simulation for design and operation*, Taylor and Francis, London, UK, 512 p.

Italian National Institute of Statistics (ISTAT), 2011, *15° Censimento popolazione e abitazioni 2011* (data published online at http://www.istat.it, accessed in January 2014).

Judkoff, R.D., Neymark, J., 1995, *International Energy Agency Building Energy Simulation Test (BESTEST) and Diagnostic Method. IEA SHC Task 12/IEA BCS Annex 21*, NREL, Golden, 296 p.

Pernigotto, G., Gasparella, A., 2013, Extensive comparative analysis of building energy simulation codes: Heating and cooling energy needs and peak loads calculation in TRNSYS and EnergyPlus for southern Europe climates, *HVAC&R Research*, vol. 19, no. 5: p. 481-492.

Saltelli, A., Chan, K., Scott, E. M., 2008, *Sensitivity Analysis*, Wiley, Chichester, UK, 494 p.