Consistency in early simulation inputs with statistical analyses of historical models: Reducing the impact of mental models and simulation settings

Nanna Dyrup Svane¹², Torben Østergård¹², Rasmus Lund Jensen¹
¹Aalborg University, Aalborg, Denmark
²MOE A/S, Copenhagen, Denmark

Abstract

Simulation inputs, especially in early design phases, are associated with great uncertainty and, on top of that, dependent on the modeller. Minor setting variations can significantly impact the simulation output, making it difficult to evaluate the design quality and compare different Building Performance Simulations (BPS).

In this study, we have made a statistical data analysis of design inputs and simulation settings for 1320 selected BPS files created within a single company in the period 2009-2020.

This extensive historical dataset provides insight into modellers' behaviour, common practice and typical building design variations. The former can help streamline future BPS and make them less modeller-dependent, whereas the latter provides realistic starting points in the early design phases.

Key Innovations

- Streamlining BPS-inputs
- Modeller dependency
- Statistical analyses
- Guidance and streamlining

Practical Implications

A higher degree of guidance and consistency in BPS inputs will reduce the modeller-dependent uncertainties and preparation time and thereby promote the use of BPS in early estimations.

Introduction

"Design evolution involves a chain of design decisions. Each decision is supported by input supplied by the various domain experts to the design team at large" (Wit & Augenbroe, 2001)

The building simulation discipline has gradually evolved and matured to be a powerful and mainstream resource in close to all larger building projects. However, building design is still a multivariate and multicriteria problem that continues to challenge the design team. Designing sustainable buildings that fulfill the requirements, needs and wishes of the building owner, building users and the designers involved requires a high level of awareness, understanding of building physics and time. Furthermore, the BPS discipline has evolved from being in the hands of a few experts to being an expected competence in recent building design graduates (Gentile, Kanters, & Davidsson, 2020).

With an increased level of detail and sophistication, a higher amount of user input is necessary to create a functional BPS model. These user inputs are subject to various uncertainty sources, especially in the early design phases (Eisenhower, O'Neill, Fonoberov, & Mezić, 2012). These uncertainties result from missing data on the physical properties and variables such as occupant behaviour and the process of modelling (Malkawi & Augenbroe, 2004). Despite increasing automation assistance, e.g., IFC data transfer and industry guidelines, research shows that the modelling process is still highly dependent on the modeller and their culture (Guyon, 1997) (Imam, Coley, & Walker, 2017) (Hitchcock & Wong, 2011) (Williamson, 2010).

One could argue that errors caused by modellers are due to a lack of experience and education. However, studies show that even expert users misinterpret data that induce considerable deviations in the final results. From a validation study of the DOE-2 Building Energy Simulation Program in 1998, it was concluded: "...when the input is uncontrolled, considerable scatter in monthly results can be expected among expert users of building energy analysis computer programs such as DOE-2. The most significant reduction in scatter can be obtained by having an independent observer check the input for errors and by eliminating gross ambiguities in the input." (Sullivan & Winkelmann, 1998).

In a more recent study by Imam et al. (2017), a total of 108 building modellers were asked to rank the importance of 21 common modelling input variables. The survey's overall goal was to identify the so-called 'mental models' of professionals in the construction industry and compare those with a validated thermal model. The author's analysis showed little correlation between the modellers' ranking of important parameters and those proved to be objectively important. Furthermore, the study revealed that qualification level and years of experience did not improve the prediction accuracy. Imam et al. (2017) suggest that personal performance might drift over time due to the engineering consultancy culture and lack of comparison of results and actual performance.

Østergård, Lund, & Søndergaard (2020) describes the 'mental model'-phenomena as a potential bias, where experience or preferences could hinder the modeller from finding optimal solutions. In their study, 22 graduate
students were asked to optimise a BPS solution with a maximum of 28 iterations. The student optimisations are compared with a full factorial sample where a large design space of both favourable and unfavourable designs is identified. The students manage to avoid some unfavourable input options but also tend to prefer solutions with no mechanical cooling. Contradictory, more of these solutions are in the Monte Carlo high-performing solution set. This avoidance could indicate an unintentional handover of 'mental models' from teacher to students because of a general, perhaps outdated, 'mentality' to avoid mechanical cooling in Danish buildings.

Comparably, Bannister (2015) describes a knowledge gap, a gap in understanding the correct use and interpretation of simulations. Via inspection of Australian performance-rated building projects, he has identified issues with incomplete modelling, unrealistic schedules, lack of reporting and inappropriate use of HVAC systems. This knowledge gap has led to developing a protocol to help determine input parameters and understand the meaning and reliability of the BPS results.

**Aim of current research**

Current research is part of an industrial PhD project with the overall goal to improve the quality and uniformity while reducing the production time of BPS in both early and main design phases. The quality improvement concerns five areas:

1. Semi-automatic geometrical modelling
2. Consistency in and sufficient accuracy of BPS input
3. Exploration of the entire solution space
4. Estimations in early design phases
5. Dissemination and communication of results

Present research concerns area 2, where the overall goal of the statistical data analysis is to evaluate previous and hence uniformise future BPS inputs. We want to document the existence and influence of 'mental models' in simulations from the industry with the historical dataset. With this knowledge, we wish to establish room templates comprising most represented variations for both geometry, systems and settings, along with national guidelines and building code requirements. The goal, however, is not to diminish the freedom of design but rather to frame and support the needed decision-making in the early phases with sufficient accuracy.

In future work, we plan to perform thousands of Monte Carlo Simulations with the room template variations to enable profound exploration of the design and solution space and promote simulation-based early building performance estimations.

These early estimations with reduced modeller dependency can potentially phase out the use of rules of thumb and antiquated experience values and entail quicker modelling of more detailed simulations when the project progresses.

**Method**

In this study, we have analysed 1320 BPS produced in the simulation software BSim in 2009-2020. Starting from a total of 10,882 files, we performed a manual sorting of the files in order to have one representative and quality-assured file per analysed room, leading to 1320 selected files. Through automation, we have extracted the most necessary input information. Unfortunately, some data, such as SHGC on windows and construction layers, are stored in a database file shared across projects, why it has not been possible to extract this data with sufficient trustworthiness. The files originate from five offices in Denmark, named dep.1 to 5, in which the number of employees working with BPS varies considerably, see Table 1.

**Results**

In the following section, we present the analysis based on the historical dataset. We investigate and present the historical variation for various simulation settings and design inputs. In the concluding part, the findings are summarised in Table 2, which comprise an example of a conceptual room template for a single-façade office to use for future BPS, where input uncertainty is high.

**Critical rooms**

The Danish building code imposes documentation of the thermal indoor environment based on conditions in critical rooms. These critical rooms are subjectively selected by the engineer/modeller based on a quick assessment of the window-wall ratio, orientation, shadow conditions and internal loads.

Examining the number of selected files per project, the company's modellers calculates an average of 4.4 rooms per project, varying from 1 to 182 rooms. The company has been involved in constructing two large hospitals (2013 and 2016), which requires analysis of many rooms. Excluding these projects from the dataset, we see that in most cases, the company simulates 1-6 rooms, with an average of 3.7 critical rooms per project, without notable variation over time. With the increasing complexity of buildings, we had expected the number of critical rooms to increase with time. The relatively fixed amount of analysed rooms could result from nationwide efficiency improvement where a sustainable level between sufficient accuracy and allotted time has been found.

| Table 1: Overview of files and modellers in the company's five departments. The number in parenthesis is the number of files attached to a known modeller |
|-------------------------------|---------------|---------------|---------------|---------------|-----------------|-----------------|
| **Years**                     | **Dep_1**     | **Dep_2**     | **Dep_3**     | **Dep_4**     | **Dep_5**     | **AVERAGE/SUM**  |
| 7447 (5317)                   | 2687 (880)    | 556 (4)       | 112 (4)       | 80 (56)       | 10.882 (6260)  |
| Rooms, total                  | 747 (560)     | 473 (153)     | 60 (1)        | 26 (2)        | 14 (10)        | 1320 (725)       |
| Known modellers, total        | 22            | 12            | 1             | 1             | 4              | 40              |
Variations per room

Looking at the total number of files and the selected number of rooms, the company modellers calculate, on average, 6.5 variations per room, varying from 0-60. However, this number is a rough estimation since there are different ways to make variation simulations. Some save a new variant file, whereas others make changes in the same file.

Simulation options

The BPS software BSim, used in the company, requires the modeller to set a number of simulation options (SO) before running a simulation. These options involve choosing from four solar radiation algorithms and possible opt-in on nine different calculation models, for instance, Multizone and Thermal Bridge calculation. Sometimes the modeller begins from scratch, for which the software defaults are chosen and may be adjusted. But often, the modeller copies a model/file from another building project with similar characteristics. In the latter case, simulation options will be inherited from that project and will persist unless actively changed. In the 1320 files, we found 68 different combinations of SO. Three of these combinations are highly represented and account for about 57% of the files. In principle, all these combinations could be suitable for specific cases, e.g., the Multizone Model should be turned on only when investigating the airflow between rooms; however, it is also possible to opt-out some calculation models and thereby make the simulation faster but less accurate. Therefore, it would make it easier to compare models and make models more trustworthy if common defaults are defined throughout the company or even throughout the Danish industry. When sorting the option combinations by modeller (file owner), we see that each identified modeller uses an average of 3.4 different combinations of design options, varying from 1 to 11; however, we see no particular pattern in the different departments.

To evaluate the potential effect on the final result of choosing the wrong combination of SO, we have simulated a simple one-façade-room with 63 most used SO combinations while keeping all other inputs constant. However, all combinations were simulated for all four main orientations to evaluate if some orientations are more sensitive to the SO.

The 63 combinations lead to 26 different results for each orientation when evaluating energy demand and hours with overtemperature. As illustrated in Figure 2, the cooling demand is almost constant, whereas heating and fan power fluctuate more. The relation between these two demands is relatively constant for most combinations. Still, for ten combinations, this relationship changes significantly. On average, the heating demand in these ten combinations is more than 28% higher than the average heating demand in the other 53 combinations. Looking at the number of hours with temperatures above 26°C (h>26°C), there seems to be no direct relationship with the energy demands, but on the other hand, a definite dependency on the SO. The number of hours varies from 18 to 230 hours, with the highest frequency around 136-139 hours, as illustrated in Figure 3. As earlier mentioned, three SO combinations were preferred in the analysed files. These three combinations result in the same output, equivalent to the median, and close to the mean values of all 63 combinations for both energy demand and overheating.

Furthermore, our investigations reveal that the heating and cooling demands are most sensitive to SO variations when the window orients north. In contrast, fan power demand and overtemperature hours are more sensitive to SO variations with a southward orientation.

Typically, the simulation aim is a maximum of 100 hours with temperatures above 26°C, which has not complied with the three prevalent SO but is obtained in 24% of the SO combinations. Contrary to this, 38% of the SO combinations yield a higher degree of overtemperature.

The SO analysis uncovers a non-negligible potential of making faulty conclusions on the simulation result solely due to simulation settings.
Furthermore, the consequences of selecting and deselecting calculation models are not apparent for the typical modeller. It cannot be expected that the modeller performs impact assessment of different SO combinations for each project. Therefore, we conclude that a common template and guidance for the SO could reduce a high degree of uncertainty and inaccuracy related to modeller-dependent simulations. Naturally, this guidance is dependent on and must be adjusted to preferred software.

**Timesteps**

Besides the above-described options, the modeller can also adjust the number of time steps per hour. Neither the software nor the building code sets requirements to the timestep size, besides a stable simulation recommendation. Due to increased simulation time with increasing timesteps per hour, it may be important to balance the number of time steps with the result accuracy, especially when performing thousands of Monte Carlo simulations.

When analysing the effect of time steps on three SO combinations that lead to a high, low and mean degree of h>26°C respectively, we see a risk of making significant result misinterpretations if simulating with only 2-10 timesteps per hour. As illustrated in Figure 3, the relative error if simulating with 2 timesteps per hour is up to 127% compared to simulating with 90 or more timesteps per hour.

**Figure 3:** Significance of timesteps for SO combinations with high (black), low (grey) and mean (blue) degree of h>26°C. The relative error in dotted lines.

Though the number of timesteps per hour is not equally sensitive to the three SO combinations, we see that the relative error is below 1% for all three combinations at about 60 timesteps per hour. From the statistical analysis, we found that the number of timesteps used in the files varied from 2 to 250 per hour, with about 25% of the files using less than 60 timesteps per hour.

**Figure 4:** Boxplot showing first quartile, median and third quartile of timesteps per hour over time

Building code developments

Out of the 1320 selected files, 634 (48%) were created in or after 2018, where the current building code commenced. For the first time, the code stated that the outdoor concentration of CO₂ in the air should be set to 400 ppm. Before 2018, the regulations did not impose requirements hereto, but it was typically accepted to use 350 ppm, which was and still is, the default value in the software. The increased outdoor concentration of CO₂ will increase the ventilation demand to maintain the same internal CO₂ y level.

We have compared the creation year and the user-defined outdoor CO₂ concentration. The analysis, illustrated in Figure 5, shows an improvement from 2018 towards 2020, where the use of the incorrect CO₂ concentrations is reduced to 23 % of the files. One employee, who was hired in 2018 and had no previous work experience, accounts for 44% of these files. The other cases are also primarily allocated to employees with a short background of 0-2 years, suggesting a lack of update of the induction training and the company's quality assurance program.

**Figure 5:** CO₂ concentration in simulations over time.

In fall 2013, the Danish climate file was updated to 'Denmark 2013', and from January 2015, the building code required its use in BPS. Not until three years later, in the current building code, was it specified to use the new climate file along with the calendar year 2010. Until then, it was usually accepted to use the year 2002. Choosing another year can influence the results related to thermal comfort since weekdays, and thus extreme weather conditions, fall differently from year to year.

**Figure 6:** Calendar year used in simulations

Looking at Figure 6, we see that from requiring using the updated weather file in 2015, but before the building code
stated a specific calendar year in 2018, there is significant variation in the year used. However, almost all files from 2018 and forward use the correct calendar year. We see this as an excellent example of how explicit guidance can eliminate modeller-dependent uncertainty.

**Room templates**

The above analyses have used the historical dataset to examine and exemplify modeller dependencies and their influence. The following analyses will examine typical geometrical and HVAC design in a Danish context with the same data set. These analyses lead to a conceptual room template for an office room, presented in Table 2, to use for realistic estimations of uncertain simulation inputs in early design phases and quality assurance.

All the selected models are manually labelled with a room type, leading to 33 different room categories, of which the top 10 are illustrated in Figure 7. These ten room types characterise 86 % of the selected models.

![Figure 7: Distribution of room types](image)

By labelling the files by room types, we can further investigate the categories and evaluate typical patterns, spans and layouts to form a future BPS template.

The room type 'Office' accounts for 24.6 % of the selected files why the roll-out of an 'office'-template is described in the following sections and joined in Table 2. For all input parameters known to have high uncertainty in early phases, we have used histograms to illustrate frequent values and the 10th and 90th percentiles to establish reasonable spans for the template.

**Façade design**

83% of the office rooms have 1 or 2 façades, with the majority, 63%, having a single façade, Figure 8.

![Figure 8: Number of facades in office rooms](image)

In most cases, the window orientation is south-facing (south ± 45°) but is closely coincident with an east- and west-facing orientation, as shown in Figure 9.

![Figure 9: Orientation of window in one-façade offices](image)

In the cases of only one façade, we found an average window-wall ratio of 41% and the ratio almost follows a normal distribution (Table 2). Only 4 % of the modelled offices have skylights, why these cases are not further investigated for the time being.

The U-value analysis for the façade and windows shows a preference for high-insulated facades and triple-glazed energy windows.

The room depth is an interesting measure in offices since daylight from the façade often sets the limit for workplace utilisation. The histogram shows a preference of room depths around 6 meters.

**People load**

The schedule data point to two kinds of offices; a standard daytime office and a more distinctive office, e.g., a hospital office, with both day- and night-time presence.

Yet, the daytime offices make up the majority of the files, of which we can read that it is often assumed that the office populates between 7 and 8 am and empties at 4-5 pm.

In October 2017, the Danish Building Research Institute published the *Industry Guidelines for indoor climate calculations*, which was prompted and espoused by the industry itself (Mette Havgaard Vorre et al., 2017). The Guidelines provide instructions for the definition of schedules and loads and documentation of indoor climate simulations. The guideline suggests a people load schedule for offices, starting at 7 am and ending at 5 pm with a reduced presence in the first and last hour.

**Equipment load**

We have analysed the equipment load with respect to both floor area and people load, and the results imply that the equipment load is designed as a function of people load, with a predominant value of 100 W/person.

**Systems**

The systems Heating, Mechanical Ventilation, Infiltration and Lighting were active in all the selected office files. Mechanical Cooling was active in 40% of the files, and Shading was activated for 77 % of all windows.

From the infiltration analysis, we see a high frequency of high infiltration air changes, suggesting a general problem with converting the requirement for air change test at 50 Pa pressure difference to the infiltration air change, and therefore a need for guidance.

The room templates have 'ideal' systems, allowing for the heating, cooling and air supply that the room demands. All the above analyses are summarised in Table 2, proposing room template spans for early design.
Table 2: Historical input distributions for one-façade offices. Boxes indicate spans for the room template (early design inputs) based on 10-90% quantiles, industry guidelines and code requirements.

<table>
<thead>
<tr>
<th>Description</th>
<th>Histogram</th>
<th>Room template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room depth [m]</td>
<td><img src="image" alt="Room depth histogram" /></td>
<td>4-9</td>
</tr>
<tr>
<td>WWR [%]</td>
<td><img src="image" alt="WWR histogram" /></td>
<td>0.02-0.65</td>
</tr>
<tr>
<td>U-value, facade [W/m²K]</td>
<td><img src="image" alt="U-value, facade histogram" /></td>
<td>0.11-0.30 Building code requirement: 0.30</td>
</tr>
<tr>
<td>U-value, windows [W/m²K]</td>
<td><img src="image" alt="U-value, windows histogram" /></td>
<td>0.8-1.8</td>
</tr>
<tr>
<td>People load [W/m²]</td>
<td><img src="image" alt="People load histogram" /></td>
<td>5-20 Industry guidelines: 10-17</td>
</tr>
<tr>
<td>People schedule First and last hour</td>
<td><img src="image" alt="People schedule histogram" /></td>
<td>7am-5pm 50% at 7am, 12pm and 5 pm As stated in Industry guidelines</td>
</tr>
<tr>
<td>Equipment load [W/person]</td>
<td><img src="image" alt="Equipment load histogram" /></td>
<td>50-150 Industry guidelines: min. 80 W incl. light</td>
</tr>
</tbody>
</table>
**Discussion**

The company and the industry continuously try to address time-consuming model construction, modeller dependency and overall quality and accuracy of early BPS. However, our presented analyses revealed a need for guidance and templates to ensure a more uniform BPS quality.

With the use of a sizeable statistical dataset, we have proposed room templates to tackle some of the before-mentioned problematic areas of early BPS. We have found it important to base the room templates on the statistical dataset to avoid our own mental models biasing in the template definition.

The statistical dataset has helped us prioritise both the most important room categories and the spans to be included in the room templates and future Monte Carlo Simulations. We have found it necessary to make a weighting between making the room templates encase the majority of future designs with enough accuracy to avoid significant redesign, but at the same avoid spending too much effort and simulation time on extreme or infrequent cases. However, due to constant system and design developments, we might need continuous devaluation of the spans. Indeed, some rooms will not fit into the templates and require individual modelling and system setup. But if we can outline room templates for the most prevalent and homogenous rooms, we see a solid foundation for exploring design space using structured Monte Carlo simulations and guidance.

We plan to compare the Monte Carlo-output tendencies found with room templates to various explicit simulations to further evaluate the room templates. This will help further expand our knowledge about the potential and broadness of the room templates. Moreover, we wish to compare the timeframe, results and tendencies obtained with the room templates with modelling and optimisations produced by various industry modellers. With this, we hope to get a more thorough picture of how mental models can affect the choice of design solutions.

We believe the field of 'mental models' in BPS is not adequately articulated and emphasised in the industry nor academia. Therefore, we see the current research along with further described developments as a small step towards a higher focus on and understanding of the modellers' influence on the BPS result. It is, however, unlikely to define simulation options and settings as building code requirements since companies use different software with varying settings. Instead, we see a need for dissemination and visibility of simulation settings in the documentation for the simulations performed, thus ensuring that a position on the settings is taken and that their significance for the result is stated.

The room templates are based on a large statistical dataset from a relatively long period of time, created by a large
studies of the DOE-2 building energy simulation program: Final report.


