Interactive Building Design Space Exploration Using Regionalized Sensitivity Analysis

Torben Østergård[1,2], Rasmus Lund Jensen[1], Steffen Enersen Maagaard[2]
[1]Department of Civil Engineering, Aalborg University, Aalborg, Denmark
[2]MOE, Consulting Engineers, Aarhus, Denmark
First author's e-mail address: to@civil.aau.dk

Nomenclature

D_i	KS2 maximum distance between two cumulative distributions for i^{th} parameter
$D_{i\text{-}AB}$	D_i between cumulative distribution sets S_A and S_B
D_{ij}	D_i for j^{th} repetition in the TOM method
EE	method of elementary effect (Morris' SA)
FF	factoring fixing (SA setting based on total effects)
FM	factor mapping (SA setting)
FP	factor prioritization (SA setting based on main effects)
FR	factor ranking (SA setting based on total effects)
J	number of repetitions in TOM method
KS2	Kolmogorov-Smirnov two-sample statistics
N	total number of Monte Carlo simulations
Normative model	Danish simulation software Be10 based on ISO 13790 (here combined with regression model for daylight)
Overtemperature	thermal comfort penalty output in normative model [kWh/m² floor area]
PCP	parallel coordinate plot (for real-time analysis)
PEAR	Pearson's product-moment correlation coefficient
Q	number of simulations in random selected subset
RSA	regionalized sensitivity analysis
SA	sensitivity analysis
SRC	standardized regression coefficients (linear regression)
S_A	set of all simulations
S_B	set of behavioural simulations meeting all criteria
S_N	set of non-behavioural simulations
S_i	first order effect (Sobol's variance-based SA)
S_T	total effect (Sobol's variance-based SA)
SA_{TOR}	comparable SA measure based on TOR [0; 100%]
SA_{TOM}	comparable SA measure based on TOM [0; 100%]
TOR	proposed RSA method used for real-time analysis – both inputs and outputs
TOM	proposed RSA method to rank inputs according to sensitivity towards multiple outputs

Abstract

Monte Carlo simulations combined with regionalized sensitivity analysis provide the means to explore a vast, multivariate design space in building design. Typically, sensitivity analysis shows how the variability of model output relates to the uncertainties in models inputs. This reveals which simulation inputs are most important and which have negligible influence on the model output. Popular sensitivity methods include the Morris method, variance-based methods (e.g. Sobol's), and regression methods (e.g. SRC). However, such methods only address one output at a time, which makes it difficult to prioritize and fixate inputs when considering multiple outputs. In this work, the primary outcome is a novel sensitivity method denoted TOM, which relies on Kolmogorov-Smirnov two-sample (KS2) statistics to rank inputs due to their influence on multiple outputs. A secondary method, denoted TOR, provides a real-time sensitivity measure when exploring data with the interactive parallel coordinate plot (PCP). The latter is an effective tool to explore stochastic simulations and to find high-performing building designs. The proposed methods help decision makers to focus their attention to the most important design parameters. As case study, we consider building performance simulations of a 15.000 m² educational centre with respect to energy demand, thermal comfort, and daylight.

Introduction

Sensitivity analysis (SA) plays a valuable role in the field of building performance simulations. Its extensive applications have been reviewed in-depth by Tian (2013). Other works compare sensitivity methods with respect to accuracy, applicability, convergence, and visualization in relation to building performance (Burhenne 2013, Das et al. 2014, Nguyen & Reiter 2015). Similar comparisons been conducted within other engineering disciplines (Confalonieri et al. 2010, Mara et al. 2017, Pianosi et al. 2016, Song et al. 2015, Yang 2011). A textbook on SA by Saltelli et al. (2008) state that the purpose of SA may be the following:

- Factor Prioritization (FP), which is used to rank inputs according to their individual contributions to output variance
- Factor Fixing (FF) or screening, which is used to fixate uncertain inputs which have negligible contribution to output variance – even when considering interactions with other inputs
- Factor Mapping (FM), which is used to identify input values that lead to model realizations in a specific output range

FP is a measure of the input's individual contribution to output variance, which is often referred to as main effects or first order effects. This setting is used to identify uncertain inputs which, when kept fixed, will lead to the greatest reduction in output variance. This is desirable in uncertainty analysis, if the analyst wish to reduce uncertainty of the results. In contrast, FF is based on the inputs' total effects, which is a measure of the variance induced by the input's individual contribution along with its interactions with other inputs. If the total effect is small, the input makes no significant contribution to the variance and it may be fixated. For this work, we define another setting called "Factor Ranking" (FR), which is based on the total effects. This setting is used to rank inputs according their overall influence, which help the analyst (or multi-actor design team) focus on the inputs that matter and interact the most.

Other purposes of SA include the study of input interactions (interdependencies), robustness assessment, and error detection. The intent of the analysis, along with computational effort and model complexity, is important when choosing among the many sensitivity methods. The global methods may be classified as regression-based, screening-based, variance-based, and regionalized sensitivity analysis. In the following, we discuss the deficiencies of popular methods when guiding decision makers towards building design with high overall performance.

Building simulations involve hundreds of inputs. When varying design parameters in Monte Carlo experiments, it is desirable to fixate the least significant inputs (FF) and thus simplify the analysis. For this purpose, the Morris method (EE) has been widely used because it is model independent and computationally cheap (Morris 1991). However, its one-at-a-time sampling strategy cannot be used for design space exploration, which is an important aspect of building design. Variance-based methods are also popular for SA, since they are model-independent and they can assess first order effects (for FP), higher order effects, and total effects (for FF). Higher order effects reveal input interactions. Though, variance-based methods have high computationally costs (Pianosi et al. 2016).

Common for (perhaps all) screening-based, variance-based, and metamodel sensitivity methods is that they address only one model output at a time. Hence, inputs contribute and rank differently for each output of interest. This makes it difficult to determine, which inputs should be kept fixed, and which inputs are the most important overall. In addition, their sensitivity measures represent the entire set of simulations, whereas the modeller may be interested in different parts of the simulated design space (FM). To address these issues, we propose to apply regionalized sensitivity analysis using two-sample Kolmogorov-Smirnov test statistics.

In this paper, the primary objective is to rank inputs with respect to multiple outputs (FR). This is particularly helpful in holistic building design that involves multiple performance outputs, such as energy demand, thermal comfort, and daylight. The secondary objective is to highlight, in real-time, the parameters affected the most by user-defined filters (FM). The latter builds on previous work, in which a multi-actor design team filters Monte Carlo simulations, using an interactive parallel coordinate plot (PCP), to investigate different regions of the design space (see Figure 1) (Østergård et al. 2017). The PCP is intuitive and easy to interpret, but if the analysis contains more than approximately 10 parameters, it becomes difficult to see which parameters have been affected by the applied filters.

Methodology

A precondition for our work is the Monte Carlo method. This is used to run a large number, *N*, of building simulations, which are explored using several SA methods. In the Monte Carlo workflow, the modeler first defines input distributions and sampling strategy. Next, simulations are run with respect to various outputs such as energy demand, thermal comfort, and daylight. The modeller may perform sensitivity analysis to fixate non-significant inputs (FF). In that case, the Monte Carlo

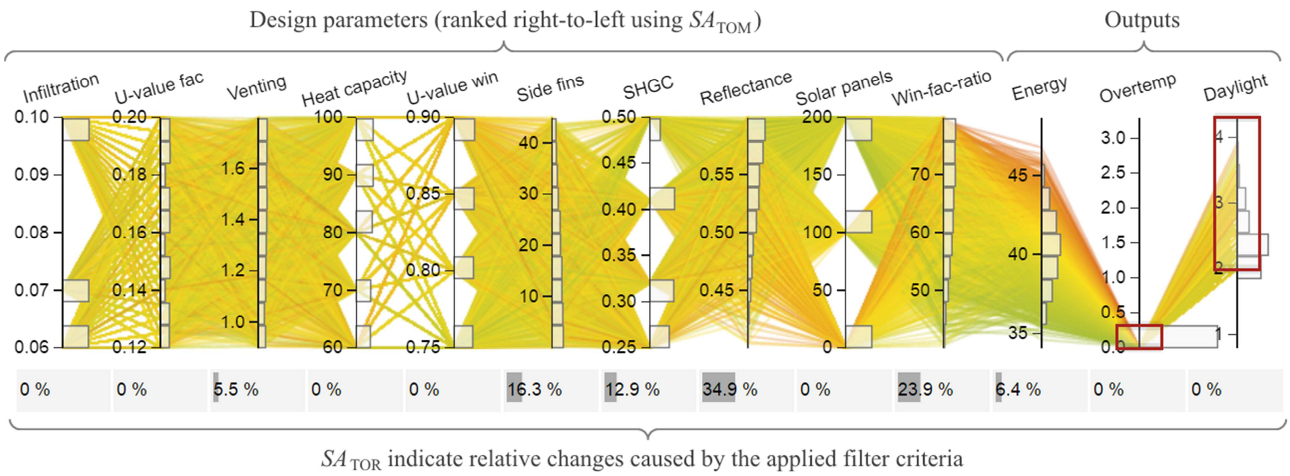

Figure 1: Parallel coordinate plot (PCP) with histograms showing distribution of the simulations, which remain after filtering. The bar plots how much the distributions have been affected by the filters (red rectangles). Each line in the PCP represents one simulation and is coloured according to its energy demand (green – yellow – red).

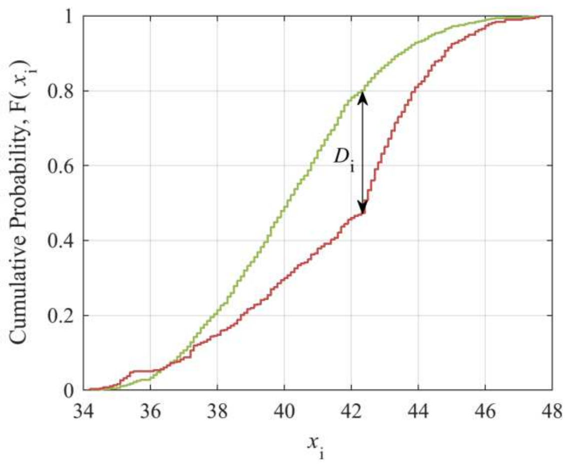

Figure 2: The maximum distance D between two cumulative distributions.

experiment is repeated with the reduced set of variable inputs. A large number of simulations must be sampled in order to represent a sufficiently large part of the multivariate design space. The modeller then ranks the inputs according to their combined output importance as explained below (FR). Finally, the multi-actor design team explores the design space by filtering inputs and outputs using an interactive parallel coordinate plot (FM). Figure 1 shows an example of the PCP, where histograms illustrate input and output distributions of simulations that meet the user defined filter criteria.

We will now briefly introduce the general concept of regionalized sensitivity analysis (RSA) when using Kolmogorov-Smirnov statistics. In the following subsection, we explain the novel sensitivity measures, TOM and TOR, which we have developed based on the Kolmogorov-Smirnov two-sample statistics (TOM and TOR are derived from the first author's name with the last letter referring to *Multiple* and *Real-time*).

The essential part of RSA is filtering (also known as Monte Carlo Filtering). The filtering is typically applied to model outputs based on specific constraints, e.g. maximum value for energy demand or minimum criteria for daylight availability. The filter criteria split the simulations into two groups: 1) the *behavioural* simulations meeting the filtering criteria, and 2) the *non-behavioural* simulations (Saltelli et al. 2008). The reason for doing so is to identify input values that most likely will result in behavioural simulations. These behavioural simulations represent building designs with high performance. After filtering, for each parameter, there is a distribution of values belonging to the behavioural simulations and likewise for the non-behavioural simulations. The two-sample Kolmogorov-Smirnov test provides a measure of how much two distributions differ (Saltelli et al. 2008). This measure, denoted *D*, is the maximum distance between two cumulative distributions as illustrated on Figure 2. If the maximum distance is large for the i^{th} input, then this input is important in driving the model into the desired output range, and vice versa. A comparable sensitivity measure, $SA_{KS2,i}$, for the i^{th} parameter is obtained from the size of D_i relative to the summed D_i's, see equation (1). Comparison of the D_i's shows which inputs are important and which are not.

$$SA_{KS2,i} = \frac{D_i}{\sum_i D_i} \tag{1}$$

TOM – Factor ranking for multiple outputs

Here, we present a novel SA method denoted TOM, which ranks inputs according to their influence on *multiple* outputs (FR). The method builds upon the above concept of splitting a large set of simulations, S_A, into two subsets, S_B and S_N. Key to this approach is that filter criteria may be applied to any number of outputs (and inputs) and still two subsets remain. Of course, the number of "behavioural" simulations decreases for each additional constraint. The novelty here is to do this "splitting" by applying filter criteria to all outputs without knowing actual, project-specific constraints. Hence, the task is to develop a strategy to define criteria values for all outputs in a generic way. Afterwards, we define a sensitivity measure SA_{TOM} based on the KS2 statistics D_i.

In the proposed methodology, we first assign an index to each Monte Carlo simulation. Next, we sort each output in ascending order while keeping a reference to the simulations' indices (Figur 3 top left). For each sequence of output values, we now choose a random starting point (corresponding to a minimum criterion) and select *Q* number of simulations above this value (see arrows on Figure 3). If this selection exceeds the maximum value,

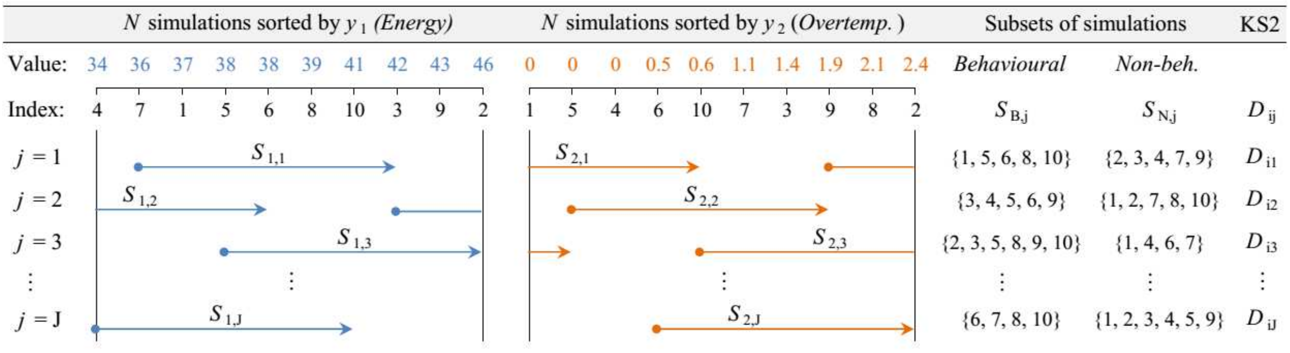

Figure 3: 10 simulations sorted for two outputs. Subsets S_1,j and S_2,j are randomly selected for each repetition, j. The subsets are illustrated with arrows, which have random starting points but same length (Q simulations).

the remaining simulations are chosen from the lowest output value. These steps are repeated J times. For each repetition j, we obtain a subset of "behavioural" simulations $S_{B,j}$ from the intersections of the subsets $S_{1,j}$ and $S_{2,j}$. For example, the simulations with indices 1, 5, 6, 8, and 10 all occur in both subsets $S_{1,1}$ and $S_{2,1}$. At the same time, we get a subset of "non-behavioural" simulations, $S_{N,j}$, from the difference of $S_{B,j}$ and $S_{A,}$. Using two subsets, we calculate the D_{ij} for all inputs. Finally, we use the average values of D_{ij}'s to establish the sensitivity measure $SA_{TOM,i}$ for all input parameters – equation (2). This measure indicates the i^{th} input's relative importance with respect to all outputs. To use this TOM method, we first need to assess how large the random subsets must be and how many repetitions are necessary, i.e. estimate Q and J.

$$D_{i,av} = \sum_{j=1}^{J} D_{ij} \quad \rightarrow \quad SA_{TOM,i} = \frac{D_{i,av}}{\sum_{i} D_{i,av}} \qquad (2)$$

Figure 4 illustrates how the size of the randomly picked subsets affects how much the randomly chosen subsets will intersect. If the subsets are too small, there will often be no intersection (Figure 4 top left). In those cases, the "non-behavioural" set, $S_{B,j}$, will equal the total set of simulations, S_A, and consequently all D_{ij}'s will be zero. If so, the step is repeated until a non-empty intersection is obtained. To get the most distinctive maximum distances, we want the "non-behavioural" set to constitute roughly half the size of the total set. From logical reasoning and experience, it seems that for m uncorrelated outputs and infinitely many repetitions J, the number of "non-behavioural" simulations converges to 50% of N when defining the subset size Q as in equation (3).

$$Q = N \cdot 0.5^{1/m} \qquad (3)$$

As mentioned, we also need to estimate the number of repetitions, J, necessary for convergence of the distance $D_{i,av}$'s. In the "Results and discussion" section, we show that the sensitivity measures converge after ~300 repetitions for the case study. However, to recommend a general value for J, we need to test additional models with different levels of complexity and number of inputs.

In the "Results and discussion" session, we apply the TOM method to several benchmark models. The results show that the method estimates the inputs' total effects. Thus, TOM can be used for Factor Ranking, which was the intention of this approach. The method may also be used for Factor Fixing. One approach is to apply the null hypothesis of KS2, which checks whether the cumulative distributions for the subsets $S_{B,ij}$ and $S_{N,ij}$ for the i^{th} input are the same at a given significance level, α. If the null hypothesis is accepted (for all J), the i^{th} input is non-influential. However, our experience has shown it difficult to find a specific significance level that avoids type I and type II errors for different models. Instead, we propose to include a "dummy" input, which does not affect the output (Mara et al. 2017). If $D_{i,av}$ for the i^{th} input is similar to $D_{dummy,av}$, then the i^{th} input must have limited or no influence and may be fixated.

A final remark relates to our choice of comparison of cumulative distributions. For the TOM method, we compare the non-behavioural set, S_N with the entire set, S_A, in order to calculate D_{ij}. Instead, we could have chosen to compare S_B with S_A or S_B with S_N. The differences are illustrated on Figure 5. However, our initial testing have indicated that the relative measures $SA_{TOM,i}$ are almost identical, no matter which two sets are used to calculate D_i. We have chosen to compare S_N with S_A since S_B might be an empty set if there is no intersection of the subsets (if Q is small).

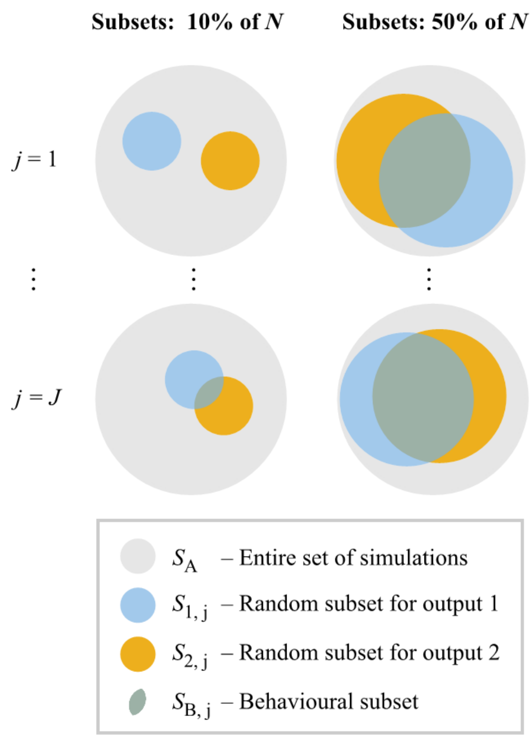

Figure 4: Conceptual illustration of how intersections of randomly selected subsets change. Smaller subsets may have no intersection and thus no behavioral simulations

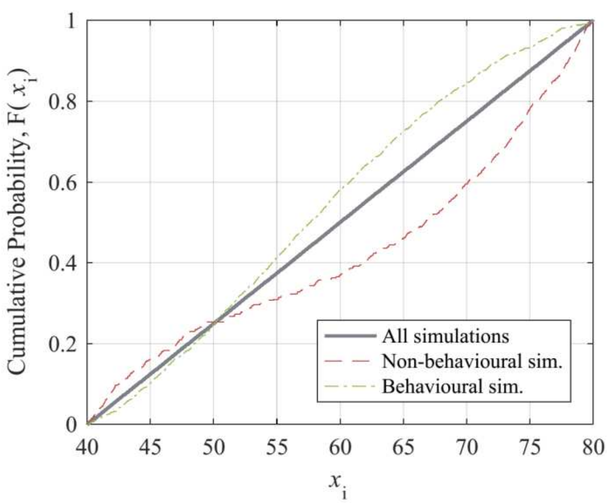

Figure 5: Cumulative distributions for S_A, S_B, and S_N for a uniformly distributed input, x_i.

Conclusively, we have now established the RSA method, TOM, which ranks inputs according their importance towards multiple outputs. Before testing TOM on a building simulation case, we present another RSA approach, TOR, which is used to highlight important parameters, which have been affected the most during real-time Monte Carlo filtering in the PCP.

TOR – Real-time highlight of important parameters

This RSA method, denoted TOR, helps highlight the parameters affected, when users add constraints to simulation data in a parallel coordinate plot (PCP). The interactive PCP is a powerful tool to analyse multivariate data in "real-time". In a building design context, the design team may filter the output coordinates in accordance with building code criteria. The remaining "behavioural" simulations indicate regions or limits of the input space that meet the criteria. For example, the distribution of the mean room *reflectance* in Figure 1 is highly skewed and favours high reflectance values after applying building code criteria. In addition, the design team may assess design choices by applying filters to input coordinates. The remaining distributions reveal the consequences of such design choices. In the same example, the design team may test if it is possible to avoid solar panels and at the same time have a high window-to-facade-ratio. Despite its strengths, the PCP becomes difficult to interpret when the number of parameters increases or when the distributions are non-uniform.

Here, we suggest using KS2 to assess how much the behavioural distributions differ from the initial distributions, when applying filters in the PCP. The user-defined filters split the simulations into a behavioural set and non-behavioural set. Therefore, we do not need to do this splitting in a generic way as in the TOM method. In this approach, we calculate the D_i's for the distributions of the behavioural set, S_B, and the entire set of simulations, S_A. In real-time, we calculate and visualize the relative distances D_i's each time a filter is applied. We suggest using bar plots to visualize the relative D_i's and thus direct the user's attention towards the parameters, which have been affected by the user-defined constraints. Notably, this method works for both inputs and outputs. Moreover, it enables the modeller to include more parameters in the Monte Carlo method, which is beneficial for building simulations that contain many design parameters and performance criteria.

Results and discussion

First, we use four benchmark models to compare the TOM method against the well-established methods of Sobol and Morris. Next, we use a building case study to test the method when considering multiple outputs. In addition, this we assess how much the sensitivity measure, SA_{TOM}, depends on the sample size N and the number of repetitions J. Finally, we exemplify how to use the TOR approach together with the parallel coordinate plot.

Benchmark models with single output

To assess the TOM method, we apply it to two non-linear and non-additive benchmark equations referred to as *"Primer"* (Saltelli et al. 2008) and *Ishigami* (Saltelli et al. 2000), respectively.

$$y = \sum_{i=1}^{4} W_i Z_i \tag{4}$$

where $Z_i \sim N(\mu_Z, \sigma_i)$, $W_i \sim N(\mu_{W,i}, \sigma_i)$, $\mu_Z = 0$, $\mu_{W,i} = 0.5i$, and $i = \sigma_i = 1,2,3,4$.

$$y = \sin(X_1) + 7\sin^2(X_2) + 0.1\ X_3^{\ 4}\sin(X_1) \tag{5}$$

where $X_i \sim U(-\pi, \pi)$. The three SA methods require different sampling techniques. For TOM, we use 1.000 and 10.000 calculations. The error bars indicate one standard deviation when repeating the method 50 times with J = 200. For Sobol' variance decomposition, we apply 100.000 calculations. For Morris, the number of

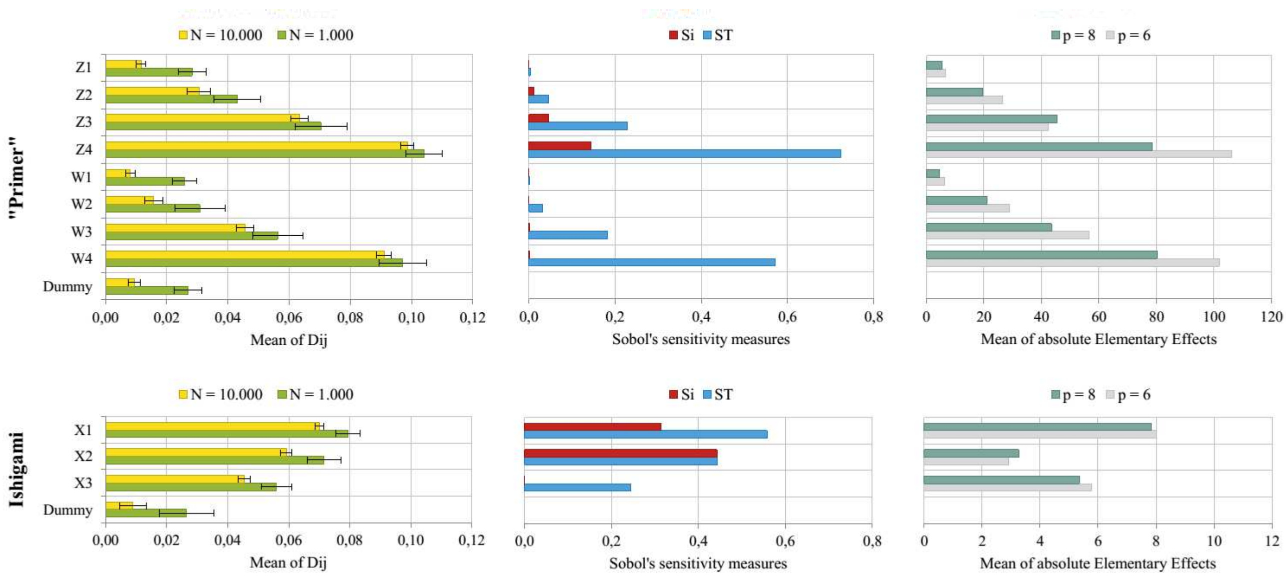

Figure 6: Results of sensitivity analysis for two benchmark models, using the TOM, Sobol', and Morris method.

trajectories is adjusted, such that 1.000 calculations are used.

Figure 6 shows how the sensitivity measures compare. The TOM method provides the same ranking of inputs as Sobol' total effects, S_T. However, the dummy variable is not zero. For the "Primer" model, the dummy has approximately the same size as Z_1 and W_1, and thus indicate that these may be fixated in a FF setting. This corresponds well with the results from Sobol' and Morris. However, the Morris method wrongfully ranks X_3 as more sensitive than X_2 for the Ishigami model. For the "Primer" model, W_3 ranks higher than Z_3 when the p-level is 6. Since the inputs in this model are normally distributed, it is necessary to truncate them (using three standard deviations) in order to apply Morris. This explains the dependence on p-level. Another issue of Morris arises when the input distributions are discrete due to a possible "misfit" with the p-level. In building performance simulations, inputs are often discrete.

Two additional benchmark models, Sobol's g-function and the Dixon-Price function, have been tested in similar manner. For those, the TOM method also produces the same ranking as the ones obtained from Sobol's total effects. For all four benchmark models, we try to estimate the number of samples needed to achieve the same ranking as Sobol'. To do so, we start with a sampling size of 10.000 and then reduce the sample size in steps of 100 (until the size is 1.000 and then in steps of 10) until the ranking differs from Sobol'. This procedure has been repeated 10 times for each model. On average, the ranking starts to differ from Sobol' when the sampling size becomes less than 800, 420, 340, and 960, for the four models respectively.

In the following, we use a building performance model to assess the TOM method for multiple outputs.

Building case study with multiple outputs

As case study, we consider a 15.000 m² educational institution during a conceptual design stage (Figure 7). The design proposal contains a floor plan, but fenestration, shading, and more, have not been defined. We may describe the "variability" of these "undecided"

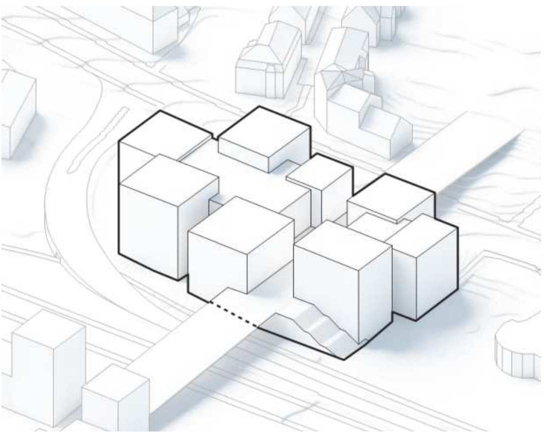

Figure 7: Early design draft of the educational institution. Illustration: EFFEKT Architects.

Table 1: Input probability distributions for case study.

Input parameters	Unit	Uniform (min – max)	Discrete
Window-facade-ratio	% (m² / m²)	40 – 80	
Solar panels	m²		0; 100; 200
Reflectance (room mean)	-	0.4 – 0.6	
Solar Heat Gain Coefficient	-		0.25; 0.32; 0.41; 0.5
Side fins (louvres)	°	0 – 45	
U-value, windows	W/m² K		0.75; 0.8; 0.85; 0.9
Heat capacity (building mean)	Wh/m²		60; 70; 80; 90; 100
Venting	l/s m²	0.9 – 1.8	
U-value, facade	W/m² K	0.12 – 0.20	
Infiltration	l/s m²		0.06; 0.07; 0.10
"Dummy" input	-	0 – 1	

design parameters using uniform distributions. For example, the design team have estimated the *windows-to-facade-ratio* to be at least 40% and no more than 80%. Another variable, *infiltration*, has been varied in three discrete steps corresponding to different levels of airtightness based on Danish building regulations. Ten design parameters have been defined using such continuous, or discrete, uniform distributions. Every possible combination of these variables constitutes an infinitely large design space. A Monte Carlo experiment is conducted to evaluate 5.000 different designs options, which is assumed to represent a sufficiently large part of this global design space. Quasi-random sampling is applied using Sobol's low discrepancy sequences, LP_τ (Sobol' & Shukman 1993). This technique reduces "gaps" and "clusters" in the simulated design space, and it reaches convergence faster than ordinary random sampling. We use a "simulation engine" based on ISO 13790 to evaluate energy demand (*Energy*) and thermal comfort (*Overtemperature*). A regression model is used to assess the average daylight factor (*Daylight*) in a typical classroom. Hence, for each simulation we obtain three performance objectives, which are often contrary. That is, improving one of them often worsens one of the others.

Dependency on repetitions and simulations (TOM)

As described above, we select a random subset of "non-behavioural" simulations J number of times. For each repetition, the subset is compared to the entire simulation set by calculating the maximum distances D_{ij} between the cumulative distributions for each input, i. Here, we determine how many repeated samples J is required to reach convergence of the mean values of D_{ij}. We consider three outputs and all of the 5.000 simulations from the case study. Thus, the number of simulations in the subsets, Q, is $0.79 \cdot N$ = 3.950 in accordance with equation (3).

From Figure 8, it seems the mean values converge after ~300 repetitions. The ranking is consistent after 25 repetitions. Note that, equally sensitive inputs may occasionally change positions. The computational time grows linearly with both N and J. For 5.000 simulations with 300 repetitions, it was less than 6 seconds using a standard laptop with Matlab R2016a. Thus, the computational time is negligible compared to that of building performance simulations.

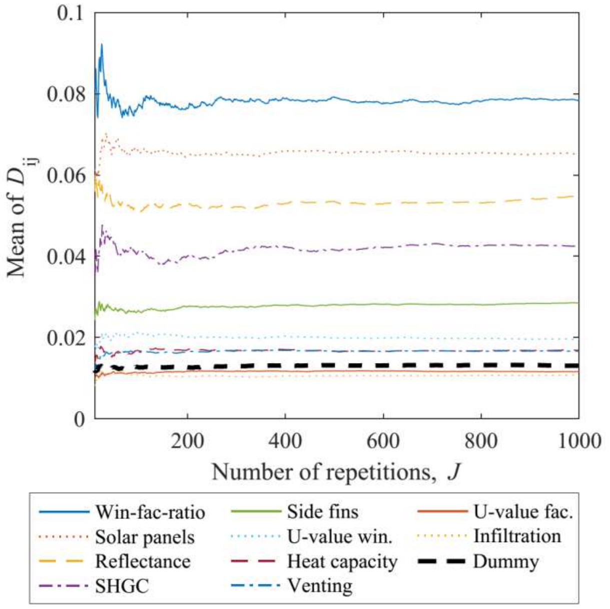

Figure 8: Convergence of the mean values of D_{ij}.

The number of simulations, *N*, required for SA is often important to the modeller, when choosing which sensitivity method to apply. *N* typically depends on the number of inputs, the complexity of the model, and the sampling strategy. In regionalized sensitivity analysis, *N* must typically be 100 times the number of inputs (Pianosi et al. 2016). Figure 9 illustrates how the mean of D_{ij}'s converge with increasing number of simulations, *N*. The former involves only one output, whereas the latter involves all three outputs. With the exception of the dummy, the mean values seem to converge when *N* exceeds 1.000 for case a single output. This fits well with the aforementioned "rule-of-thumb" suggested by Pianosi et al. Roughly three times as many is needed in the case of three outputs (for this case study). Both plots show some fluctuations of the mean value, but the ranking of the most important inputs is consistent when *N* is larger than 1.000.

Sensitivity analysis for multiple, correlated outputs

As described earlier, the main purpose of TOM is to rank inputs with respect to their sensitivity towards multiple outputs (FR). However, a multiple output measure may also be obtained from ordinary sensitivity methods by combining the sensitivity measures for the individual outputs using a weighting system. In the context of building performance simulation, some outputs may be highly correlated, since the design has to comply with several, correlated performance indicators. Now, we compare sensitivity measures for the case study in the following steps:

1. For each output, we rank inputs using the relative sensitivity measures obtained from SRC, Morris, and TOM.
2. The results from the TOM method for three outputs are discussed.
3. The TOM method is compared to a weighted SRC method in the case of 7 outputs from which 5 are identical (thus correlated).

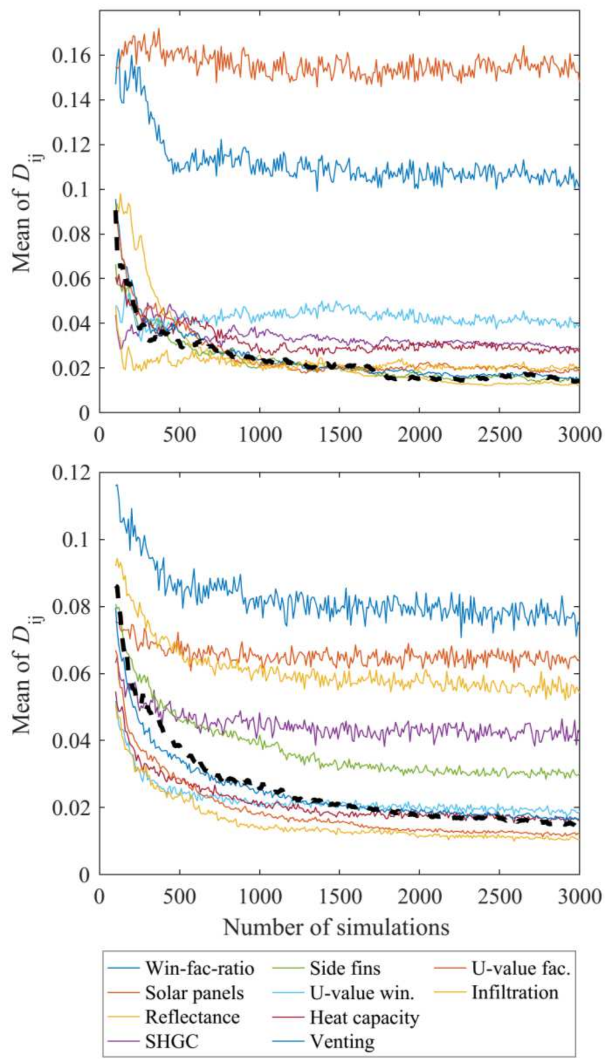

Figure 9: Mean of D_{ij} in steps of 10 simulations with respect one output, Energy demand (top), and all three outputs (bottom).

The sensitivity measures from TOM and SRC are based on 5.000 simulations. The number of repetitions for TOM is 500. For Morris, we discretize all inputs into 8 levels and run 450 trajectories (4.950 simulations). The results are shown in Table 2, which also include the results from TOM with respect to all three outputs.

In Table 2, the inputs have been sorted with respect to their sensitivity towards all three outputs (TOM). First, we consider one output at a time only. From SRC, we obtain the coefficients of determination, R^2, which are 0.96, 0.42, and 0.96, respectively. Thus, the outputs *Energy Demand* and *Daylight Factor* are nearly linear for these idealized building performance models. We observe that the different methods provide the same ranking of the three highest ranked inputs. However, we do not necessarily expect the same ranking for SRC as for the two others, since the SRC measures are obtained from linear regression, and, therefore, they only include linear effects. The ranking of less important inputs differ slightly. For example, *SHGC* ranks 7 with SRC, 4 with Morris, and 5 with TOM (SRC does not capture *SGHC*'s

Table 2: Sensitivity measures obtained from SRC, Morris (EE), and TOM.

Parameter	**E, O, D**		**Energy Demand (E)**						**Overtemperature (O)**						**Daylight Factor (D)**					
	TOM		SRC		EE		TOM		SRC		EE		TOM		SRC		EE		TOM	
Win-fac-ratio	1	***23%***	2	*27%*	2	*28%*	2	*24%*	2	*31%*	2	*33%*	2	*27%*	1	*34%*	1	*35%*	1	*34%*
Solar panels	2	***19%***	1	*31%*	1	*29%*	1	*36%*	-	*0%*	-	*0%*	10	*1%*	-	*0%*	-	*0%*	10	*1%*
Reflectance	3	***17%***	9	*2%*	9	*2%*	8	*3%*	7	*1%*	-	*0%*	5	*5%*	2	*34%*	2	*35%*	2	*32%*
SHGC	4	***13%***	7	*4%*	4	*8%*	5	*6%*	1	*40%*	1	*42%*	1	*40%*	4	*13%*	4	*11%*	4	*9%*
Side fins	5	***9%***	10	*1%*	10	*2%*	10	*2%*	5	*4%*	5	*4%*	6	*4%*	3	*18%*	3	*19%*	3	*15%*
U-value win.	6	***5%***	3	*13%*	3	*11%*	3	*9%*	6	*3%*	6	*2%*	7	*4%*	-	*0%*	-	*0%*	9	*1%*
Heat capacity	7	***5%***	4	*9%*	5	*8%*	4	*6%*	4	*8%*	4	*7%*	4	*6%*	-	*0%*	-	*0%*	5	*3%*
Venting	8	***4%***	8	*3%*	8	*3%*	9	*3%*	3	*12%*	3	*12%*	3	*10%*	-	*0%*	-	*0%*	6	*3%*
U-value fac.	9	***3%***	6	*5%*	7	*4%*	7	*4%*	-	*0%*	7	*1%*	8	*1%*	-	*0%*	-	*0%*	7	*1%*
Infiltration	10	***3%***	5	*6%*	6	*5%*	6	*5%*	-	*0%*	-	*0%*	9	*1%*	-	*0%*	-	*0%*	8	*1%*
Dummy	6.5						11						4.5						4.5	

interaction effects with e.g. *Win-fac-ratio*). The bottom row shows how the dummy variable would rank for the TOM method. For example, the dummy would be placed between the fourth and fifth highest ranked inputs for *Overtemperature* and *Daylight Factor*. Thus, the last six inputs may be considered non-influential for these outputs according to TOM. The dummy ranks last (11) for *Energy Demand*.

We now turn our attention to the overall ranking towards multiple outputs, which are shown in the leftmost column in Table 2. We notice that each of the individually most important inputs, *Win-fac-ratio*, *Solar panels*, and *SGHC*, end up on the first, second, and fourth place. Remarkably, *Reflectance* ranks third overall even though it only ranks second for *Daylight Factor* and it is nearly insignificant for *Energy Demand* and *Overtemperature*. The reason is that only a few inputs affect *Daylight Factor* and *Reflectance* is a major contributor to the variance of this output. For the case study, this high ranking of *Reflectance* stresses out its importance to the design team. Thus, the design team must consider this interior design parameter at the early stages even though such parameters are often not determined before the late design phases. For example, the design team may search for a lower limit for *Reflectance* from Factor Mapping (see Figure 10).

Finally, we wish to assess how the TOM method ranks inputs when some of the outputs are correlated. This is often the case in the context of building performance, since the design has to comply with several, correlated performance indicators. Examples of such indicators are the number of hours with indoor temperatures above 26 and 27 °C, the number of hours the indoor climate falls into different categories, and heating demand, cooling demand, and total energy demand. In a holistic building design context, it is desirable to give less weight to such performance indicators since we wish to optimize the overall performance of the building. Here, we construct five "artificial" and 100% correlated outputs by including the output *Daylight Factor* five times for the TOM analysis. We also consider the outputs *Energy Demand* and *Overtemperature*. For comparison, we create an overall "weighted-sum" measure from SRC. Table 3 shows the rankings obtained from the TOM method and the weighted-sum SRC approach (WS-SRC) together with sensitivity measures for the single output, *Daylight Factor*. Naturally, the percentages from WS-SRC are close to those from SRC for *Daylight Factor*. In contrast, the TOM method puts less weight to these fully correlated outputs. For example, *Solar panels* (sensitive to *Energy demand*) ranks third and *Venting* (sensitive to *Overtemperature*) ranks sixth. The reason is that the randomly selected subsets for the correlated outputs will often intersect and therefore their contributions to the behavioural subset will often be very similar. In conclusion, the TOM method helps rank inputs with respect to multiple outputs with less weight on correlated outputs, which is a desirable feature in holistic building design.

Table 3: Ranking with respect to multiple, correlated outputs (blue). For comparison, the sensitivity measures for the "duplicated output" Daylight Factor are shown to the right.

Parameter	**E, O, 5 x D**				**Daylight Factor**			
	TOM		WS-SRC		TOM		SRC	
Win-fac-ratio	1	*26%*	1	*33%*	1	*34%*	1	*34%*
Reflectance	2	*22%*	2	*25%*	2	*32%*	2	*34%*
Solar panels	3	*11%*	5	*4%*	10	*1%*	-	*0%*
Side fins	4	*11%*	4	*14%*	3	*15%*	3	*18%*
SHGC	5	*10%*	3	*15%*	4	*9%*	4	*13%*
Venting	6	*5%*	8	*2%*	6	*3%*	-	*0%*
U-value win.	7	*4%*	7	*2%*	9	*1%*	-	*0%*
Heat capacity	8	*4%*	6	*3%*	5	*3%*	-	*0%*
U-value fac.	9	*4%*	10	*1%*	7	*1%*	-	*0%*
Infiltration	10	*3%*	9	*1%*	8	*1%*	-	*0%*

Real-time highlight of importance (TOR)

Now, we demonstrate how the TOR method improves the use of the interactive parallel coordinate plot. As mentioned, the PCP is very intuitive and effective when exploring and analysing multivariate data. However, changes may be difficult to observe – especially if the plot contains many parameters. Here, the Kolmogorov-Smirnov maximum distances, D_i's, are based solely on the user-defined filter criteria. Therefore, we need not define subset size Q or number of repetitions J.

Figure 10 shows examples of a PCP with different filters applied. Bar plots show the relative sizes of the D_i's for the parameters with no filters applied. The 10 input parameters have been arranged according to the ranking obtained from the TOM method, such that the left-most inputs are the least important, and vice versa. In the topmost plot, we have removed all simulations, which have *Overtemperature*-values larger than zero. This constraint largely affects the remaining distributions of *SHGC* (30.7%) and *Window-to-facade-ratio* (20.2%). In addition, it affects the remaining distributions for *Energy Demand* (18.7%) and *Daylight Factor* (17.6%).

In the middle plot, we have added constraints to all three outputs in accordance with Danish building code regulations. Noteworthy, the TOR sensitivity measures do not provide the exact same ranking as the initial ranking from TOM, because the user-defined filters are different from the J random applied filters used for TOM. For example, *Reflectance*, and not *Win-fac-ratio*, has been affected the most by the filters applied to the three outputs in the middle plot.

In the bottom plot, we assume the design team aims for a mean room *reflectance* larger than 0.5, because of its importance. Moreover, we assume the design team strives for a *window-to-facade-ratio* larger than 60%. The TOR measures and histograms show that this combination of criteria greatly affects the remaining distributions of values for *Solar Panels* and *SHGC*. The TOR measures also indicate some influence from *Heat Capacity* and *U-value windows*, which is harder to notice from the histograms.

Conclusively, the TOR method helps decision makers focus on parameters that matter the most, and see the consequences of design choices. Especially, if the initial distributions are not uniform, changes are difficult to observe. However, when few simulations remain, the KS2 statistics will become inaccurate and it may be erroneous to draw conclusions about trends based on the histograms. To overcome this, metamodeling may be applied to create new predictions in the reduced

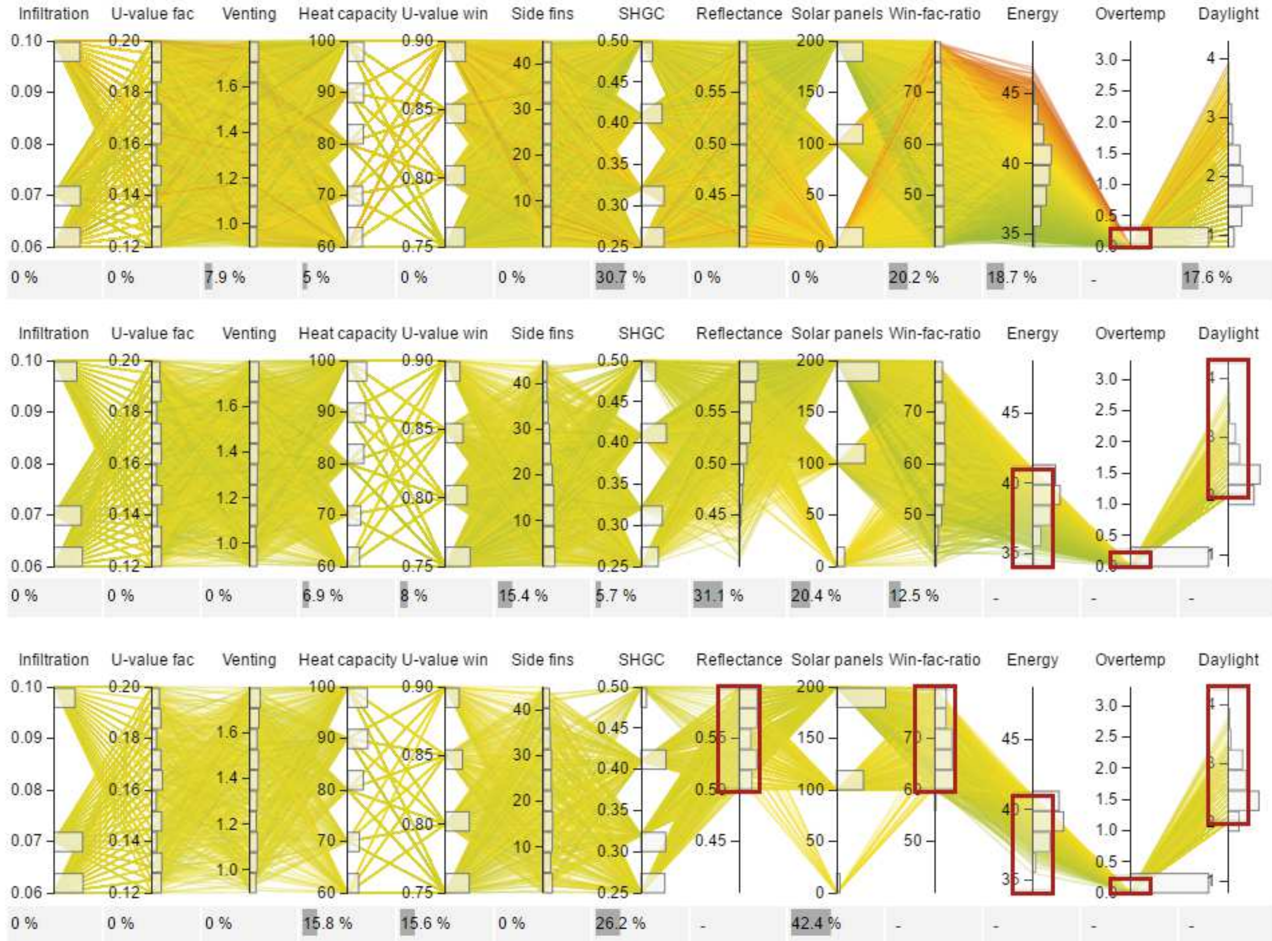

Figure 10: PCP's with user-defined filters illustrated with red rectangles. Based on TOR, bar plots indicate which parameters have been affected the most by the filtering. The inputs have been ranked right-to-left using TOM.

subspace as discussed by Østergård et al. (2017).

Combining TOM and TOR

TOM and TOR may both be used in a design process with Monte Carlo simulations. First, the modeller runs ~1.000·m simulations and use TOM to fixate the least influential parameters with respect to all m outputs (FF). Afterwards, a very large set of simulations is run to represent the global design space (optionally using fast metamodels). Then, TOM is used to rank the inputs, e.g. for positioning in the PCP (FR). Finally, TOR is used to highlight changes during real-time exploration in the PCP (FM).

Conclusion

We have presented two novel sensitivity methods, denoted TOM and TOR, which help decision makers focus on the most important parameters during building design. A precondition is the use of the Monte Carlo method to perform thousands of simulations to explore the multivariate design space. In contrast to the popular Morris and variance-based methods, TOM and TOR can be used for multiple outputs and they work with random or quasi-random sampling.

To test the TOM method, we used four non-linear and non-additive benchmark models and compared with Morris and Sobol'. The TOM method provided the same ranking of inputs as Sobol', even when Morris did not. A building case study showed that TOM puts less weight on correlated outputs, which is preferable in holistic building design. The TOR method makes it easier to perform real-time exploration of multivariate data in the parallel coordinate plot. TOR highlights the parameters, which are most affected by user-defined criteria. This allows more parameters to be included in the analysis without the PCP becoming unmanageable. The reader may download Matlab code for TOM or test the combination of PCP and TOR on:
http://buildingdesign.moe.dk/phd/ibpsa.html

In future work, we wish to investigate larger case studies with more inputs and outputs. In addition, we will assess how to use the dummy variable or hypothesis tests to identify truly non-influenial inputs. Alternatives to the KS2 test, such as the Anderson-Darling test, may improve the accuracy of the methods (Engmann & Cousineau 2011). Finally, the methods may be combined with the regionalized sensitivity measure, PAWN, to detect interaction effects (Pianosi & Wagener 2015).

Acknowledgements

Funding was provided by Innovation Fund Denmark (grant number 4019-00009) and MOE A/S. The work was part of an industrial doctorate program with Aalborg University and consultancy company

References

Burhenne, S. (2013). Monte Carlo Based Uncertainty and Sensitivity Analysis for Building Performance Simulation. *PhD dissertation, Karlsruhe Institute of Technology.*

Confalonieri, R. et al. (2010). Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecological Modelling 221*. 1897–1906.

Das, P. et al. (2014) Using probabilistic sampling-based sensitivity analyses for indoor air quality modelling. *Building and Environment 78*. 171–182.

Engmann, S. and Cousineau, D. (2011). Comparing Distributions : the Two-Sample Anderson-Darling Test As an Alternative To the Kolmogorov-Smirnoff Test. *Journal of Applied Quantitative Methods 6(3)*, 1–17.

Mara, T.A. et al. (2017), Addressing factors fixing setting from given data: A comparison of different methods. Environmental Modelling & Software 87, 29–38.

Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. *Techno-metrics 33(2)*, 161–174.

Nguyen, A. and Reiter, S. (2015), A performance comparison of sensitivity analysis methods for building energy models. *Building Simulation 8(6)*, 651–664.

Pianosi, F. et al. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software 79*, 214–232.

Pianosi, F. and Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software 67*, 1–11.

Saltelli, A. et al. (2000). Sensitivity analysis. *Wiley.*

Saltelli, A. et al. (2008). Global sensitivity analysis: the primer. *Wiley & Sons.*

Sobol', I.M. and Shukman, B.V. (1993). Random and quasirandom sequences: Numerical estimates of uniformity of distribution. *Mathematical and Computer Modelling 18(8)*, 39–45.

Song, X. et al. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology 523*, 739–757.

Tian, W. (2013), A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews 20*, 411–419.

Yang, J. (2011), Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. *Environmental Modelling & Software 26(4)*, 444–457.

Østergård, T. et al. (2017). Early Building Design - Informed decision-making by exploring multidimensional design space using sensitivity analysis. *Energy and Buildings (submitted June 2016).*