Proactive building simulations for early design support

Multi-actor decision-making based on Monte Carlo simulations and global sensitivity analysis

Østergård, Torben

DOI (link to publication from Publisher):
10.5278/vbn.phd.eng.00017

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

PROACTIVE BUILDING SIMULATIONS FOR EARLY DESIGN SUPPORT

MULTI-ACTOR DECISION-MAKING BASED ON MONTE CARLO SIMULATIONS AND GLOBAL SENSITIVITY ANALYSIS

BY
TORBEN ØSTERGÅRD

DISSERTATION SUBMITTED 2017

AALBORG UNIVERSITY
DENMARK

PROACTIVE BUILDING SIMULATIONS FOR EARLY DESIGN SUPPORT

MULTI-ACTOR DECISION-MAKING BASED ON MONTE CARLO SIMULATIONS AND GLOBAL SENSITIVITY ANALYSIS

by

Torben Østergård

AALBORG UNIVERSITY
DENMARK

Dissertation submitted

June 2017

.

Dissertation submitted: October 2017

PhD supervisor: Associate Professor Rasmus L. Jensen
Aalborg University

Company PhD supervisor: Technical Director Steffen E. Maagaard
MOE A/S

PhD committee: Professor Per Heiselberg (chairman)
Aalborg University

Professor Jan Hensen
Eindhoven University of Technology

Director Duncan Horswill
Bjarke Ingels Group

PhD Series: Faculty of Engineering and Science, Aalborg University

Institut: Department of Civil Engineering

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-091-3

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2017

CURRICULUM VITAE

Name	Torben Østergård
Date of birth	16.05.1982
Phone	+45 6130 2798
E-mail	torbeniha@gmail.com
ResearchGate	www.researchgate.net/profile/Torben_Ostergard
LinkedIn	www.linkedin.com/in/torben-oestergaard
Facebook	www.facebook.com/ostergard

PROFESSIONAL EXPERIENCE

2014 – 2017	Industrial PhD Student MOE A/S and Aalborg University
2013 – 2014	Consulting Engineer, Energy and Indoor Climate MOE A/S
2011 – 2013	Assistant Engineer MOE A/S
2008 – 2012	Assistant Teacher, Mathematics and Physics Aarhus University | Engineering College of Aarhus

EDUCATION

2011 – 2013	MSc in Engineering, Architectural Engineering Aarhus University
2007 – 2011	BEng in Civil Engineering, Energy and Indoor Climate Engineering College of Aarhus
2004 – 2007	Physics (Minor without diploma) Aarhus University

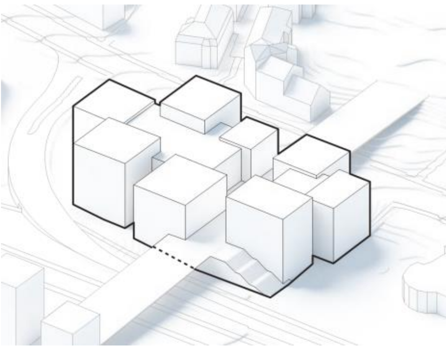
HONORS AND AWARDS

Best Paper Award	14th International Conference of IBPSA, Hyderabad, India
IDA prisen 2011	The year's most innovative and best-communicated graduation project, Engineering College of Aarhus

ENGLISH SUMMARY

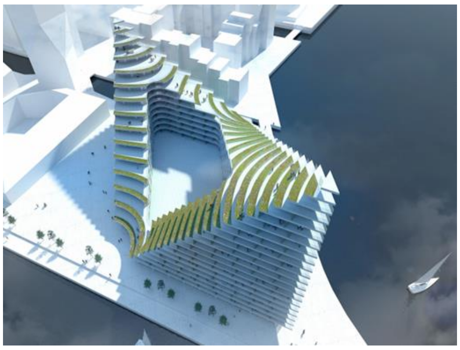
This industrial PhD study is motivated by the many challenges experienced by practitioners of building design. The design team must meet increasing requirements for energy demand and indoor environment, while addressing demands for aesthetics, functionality, and building costs. Especially in the early design stages, it is difficult to guide decision-makers in improving building performance. The *early design* begins with the initial drafts of building form and room layout and finishes with the transition to the detailed project stages. These early stages are characterized by large uncertainties and considerable design freedom, and, moreover, wide-ranging changes occur frequently. Decision-making are influenced by multiple stakeholders, e.g. building owners, architects, and engineers, which have diverse ideas and demands. Simulation-based support for this design team often relies on deterministic simulations in a time-consuming, evaluative manner. The ambition with this project is to facilitate proactive and holistic guidance which points out the most important parameters and identifies favorable parts of an enlarged solution space.

Initial draft of educational building



EFFEKT architects

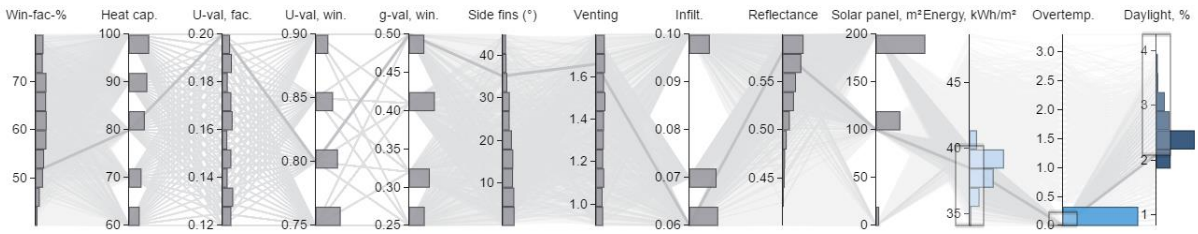
Schematic design of residential building



BIG architects

The PhD study has led to the development of a novel simulation framework, and various tools, that help overcome the above challenges. The many variable design parameters span a high-dimensional *design space* which is assessed by thousands of statistically chosen building performance simulations (BPS). This allows the design team to identify a high-performing *solution space* instead of just assessing a single design and adjusting it in a manual trial-and-error approach until minimum criteria is met. The proposed method expands the “architectural freedom” and helps avoid inappropriate decisions leading to poor, or costly, performance.

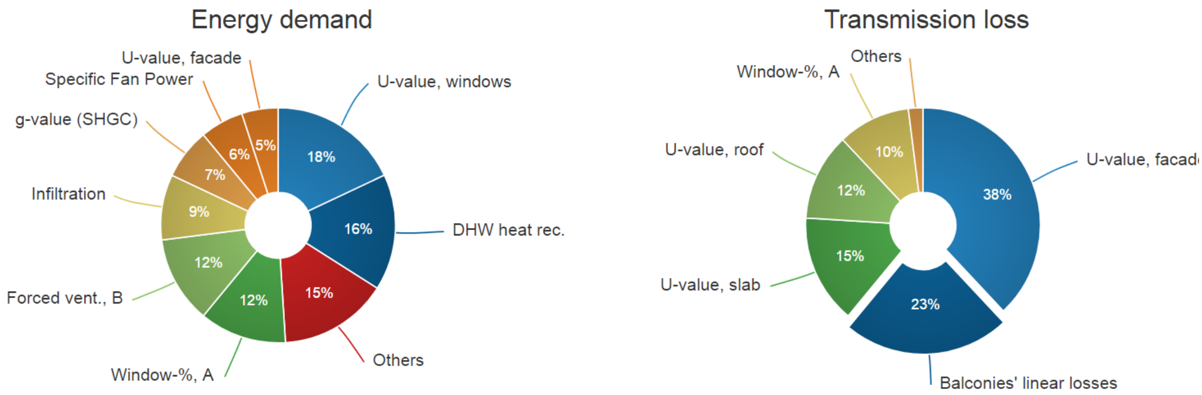
Interactive plot for design space exploration



Interactive visualizations have been implemented to encourage multi-actor collaboration and design space exploration. These interactive plots makes it easier to change various design parameters and observe, in real-time, their effects on performance and consequences for other designs. The method was applied to the preliminary design of a 15.000 m² educational building. Based on 5.000 simulations, the design team pursued the maximum window-to-wall-ratio and observed a need for renewables, i.e. the amount of photovoltaics to balance the energy frame. To avoid renewables, the design team could immediate find a suitable window-to-wall-ratio and at the same time notice the consequences for insulation level and room reflectance to meet energy and daylight requirements.

A crucial element of the proposed simulation framework is the use of sensitivity analysis. At a given design stage, such analysis calls attention to the most influential design parameters and it shows insignificant parameters which can be ignored during design meetings. Sensitivity analysis can also reveal interdependent design parameters that should be treated with care since optimizing on one them depends on the choices, or values, of the others. During a multi-collaborator design meeting related to a 24.000 m² residential building, sensitivity analysis helped stress the importance of the balconies' heat losses which therefore received the necessary attention.

Relative parameter sensitivity for two performance criteria

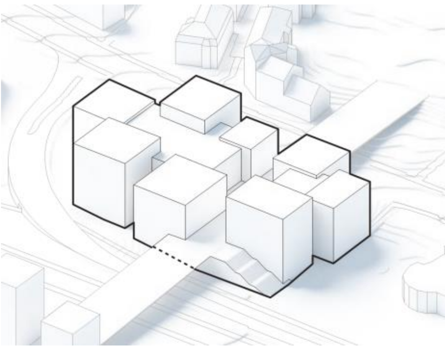


The above concepts rely on the use of Monte Carlo simulations aided by sensitivity analysis and fast metamodels. The variability of design parameters is defined using probability distributions, from which thousands of combinations are chosen, using quasi-random sampling, to represent the global design space. This PhD study advocates a paradigm shift from the traditional, deterministic simulations to this extended, stochastic approach. The reader is encouraged to visit the website *buildingdesign.moe.dk* for more examples and to get a hands-on experience with design space exploration using interactive visualizations.

DANSK RESUME

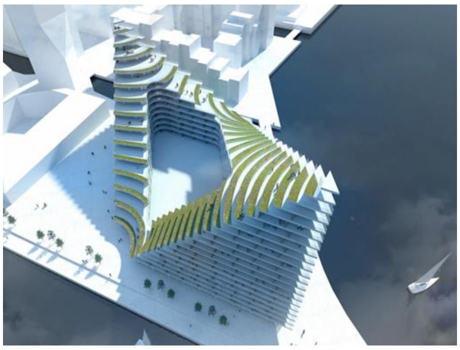
Nærværende ErhvervsPhD-studium er motiveret af en række udfordringer, som opleves af aktører inden for bygningsdesign. Designteamet skal opfylde stadigt stigende krav til energiforbrug og indeklima, samtidig med man skal imødekomme ønsker til arkitektur, funktionalitet og bygningsomkostninger. Særligt i de tidlige faser er det vanskeligt at vejlede beslutningstagere i forhold til at forbedre bygningens performance. Med angivelsen *tidligt design* refereres til en række overlappende faser startende fra den indledende konceptuelle skitseringsfase til og med forprojektet, hvor detaljeringsniveauet præciseres tilstrækkeligt til myndighedsgodkendelse. Disse tidlige faser er kendetegnet ved mange usikkerheder og stor designmæssig frihed og der forekommer ofte væsentlige ændringer i designet. Beslutninger vedrørende bygningens design og performance er influeret af adskillige interessenter, såsom bygherrer, arkitekter og ingeniører, der typisk har forskellige ideer og ønsker. Beslutningstagningen støttes ofte af tidskrævende iterative bygningssimuleringer. Denne tilgang har en evaluerende, bagudskuende karakter frem for proaktivt at anvise vejen mod bedre løsninger. Hensigten med nærværende studium er at anvende helhedsorienterede bygningssimuleringer til guide beslutningstagningen mere proaktivt og dermed identificere favorable designvalg og forbedre bygningens performance.

Skitse af uddannelsesinstitution (konkurrence)



EFFEKT architects

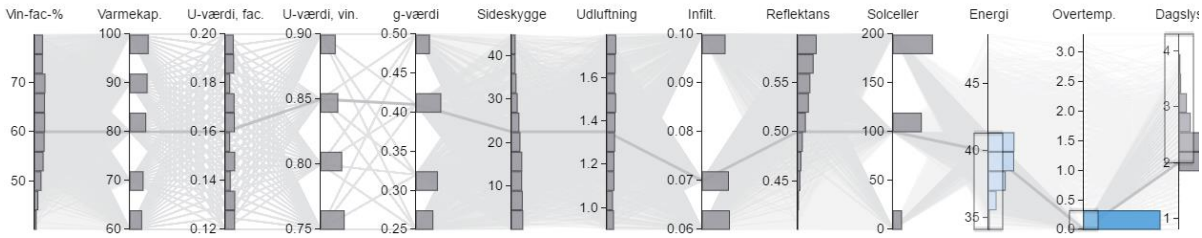
Illustration af beboelsesbygning (forprojekt)



Bjarke Ingels Group

PhD-studiet har ført til udviklingen af en ny simuleringsmetodik, samt diverse værktøjer, der hjælper til at imødekomme de ovennævnte udfordringer. De variable parametre udgør tilsammen et multidimensionalt *design rum*, som undersøges ved tusindvis af simuleringer af bygningens performance. Herved kan designteamet identificere et stort løsningsrum med høj performance. Dette ses i kontrast gængs praksis med evaluering af et specifikt design, der derefter tilpasses i en iterativ ”trial-and-error” tilgang indtil minimumskravene opnås. Den foreslåede metode udvider den arkitektoniske frihed og hjælper til at undgå dårlige beslutninger, der fordyrer byggeriet eller fører til utilstrækkelig performance.

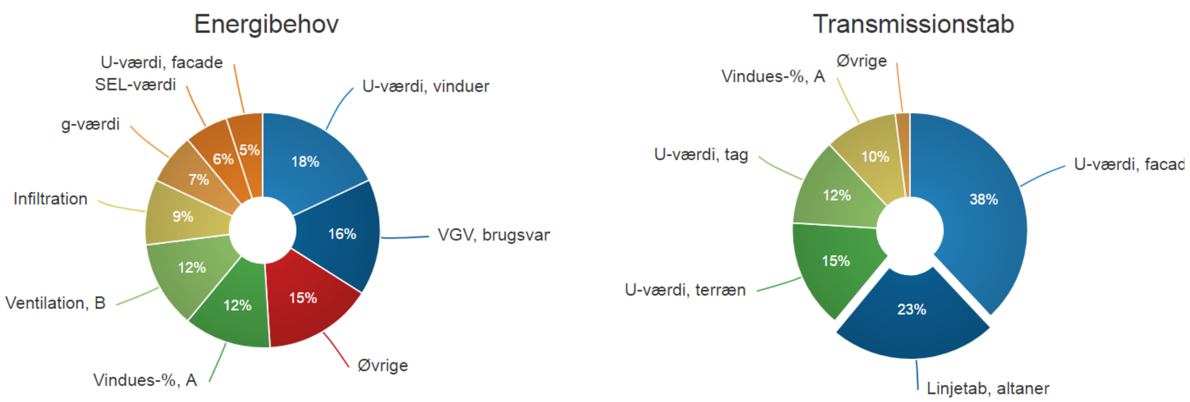
Interaktivt plot til udforskning af et 10-dimensionelt designrum med tre performance kriterier



Som en del den foreslåede designmetode avendes interaktive visualiseringer, der tilskynder forskellige fagpersoner at samarbejde til at analyse og udforske designrummet. Sådanne interaktive plots gør det lettere at afprøve designstrategier og strakt observere deres betydning for performance og vurdere, hvorvidt en specifik strategi begrænser det tilbageværende løsningsrum. Designmetoden blev anvendt ved udarbejdelsen af et konkurrenceforslag til en 15.000 m² uddannelsesinstitution. På baggrund af 5.000 simuleringer søgte designteamet den maksimalt tilladte vinduesandel i facaden, hvilket ville nødvendiggøre brug of solceller for at opfylde energirammen. Ved at bruge det interaktive plot til at sortere simuleringerne kunne teamet straks finde et mere passende spænd for vinduesandelen, hvorved solceller kunne undgås. Samtidigt fremgik det af plottet, hvilke konsekvenser den moderate vinduesandel uden solceller ville have for isoleringsmængde og indvendig reflektans for at overholde kravene til energiramme og dagslys.

Et vigtigt element i den foreslåede designmetode er anvendelse af følsomhedsanalyse. I hvert stadie vil sådan en analyse kunne påpege de mest betydningsfulde designparametre samt de ubetydelige parametre, der kan ses bort fra under designmøder. Følsomhedsanalyse kan også afsløre indbyrdes afhængige designparametre, der kræver særlig opmærksomhed, da optimering af en af disse vil afhænge af værdierne for de øvrige. Under et møde relateret til et 24.000 m² boligbyggeri hjalp følsomhedsanalysen til at understrege vigtigheden af linjetabene ved de mange altaner. De deltagende bygherrer, arkitekter, ingeniører og entreprenører kunne så agere derefter og på et tidligt stadie.

Relative følsomheder for inputparametre i forhold til to normkrav



De ovenstående metoder beror på anvendelse af Monte Carlo simuleringer suppleret med følsomhedsanalyse og hurtige regressionsmodeller. Den omtalte variabilitet af designparametre defineres ved tæthedsfunktioner, hvorfra statistisk effektive metoder anvendes til at udvælge tusinde kombinationer, der repræsenterer det højdimensionale designrum. PhD-studiet advokerer et paradigmeskifte fra traditionelle, deterministiske simuleringer til anvendelse af den omtalte stokastiske tilgang. Læseren opfordres til at besøge hjemmesiden buildingdesign.moe.dk for at eksperimentere med udforskning af globale designrum ved brug af interaktive visualiseringer.

PREFACE

The work presented in this thesis is part of an Industrial PhD project funded by MOE A/S and Innovation Fund Denmark. The work has been carried out at MOE A/S and Aalborg University in the period from July 2014 to June 2017. The author greatly appreciates these organizations, which have made the PhD possible.

THESIS OUTLINE

The core of this thesis is the following collection of articles:

Paper A *"A stochastic and holistic method to support decision-making in early building design"*
Proceedings of Building Simulation 2015.

Paper B *"Building simulations supporting decision making in early design – A review"*
Renewable and Sustainable Energy Reviews.

Paper C *"Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis"*
Energy and Buildings.

Paper D *"A comparison of six metamodeling techniques applied to building performance simulations"*
Applied Energy, revised manuscript submitted 2017.

Paper E *"Interactive building design space exploration using regionalized sensitivity analysis"*
Proceedings of Building Simulation 2017.

Paper F *"Thermal comfort in residential buildings by the millions - early design support from stochastic simulations"*
CLIMA 2016 - proceedings of the 12th REHVA World Congress

Even though, the thesis is paper-based it is presented as monograph to avoid endless self-citations and unnecessary duplication of work. Therefore, the papers A to D have been integrated directly into the main body of the text. Paper E and F are located in the appendix with references in the main text. The glossary, following Chapter 5, provides definitions and explanations of abbreviations and terms used in this thesis.

ACKNOWLEDGEMENTS

I wish to thank my company supervisor, Steffen E. Maagaard, and university supervisor, Rasmus L. Jensen, for their guidance and for many fruitful discussions. Their mutual understanding and respect for each other's interests made it easy for me to navigate in the cross field between research and industry. Steffen has extensive experience with building design and construction, which has added much relevance to the project. Rasmus has provided crucial feedback and guidance during the writing of the articles and the thesis.

In addition, I would like to thank my colleagues at MOE and Aalborg University for assistance and an enjoyable working environment. Special thanks go to Jesper Kjær, Peter Johansen, Morten Andersson, and Kristian Erlandsen along with friends who have let me stay overnight during weeklong PhD courses.

Finally, I am grateful for the love and support from my family. In particular from my girlfriend, Nanna, who has listened to all the geeky stuff and the (seldom) complaints about work. Thank you for nursing our baby son day and night – when I am sleeping or spending time away from home. Thank you for your patience, support, and trust in me. Love you always.

Torben Østergård
Aalborg University, June 2017

TABLE OF CONTENTS

GLOSSARY

Be10 (software)	Danish code-compliance software superseded by Be15.
Be15 (software)	Danish code-compliance software based on the monthly quasi-steady-state method described in ISO 13790 [5-6]. Developed by the Danish Building Research Institute.
behavioural	Subset of simulations which meet user-defined constraints or filter criteria.
BPS	Building Performance Simulation(s).
BSim (software)	Hygro-thermal bulding simulation software developed by Danish Building Research Institute [7].
deterministic	Typical simulation approach, in which a design is defined by a set of fixed inputs, i.e. without uncertainty or variability, and the resulting output has no variance, i.e. the output consists of a specific value.
DF	Daylight Factor, i.e. a measure of daylight availability under fixed, cloudy conditions.
early design (stages)	Design phases ranging from the initial drafts until the transition to detailed design.
Energy frame	Whole-building energy demand under predetermined conditions used for building code compliance in Denmark. Unit: kWh/m² floor area (primary energy).
global	Approach to explore or analyse a multidimensional input (design) space in such a way that this "global" space is represented equally by the sampled simulations. This contrasts the one-at-a-time (OAT) approach.
h>26 °C	Performance indicator for thermal comfort measured by the number of hours, in which the operative temperature exceeds 26 °C.
holistic	Designing with emphasis on overall building performance taking into account multiple building performance criteria – both qualitative and quantitative.
non-behavioural	Subset of simulations that do not meet user-defined constraints or filter criteria.
one-at-a-time (OAT)	Design space exploration and sensitivity analysis in which only one parameter is varied between consecutive, deterministic simulations.
output	Quantitative performance indicator obtained from building performance simulations.
Overtemperature	Penalty term used in Be10/Be15, which is added to the energy demand as a way to punish designs, for which the building mean temperature exceeds 26 °C. Unit: kWh/m² floor area.
PCP	Parallel Coordinate Plot.
performance indicator	Qualitative and quantitative measures of building performance, e.g. buildability and energy demand.
SA	Sensitivity Analysis, i.e. study of how the uncertainty in the output of a (BPS) model relates to the uncertainties in the input parameters.
stochastic	Simulation approach based on Monte Carlo simulations with random or quasi-random sampling to represent a multidimensional input (design) space. Contrasts the "deterministic" approach.
uncertainty	The uncertainty of a parameter due to user behaviour, unpredictable weather, deviations in materials' properties, etc. (see related term "variability").
variability	The possible range of values for a design parameter which may be assigned by the design team (see related term "uncertainty").

1 INTRODUCTION

1.1 BACKGROUND

This thesis is part of an industrial PhD project made in collaboration between the consultancy company MOE and Aalborg University. The main goal is to improve the use of building simulations to assist multi-actor decision-making during early design. Before we dive into the technicalities of building simulations and statistical methods, we wish to describe the role of the host company and the characteristics of building design in a Danish context. These settings have played a vital role for the direction and outcome of the project. We therefore introduce this project with a description of motivations and ambitions from the perspective of the host company.

The primary initiator of this project is the engineering and consulting company MOE. This Danish based company employs more than 600 people divided into the sections *buildings*, *energy and industry*, and *infrastructure* [1]. Measured by ongoing building constructions, MOE was the largest player among engineers in the Danish marked anno 2016 [2]. A branch of the building section consists of engineers specialized in energy performance, indoor climate, and sustainability. This rapidly evolving area has experienced substantial growth and attention during the last decade. Worth to mention, this is the branch in which the PhD student (the author) and the company supervisor are employed.

Motivational factors for the PhD project include the many challenges that face practitioners of building design and construction. Over the last decades, legislative directives and building codes have led to ever-stricter energy demands along with additional requirements for indoor climate and sustainability. This tendency is supplemented by the increasing popularity of holistic assessment schemes [3]. Thus, the design team needs to address a large number of opposing, and steadily tightened, performance criteria which makes most rule-of-thumbs and heuristics obsolete. These circumstances pose a particular challenge during the critical, early design stages in which the frequent and iterative design changes call for immediate feedback on building performance. Moreover, the early design is characterized by a large design freedom and great uncertainties. Building performance simulations (BPS) are commonly used to assess building performance but the deterministic, evaluative nature of most software makes it difficult to explore a multivariate design space and provide proactive and timely guidance. Finally, the complex relationships between design parameters and the diverse, qualitative and quantitative, objectives complicate decision-making in a design process with many stakeholders.

Concurrent with these developments, MOE have had a growing need to address the emerging challenges. Potential solutions, to some of them, have been identified during a preliminary study as described in the following section. A key concept is to shift from evaluative, deterministic simulations to a more proactive and global approach by means of Monte Carlo simulations and sensitivity analysis [4]. This allows for a more comprehensive exploration of the vast design space, which helps reveal high-performing designs that satisfy the diverse requirements. Many obstacles remained, but these initial investigations formed the platform for the PhD study.

The host company has great ambitions within the field of energy, indoor climate, and sustainability of buildings, which the industrial PhD project, and other concurrent developments, should help to achieve. One ambition is to become the architects' preferred partner and consultant. A mean to do so is the extensive investigation of the multivariate design space, which is assumed to reveal numerous diverse solutions and thus provide more "design freedom" for the architect. Moreover, the intention is to provide immediate feedback and proactive guidance. The latter includes the ability to focus on things that matters most and ignore insignificant parameters. A holistic approach is ever-present and essential, since we cannot optimize on one criterion without affecting others. The project includes the development of novel methods, and potentially the required software, to assist during design competitions and in the regular design process. Finally, a desired outcome of the PhD project is to brand the host company as one of the industry leaders in this field of expertise and as a contributor to science.

Special features of the industrial PhD setup involve communication form and intellectual property rights (IPR). Here, the company and university have different preferences. In terms of communication, the university favors knowledge sharing through peer-reviewed journals and conference proceedings. Teaching is appreciated but not mandatory for an industrial PhD. In contrast, the company encourages a more immediate exposure of potential achievements through webpages, social media, and special media. This is preferably supplemented by direct contact with business partners and clients. Regarding IPR, the company controls the rights for developed software, which has affected the choice of platform and distribution.

To round off this background presentation, we discuss some of the implications of carrying out the project in a Danish setting. The Danish architecture has a strong legacy and the design is often complex and experimental making each building a challenging endeavor. Architects, engineers, and other decision makers typically work together in an integrated design process. The architects focus on aesthetics, functions, and logistics; whereas engineers are responsible for ensuring that the design meets the requirements related to energy and indoor climate. Though, the responsibilities for some aspects vary to a greater extent, e.g. the assessment daylight availability and holistic performance. A typical approach is to ensure code compliance of the whole building energy demand using the mandatory software, Be15, which is based on the monthly steady state version described in ISO 13790 [5-6]. Parallel to this, the indoor climate is often evaluated for “critical” or “representative” rooms using dynamic building simulation software (e.g. BSim, IDA-ICE, or IES-VE) [7]. For this project, the performance criteria used in case studies reflect the Danish building code regulations and recommendations. This explains the use of performance metrics such as “overtemperature”, “hours above 26°C”, and daylight factor. However, we conclude by mentioning that the methods are universal. Presumably, they are also applicable to other research disciplines and industries involving multi-actor decision-making and multidimensional domains with large uncertainties.

1.2 PRELIMINARY STUDY

Our first article describes a preliminary study, which has formed the basis for the PhD study. The article introduces the use of Monte Carlo simulations to take into account the large variabilities of inputs in early design [4]. Holistic scoring functions combined with statistical methods provide the means to digest the large number of simulations and to make informed decisions to improve building performance.

1.2.1 PAPER A

The following article, denoted Paper A, is titled “*A stochastic and holistic method to support decision-making in early building design”*. The paper has been published in the Proceedings of Building Simulation 2015, Pages 1885 – 1892, Dec. 2015.

A STOCHASTIC AND HOLISTIC METHOD TO SUPPORT DECISION-MAKING IN EARLY BUILDING DESIGN

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

ABSTRACT

The use of holistic certification tools is increasing and requirements in legislation are continuously being tightened. This calls for a holistic simulation approach in the early design phase where input uncertainties are large and decisions are crucial to the performance. An iterative parametric method is proposed: 1) Assign uniform distributions to uncertain design inputs of interest; 2) Perform sensitivity analysis (SA) by the method of Morris to rank input by relative importance; 3) Run Monte Carlo simulations to explore the entire design domain; 4) Apply Monte Carlo filtering to identify preferable input domains for the most influential parameters.

To enable computationally fast simulations, we combined calculations of energy demand and thermal comfort based on ISO 13790 (CEN 2008) with a regression model for daylight factor. We constructed scoring functions for the three outputs and applied weighting to combine the three scores into a single holistic score ranging from 0 to 100.

The method was tested on a simple office building. An initial run of 3000 simulations was performed using a Quasi-Random LpTau sampling strategy for 22 variable inputs. A filter was applied to the holistic score to collect the 10 % best performing simulations. From this collection, histograms were used to identify favourable and adverse input spans for a selection of the most sensitive parameters. Subsequently, two runs of each 3000 simulations were performed – one using the favourable input spans and the other using the adverse spans. The results showed that the distribution related to favourable input spans was shifted significantly towards higher holistic scores. The authors conclude that the use of a stochastic, holistic method can guide decision-making by identifying favourable input regions, and thereby increase the remaining solution space and overall building performance.

INTRODUCTION

The building design community is challenged by continuously increasing energy demands, which are often combined with ambitious goals for the indoor environment. By 2020, all new buildings are required to by "nearly zero energy" buildings in the European Union (European Parliament 2010). In Denmark, the authorities are gradually tightening the energy requirements by reducing energy demands by 25 % in 2015 and again in 2020 (Energistyrelsen 2010). Concurrently, thermal comfort in dwellings must be assessed in terms of overheating hours to achieve 2015 classification, and daylight demands are being sharpened by increasing the daylight factor threshold in order to reach 2020 classification. To meet the ever-stricter demands and create high performing buildings, we suggest a holistic approach and exploration of a vast design space.

The three objectives; energy demand, thermal comfort, and daylight, receive much attention due to the legislative requirements and their strong interdependencies. These objectives are especially difficult to address since improving one of them often worsens another. At the same time, the design team must address many less quantifiable objectives such as logistics, aesthetics, and function. Above all is the budget, which is perhaps the most important design constraint. Therefore, a holistic approach is crucial in the multi-collaborator design process, where decision-making involves building owners, architects, engineers, and contractors.

Engineers typically rely on deterministic building simulations to evaluate design options, though the software gives little or no guidance on how to improve the design. Most detailed simulation software is used to evaluate design options to ensure building code compliance and has not been developed to take into account the large uncertainties and rapid change of design which are characteristic of early design (Petersen 2011). Instead, we suggest exploration of a global design space by stochastic methods while considering multiple objectives. This approach creates pro-active information that supports the design team in the decision-making process. Moreover, the approach gives more room for decisions across disciplines and performance objectives.

Literature shows many uses of stochastic building simulations combined with uncertainty analysis and sensitivity analysis (Tian 2013). However, emphasis is often on uncertainties associated with user behaviour, weather scenarios, and physical properties of materials (de Wit & Augenbroe 2002)(Hopfe &

Hensen 2011)(Struck et al. 2009). An alternative approach is to focus on design variability, where possible ranges in design parameters are treated as uncertainties by applying uniform probability distributions (Yildiz et al. 2012). Similar approach was taken by Heiselberg et al. (2009), who adopted the method of Morris to perform sensitivity analysis during early design to identify inputs that have the largest impact on energy consumption and as a consequence these inputs deserved most attention (Heiselberg et al. 2009). In this work, we expand this method by adding thermal comfort and daylight to the objectives of interest. Moreover, we apply Monte Carlo filtering techniques to identify favourable design input spans. The proposed method is demonstrated for an office building.

METHODOLOGY

The goal of our research is to develop a method that can be used in the iterative design process to guide the design team in creating high performing buildings. Based on the architect's design proposal, the engineer builds a simulation model and performs stochastic calculations of energy demand, thermal comfort, and daylight. The simulation results are combined into a holistic score and analysed using uncertainty analysis and sensitivity analysis. The design team is informed about best and worst case scenarios and favourable spans for the most influential design variables. The process may be repeated to decrease variability or the design may continue to the detailed design stage.

The development of the methodology involves the following tasks: i) creating a parametric simulation model; ii) creating holistic scoring functions; iii) applying suitable methods for uncertainty and sensitivity analysis. These efforts are described below.

Idealized simulation model

To develop, test, and evaluate the proposed method, we needed to construct a simulation model satisfying the following properties:

- Assignment of probability density functions to inputs
- Execution of Monte Carlo simulations
- Evaluation of whole building energy demand, thermal comfort, and daylight factor for selected zones
- Calculation detail level appropriate for early design

In addition to these requirements, emphasis was on calculation speed in order to do thousands of simulations in minutes rather than hours or days.

To calculate energy demand we chose the normative model Be10 used for building code compliance in Denmark (SBI - Danish Building Research Institute 2014). The model is based on the simplified quasi-steady-state monthly method provided by ISO 13790 (CEN 2008). An advantage of this model is computational speed, which is measured in milliseconds.

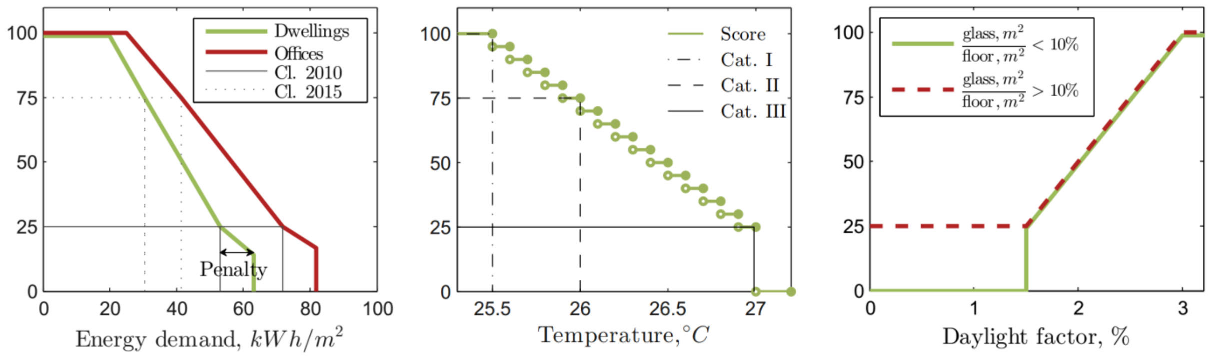
To assess thermal comfort, we use an hourly-based idealized model "Summer Comfort", which is also developed by the Danish Building Research Institute on the basis of ISO 13790. The model was developed to evaluate operative temperature in the critical rooms of dwellings by calculating the number of hours above 26 °C and 27 °C, respectively.

Daylight simulations are often computational heavy compared to energy and thermal calculations. For this study, we applied a simplified regression model based on Danish guidelines made for building code compliance (Johnsen & Christoffersen 2008). In this, daylight factor is estimated for rectangular rooms using pre-calculated line curves and correcting these to take into account shading effects, glazing properties, room reflectance, etc.

To sum up, we have put together a fast, idealized simulation model based on Danish building code and practice. The required input for this model fits the level of detail in early building design. It enables fast computation of some of the most correlated quantifiable output; energy consumption, thermal comfort and daylight. We omitted other measures such as embodied energy and acoustics, since they relate more to materials which are typically specified at later design stages.

Creating holistic scoring functions

A way to encourage holistic design is to construct scoring functions for each output and combining these scores into one overall, holistic score. This approach is seen in several sustainability assessment methods including LEED, BREEAM, and DGNB. A holistic score eases comparison when comparing a large number of design options. Moreover, it helps interpretation of sensitivity analysis and allows more consistent filtering of Monte Carlo simulations. In this work, we apply scoring function to energy, thermal comfort, and daylight such that each of these objectives is evaluated in the range 0 – 100 points. Afterwards, a combined scored is constructed using a user-defined weighing system.



a) Score Energy b) Score Thermal Comfort c) Score Daylight

Figure 1 Scoring functions for energy demand, thermal comfort, and daylight.

Energy
In general, Danish building code differentiates between dwellings, hotels, et al. and offices, schools, et al. A function for each category is constructed as shown on Figure 1 (a). Energy demands lower than 2020 requirements results in 100 points whereas demands equal to 2010 requirements results in 25 points. A penalty function allows for a slight overconsumption, which may be compensated for by adding renewables such as photovoltaics.

Thermal comfort (summer)
The scoring function for thermal comfort is based on EN 15251 (CEN 2007) and shown on Figure 1 (b). This standard divides indoor climate into four categories: I, II, III, and "out of category". The upper limits for room operative temperature in classes I to III are 25.5, 26.0, and 27.0 °C, respectively. However, an excess of these thresholds may be allowed for either 3 % or 5 % of the time.

Daylight
Danish building regulations offer two ways to evaluate daylight in workrooms, occupiable room, and similar. One option is to ensure that the glass-to-floor ratio is at least 10 % while adjusting for the light transmittance value of the glazing. The other option is to calculate the daylight factor at workplaces, which must be at least 2 %. To meet 2020 requirements the daylight factor must be 3 %. Using these requirements, we established a scoring function that depends on both daylight factor and glass-to-floor ratio as shown on Figure 1 (c).

Finally, the three scores are combined into a single, holistic score, *EDT*, by assigning user-defined weighting factors. In the case study below, we use weighting factors of 50, 25, and 25 % for energy, daylight, and thermal comfort, respectively. Similar approach and weighting is used by Bjørn and Brohus (2006) when combining energy use, atmospheric and thermal comfort into one score called "Eco-factor". Moreover, the chosen weightings are similar to the ratios between the maximum score for energy, thermal comfort (summer), and visual comfort in the Danish DGNB assessment system. An additional property is that the holistic score becomes zero if either one of the scoring functions is zero. This prevents possible high holistic scores at the expense of a single objective.

Uncertainty and sensitivity analysis

Uncertainty and sensitivity analysis play an essential role in turning data from a large number of simulations into design information that support decision-making. In the proposed methodology, we first apply the method of Morris to rank variable input in accordance to sensitivity. Secondly, we perform thousands of Monte Carlo simulations and apply Monte Carlo Filtering to assess the outcome. The two methods may be performed individually or in combination depending on the scope of the analysis. Both methods consist of the steps described in Figure 2.

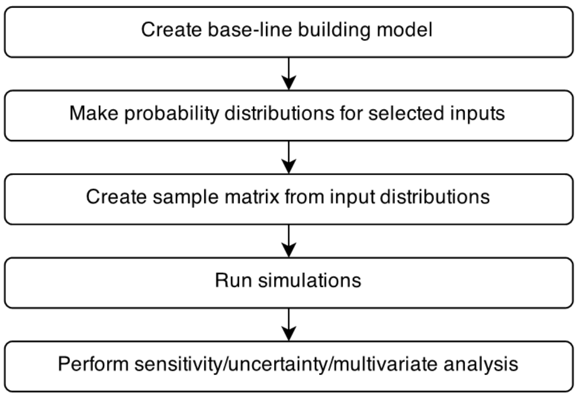


Figure 2 Workflow for automated building simulations.

In this work, emphasis is on early building design and investigation of design space. Hence, variable inputs are described using uniform probability distributions. For example, insulation thickness may vary from minimum 150 mm to 300 mm with even probability since the designer may choose this value. This variability analysis is in contrast to regular uncertainty and robustness analysis, where inputs related to user behaviour, weather, and physical inaccuracies, often follow a normal or log-normal distribution.

Sensitivity using method of Morris

The method of Morris is a computational effective way to screen a large number of inputs in order to find those, which show: negligible, linear and additive, or nonlinear or interaction effects (Morris 1991). Morris (1991) introduces the concept of elementary effect *EE* of a model $Y(X_1, \ldots, X_k)$ with *k* inputs. The *k*-dimensional input space is discretized into *p* levels by splitting their values into *p* quantiles. Then the elementary effect for the i^{th} input factor in a point **X** is defined as (Saltelli et al. 2008):

$$EE_i = \frac{\begin{bmatrix} Y\left(X_1, X_2, \ldots, X_{i-1}, X_i + \Delta, \ldots, X_k\right) \\ -Y\left(X_1, X_2, \ldots, X_k\right) \end{bmatrix}}{\Delta} \quad (1)$$

where $\Delta \in [1/(p-1), \ldots, 1-1/(p-1)]$. Distributions of elementary effects are sampled from the global, discretized input space by following so-called trajectories where only one factor is changed at-a-time. For each input *i*, we calculate the absolute mean and standard deviation of these distributions as follows:

$$\mu_i^* = \frac{1}{r}\sum_{j=1}^{r}\left|EE_i^j\right| \quad (2)$$

$$\sigma_i^2 = \frac{1}{r}\sum_{j=1}^{r}\left(EE_i^j\right)^2 \quad (3)$$

where *r* is the number of samples. The mean μ_i^* indicates the overall influence of the i^{th} input on the output *Y*. Thus, sensitive inputs have relatively high values of μ_i^* whereas negligible inputs have low values. If the standard deviation is large compared to the mean, then the computation of *EE* is strongly affected by choice of sample point at which it is computed. That means this input depends on the values of other inputs, or the input has non-linear relation with the output *Y*.

Monte Carlo simulations

Following the screening exercise above, we want to explore the global design space in depth by performing thousands of Monte Carlo simulations. When doing so, we may use the same design inputs as before or narrow down the number of variables by omitting inputs with little or no influence on the output as indicated by the method of Morris.

The sampling strategy is to investigate the largest possible design space. If design variables are divided into discrete values with equal probability, all combinations could be investigated when using full factorial sampling. Unfortunately, the required simulations grow exponentially with the number of design variables. Instead, we may choose between different sampling techniques to reduce the number of simulations by investigating a subspace, which is still representative of the entire design space. These methods include amongst others; fractional factorial sampling, Latin hybercube sampling, stratified random sampling and quasi-random sampling using low-discrepancy sequences (Saltelli et al. 2008). In this work, we use the low-discrepancy sequence LP_τ since generation of quasi-random numbers are independent of the number of variables. And more importantly, it is possible to increase statistical convergence when compared to randomly generated numbers (Sobol' & Shukman 1993). From the quasi-random numbers and the uniform probability distributions, we construct an input matrix and run simulations.

From the Monte Carlo experiment, we yield thousands of input-output relationships to be analysed using various statistical techniques. Scatterplots can reveal both linear and non-linear correlations including the strength of the correlations. Boxplots show minimum and maximum values, which correspond to best-case and worst-case scenarios of the current design. In addition, quantitative sensitivity measures can be calculated using Pearson's product-moment correlation or Spearman's rank correlation (Joint Research Centre n.d.). Finally, we can apply Monte Carlo filtering to achieve valuable design information as stressed out in the following.

Monte Carlo Filtering

A key element of the proposed methodology is to apply Monte Carlo Filtering to identify regions of the design space, which are more likely to produce acceptable results. Furthermore, it helps showing the effect of constraints for various objectives, such as energy demand, daylight, and thermal comfort. For a model $Y = (\mathbf{X})$ with *k* variables, such that $\mathbf{X} = (X_1, X_2, \ldots, X_k)$, we split the output into two subspaces referred to herein as B and $\bar{B}$ corresponding to behavioural and non-behavioural results, respectively (Saltelli et al. 2008). Likewise, each input X_i domain is divided into two subspaces, $(X_i \mid B)$ and $(X_i \mid \bar{B})$, depending on whether they produce behavioural or non-behavioural output. The benefits of applying Monte Carlo Filtering will become apparent in the case study below.

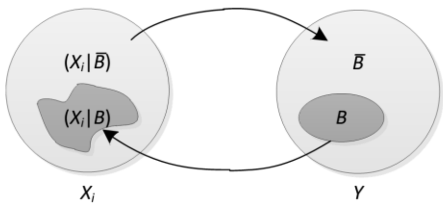


Figure 3 Monte Carlo filtering of input X_i and output Y into behavioural and non-behavioural subspaces.

CASE STUDY

Office building description

The proposed methodology is tested on a simple, office building to illustrate how the method may be applied in praxis. Furthermore, we show how the gained knowledge makes it plausible to create better performing building design. The three stories office building with basement has a rectangular shape measuring 60 x 15 x 9.9 m. Rectangular offices measuring 4 x 6,25 m are situated along the elongated facades oriented towards south and north, respectively. Heat loads from persons and equipment are uniformly distributed with values 4 and 6 W/m² according to Danish building code.

Benefits of Monte Carlo Filtering

First, we demonstrate the effect of adding filters for different objectives in relation to the distribution of "behavioural" simulations. Using the office building above, we run 3000 simulation with 18 uncertain inputs with uniform probability distributions. Figure 4 shows distributions of behavioural simulations related to a highly sensitive input – the size of the windows' overhang. Without filters, we see an even distribution, which is due to the uniform probability distribution used when sampling. First, we remove simulations not meeting Danish energy requirements. Approximately one third of simulations disappear but the behavioural simulations are still distributed evenly with respect to windows' overhang. Secondly, we remove simulations not meeting the thermal requirement of maximum 100 h above 26 °C. This has large impact on the distribution, which shows that the larger the overhang, the better. Finally, we filter out the simulations with daylight factors below 2 %. Since overhang reduces daylight, many of the simulations with large overhangs are removed. In this example, we conclude that the overhang must be in the range 5 – 43° and preferably more than 20°. Generally, the example shows the importance of a holistic design approach, in which we consider interdependent objectives simultaneously.

Another benefit from Monte Carlo Filtering becomes apparent when analysing relationships between mutual interdependent and sensitive inputs. To inspect such a relationship we use a scatterplot as shown on Figure 5.

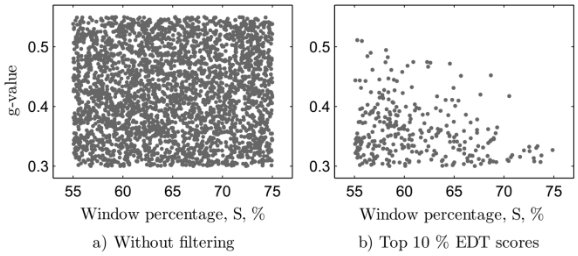


Figure 5 Input-input scatterplots with and without filtering.

In this case, we used filtering to select the 10 % best performing results using the holistic score, *EDT*, defined above. Without filtering, we see that the points are evenly distributed in the two-dimensional space. After applying the filter, we see strong dependency between g-value and window size as expected. The plot shows how much we can expect to reduce the g-value when increasing window size. If the designer chooses a combination of low g-value and small window percentage, there are many options to choose from, which mean the designer has lot of freedom to vary other uncertain input as well. The "outliers" along the boundary are also feasible design options but choosing such outliers will put constraints on other inputs, e.g. overhang and solar shading.

Combining SA and Monte Carlo Filtering

In this example, we demonstrate how Morris analysis followed by Monte Carlo Filtering improves building performance. After setting up the baseline model for the office, we choose 22 design parameters to which we assign uniform probability distributions as shown in Table 1. Relatively wide spans are used in order to investigate a very large design space. The 22 parameters are divided into three categories: building form, built quality, and technical systems.

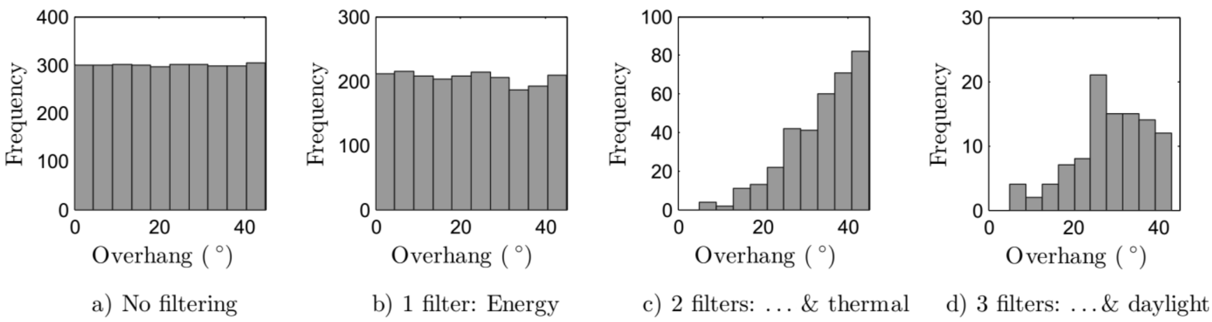


Figure 4 Histograms for behavioural simulations when gradually applying filters.

Table 1 Input spans for 22 design parameters.

Parameter		Span Min	Span Max	Unit
11	Window-%, N	25	75	%
12	Window-%, S	25	75	%
13	Overhang	0	45	°
14	Side fins	0	30	°
15	Window opening	0	10	%
16	Solar shading, Fc	0.2	1	-
17	Mean reflectance	0.4	0.6	-
18	Wall thickness	0.4	0.6	m
21	Heat capacity	60	140	Wh/Km²
22	g-value	0.3	0.7	-
23	U-value, windows	1.0	1.6	W/m²K
24	U-value, walls	0.10	0.15	W/m²K
25	U-value, terrain	0.08	0.13	W/m²K
26	U-value, roof	0.08	0.18	W/m²K
27	Lin. heat loss, base	0.1	0.4	W/mK
28	Lin. heat loss, windows	0.02	0.06	W/mK
31	Mech. ventilation, qm,s	0.9	3.6	l/s m²
32	Venting, qn,n	0.9	1.2	l/s m²
33	Heat recovery, η	0.70	0.95	-
34	Specific fan power	1.5	2.1	kJ/m³
35	Lighting, stand-by	0	1	W/m²
36	Lighting, installed	4	10	W/m²

Sensitivity using method of Morris

First, the method of Morris is used to rank the parameters in relation to their sensitivity and identify parameters that are non-linear or highly correlated with others.

As input for the method of Morris we used $r = 10$ samples, $p = 8$ levels and $M = 100$ possible trajectories. The method is applied for the holistic score *EDT* as shown on Figure 6. The most influential parameters are: g-value, window percentage (south), mean reflectance, window percentage (north), overhang, and heat capacity. Addtionally, the plot reveals non-linear or correlated behavior of the parameters close to the standard-error-of-mean line: g-value, windows percentages, and solar shading. In

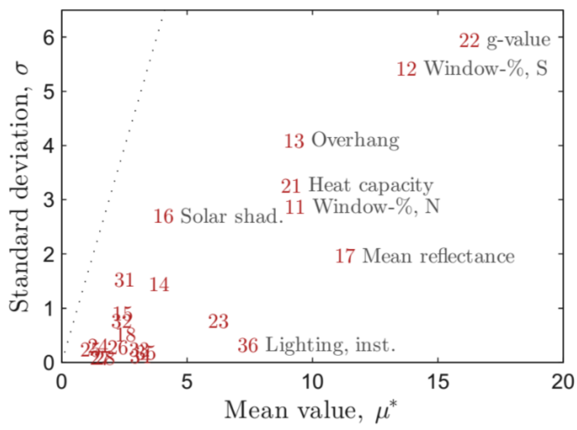


Figure 6 Estimated means and standard deviations of the distributions of EE's in relation to the EDT-score. Line corresponds to standard-error-of-mean.

contrast, the mean reflectance is the third most sensitive parameter but shows little correlation or non-linearity, which fits well with the algorithms in use.

Monte Carlo simulations

Following the Morris analysis, we perform 2500 Monte Carlo simulations using the same variable inputs and spans. Alternatively, we could have excluded the parameters showing little or no influence according to the sensitivity analysis. Since we are interested in initiatives to improve the overall performance of the design, we apply Monte Carlo Filtering to select the 10 % highest holistic scores. Hereafter, we create a histogram for each variable input to see how the 10 % best scores are distributed in relation to the different inputs. In Table 2 the parameters are ranked according to their sensitivity. For about 5 – 7 of the most sensitive the histograms show tendencies, from which we can make recommendations in terms of favourable input spans and non-favourable spans. Recommended spans are listed to the right in Table 2.

As proof of concept, we performed two additional runs of each 3000 Monte Carlo simulations – one using the recommended input spans and the other using the adverse spans. Initial spans were maintained for inputs where no recommendations were made.

Table 2 Histograms for top 10 % best performing simulations along with recommended spans. Parameters are ranked according to sensitivity.

Parameter	Initial spans Min	Initial spans Max	Hist.	Recommended Min	Recommended Max
g-value	0.3	0.7		–	–
Window-%, S	25	75		33	60
Mean reflectance	0.4	0.6		0.5	0.6
Window-%, N	25	75		30	65
Overhang	0	45		–	–
Heat capacity	60	140		85	140
Lighting, inst.	4	10		–	–
U-value, windows	1.0	1.6		–	–
Solar shading, Fc	0.2	1		0.2	0.6
Side fins	0	30		–	–
Lighting, stand-by	0	1		–	–
Specific fan power	1.5	2.1		–	–
Heat recovery, η	0.7	0.95		–	–
Wall thickness	0.4	0.6		–	–
Mech. ventilation	0.9	3.6		–	–
Window opening	0	10		–	–
Venting, qn, n	0.9	1.2		–	–
U-value, roof	0.08	0.13		–	–
Lin. heat loss, win.	0.02	0.06		–	–
Lin. heat loss, base	0.1	0.4		–	–
U-value, walls	0.1	1.5		–	–
Lighting, installed	0.08	0.13		–	–

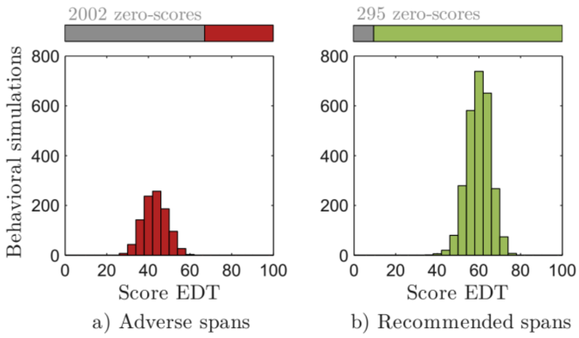


Figure 7 Distributions of holistic scores when using adverse spans (left) and recommended spans (right).

Distributions of the resulting holistic scores for each of the two runs are shown on Figure 7. Comparing the two runs, we observe a clear shift towards higher scores when using the recommended spans. The zero score means that at least one of the requirements is not met. The number of simulations resulting in a zero score is reduced from more than 2000 to less than 300 simulations, which means the remaining design space will be increased significantly when following the design recommendations. Since we still observe zero scores for the recommended spans, we cannot ensure that requirements will be met. For example, the EDT score may result in zero if the design team use the "worst" combinations of the remaining spans such as high U-values combined with poor ventilation performance.

Figure 8 shows scatterplots and histograms for the 10% best performing simulations in relations to the three most sensitive inputs and the least sensitive input. In addition to showing distributions of the best simulations, these small multiples of plots help to identify direction, form and strength of important input-output relationships. As expected, we observe relatively strong positive relation between daylight and the two inputs: reflectance and window percentage. For energy consumption, there are moderate linear relationships with g-value and window percentage where large g-values and low window percentages reduces energy demand. However, some of the highest holistic scores are seen at lower g-values and relatively high window percentages. For the least sensitive input, U_{terrain}, we observe neither direction nor form.

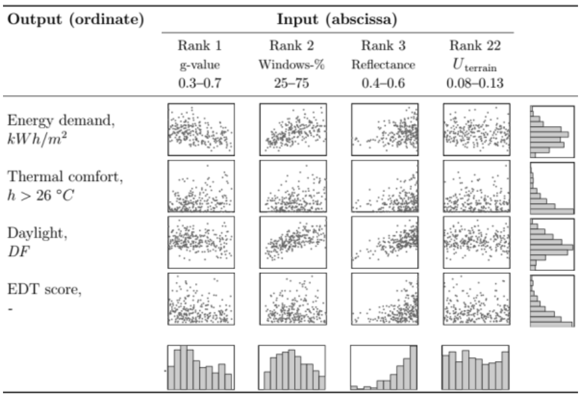


Figure 8 Scatterplots and histograms for the 10 % highest EDT-score related to the three most sensitive inputs and the least sensitive input.

CONCLUSION

This study showed how uncertainty analysis and sensitivity analysis could be applied to support decision-making in early building design. Emphasis was on addressing correlated objectives simultaneously – in this case energy demand, thermal comfort, and daylight. A holistic score was shown to ease comparison of designs and rank inputs after sensitivity on overall performance. The method of Morris proved useful to screen a large number of uncertain inputs to focus on the most influential ones and reveal possible correlations and non-linearity. Monte Carlo simulations made it possible to investigate a large design space defined by inputs with uniform probability distributions. Such distributions works well with subsequent Monte Carlo Filtering making it possible to identify favourable and adverse input spans for important inputs. In conclusion, the methodology make the use of building simulations more pro-active compared to the widespread evaluative use, that give little or no guidance on how to improve the building design. Practical implications of the proposed methodology include: (a) continuously identification of design inputs that matters most, (b) awareness of interdependent inputs, (c) global investigation of a large design space to achieve higher performing designs, (d) holistic approach to ease comparison that ensures well-balanced design, (e) handling of design uncertainty and variability.

Limitations and further research

Often the most important design objective is building costs, which was omitted in this study. Though, estimating cost when doing stochastic simulations may be an impossible task, since there are no unique way to calculate cost of variable overhangs, shading systems, glass quality, etc. Instead, the proposed method help identify feasible regions in design space, after which the design team use experience to determine what is feasible when considering cost, aesthetics, construction, etc.

Using weighting factors to combine outputs into a single score emphasizes holistic design but the different choices of weighting factors may lead to different conclusions. As an alternative to fixed weighting factors, the design team may apply and vary filters values for each objective independently. Again, this will split the simulations into behavioural and non- behavioural regions from which favourable inputs spans can be identified. The filter values may be varied until a suitable number of behavioural inputs are obtained or until the desired level of performance is reached. When seeing the consequences of the filtering, the design team may want to change the initial requirements (filter values), e.g. seeking higher performance for one objective while accepting a slightly lower performance for another.

For this study, we used idealised models where speed and level of detail where suitable for comparison of stochastic simulations in early building design. Further research is needed to incorporate advanced simulation model to improve validity of calculations and allow for analysis of advanced systems, fenestration, etc. Such effort would presumably require the use of computer clusters or cloud computing to account for the vast increase in computational effort.

ACKNOWLEDGEMENTS

Innovation Fund Denmark and MOE A/S provided funding. The work was part of an industrial doctorate program with Aalborg University and consultancy company MOE A/S.

REFERENCES

Bjørn, E. & Brohus, H., 2006. Overall evaluation of indoor climate and energy for alternative office designs using the Eco-factor. Journal of civil engineering and management, XII, pp.43–49.

CEN, 2007. EN 15251 Indoor Environmental Input Parameters for Design and Assessment of Energy Performance of Buildings -- Addressing Indoor air Quality, Thermal Environment, Lighting and Acoustics.

CEN, 2008. ISO 13790:2008 Energy performance of buildings -- Calculation of energy use for space heating and cooling, Geneva, Switzerland.

Energistyrelsen, 2010. Danish Building Regulations 2010 - BR10. www.bygningsreglementet.dk.

European Parliament, 2010. Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings. Official Journal of the European Union, (L 153), pp.13–35.

Heiselberg, P. et al., 2009. Application of sensitivity analysis in design of sustainable buildings. Renewable Energy, 34(9), pp.2030–2036.

Hopfe, C.J. & Hensen, J.L.M., 2011. Uncertainty analysis in building performance simulation for design support. Energy and Buildings, 43(10), pp.2798–2805.

Johnsen, K. & Christoffersen, J., 2008. SBi-anvisning 219 Dagslys i rum og bygninger,

Joint Research Centre - Econometrics and Applied Statistics Unit, Simlab 2.2: Reference Manual

Morris, M., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2), pp.161–174.

Petersen, S., 2011. Simulation-based support for integrated design of new low-energy office buildings. Technical University of Denmark.

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

SBI - Danish Building Research Institute, 2014. SBi anvisning 213 - Bygningers energibehov,

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

Struck, C., Hensen, J. & Kotek, P., 2009. On the Application of Uncertainty and Sensitivity Analysis with Abstract Building Performance Simulation Tools. Journal of Building Physics, 33(1), pp.5–27.

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

De Wit, S. & Augenbroe, G., 2002. Analysis of uncertainty in building design evaluations and its implications. Energy and Buildings, 34(9), pp.951–958.

Yildiz, Y. et al., 2012. An approach for developing sensitive design parameter guidelines to reduce the energy requirements of low-rise apartment buildings. Applied Energy, 93, pp.337–347.

1.2.2 LIMITATIONS

Paper A has demonstrated the necessity of a holistic approach and the potential of Monte Carlo simulations combined with sensitivity analysis [4]. To follow up, we give some thoughts to possible limitations and alternative ways for communication and visualization.

One disadvantage of the holistic approach in Paper A is the use of weights to combine performance measures into a combined holistic score. Defining such arbitrary weights is tricky since decision-makers often have different preferences. This weighting system is problematic because the weights influence both the ranking from sensitivity analysis and the distributions of “behavioral” simulations. Ultimately, changing the weights alters the recommendations. In addition, the proposed filtering approach, which was demonstrated with a “top 10%” threshold, does not reflect the actual criteria. Such issues may be remedied by the ability to apply user-defined, and project-specific, filter criteria. The method would also benefit from the ability to apply filters to model inputs to investigate specific regions of input space and observe the consequences of such constraints. In conclusion, these lessons call for flexible, interactive ways to analyze and explore the global design space. Moreover, we would appreciate a sensitivity analysis technique that works for multiple outputs but does not depend on a weighting system.

Based on the preliminary study, we broaden our perspective by performing an extensive literature in the next chapter.

2 LITERATURE REVIEW

2.1 SCOPE AND MOTIVATION

In the previous chapter, we presented various challenges of building design as experienced by MOE engineers. The preliminary study showed potential ways to address some of the issues related to building performance. Throughout this PhD project, the use of building performance simulations (BPS) is assumed a prerequisite for aiding decision-makers in the design of buildings with high performance. The following article, denoted Paper B, contains a literature review of the use of BPS with emphasis on implementation during the early design stages. The motivation is to identify current state of knowledge and research areas that relate to challenges hindering the adoption of building simulations in early design [8]. The review contains a proposal for an “ideal” simulation framework based on identified knowledge gaps. Finally, it provides an overview of BPS software in the perspective of the proposed framework.

Following Paper B, we describe the research areas to be addressed in the subsequent studies in the PhD project. In addition, we explain how these fit into concurrent efforts made in the host company and at Aalborg University. We round off with a discussion of the BPS software to be used in this project.

2.1.1 PAPER B

Paper B refers to the review article titled “*Building simulations supporting decision making in early design – A review*”, which has been published in Renewable and Sustainable Energy Reviews, Volume 61, Pages 187 – 201, 2016.

Renewable and Sustainable Energy Reviews 61 (2016) 187–201

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Building simulations supporting decision making in early design – A review

Torben Østergård [a,b,*], Rasmus L. Jensen [a], Steffen E. Maagaard [b]

[a] Aalborg University, Department of Civil Engineering, Sofiendalsvej 9-11, DK-9200 Aalborg SV, Denmark
[b] MOE A/S, Åboulevarden 22, DK-8000 Aarhus, Denmark

ARTICLE INFO

Article history:
Received 7 July 2015
Accepted 13 March 2016

Keywords:
Building performance
Uncertainty analysis
Sensitivity analysis
Interoperability
Optimisation
Knowledge based input generation

ABSTRACT

The building design community is challenged by continuously increasing energy demands, which are often combined with ambitious goals for indoor environment, for environmental impact, and for building costs. To aid decision-making, building simulation is widely used in the late design stages, but its application is still limited in the early stages in which design decisions have a major impact on final building performance and costs. The early integration of simulation software faces several challenges, which include time-consuming modeling, rapid change of the design, conflicting requirements, input uncertainties, and large design variability. In addition, building design is a multi-collaborator discipline, where design decisions are influenced by architects, engineers, contractors, and building owners. This review covers developments in both academia and in commercial software industry that target these challenges. Identified research areas include statistical methods, optimisation, proactive simulations, knowledge based input generation, and interoperability between CAD-software and building performance software. Based on promising developments in literature, we propose a simulation framework that facilitates proactive, intelligent, and experience based building simulation which aid decision making in early design. To find software candidates accommodating this framework, we compare existing software with regard to intended usage, interoperability, complexity, objectives, and ability to perform various parametric simulations.

Contents

Abbreviations: BPS, Building performance simulation; OAT, one-at-a-time; UA, uncertainty analysis; SA, sensitivity analysis; LCC, life cycle costs; LCA, life cycle analysis

* Corresponding author at: Aalborg University, Department of Civil Engineering, Sofiendalsvej 9-11, DK-9200 Aalborg SV, Denmark. Tel.: +45 2540 0325.
E-mail address: to@civil.aau.dk (T. Østergård).

http://dx.doi.org/10.1016/j.rser.2016.03.045

1. Introduction

The building design community is challenged by continuously increasing energy demands which are often in conjunction with ambitious goals for the indoor environment. The recast of the European Performance of Buildings Directive (EPBD) requires all new buildings in the European Union to be "nearly zero energy" buildings by 2020 [1]. In addition to stricter energy demands, the use of environmental assessment methods has increased considerably [2,3]. As a result, the design team must try to optimize on a large number of criteria, such as energy demand, indoor environment, materials, life cycle cost, etc., which are often conflicting. Supporting decision making and guiding the design towards high performance is of utmost importance in the early design phase where decisions have the highest impact on final performance and costs [4–6][1]. Predicting the consequences of early decisions is particularly difficult, but crucial, since adverse decisions will reduce the remaining design space and make it more strenuous and expensive to meet high performance goals. For example, the design team may early on decide on a design concept with a highly transparent facade favoring daylight (high window-to-wall-ratio) where potential issues, concerning cooling energy, thermal comfort, and glare, are avoided by a combination of hybrid ventilation and automatic, external shading. If the initial conditions later turn out to be too optimistic or unrealistic (e.g. solar shading in use more that 80% of the time, venting needs an air change of more than 10 h^{-1} to keep temperatures within limits), it will have major impact on both cost and design to remedy this early decision and reach ambitious goals. Despite the potential of performing building simulations, the information obtained from building performance simulation software is often evaluative instead of proactive [5,7]. Even when the software is sophisticated, accurate, and capable of assessing a wide range of different performance indicators, it is often most suitable for code compliance, benchmarking, and quality control. There is a lack of tools that provide timely feedback on performance implications and help compare and rank multiple design variations [8,5]. The software's ability to provide this kind of active support is sometimes referred to as "intelligence" [5,9]. In a survey among 230 architects, "intelligence" and "usability" ranked higher than "interoperability" and "accuracy" when selecting BPS tools [5]. In other words, the software's ability to inform and guide the design has the highest priority by the majority of the architects. According to Batueva and Mahdavi [9], less than 8 percent of more than 400 building simulations tools listed by the U.S. Department of Energy [10] have potential for early design deployment potential. In summary, challenges of performing building simulations at the early stages, identified by the authors, include: lack of information, uncertainty, vast design space, increasing levels of model resolution (level of detail), time-consuming modeling, and rapid change of design. In general, challenges affecting all stages of building design include: contradicting and stricter requirements, interoperability, limited reuse of knowledge, discrepancy between simulations and real-life measurements, and lack of simulation guidance.

The main focus of this review is to identify state-of-the-art within the field of building simulations addressing the challenges above. The review is part of a research project which aims to develop a simulation framework that addresses all of these diverse challenges in order to facilitate proactive, intelligent, and experience based building simulations. Another ambition of the research is to implement such a framework in the design project as early as possible. Below, we outline six research areas targeting at least one of the identified challenges, and we specify how this review differs from previous reviews related to building simulations. In chapter 2, we describe how each of the six research areas approaches the issues of BPS, and we highlight promising and trending methods. In chapter 3, we propose an ideal framework for building performance simulations based on our findings in chapter 2. In continuation of this, we carry out a software review in search for available software that fits the requirements and properties of this "ideal" framework.

In this paper, attention is drawn to developments facilitating improved assistance and guidance for the design team during the early design stages. Particular interest is given to methods that enable the designer to investigate a global design space, which is expanded from the variability of multiple design parameters. The reason for this is that a single building performance simulation only evaluates a single point in the design space without taking uncertainties and variability into account. Nor does the single evaluation guide the designer on how to improve the design. As a consequence, designers often perform manual or automatic, parametric simulations varying one parameter at a time. This one-at-a-time approach (OAT) is referred to as local analysis. In early design, many parameters may be varied at the same time which advocates exploration of a global design space, which presumable can reveal higher performing design as illustrated on Fig. 1.

1.1. Research targeting early building simulations

This review covers a wide range of research addressing the challenges related to building simulations as identified above. To create an overview, the reviewed papers have been organized into six larger groups of research areas – each of them targeting one or more specific challenges as illustrated on Fig. 2. Definitions of these intertwined research areas, and motive for their inclusion, are as follows:

- Proactive building simulations refer to a proactive exploration of the design space in order to guide the design rather than evaluate design.

[1] In this review, we distinguish between early design and detailed design. In addition, the early design stage may be split into two phases: conceptual design, in which the building concept is developed and schematic drawings are produced; and preliminary design, where schematic drawings are refined to estimate the main quantities for the building project (adopted from [11]).

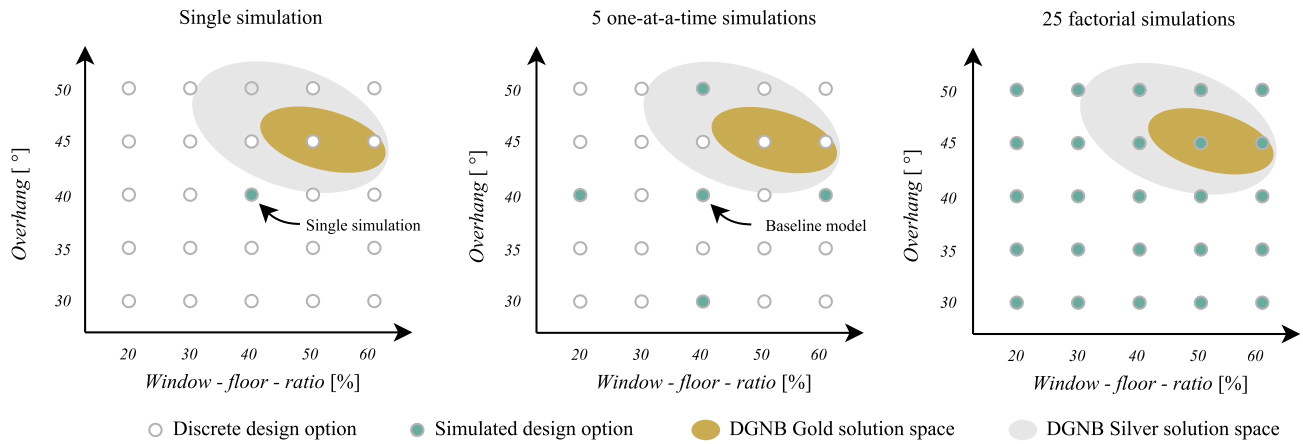


Fig. 1. Different explorations of a 2D discretized design space. Favorable solution spaces are illustrated by simulations resulting in best (Gold) and second best (Silver) awards according to the DGNB certification system for sustainable buildings [11] (similar to e.g. LEED Platinum and Gold certification [12]).

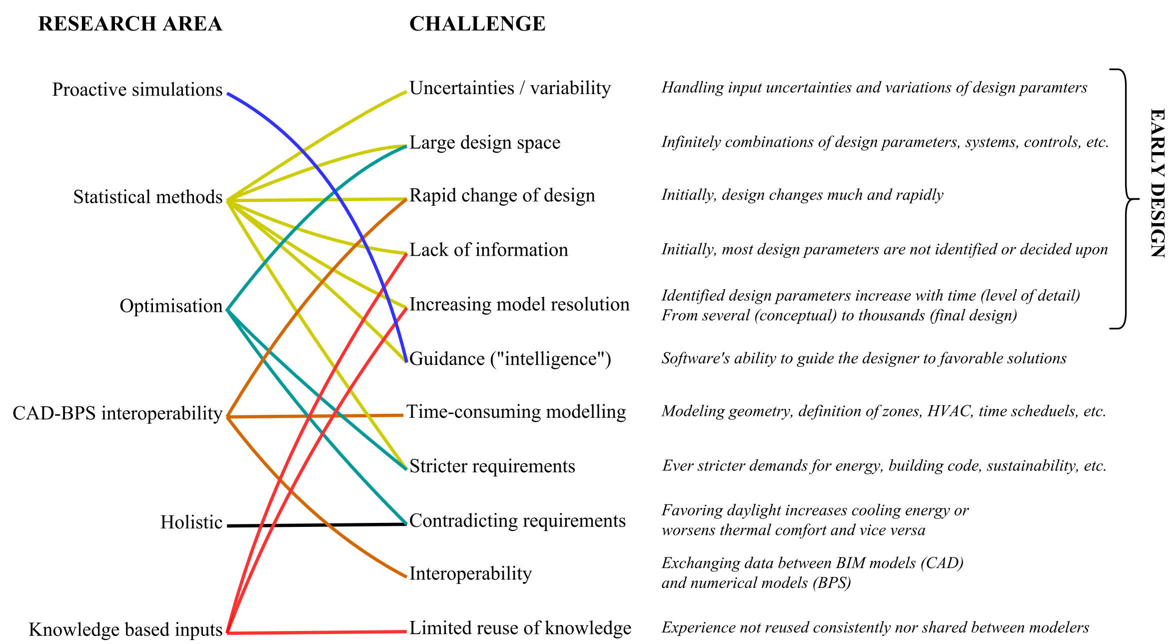


Fig. 2. The reviewed research areas and their relation to different challenges of performing building simulations in the early design stages and in general.

- Statistical methods include running large numbers of simulations and applying statistical measures. As well as coping with uncertainties, a statistical approach may facilitate exploration of a large design space and identify important inputs and favorable input domains.
- Holistic design includes calculation of many interdependent performance objectives and combining the results to support decision making. Examples of important interdependent objectives are energy demand, thermal comfort, and daylight.
- Optimisation on performance objectives helps to automate the exploration of a large design space and guide the design towards high performance.
- CAD-BPS interoperability may be achieved by integration of models, run-time coupling, and shared schemas. A common ambition is to ensure fast and consistent modeling.
- Knowledge based methods aim to reuse and share knowledge to reduce the time spent modeling, and they seek to improve consistency and validity. Moreover, knowledge databases may be used to set default values to enable simulations when the input resolution is low (model detail).

Since the main focus of the review is simulations made in the early design phases, we will not cover efforts in improved algorithms describing building physics. Nor will we address methods which primarily intend to improve detailed analysis or reduce performance gaps.

1.2. Literature reviews and comparative surveys

Prior to this study, we found a considerable amount of comparative studies and reviews concerning building performance simulations. These studies provide a comprehensive insight into a specific discipline or branch of building design, such as: energy simulation, daylight simulation, software comparison, optimisation, sensitivity analysis, etc. This review covers a more wide range of research areas to see how the industry might benefit from the combined efforts made across disciplines. The reader looking for a more in-depth review of a specific topic may look into the following:

- Kanters et al. [8]: tools and methods used by architects for solar design.
- Hopfe et al. [13]: comparison of 6 BPS tools and potential of BPS in conceptual design phase.
- Crawley et al. [14]: comparison of 20 building energy performance simulation programs.
- Attia et al. [15]: survey with 249 architects and their relation to 10 BPS tools.
- Attia and Herde [16]: comparison of 10 early design simulation tools.
- Zhao et al. [17]: review on the prediction of building energy consumption.
- Pacheco et al. [18]: review on energy efficient design of building
- Ochoa et al. [19]: review of lighting simulation for building science.
- Tian [20]: review of sensitivity analysis methods in building energy analysis.
- Evins [21]: review of computational optimization methods applied to sustainable building design.
- Machairas et al. [22]: review of algorithms for optimization of building design.
- Bucking et al. [23]: uncertainty, sensitivity, and optimisation in building simulation.
- Iwaro et al. [24]: criteria weighting framework and multi-criteria decision making.
- Fumo [25]: basics and classification of whole building energy estimations.

Primarily works after 2005 have been included.

2. Research areas

2.1. Proactive building simulations

The engineer responsible for building performance simulations regarding energy, comfort, cost, etc., are often asked various “What if…” questions by building owners, entrepreneurs, and architects. These questions refer to alternative design options, such as “what if we allow external shading”, “what if we increase window-to-wall ratio”, “what if we combine venting and overhangs to avoid mechanical cooling”. Since most simulation software is evaluative in nature, such queries are difficult to give immediate replies to – especially in the early design phase where the option space is immense. Trying to answer such queries will often require the simulation expert to run additionally simulations between meetings or workshops. When the answers are obtained, it may already be too late, since the design has evolved and new issues and questions have arisen instead. Addressing this issue of time-consuming, iterative, and evaluative nature of building simulation, Shady et al. [5] used the concept of “pre-design informative” BPS that enables proactive guidance and support for decision making during early design. According to the authors, only 1% of the then 392 tools listed on the U.S. Department of Energy homepage [26] can be categorized as pre-design informative.

In this paper, we will distinguish between the terms “pre-design informative” and “proactive” simulations. The term “pre-design informative” is applied to methods where simulations have been carried out prior to the design stage. Examples include the constructions of meta-models (see Section 2.2.3) and the creation of databases from simulations of predefined rooms or building types. “Proactive” is considered a more broad term that also applies to methods where alternative simulations are carried out in a structured way to guide, rather than evaluate, the design.

Petersen [7] recognizes the potential of the simulation environment to become more proactive and provide data-driven advice along with design implications. He therefore focuses on enabling “the support environment to generate input to the overall building design process prior to any actual design decisions”. Petersen introduces a novel tool that enables parametric, room-level simulations with respect to energy consumption, air quality, daylight, and thermal comfort. For all inputs, the user assigns a reference value and optionally two alternative values. Along with the reference model, the tool will then perform one-at-a-time simulations to evaluate two variations for each of the varied parameters. The tool was tested on three real building projects, where the actors involved found this one-at-the-time parametric analysis useful for decision support. Though, the extent to which the design information was allowed to influence design decision differed due to different opinions on the benefits from inter-disciplinary collaboration in the conceptual design stage. This demonstrates the importance of an open mind towards multi-actor collaboration and towards the implementation of novel methods and tools that may improve the design process.

Similar to Petersen's approach, Ochoa and Capeluto [27] have developed an advice tool for the conceptual design stage of intelligent facades based on energy and visual comfort. The tool employs the EnergyPlus [28] engine in order to evaluate intelligent facades and to ensure continuity with the subsequent preliminary and detailed design phases. The many EnergyPlus inputs are abstracted away by using presets that are determined from a few architectural considerations concerning location, main orientation, occupancy level, sophistication level, facade openness, surroundings, and building depth. In that way, the architect does not need to assign specific input values into the simulation tool. Instead, the designer defines relatively few properties regarding geometry and location along with some desired design concepts. On the basis of these properties and design intentions, the tool creates building design alternatives that follow a set of built-in design rules. An interesting feature is that the logic also generates an alternative with a degree of randomness to avoid locking the designer into one direction. In the end, the designer is presented with a list of detailed design alternatives.

Attia et al. [5] has developed a prototype tool for net zero-energy buildings in hot climates with the purpose to inform designers prior to decision making. The prototype consists of a simple and easy-to-use interface enabling parametric runs of the EnergyPlus simulation engine. Numerous inputs for EnergyPlus have been reduced to reflect the early design stage. This allows for fast creation and exploration of a variety of alternatives while using advanced, validated simulation software. The prototype allows for simulation of a number of predefined building types and applies sensitivity analysis to guide decision making.

The above works focus on early design integration, creation of alternatives, and guidance of the designer. Such ambitions may also be facilitated by means of statistical methods as described in the following section.

2.2. Statistical methods

In this context, statistical methods refers to a design process where the modeler runs a large number of simulations in a structured manner and use statistical techniques to achieve design support from the simulated data. This approach enables the modeler to explore a large design space in a systematic way, which potentially enlarges the solution space, and thereby improves building performance compared to typical one-at-a-time parametric analysis (see Fig. 1). Statistical analysis also allows for definition of inputs in form of possible spans, thereby addressing the issue of non-determined or uncertain inputs that is characteristic of the early design phase. Finally, statistical analyses are

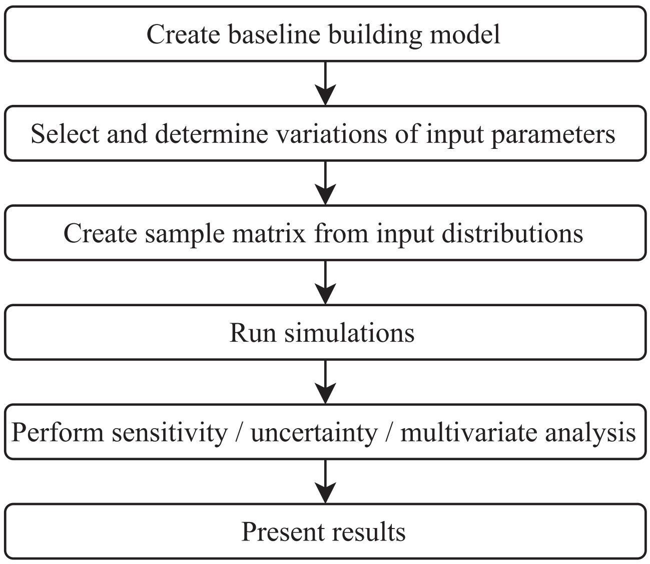


Fig. 3. Schematic flow diagram of typical implementation of statistical analyses in a building performance simulation process.

suitable for addressing the challenges related to the probabilistic nature of user behavior and weather.

Statistical building performance modeling consists of three intertwined disciplines, namely uncertainty analysis, sensitivity analysis, and multivariate analysis. Present work covers diverse uses of statistical analysis, but the following steps are common (see reduced workflow on Fig. 3):

1. A baseline model is created in building performance software capable of calculating the objectives of interest.
2. Depending on the scope of the analysis, a number of input parameters, ranging from a few to hundreds, are selected. Each parameter is assigned with a probability density function that reflects parameter uncertainty related to the numerical model, boundary condition, physical property, or design variability.
3. A sample matrix is constructed from the probability density functions. Various sampling procedures exist and their applicability depends on the analysis to be performed. Sampling procedures include random, stratified, factorial, Latin hybercube, and quasi-random with low-discrepancy sequences [29].
4. For each sample a building simulation is performed and outputs of interest are collected.
5. Results are analyzed utilizing uncertainty analysis, sensitivity analysis, multivariate analysis, or combinations of these. The results may also be used to create meta-models as described below.

This workflow is often facilitated by using statistical software packages such as SimLab [30] or the statistical programming language R [31] in combination with building performance software [20,32]. Increasing interest and need for such workflows drives developers to create extensions to the building simulation environment to facilitate parametric modelling, e.g. Parametric Analysis Tool for OpenStudio [33] and jEPlus [34]. It seems that the proliferation of scripting languages, particularly interpreted and dynamically typed languages such as Python and Ruby, makes programming more accessible for simulation specialists who want to perform very specific simulation tasks [35–37].

2.2.1. Uncertainty analysis

An early, comprehensive research of uncertainties related to building simulations was conducted by MacDonald [38], who addressed the problem of quantifying the effects of uncertainties on the predictions derived from building simulation software. More recent work focus on utilizing uncertainty analysis as part of the decision making process [39–42]. Hopfe and Hensen [41] conclude that "the integration of uncertainties in BPS provides evidence based decision support in design team meetings and dialogues with building partners." When augmented by sensitivity analysis such integration will give an idea of the significance of uncertainties and facilitate quality assurance of the model. Uncertainty analysis is useful to investigate design variation and gives insight into design robustness and possible ranges of performance indicators, i.e. minimum and maximum values for energy demand, daylight metrics, costs, etc. However, decision making under consideration of uncertainty is not straightforward. As exemplified by de Wit and Augenbroe [39] a decision maker will find it difficult to decide whether or not to implement a cooling system when such a system is required if the hours with overheating exceeds 150 but the overheating temperatures are represented by a probability functions that spans over this limit. To address this issue, the authors propose implementation of Bayesian decision theory by setting up and comparing utility functions. Another approach for decision-making under uncertainty is suggested by Rezaee et al. [42]. They estimate the level of confidence that option *A* performs better than option *B* by comparing output distributions for each of the two alternatives. Thereby, the designer gets an idea of how likely it is that one design proposal will outperform another.

Since uncertainty is inherent in all building simulations one might argue always to include uncertainty analysis. Even in late retrofit design and in model calibration, the effects of occupants' behavior and unpredictable weather impose substantial uncertainty on the model's predictions which militate against use of deterministic calculations. Various studies applied uncertainty and sensitivity analysis to study the effects of occupants' behavior and weather variability [43–48]. Brohus et al. [43] perform both a theoretical and empirical study of energy consumption of domestic buildings which shows occupant's behavior to be the major contributor to the variance. Hoes et al. [44] also include thermal analysis in an office case study and propose a methodology for better representation of user behavior. Their results show that no general design concept ensures robustness towards user behavior without applying extensive oversized active systems. According to O'Brien [45], implementation of passive systems, e.g. fixed solar shading, may reduce both energy use and uncertainty associated with occupant behavior. Applying uncertainty analysis is often accomplished by assigning probability density distributions to uncertain inputs as described above. This method, however, does not work for uncertainties related to user behavior and weather when performing whole-year simulations. To address this issue, Rodríquez et al. [46] defines three levels of both occupant load and weather load. By combining these, a total of nine scenarios are investigated which enhances the robustness of the analysis. Furthermore, the authors apply sensitivity analysis which shows that the ranking of influential inputs are similar for the nine investigated scenarios.

Summing up, uncertainty analysis may aid building design in various manners. This analysis ensures more reliability to the results, enables exploration of large design spaces, and assesses model quality and robustness. Though, design comparison becomes less straightforward when considering uncertainties as compared to evaluating deterministic calculations.

2.2.2. Sensitivity analysis

Various authors suggest to incorporate sensitivity analysis during early design to identify the input parameters with highest impact on building performance [40,49–52]. By identifying the most influential input parameters, the design team may direct their attention to these inputs in subsequent analyses, such as parameter variations and optimisation, and during construction of meta-models. Sensitivity analysis may answer "What-if" questions by calculating regression or correlation coefficients which indicate the size and direction of the change in performance when changing values for a certain input [40,53]. Different sensitivity analysis techniques are described thoroughly in an often cited book "Global Sensitivity Analysis: The Primer" by Saltelli et al. [29], while the use of sensitivity analysis in building energy analysis is covered in a comprehensive review by Tian [20].

Sensitivity analysis can be divided into local and global approaches [29]. Derivative based local methods consider the effects of uncertain inputs around a point in design space (or baseline model) by varying one parameter at a time (OAT). This approach requires few computations but is ill-suited for non-linear systems [29]. Global methods consider the uncertain inputs over the whole input space. Global methods are more versatile since they can handle nonlinear, non-additive, and non-monotone systems and consider the effects of interactions between inputs. As an example of a nonlinear and non-monotone system in BPS, we may consider energy consumption as a function of windows' *g*-value. For a given model, the heating load in winter may be reduced by increasing the g-value but only to a certain limit after which the cooling load will increase. Yet, this relationship is highly dependent on other parameters such as fenestration, solar shading, shadows, set points, internal loads, etc. These complex relations may be investigated by applying sophisticated, global sensitivity analysis methods such as decomposition of variance and other quantitative measure. Though, these approaches typically increase the amount of computational effort accordingly. Hemsath and Bandhosseini [52] argue that pre-design local sensitivity coefficients may aid early decision-making, and it may be extended to global analysis in a later design optimisation stage.

Sensitivity analysis may be applied for multiple performance indicators and thereby provide an overview of critical design parameters in a holistic design context. Using an office test case, Jin and Overend [53] calculated sensitivity indices for 14 facade design variables with respect to 13 different outputs related to energy, comfort, and cost. The resulting sensitivity coefficient charts for three different climatic zones help allocate design time and construction budget to the variables with highest impact on performance.

2.2.3. Meta-modelling

A meta-model may be defined as a simplified model of a model. In other words, if a numerical model is an abstraction of the real world, the meta-model is yet another abstraction of that numerical model. Meta-modelling involves analysis of input and output relationships in order to establish a mathematical relationship (algorithm) that is easy and fast to compute. A broad range of techniques exist, such as Artificial Neural Network (ANN) [54], Support Vector Machines (SVM) [55], Kriging [56], Multivariate Linear Regression [4,57], but in general no type is optimal in all circumstances [58].

In a building simulation context, a meta-model is typically constructed from a large set of simulations made with validated, detailed building performance software which is often computationally heavy. Alternatively, a meta-model may be constructed from experimental or observational data. For instance, meta-models can be constructed from large building performance databases [56,59]. The simplified model usually consists of a limited set of inputs and outputs that are relevant for the task at hand. The reduced set of inputs and the computationally fast algorithms makes meta-modelling attractive for early building design where only a few variables have been identified and the demand for fast feedback is crucial. Due to the fast algorithms, meta-modelling may be attractive when performing optimisation, uncertainty analysis, sensitivity analysis, and real-time simulations. Techniques based on regression analysis, sometimes considered easier and more practical [60], enable both interpretation and prediction [61]. Interpretation of regression coefficient helps understand input-output relationships as well as interactions between inputs – i.e. sensitivity analysis is easily accomplished. Additionally, regression coefficients enable prediction of building performance and hence provide proactive decision support.

The literature, reviewed here, concerns early building design, retrofit analysis, and test of the meta-modeling techniques. Performance indicators of interest include heating and cooling loads [4,55–57,62,63], thermal comfort [55], indoor air quality [54], daylight factor [60] and net cost [62]. The training set for establishing the models consist of both experimental [56,59] and simulated data. The use of meta-modelling in a holistic context will probably become highly laborious since individual algorithms must be developed for each performance indicator. Furthermore, a meta-model is only applicable in the domain of which it has been constructed, i.e. it becomes invalid if the prerequisites change, e.g. loads, orientation, constructions, etc. This characteristic is a considerable downside worth mentioning.

2.2.4. Multivariate analysis and filtering

Several authors make use of a stochastic approach to run an exhaustive set of simulations of the design space [33,51,64]. Applying filtering methods afterwards help identify favorable areas of the design space that meet certain design criteria [32]. Moreover, multivariate analysis of the vast amount of data obtained from thousands of simulations may be assisted by various visualization techniques such as scatterplots, histograms, and parallel coordinate plots. Naboni et al. [64] demonstrate the possibilities of cloud computing by running 221.184 EnergyPlus [28] simulations within 72 h. Using factorial sampling of 8 discrete design variables, all combinations are considered. The method is compared to a conventional manual approach where a practitioner is assumed to generate and run up to 50 manually configured simulations. When comparing time consumption, the additional computational time of the parametric approach is balanced out by the time spent on setting up and analyzing the manual simulations. The advantage of the parametric modelling is the exhaustive, global investigation of the design space and the possibility to apply statistical analysis. By comparing Pareto fronts, the authors show that the parametric approach may reduce both cooling and heating needs significantly. For instance, the energy savings are increased by 33% when choosing the best performing parametric design as comparing to the best performing manually configured design.

2.3. Holistic design

A building design needs to satisfy a vast range of often contradicting requirements and objectives. Certification schemes such as DGNB [11], LEED [12], and BREEAM [65] involve evaluation of up to 100 objectives. Some may be estimated quantitatively with simulation software while others can only be evaluated qualitatively. Another characteristic of the building design process is the gradually increase in identified design parameters and objectives [40]. For example, it is nearly impossible to calculate room acoustics, draught, and LCA in conceptual design. Since objectives are often correlated, a design change improving a certain objective will affect other objectives as well. These circumstances challenge the holistic design approach, especially in the early design phase. As stated by Cheung et al. [35] "There is a clear need for a designer-focused system that can give simultaneous design assessment on various aspects in the conceptual design stage."

One element of holistic design in a simulation context is to enable simultaneous calculations of as many objectives as possible. This may be facilitated by improved interoperability by common file exchange schemas (IFC, gbXML, etc.) or by integrating a multitude of algorithms into one software platform (see Section 2.5). Another element is to combine these diverse performance results and extract information that supports decision making. This lies in the extensive field of multi-criteria decision making (MCDM).

Pohekar and Ramachandran [66] and Wang et al. [67] have made reviews on MCDM in neighboring research areas and describe different techniques that aid decision making when considering conflicting and multiple objectives. These methods are based on weighting averages, priority setting (Analytical Hierarchy Process (AHP)), outranking (ELECTRE, PROMETHEE), and fuzzy principles [66,67]. In the field of sustainable energy decision-making, the simple method of equal criteria weights are the most popular followed by the more comprehensive Analytical Hierarchy Process [67]. Similar trends may apply to the field of building design, where weighting systems is demonstrated by Bjørn and Brohus [68], Iwaro et al. [69], and Østergård et al. [24]. Moreover, weighting frameworks, such as DGNB, LEED, and BREEAM, are getting increasingly popular. These weighting systems involve prioritization and establishment of comparable performance measures. Such systems compel the design team to think holistic and they reveal which objectives may be improved.

In the Analytical Hierarchy Process the decision problem is decomposed into a hierarchy of sub-problems. Decision makers compare these sub-problems pairwise by assigning numbers from 1 as 'equally important' up to e.g. 9 for 'extremely more important' [70]. A matrix consisting of all pairwise comparisons is used to calculate numerical weights for all objectives in the hierarchy, allowing diverse objectives to be compared in a consistent way. Hopfe et al. [70] use AHP to support multi-criteria decision making under uncertainty based on stakeholders preferences. By propagating uncertainty from design parameters into probability distributions of performance indicators, much information is generated but it complicates decision making (see example in Section 2.2). Applying AHP helps rank design options where uncertainty is included and thereby aids decision making while reaping the benefits from uncertainty analysis. According to Iwaro et al. [24], the majority of the subjective criteria weighting frameworks, such as AHP, fail to consider objective information. Therefore, Iwaro et al. suggest an integrated frame where AHP is combined with an objective weighting approach to assess life cycle performance. The framework was concluded to provide a robust methodology for weighting and assessment of the sustainable performance of residential building designs.

Another research dealing with uncertainties and multiple objectives is that of Jin and Overend [53]. As described in Section 2.2, the authors take into account the large uncertainties related to early design to create façade sensitivity charts for 13 output variables describing the performance of two office scenarios in three geographical locations.

Holistic design promotes evaluation of a vast number of opposing performance indicators. Since design comparison becomes more troublesome when considering multiple objectives, the design team may want to exclude objectives having little importance or having large correlation with other objectives. An example of the latter, in a Danish context, is the evaluation of overheating hours above 26 °C and 27 °C, which are required by building code. From a design perspective, the two measures will show similar behavior and addressing either one of them will most likely have similar consequences on building design. These nearly redundant objectives may be excluded to reduce the information load. To identify such objectives, the following methods are listed by Wang et al. [67]: the least mean square method, the min-max deviation method, and the correlation coefficient method. These methods are simple to apply and may help to focus on the most important parameters in a holistic design process.

In this brief overview of holistic design, we have left out multi-criteria optimisation which will be covered in the following. In holistic design, we stress that optimisation requires caution since building design is a high-dimensional and complex task where a single best holistic solution (or single Pareto front) does not exists.

2.4. Optimisation

In this context, optimisation refers to the automated use of mathematical optimisation in combination with building performance simulations. The aim of this section is to give an overview of trends, benefits, and challenges based on five reviews [21,22,71–73] related to building design optimisation. A building optimisation analysis typically consists of the following steps that may be repeated in an iterative design process (combined from Machairas [22] and Nguyen et al. [72]):

1. Identification of design variables and constraints.
2. Selection of simulation tool and creation of a baseline model.
3. Selection of objective function(s).
4. Selection of optimisation algorithm.
5. Running simulations until optimisation convergence is achieved.
6. Interpretation and presentation of data.

Since the turn of the millennia, publications about building optimisation have roughly increased tenfold [21,22,72]. This development is aided by advances in computer science in terms of parallel and cloud computing as well as advancements in optimisation theory where genetic algorithms (GA) and particle swarm optimisation are prevalent [71,72]. Based on keyword searches in the scientific database "ScienceDirect", Machairas et al. [22] conclude that optimisation on HVAC and controls represent the majority of the publications. Though, optimisation of parameters influencing building design has become increasingly popular during the last decade. Applying optimisation to building design is often motivated by the stringent and often divergent requirements of high-performance buildings. Interviews with researchers and practitioners emphasize that optimisation of building design is not about finding the "best" solution but rather to find alternative solutions from automated exploration of a large design space [71]. Arguably, "parameter variations" may be a better term when this is the purpose of the optimisation.

Building designers seek to design buildings that perform well on a wide range of both quantitative and qualitative measures. While early building optimisation studies were dominantly single objective, the trend is towards multi-objective optimisation [21]. One way to include more objectives is to apply the weighted-sum method which reduces the optimisation problem to single-objective at the cost of introducing arbitrary fixed weights to all objectives. Otherwise, multi-objective optimisation consists of quantifying trade-offs curves of solutions, known as Pareto Fronts, where objectives cannot be improved further without worsening others. Typically, multi-optimisation addresses only two objectives though a few recent works applied full 3-objective optimisation [74,75].

According to two different reviews [21,76], the common objectives to optimize, in decreasing order, are energy, cost, thermal comfort, and carbon dioxide. Often, optimisation of one or two objectives is performed while setting constraints for other objectives to make sure the constrained objectives comply with relevant standards. Arguably, this approach is inadequate when the designer needs to score high in holistic assessments such as LEED [12], BREEAM [65], and DGNB [11] where the overall score depends on a wide range of opposing objectives. In such cases, the weighted sum method seems more appropriate. Furthermore, since building simulations lack qualitative measures, such as aesthetics, space layout, and logistics, optimisation on a few objectives may be at the cost of equally important qualitative measures.

Despite the growing interest for building performance optimisation in academia, adoption in practice is still limited [71]. Barriers to widespread implementation consist of various issues that need be addressed. Time-consuming computations have long been a well-known obstacle. This may be overcome by the proliferation of parallel and cloud computing or by constructing computationally fast meta-models [21]. Another issue is the inability of optimisation algorithms to cope with uncertainties [21], which are especially large for early building design. Performing optimisation is not trivial, and it requires knowledge and experience to formulate the problem properly and select appropriate software and algorithms which calls for education of practitioners [71]. Another hurdle is that available optimisation tools, either generic[2] or customized for building simulations[3], require time-consuming and error-prone linking to the simulation software [71]. A solution for this could be full integration of optimisation techniques into commercial software. In addition, interviewed researchers and practitioners desired the following features: better GUIs, parallel computing, and coupling of simulation software and optimisation tools to do real-time optimisation within BIM models [71]. Simulation experts also need to prove legibility of optimisation to architects, building owners, and contractors since building design is a multi-collaborative iterative process where stakeholders have different areas of responsibility.

2.5. CAD-BPS interoperability

For many years, the field of data exchange and interoperability between CAD models and building performance simulations (BPS) has received a lot of attention among software developers and researchers. Table 1 illustrates different ways of combining CAD and BPS, i.e. representations of the physical world (CAD) and analytical, numerical models (BPS). Note that the illustrations in Table 1 only show one numerical model although there often exists a number of such models of varying sizes. Moreover, interoperability may be a mix of the methods shown. For example, run-time interoperability often only works when there exists a common file exchange format. Furthermore, one type of analysis may be performed using an integrated, simplified algorithm for early design support, whereas a more detailed analysis might prove necessary in later stages. For example, detailed simulations using CFD software.

2.5.1. Integration and direct links in early design

Improved interoperability would address several of the early design issues identified in the introduction, e.g. time-consuming iterative modelling, and need for rapid feedback. Since the early design stages are dominated by architects, who create building models using CAD software, this section focus on the features of the integrated and run-time coupled approaches.

During the last decades, the CAD industry has evolved from 2D drawings to 3D models and now "4D" models where more and more semantic data is integrated into the CAD environment. Moreover, advanced CAD software tends to integrate an increasingly amount of analyses, such as collision control, duct sizing, and solar analysis. In addition, various software vendors facilitate BPS through dynamically coupled tools or add-ons. Examples include Autodesk's Green Building Studio [78] for Revit [79] while Graphisoft's has EcoDesigner Star [80] for ArchiCAD [81]. Third party vendors also enable direct links to BPS through application programming interfaces (API) to promote early design decision support and rapid analyses. These include Sefaira [82], IESVE [83], and OpenStudio [84] that may be linked to SketchUp [85]. Several of these couplings rely on common file formats to do so, i.e. IFC, gbXML, osm, etc. Various plug-ins and API's make use of detailed software engines, such as EnergyPlus [28], Daysim [86], and Radiance [87], which are computationally heavy and require lots of inputs. As a consequence, most inputs are assigned to defaults values related to specific building types. The challenge of running time-consuming BPS, while designing in CAD, may be overcome by applying cloud computing. Such development may facilitate run-time analysis, enable rapid feedback, ease iterations, and reduce amount of (re)modeling. Moreover, zoning may be set up in the CAD environment after which changes in geometry automatically updates zoning as well. To test such an integrated framework, Batueva and Mahdavi [9] assessed the use of Graphisoft's EcoDesigner [80] which has been integrated into ArchiCAD. The authors acknowledge the effortless interoperability but desire more intelligence in terms of guidance and comparison features [9].

Much of these efforts rely on software vendors to incorporate BPS into the CAD domain, or link the two, but similar work is carried out in the scientific community [88–92]. Jakubiec and Reinhart [89] describe a plugin for Rhinoceros [93] which combines daylight analysis, using Radiance and Daysim, with thermal load calculations, using EnergyPlus. Muehleisen and Craig [90] implement the ISO 13790 monthly energy model into the OpenStudio environment, which is available as a plug-in for SketchUp. The authors conclude that this particular plug-in is suitable for parametric simulations and Monte Carlo analysis during early design, because the simulation time is five orders magnitude faster than the equivalent EnergyPlus model and the simplified algorithm requires far fewer inputs.

2.5.2. Parametric geometric modelling

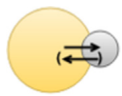
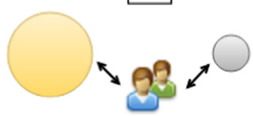
As exemplified by various authors [94–96], parametric modeling are increasingly adopted in design practice by means of tools like Grasshopper [97], Dynamo [98], and GenerativeComponents [99]. Concurrently, plug-ins are developed to link these tools with BPS thereby enabling data-driven support for early stage, parametric, and geometric modeling. Examples of plug-ins for the probably most widely used tool, Grasshopper, include: a) Honeybee which links to Radiance, Daysim, EnergyPlus, and Openstudio [100]; b) Mr. Comfy which facilitate interactive visualizations of thermal simulations results [91]; c) ICEbear that integrates indoor climate, daylight, and energy performance [92]; and d) Tortuga [101] which estimates LCA and a global warning potential based on the Ökobau database [102]. Comprehensive libraries of applications relevant to the architecture, engineering, and construction industries can be found on the sites aec-apps.com [103] and Food4Rhino [104]. The ability to add several plug-ins to parametric modeling could be a feasible way to facilitate holistic simulation support. Though, even if plug-ins ensure smooth CAD-BPS interoperability during the early design phases such plug-ins may not be suitable for detailed analysis. Therefore, it is desirable that plug-ins make use of detailed software engines or common exchange formats to avoid complete remodeling, and to avoid inconsistent results, when the design evolves to detailed stages.

Despite improvements with interoperability, plenty obstacles remain. Most of the couplings illustrated by arrows in Table 1 are

[2] E.g. GenOpt, ModelCenter, modeFRONTIER, DAKOTA, iSIGHT, Matlab optimisation toolbox [136].

[3] E.g. BEopt, TRNOPT, MultiOpt, jEPlus + EA, GENE-ARCH, Opt-E-Plus [136].

Table 1
Characteristics and examples of four different methods to combine CAD (large disk) and BPS (small disk). Categorization adopted from Petersen [7] and Citherlet [77].

Method		Characteristics	Examples
Integrated		Numerical calculations integrated into CAD environment.	Collision control, duct sizing, and solar analysis
Run-time interoperable		Links between CAD software and analytical models established by add-on or API. Simulations performed at run-time or in a concurrently running desktop or web edition of the BPS tool.	Grasshopper and Dynamo plugins. SketchUp & Revit with Sefaira, OpenStudio
File exchange		Common file exchange format readable and sometimes writable from both CAD and BPS tools – i.e. Building Information Modelling (BIM).	Proprietary: dwg, rvt, gbXML, osm Public: IFC, XML
Standalone (users interpret)		Data interpreted by users. Building simulationist remodels building or selected rooms by interpreting CAD models or drawings and eventually presents results orally or in reports.	EnergyPlus, Radiance

uni-directional. It is very seldom that properties derived from BPS are transferred back to architectural or BIM model. Moreover, BIM are still challenged by the complexity of the heterogeneous BPS data which requires user interpretation as well as extensive pre-processing and enrichment of incomplete building information [105,106]. Moreover, this central framework with a shared schema has to be operated in consensus with all stakeholders, i.e. architects, engineers, and contractors [107]. Aforementioned examples of coupled and integrated models are often limited to single user use, since the coupled programs normally have to be installed on the same computer. This is troublesome in a multi-actor, interdisciplinary collaboration where different actors possess expertise and responsibility over different areas [105].

In conclusion, much effort is made in academia and by software developers to improve interoperability between CAD and BPS in the early design stages. Achieving effortless interoperability and smooth transition between design stages will make life easier for all parties involved.

2.6. Knowledge based input generation

Building performance software requires hundreds or thousands of inputs which may be assigned manually by the user or by importing data from CAD models, shared schemas (BIM), and databases within the software. Databases may include constructions, HVAC components, load and user profiles, weather data, etc. They play an important role in terms of modelling time and reliability. The quality and applicability of such databases depend on their ability to address several issues such as:

- Ease of implementation.
- Scalability and updatable.
- "Best practice", i.e. in accordance with code compliance or prior experience.
- Flexibility, e.g. usable for both early and detailed analysis and across different tools.
- Ability to be varied in multiple (parametric, batch, or stochastic) simulations.
- Documentation and validity.

Vendor supplied libraries often serve as the only or main source of information for practitioners and are often poorly documented and difficult to share and reuse across applications [106]. Such issues are addressed by National Renewable Energy Laboratory that are developing a comprehensive, online, searchable library of energy building blocks and descriptive metadata which works for different applications [108,109], e.g. EnergyPlus [28], OpenStudio [84], and DOE2 [110]. Flexible and extensible set of attributes provide the opportunity to add metadata such as *U*-value, cost, and images. In addition, the attributes "user ratings" and "number of downloads" may support the selection of materials, components, and systems across fields and practitioners.

Another large online database is the "building performance database" which contains information about physical and operational characteristics of hundreds of thousands of real commercial and residential buildings in the U.S. [26,111]. Aimed at the vast retrofit market, this database enables assessment of energy retrofitting opportunities and helps to quantify risk related to project performance. A statistics tool is integrated to estimate expected changes in energy performance due to changes in component technologies. Though, since the database primarily concerns energy from existing buildings, the effects on indoor climate performance resulting from retrofitting remain unknown.

Performing simulations in the early design phase is challenged by lack of data. This is especially the case for detailed simulation software that requires a high level of information. This difficulty may be overcome by a macro-component approach where predefined constructions allows for energy and LCA assessments in the early phase using detailed software [112]. Similarly, Rodríguez et al. [46] aggregates macro-parameters of occupancy and weather data to enable uncertainty and sensitivity analysis in detailed models. Hiyama et al. [113] propose a method to automatically generate default configuration for simulations in the early stage thereby making the design process more efficient and consistent. The configurations are based on past experience in combination with objectives and constraints of the current project.

Pont et al. [114] make use of semantic web technologies to acquire and utilize building related data available on the Internet. Semantic rules and reasoning enable restructuring of ill-structured "web of documents" to machine-readable "web of data" by means of interlinking data from various web sources and by re-categorizing the data using consistent logic. Such methods can in theory be applied to any web-based resources such as databases and manufacturer sites. Data from different sources may be merged into one rich library with links to original data and providing opportunity for regular updates and acquisition of new information. This could be information about construction types, materials' properties, and prices,

The increasing use of uncertainty analysis and sensitivity analysis calls for development of databases that facilitate stochastic simulations. In contrast to deterministic defaults, the designer needs recommendations in terms of appropriate input distributions, input spans, and sampling strategies. Lee et al. [115] present an uncertainty and risk analysis toolkit that give energy modelers

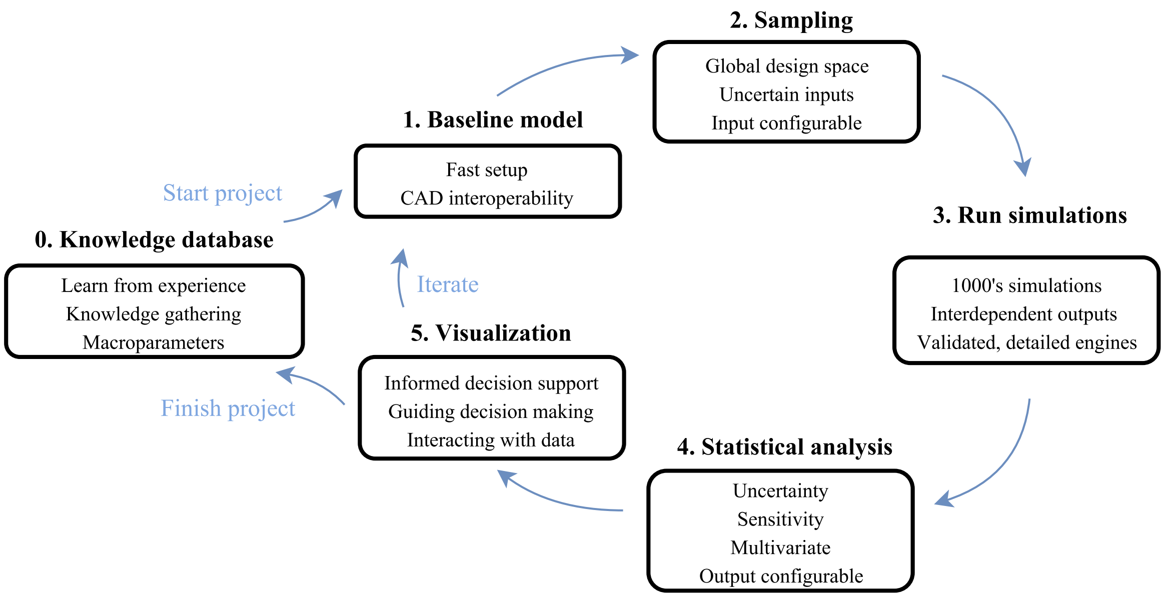


Fig. 4. Desired workflow and properties as facilitated by the proposed simulation framework.

access to previously defined uncertainty distributions for a variety of parameters and models. Furthermore, the toolkit provides automatic identification and modification of parameters values in simulation input files. Such efforts might make uncertainty analysis more accessible for non-specialists and help to increase the use of UA and SA.

To sum up, databases may be employed in a variety of ways to support and improve the building design process. When used for setting up initial configurations for building simulations, the practitioner must be aware of certain inherent risks: a) the configuration may return results in local "optimum" causing the designer to stop exploring a sufficiently large design space, b) default configuration may lead to misleading baseline models if there is a big discrepancy between database values and measurements, new requirements, and codes, c) initial configurations used in architectural design software may guide the architect in wrong directions if these configurations are not aligned with engineers who are responsible for code compliance in detailed design phase.

3. Software comparison

As stated in the introduction, the motivation for this review is to identify state-of-the-art within the field of building simulations with emphasis on early design. In chapter 2, we covered developments in literature across six research areas. In this chapter, we will propose a simulation framework combining the most promising methods found in literature after which we compare existing software packages that may satisfy some of the requirements of such a framework.

In the introduction, we identified a number of challenges related to building simulations in the design process (see Fig. 2). To address these diverse issues, it is necessary to combine several of the methods and developments described in the literature review in chapter 2. Based on those findings, we describe a framework that, presumably, facilitates proactive, intelligent, and experience based building simulation which aid decision making in early design. The proposed framework contains the following properties, which are combined in an iterative design process as illustrated on Fig. 3:

1. A knowledge based database represents the starting and finishing point for each project. It must facilitate fast input generation, consistency, and collection of experience. Moreover, the database should contain macro-parameters to enable the use of detailed software in early design stages (macro-parameters represent predefined sets of constructions, HVAC systems, time schedules, etc. that contain the input values necessary to run a detailed simulation). Finally, it should ease the definition of uncertain inputs.

2. A baseline model is swiftly set up by a combination of database inputs and suitable CAD interoperability.
3. Uncertainties are assigned to inputs and a sampling strategy is applied to explore the global design space and to facilitate uncertainty and sensitivity analysis.
4. Thousands of simulations are run using a validated and detailed software engine(s) that evaluate important, interdependent design objectives.
5. Data is analyzed using UA and SA.
6. On the basis of UA and SA, attention is drawn to the most important design parameters and the design team is informed of benefits and consequences of various design strategies. Interactive visualization allows for interaction with the simulated data where different stakeholders preferences may be explored.

The properties of desired framework shown on Fig. 4 entail various requirements of the simulation software. Therefore, we carry out a software review to assess features and limitations of current building simulations software packages. If no software satisfies all requirements, we aim to find software candidates that may be combined into the desired framework. Important properties of the reviewed software are[4]:

A. **Users:** Is the software primarily intended for architects, engineers, or both?
B. **Design stage:** In which design stages are the software typically used?

[4] Properties, omitted in this work, include: licensing, price, version, status (beta, deprecated), and number of users/downloads.

C. **Interoperability:** How does the BPS software connect to CAD environment and other software packages (see Table 1 for definitions)?
D. **Level of complexity of the core algorithms:** The complexity set the constraints of design options that the software enables to investigate and to what level of detail. For energy and thermal calculations, the monthly averaged ISO 13790 [116] is considered to have a "low" complexity level, as opposed to detailed software with "high" level of complexity due to features like multi-zones, advanced fenestration, HVAC and lighting control strategies, moisture transport, etc. Somewhere in between, we have the hourly averaged ISO 13790 [116] and RC models. For daylight calculations, simplified regression models have "low" complexity compared to advanced algorithms that, for instance, use ray tracing or radiosity to evaluate illuminance, luminance, and glare under various sky conditions and at different times a year [19].
E. **Objectives:** Important, interdependent objectives must be evaluated to ensure holistic design.
F. **Parametric:** Ability to run global parametric calculations and to perform UA and SA – either by using integrated features or by configuring input text files and accessing output text files. Option to enable cloud computing is desirable.

In the search for relevant, existing software, we rely on various resources: the tools directory list on U.S. department of energy homepage [10], the AEC-apps homepage [103], the BLDG-SIM mailing list [117], and prior knowledge of novel and trending software in Scandinavia. A reduced set of programs have been selected for further investigation. Deprecated software packages (Ecotect, Vasari) have been excluded along with software that did not seem to fit into the proposed framework (Modelica and TrnSys). The selected programs differ greatly in scope, validation, purpose, price, level of detail, and more, but each of them can potentially fulfill a specific purpose in the framework described in Fig. 4. Table 2 shows how the software compares. In the evaluation of the software, we rely on vendors' homepages, webinars, manuals, colleagues, and other reviews from academia [8,14,118,119]. Readers are reminded that both table structure and table inputs are very much governed by our subjective perceptions of the programs' capabilities.

According to our limited review, no existing software package satisfies all requirements of the proposed framework described Fig. 4. Though, the following three software setups may be used as starting point to test the framework.

Riuska [131] has integrated UA and SA into a standalone application which removes the challenges of linking the processes "sampling" and "statistical analysis" with the execution of the simulations (the links illustrated by arrows between 2 and 3, and 3 and 4 on Fig. 4). Supposedly, the lack of several important objectives (daylight, LCA, LCC) will be difficult to remedy by combining Riuska with other applications since UA and SA are constrained to Riuska.

OpenStudio [84] is a collection of software tools which include the validated, detailed applications EnergyPlus and Radiance. The packages "parametric analysis tool" (PAT) and "large scale analysis" extends OpenStudio's capabilities by enabling large parametric studies and cloud computing. A SketchUp plug-in facilitate use in early design whereas gbXML compatibility allows for geometry

Table 2
Comparison of software in terms of fulfilling the requirements of the proposed software framework. Checkmarks indicate fulfilment of the requirement. Checkmarks in parenthesis indicate that software include the specific feature without satisfying the requirement. See explanations of headers A–F in the text [120–130,132–135].

	Software	A. Users		B. Design stage				C. Interoperability	D. Core complexity	E. Objectives						F. Parametric sim.						Ref.
		Architects	Engineers	Conceptual	Preliminary	Detailed	Management			Energy	Thermal	Daylight	Air Quality	LCA	LCC	Cloud	I/O Configurable	UA	SA	OAT	Optimisation	
BPS (own engine)	Be10 (ISO 13790 monthly)[2]	(A)	E		✓			Standalone	Low	✓	(✓)						✓					[120]
	BSim		E		✓	✓		Standalone	High	✓	✓	✓	✓							(✓)		[121]
	DOE2		E			✓		Standalone	Medium	✓							✓					[110]
	EnergyPlus (E+)		E			✓		Standalone	High	✓	✓		✓				✓					[28]
	EPC (ISO 13790 hourly)	(A)	E	✓	✓			Standalone	Medium	✓							✓					[122]
	ESP-r		E			✓		Standalone	High	✓	✓		✓				✓					[123]
	IDA-ICE		E		✓	✓		File exchange	High	✓	✓	✓	✓				✓			✓		[124]
	iDbuild		E	✓	✓			Standalone	Medium	✓	✓	✓	✓							✓		[125]
	IESVE	(A)	E		✓	✓	✓	File exchange	High	✓	✓	✓	✓	✓	✓							[83]
	Radiance	(A)	E		✓	✓		Standalone	High			✓					✓					[87]
	VELUX Daylight Visualizer	A	E	✓	✓	✓		File exchange	High			✓										[126]
BPS (external engine)	A+E3D	A		✓	✓			Integrated	Be10	✓	(✓)	(✓)										[127]
	Daysim	(A)	E	✓	✓	✓		Run-time	Radiance			✓					✓					[86]
	DesignBuilder	(A)	E		✓	✓		File exchange	E+, Radiance, jE+	✓	✓	✓	✓		✓	✓					✓	[128]
	eQuest		E		✓	✓		Standalone	DOE2	✓												[129]
	N++		E	✓	✓	✓		Separated	E+, jE+, GenOpt	✓	✓		✓			✓				✓	✓	[130]
	OpenStudio		E		✓	✓	✓	File exchange	E+, Radiance	✓	✓	✓	✓		✓	✓	✓					[84]
	Riuska		E		✓	✓		File exchange	DOE 2, own engine	✓	✓		✓					✓	✓			[131]
	Sefaira	A		✓	✓			Run-time	E+, Radiance	✓	✓	✓				✓		(✓)	(✓)		✓	[82]
Plug-in	DIVA for Rhino	A	(E)	✓	✓	✓		Run-time	Radiance			✓										[132]
	Green Building Studio	A	(E)	✓	✓			File exchange	DOE2	✓						✓						[78]
	HoneyBee (GH)	A	E	✓	✓	✓		File exchange	OpenStudio, E+, Radiance	✓	✓	✓			✓		✓					[100]
	jEPlus (+ JESS)		E		✓	✓		Run-time	E+, DesignBuilder, N++	✓	✓	✓	✓		✓	✓		(✓)	(✓)	✓		[133]
	Parametric Analysis Tool		E		✓	✓		File exchange	OpenStudio	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓	[134]
	Solon	(A)	E		✓	✓		File exchange	Green Building Studio	✓						✓			(✓)	✓		[135]
Glue[1]	Dynamo	A	E	✓	✓			-	-													[98]
	Grasshopper (GH)	A	E	✓	✓			-	-													[97]

[1] Glue refers to software that enables linking between BPS and geometrical modeling through graphical programming (also refered to as algorithmic modeling)

[2] Be10 is mandatory to use for code compliance in Denmark

import from e.g. detailed Revit [79] CAD models for the late design stage. Through a SketchUp plug-in, OpenStudio may access the online, searchable library of user-rated building blocks described in Section 2.6 [108] and thereby include several features of the desired knowledge-based database. The combined set of tools seems to contain most of the properties needed by the proposed framework. Though, several features are still under development (beta-versions) and the use of all the packages mentioned (PAT, online database, large scale analysis, and SketchUp plugin) may be precarious and error-prone.

Honeybee [100] connects the Grasshopper and Rhino framework with OpenStudio and thereby combines the strengths of these packages. The former enables parametric studies of building geometry while the link to OpenStudio allows for building performance evaluation. However, Honeybee cannot access all features of OpenStudio – namely the "parametric analysis tool" and the "large scale analysis. A possible drawback is that the Rhino software is often not detailed enough for the final design models, which complicates data interoperability in the transition from preliminary to detailed design.

In conclusion, it is still not possible to perform global and holistic UA and SA that simultaneously vary geometry, zoning, materials, and systems. Riuska seems like a suitable fit for the engineer who wants to learn about, and experiment with, global parameter variations with emphasis on energy and thermal comfort. The OpenStudio framework expands these possibilities even further by accessing a knowledge based database, assessing most performance metrics, and enabling cloud computing. Though, obstacles remain in order to combine these capacities with the parametric tools, Dynamo and Grasshopper, which are growing increasingly popular among architects in particular. We emphasize that geometric parameter variations should be done while varying other sensitive inputs as well, i.e. global variations (see Fig. 1). Otherwise, the results from the BPS will only be valid around the specific baseline with fixed HVAC system, controls, materials, etc.

4. Conclusion and discussion

This paper provides an overview of the developments in academia and in the software industry related to the use of building simulations in early building design. As identified in the introduction, challenges to early stage deployment include lack of information, uncertainties, model resolution, and rapid change of design. In addition, general challenges include interoperability, time-consuming modeling, stricter and opposing requirements, limited reuse of knowledge, and simulation guidance. We identified six areas of research addressing one or more of these challenges: proactive building simulations, statistical methods, holistic design, optimisation, CAD-BPS interoperability, and knowledge based input generation. Below, we describe promising developments within these research areas along with our perception of how these developments may be used to improve building simulation in the early stages.

4.1. Proactive building simulations

Building simulation software is typically used to ensure building code compliance or to evaluate the performance of a few alternative designs or systems. Therefore, most software lacks the ability to guide the designer towards better performing buildings. To remedy this, a few authors have developed design tools to perform proactive building simulations. The three prototypes, reviewed here [5,7,27], allow fast creation of a number of alternative designs with emphasis on the early design phase. Such efforts contrast the typical, time-consuming trial-and-error approach. To avoid locking the designer in one direction, one tool [27] included a degree of randomness into the logic creating design alternatives.

4.2. Statistical methods

In academia, there is a growing interest in stochastic simulations supported by statistical analysis. This approach enables the design team to handle uncertainties and to explore large design spaces. Several works apply sensitivity analysis to identify correlations and interdependencies between inputs, and to rank design inputs of importance [49,50,52]. Other works uses parametric simulations or building performance databases [56,59] to construct fast meta-models which have few inputs and are suitable for rapid simulations. However, meta-models are only valid in the domains in which they were constructed. Applying uncertainty analysis are shown to add reliability to results, help explore vast design spaces [41], and assess model quality and robustness [44,45] (e.g. against uncertainties related to user behavior and weather [46]). Though, the inclusion of uncertainties makes design comparisons less straightforward. Finally, multivariate analysis and filtering techniques are effective when analyzing large dissect large amount of simulation data to guide decision makers [51,64].

4.3. Holistic design

The need to address multiple, contradicting objectives emphasizes a holistic approach during all stages of the design process. The means to do so are diverse and include weighted scoring systems [68,69], improved CAD-BPS interoperability, analytical hierarchy processes [24,70], and sensitivity charts of multiple objectives [53].

4.4. Optimisation

Motivated by the stringent and often divergent requirements of high performance buildings, optimisation algorithms have become increasingly popular in academia over the last ten years. The trend is towards multi-objective algorithms which focus on energy, cost, thermal comfort, and CO_2 [21,76]. However, algorithms are still limited to two or three variables at a time. A more important drawback is that optimisation lacks qualitative measures such as aesthetics, space layout, and logistics, which are critical parameters in early design. Thus, optimisation may favor solutions that come at the cost of other equally important qualitative measures.

4.5. CAD-BPS interoperability

For decades, academia and software developers have given much attention to the interoperability between CAD and BPS. These efforts address the issues of time-consuming modeling, continuity, and interdisciplinary collaboration. The different approaches to CAD-BPS interoperability may be split into four categories: a) integrated, b) run-time interoperable, c) file exchange, and d) standalone. Dominant vendors gradually integrate algorithms directly into the CAD software [80], or they develop proprietary BPS software to ease interoperability [78]. Concurrently, a wide range of add-on applications come to life in academia and in open-source communities. Much attention is put on run-time coupling to ensure fast feedback and enable parametric analysis [82,100]. The field is rapidly evolving, but still needs to overcome difficult obstacles (for instance, project configuration changes from one project to the next, and project members rely on different software packages and modeling tradition). An important challenge is the multi-actor collaboration in building design where companies team up differently for each project and have different software tools and design approaches.

4.6. Knowledge based input generation

Input generation for building simulation is often time-consuming and lacks reusability of best practice. Vendor supplied input databases are often rigid and have been made for detailed simulations in the late design stages. The works reviewed here cover "the development of flexible, online database with optional user ratings" [108,109]; "the definition of macro-components for level of detail in early design" [112]; and "input generation using semantic web technologies" [114]. However, the use of default inputs may limit the exploration of the design space since default configurations act as constraints for possible solutions. Further work is needed to improve input databases to account for the vast possibilities in early design and to enable stochastic modelling.

Based on the literature review, we have proposed a simulation framework with the ambition to facilitate proactive, intelligent, and experience based building simulations (see Fig. 4). Though applicable during all design stages, emphasis is on assisting the design team to explore the vast design space in the early phases. Another essential element, of the framework, is to ensure holistic design thinking in order to create buildings with high overall performance and with respect to different stakeholders' preferences. The proposed framework incorporates promising methods and ideas from literature, among others: flexible and experience based database for consistency and fast setup; uncertainty and sensitivity analysis to explore design space and ensure robustness; and a holistic approach considering multiple, contradicting objectives (e.g. energy, thermal comfort, and daylight). Finally, the proposed exploration of a vast, global design space using thousands of detailed simulations requires cloud computing to ensure sufficiently fast response time in the early phases.

We may test the framework hypothetically using the example from the introduction, in which a highly transparent design is justified by a combination of venting and solar shading with unrealistic preconditions. First of all, sophisticated (detailed) algorithms are needed to model venting and shading systems appropriately. Secondly, a holistic approach ensures that emphasis on certain objectives, such as daylight and transparency, does not come at the expense of other important objectives, like energy and thermal comfort. Uncertainty analysis may reveal insufficient robustness towards uncertainties related to control strategy, user behavior, and weather. Sensitivity analysis can help the designer to identify the most important simulation inputs, on which he can direct his attention. And finally, a knowledge based database would reduce the risk of starting out with unrealistically inputs.

To identify potential software satisfying the properties of the proposed framework, we have compared 27 software packages, plug-ins, and environments (see Table 2). From these, we highlighted three different setups, consisting of the standalone software Riuska [131], the OpenStudio framework [84], and the plugin Honeybee [100] that links Grasshopper [97] and OpenStudio. Since, currently these tools do not satisfy all requirements of the framework, further research and development is needed to enable setups that fulfil the full potential of proactive, holistic building simulations aiding decision making in the early design stages.

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 MOE A/S.

References

[1] European Parliament and Council. Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings. Off J Eur Union 2010;L153:13–35.
[2] Cole RJ, Valdebenito M Jose. The importation of building environmental certification systems: international usages of BREEAM and LEED. Build Res Inf 2013;41(6):662–76.
[3] BRE Global Ltd, The Digest of BREEAM Assesment Statistics, 2014.
[4] Hygh JS, DeCarolis JF, Hill DB, Ranjithan SR. Multivariate regression as an energy assessment tool in early building design. Build Environ 2012;57: 165–175.
[5] Attia S, Gratia E, De Herde A, Hensen JLM. Simulation-based decision support tool for early stages of zero-energy building design. Energy Build 2012;49: 2–15.
[6] Kanters J, Horvat M. The design process known as IDP: A discussion. Energy Procedia 2012;30:1153–62.
[7] Petersen S. Simulation-based support for integrated design of new low-energy office buildings. PhD Dissertation. Kgs. Lyngby: Department of Civil Engineering, Technical University of Denmark; 2011.
[8] Kanters J, Horvat M, Dubois M-C. Tools and methods used by architects for solar design. Energy Build 2014;68:721–31.
[9] Batueva, E. Mahdavi, A. Assessment of a computational design environment with embedded simulation capability, in eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014, 2015, pp. 197–202.
[10] Building Energy Software Tools Directory: Tools Listed Alphabetically. [Online]. Available: ⟨http://apps1.eere.energy.gov/buildings/tools_directory/alpha_list.cfm⟩. [accessed 14.08.14].
[11] German Sustainability Building Council, DGNB System, 2015. [Online]. Available: ⟨http://www.dgnb-system.de/en/system/gold-silver-bronze/⟩. [accessed 07.07.15].
[12] U.S. Green Building Council, LEED Certification, 2015. [Online]. Available: ⟨http://www.usgbc.org/certification⟩. [accessed 07.07.15].
[13] Hopfe, C.J. Struck, C. Harputlugil, G.U. Hensen, J. and De Wilde, P. Exploration of the use of building performance simulation for conceptual design, Proceeding IBPSA-NVL Conf., pp. 1–8, 2005.
[14] Crawley DB, Hand JW, Kummert M, Griffith BT. Contrasting the capabilities of building energy performance simulation programs. Build Environ 2008;43 (4):661–73.
[15] Attia S, Beltrán L, De Herde A, Hensen J. Architect friendly: a comparison of ten different building performance simulation tools. In: Proceedings of building simulation; 2009. p. 204–211.
[16] Attia, S. De Herde, A. Early design simulation tools for net zero energy buildings : a comparison of ten tools design process & tools of NZEB, in: Proceedings of Building Simulation, 2011, pp. 94–101.
[17] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16(6):3586–92.
[18] Pacheco R, Ordóñez J, Martínez G. Energy efficient design of building: a review. Renew Sustain Energy Rev 2012;16(6):3559–73.
[19] Ochoa CE, Aries MBC, Hensen JLM. State of the art in lighting simulation for building science: a literature review. J Build Perform Simul 2012;5(4):209–33.
[20] Tian W. A review of sensitivity analysis methods in building energy analysis. Renew Sustain Energy Rev 2013;20:411–9.
[21] Evins R. A review of computational optimisation methods applied to sustainable building design. Renew Sustain Energy Rev 2013;22:230–45.
[22] Machairas V, Tsangrassoulis A, Axarli K. Algorithms for optimization of building design: a review. Renew Sustain Energy Rev 2014;31(1364):101–12.
[23] Bucking S, Zmeureanu R, Athienitis A. A methodology for identifying the influence of design variations on building energy performance. J Build Perform Simul 2014;7(6):411–26.
[24] Iwaro J, Mwasha A, Williams RG, Zico R. An integrated criteria weighting framework for the sustainable performance assessment and design of building envelope. Renew Sustain Energy Rev 2014;29:417–34.
[25] Fumo N. A review on the basics of building energy estimation. Renew Sustain Energy Rev 2014;31:53–60.
[26] U.S. Department of Energy, Building Performance Database, 2015. [Online]. Available:⟨http://energy.gov/eere/buildings/building-performance-database⟩. [accessed 04.02.15].
[27] Ochoa CE, Capeluto IG. Advice tool for early design stages of intelligent facades based on energy and visual comfort approach. Energy Build 2009;41 (5):480–8.
[28] U.S. Department of Energy | Energy efficiency and renewable energy, EnergyPlus, 2015. [Online]. ⟨Available: https://www.energyplus.net/⟩. [accessed 26.06.15].
[29] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. Global sensitivity analysis: the primer. Chichester, England: John Wiley & Sons Ltd; 2008.

[30] The European Commission's in-house science service, SimLab. [Online]. Available: ⟨https://ec.europa.eu/jrc/en/samo/simlab⟩. [accessed 07.07.15].
[31] R Development Core Team, The R Project for Statistical Computing. [Online]. Available: ⟨http://www.r-project.org/⟩. [accessed: 09.02.15].
[32] Burhenne S, Tsvetkova O, Jacob D, Henze GP, Wagner A. Uncertainty quantification for combined building performance and cost-benefit analyses. Build Environ 2013;62:143–54.
[33] Macumber, D.L. Ball, B.L. Long, N.L. A graphical tool for cloud-based building energy simulation, in: 2014 ASHRAE/IBPSA-USA Building Simulation Conference, 2014, pp. 87–94.
[34] Zhang, Y. Korolija, I. Performing complex parametric simulations with jEPlus, in: SET2010 - 9th International Conference on Sustainable Energy Technologies, 2010.
[35] Cheung FKT, Rihan J, Tah J, Duce D, Kurul E. Early stage multi-level cost estimation for schematic BIM models. Autom Constr 2012;27:67–77.
[36] Miller, C. Hersberger, C. Jones, M. Automation of common building energy simulation workflows using Python, in: Proceedings of Building Simulation, 2013, pp. 210–17.
[37] Nembrini J, Samberger S, Labelle G. Parametric scripting for early design performance simulation. Energy Build 2014;68:786–98.
[38] Macdonald, I. Quantifying the effects of uncertainty in building simulation, PhD Dissertation, Department of Mechanical Engineering, University of Strathclyde, United Kingdom, 2002.
[39] de Wit S, Augenbroe G. Analysis of uncertainty in building design evaluations and its implications. Energy Build 2002;34(9):951–8.
[40] Struck C, de Wilde PJCJ, Hopfe CJ, Hensen JLM. An investigation of the option space in conceptual building design for advanced building simulation. Adv Eng Informatics 2009;23(4):386–95.
[41] Hopfe CJ, Hensen JLM. Uncertainty analysis in building performance simulation for design support. Energy Build 2011;43(10):2798–805.
[42] Rezaee, R. Brown, J. Augenbroe, G. Kim, J. A new approach to the integration of energy assessment tools in CAD for early stage of design decision-making considering uncertainty, in: eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014, 2015, pp. 367–373.
[43] Brohus, H. Heiselberg, P. Simonsen, A. Sørensen, K.C. Uncertainty of energy consumption assessment of domestic buildings, in: Proceedings of Building Simulation, 2009, pp. 1022–29.
[44] Hoes P, Hensen JLM, Loomans MGLC, de Vries B, Bourgeois D. User behavior in whole building simulation. Energy Build 2009;41(3):295–302.
[45] O'Brien, W. Occupant-proof buildings : can we design buildings that are robust against occupant behaviour?, in: Proceedings of Building Simulation, 2013, pp. 1746–54.
[46] Calleja Rodríguez G, Carrillo Andrés A, Domínguez Muñoz F, Cejudo López JM, Zhang Y. Uncertainties and sensitivity analysis in building energy simulation using macroparameters. Energy Build 2013;67:79–87.
[47] Silva AS, Ghisi E. Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation. Energy Build 2014;76:381–91.
[48] Azar E, Menassa CC. A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings. Energy Build 2012;55:841–53.
[49] Heiselberg P, Brohus H, Hesselholt A, Rasmussen H, Seinre E, Thomas S. Application of sensitivity analysis in design of sustainable buildings. Renew. Energy 2009;34(9):2030–6.
[50] Shen H, Tzempelikos A. Sensitivity analysis on daylighting and energy performance of perimeter offices with automated shading. Build Environ 2013;59:303–14.
[51] Laine, T. Forns-Samso, F. Katranuschkov, P. Hoch, R. and Freudenberg, P. Application of multi-step simulation and multi-eKPI sensitivity analysis in building energy design optimization, in: eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014, 2015, pp. 799–04.
[52] Hemsath TL, Bandhosseini KA. Sensitivity analysis evaluating basic building geometry's effect on energy use. Renew Energy 2015;76:526–38.
[53] Jin Q, Overend M. Sensitivity of facade performance on early-stage design variables. Energy Build 2014;77:457–66.
[54] Das P, Shrubsole C, Jones B, Hamilton I, Chalabi Z, Davies M, et al. Using probabilistic sampling-based sensitivity analyses for indoor air quality modelling. Build Environ 2014;78:171–82.
[55] Eisenhower B, O'Neill Z, Narayanan S, Fonoberov V a, Mezić I. A methodology for meta-model based optimization in building energy models. Energy Build 2012;47:292–301.
[56] Manfren M, Aste N, Moshksar R. Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation. Appl Energy 2013;103:627–41.
[57] Asadi S, Amiri SS, Mottahedi M. On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. Energy Build 2014;85:246–55.
[58] Gorissen D. Evolutionary model type selection for global surrogate modeling. J Mach Learn Res 2009;10:2039–78.
[59] Cheng M-Y, Cao M-T. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. Appl. Soft Comput. 2014;22:178–88.
[60] Mavromatidis LE, Marsault X, Lequay H. Daylight factor estimation at an early design stage to reduce buildings' energy consumption due to artificial lighting: a numerical approach based on Doehlert and Box–Behnken designs. Energy 2014;65:488–502.
[61] Tian W, Song J, Li Z, de Wilde P. Bootstrap techniques for sensitivity analysis and model selection in building thermal performance analysis. Appl Energy 2014;135:320–8.
[62] Van Gelder L, Janssen H, Roels S. Probabilistic design and analysis of building performances: methodology and application example. Energy Build 2014;79:202–11.
[63] Geyer P, Schlüter A. Automated metamodel generation for Design Space Exploration and decision-making – A novel method supporting performance-oriented building design and retrofitting. Appl Energy 2014;119:537–56.
[64] Naboni, E. Zhang, Y. Maccarini, A. Hirsh, E. Lezzi, D. Extending the use of parametric simulation in practice through a cloud based online service, in: Proceedings of first IBPSA-Italy conference BSA 2013, 2013, pp. 105–12.
[65] BRE Global Ltd, BREEAM: the world's leading design and assessment method for sustainable buildings, 2015. [Online]. Available: ⟨http://www.breeam.org/⟩. [accessed 07.07.15].
[66] Pohekar SD, Ramachandran M. Application of multi-criteria decision making to sustainable energy planning - A review. Renew Sustain Energy Rev 2004;8(4):365–81.
[67] Wang JJ, Jing YY, Zhang CF, Zhao JH. Review on multi-criteria decision analysis aid in sustainable energy decision-making. Renew Sustain Energy Rev 2009;13(9):2263–78.
[68] Bjørn E, Brohus H. Overall evaluation of indoor climate and energy for alternative office designs using the Eco-factor. J. Civ. Eng. Manag. 2006;12(1):43–9.
[69] Østergård, T. Maagaard, S.E. Jensen, R.L. A Stochastic and Holistic Method to Support Decision-Making in Early Building Design, in: Proceedings of Building Simulation [Submitted], 2015.
[70] Hopfe CJ, Augenbroe GLM, Hensen JLM. Multi-criteria decision making under uncertainty in building performance assessment. Build Environ 2013;69:81–90.
[71] Attia S, Hamdy M, O'Brien W, Carlucci S. Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design. Energy Build 2013;60:110–24.
[72] Nguyen A-T, Reiter S, Rigo P. A review on simulation-based optimization methods applied to building performance analysis. Appl Energy 2014;113:1043–58.
[73] Stevanović S. Optimization of passive solar design strategies: a review. Renew Sustain Energy Rev 2013;25:177–96.
[74] Jin, Q. Overend, M. Facade renovation for a public building based on a whole-life value approach, in: First Building Simulation and Optimization Conference, Loughborough, UK, 2012, pp. 378-385.
[75] Chantrelle FP, Lahmidi H, Keilholz W, El Mankibi M, Michel P. Development of a multicriteria tool for optimizing the renovation of buildings. Appl Energy 2011;88(4):1386–94.
[76] Attia, S. Computational optimisation for zero energy building design – Interviews with twentyeight international experts, International Energy Agency(IEA) Task 40: Towards Net Zero Energy Buildings Subtask B, 2012.
[77] Citherlet S. Towards the holistic assessment of building performance based on an integrated simulation approach. PhD Dissertation], Architecture Department. Lausanne: Swiss Federal Institute of Technology; 2001.
[78] Autodesk, Green Building Studio, 2015. [Online]. Available: ⟨https://gbs.autodesk.com/GBS/⟩. [accessed 17.06.15].
[79] Autodesk, Building Design Software | Revit Family. [Online]. Available: ⟨www.autodesk.com/revit⟩. [accessed 02.07.15].
[80] Graphisoft, EcoDesigner Star, 2015. [Online]. Available: ⟨http://www.graphisoft.com/archicad/ecodesigner_star/⟩. [accessed 07.07.15].
[81] Graphisoft, About ARCHICAD - A 3D CAD software for architectural design and modeling. [Online]. Available: ⟨http://www.graphisoft.com/archicad/⟩. [accessed 02.07.15].
[82] Sefaira, Sefaira, 2015. [Online]. Available: ⟨http://sefaira.com/⟩. [accessed 16.07.15].
[83] Integrated Environmental Solutions Ltd., IESVE. [Online]. Available: ⟨https://www.iesve.com/⟩. [accessed 27.06.15].
[84] National Renewable Energy Laboratory, OpenStudio, 2015. [Online]. Available: ⟨https://www.openstudio.net/⟩. [accessed 27.06.15].
[85] SketchUp, 3D for Everyone | SketchUp. [Online]. Available: ⟨http://www.sketchup.com/⟩. [accessed 02.07.15].
[86] Reinhart, C. DAYSIM Advanced Daylight Simulation Software. [Online]. Available: ⟨http://daysim.ning.com/⟩. [accessed 27.06.15].
[87] Fritz, R.M. McNeil, A. Radiance. [Online]. Available: ⟨http://www.radiance-online.org/⟩. [accessed: 01.01.15].
[88] Schlueter A, Thesseling F. Building information model based energy/exergy performance assessment in early design stages. Autom. Constr. 2009;18(2):153–63.
[89] Jakubiec, J.A. Reinhart, C.F. DIVA 2.0: Integrating daylight and thermal simulations using Rhinoceros 3D, Daysim and EnergyPlus, in: Proceedings of Building Simulation, 2011, pp. 2202–2209.
[90] Muehleisen, R.T. Craig, B. Integration of the CEN/ISO monthly building energy model into OpenStudio, in: ACEEE Summer Study on Energy Efficiency in Buildings, 2014, pp. 247–259.

[91] Doelling, M.C. Space-based thermal metrics mapping for conceptual low-energy architectural design, in: Building Simulation and Optimization BSO14, 2014.
[92] Lauridsen PKB, Petersen S. Integrating indoor climate, daylight and energy simulations in: parametric models and performance-based design. in: Proceeding of the 3rd international workshop on design in civil and environmental engineering; 2014. p. 111–118.
[93] Robert McNeel & Associates, Rhinoceros, 2014. [Online]. Available: ⟨https://www.rhino3d.com/⟩. [accessed: 07.07.15].
[94] Naboni, E. Integration of Outdoor Thermal and Visual Comfort in Parametric Design, in: Proceedings of the 30th International PLEA Conference, 2014, pp. 360–369.
[95] Lin S-H, Gerber DJ. Evolutionary energy performance feedback for design: multidisciplinary design optimization and performance boundaries for design decision support. Energy Build 2014;84:426–41.
[96] Banke T. Parametri i praksis, PhD Dissertation. Copenhagen: The Royal Danish Academy of Fine Arts Schools of Architecture, Design and Conservation; 2013.
[97] Davidson, S. Grasshopper - Algorithmic Modeling for Rhino. [Online]. Available: ⟨http://www.grasshopper3d.com/⟩. [accessed 27.06.15].
[98] Autodesk, Dynamo BIM, 2015. [Online]. Available: ⟨http://dynamobim.com/⟩. [accessed 17.06.15].
[99] Bentley, GenerativeComponents. [Online]. Available: ⟨http://www.bentley.com/en-US/Promo/Generative Components/default.htm⟩. [accessed 30.06.15].
[100] Roudsari, M.S. Honeybee for Grasshopper. [Online]. Available: ⟨https://github.com/mostaphaRoudsari/Honeybee/⟩. [accessed 27.06.15].
[101] Food4Rhino | Apps for Rhino and Grasshopper, Tortuga - LCA in Grasshopper. [Online]. Available: ⟨http://www.food4rhino.com/project/tortuga?ufh⟩. [accessed 30.06.15].
[102] Federal Ministry for the Environment Nature Conservation Building and Nuclear Safety, Ökobaudat | Sustainable Construction Information Portal. [Online]. Available: ⟨http://www.oekobaudat.de/en.html⟩. [accessed: 30.06.15].
[103] AEC-APPS, A Community-Driven Library of Apps for AEC Professionals, 2015. [Online]. Available: ⟨https://aec-apps.com/⟩. [accessed 21.01.15].
[104] Food4Rhino | Apps for Rhino and Grasshopper, Food4Rhino. [Online]. Available: ⟨http://www.food4rhino.com/?ufh⟩. [accessed 30.06.15].
[105] de Wilde P, Augenbroe G, van der Voorden M. Design analysis integration: supporting the selection of energy saving building components. Build. Environ 2002;37(8–9):807–16.
[106] Gudnason, G. Katranuschkov, P. Scherer, R.J. Balaras, C.A. Framework for sharing and re-use of domain data in whole building energy simulation, in: eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014, 2015, pp. 495–502.
[107] Negendahl K. Building performance simulation in the early design stage: an introduction to integrated dynamic models. Autom. Constr. 2014;54:39–53.
[108] National Renewable Energy Laboratory, Building Component Library, 2015. [Online]. Available: ⟨https://bcl.nrel.gov/⟩. [accessed 03.02.15].
[109] Long, N. Fleming, K. Brackney, L. An object-oriented database for managing building modeling components and metadata, In: Proceedings of Building Simulation, 2011, pp. 2356–2362.
[110] EnerLogic and James J. Hirsch & Associates, DOE2, 2012. [Online]. Available: ⟨http://www.doe2.com/⟩. [accessed 26.06.15].
[111] Brown, R.E. Walter, T. Dunn, L.N. Custodio, C.Y. Mathew, P.A. Berkeley, L. getting real with energy data : using the buildings performance database to support data-driven analyses and decision-making, in: ACEEE Summer Study on Energy Efficiency in Buildings, 2014, pp. 49–60.
[112] Gervásio H, Santos P, Martins R, Simões da Silva L. A macro-component approach for the assessment of building sustainability in early stages of design. Build. Environ 2014;73:256–70.
[113] Hiyama K, Kato S, Kubota M, Zhang J. A new method for reusing building information models of past projects to optimize the default configuration for performance simulations. Energy Build 2014;73:83–91.
[114] Pont, U. Ghiassi, N. Shayeganfar, F. Mahdavi, A. Fenz, S. Heurix, J. et al., SEMERGY: Utilizing semantic web technologies for performance-guided building design optimization, in: eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014, 2015, pp. 209–214.
[115] Lee, B.D. Sun, Y. Augenbroe, G. Paredis, C.J.J. Towards better prediction of building performance : a workbench to analyze uncertainty in building simulation, in: Proceedings of Building Simulations, 2013, pp. 1231–1238.
[116] ISO 13790, Energy performance of buildings – calculation of energy use for space heating and cooling, Geneva, Switzerland, 2008.
[117] onebuilding.org, Bldg-sim – Users of building energy simulation tools, 2015. [Online]. Available: ⟨http://onebuilding.org/guidelines.html⟩. [accessed 19.06.15].
[118] Venancio, R. Pedrini, A. Van Der Linden, A.C. Van Den Ham, E. Understanding envelope design: survey about architectural practice and building performance, in: Proceedings of Building Simulation, 2011, pp. 514–521.
[119] Abdullah, A. Cross, B. Aksamija, A. Whole building energy analysis: a comparative study of different simulation tools and applications in architectural design. In: ACEEE Summer Study on Energy Efficiency in Buildings, 2014.
[120] Danish Building Research Institute, Be10 (Danish), 2015. [Online]. Available: ⟨http://www.sbi.dk/miljo-og-energi/energiberegning⟩ [accessed 26.06.15].
[121] Danish Building Research Institute, BSim, 2015. [Online]. Available: ⟨http://sbi.dk/en/bsim⟩. [accessed: 26.06.15].
[122] Ahuja, S. Chopson, P. Energy Performance Calculator, 2014. [Online]. Available: ⟨http://zeroenergygreen.com/2014/06/12/urban-epc/⟩. [accessed: 05.07.15].
[123] University of Strathclyde, ESP-r. [Online]. Available: ⟨http://www.esru.strath.ac.uk/Programs/ESP-r.htm⟩. [accessed: 26.06.15].
[124] EQUA Simulation AB, IDA Indoor Climate and Energy. [Online]. Available: ⟨http://www.equa.se/en/ida-ice⟩. [accessed 26.06.15].
[125] Petersen, S. Hviid, C.A. iDbuild, 2013. [Online]. Available: ⟨http://www.idbuild.dk/⟩. [accessed 26.06.15].
[126] Velux, Daylight Visualizer. [Online]. Available: ⟨http://viz.velux.com/⟩. [accessed 27.06.15].
[127] A+E3D, AE3D. [Online]. Available: ⟨http://apluse.biz/⟩. [accessed 27.06.15].
[128] DesignBuilder Solutions Ltd, DesignBuilder. [Online]. Available: ⟨http://www.designbuilder.co.uk/⟩[accessed 27.06.15].
[129] EnerLogic and James J. Hirsch & Associates, eQUEST - the QUick Energy Simulation Tool. [Online]. Available: ⟨http://www.doe2.com/equest/⟩. [accessed 27.06.15].
[130] ExpertApp, N++. [Online]. Available: ⟨http://expertapp.com/npp.php⟩. [accessed 27.06.15].
[131] Granlund, RIUSKA. [Online]. Available: ⟨http://www.granlund.fi/en/software/riuska/⟩. [accessed 27.06.15].
[132] Solemma LCC, DIVA for Rhino. [Online]. Available: ⟨http://diva4rhino.com/⟩. [accessed 27.06.15].
[133] Zhang, Y. Korolija, I. jEPlus - An EnergyPlus simulation planner for parametrics. [Online]. Available: ⟨http://www.jeplus.org/wiki/doku.php⟩. [accessed 27.06.15].
[134] National Renewable Energy Laboratory, OpenStudio User Docs - Parametric Analysis Tool (PAT). [Online]. Available: ⟨http://nrel.github.io/OpenStudio-user-documentation/reference/parametric_studies/⟩. [accessed 27.06.15].
[135] Autodesk, Autodesk Building Performance Analysis | Help - Project Solon. [Online]. Available: ⟨http://help.autodesk.com/view/building_performance_analysis/enu/?guid=guid-80c2b132-8472-4c0b-a711-699ab22a4b90⟩. [accessed 27.06.15].
[136] Palonen M, Hamdy M, Hasan A. MOBO a new software for multi-objective building performance optimization. Proc. Build. Simul 2013:2567–74.

2.2 FOCUS IN THIS THESIS AND AT MOE

Here, we follow up on the concluding remark in Paper B which states a need for further research and development to realize the proposed “ideal” simulation framework. The required efforts to realize it are too extensive for this PhD project. Some of the work has therefore been split into various subtasks, which have been taken on by other MOE employees and students at Aalborg University. Figure 2-1 provides an overview of this allocation of subtasks, which we elaborate on below.

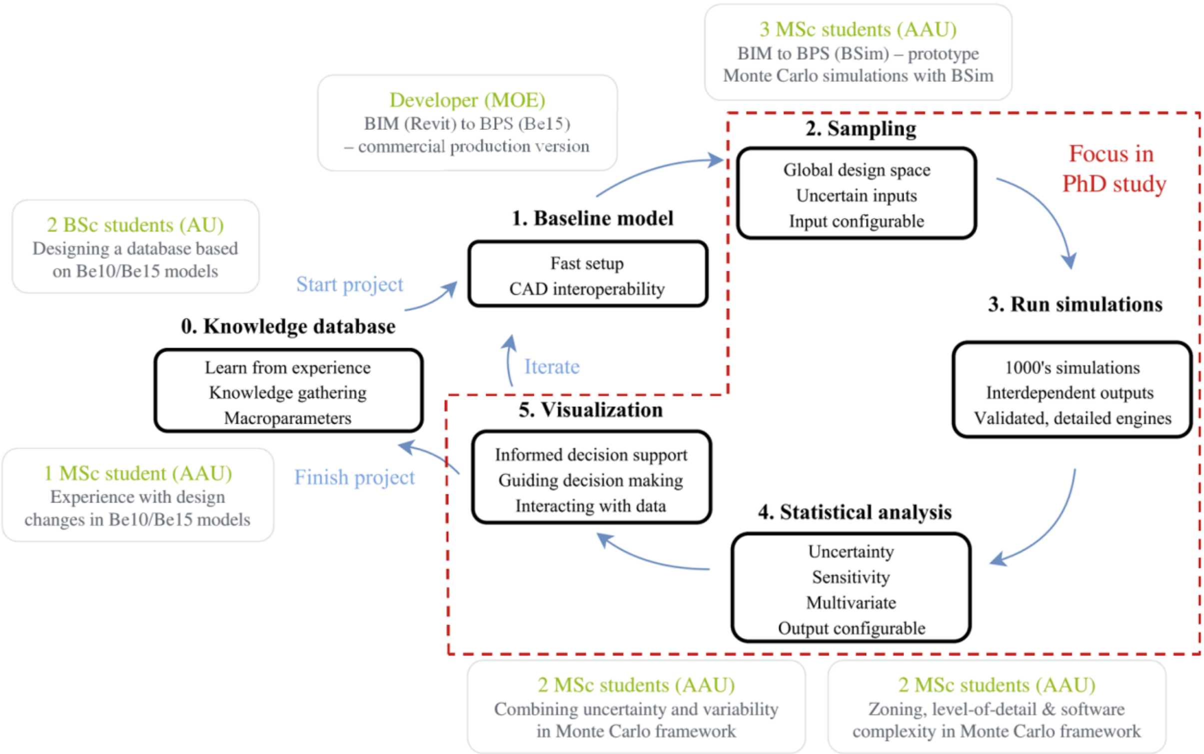


Figure 2-1 Overview of studies and developments related to the simulation framework proposed in Paper B. AU and AAU are abbreviations for Aarhus University and Aalborg University, respectively.

In the continued work of the PhD student, emphasis is on further development of the design approach with Monte Carlo simulations assisted by statistical methods and interactive visualizations (as indicated by the red dashed line in Figure 2-1Figure 2-1). This research area is believed to have the largest potential for increasing building performance and for extending the use of BPS during early design. In comparison, the other topics relate more to the challenges of reducing errors and time consumption. However, the preferred area is still too wide for the PhD project. As a result, three groups of thesis students at Aalborg University have looked at different research questions within this field. One group investigates how to deal with variability and uncertainty at the same time [9]. The former refers to design parameters, such as geometry and HVAC system, which can be defined by the design team, whereas the latter refers to uncertain parameters and circumstances related to user behavior and weather. Another group assesses the influence of different zoning, level of detail, and software complexity in a Monte Carlo framework [10]. The third group has developed a tool that allows for Monte Carlo simulations using the Danish BSim software [11]. Though, limited to a single zone it has revealed both potentials and obstacles of adding the ability of automated, multivariate simulations to existing BPS software.

Another important topic is the development of seamless links between CAD and BPS enabling fast model setup and updates. As explained in Paper B, this has received a lot of attention by software vendors. However, Danish building code requires specific measurement methods and it is mandatory to use the aforementioned Be15 software. A developer within MOE has developed an API for Revit facilitating fast setup and updates of Be15 models [12]. Time-consumption is reduced to minutes rather than days, which is normally the case for manual and error-prone setup in complex building projects. Still, this new tool cannot be applied in the early stages where a BIM (Revit) model has not yet be created.

A final note on the PhD related tasks concerns the knowledge based databases. An appealing example from Paper B is the Building Component Library [13]. Though, it applies only to OpenStudio and EnergyPlus and is therefore not quite applicable in a Danish context. Since Be15 is relatively simple, it allows for experimenting with database structure and knowledge gathering based on hundreds of projects from MOE. This has led to a couple of student projects which have revealed various difficulties [14][15]. For example, the logic behind design decisions and the actual performance of the constructed building do not appear from BPS models made with Be15 or BSim. Moreover, it is troublesome to provide experience based proposal for variability spans when the requirements are repeatedly tightened.

As alluded above, the BPS software used throughout the PhD project is Be15 and BSim. This contrasts the findings of Paper B, where Riuska, OpenStudio, and Honeybee have been pointed out. Part of the reason is that Riuska does not address daylight and Honeybee does not facilitate Monte Carlo simulations. OpenStudio has these capabilities but large scale analysis is still experimental and under development. Though, the main reason for choosing Be15 and BSim is that we, the PhD candidate and supervisors, are experienced users. Learning new, complex BPS software is a very time consuming task not to be underestimated.

3 BUILDING SIMULATIONS WITH A MONTE CARLO APPROACH

In this chapter, we elaborate on building simulations with a Monte Carlo approach. The preliminary study indicated a promising potential for this approach and, in the subsequent literature review, it was placed into a broader simulation framework. On this basis, we continue with a discussion of suitable ways to analyze and communicate the extensive amount of data from Monte Carlo simulations. This is followed by Paper C, which presents a refined simulation framework based on the experience from three sequential case studies. In Paper C, we also describe available sensitivity methods and we observe a potential for metamodeling. We subsequently experience a need for sensitivity methods which work with real-time analysis or rank inputs according to multiple outputs. Another lesson learnt, is the curse-of-dimensionality in building design that contains many design parameters and criteria. The use of rapid metamodels is considered a necessity. We therefore conclude with a comprehensive comparison of metamodeling techniques in Paper D including their application for different BPS scenarios.

3.1 FROM STATIC GUIDANCE TO REAL-TIME, MULTI-ACTOR DESIGN EXPLORATION

From practical experience, we know the importance of clear communication and intuitive visualization when collaborating with multiple decision-makers, which have diverse backgrounds and interests. This becomes particularly evident when dealing with a novel, stochastic approach, which differs considerably from a conventional, deterministic approach. In paper A, both scatterplots and histograms are used to visualize and analyze the simulations [4]. Since the number of inputs and outputs, denoted D, is often large, the resulting $D(D-1)$ scatterplots becomes unmanageable. Moreover, tendencies get obscured and difficult to observe – especially for non-experienced decision-makers. Instead, the histograms seem more appropriate, because they "sum up" the tendency for a given parameter and the number of histograms only grows linearly with the number of dimensions, D. However, the histograms shown in Paper A relate to static constraints of the simulation outputs. This calls for more interactive ways to analyse multivariate data in real-time during design meetings with multiple decision-makers.

With the above concerns in mind, a "Eureka" moment occurred when stumbling upon the D3.js JavaScript library [16]. This rich source provides free access to numerous interactive and data-driven visualizations for web-based applications. Especially, the parallel coordinate plot (PCP) is an ideal companion when exploring multivariate simulation data. Using "filters", a multi-actor design team can observe the consequences of applying different performance criteria or investigate specific "subspaces" of the input domain. An example of the PCP combined with histograms is shown on Figure 3-1. A detailed description of the usage and benefits of this interactive plot is given in Paper C, which follows this section. In general, novel interactive plots are likely to become more widespread in the building sector which is otherwise dominated by static documentation and reports. Another example is the use of a *dependency diagram* to explain complex input-output relations as described next.

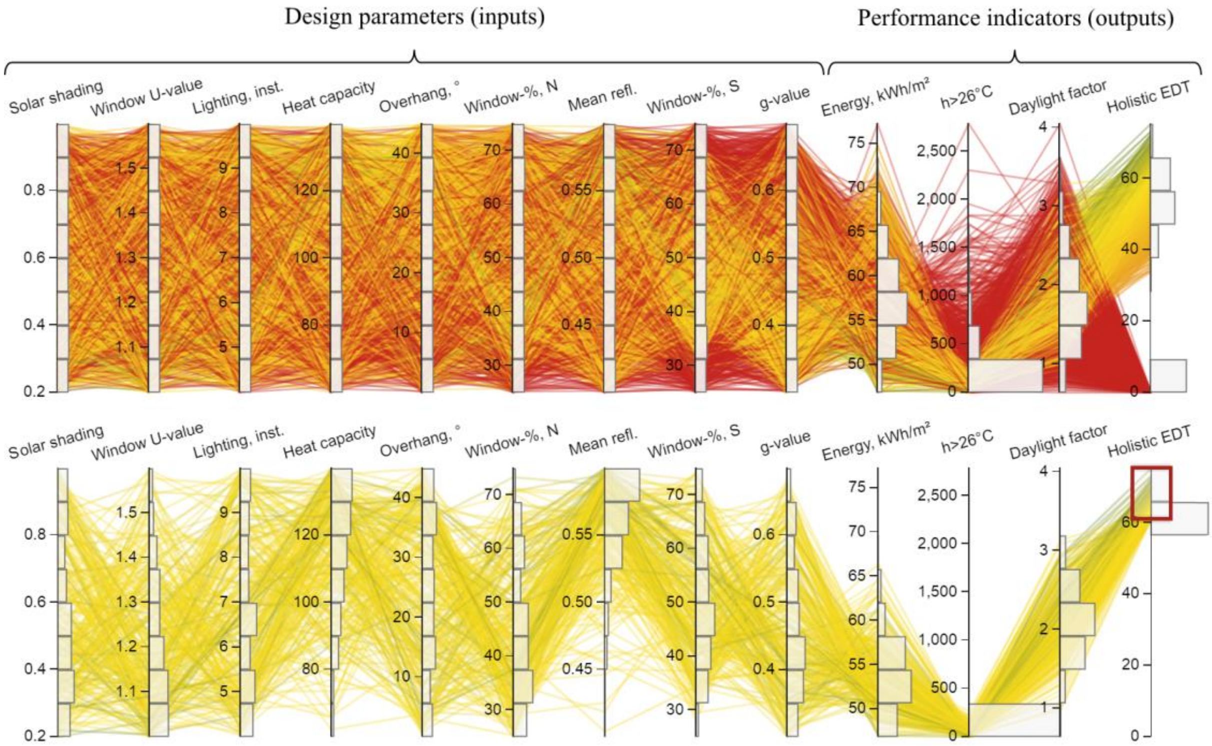


Figure 3-1: PCP for the case study from Paper A. The nine most sensitive inputs are included. On the bottom plot, a filter is added to include only the top 10% highest holistic EDT-scores.

In paper A, we briefly mentioned the necessity to keep in mind non-included performance indicators, such as aesthetics, functionality, and building costs. In subsequent work, we have created an interactive "dependency diagram", which helps illustrate the dependencies between design parameters and performance indicators – both quantitative and qualitative. Figure 3-2***Figure*** shows screenshots of this diagram. The green connections indicate dependencies between design inputs (grey) and qualitative performance indicators. These dependencies are based on experience but we cannot express the strength of these qualitative relations. In contrast, the blue connections represent dependencies between design inputs and quantitative indicators. Here, the varying line thickness relates to the strength of the correlation, which is based on first-order sensitivity analysis made for each of the three quantitative outputs. Thus, a thick line represents a strong relationship and vice versa.

The leftmost plot in Figure 3-2***Figure*** shows all connections and looks rather confusing, which emphasizes the complexity of building design. By "hovering the cursor" over a specific parameter, the rightmost plot displays only the connections relevant for that parameter [17]. We believe this interactive diagram helps bridge the gap between qualitative and quantitative measures. Moreover, it can be used to highlight the indicators, which are likely to be affected by a given design change, or to show inputs to focus on when needed to address a particular output.

More examples of interactive plots are available at the website buildingdesign.moe.dk. Now, we return our attention to multivariate building simulations aiding decision-making in early design.

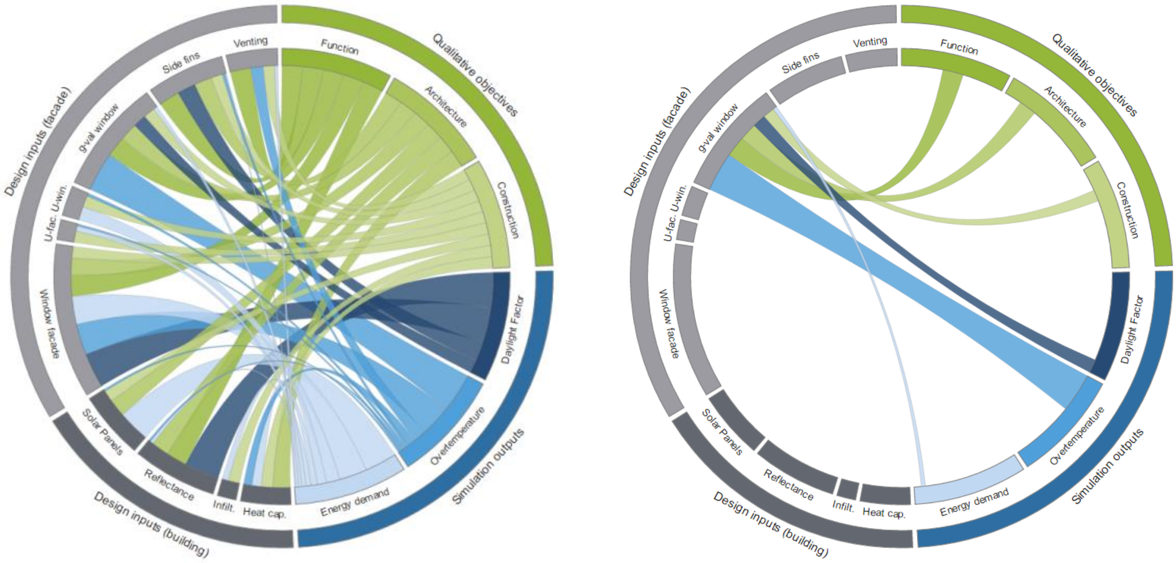


Figure 3-2: Screenshots of the interactive dependency diagram for a project-specific building [17]*. Left plot shows all connections are present. Right plot shows connections to performance indicators affected by the windows' g-value. The reader is encouraged to try it on buildingdesign.moe.dk.*

3.1.1 PAPER C

Paper C describes the development of a refined simulation framework, which relies on Monte Carlo simulations in combination with sensitivity analysis and metamodeling. The article is titled *"Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis"*, which has been published in Energy and Buildings, Volume 142, Pages 8 – 22, 2017.

Energy and Buildings 142 (2017) 8–22

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis

Torben Østergård [a,b,*], Rasmus L. Jensen [a], Steffen E. Maagaard [b]

[a] Aalborg University, Department of Civil Engineering, Thomas Manns Vej 23, DK-9220 Aalborg Ø, Denmark
[b] MOE Consulting | Engineers, Mariane Thomsens Gade 1C, DK-8000 Aarhus, Denmark

ARTICLE INFO

Article history:
Received 16 June 2016
Accepted 26 February 2017
Available online 2 March 2017

Keywords:
Early design stages
Building performance simulation
Building design
Sensitivity analysis
Monte Carlo simulations
Decision making support
Parallel coordinate plot
Multivariate analysis

ABSTRACT

This paper describes a novel approach to explore a multidimensional design space and guide multi-actor decision making in the design of sustainable buildings. The aim is to provide proactive and holistic guidance of the design team. We propose to perform exhaustive Monte Carlo simulations in an iterative design approach that consists of two steps: 1) preparation by the modeler, and 2) a multi-collaborator meeting. In the preparation phase, the simulation modeler performs Morris sensitivity analysis to fixate insignificant model inputs and to identify non-linearity and interaction effects. Next, a representation of the global design space is obtained from thousands of simulations using low-discrepancy sequences (LP_τ) for sampling. From these simulations, the modeler constructs fast metamodels and performs quantitative sensitivity analysis. During the meeting, the design team explores the global design space by filtering the thousands of simulations. Variable filter criteria are easily applied using an interactive parallel coordinate plot which provide immediate feedback on requirements and design choices. Sensitivity measures and metamodels show the combined effects of changing a single input and how to remedy unwanted output changes. The proposed methodology has been developed and tested through real building cases using a normative model to assess energy demand, thermal comfort, and daylight.

1. Introduction

The field of building design is often challenged by strict requirements for energy and costs that come together with ambitious goals for building performance. Moreover, the design team consists of multiple stakeholders, e.g. building owner, architects, engineers, and contractors, which have different opinions and areas of responsibility. To complicate matters further, design decisions made in the early stages have major impact on the final building costs and performance [7]. The early stages are characterized by large variability for a large number of design parameters which together form a vast design space. Despite the advancements of building performance tools, these tools typically provide deterministic results that *evaluate* the design rather than *guide* the design proactively [8]. This makes it almost impossible to explore the large design space sufficiently and it is difficult to guide decision makers – especially during the early stages.

To promote the use of building performance simulations (BPS) during the early stages, various researchers have developed novel tools with emphasis on reducing simulation time and complexity of inputs [9–11]. A different approach is to focus on establishing a more seamless interoperability between CAD models and BPS tools [12]. Another trend is the development of a complete integration or merging of performance simulations with CAD software. While such developments enable faster feedback at the early stages, most tools are still evaluative, give little or no guidance, and do not consider variability and uncertainty. Hygh et al. [13] argue that such point estimates are not meaningful for conceptual design. Likewise, Picco et al. [14] notice that the extreme accuracy and level of detail from sophisticated software is unrealistic during early design due to sparse information and large uncertainties. The aforementioned instant feedback allows the modeler to quickly perform a series of parameter variations around a baseline model or by varying one parameter at a time. However, such one-at-a-time (OAT) variations depend highly on the order, in which the parameters are varied, and on valid assumptions regarding the parameters kept fixed. For example, a design variation with a high window-to-facade-ratio might favor a low g-value whereas the opposite is true if the window-to-facade-ratio is low. The analysis is further complicated if other design parameters are still unknown and may vary. These

* Corresponding author at: Aalborg University, Department of Civil Engineering, Thomas Manns Vej 23, DK-9220 Aalborg Ø, Denmark. Tel.: +45 2540 0325.
E-mail address: to@civil.aau.dk (T. Østergård).

http://dx.doi.org/10.1016/j.enbuild.2017.02.059
0378-7788/

Nomenclature

Be10 or "idealized model"	Danish simulation software be10 based on ISO 13790 (here combined with a regression model for daylight)
BPS	Building performance simulation
EE	Elementary effect (in Morris method) [1]
g-value	Glazing's solar heat gain coefficient (SHGC)
LPτ	Low-discrepancy sequences by sobol [2]
MCF	Monte Carlo filtering
Morris	SA method aka. method of elementary effects [1]
OAT	One-at-a-time parameter variation
Overtemperature	Penalty output in idealized model [kWh/m^2 floor area]
PCP	Parallel coordinate plot
PEAR	Pearson's product-moment correlation coefficient used for SA
R^2	Coefficient of determination
RSA	Regionalized sensitivity analysis [3]
S_i	First order effect of X_i on Y (SA) [3]
S_{ij}	Second order effect of X_i and X_j on Y (SA) [3]
S_{Ti}	Total effect of factor X_i on Y (SA) [3]
SA	Sensitivity analysis
SDP	State-dependent parameter metamodeling and SA [4]
Sobol	Sobol's variance based sensitivity analysis (decomposition of variance) [5]
SPEA	Spearman's rank correlation coefficient used for SA
SRC	Standardized regression coefficients (from multivariate linear regression)
SRRC	Standardized rank regression coefficients
U-value	Heat transfer coefficient [W/m^2 K]
Win-fac-%	Windows-to-facade-ratio
μ^*	Mean of absolute elementary effects (Morris) [6]
μ	Mean of elementary effects [1]
σ	Standard deviation of elementary effects (Morris) [1]

can be internal loads, type of shading system, size of overhang, etc. Ultimately, the design team may draw wrong conclusions depending on the order of the parameters being varied and on the validity of the baseline model.

Sensitivity analysis may be applied to identify the order of which the design parameters should be addressed [15,16]. Though, in continuation of the above remarks on OAT variations, Saltelli and Annoni argue strongly against using local sensitivity measures obtained from OAT variations [17]. Instead, they advocate the use of global sensitivity analysis where model inputs are varied simultaneously. In addition, global sensitivity analysis can: (a) reveal interaction effects, (b) provide insight into input-output relationships, (c) reveal non-influential inputs, and (d) show regions of input space that meet certain criteria [3]. The application of sensitivity analysis in the context of building performance simulations is covered in depth by Tian (2013) [18].

One way to remedy the shortcomings of OAT parameter variations is to explore the global design space using Monte Carlo simulations. This approach facilitates global sensitivity analysis including *Monte Carlo filtering* which help identify regions of input space that meet certain criteria [19–21]. Some authors exploit the Monte Carlo simulations to create fast metamodels that may be used for design space exploration, optimisation, or sensitivity analysis [13,22–25]. The potentially time-consuming computational load required to run thousands of simulations may be overcome by using idealized models, creating fast metamodels, or by utilizing cloud computing [26].

Fig. 1 illustrates the conceptual difference between OAT parameter variations and global exploration when considering two or three inputs with large variability. The green areas indicate *high* performing design spaces whereas the yellow areas represent *acceptable* design spaces that, only just, meet the design criteria. In an OAT approach, a first simulation (1) is performed on basis of the early-stage geometric model. No matter how detailed or complex the simulation engine, this single deterministic evaluation gives no guidance on which parameter to adjust and in which direction. The modeler may start by varying the *g-value* (2 and 3) and conclude it has limited influence and thus keep it fixed at 0.4. Next, the modeler may choose to manually vary the windows' *overhang* (4–6) and find a suitable value around 30°. This design meets the requirements and may be proposed as a viable solution. However, the high performing solutions are not found unless the modeler returns his interest back to *g-value*. Now imagine that other parameters may also be varied, e.g. fenestration, internal loads, air change rates, shading system, etc. Changing a third variable may cause this cooling dominated design to be heating dominated instead. For example, if the *window-facade-ratio* is reduced from 70% to 50%, the acceptable and high performing areas change a lot (Fig. 1 bottom). Such a change or variation can make the previous OAT parameter variations invalid and misleading.

In a global approach, the modeler investigates a large design space through proper sampling and automated simulations (Fig. 1 right). This enables the design team to identify many high performing solutions in a large multidimensional design space. In our simplified example, 125 simulations identify the high performing solutions in a 3-dimensional space (Fig. 1 bottom-right). This global approach facilitate *global sensitivity analysis* which help the modeler to address the most influential design parameters while knowing which inputs are interdependent and which outputs are affected. Moreover, favorable input regions can be identified under *variable constraints* since the global design space is covered by exhaustive simulations. Note that this is not possible with optimisation routines which rely on fixed constraints. For example if the building owner changes the requirement for energy demand, due to higher ambitions, or add a new constraint on thermal comfort, the acceptable and high performing regions will most likely change. Similarly, optimisation cannot take into account qualitative objectives, such as aesthetics and constructability, which may be important to other stakeholders than the modeler.

The work presented here is part of an industrial PhD project in which the goal is to improve the use of building performance simulations (BPS) during the critical early design stages. For a complete background on this project, we refer to "Building simulations supporting decision making in early design – a review" [12]. The review also involves CAD-BPS interoperability, time-consuming model generation, limited reuse of knowledge, and optimisation. However, in this paper, we focus on the following topics:

- Tackling large variability and uncertainty in the early design stages
- Exploration of a vast design space with regard to multiple, conflicting requirements
- Support and guidance of multiple decision-makers

The outcome of this paper is a proposed methodology to perform, analyze, and visualize simulations of a global design space in order to guide decision-making in a multi-collaborator context. The method has been developed through a number of case studies and through in-depth theoretical analyses. In chapter 2, we present the lessons learnt from, both hypothetical and real, building cases along with feedback from different stakeholders. In chap-

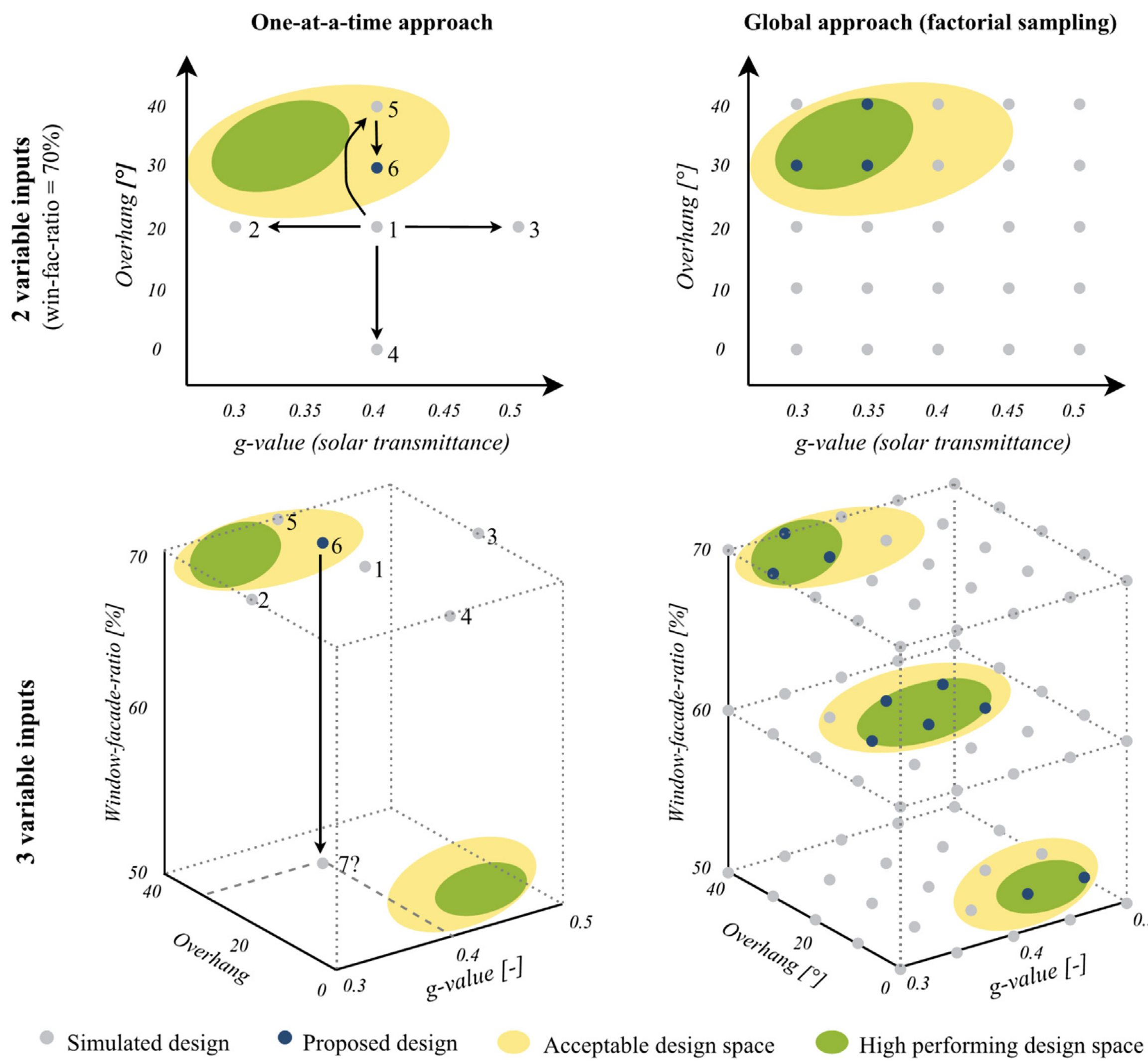


Fig. 1. Comparison of a One-at-a-time approach (left) and global design space exploration (right) when considering only two (top) and three variables (bottom). Yellow and green circles represent favorable design spaces in 2D plane (overhang – g-value). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ter 3, we describe the analyses made to identify the most efficient techniques in terms of sampling, sensitivity analysis, and visualization. In addition, a brief description is given of the software model used to simulate building performance. Readers with limited background in sensitivity analysis may benefit from reading chapter 3 before chapter 2. In chapter 4, we sum up and present the proposed methodology, its limitations, and future research.

2. Case studies

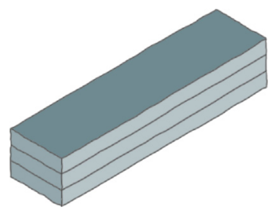
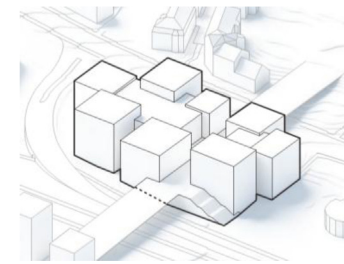
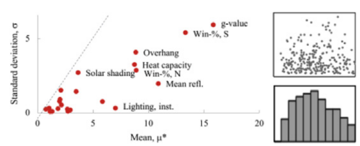
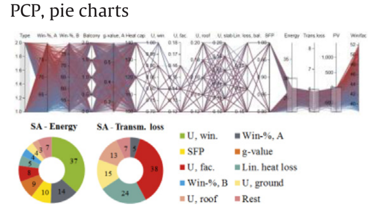
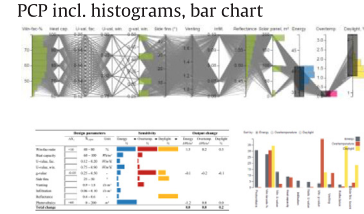
The proposed methodology has been developed through a series of case studies over the last two years. In this paper, the development has been divided into three phases to highlight the main objectives, the case buildings, and the lessons learnt. An overview is given in Table 1 which also includes visualizations that have been applied to analyze and communicate the many simulations and the derived sensitivity analyses and metamodels.

2.1. First step

In the development of the proposed design methodology, the initial step was to investigate a global design space with a holistic approach [27]. To evaluate energy demand and thermal comfort, we used an idealized model based on ISO 13790 [28] together with a regression model to estimate daylight. In general, early design decisions have major impact on these objectives. Moreover, daylight, thermal comfort, and energy demand are difficult to address since optimizing one often worsens another. Therefore, we suggested a holistic scoring function that combines their performance into one overall score.

From this initial work, we learnt that the global design space might preferably be explored by assigning uniform distribution to design inputs followed by Monte Carlo simulations using quasi-random sampling [2]. By applying Morris sensitivity analysis, the design parameters were ranked in accordance to their influence on the holistic score, i.e. the overall performance. The design team could then direct its attention to the most influential inputs. Perhaps, the most important finding was to apply *Monte Carlo filtering* which helped identify favorable input spans, e.g. illustrated by histograms. The advantage of the holistic approach became obvious when gradually applying filters that correspond to the requirements for each output. An example of this is shown on Fig. 2 which is based on stochastic simulations for a shoebox shaped office building. The figure shows the input distribution for the windows' overhang which was varied along with 17 other inputs. The distribution changed when we gradually removed simulations

Table 1
Overview of the development of the proposed design methodology (Sketches: authors, BIG architects, and EFFEKT architects). Identified developments points are addressed in the subsequent steps or in future work.

	First step	Second step	Third step
Building	2.700 m^2 office	24.000 m^2 residential building	15.000 m^2 educational center
Conceptual design sketches			
Methodology developments	• Morris method (SA) • Monte Carlo filtering (fixed) • Holistic scoring function	• Quantitative sensitivity analysis • Responsive Monte Carlo filtering • Multi-actor decision making	• Metamodels for: – Additional simulations – What-if questions
Applied visualizations	Morris plot, scatterplots, histograms  	PCP, pie charts  	PCP incl. histograms, bar chart 
Identified development points	• Constraints are fixed • Scoring functions are fixed • Difficult to explore subspaces	• Insufficient simulations • Effect of specific input *change* • Difficult to see distributions on PCP	• Qualitative RSA to show changes • Other simulation engines • Uncertainty analysis

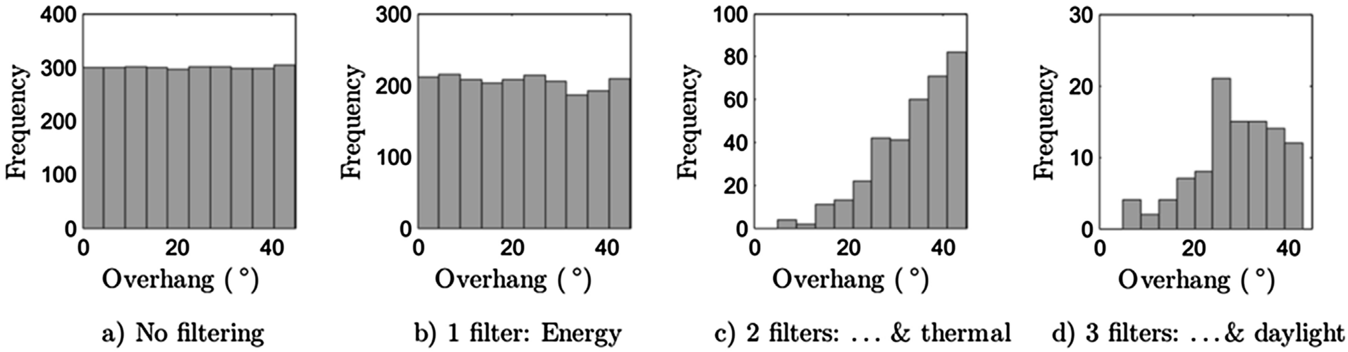


Fig. 2. Distribution of the design variable, *overhang*, when gradually adding constraints to a set of 3.000 stochastic simulations (from Ref. [27]).

not meeting the requirements related to energy demand, thermal comfort, and daylight. Ultimately, this design parameter had to be in the interval 5–43° and preferable higher than 20°. Knowing the construction costs of each design would add extra strength to this analysis. However, it may be near impossible to define valid cost functions for all design parameters during early design. For example, the overhang may be constructed in a great number of ways which affect its costs. In conclusion, it is recommended to include as many opposing quantitative objectives in the analysis and suggest favorable input spans instead a single deterministic solution. Suggesting input spans leave room to meet other qualitative requirements and other stakeholders' opinions. This first step had some shortcomings; the holistic score relies on weighting and the input spans and output criteria are kept fixed. This inflexibility makes it difficult to answer the many questions that arise during design meetings, i.e. "what if we lower the ambition for energy demand", or "what if we decrease the input span for the overhang to see possible solutions when there is almost no overhang?"

2.2. Second step

During the second step of our research, we learnt that the above shortcomings might be remedied using an interactive parallel coordinate plot (PCP) as shown on Fig. 3. This plot is very intuitive to use and the reader is encouraged to try it on *buildingdesign.moe.dk* [29]. Essentially, the interactive plot enables the design team to apply filters in real-time and thus immediately see the consequences of different requirements or design choices. Among other cases, the method was tested during the preliminary design of a complex multi-story residential building with a floor area of 24.000 m^2. Several stakeholders were present at the meetings: 2 building owners, 2 contractors, 3 architects, and 4 engineers. Despite the short introduction lasting a few minutes, everybody quickly understood the concept and eagerly participated in applying different filters to the PCP to identify input limits related to different design criteria. At the same time, quantitative sensitivity analysis was illustrated using pie charts to stress out the importance of certain inputs. Particularly, the sensitivity analysis demonstrated unexpectedly large impact of the *linear heat loss from balconies* on the *building transmission loss*. Subsequent feedback was very positive. The participants agreed that the method aided multi-actor decision-making and helped illustrate the most important design parameters. Also, the speed of the idealized model was advantageous since a new set of 10.000 simulations could be run during the two-hour meeting in which the baseline model had changed. However, despite the many simulations, the method failed to answer simple "what-if" answers such as "how much does the energy demand decrease if we add another 200 m^2 of photovoltaics". Moreover, it was sometimes difficult to see which inputs were affected when applying a filter. In particular, *discrete distributions* are troublesome since the user cannot visually see whether the plot shows one connecting line or many overlapping lines (notice the *discrete distributions* on Fig. 3).

2.3. Third step

To address the above issues, we added several features which help analyze the vast design space. First, the interactive parallel coordinate plot was combined with histograms (shown for some parameters on Fig. 3). The combined plot illustrates the distributions of the remaining simulations and thus removes the issue of overlapping lines related to discrete inputs. Moreover, the histograms helped indicate those input distributions that were affected the most by the applied filters. Presumably, quantitative regionalized sensitivity analysis (RSA) would help highlight the changes and thus help navigate the design team when exploring the design space through filtering. Finally, a metamodel constructed from multivariate linear regression helps answer how much a specific *input change* affects the outputs.

This enhanced approach was successfully tested during the conceptual design of a 15.000 m^2 educational institution. Moreover, this case was a good example of how early in the design process this methodology can be applied. Fig. 4 illustrates how the architects approached the contextual design. Using the proposed method with uniform input distributions, it was possible to perform simulations based on the simple conceptual design without knowing the distributions of windows, the size of overhangs, the type of shading systems, etc. (Fig. 4 right).

The design methodology has been applied to other real building cases and feedback has been very positive. The next steps involve integrating uncertainty analysis and quantitative regionalized SA and replacing the normative model with a more advanced simulation engine.

3. Development of the design method

This chapter covers the analyses and reasoning made during the development of the proposed design methodology. First, we describe the idealized model used for the stochastic simulations. Only a brief description is given since the design method is believed to work with most BPS software. Instead, emphasis is on sampling strategy, sensitivity analysis, and metamodeling.

3.1. Idealized simulation model

For this research, we use an idealized, normative model to evaluate energy demand, thermal comfort, and daylight. To evaluate whole-building energy demand, we use a monthly normative model [30] developed for code compliance in Denmark based on ISO 13790 [28]. An hourly-based model [31], also developed on basis

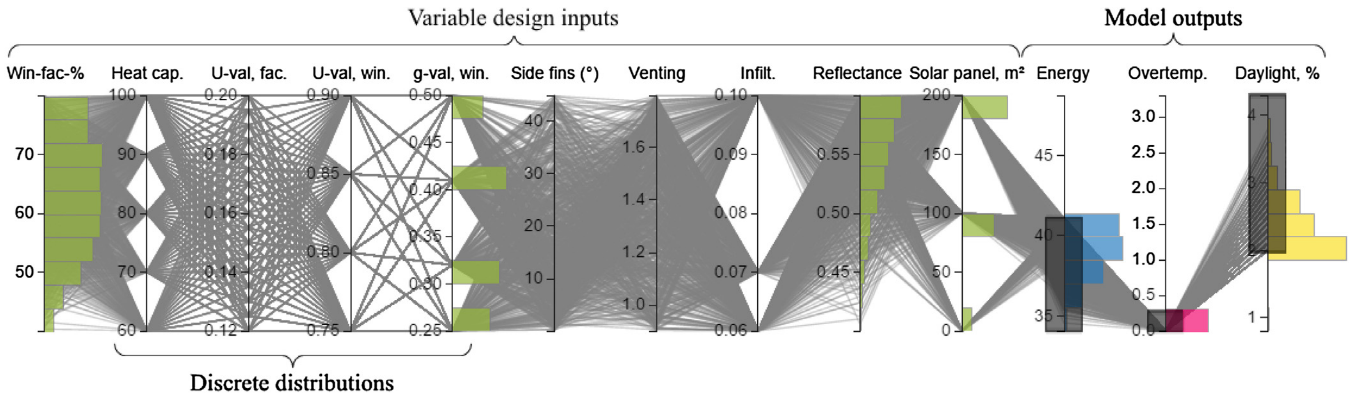


Fig. 3. The interactive parallel coordinate plot combined with histograms (from Ref. [29][29]). Only some histograms have been included to show the difference. Here, the initial 5.000 simulations for the educational institution are reduced to 988 when applying filters according to target criteria.

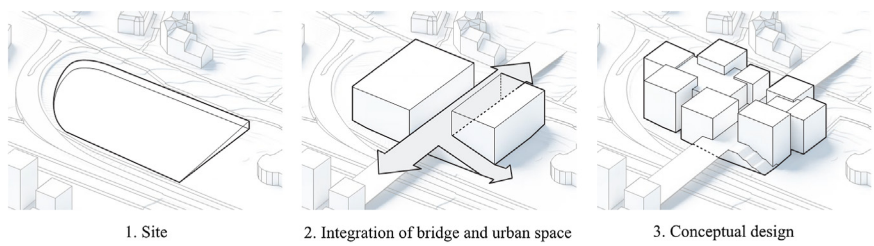


Fig. 4. The conceptual design contained a rough room layout but fenestration, shading, and more were not known. The global simulation approach enabled us to perform simulations at this early stage (Illustration: EFFEKT Architects).

of ISO 13790, has been used to assess thermal comfort by calculating the number of hours above 26 °C and 27 °C for critical rooms. Alternatively, *thermal comfort risk* is estimated from the penalty function, *overtemperature*, which penalizes designs for which the average building temperature exceeds 26 °C. To assess daylight, we are using a regression model created for building code compliance in Denmark [32]. In this paper, the combined, idealized model is referred to as *Be10* which is the name of the software using the two former models. This combined model matches the level of detail in the early design stages and they are computationally cheap. This allows us to run thousands of simulations in minutes rather than hours or days.

Simulation results obtained from idealized models may not precisely match the equivalent results from more advanced simulation software. However, exact values are not always necessary when comparing design options. It is more important that the ranking of results from simplified models match the ranking obtained from detailed software [33,34]. When rankings are consistent, the decision-makers will favor the highest ranked option regardless of the exact output values. For this work, the idealized model has been useful to test and develop the proposed design methodology. It is hypothesized that the methodology also works with complex, dynamic simulation engines.

3.2. Modeling and analyzing a global design space

The main idea is to run a large set of building performance simulations that adequately represent the vast design space available during the early design. Aided by sensitivity analysis, the design team explores this large set of simulations and tests a great number of different designs. The consequences of different design choices are then provided immediately. As shown on Fig. 5, the stochastic modeling and subsequent analysis consist of the following steps which may be iterated:

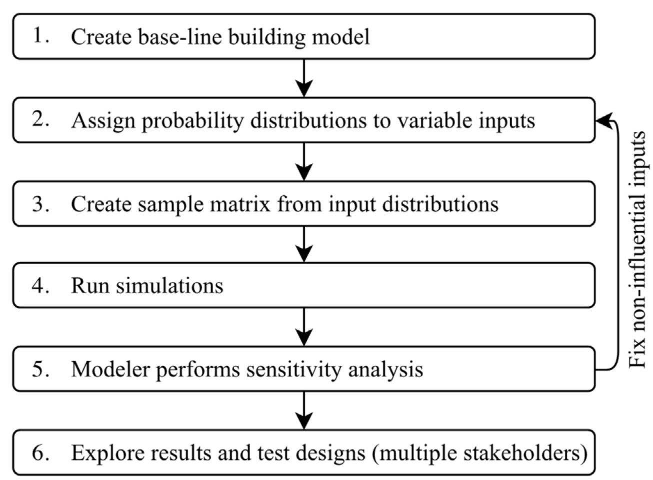


Fig. 5. Proposed workflow when modeling a global design space using Monte Carlo simulations.

1. A baseline building performance model is created in such way that important design parameters may be varied. For example, the modeler must consider how to vary window size, constructions, geometry, etc.
2. The design team assigns uniform distributions to the variable inputs. The number of variable inputs may range from a few to hundreds.
3. A sample matrix is constructed from the probability functions. Sampling strategy and the number of samples (simulations) depend on the model, the number of variable inputs, and the type of sensitivity analysis.
4. Building simulation is performed for each sample and relevant outputs are stored.
5. The modeler applies sensitivity analysis to analyze the large set of building simulations.

6. Guided by sensitivity analysis, the design team explores the design space and tests different designs.

The modeler is advised to carry out an initial sensitivity analysis (steps 2–5) to identify potential errors and reduce the design problem by fixating variable inputs with negligible influence. The final step (6) represents a design meeting in which multiple stakeholders explore the design space and make decisions assisted by sensitivity analysis. The design parameters must be addressed in accordance with their influence on the overall importance and with knowledge about their mutual interactions. The decision-makers will fixate or reduce the variation of the most influential design parameters. Following the meeting, the design can be further developed and explored within the limits of the remaining design space. When a refined design is available, the above workflow is repeated to identify which parameters have become most influential and to explore the design space related to the refined design.

3.3. Selection of parameters and sampling strategy

Prior to any modelling (CAD or BPS), the design team must consider how to evaluate building performance and how to set up variable design parameters. For instance, constraints towards energy demand, daylight, and thermal comfort may be assessed on whole-building level, floor level, or room level and different metrics may be applied. Similarly, input parameters can be varied in different ways. This applies for the distribution and sizing of windows which may be varied on whole-building level, facade level, or room level. It may be necessary to create *intermediate design variables* in order to vary a large number of simulation inputs simultaneously. An example of such a design variable could be *window percentage* used to vary the size of all windows and thus *bundle* these inputs together. Some simulation software allows for the creation of *macro-parameters* [35]. For example, OpenStudio's Measures can represent pre-defined constructions or entire HVAC systems. The use of such *bundles* or *macro-parameters* can be of assistance during early design to reduce the complexity and lower the level of detail.

The variations of the selected inputs can be defined using uniform distributions – either continuous or discrete. The combination of these variations represents the design space. Uniform

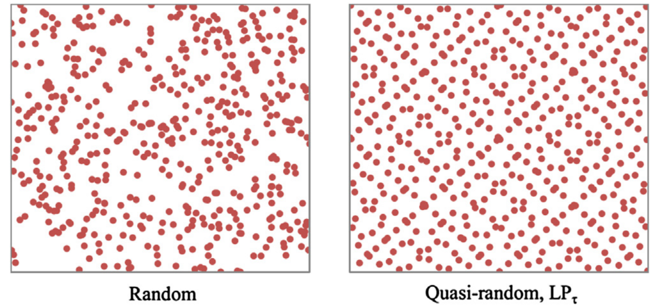


Fig. 6. Examples of 500 points made with random and quasi-random sampling, respectively.

input distributions are necessary to identify favorable input regions when applying Monte Carlo filtering (see e.g. Fig. 2). Moreover, uniform distributions cause poorly performing designs to be evaluated which may help convince the design team that some design strategies are disadvantageous or even impossible. A sampling strategy is required to create input samples from the assigned probability distributions. The most appropriate sampling strategy depends on the subsequent analysis. Some sensitivity analysis techniques require specific sampling procedures, such as Sobol's variance based method and the Morris method (see Table 2). Others only require generation of random or quasi-random numbers. The advantage of the latter is their ability to cover the design space more quickly and evenly as compared to the use of random numbers that results in lumps and gaps of the sampled inputs (Fig. 6). For this work, we apply Sobol's LP_τ low discrepancy sequences when modeling the design space [2].

Another aspect of the sampling strategy is sampling size, i.e. the number of Monte Carlo simulations to perform. Between 100 and 1.000 simulations are typically required to achieve stable, converged sensitivity measures [36,37]. Convergence depends on the sensitivity analysis method, the sampling method, and the model. However, a far greater amount of simulations are needed for a comprehensive representation of a global design space. To illustrate this, consider a *N*-dimensional model for which each variable input is discretized into *p* levels. Performing all simulations using full factorial sampling would require N^p simulations. Thus the exponential

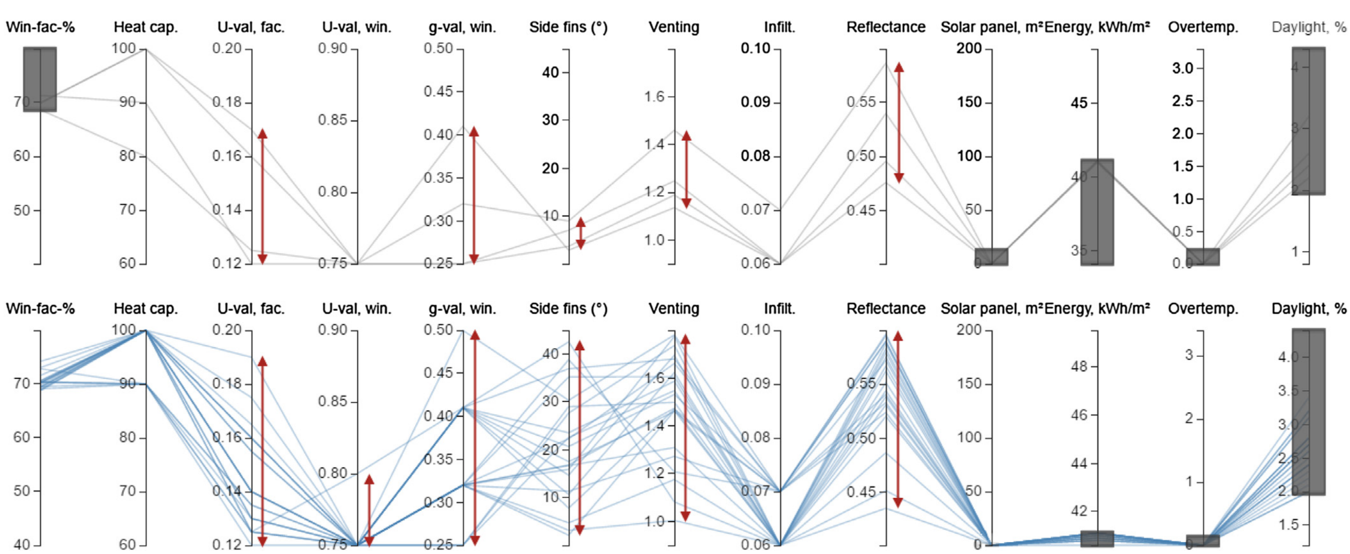


Fig. 7. Top: Only four simulations remain from initial 5.000 simulations after adding constraints to the three outputs and the inputs *win-fac-%* and *solar panel* (filters represented by black boxes). Bottom: When running another 5.000 simulations within the reduced spans for *win-fac-%* and *solar panel*, 22 simulations meet the criteria. The distribution spans are now particularly larger for the inputs with low influence on energy demand, i.e. *side fins* and *venting*.

Table 2
Comparison of global sensitivity methods (setup based on Ref. [39]).

	Correlation	Linear regression	RSA	SDP	Morris	Sobol's
Subtypes or aliases	Pearson's (PEAR) Spearman's (SPEA)	SRC SRRC	Monte Carlo Filtering Kolmogorov-Smirnov	Recursive splines	Elementary Effects	Variance Based
Computational effort[a]	Cheap m	Cheap m	Cheap m	Moderate m	Cheap $r\cdot(n+1)$	Very expensive $m\cdot(2\cdot n+2)$
Model dependence	Linear (PEAR) Monotonic (SPEA)	Linear (SRC) Monotonic (SRRC)	Independent	Independent	Independent	Independent
Sampling strategy	(Quasi) random	(Quasi) random	(Quasi) random	Quasi random LPτ	Morris	Sobol's
Interaction effects	No	No	(No)	Yes (to 2nd order)	Yes (qualitative)	Yes (quantitative)
Reliability	Only linear models	Only linear models	High (if m is not small)	High (to 2nd order)	High	Very high
Metamodel	No	Yes (linear, additive)	No	Yes (to 2nd order)	No	No
Characteristic and applicability	Easy to implement and understand	Easy to implement and understand Simple metamodel	Handles multiple objectives Fits well with PCP	Reliable 2nd order metamodel Difficult to compute	Preliminary screening (factor fixing)	Suitable when high reliability and interactions important

[a] Model runs requireded depends on number of factors n and sample size m. The sample size necessary to reach convergence depends on model complexity and sampling strategy (generation of random and quasi-random numbers). For Morris, r is the number of trajectories.

increase quickly exceeds millions of simulations in the early design stage due to the many variable design parameters. To remedy this, we strive to find a reduced set of samples that is still sufficiently large for the design team to test different scenarios and observe the consequences.

In the case studies, we performed between 3.000 and 10.000 simulations which seemed sufficient for these cases which involved 12–23 variable inputs. However, the design team may add additional filters to investigate a particularly interesting subspace until only a few simulations remain (see Fig. 7). The design team must be aware that the few remaining simulations show possible design solutions but they do not rule out other design possibilities. When so few simulations remain, we advise to perform a new set of simulations within the particular subspace and see if the limits and distributions of the simulations change. An example based on the aforementioned institutional center is shown on Fig. 7. The design team wanted to search the global design space for solutions that did not include *photovoltaics* and at the same time ensured a high *window-facade-ratio* of roughly 70%. Sensitivity analysis showed that these inputs affect energy demand considerably, i.e. *photovoltaics* represents 31%, and *win-fac-ratio* 27%, of the variance of *energy demand*. Surprisingly, the spans of less influential inputs, i.e. *side fins* and *venting*, were quite narrow when only four simulations remained (as illustrated with red arrows on Fig. 7). Another 5.000 simulations were run within the subspace with no *photovoltaics* and a high *win-fac-ratio*. Afterwards, these additional simulations showed that the inputs spanned greater than indicated on the initial plot (Fig. 7 bottom). The lessons learnt are to take advantage of knowledge gained from the sensitivity analysis and be careful when interpreting the parallel coordinate plot when only a few simulations remain. Therefore, to achieve more information, we advise to run an additional set of simulations for that particular subspace. Another approach is to compute "all" possible design combinations from the beginning, e.g. using full-factorial sampling for discrete inputs. This, however, quickly becomes an overwhelming task. Even for a simple thermal comfort model with only 9 inputs, this approach involved 7.5 million calculations [38].

3.4. Sensitivity analysis

Sensitivity analysis can be defined as the study of how the uncertainty in a model's output is related to the uncertainties in the model's inputs [3]. It has many uses related to model development, calibration, uncertainty analysis, and scenario analysis. In this work, we have looked for methods with the ability to:

- Simplify the design problem by identifying non-influential inputs (screening)
- Help the modeler to understand and calibrate the model
- Highlight inputs that deserves most attention in a multi-actor design process
- Identify regions of design space that meet design criteria
- Create reliable, fast metamodels

Various authors have compared sensitivity analysis methods and it can be concluded that no single method is superior and that many methods are valuable during the design phase [3,18,36,37,39–41]. Based on these works, characteristics and applicability of popular, global sensitivity methods are shown in Table 2. Below, we give a short description of some the methods including their applicability in the proposed design methodology.

3.4.1. Method of Morris

In the preparation phase, we recommend a preliminary screening analysis if the number of design variables and objectives are unmanageable large. The aim is, at different design stages, to

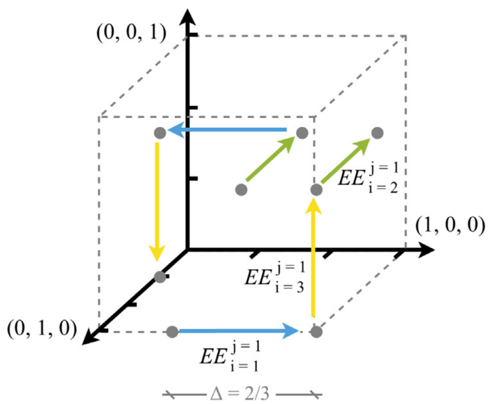


Fig. 8. Example of two trajectories (j) in a 3-dimensional design space where each design parameter is scaled to [0;1] and discretized into p = 4 levels. The blue arrows illustrate changes in the 1st dimension, i.e. the EE_1's, green in the 2nd, and yellow in the 3rd. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

identify and fixate insignificant inputs that have no or negligible influence on the building's performance (see *factor fixing* in [3]). For this purpose, the method of Morris [1] and its extended version [6] have been widely used in the field of building simulations [18]. The idea is to create *r* different trajectories in the *N*-dimensional design space (Fig. 8). For this space, each dimension is scaled to [0;1] and discretized into *p* levels by splitting their values into *p* quantiles. Each trajectory contains *N* + 1 calculations for which only one parameter changes at a time and with equally big steps, Δ. Ultimately, each input is related to *r* so-called *elementary effects* (*EE*) that describe the output change measured at *r* different places in the design space. Thus, the *elementary effect* for the i^{th} input factor in a point **X** can be defined as [3]:

$$EE_i = \frac{\begin{bmatrix} Y(X_1, X_2, \ldots, X_{i-1}, X_i + \Delta, \ldots, X_N) \\ -Y(X_1, X_2, \ldots, X_N) \end{bmatrix}}{\Delta}$$

where $\Delta \in [1/(p-1), \ldots, 1-1/(p-1)]$ is the input change. For each input *i*, we obtain the following three sensitivity measures:

- Mean of elementary effects: $\mu_i = \frac{1}{r}\sum_{j=1}^{r} EE_i^j$
- Mean of elementary effects' absolute values: $\mu_i^* = \frac{1}{r}\sum_{j=1}^{r} |EE_i^j|$
- Standard deviation of EE's: $\sigma_i^2 = \frac{1}{r}\sum_{j=1}^{r}\left(EE_i^j\right)^2$

If the mean of the absolute values of the *EE*'s, denoted μ^*, is large for the i^{th} input, then the i^{th} input has high influence, and vice versa. If the *EE*'s for the i^{th} input has a large standard deviation, denoted σ, then the model is either non-linear or the influence of the i^{th} input depends on the values of other inputs (interaction effects). The mean value, μ^*, is a good approximation of the total sensitivity measure, i.e. the total sensitivity of the parameter including the contributions from interaction with other parameters. Therefore, this measure is suitable for detecting non-influential inputs that may be left out of the subsequent exhaustive simulations.

3.4.2. Variance-based methods

If the modeler seeks a deeper understanding of the model, Sobol's variance decomposition can be applied to identify interaction effects[1] [3]. These effects describe how much particular combinations of inputs contribute to the variance of the output. Variance-based methods are often considered the most accurate and information-rich but they are computationally expensive [39]. High accuracy obtained at high computational costs are arguably not necessary during design meetings with non-experts. However, the modeler may apply variance-based methods to validate other sensitivity methods for the particular BPS software in use.

3.4.3. Correlation and linear regression

Since the methods of Morris and Sobol both rely on specific sampling strategies, they cannot at the same time be used to perform simulations that uniformly cover the design space. Instead, the global design space is modeled from Monte Carlo simulations with quasi-random sampling. Such simulations can easily be analyzed using either *correlation* or *linear regression* techniques which provide quantitative estimates of the inputs' linear effects on the outputs first order sensitivity indices, first order sensitivity indices, . These sensitivity indices can be converted into percentiles of their relative importance of the output variance. The percentages are easy to compare and they are intuitive to non-specialists. However, these methods do not account for interaction effects and thus only cover the variance proportional to the coefficient of determination, R^2. For example, if R^2 equals 0.90, the standardized regression coefficients (SRC) from linear regression explain only 90% of the total variance. From our experience, approximate sensitivity estimates or ranking of parameters are sufficient to guide decision-makers towards the most important design parameters.

3.4.4. Monte Carlo Filtering

Essential for the proposed design methodology is the use of Monte Carlo Filtering (MCF) also known as *factor mapping*. As described above, filtering of Monte Carlo simulations reveal areas of input space which are most likely to produce behavioral results meeting the desired criteria. The parallel coordinate plots and histograms are visually effective to show the consequences of applying filters. However, changes may be difficult to observe during a busy meeting with multiple team members when analyzing a model with many inputs and outputs. To remedy this difficulty, quantitative *regionalized sensitivity analysis* (RSA) can indicate which input distributions have been affected the most by the applied filters. For each input, we compare the distribution of *behavioral* simulations (those remaining after filtering), denoted f_b, with the distribution of *non-behavioral* simulations, denoted f_n [3]. If the distributions f_b and f_n for a given input, X_i, differ significantly, then this input is largely responsible for splitting the simulations into behavioral and non-behavioral realizations. This qualitative measure of the distributions' difference is obtained by applying the *Kolmogorov-Smirnov two-sample test* to each input. In contrast to the other methods, RSA can be applied to multiple outputs at the same time. In other words, RSA can display the combined changes on the input distributions when adding filters to multiple outputs.

[1] The conditional variance $V_{X_i}\left(E_{X_{\sim i}}(Y|X_i)\right)$ is called the first-order effect of X_i on Y and S_i is the first-order sensitivity index defined as $S_i = \frac{V_{X_i}\left(E_{X_{\sim i}}(Y|X_i)\right)}{V(Y)}$ [3]. The total sensitivity index S_{Ti} for X_i on Y may be defined as $S_{Ti} = 1 - \frac{V_{X_i}\left(E(Y|X_{\sim i})\right)}{V(Y)}$ or $S_{Ti} = S_i + \sum_{j>i}^{k} S_{ij} + \sum_{l>j>i}^{k} S_{ijl} + \cdots + S_{1\ldots i\ldots k}$ where the latter terms describe higher dimensions sensitivity indices from interaction effects [3,44].

3.5. Metamodeling

During design meetings, we encountered the following two issues despite having access to thousands of simulations:

1. Insufficient design space exploration: Sometimes only a few simulations remain, when applying filters to investigate a design subspace. In that case, there is a risk of drawing improper conclusions based on the parallel coordinate plot and histograms (Fig. 7).

2. What-if questions: Scatterplots, parallel coordinate plots, and histograms help analyze a multidimensional design space but they do not reveal *what* happens to the outputs *if* an input is changed by a specific amount. For example, "what happens if we increase the window-to-facade-ratio by 10%?"

In the first case, we immediately needed more simulations representing that particular subspace even though we had removed insignificant parameters using the Morris method. In the second case, we needed a rapid model that relates an input change to output changes in the multidimensional domain. To solve these issues, metamodeling is a viable solution as we will now demonstrate.

In general, metamodeling can be applied to construct a simplified, fast model from input-output relationships obtained from complex mathematical models such as those used in building performance software. A wide range of metamodeling techniques exist, such as regression, Kriging, Artificial Neural Network, and Support Vector Machines [42]. Similar to sensitivity analysis methods, each metamodeling technique has its pros and cons. Here, we demonstrate and compare the applicability of two methods, SRC (linear regression) and SDP (state-dependent parameter regression), which were also tested for sensitivity analysis purposes. Here, emphasis is on demonstrating the use of metamodeling in the proposed methodology whereas a comprehensive comparison study is needed to identify the most suitable technique in terms of reliability and applicability.

As case study, we adopt the aforementioned educational center for which 10 design parameters are varied simultaneously. First, Be10 is used to run 2.000 simulations which constitute the training set for the metamodels. For each metamodel, we make 2.000 new predictions of *energy demand*, *overtemperature*, and *daylight factor*. Fig. 9 shows how the predicted values compares to the corresponding values obtained from the original Be10 software. Judged by R^2, RMSE, and the points' closeness to the ideal $y = x$ line, the SDP metamodel outperforms SRC for all three objectives. Though, both models provide inaccurate estimates of the output *overtemperature* which is used to penalize designs with unacceptable high indoor temperatures. In the above description of the idealized simulation model, Be10, we argued that proper ranking of design options is more important than accurate simulation values (Section 3.1). Similarly, we will now argue that ranking is also more important than accurate values of metamodels when applying Monte Carlo filtering to investigate a design space.

3.5.1. Design space exploration using metamodels

As described earlier, we suggest using the parallel coordinate plot with histograms to quickly explore the design space and see how filtering affects the input distributions of the remaining simulations. Fig. 10 shows examples of such distributions after filtering the outputs corresponding to building code criteria. The unfiltered dataset for the idealized model (Be10) and the two metamodels each consist of 5.000 simulations with uniform input distributions. After the removal of simulations not meeting the requirements, the number of remaining simulations and their distributions differ considerably. This is mainly due to the metamodels' lacking ability to correctly represent the output *overtemperature*. Thus, too many simulations are removed when adding the constraint *overtemperature* ≤ 0 kWh/m^2. In comparison, only 220 and 406 simulations remain for the metamodels whereas 959 simulations remain in the case of the Be10 model (see grey highlight on Fig. 10). Another consequence is that the input distributions differ – particularly for *g*-value and *win-fac-ratio*. Hence, design decisions based on these distributions may be misguided. For example, the histogram for *g-value* based on SRC shows that low values are preferable since most remaining simulations lie in the bins representing low *g*-values (0.25 and 0.32). But the corresponding histogram, based on the original Be10 model, shows that most simulations remain in the third bin (*g*-value of 0.41) and many remain in the fourth bin (*g*-value of 0.5).

To remedy the above behavior, we try to adjust the filter values in such a way that an equivalent number of simulations are removed when applying a filter to a specific output. For example, when adding the constraint *overtemperature* ≤ 0 *kWh/m*^{2} to the Be10 dataset, 14.6% of the simulations get filtered out, whereas the same constraint removes 70.6% in the case of SRC. By adjusting the filter value to 0.66 for the SRC dataset, 14.8% of the simulations get filtered out, i.e. almost the same as 14.6%. Similar adjustments of filter criteria are made for both metamodels. The adjusted filters result in new histograms as shown on Fig. 11. Now, the distributions for each input are nearly indistinguishable when comparing across the different models. Even the easily computable SRC models perform well despite a low R^2 value of 0.42 for the output *overtemperature*. Conclusively, metamodels can be applied to remedy the issue of *insufficient design space exploration*. In this work, each calculation takes less than 0.5 s using the original Be10 software so additional simulations can be computed relatively fast. However, if computational demanding simulation software is used, metamodels provide plausible means to decrease the computing time of design space exploration.

3.5.2. Answering what-if questions using metamodels

To deal with "what-if" questions, we apply metamodeling in a different manner. The aim is to answer questions such as "what if we add another 100 m^2 of photovoltaics?" and "what if we increase the window-to-floor-ratio by 10%?" In other terms, what are the effects on Y_1, Y_2, etc. if we change input X_i by the amount ΔX_i? This issue appeared during a meeting related to the multi-story residential building. Despite having access to 10.000 simulations, we could not immediately answer such questions since all design inputs had been varied between simulations. To remedy this issue for the subsequent case, we constructed metamodels using SRC. The regression coefficients for SRC can be used to estimate how much the outputs are affected by a specific input change.

Fig. 12 shows how the SRC metamodels can be applied to illustrate the effects of input changes. The example is based on a request to know the consequences of raising the *win-fac-%* by 10 percentage points. The standard regression coefficients are used to calculate the averaged output changes in the multidimensional domain. Aided by the sensitivity measures, the design team may choose to counteract the increase in *overtemp.* by decreasing the windows' *g-value* by 0.05 (green boxes). Finally, the combined changes to the energy demand may be counteracted by adding 60 m^2 solar cells (grey boxes).

The validity of the predictions, obtained from this approach, relies on the accuracy of the metamodels and the complexity of the original simulation model. The accuracy of the SRC method can be assessed by the R^2 values.[2] The changes predicted by the SRC model are averaged changes obtained from the global design space. An example is to estimate the average effect of changing the *g-value* by 0.1 in a global design space in which the *window-to-facade*-ratio may be anywhere between 40 and 80%, the *heat capacity* between

[2] If the R^2 is low, a comparison of the measures μ and μ^* from Morris analysis can reveal which input that show non-linear relationships (see Fig. 15).

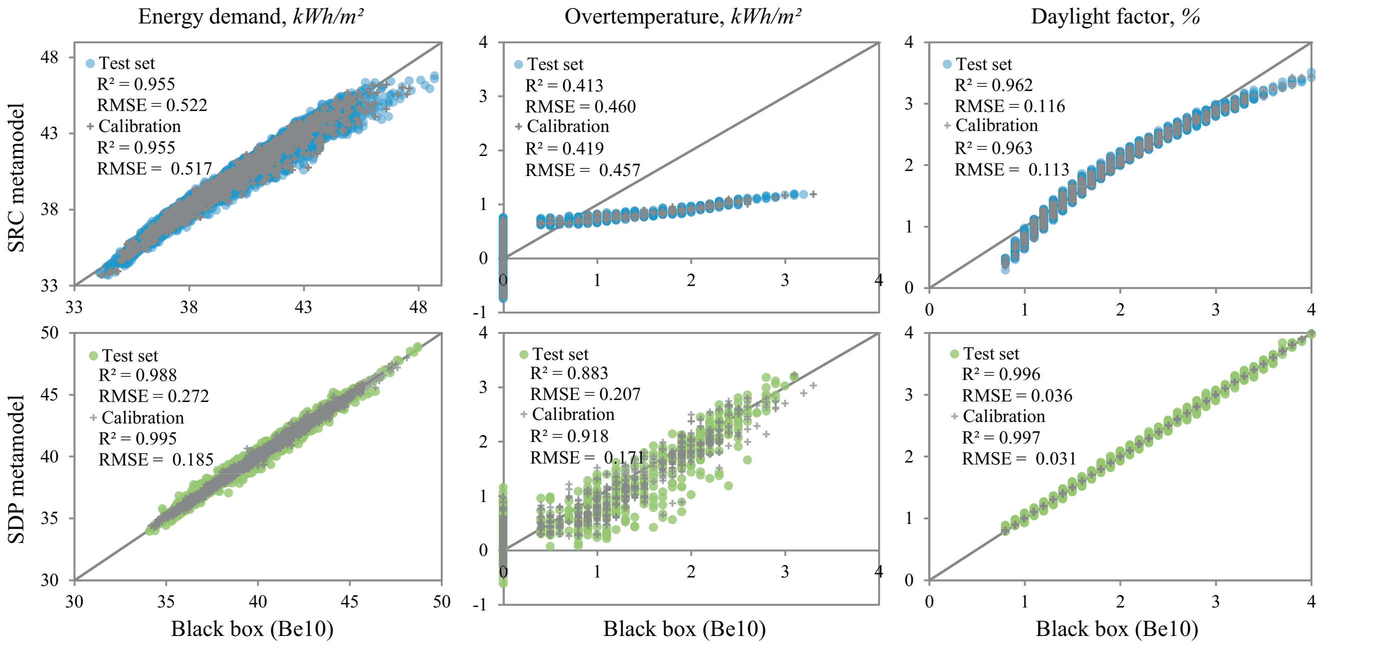


Fig. 9. Validation of two metamodeling techniques, SRC and SDP, applied to three building performance measures. Characteristics of a good metamodel include a high coefficient of determination, R^2, low Root-Mean-Square-Error (RMSE), and points close to the ideal line, metamodel = original model ($y = x$).

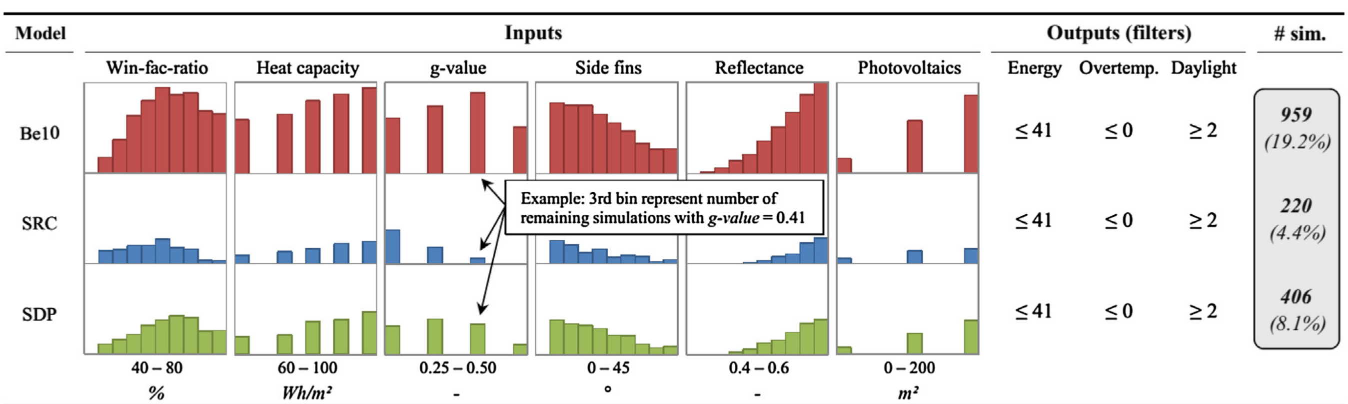


Fig. 10. Selected input distributions for the original model, Be10, and the two metamodels, SRC and SDP, when using the same filter criteria. The numbers of remaining simulations differ significantly (highlighted with grey box).

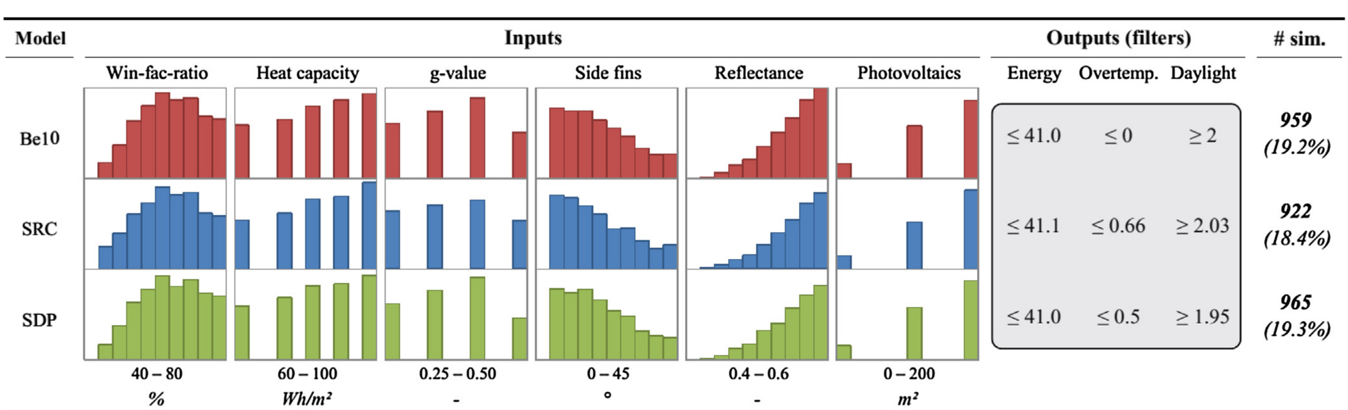


Fig. 11. Input distributions when adjusting the filter values such that the numbers of remaining simulations are roughly the same for all three models. The adjusted filter criteria are highlighted with a grey box.

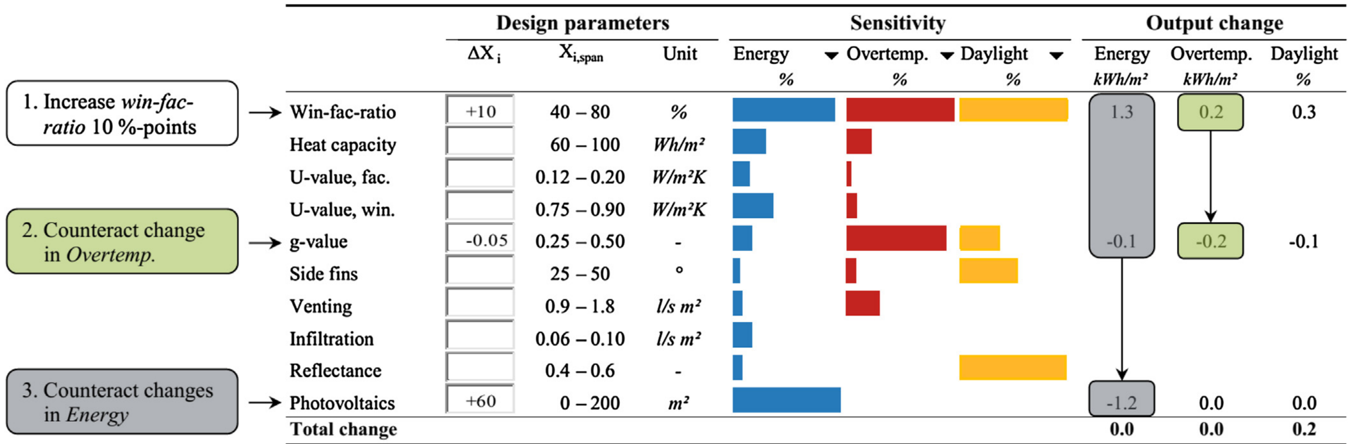


Fig. 12. Combining sensitivity and SRC metamodeling to observe and counteract consequences on input changes with regard to the performance objectives. Boxes highlight the applied changes and their consequences.

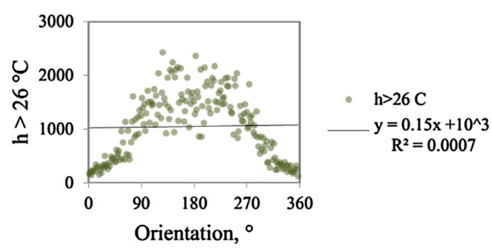


Fig. 13. Scatterplot show non-linear relationship between two parameters which cannot be described using linear regression.

80 and 120 Wh/K m^2, etc. These averaged changes calculated on basis of a global design space are more valid than calculating the change at a specific point in design space, e.g. at the design option where the *win-to-fac* is 40%; *heat capacity* is 120 Wh/K m^2, and so on. Even though SRC seems to work for this case, additional research is needed to assess the best applicable metamodeling technique and to identify appropriate training set sizes.

3.6. Scatterplots and non-linearity

A valuable supplement to the modeler's toolbox is the scatterplot. A scatterplot may reveal important characteristics, such as trends, form, outliers, and clusters, which are difficult to observe from statistical values from correlation, linear regression, or RSA. This is exemplified on Fig. 13 which shows the relationship between a design variable, *orientation*, and the number of hours with indoor temperature exceeding 26 °C. For this case, sensitivity analysis methods provide completely different results. Among eight varying inputs, *orientation* ranks last according to PEAR and SRC (<1%) as opposed to Morris in which *orientation* ranks first. The scatterplot shows a strong relationship, and a clearly visible form, which corresponds well with the Morris results. PEAR is based on Pearson product-moment correlation coefficient, *r*, which measures the linear relationship between two variables. In this case, Pearson's *r* is 0.026 which implies no relationship. Similarly, linear regression fails to describe the relationship which is evident from the R^2 value of 0.0007. However, in case of high-dimensional design problems, visual inspection of scatterplots becomes laborious and inefficient. The reason is that the possible combinations of scatterplots grow with N^2 with increasing dimensionality N. Often, the majority of such scatterplots show randomly scattered points and reveal little information (see the two right-most plots on Fig. 14). To identify the most informative input-output scatterplots, we recommend applying either Morris or SDP. As shown next, these reveal non-linear behavior and interaction effects – in contrast to PEAR and SRC.

3.6.1. Understanding non-linearity using scatterplots

In the following example, we consider a simple shoebox shaped building for which we calculate the energy demand while varying five input parameters. Fig. 14 shows scatterplots for each input in relation to the output. The two left-most plots show clear increasing tendencies, whereas the other distributions appear more random. Though, more information, which is well-hidden, is revealed when applying Monte Carlo filtering. First, Morris analysis shows difference between μ and μ^* and a high standard deviation, σ, of Elementary Effects for the input, *g-value* (Fig. 15). The difference between μ and μ^* means that a given change of *g-value*, at different points in the multidimensional space, can cause both negative and positive changes of the output. Similarly, this non-linear behavior can be detected by looking at the S_i and S_{ij} indices available from SDP and variance-based methods. For example, SDP analysis reveals that for the parameter *g-value* $S_i = 0.004$ whereas $S_{ij} = 0.024$

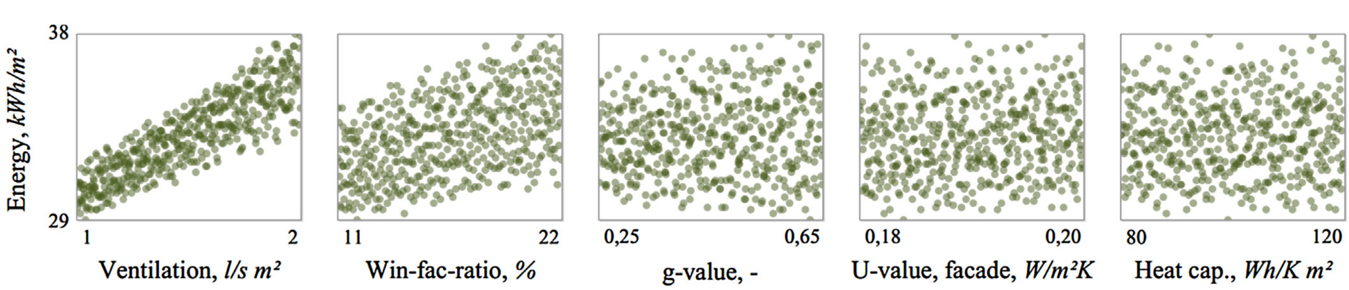


Fig. 14. Five scatterplots positioned left to right according to their influence on *Energy demand* as indicated by Morris. The transparency is set to 50% to show overlapping points.

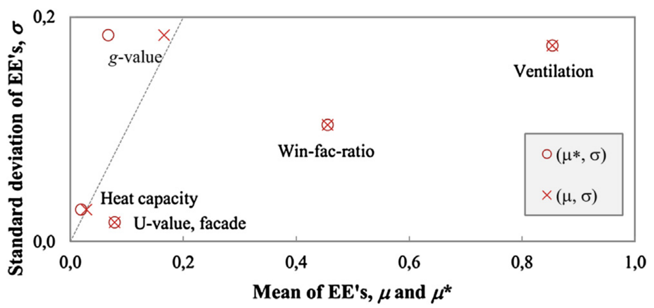


Fig. 15. According to the Morris sensitivity measures, the most important inputs are *ventilation* and *win-fac-ratio* whereas *g-value* contains high degree of non-linear and interaction effects.

for *g*-value and *ventilation*. This means that the variance of *g-value* in itself has almost no influence but its interaction with *ventilation* has some influence. We therefore investigate the relationship between *g-value* and *ventilation* for different subsets by applying filters to the most sensitive parameter, *ventilation*. Fig. 16 (right) shows how the slope of the regression line for *g-value* differs when applying two different filter criteria to *ventilation*. The shifting slopes indicate that the energy demand increases with *g-value* for low ventilation rates and vice versa. This emphasizes the importance of modeling a global design space and addressing the design parameters in accordance with their influence on the outputs. Conclusively, a suitable sensitivity analysis combined with filtering and visual inspection of scatterplots helps identify and understand non-linear behavior of models.

4. Results and further work

In the previous sections, we covered our experiences so far including the development of the proposed design methodology. We covered a wide range of statistical methods which differ in applicability and complexity. Fig. 17 sums up how the different methods and visualizations may be applied during the design process. The proposed method is iterative and contains two distinguishable parts: 1) a pre-meeting preparation part performed by the modeler, and 2) a multi-collaborator meeting for design space exploration and decision-making. Prior to these, it is assumed that a design concept has been formulated including the definition of target goals, variable design parameters, and a geometric model (as in Fig. 4 right).

During the meeting, the design team will gradually fixate or narrow the variability of the most important design parameters. Some parameters may be left free to vary in order to provide design 'freedom' for the next design iteration. Another outcome of the meeting is, hopefully, a better understanding of the model's behavior and the design parameters' influence on building performance. When a refined design is available, the proposed workflow may be repeated. For each iteration, the level of detail increases while the design space shrinks. Eventually, in the detailed design stages, the building physicist is less dependent on other stakeholders and may try to optimize on HVAC systems and control strategies or perform uncertainty analysis to consider the effects of weather and user behavior.

4.1. Further work

The methodology was tested, and gradually improved during three case studies, using an idealized model, Be10, to assess whole-building performance. The next step in our work will be to substitute the idealized model with detailed BPS software. Emphasis will be on how to deal with multiple zones and the increased number of simulation inputs and outputs. Possible initiatives are the definition of "macro-parameters" [35] and bundling of inputs and zones. To meet the increased computational requirements, we have already mentioned metamodels as a possible way to rapidly evaluate thousands of design options. As mentioned, this approach calls for a thorough comparison of metamodels techniques. This includes an investigation of the number of building simulations required to construct viable metamodels. Alternatively, cloud computing is becoming more widespread and may be the best choice for extensive simulations.

In the proposed workflow, we have considered *design variability* – also referred to as *design uncertainty*. Further research is needed to assess the design approach in terms of robustness of design decisions when also considering uncertainties related to user behavior, weather conditions, and modeling abstraction. One approach suggested by Rodríquz et al. [43] is to define different levels for both occupant load and weather load and then run global simulations for each combination. Comparison of different design options under uncertainty is not straightforward. Output distributions for different designs may be compared using histograms and statistics may help to estimate the level of confidence of a design choice [33]. However, this issue of comparison under uncertainty is further challenged when assessing not just one objective, such as energy demand, but a range of objectives (energy, daylight, thermal comfort, cost, etc.). Therefore, more work is needed to provide more confidence of design choices when considering uncertainty and multiple, opposing design criteria.

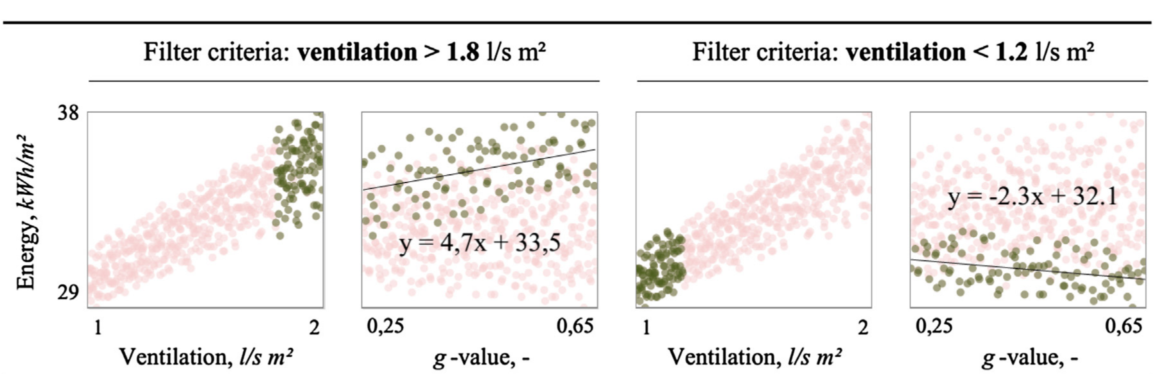


Fig. 16. The scatterplots show relationships between *energy demand* and two interdependent design parameters, *ventilation* and *g-value*. Green points represent simulations that meet the filter criterion. The slope of the regression line for *g-value* change sign when analyzing different subsets using filters on *ventilation*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

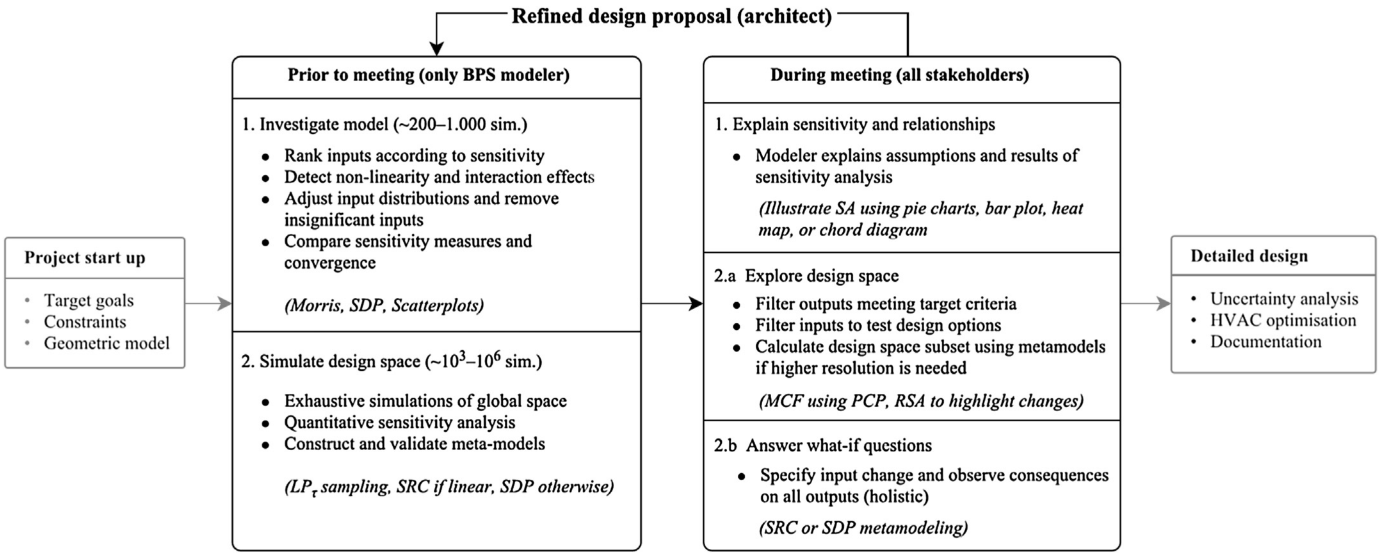


Fig. 17. In the iterative workflow (black), the BPS modeler generates an exhaustive set of simulations which is explored in plenum during design meetings. The architects refine the design and the workflow is repeated. Finally, the design evolves to the detailed design stages where the modeler may focus on HVAC systems, uncertainty analysis, and documentation.

5. Conclusion

In the section above, we described the workflow of the proposed design methodology along with issues that still remain. Here, we sum up the lessons learnt from the development of the methodology and the case studies:

- Building performance simulations are commonly computed using deterministic, non-linear, and complex simulation models. At the same time, building design is characterized by large number of variable design parameters which constitute a vast multidimensional design space. These circumstances speak for a global design exploration through Monte Carlo simulations.
- Global sensitivity analysis is necessary to address design parameters according to their influence on building performance. The modeler must have in-depth knowledge to select between various advanced methods and to interpret their results. During the multi-collaborator meetings, the complexity, and method-specific results, may preferably be communicated through simplified and intuitive visualizations.
- Prior to multi-collaborator decision-making, it is advisable to calculate important, opposing performance measures while making room for qualitative measures. A holistic, global approach helps ensure a high-performing and well-balanced design.
- The variability of the design parameters may be described using uniform distributions to represent the multidimensional design space evenly. This makes it easy to observe favorable regions of input space when using interactive parallel coordinate plots and histograms. The latter facilitate fast design space exploration, easy interpretation, and flexibility of design constraints.
- Metamodels can estimate how an input change affects multiple outputs in a global design space. In addition, metamodels can generate additional evaluations within a limited subset of the original design space. Despite a low accuracy, the simple metamodeling technique, multivariate linear regression (SRC), provided similar ranking as the original model. It could therefore be used for Monte Carlo filtering and decision support when modifying the design constraints accordingly.

This study was undertaken within the field of building performance simulations. However, the proposed modeling approach is relevant to other disciplines in which multi-collaborator decision-making is based on complex models with variable inputs.

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 MOE A/S.

References

[1] M. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (2) (1991) 161–174.
[2] I.M. Sobol', B.V. Shukman, Random and quasirandom sequences: numerical estimates of uniformity of distribution, Math. Comput. Model. 18 (8) (1993) 39–45.
[3] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis: the primer, John Wiley & Sons Ltd., Chichester, England, 2008.
[4] M. Ratto, A. Pagano, P. Young, State Dependent Parameter metamodelling and sensitivity analysis, Comput. Phys. Commun. 177 (11) (2007) 863–876.
[5] I.M. Sobol', Sensitivity estimates for nonlinear mathematical models, Math. Model. Comput. Exp. 1 (1993) 407–414.
[6] F. Campolongo, J. Cariboni, A. Saltelli, An effective screening design for sensitivity analysis of large models, Environ. Model. Software 22 (10) (2007) 1509–1518.
[7] G. Löhnert, A. Dalkowski, W. Sutter, Integrated Design Process: a guideline for sustainable and solar-optimised building design, IEA Int. Energy Agency, no. April, 2003.
[8] S. Attia, E. Gratia, A. De Herde, J.L.M. Hensen, Simulation-based decision support tool for early stages of zero-energy building design, Energy Build. 49 (2012) 2–15.
[9] B.J. Urban, L.R. Glicksman, A rapid building energy model and interface for non-technical users, Build. X Proc. (2007).
[10] A. Schlueter, F. Thesseling, Building information model based energy/exergy performance assessment in early design stages, Autom. Constr. 18 (2) (2009) 153–163.
[11] S. Petersen, S. Svendsen, Method and simulation program informed decisions in the early stages of building design, Energy Build. 42 (7) (2010) 1113–1119.
[12] T. Østergård, R.L. Jensen, S.E. Maagaard, Building simulations supporting decision making in early design–a review, Renewable Sustainable Energy Rev. 61 (2016) 187–201.
[13] J.S. Hygh, J.F. DeCarolis, D.B. Hill, S.R. Ranjithan, Multivariate regression as an energy assessment tool in early building design, Build. Environ. 57 (2012) 165–175.
[14] M. Picco, R. Lollini, M. Marengo, Towards energy performance evaluation in early stage building design: a simplification methodology for commercial building models, Energy Build. 76 (2014) 497–505.

[15] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinre, S. Thomas, Application of sensitivity analysis in design of sustainable buildings, Renewable Energy 34 (9) (2009) 2030–2036.
[16] H. Shen, A. Tzempelikos, Sensitivity analysis on daylighting and energy performance of perimeter offices with automated shading, Build. Environ. 59 (2013) 303–314.
[17] A. Saltelli, P. Annoni, How to avoid a perfunctory sensitivity analysis, Environ. Model. Software 25 (12) (2010) 1508–1517.
[18] W. Tian, A review of sensitivity analysis methods in building energy analysis, Renewable Sustainable Energy Rev. 20 (2013) 411–419.
[19] F. Ritter, P. Geyer, A. Borrmann, Simulation-based decision-making in early design stages, 32nd CIB W78 Conference 2015 (2015).
[20] E. Naboni, Y. Zhang, A. Maccarini, E. Hirsh, D. Lezzi, Extending the use of parametric simulation in practice through a cloud based online service, Proceedings of first IBPSA-Italy conference BSA 2013 (2013) 105–112.
[21] T. Laine, F. Forns-Samso, P. Katranuschkov, R. Hoch, P. Freudenberg, Application of multi-step simulation and multi-eKPI sensitivity analysis in building energy design optimization, in eWork and eBusiness, Architecture, Engineering and Construction – Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM, 2014 (2015) 799–804.
[22] B. Eisenhower, Z. O'Neill, S. Narayanan, V.a. Fonoberov, I. Mezić, A methodology for meta-model based optimization in building energy models, Energy Build. 47 (2012) 292–301.
[23] P. Geyer, A. Schlüter, Automated metamodel generation for Design Space Exploration and decision-making–a novel method supporting performance-oriented building design and retrofitting, Appl. Energy 119 (Apr. 2014) 537–556.
[24] S. Asadi, S.S. Amiri, M. Mottahedi, On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design, Energy Build. 85 (2014) 246–255.
[25] L. Van Gelder, H. Janssen, S. Roels, Probabilistic design and analysis of building performances: methodology and application example, Energy Build. 79 (2014) 202–211.
[26] D.L. Macumber, B.L. Ball, N.L. Long, A graphical tool for cloud-based building energy simulation, 2014 ASHRAE/IBPSA-USA Building Simulation Conference (2014) 87–94.
[27] T. Østergård, S.E. Maagaard, R.L. Jensen, A stochastic and holistic method to support decision-making in early building design, Proceedings of Building Simulation (2015) 1885–1892.
[28] CEN, ISO 13790:2008 Energy performance of buildings – calculation of energy use for space heating and cooling, Geneva, Switzerland, 2008.
[29] MOE A/S, Demonstration of Proactive Building Simulations, 2016. [Online]. Available: http://buildingdesign.moe.dk/PhD-Project/Demonstration-of-Proactive-Building-Simulations. (Accessed: 26 May 2016).
[30] SBI. – Danish Building Research Institute, SBi anvisning 213 – Bygningers energibehov, Hørsholm, 2014.
[31] L.H. Mortensen, S. Aggerholm, Simplified hourly method to calculate summer temperatures in dwellings, 33rd AIVC and 2nd TightVent Conference (2012).
[32] K. Johnsen, J. Christoffersen, SBi-anvisning 219 Dagslys i rum og bygninger, Hørsholm, 2008.
[33] R. Rezaee, J. Brown, G. Augenbroe, J. Kim, A new approach to the integration of energy assessment tools in CAD for early stage of design decision-making considering uncertainty, eWork and eBusiness in Architecture, Engineering and Construction – Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM, 2014 (2015) 367–373.
[34] J.-H. Kim, G. Augenbroe, H.-S. Suh, Comparative study of the leed and iso-cen building energy performance rating methods, Proc. Build. Simul. (2013) 3104–3111.
[35] H. Gervásio, P. Santos, R. Martins, L. Simões da Silva, A macro-component approach for the assessment of building sustainability in early stages of design, Build. Environ. 73 (2014) 256–270.
[36] S. Burhenne, Monte Carlo Based Uncertainty and Sensitivity Analysis for Building Performance Simulation, Karlsruhe Institute of Technology, 2013.
[37] A.-T. Nguyen, S. Reiter, A performance comparison of sensitivity analysis methods for building energy models, Build. Simul. 8 (6) (2015) 651–664.
[38] T. Østergård, R.L. Jensen, S.E. Maagaard, Thermal comfort in residential buildings by the millions – early design support from stochastic simulations, CLIMA 2016 – Proceedings of the 12th REHVA World Congress vol. 6 (2016).
[39] J. Yang, Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis, Environ. Model. Software 26 (4) (2011) 444–457.
[40] R. Confalonieri, G. Bellocchi, S. Bregaglio, M. Donatelli, M. Acutis, Comparison of sensitivity analysis techniques: a case study with the rice model WARM, Ecol. Modell. 221 (16) (2010) 1897–1906.
[41] F. Pianosi, K. Beven, J. Freer, J.W. Hall, J. Rougier, D.B. Stephenson, T. Wagener, Sensitivity analysis of environmental models: a systematic review with practical workflow, Environ. Model. Software 79 (2016) 214–232.
[42] S. Razavi, B.A. Tolson, D.H. Burn, Review of surrogate modeling in water resources, Water Resour. Res. 48 (7) (2012).
[43] G. Calleja Rodríguez, A. Carrillo Andrés, F. Domínguez Muñoz, J.M. Cejudo López, and Y. Zhang, Uncertainties and sensitivity analysis in building energy simulation using macroparameters, Energy Build. 67 (2013) 79–87.
[44] B. Eisenhower, Z.O. Neill, V.A. Fonoberov, I. Mezi, Uncertainty and sensitivity decomposition of building energy models, J. Build. Perform. Simul. 5 (3) (2012) 171–184.

3.2 AFTERMATH

In the "Further work" section of Paper C, we outlined several tasks to be addressed posterior to this paper. Some of these have already been discussed in section 2.2. This includes investigation of uncertainty, zoning, and level of detail, which have since been addressed in various Master's theses using advanced BPS software. Here, we focus our attention to sensitivity analysis and metamodeling after a brief discussion of the lessons learnt from the work presented in Paper C.

3.2.1 LESSONS LEARNT

For the second case study of Paper C, we asked the design team to answer a questionnaire on the use of building simulations. Both questionnaire and feedback from five participants are shown in Appendix A. Here, we sum up some of the feedback. Firstly, there was broad agreement that the most important property of building simulation tools is "intelligence", i.e. the ability to aid decision making. This corresponds well to a survey of roughly 450 architects and engineers, where a majority of the architects favored "intelligence" as a tool selection criterion [18]. The engineers preferred, however, "accuracy and the ability to simulate complex components". Regarding which building performance criteria should be prioritized, the responses to our questionnaire may be summed up in the following ranking: energy, thermal comfort, costs, daylight, air quality, environmental impact, and acoustics. This fit well with the prioritization made in this project, where emphasis is on energy demand, thermal comfort, and daylight. As mentioned, building costs are troublesome to quantify in early design characterized by great "variability". Another valuable feedback is that the respondents appreciated the sensitivity analysis and considered it important. In general, the feedback suggested that the proposed simulation method and the real-time analysis have great potential and aid collaboration. These responses are consistent with the oral feedback received after presentations to various building investors, architects, and researchers.

Despite the potential of Monte Carlo simulations, we have learnt a few lessons regarding the cumbersome definition and sampling of inputs. These conditions relate to the *correlation of inputs*, the *aggregation of input distributions*, and the *heterogeneity of sub settings* which we will elaborate on next.

In all case studies, the sampling strategy implies independent inputs even though some inputs are correlated. For example, a variation of the windows' *Solar Heat Gain Coefficient* (SHGC) affects the windows' *Light Transmittance* and *Thermal Resistance*. There are different, software-dependent, ways to handle this. If such inputs are defined separately (e.g. in Be15), the modeler may allow one input to vary freely and describe dependent inputs using regression formulas or a correlation matrix. Such dependencies may be derived from a materials database or product catalogues. For other BPS tools (e.g. BSim), such properties are defined for a specific construction element (macro-parameter). In that case, the modeler must define a number of elements, e.g. windows, with varying properties. Another "correlation" is the connection between a variable window-facade-ratio and the resulting facade area, which may need to be defined explicitly. A final remark on these correlations is that they may be troublesome to handle correctly during sensitivity analysis.

In Paper A and C, uniform probability distributions have been used to model the variability of design parameters. In this regard, the modeler must be aware of how uniform distributions "aggregate" or "sum up". For example, the building's window-to-wall-ratio (WWR) may be chosen to vary uniformly in the range from 40 to 80%. Alternatively, the modeler may choose to vary the north-oriented windows separately from south-oriented windows – each in the range from 40 to 80%. Similarly, WWR may be varied individually for North, South, East, and West. When sampling with these three approaches, the resulting distributions of the buildings average WWR differs significantly as shown on Figure 3-3Figure . The building facades are assumed to be equally large in all directions. This behavior is worth to remember, if the intention is to sample the "overall" or "aggregated" design parameter uniformly, i.e. with equal probability across its range. Moreover, this exemplified division of a design parameter into multiple design parameters also affects sensitivity analysis. That is, if 50 windows are varied individually in the same range, each window produces a negligible contribution to the variance of the simulation output. Sensitivity analysis would then consider them all as insignificant. If they were varied as a group, this could be the most important (sensitive) design parameter.

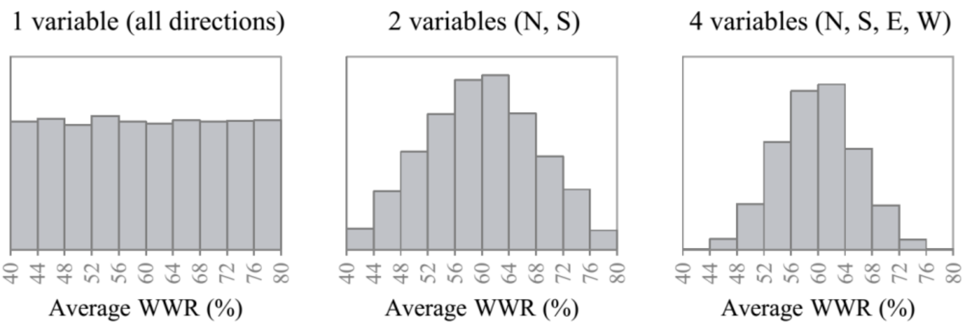


Figure 3-3: Distribution of 10.000 samples of the building's average window-to-wall-ratio (WWR) when varying WWR for one, two, and four uniformly distributed parameters, respectively.

The final "lesson learnt" is referred to as the *heterogeneity of sub settings*. Consider a scenario, in which the design team wishes to vary multiple parameters including a *ventilation strategy* with three options: 1) mechanical ventilation only, 2) mechanical ventilation with cooling, and 3) mechanical ventilation and venting. Each of these settings may induce important sub settings which should also be varied. For the second option with cooling, important sub settings could be *coefficient-of-performance*, *cooling power*, and *set point*. For the third option with venting, important sub settings might be *opening area* and *schedule*. Obviously, such heterogeneous sub settings are difficult to handle in a Monte Carlo simulation framework. Later, we give an example on how to circumvent this issue for a scenario, in which the mechanical cooling, with one sub setting, is either "on" or "off".

3.2.2 SENSITIVITY ANALYSIS

As mentioned, feedback has been positive when using sensitivity analysis to reveal important and insignificant design parameters. Paper C provided a brief overview of popular sensitivity analysis methods. Despite their different assets, there are still features missing that would be beneficial in a building design context. First of all, building design involves multiple performance criteria but popular sensitivity methods only addresses one output at a time. Thus, we need a way to prioritize design parameters according to the overall building performance as attempted in Paper A, but without the troublesome weighting system. Another typical "trait of" of sensitivity analysis is that they provide fixed sensitivity measures that are not necessarily of interest to the decision-maker. To overcome these issues, we have developed two novel sensitivity techniques denoted TOM and TOR [19]. The TOM method ranks inputs according their influence on multiple outputs, whereas TOR provides real-time sensitivity measures during Monte Carlo Filtering. The two methods are presented in the conference paper "Interactive building design space exploration using regionalized sensitivity analysis" which is denoted Paper E placed in appendix A [19].

3.2.3 CURSE OF DIMENSIONALITY

As mentioned in Paper C, 5.000 simulations easily become insufficient when applying filters to explore specific regions of the design space. In general, the design space expands exponentially with the number of dimensions, which makes it near impossible to perform enough simulations to cover the multidimensional space sufficiently. This dimensional issue is sometimes referred to as the "curse of dimensionality" [20]. To exemplify this, imagine that 5 simulations represent a sufficiently dense sample for a univariate problem. The equivalent sample density for 10 dimensions requires $5^{10} \sim 10$ million simulations. We elaborate on this issue in Paper F, in which we argue that for a simple model, used for evaluation of thermal comfort according the Danish building regulations, it is possible to simulate (nearly) all of the design space. The reader may go to Appendix A to read Paper F which refers to the conference paper "Thermal comfort in residential buildings by the millions - Early design support from stochastic simulations" [21]. The article has led to the development of a free, online design tool [22]. This tool relies on millions of simulations calculated in advance. It is used to guide designers and meet thermal comfort requirements in dwellings according to Danish building code.

However, the approach of performing millions of simulations to cover "all" of the design space is only viable for BPS software with few influential inputs and near-instant calculation time. Thus, for most BPS software, such as

BSim, this is not a suitable approach. Instead, the “curse of dimensionality” calls for the use of metamodeling as described in the following.

3.3 DEALING WITH MULTIPLE DIMENSIONS USING METAMODELS

As argued in Paper C, metamodeling may be the solution to various issues related to building performance simulations. To recap, a *metamodel* can be defined as a model generated from a more complex model by means of *metamodeling*. Since metamodels provide much faster calculations than the complex model, they are often used to explore design space, optimize performance, and provide immediate feedback. As described in Paper C, even linear regression models with low accuracy can be used to generate useful predictions of building performance [23]. However, this only worked under assumption of monotonic models and by adjusting the filter criteria. To avoid such inconvenient circumvention, we desire more accurate and robust metamodels which also works for detailed BPS models such as BSim. For this reason, we have made a comprehensive review and comparison of metamodeling techniques in the context of building simulations. This work is presented in Paper 4, which follows a brief description of the BSim model used to test the different metamodeling techniques.

As part of ongoing research, we have defined a BSim model for a typical office space with 14 variable design parameters. Examples of variable parameters are window-to-wall ratio, internal load, room depth, and minimum air supply. For this “generic” office space, the BSim software is used to assess energy demand, daylight, thermal comfort, and air quality. The logic behind such “generic” spaces is similar to the design approach presented in Paper F, in which we have “pre-calculated” millions of designs to guide decision-makers in early design [21]. However, in that case, we used a rapid model to assess thermal comfort for millions of predetermined designs with 10 variable inputs. This is not possible here, since the time consumption for each simulation is at least a hundred times greater for BSim than for Be15 and this case involves more dimensions. With BSim, we could perform 10.000 simulations over the course of a weekend using a laptop [24], whereas with Be15 100.000 calculations were made in a couple of hours. Thus, the ratio of time consumption is approximately 1:200. We return to the concept and applicability of the “generic spaces” in the chapter 4.1, but now we continue with the comparison of metamodels in Paper D

3.3.1 PAPER D

In Paper D, we apply and compare six metamodeling techniques using supervised learning to construct fast regression models. The article is titled *“A comparison of six metamodeling techniques applied to building performance simulations”* which has been submitted to Applied Energy.

Revised manuscript submitted to Applied Energy

A comparison of six metamodeling techniques applied to building performance simulations

Torben Østergård[a,b], Rasmus L. Jensen[a] and Steffen E. Maagaard[b]

[a] *Aalborg University, Department of civil engineering, Thomas Manns Vej 23, DK-9220 Aalborg Ø, Denmark*
[b] *MOE Consulting | Engineers, Mariane Thomsens Gade 1C, DK-8000 Aarhus, Denmark*

Building performance simulations (BPS) are used to test different designs and systems with the intention of reducing building costs and energy demand while ensuring a comfortable indoor climate. Unfortunately, software for BPS is computationally intensive. This makes it impractical to run thousands of simulations for sensitivity analysis and optimization. Worse yet, millions of simulations may be necessary for a thorough exploration of the high-dimensional design space formed by the many design parameters. This computational issue may be overcome by the creation of fast metamodels. In this paper, we aim to find suitable metamodeling techniques for diverse outputs from BPS. We consider five indicators of building performance and eight test problems for the comparison six popular metamodeling techniques – linear regression with ordinary least squares (OLS), random forest (RF), support vector regression (SVR), multivariate adaptive regression splines, Gaussian process regression (GPR), and neural network (NN). The methods are compared with respect to accuracy, efficiency, ease-of-use, robustness, and interpretability. To conduct a fair and in-depth comparison, a methodological approach is pursued using exhaustive grid searches for model selection assisted by sensitivity analysis. The comparison shows that GPR produces the most accurate metamodels, followed by NN and MARS. GPR is robust and easy to implement but becomes inefficient for large training sets compared to NN and MARS. A coefficient of determination, R^2, larger than 0.9 have been obtained for the BPS outputs using between 128 and 1024 training points. In contrast, accurate metamodels with R^2 values larger than 0.99 can be achieved for all eight test problems using only 32 to 256 training points.

Keywords: Gaussian process regression (kriging), Random Forest, Neural Network, Support Vector Regression, Sensitivity analysis

1 Introduction

1.1 Motivation for metamodeling

The building sector accounts for roughly 40% of the total energy consumption and 38% of the CO_2 emissions in the European Union [1]. On a global scale, the energy savings potential is estimated to 53 exajoules each year by 2050 [2]. Building designers play a vital role in realizing this enormous energy savings potential. Architects and engineers use building performance simulations (BPS) to assess and reduce the environmental impact of buildings and, at the same time, meet strict requirements related to indoor climate. Examples of performance objectives are energy demand, CO_2 footprint, thermal comfort, daylight availability, and construction costs. To find possible solutions, the design team may vary a large number of design parameters such as building geometry, insulation thickness, glazing properties, and HVAC systems. The variations of these parameters constitute an enormous multi-dimensional "design space". The many design parameters (model inputs) and requirements (model outputs) makes it difficult and time-consuming to explore the design space efficiently and find favorable solutions that meet all requirements. To address this multivariate problem, it is becoming increasingly popular to perform a large number of simulations using Monte Carlo methods or optimization routines [3][4][5][6][7]. Unfortunately, most BPS software are computational demanding. A single simulation often takes minutes to compute or even hours in the case of CFD simulations. For real-life applications, the task of running thousands or millions of simulations is an obstacle for widespread adoption of design space exploration, uncertainty analysis, sensitivity analysis, and optimization.

The computational obstacle of BPS may be overcome by supercomputers, cloud computing, or metamodeling. Supercomputers are expensive if not managed efficiently to avoid downtime. Cloud computing is presumably a cheaper alternative, and several popular BPS tools provide this feature for optimization or uncertainty analysis [3]. Still, if the design team wish to perform sensitivity analysis to identify important inputs or interaction effects, such analysis easily requires thousands of simulations. This is likely to take hours or days – even with access to cloud computers [8][9]. An extensive set of Monte Carlo simulations allows the design team to explore a high-dimensional design space under various constraints and immediately observe the consequences of different design choices [10]. A useful tool for such analysis is the interactive parallel coordinate plot, which enables rapid and visual exploration of multivariate data (see Figure 5) [11]. In terms of computational effort, Østergård et al. (2017) demonstrated that 5,000 simulations were insufficient when applying five constraints representing building legislation and architects' ambitions. Similarly, in the development of a design tool for thermal comfort evaluation, it was necessary to run millions of simulations to cover the design space sufficiently, even though the model only contained nine design parameters and two objectives [12]. Conclusively, this "curse of dimensionality" advocates the use of fast metamodels to overcome the computational challenges.

1.2 Performance requirements of metamodels

Above, we have argued for a potential of using metamodeling in the context of BPS. The type of metamodeling, addressed in this paper, is also referred to as supervised learning, which overlap with regression analysis. Hence, the metamodels are constructed from a set of input and output data to enable predictions of future outputs. That means we need to run a number of building simulations from which we construct the metamodel to be used for faster predictions of building performance. The metamodels may be constructed in a wide variety of ways using many different techniques. These may differ substantially with respect to predictive accuracy, computational efficiency, ease-of-use, transparency,

Be15	Idealized quasi-steady-state BPS software
BPS	Building performance simulation
BSim	Dynamic, multi-zone BPS software
CART	Classification and regression trees
D	Number of input dimensions
GPR	Gaussian process regression aka. kriging
hyperparameter	Model specific setting that may be changed or tuned to increase accuracy
LP_τ	Low-discrepancy sequences by Sobol
MARS	Multivariate adaptive regression splines
N	Number of samples/simulations
NN	Artificial neural network
OLS	Ordinary least squares linear regression
PR	Polynomial regression
problem	Output (aka. dependent variable aka. response variable) from BPS or theoretical test functions
Pruning	Reduction of model complexity to avoid overfitting
regularization	Addition of regularization term or penalty term to avoid overfitting
R^2	Coefficient of determination
RBF	Radial basis function
RF	Random forest incl. regression trees
SVR	Support vector regression

and robustness (see 4.1). Unfortunately, there is no "perfect" and one-fits-all method. Thus, in searching for the most suitable metamodel for BPS, we have to describe the characteristics of our building simulations and define requirements, and optionally desirable features, of the metamodels. Emphasis is on applicability for novel users and the applied algorithms are all available with Matlab.

Building simulations are typically performed with complex "black box" models with many variable design parameters. Therefore, the users normally have no clear knowledge of the underlying equations and how the inputs interact with each other. In addition, the design team need to assess quite different building performance indicators, such as energy, thermal comfort, and daylight. These may vary substantially in complexity and in the shape of the output distribution (see 4.2). Thus, the metamodeling technique must be *robust* to the number of inputs and the type of output. We have identified three separate cases for which metamodeling can be applied to enable design space exploration, optimization, and sensitivity analysis. The three cases denoted A, B, and C can be distinguished by their diverse requirements, which may be characteristic for many other applications across scientific disciplines:

A. *Expert with considerable time (~days) and emphasis on accuracy.*

   The task is to develop a generic, reusable tool for early design support based on a predefined room types that often occur in buildings, e.g. open-offices and meeting rooms. Time for training and construction of the metamodel is not critical.

B. *Non-expert with limited time (~hours) and need for ease-of-use and robustness*

   For each building project, an engineer (or architect) with presumably limited knowledge of metamodeling needs to construct metamodels, which represent the particular building and its desired performance. The applied technique must be robust towards different objectives and easy to use.

C. *Automated metamodeling requiring a minimum of training points (obtained in minutes) and high robustness*

   These conditions apply to the development of a joint CAD and BPS framework, in which the design team selects specific rooms in a CAD environment. For the selected room, BPS and metamodels are automatically performed with no user interaction and within a limited time frame.

For all cases, a large number of new predictions must be performed rapidly for real-time design-space exploration during meetings with multiple stakeholders [10].

The purpose of this study is to perform a comprehensive comparison of metamodeling techniques and thereby identify the techniques most suitable to accommodate the requirements listed above. We have strived for an extensive comparison by considering diverse building performance metrics and well-known mathematical test functions. The number of training data points has been varied substantially, i.e. from 2^5 to 2^{13} (32–8,192). To improve transparency and reproducibility, we show all hyperparameter variations and data online [13]. In addition, we use the unit-less R^2 values to report accuracies, which makes it easier to compare the accuracies for diverse problems encountered in different disciplines. With these intensions, we strongly believe that this study is relevant to all research areas and industries, which apply metamodeling in the form of supervised learning.

2 Literature review

First, we investigate earlier uses of metamodeling in the context of building simulations. Afterwards, we look for promising metamodeling techniques based on comparisons made across scientific disciplines. Emphasis is on accurate, rapid techniques that can be applied by a practitioner of building performance simulations with limited knowledge of metamodeling. Moreover, non-Gaussian distributions of aggregated outputs (see 4.2) and interaction effects must be captured without knowing the underlying equations, which are mostly hidden in the commercial "black box" software.

2.1 Metamodeling in the field of building performance simulations

Metamodeling has been applied to building simulations for a variety of reasons, which include early design decision-making [14][15][16][17], uncertainty and sensitivity analysis [15][17][18][19][20][21][22], design optimization [16][18][19], and model calibration [21][23][24]. Most research addresses energy consumption, though metamodels have also been used to emulate thermal comfort [18][25][22], daylight [16][22], and financials costs [19]. A wide range of metamodeling techniques have been applied in the reviewed studies: linear regression (OLS) [14][15][22][24], polynomial regression (PR) [16][17][24], multivariate adaptive regression splines (MARS) [22], support vector regression (SVR) [18], neural network (NN) [20], and Gaussian processes regression (GPR) [21][25]. Additional methods, such as step-wise linear regression, decision trees (CART) and random forests (RF), have been found in works that compare metamodel methods in relation to BPS [26][27][28]. To sum up, a wide variety of metamodeling methods have been applied for diverse applications but typically only one method has been applied and the basis for choosing it is not always well-founded.

The accuracies reported in the reviewed papers are relatively high. Surprisingly, linear methods (OLS) provide R^2 values as high as 0.94–0.96 [14][15]. Slightly higher values, 0.97 and 0.98, have been obtained from MARS and NN, respectively [19][20]. Chen et al. (2017) apply MARS to improve accuracy compared to stepwise linear regression; 0.70 to 0.77 for illuminance level, 0.93 to 0.97 for air change rate, and 0.87 to 0.9 for thermal comfort [22]. The accuracies from the other studies are difficult to summarize, since the metrics used to report accuracy are not unit less nor problem independent. The complex BPS problems addressed in this work lead to R^2 values in the range of 0.48–0.85 using linear regression (see 5.2.1). This calls for the use of more advanced statistical methods that can handle non-linearity and interaction effects. Even if more advanced methods are used, we encourage reporting R^2 values obtained from OLS. Thereby, it is possible to assess the complexity of the problem and observe how much the accuracy is improved.

2.2 Previous comparisons of metamodeling techniques

We now consider 10 papers, which compare at least three metamodeling techniques without restricting the scope to building performance [26][27][28][29][30][31][32][33][34][35]. Table 1 summarizes recurrent features which include *applied techniques*, *training set sizes*, *number of engineering and test problems*, *number of inputs*, and *conclusion*. Note that the applied techniques may include variants or optimized versions. For each article, a checkmark indicates the “best” technique while a checkmark in parenthesis represents an alternative technique, which is almost as good or even better if another performance criterion, such as *time consumption* or *ease-of-use*, is more important than e.g. *accuracy*. Since each of the applied metamodeling techniques has at least one checkmark, it is difficult to rule anyone out. To sum up, the reviewed articles provide very different recommendations and the “best” technique depends on the problem at hand. Nevertheless, GPR seems to stand out in terms of accuracy but, at the same time, it is among the slowest algorithms.

In addition to the variations shown in Table 1, the reviewed works differ in the extent of which *hyperparameters* have been optimized. Some methods can be “tweaked” only a little, such as RF. Other methods, especially NN, have a large number of possible configurations, which induce a potential for improving accuracy at the expense of time-consuming optimization. This makes it difficult to conduct a fair comparison. To address this issue, we aim to make the selection and variations of hyperparameters clear and reproducible (see 4.4.1). Finally, we try to combine the best practices found in these comparison papers to produce an all-encompassing and improved metamodeling comparison. We, therefore, assess popular techniques under various requirements while varying the number of training points, problems, and inputs.

3 Metamodeling techniques

In this chapter, we present the basic principles behind the six metamodeling techniques, which have been selected based on popularity and the recommendations in the reviewed literature (Table 1). We have not included all the methods identified in Table 1. For example, extensions of linear regression, such as polynomial regression, are assumed to be computational inefficient when compared to more sophisticated high-dimensional expansions made with e.g. SVR and GPR. As for RBF, this is considered a special case of NN and we confine the NN analysis to the most common single hidden layer perceptron, since covering all NN structures would be too overwhelming. The reviewed methods are available in MATLAB R2017A except from MARS, which we test using an independently developed toolbox, ARESlab [36]. Each method has a variety of adjustable settings that affect both accuracy and computational effort. Examples are fitting methods, tuning parameters, regularization terms, convergence criteria, and so on. We try to identify the most influential settings based on the theoretical background and software documentation. In the next chapters, we perform sensitivity analysis of these settings to assess their relative importance and to observe if some settings seem to provide more robust and accurate predictions.

3.1 Ordinary least squares linear regression

In this paper, OLS refers to linear regression models fit by ordinary least squares. Such commonly used linear regression models are very fast, simple to apply, and easy to interpret [37]. Despite their simplicity, they may outperform complex, non-linear models in situations when data is sparse, noise is large, or the number of training samples is low [37]. The linear methods may also be applied to variables, which are produced by the original inputs by the use of transformations, basis expansions (polynomial regression), or inclusion of interaction terms. In our work, we apply the simple version in which the variables are identical to the training set data. This produces a measure of the linearity of the modelled problem, and we observe the potential benefits of using more complex and time-consuming machine learning methods.

Table 1: Summary of articles which compare metamodeling techniques.

Reference and discipline	Metamodeling techniques									Training set	# problems		# inputs	Conclusion
	GPR	MARS	NN	OLS	PR	RBF	RF	SVR	Others*		Engineer	Benchmark		
Cheng & Cao, 2014 [26] → Building performance		✓	×			×	×	×		768	2		8	MARS combined with optimization algorithm best for both heating and cooling
Wei et al., 2016 [27] → Building performance	(✓)	(✓)		×			(✓)	×		100	3		8	MARS and RF effective for annual heating GPR and MARS effective for cooling and electricity
Yildiz et al., 2017 [28] → Building performance			(✓)	(✓)			×	×	×	≤ 8760 †	20 ‡		11	Linear regression easier to use and implement. Higher accuracies obtained with e.g. NN
Lim & Zhai, 2017 [29] → Building performance	(✓)	×	×	×				×		100	2		6	GPR most accurate but also slowest
Duarte et al., 2017 [30] → Building performance			✓				✓	(✓)	(✓)	768	2		8	R^2 values above 0.98 for all models except decision tree for cooling load (0.96)
Jin et al., 2001 [31] → Optimization	(✓)	×			(✓)	✓				scarce ($3D$), small ($10D$), large ($\sim D^2$)	1	13	2, 3, 10, 14, 16	RBF performs best overall PR/GPR performs best with scarce data
Kim et al., 2009 [32] → Mechanical science	✓					×		×	✓	$3D$, $5D$, $7D$		6	2, 4, 6, 8	GPR best with 2 inputs, otherwise MLS
Li et al., 2010 [33] → Decision Support	(✓)	(✓)	×			×		✓		100	2	2	2, 3, 4, 8	SVR most robust and accurate, followed by GPR, though MARS is fastest
Villa-Vialaneix et al., 2012 [34] → Agriculture	×		×	×			✓	×	×	100, 200, 500, … , 15.000	2		11	RF is most reliable and has low computational cost while accuracy is acceptable (not highest)
Ju et al., 2016 [35] → Mechanical engineering			×			×		✓		scarce (11–19), small (9–50), large (20–100)		10	2, 3, 5, 8, 10	SVR most robust and accurate

✓ indicates the preferred technique and × indicates a reviewed technique with a poorer performance.

* Other reviewed techniques include moving least squares regression (MLS), step-wise linear regression, and state-dependent parameter regression.

† Hourly time series for a maximum of one year. ‡ Metamodels were built for hourly and daily peak electrical loads for five seasons and two case studies.

3.2 Random forest

Random forest is an extension of a simple and intuitive metamodeling approach called *classification and regression trees* (CART) [38]. A *decision tree* is constructed by recursive, binary splitting of the predictor space into non-overlapping regions (*leafs*), such that new predictions equal the mean response value in those regions. Starting with the whole training set (*trunk*), the tree is split into two child nodes (*branches*) at a threshold value of the predictor X_i, which leads to the largest reduction in the residual sum of squares. In other words, each split maximizes the difference of the average response values in the two child nodes. The splitting is repeated recursively until a stopping criterion is met, e.g. if the split produces child nodes with a number of observations less than a given value (MinLeafSize) or if a predefined maximum number of splits is reached. Alternatively, the splitting continues until all leafs contain a single observation after which the tree is "pruned" to a simpler sub-tree to avoid overfitting.

Decision trees are easy to interpret and visualize, but they are not very accurate nor robust regarding changes in the data [38]. To remedy this, bootstrapping techniques, such as bagging or random forest, are used to grow many trees, which are combined and averaged to provide better predictions. Random forest differs slightly from bagging in such a way that, for each split, only a subset of predictors may be considered. This results in de-correlated trees, which are less dependent on strong predictors and provide more reliable predictions compared to bagging [38].

To apply random forest we use the Matlab function treebagger and vary the following hyperparameters:

- The number of predictors to consider at each split (NumPredictorsToSample)
- The minimum number of observations per tree leaf (MinLeafSize)

For all problems, we grow 100 trees since larger numbers seem to increase computational with an insignificant improvement of accuracy. For better time comparison with other techniques, the computations were not run in parallel.

3.3 Multivariate adaptive regression splines

Multivariate adaptive regression splines (MARS) is a regression technique, which combines recursive partitioning and spline regression [39]. Basically, MARS builds a model of weighted basis functions. The two-stage construction consists of a forward phase and backward phase similar to that of CART with pruning. In the forward phase, MARS recursively partitions the training data by adding a pair of basis functions, which maximize the reduction of the residual sum of squares (RSS). To elaborate, the data is split at the optimal knot location using hinge functions, which results in a regression formula for each of the two split regions. This "greedy" forward approach is stopped when a maximum number of basis functions is reached or if the reduction of RSS is too small. In the backward "pruning" phase, a generalized cross validation (GCV) criterion is used to successively remove the least effective terms. This reduces the complexity of the model to avoid overfitting. GCV includes a user-defined penalty term to approximate the error from leave-one-out cross validation.

Since MARS is not implemented in Matlab, we use the independent toolbox ARESlab [36]. The model configuration includes more than 20 properties. As recommended by the developer, we pay most attention to the following:

- The maximum number of basis functions
- The maximum number of basis function after pruning
- The maximum number of factors, i.e. interaction order
- Utilization of either piecewise-linear or piecewise-cubic modelling
- Generalized cross validation penalty per knot (c)

3.4 Support vector regression

Support vector machine (SVM) is a machine learning approach, which is popular for *classification* problems, but SVM may also be used for *regression* (SVR) [37]. In a two-class *classification* setting (SVM), the algorithm seeks the boundary or hyperplane with the maximal margin, which separates the training data into its respective classes with a minimum of misclassifications. Only observations exactly on the margin, or on the wrong side of the margin, are the so-called support vectors that define the hyperplane [38]. Analogously, in *SVR*, a "best fit" hyperplane is constructed from support vectors, which describe observations that lie far from the regression hyperplane. SVR can be applied to non-linear problems using the kernel trick which maps inputs into a high-dimensional feature space. This is possible since the solution of the support vector problem depends on the inner product of observations which can be replaced by a kernel function [40]. Popular kernel functions include the Gaussian radial basis function and polynomial kernels of degree d [37].

In this work, we use the Matlab function fitrsvm. This function takes more than 30 arguments of which most are optional and do not affect the accuracy. Based on literature recommendations and Matlab documentation, the following properties are varied [32][33][34]:

- Three kernel functions are used: a Gaussian, a linear, and a polynomial kernel
- The scaling factor (KernelScale)
- The regularization term C, which controls the influence of each support vector (BoxConstraint)
- The width of the margin ε, which affects the number of training points used as support vectors.

3.5 Gaussian process regression (kriging)

Gaussian process regression (GPR), also known as kriging, is a non-parametric Bayesian approach to supervised learning [41]. The Gaussian process is a collection of random variables which has multivariate Gaussian distribution. The main idea is to define a prior probability to infinitely many functions and then add a finite number of training points, which result in a posterior distribution over functions that pass exactly through the data points. Hence, any complex model can be fit by this non-parametric approach. Typically, closely connected points in the input space have similar response values. These similarities can be expressed using the covariance function, also referred to as kernel function, which specifies the covariance between pairs of random variables. The Gaussian process is specified by this covariance function along with a mean function. The choice of covariance function introduces different hyperparameters to be optimized. Since GPR is a probabilistic method, it provides uncertainty information for new predicted values since every point is described by a normal distribution.

In this work, we apply the Matlab function fitrgp and vary the following properties:

- The kernel function
- The basis function
- Signal variance (sigma)

3.6 Neural Network

The term *neural network* covers a broad class of models and learning algorithms. In this work, we focus on the common *feedforward neural network* with a *single hidden layer*. This is also referred to as a single layer perceptron. This network consists of an input layer, one hidden layer, and an output layer. Each layer consists of a number of *nodes* or *neurons*, which propagate information to each of the nodes in the subsequent layers using weighted connections and transfer functions. The number of nodes in the input layer and output layer equals the number of model inputs and the number of model outputs, respectively. The number of neurons in the hidden layer is predetermined by the modeler. Increasing the number of neurons allows for more complexity at the risk of overfitting unless regularization or penalty functions are applied. The hidden layer neurons may be interpreted as basis expansion of the original inputs X with the neural network acting as a standard linear model using these transformations as inputs [37]. When the network and initial conditions are defined, a training vector is presented to the input layer and propagated to the output layer. The network output is compared to the desired output using an error function, typically based on the mean square error. An optimization algorithm is then used to update the weights and biases to minimize the training error. Unfortunately, the optimization problem typically contains many local minima. Therefore, different initial conditions and optimization algorithms may lead to different non-optimal solutions.

For the investigation of the neural network settings, we constrict the analysis to the one-hidden layer perceptron for which the following parameters are varied:

- Number of neurons in the hidden layer
- Training function minimizing the mean-square-error
- Transfer function for the hidden layer
- Transfer function for the output layer

4 Methodology

As stated in the introduction, we strive for a comprehensive comparison study of six popular metamodeling techniques with emphasis on building performance simulations. This chapter explains our approach, which is based on the lessons learnt from the literature review. First, we establish which measures to use when assessing the performance of the metamodels. Accuracy is of particular interest, and we report it using the unitless and problem-independent R^2 values. Next, we describe 13 "problems" from BPS and theoretical benchmark functions, which provide diversity and transparency to the study. For all problems, we construct metamodels for a great range of training points, which allows us to determine when the accuracy converges and when the time-consumption exceeds different thresholds. This helps balance accuracy and time-consumption for different applications. Lastly, we describe how to use a grid based approach and sensitivity analysis to test a variety of model settings for different techniques in a consistent manner.

4.1 Performance measures

The performance of a metamodeling technique can be assessed using a variety of qualitative and quantitative measures [42]:

- **Accuracy** describes how much the metamodel outputs deviate from the real model outputs.
- **Computational efficiency** relates to the computational effort of constructing and tuning the metamodel along with the calculation of new predictions
- **Robustness** refers to the ability to provide acceptable accuracy for diverse problems with varying levels of complexity and dimensionality
- **Simplicity** or ease-of-use. The method should be simple to implement with few user inputs and easy to configure for each problem
- **Interpretability** or transparency describes the amount of insight which the method gives into model behavior such as parameter importance and interaction effects

To some extent, we address all of these in our comparisons. The criteria for selecting a metamodeling technique differ with the problem at hand, as explained in the introduction. However, accuracy is often a decisive factor. We apply the well-known coefficient of determination, R^2, which is a unit less measure of the global error with an optimal value of 1 making the measure practical for comparison. R^2 is correlated with two other common metrics, the mean-square-error (MSE) and the root-mean-square-error (RMSE) [42]. There are no universal threshold values to determine the accuracy. In the context, we arbitrarily consider a metamodel with $R^2>0.9$ as *reasonable*, $R^2>0.95$ as *accurate*, and $R^2>0.99$ as *near perfect*, respectively. However, a metamodel based on OLS with a low value of 0.4 has been shown to provide sufficient information and guidance in a Monte Carlo filtering process [10]. Though, this requires an adjustment of filter criteria and causes a problem with a monotone response.

Computational efficiency is another important aspect of metamodeling. The underlying equations may indicate how the computational requirements depend on dimensionality of input space, observations, expansion functions, interactions, etc. However, the efficiency is also highly dependent on the software, amount of tuning, convergence criteria, parallelization, and so on. Therefore, we strive to provide rough measures of the computational efficiency with respect to the applied technique, the input size, and the number of training point and new predictions. We distinguish the computational efficiency by the units *milliseconds*, *seconds*, *minutes*, and *hours*. For real-time analysis and design space exploration, computations are preferably measured in milliseconds. Computational time varying from seconds to about a minute may be acceptable for a modeler doing manual tuning and analysis. When the time frame is several minutes or hours, the user cannot be expected to wait for the results, so the computations must be run in the background, during breaks, or overnight, which limits the applicability.

Table 2: Probability distributions describing the variability of the design parameters for the generic office room.

Input parameters	Unit	Uniform	Discrete	
		min – max	values	weights
Minimum air supply	*l/s m²*	0.5 – 2		
Load, people and equipment	*W/m²*	5 – 30		
Maximum cooling	*W/m²*		0, -20, -40, -60	0.5, 0.16, 0.16, 0.16
Window-wall-ratio (WWR)	*%*	30 – 90		
Room depth	*m*	4 – 8		
Solar Heat Gain Coefficient	-		0.25, 0.32, 0.37, 0.43, 0.5, 0.57, 0.63	equal
Overhang	°		0, 20, 40, 60	equal
Activate shading	*klux*		38, 45, 52, 61, 70, 100	equal
Reflecance, room average	-		0.4, 0.55, 0.65, 0.78	equal
Night ventilation	*l/s m²*	0.5 – 4		
Shading factor	-		0.2, 0.4, 0.6, 0.8	equal
Load, lighting	*W/m²*		2, 4, 6, 8	equal
U-value, facade	*W/m² K*		0.12, 0.14, 0.16, 0.18, 0.2	equal
Thermal mass, averaged	*Wh/m²*		60, 140	equal

4.2 BPS models and test problems

For this work, we consider five types of output from building performance simulations related to two separate cases; a generic office room and a project-specific educational building. In addition, we test the metamodeling techniques for eight theoretical test problems with varying degree of complexity and dimensionality. The data sets can be accessed online from Mendeley Data [13].

First, we describe the generic office room, which can be modified by 14 variable inputs and the performance of the design is assessed by four aggregated simulation outputs. The variable inputs, which include window size, room depth, internal loads, and airflow, are important to building performance but difficult to assess and fixate during early design stages. Their variabilities are described by probability distributions as shown in Table 2. Using quasi-random sampling, we perform a Monte Carlo experiment to make performance simulations in the multi-dimensional feature space. The BPS software BSim, version 7.16.1.19, is used for whole-year, dynamic simulations of the indoor climate and energy demand and to evaluate the daylight factor (DF) [43]. For each simulation, we consider four aggregated outputs:

- **Energy** sums up the yearly energy demand related to heating, cooling, ventilation, and lighting. Measured in kWh/m² floor area.
- **DF>2%** is an indicator of daylight availability measured by the percentage of the room working plane with a daylight factor above 2%.
- **h>26°C** is the number of hours with indoor temperature above 26°C, which is a measure of thermal (dis)comfort.
- **Max CO_2** refers to the yearly maximum CO_2 level. It is measured in parts per million (ppm) and indicates air quality.

These performance objectives are problematic to optimize, since improving one of them often worsens the others and, in addition, the design parameters affect other objectives, such as aesthetics and costs. In addition, some of the 14 inputs are redundant with respect to the individual outputs, e.g. only a few inputs affect the CO_2 level. This may pose a challenge for some metamodeling techniques.

Now, we consider the building specific BPS problem, which estimates the yearly energy demand for a 15,000 m² educational building. For this case, 10 variable inputs describe the variability during early design. The energy demand is computed with Be15, version 8.16.2.4, which is based on the quasi-steady-state monthly method from ISO 13790 [44]. The underlying equations and the level of detail contained in the Be15 software are much simpler than the differential equations and complexity used in BSim. From a BPS perspective, it is of interest to compare the metamodeling of

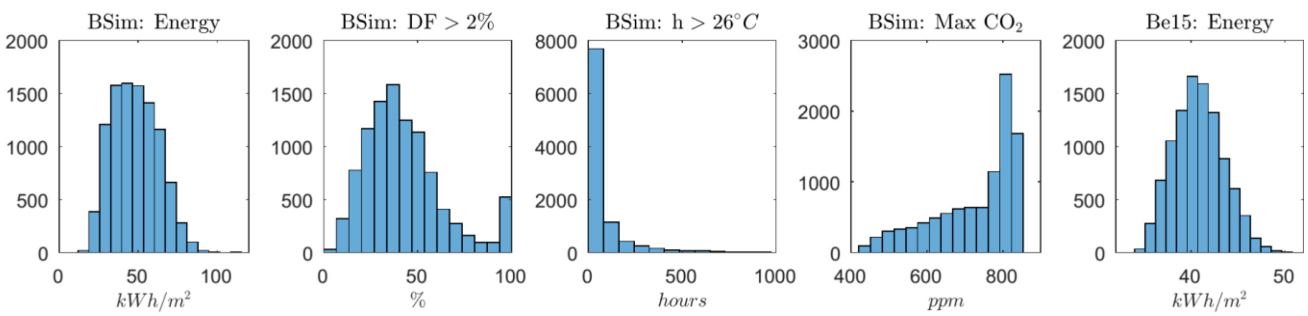


Figure 1: Output distributions for the four aggregated performance indicators for the generic office room.

Table 3: List of theoretical test problems. $x \sim U(a,b)$ denotes an input *x* distributed uniformly in the range (*a*,*b*), and $x \sim N(\mu,\sigma)$ denotes a normally distributed input *x* with mean μ and standard deviation σ.

Id	Function name	Inputs	Formula	Variations
A	McCormick	2	$y = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$	$x_1 \sim U(-1.5,4)$ $x_2 \sim U(-3,4)$
B	Six Hump Camel	2	$y = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (4 + 4x_2^2)x_2^2$	$x_1 \sim U(-2,2)$ $x_2 \sim U(-1,1)$
C	Ishigami	3	$y = \sin(x_1) + 7\sin^2(x_2) + 0.1\, x_3^4 \sin(x_1)$	$x_i \sim U(-\pi,\pi)$
D	Miele-Cantrell	4	$y = (e^{-x_1} - x_2)^4 + 100(x_2 - x_3)^6 + \tan^4(x_3 - x_4) + x_1^8$	$x_i \sim U(-1,1)$
E	Egg-Holder	6	$y = \sum_{i}^{n-1}\left(-x_i \sin\left(\sqrt{\lvert x_i - x_{i+1} - 47\rvert}\right) - (x_{i+1} + 47)\sin\left(\sqrt{\lvert 0.5x_i + x_{i+1} + 47\rvert}\right)\right)$	$x_i \sim U(-512,512)$
F	The Primer	8	$y = \sum_{i=1}^{4} W_i Z_i$	$Z_i \sim N(\mu_Z, \sigma_i)$ $\mu_Z = 0$ $W_i \sim N(\mu_{W,i}, \sigma_i)$ $\mu_{W,i} = 0.5i$
G	Michalewicz	10	$y = -\sum_{i=1}^{10} \sin(x_i)\sin^{2m}\left(\frac{ix_i^2}{\pi}\right)$	$x_i \sim U(0,\pi)$ $m = 10$
H	Rosenbrock	12	$y = \sum_{i}^{n-1}(x_i - 1)^2 + 100(x_i^2 - x_{i+1})^2$	$x_i \sim U(-2.048, 2.048)$ $i = 1, 2, \ldots, 11$

the energy demand calculated by two models of different complexity.

The five BPS "problems" represents common ways to aggregate results from whole-year building simulations. Figure 1 shows the output distributions of 10,000 samples. Except for the energy related outputs, the distributions are somewhat irregular judged by their multi-modality, skewness, and peak behavior. Such characteristics may pose a challenge to the metamodeling techniques.

To provide a more thorough comparison, we also consider the eight theoretical test problems listed in Table 3. Initially, we looked for recurrent problems in earlier works, but found that accurate metamodels could easily be obtained for some of these problems [31][32][35][45]. Presumably, this is achievable since the specific problems are best suited for testing of optimization algorithms and not for metamodeling. Thus, the general behavior may be easily reproduced leading to high accuracy while small subtleties, describing the function optima, are overlooked, see e.g. the McCormick optimization problem. Therefore, we searched for test problems better suited for metamodeling by emphasizing non-linearity, input interactions, and varying dimensionality [46][47]. Though, two problems, Ishigami and The Primer, are included because of their applicability to test sensitivity analysis methods [48][49].

4.3 Data generation and preparation

Training data may be generated in variety of ways. These include *design of experiments* with samples near the design space boundaries (Box-Behnken, factorial), *space-filling sampling* (Latin hypercube design, low-discrepancy sequences), and *sequential* sampling where new sample points are generated adaptively in regions where more information is needed [50]. In this context, building simulations are performed in a single event using low-discrepancy sequences (Sobol's LP_τ) which are assumed to cover the design space more evenly when compared to random sampling or Latin hypercube sampling as indicated by studies of Kucherenko et al. (2015) [51][52]. This process consists of the following steps. First, Sobol sequences are used to generate *N* points in an *n*-dimensional unit hypercube where *n* equals the number of inputs in the given function. Next, the points are converted to an $N \times n$ input matrix using the probability distributions shown in Table 2 and Table 3. Finally, we perform building simulations for each row in the input matrix to obtain the building performance outputs. This simulation process is automated from Excel Visual Basic for Applications (VBA). Similarly, the test problems defined in Table 3 are evaluated using Matlab.

The *appropriate* size of training data depends on the shape and complexity of the original model along with the computational budget. At some point, more points only lead to a negligible improvement at unacceptable computational costs. Estimating the appropriate size using heuristic formulas leads to very different recommendations, e.g. in the range of $2 \cdot 10^1$–10^4 for problems with 14 dimensions [31][53]. The computationally acceptable limit in our case was 10,000 simulations, which could be done over the course of a weekend on a laptop running a single core.

For each problem and technique, a metamodel is trained using an exponentially increasing number of points in the range 32 – 8,192. To select the best model settings, we use a validation set with half the size of the training set. However, we use the same 500 test points to assess the accuracy of the chosen metamodel for all cases. Prior to training, the data for the BPS problems is standardized by subtracting means and dividing by standard deviations, since the variables are measured in different units.

4.4 Assessing model settings and convergence

The metamodeling process has been performed in two distinct steps, A and B, as listed in Table 4. In step A, emphasis is on model settings and tuning of hyperparameters. The aim is to assess how easy the methods are to apply and how much the performance varies with different settings. Next, we apply sensitivity analysis to assess the relative importance of model settings used in step A. This enables us to find suitable options and values for a reduced set of variable settings in step B. Finally, we gradually increase the training set from 32 to 8,192 points and compare accuracy, efficiency, and robustness.

4.4.1 Variation of model settings

All metamodeling techniques may be configured and optimized in various ways. These model settings or hyperparameters can be *categorical* (e.g. kernel function and neural network class), *ordinal* (e.g. polynomial order and interaction order), or *continuous* variables (e.g. margin width and signal variance). Some settings induce additional and possibly different "subparameters", such as particular tuning parameters related to different kernel functions. Obviously, it is not possible to perform a complete investigation of the infinite configurations. The process of finding a suitable configuration or optimized model is referred to as *model selection* or *hyperparameter optimization*. Different approaches, to do so, include random search, grid search, and optimization. Here, we perform grid searches of the hyperparameters identified in chapter 2. A grid search is an exhaustive approach where all combinations of discretized hyperparameters are trained. Table 5 lists the discrete options chosen for these hyperparameters for both steps, A and B. The options used in step A are based on recommendations in the software documentation, and include the default settings.

The motivation for the grid based approach, in step A, is not to find the *optimal* configurations, which may be time-consuming or even unattainable. Instead, the idea is to cover a variety of settings for each technique in a similar and fair way that indicate the potentials and ease-of-use of the methods. By including the default values, we can assess how well a method performs "out-of-the-box" without tuning and optimization. We investigate whether the best settings at low training sets also yield good metamodels with larger training sets. If so, this strategy may help to reduce the computational effort of optimizing metamodels for large training sets.

4.4.2 Sensitivity analysis

An advantage of the grid search is that it facilitates the use of sensitivity analysis. This enables us to identify the least influential hyperparameters which may be fixated in step B with a minimum decrease in the variation of the measured performance (R^2). For "factor fixing", it is common to apply a global sensitivity analysis method that estimate the inputs' total order effects, which describe inputs' combined contributions to output variance [49]. We apply a regionalized sensitivity analysis method denoted SA_{TOM} [54]. This method ranks inputs by repeatedly performing random splits of the sorted output distribution and compare the averages of the Smirnov two-sample test statistics for the split input distributions. The method can be used for factor ranking and factor fixing, since the SA_{TOM} measures rank similar to total order effects of the inputs [54]. An important feature is that SA_{TOM} works with the grid based sampling, which is not attainable with the more commonly used methods of Morris and Sobol (decomposition of variance).

When the least significant settings have been identified using SA_{TOM}, we broaden the sensitivity analysis using Monte Carlo filtering to fixate these settings at their most suitable options and reduce the variations of others. For each metamodeling technique, we use an interactive parallel coordinate plot (PCP), combined with histograms showing parameter distributions, to identify settings that provide the most accurate models (see Figure 5). In the PCP, each line represents a specific metamodel configuration and the resulting R^2 values (truncated to zero). Histograms represent the parameter distributions that remain after filter criteria are applied to one or more of the parallel axes. If the histogram shows a large bin for a particular value of an insignificant hyperparameter, we can fixate the hyperparameter at this suitable value.

4.4.3 Accuracy, efficiency, and robustness

Following the sensitivity analysis, we can proceed to step B and construct metamodel for much larger training sets but with less hyperparameter configurations (Table 5). The purpose is to assess accuracy, efficiency, and robustness with respect to variable training sets and diverse problems. Presumably, the variation of hyperparameters now only include the most suitable settings. We therefore assume that the accuracy obtained with the limited variations is still a reasonable indicator of the potential of the metamodeling technique. Presumably, the accuracy of each metamodel can be improved slightly but it comes at the expense of more tuning and increased time consumption.

5 Results and comparison

5.1 Model settings

This first step of the two-step approach defined in Table 4 concerns the ease-of-use of the metamodeling techniques with respect to model settings. To do so, we assess the accuracy obtained using default settings and we test whether the best settings at small training sets also produce the most accurate metamodels for large training sets. Afterwards, we apply sensitivity analysis to reduce the number of configurations to be used in the second step presented in section 5.2.

Table 4: Extent and purpose of the two-step investigations.

Investigation	A. Model settings	B. Convergence
Problems	13	13
Configurations	5 – 109	1 – 5
Training sets	64, 256, 1.024	32, 64, 128, 256, 512, 1.024, 2.048, 4.096, 8.192
Purpose	Assess ease-of-use and find suitable settings	Assess accuracy, efficiency, and robustness

Table 5: Model settings varied in the grid searches in step A and B.

	Model setting (hyperparameter)	Step A (variations)	Defaults	Step B
RF	*1. # variables to sample*	2, 3,…, D	$D/3$	2, 4, …, D
	2. Minimum leaf	2, 3, 4, 5	5	3
	Combinations	(D-1) 5	1	~$D/2$
MARS	*1. Max. # basis functions*	-1, 50, 100	-1	100
	2. Max. # final functions	Inf, 5, 15	Inf	Inf, 10
	3. Max. # interactions	-1, 3	1	3
	4. Piece-wise cubic	true, false	true	true
	5. GCV penalty pr. knot, c	0, 2, 3	3	3
	Combinations	109 *	1	2
SVR	*1. Kernel function*	a. Gaussian (radial basis) b. Linear c. Polynomial	b	c
	2. Kernel Scale	1, auto	1	auto
	3. C (BoxConstraint)	0.001, 1, 1000	1	1
	4. Epsilon	0.001, 0.1, 100	0.1	0.001, 0.1
	5. Polynomial order (step B only)	3	3	3, 5
	Combinations	54	1	4
GPR	*1. Kernel function*	a. Squared exponential † b. Squared exponential c. Matern kernel with 3/2 d. Matern kernel with 5/2 e. Rational Quadratic	a	d
	2. Basis function	i. Constant, ii. Linear, iii. Pure quadratic	i	i
	2. Sigma	α. $0.01\sigma_y$, β. $\sigma_y/\sqrt{2}$, γ. $10\sigma_y$	β	β
	Combinations	45	1	1
NN	*1. Neurons*	5, 10, 15, 20	10	5, 10, 15, 20
	2. Train function (backpropagation)	a. Levenberg-Marquardt b. Bayesian regularization c. Scaled conjugate gradient d. Resilient backpropagation	a	b
	3. Transfer function, hidden layer	i. tan-sigmoid, ii. log-sigmoid	i	i
	4. Transfer function, output layer	α. linear, β. saturating linear	α	α
	Combinations	64	1	4

** includes one configuration with defaults settings. † uses same length scale per predictor.*

5.1.1 Ease-of-use

Figure 2 and Figure 3 show the results from the grid search approach applied to the BSim outputs *Energy* and *h>26°C* using 64, 256, and 1024 training points, respectively. On each box plot, the median is indicated by the central mark, the 25th and 75th percentiles correspond to the box edges, and the minimum and maximum values lie at the whiskers. "Extreme" data is compressed below the dashed line when some R^2 values lie considerably below 0.5. On each plot, orange diamonds show the results when using default settings. Blue crosses show the results when reusing the settings that provide the highest accuracy for the smallest training set of 64 simulations. The tables below show the hyperparameter settings that provide the highest accuracy for each of the grid searches.

The box plots show relative small spreads for RF and GPR and large spreads for SVR and NN. The former methods are thus less sensitive to the chosen settings, and GPR even provides the most accurate results for both problems. Reasonably accurate models may also be obtained using SVR and NN, though it requires more tuning to find the most suitable settings – especially with SVR. In general, the default values do not yield the best performance, which makes the importance of model selection and tuning more critical for SVR, MARS, and NN. For GPR, only the default kernel uses the same length scale for each predictor whereas the other kernels, in our grid search, use separate length scales, which results in more accurate models.

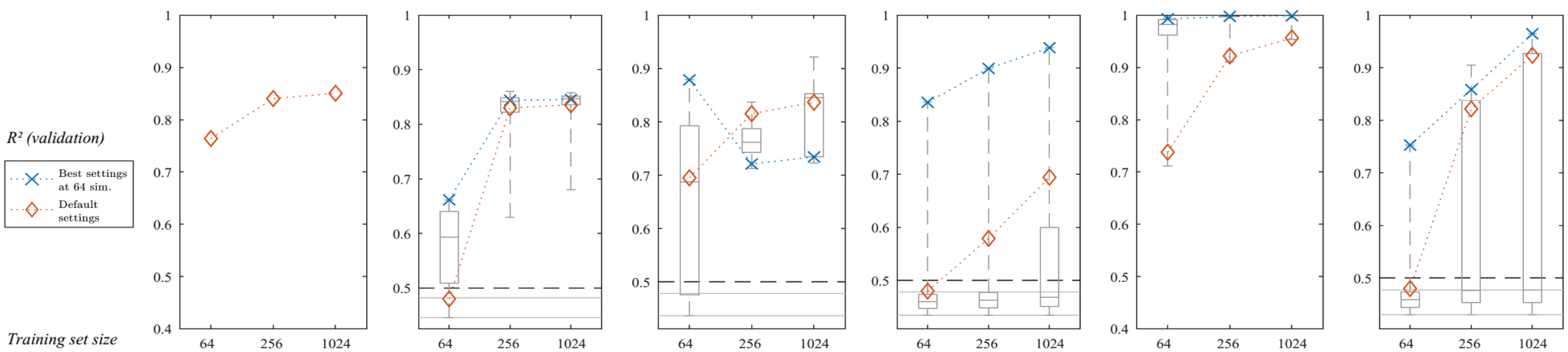
The rationale for reusing the configuration that leads to the best results for the smallest training set (blue crosses), is to reduce the computational effort of the hyperparameter optimization. Indeed, this approach provides better results than the defaults values except from the MARS metamodels of the *Energy* output. Particularly for SVR and GPR this could be a time-saver and assist the modeler in choosing suitable hyperparameters before training metamodels with large datasets. Finally, we observe that the best settings, for each grid search, may vary considerably as shown below the plots.

5.1.2 Sensitivity analysis of model settings

The next step is to perform sensitivity analysis to identify the least influential variables used in step A and to find suitable settings for step B, which includes more training set sizes for all BPS and test problems. For the SA_{TOM} sensitivity analysis, we use the three training set sizes, 64, 256, and 1024, for all BPS and test problems. The relative sensitivity measures are shown on Figure 4. The "*Dummy"* variables help identify inputs with insignificant influence on the R^2 value. Such insignificant inputs include *Minimum number of leafs* for RF, *Kernel Scale* for SVR, *Maximum interactions* and *GCV knot penalty* for MARS, *Transfer function for hidden layer* for NN, and *Basis function* for GPR. These settings may be fixated with limited effect on the metamodel accuracy. Instead, we pay more attention to the most sensitive settings. Note that the tuning parameter, *neurons*, for NN seems unimportant. This is probably caused by the varying number of variables in the various problems. The same may apply to e.g. the *Minimum number of leafs* for RF.

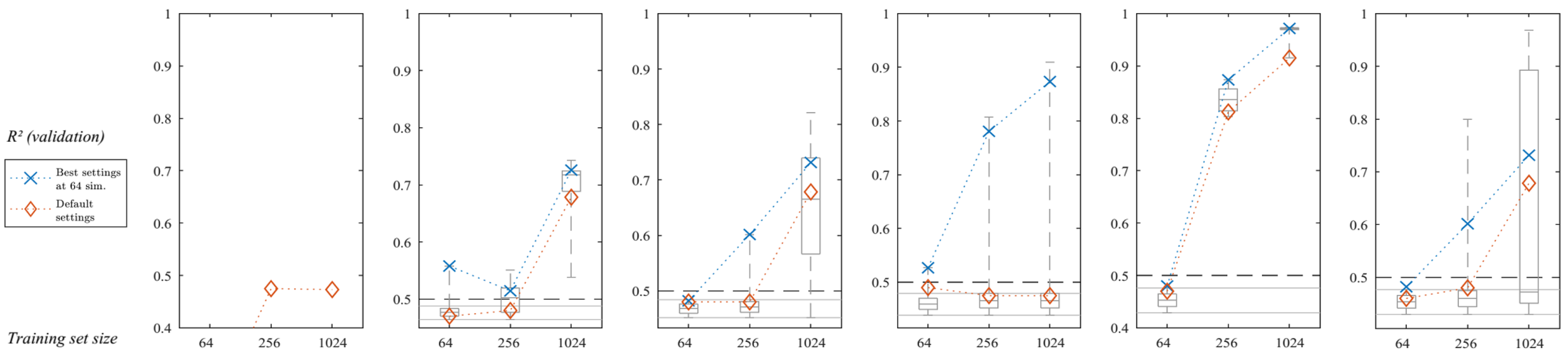
Now, we apply Monte Carlo filtering using the interactive parallel coordinate plot to identify suitable for settings for the reduced number of configurations in step B. An example of this filtering approach is shown on Figure 5, which concerns the NN metamodels. The top plot shows all configurations for a total of 2496 NN metamodels along with the achieved R^2 values and how the metamodels rank with respect to each problem. In the bottom plot, a filter criterion is applied to the coordinate "Rank" to focus on the 10% most accurate metamodels for each problem. Note that "Rank" is strongly correlated with R^2 but problem-independent. From the histograms, we observe that none of the top 10% ranked metamodels have a "saturating linear" transfer function for the output layer (*TF output*). This is the reason that setting *TF output* stands out as the most sensitive parameter in the SA_{TOM} analysis shown on Figure 4. Hence, this setting is set to the option "linear" in the extended analysis. In contrast, the histogram bins are equally large for the two options for the least sensitive setting, *TF hidden*. Therefore, this setting can be set to the default option "tan-sigmoid" without affecting accuracy notably. The second-most influential setting is the type of *training function* for which Bayesian regularization (BR) and alternatively Levenberg-Marquardt (LM) are the best options based on their bin sizes. In the extended analysis, we choose to use BR and only vary the number of neurons to tune the models.

Similar investigations have been performed for all metamodeling techniques except OLS. The fixated options and reduced variations for the extended analysis in step B are shown in Table 5. Note that by filtering the R^2 values, we have been able to assess whether some settings work better for the test problems and other settings are more suitable for the BPS problems. Indeed, this was the case for the kernel function applied to SVR, where the Gaussian radial basis kernel generally performed better for the test problems whereas the polynomial kernel was preferable for the BPS problems. Since focus is on the BPS problems, we use the polynomial kernel for step B and vary the degree of the polynomial between 3 and 5 to allow for more complexity.



Technique	OLS			RF			MARS			SVR			GPR			NN		
Training set size	64	256	1024	64	256	1024	64	256	1024	64	256	1024	64	256	1024	64	256	1024
Best hyperpar. 1	-	-	-	12	9	9	100	50	100	c	c	c	b	c	c	20	5	20
Best hyperpar. 2	-	-	-	5	5	2	3	3	3	auto	auto	auto	ii	iii	i	b	b	b
Best hyperpar. 3	-	-	-	-	-	-	true	true	true	1	1	1	γ	β	α	i	i	i
Best hyperpar. 4	-	-	-	-	-	-	0	3	3	0.001	0.001	0.001	-	-	-	α	α	α
Best hyperpar. 5	-	-	-	-	-	-	5	15	Inf	-	-	-	-	-	-	-	-	-

Figure 2: R^2 values for the output “BSim Energy” in step A. Blue crosses show the results when reusing the best settings found for the training size of 64 simulations.



Technique	OLS			RF			MARS			SVR			GPR			NN		
Training set size	64	256	1024	64	256	1024	64	256	1024	64	256	1024	64	256	1024	64	256	1024
Best hyperpar. 1	-	-	-	8	9	9	50	50	100	c	c	c	c	c	d	5	5	15
Best hyperpar. 2	-	-	-	3	2	2	3	3	-1	auto	1	1	i	i	iii	b	b	b
Best hyperpar. 3	-	-	-	-	-	-	true	true	true	1000	1	0,001	β	β	β	i	i	ii
Best hyperpar. 4	-	-	-	-	-	-	3	3	3	0,001	0,001	0,001	-	-	-	α	α	α
Best hyperpar. 5	-	-	-	-	-	-	15	15	Inf	-	-	-	-	-	-	-	-	-

Figure 3: R^2 values for the output “BSim h>26°C” in step A.

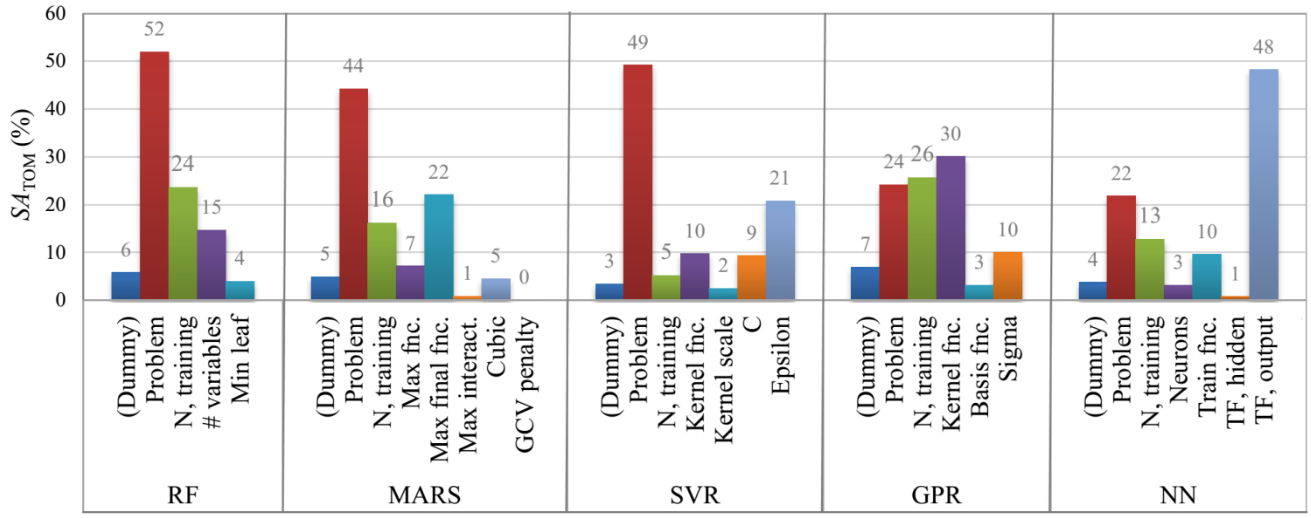


Figure 4: Relative sensitivity measures with respect to problem type, number of training points, and varied hyperparameters in step A.

5.2 Convergence with variable training sets

In step B, we construct and compare metamodels for nine training set sizes shown in Table 4. The reduced set of hyperparameter variations are shown in Table 5.

5.2.1 Accuracy

First, we consider the accuracies obtained for each of the BSim outputs modelled with variable training set size as shown on Figure 6. Again, OLS is mainly used as reference for the linearity of the problems (note that R^2 converges at 0.48 for *h>26°C*). Except for *Max* CO_2, GPR provides the most accurate models whereas RF has the poorest performance. NN, MARS, and SVR provide similar accuracies. The *Energy* output, which showed the most regular distribution on Figure 1, is the easiest to metamodel, whereas no method provided an accurate model for *Max* CO_2. In terms of convergence, the increase in accuracy levels off in the range from 512 to 2048 training points. Again, *Max* CO_2 is an exception, and remarkably, the accuracy decreases for GPR at the largest training sets.

Now, we extend the investigation to include the *Be15 Energy* output and the eight test problems. Figure 7 shows that these problems can be modeled with higher accuracies and at lower training costs. For each of these problems, a near-perfect model, with R^2 larger than 0.99, can be obtained with 256 training points or less. Note that OLS can provide an accurate model of the *Be15 Energy* output with R^2 larger than 0.95. Thus, OLS may be sufficient for this aggregated output from

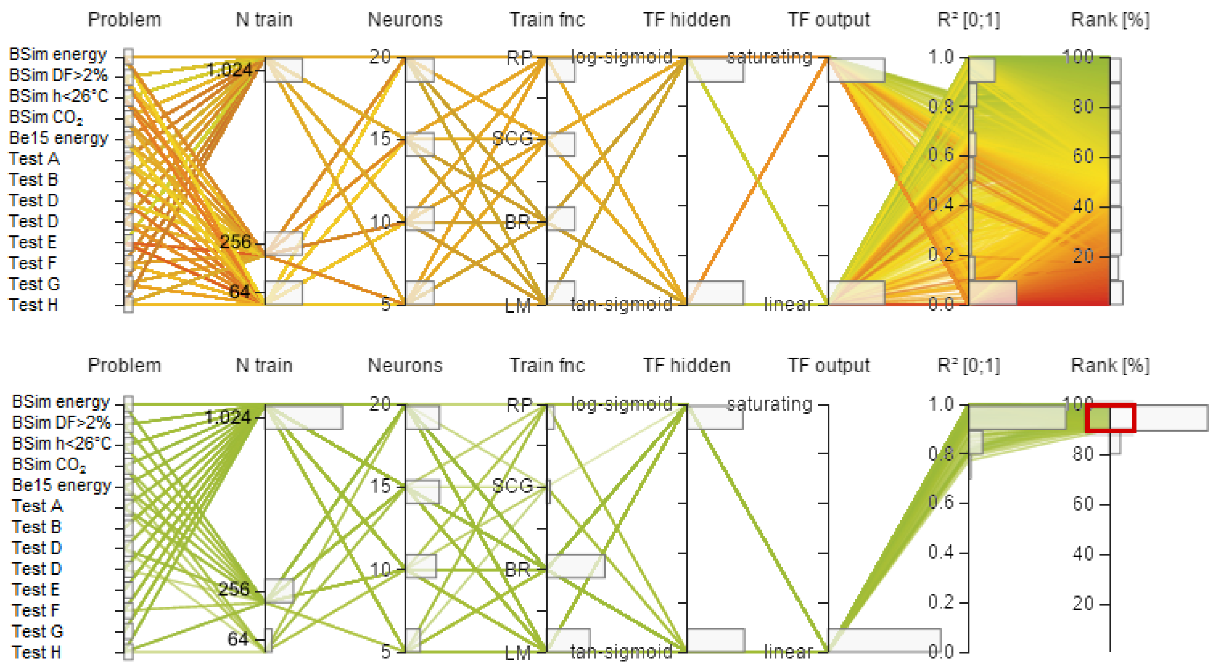


Figure 5: Interactive parallel coordinate plots used to identify suitable settings from the variations of NN in step A. The top plot has no filters whereas the bottom plot shows the top 10% best metamodels for each problem. A red rectangle indicates this filter criterion and histograms show distributions of filtered data.

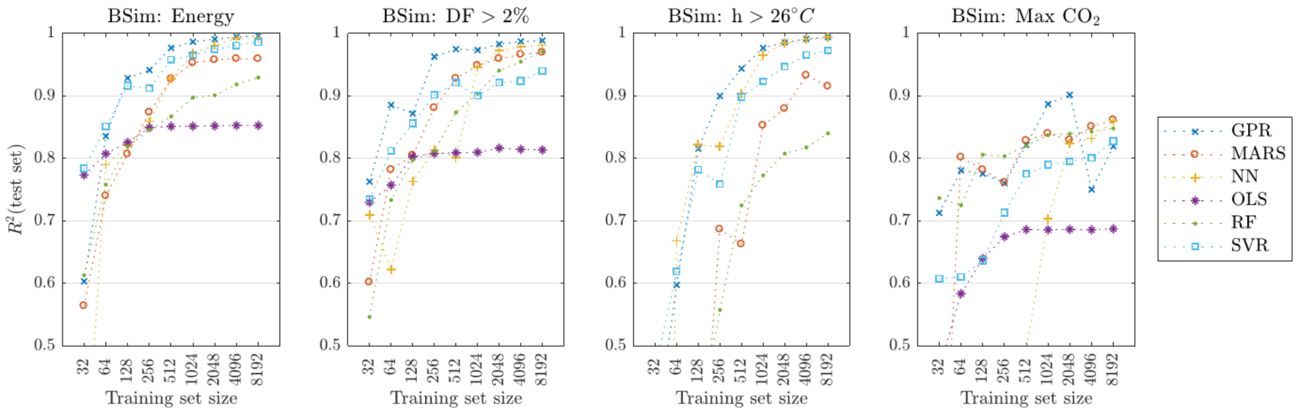


Figure 6: Convergence investigation of R^2 values with increasing training set sizes for the BSim outputs.

the idealized BPS software. On Figure 7, a grey cross is used when the accuracy of the applied technique is less than the maximum R^2 value obtained from OLS for the specific problem. This is used to indicate very poor performing metamodels (or insignificant training data), since these metamodels do not even describe the linear tendencies of the problem.

5.2.2 Robustness

Here, we wish to assess the robustness of the techniques towards diverse problems with varying levels of complexity and dimensionality. Since we concentrate on data from building simulations, we do not consider the influence of outliers, mixed data types, and noise. As indicator of robustness, previous works have used the standard deviations of the accuracies obtained from multiple problems [31][33]. Therefore, mean values and standard deviations are shown below the heat map on Figure 7. In most cases, the standard deviations decrease with increasing training sets. In general, GPR has the lowest spread, and SVR and OLS the largest[1].

A drawback of using standard deviation as robustness indicator is that it does not reveal inconsistencies such as modeling failures or decreasing accuracies. For example, the GPR method simply failed for test problem F with 4,096 and 8,192 training points. Similar errors appeared for SVR with problem D and H. Such failures are somewhat disturbing, especially when the metamodeling is to be automated as intended for the case “C” described in the introduction. Finally, we remark that the accuracies obtained for RF and NN vary when repeating the metamodeling even without changing the settings. For instance, when repeating the metamodeling of *Be15 Energy* 50 times for NN with 256 training points, the R^2 varied from 0.86 to 0.94. For RF, the corresponding approach gave results from 0.84 to 0.86. Such variations are not critical, but it might complicate the hyperparameter tuning since the “optimal” settings are not consistent.

5.2.3 Efficiency

It is troublesome to compare the computational efficiency due to software dependency, stopping criteria, amount of tuning, etc. With that in mind, we aim to compare the efficiencies with respect to training size, new predictions, and input dimensionality. Figure 8 clearly shows that OLS is computationally much more efficient than the more complex regression methods. NN also stands out since the computational effort is relatively constant and NN is less time-consuming with large training sets and many new predictions. SVR is the second-most efficient up to a few hundred training points. MARS becomes very laborious with thousands of training points, which can take hours depending on the amount of hyperparameter optimization. Training time also increases rapidly for GPR. As seen on the middle plot, GPR is the least efficient method for making new predictions when the model is trained with more than 128 observations. Again, NN is far more efficient and relatively constant. The right plot on Figure 8 shows the training time for a selection of problems with varying dimensionality. The fluctuations indicate that the training time depends much more on the complexity of the model than on the number of inputs. An increase in time can be observed for GPR. Therefore, there may be a time saving potential of removing redundant input prior to training. For example, the *Max CO2* output only depends on a few of the 14 inputs, and the redundant inputs for this output could be removed using sensitivity analysis.

5.3 Interpretability

We conclude this chapter with a brief and general description of the interpretability based on textbook material and software documentation. OLS is the simplest method and easy to interpret. The regression coefficients reveal the linear response to input changes and their magnitude indicates the relative importance of the inputs. RF is made up of a collection of regression trees (CART), which by themselves are intuitive to understand and the “tree structure”, obtained from the recursive splitting, can be illustrated with dendograms. However, the aggregation of multiple trees in RF obscures the internal structure of the model and makes RF difficult to interpret. The treebagger toolbox does, however, provide a measure of the inputs’ relative importance. The related MARS technique is fairly transparent. A MARS model consists of a sum of basis functions, which show the response in the recursively split regions. The basis functions also reveal interactions between inputs. However, understanding the hinge functions may require some practice, and the model can become unmanageable to interpret with large numbers of basis functions. A SVR model with a linear kernel function is comprehensible, but it is hard to interpret models with other kernels, which are typically necessary for sufficient flexibility of the SVR technique. Likewise, the kernel based GPR approach is difficult to comprehend. Finally, NN is possibly the least interpretable method and is often described as black box, which is flexible but provides no insight of the structure of the function being approximated.

[1] We remark that SVR can provide more accurate models for the test problems using Gaussian kernel functions.

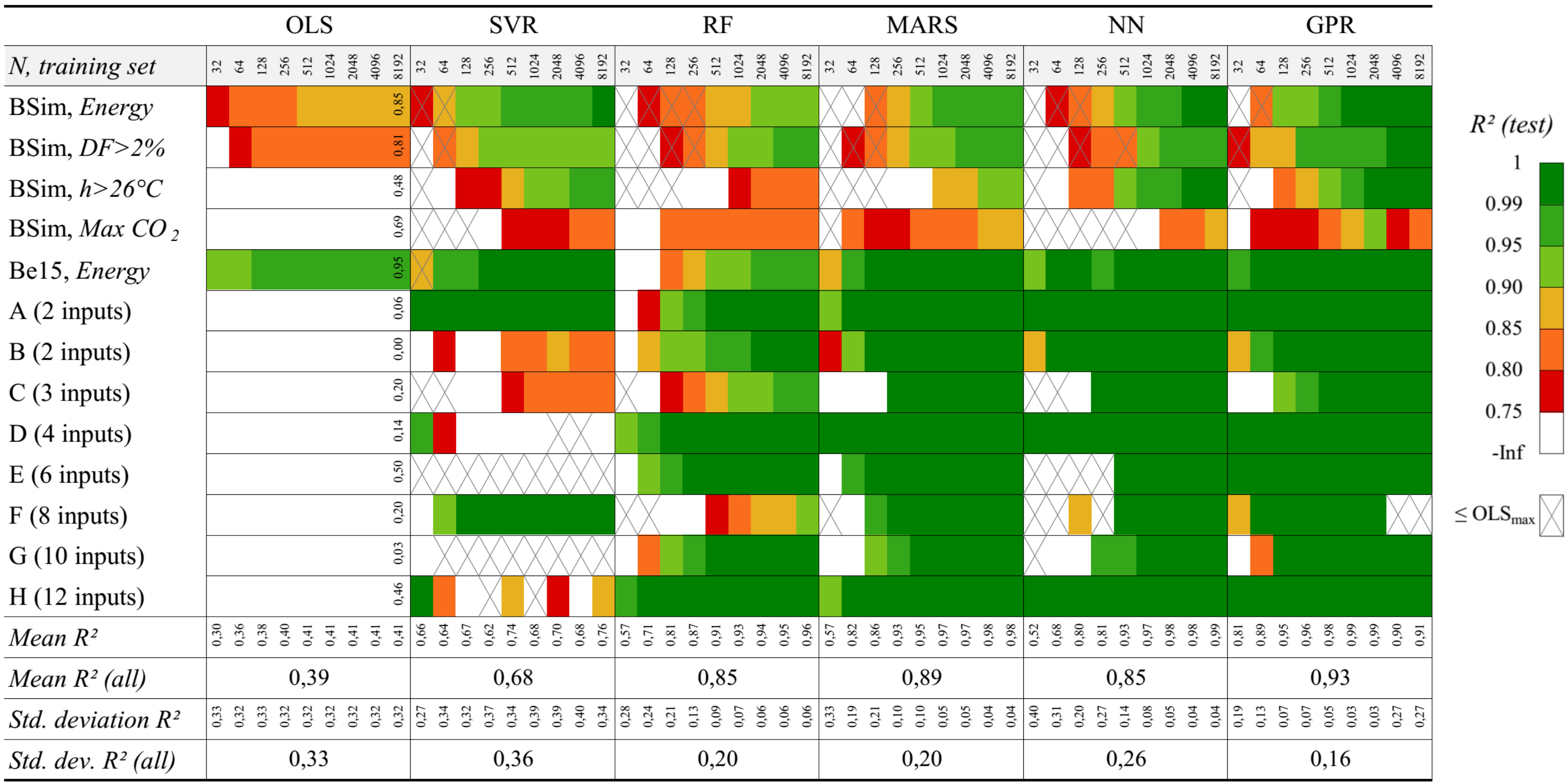


Figure 7: Accuracies obtained from test data of the metamodels constructed with varying training set sizes.

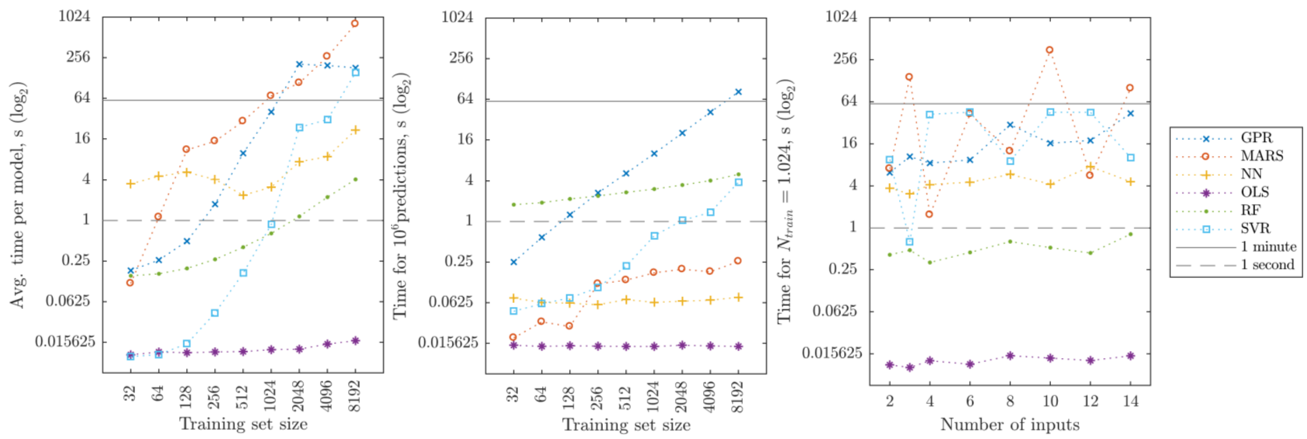


Figure 8: Left: Average time spent to train the models for the BSim problems. Middle: Time spent to perform 10^6 new predictions for the best models for the BSim problems. Right: Time spent to train models with 1024 training point for problems with varying input size.

6 Discussion

We first discuss the best metamodeling candidates for three scenarios with differing requirements of user experience, computational efficiency, and robustness (see 1.2).

A. *Expert with considerable time (~days) and emphasis on accuracy.*

In general, GPR is likely to produce the most accurate metamodels but the construction process becomes unstable and inefficient with thousands of training points. Instead, NN can be constructed relatively fast with large training sets and still provide millions of new predictions in a fraction of a second. With sufficient time and expertise, the modeler may be able to tweak the configuration of NN considerably and obtain better accuracies than presented here. In this study, the configuration was restricted to a single layer perceptron.

B. *Non-expert with limited time (~hours) and need for ease-of-use and robustness*

In a matter of hours (or overnight), it is possible to create a medium-sized training set of 128–512 building simulations. In that case, GPR seems the best candidate since it produces accurate metamodels for diverse problems with little need for hyperparameter optimization. Another advantage of GPR is its probabilistic structure, which provides uncertainty estimates, with confidence intervals, of new predictions.

C. *Automated metamodeling requiring a minimum of training points (obtained in minutes) and high robustness*

With only 32 training points, we obtained near-perfect metamodels (R>0.99) for half of the test problems. In contrast, as much as 128 points were needed from the BPS problems just to exceed the maximum accuracies obtained from OLS. It is not possible to run that many simulations within a few minutes using advanced BPS software such as BSim. In this case, it seems necessary to use faster software, e.g. the normative Be15. Combining such software with fast OLS metamodels could enable design space exploration in the matter of seconds – though, with a loss of accuracy and level of detail.

For this comparison study, we have split the data into three sets for training, validation, and testing. However, it is often advantageous to use cross-validation or bootstrapping when data is scarce, e.g. for the cases B and C. The recommended techniques differs for the three cases, which also help explain the lack of consistency of the preferred methods in the reviewed papers in Table 1. One stands out, since GPR tends to perform well in most studies when emphasizing accuracy, though it is inefficient for large data sets [27][29][31][32][33].

To round off, we discuss some pros and cons of the methodology used to compare metamodeling techniques. We set out to broaden the study when compared to the papers listed in Table 1. It has turned out rewarding to include problems from both real-life applications (BPS) and well-known theoretical functions. For example, we observed that, for SVR, the recommendable settings differed for the BPS problems and the test functions. In general, the BPS problems requires significantly more training data and high accuracies are harder to obtain. One reason is the aggregation of model outputs which causes a loss of information and leads to irregular shaped distributions, which are difficult to emulate (see 4.2). Test problems with noisy data have not been considered, since emphasis is on building simulations. However, the inclusion of "noisy" problems, and problems from other engineering disciplines, would produce better insight into the capabilities of the techniques.

Each technique have a great variety of method-specific settings[2]. This makes it hard to conduct a fair comparison to reveal the potential of each methods and to identify suitable settings for a collection of problems. We chose to discretize up to five settings and used a grid search to evaluate all combinations. An alternative approach using optimization algorithms might have produced more accurate models by identifying better values for the continuous variables. Though, optimization algorithms would add yet another layer of complexity to the study, and the chosen algorithm might work better for some techniques than others, and thus obscure the comparison.

The use of sensitivity analysis has enabled us to fixate the least significant model settings, thereby making hyperparameter optimization less burdensome at larger training sets. This use of sensitivity analysis for factor fixing may be automated in metamodeling. In contrast, the visual and user-dependent investigation, using parallel coordinate

[2] Settings may refer to continuous, discrete, and categorical variables and entire setups, such as ensemble methods for RF or network structure for NN.

plots, do not work for automated implementations of metamodels. If time allows it, it does provide valuable information on suitable settings when used in a grid based approach.

7 Conclusion

This work builds upon the learnings from 10 papers which compare a variety of metamodeling techniques for supervised learning. Across those studies, there was no consensus on a single "best" technique but we did identify six prevalent and popular methods. These six were chosen as possible candidates for the construction of fast metamodels to overcome the computational barrier related to design space exploration, design optimization, and sensitivity analysis. To perform an extensive metamodeling comparison, we combined best practices from the literature review and considered 13 diverse problems, with varying dimensionality and complexity, using nine different sizes of training data. Since the "best" technique depends on the given context and its purposes, we considered metamodeling performance for three scenarios with different requirements of time consumption and user interaction.

A methodological approach was required to conduct a fair, in-depth comparison of metamodeling techniques which may be configured endlessly to improve the accuracy of new predictions. A grid search approach was preferred to assess all combinations of several discretized model settings. The alternative use of an optimization algorithm might have produced slightly more accurate metamodels by optimizing continuous variables, but it would have added another layer of complexity to the comparison and made it less transparent and reproducible. The grid search allowed us to perform regionalized sensitivity analysis to fixate the least important settings and identify suitable options for the most influential ones. This allowed us to assess efficiency, robustness, and accuracy for training sets ranging from 2^5 to 2^{13} points.

The following summarizes the key findings related to model settings:

- In general, default settings yielded poor or mediocre accuracies. Thus, the defaults were of little use which stressed the need for hyperparameter optimization.
- Sensitivity analysis helped fixate the least influential settings and identify suitable options for important settings for a collections of problems.
- For SVR, the most suitable settings differed notably for the BPS problems and the mathematical test functions. This implies that SVR are less robust and more settings must be considered for diverse problems.

The following findings relate to accuracy and efficiency:

- In general, GPR produced the most accurate metamodels, followed by NN and MARS. Linear regression using ordinary least squares (OLS) was mainly used as a benchmark and provided poor accuracy due to the non-linearity of the problems.
- GPR metamodels with R^2>0.99 were obtained for all eight mathematical benchmark functions using between 32 and 256 training points. In contrast, the BPS problems required much larger training sets: 2,048 to achieve R^2>0.90 for *Max* CO_2, and between 1,024 and 4,096 to achieve R^2>0.99 for the other outputs obtained from the advanced BPS software (BSim).
- OLS was, by far, the fastest method for both training and new predictions. For the most accurate, non-linear methods, NN proved the most efficient for large training sets whereas GPR became slow and even "unstable".
- The dimensionality, i.e. the number of model inputs, had only a small influence on both accuracy and efficiency.

Hopefully, this comparison study will provide a good starting point for novices in metamodeling and contribute to the proliferation of supervised learning. The applied methodology and the findings are relevant to both researchers and practitioners – not limited to the field of building performance simulations. A logical next step would be to extend the comparison to address data which include noise, e.g. measurements of energy consumption or perceived indoor comfort. To gain more insight into the performance of metamodeling, we encourage others to extend this comparison by testing additional problems and techniques. For comparability, we advocate the use of the problem-independent coefficient of determination, R^2, to report accuracy.

Acknowledgements

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

References

[1] Saheb Y, Bódis K, Szabo S, Ossenbrink H, Panev S. Energy Renovation: The Trump Card for the New Start for Europe. 2015. doi:10.2790/39989.

[2] Building Energy Efficiency Taskgroup. Building Energy Performance Metrics Supporting Energy Efficiency Progress in Major Economies. 2015.

[3] Østergård T, Jensen RL, Maagaard SE. Building simulations supporting decision making in early design – A review. Renew Sustain Energy Rev 2016;61:187–201. doi:10.1016/j.rser.2016.03.045.

[4] Tian W. A review of sensitivity analysis methods in building energy analysis. Renew Sustain Energy Rev 2013;20:411–9. doi:10.1016/j.rser.2012.12.014.

[5] Evins R. A review of computational optimisation methods applied to sustainable building design. Renew Sustain Energy Rev 2013;22:230–45. doi:10.1016/j.rser.2013.02.004.

[6] Machairas V, Tsangrassoulis A, Axarli K. Algorithms for optimization of building design: A review. Renew Sustain Energy Rev 2014;31:101–12. doi:10.1016/j.rser.2013.11.036.

[7] Nguyen A-T, Reiter S, Rigo P. A review on simulation-based optimization methods applied to building performance analysis. Appl Energy 2014;113:1043–58. doi:10.1016/j.apenergy.2013.08.061.

[8] Yang J. Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environ Model Softw 2011;26:444–57. doi:10.1016/j.envsoft.2010.10.007.

[9] Naboni E, Zhang Y, Maccarini A, Hirsh E, Lezzi D. Extending the use of parametric simulation in practice through a cloud based online service. Proc. first IBPSA-Italy Conf. BSA 2013, 2013, p. 105–12.

[10] Østergård T, Jensen RL, Maagaard SE. Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis. Energy Build 2017;142:8–22. doi:10.1016/j.enbuild.2017.02.059.

[11] MOE A/S. Demonstration of Proactive Building Simulations 2016. http://buildingdesign.moe.dk/PhD-Project/Demonstration-of-Proactive-Building-Simulations (accessed October 12, 2017).

[12] Østergård T, Maagaard SE, Jensen RL. Thermal Comfort in Residential Buildings by the Millions - Early Design Support from Stochastic Simulations. CLIMA 2016 - Proc 12th REHVA World Congr 2016;6.

[13] Østergård T, Jensen RL, Maagaard SE. Test problems for metamodeling comparison: 5 building performance metrics and 8 theoretical problems 2017. doi:10.17632/7twjr7g7tv.2.

[14] Asadi S, Amiri SS, Mottahedi M. On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. Energy Build 2014;85:246–55. doi:10.1016/j.enbuild.2014.07.096.

[15] Hygh JS, DeCarolis JF, Hill DB, Ranjithan SR. Multivariate regression as an energy assessment tool in early building design. Build Environ 2012;57:165–75. doi:10.1016/j.buildenv.2012.04.021.

[16] Mavromatidis LE, Marsault X, Lequay H. Daylight factor estimation at an early design stage to reduce buildings' energy consumption due to artificial lighting: A numerical approach based on Doehlert and Box–Behnken designs. Energy 2014;65:488–502. doi:10.1016/j.energy.2013.12.028.

[17] Hester J, Gregory J, Kirchain R. Sequential Early-Design Guidance for Residential Single-Family Buildings Using a Probabilistic Metamodel of Energy Consumption. Accept Energy Build 2016;134:202–11. doi:10.1016/j.enbuild.2016.10.047.

[18] Eisenhower B, O'Neill Z, Narayanan S, Fonoberov V a., Mezić I. A methodology for meta-model based optimization in building energy models. Energy Build 2012;47:292–301. doi:10.1016/j.enbuild.2011.12.001.

[19] Van Gelder L, Janssen H, Roels S. Probabilistic design and analysis of building performances: Methodology and application example. Energy Build 2014;79:202–11. doi:10.1016/j.enbuild.2014.04.042.

[20] Turhan C, Kazanasmaz T, Uygun IE, Ekmen KE, Akkurt GG. Comparative study of a building energy performance software (KEP-IYTE-ESS) and ANN-based building heat load estimation. Energy Build 2014;85:115–25. doi:10.1016/j.enbuild.2014.09.026.

[21] Yuan J, Nian V, Su B, Meng Q. A simultaneous calibration and parameter ranking method for building energy models. Appl Energy 2017;206:657–66. doi:10.1016/j.apenergy.2017.08.220.

[22] Chen X, Yang H, Sun K. Developing a meta-model for sensitivity analyses and prediction of building performance for passively designed high-rise residential buildings. Appl Energy 2017;194:422–39. doi:10.1016/j.apenergy.2016.08.180.

[23] Manfren M, Aste N, Moshksar R. Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation. Appl Energy 2013;103:627–41. doi:10.1016/j.apenergy.2012.10.031.

[24] Yang Z, Becerik-Gerber B. A model calibration framework for simultaneous multi-level building energy simulation. Appl Energy 2015;149:415–31. doi:10.1016/j.apenergy.2015.03.048.

[25] Massa Gray F, Schmidt M. Thermal building modelling using Gaussian processes. Energy Build 2016;119:119–28. doi:10.1016/j.enbuild.2016.02.004.

[26] Cheng M-Y, Cao M-T. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. Appl Soft Comput 2014;22:178–88. doi:10.1016/j.asoc.2014.05.015.

[27] Wei L, Tian W, Zuo J, Yang Z-Y, Liu Y, Yang S. Effects of Building Form on Energy Use for Buildings in Cold Climate Regions. Procedia Eng 2016;146:182–9. doi:10.1016/j.proeng.2016.06.370.

[28] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew Sustain Energy Rev 2017;73:1104–22. doi:10.1016/j.rser.2017.02.023.

[29] Lim H, Zhai ZJ. Comprehensive evaluation of the influence of meta-models on Bayesian calibration. Energy Build 2017;155:66–75. doi:10.1016/j.enbuild.2017.09.009.

[30] Duarte, Grasiele Regina; Fonseca, Leonardo Goliatt; Capriles P, Lemonge AC de C. Comparison of machine learning techniques for predicting energy loads in buildings. Ambient Construído 2017;17:103–15. doi:10.1590/s1678-86212017000300165.

[31] Jin R, Chen W, Simpson TW. Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidiscip Optim 2001;23:1–13. doi:10.1007/s00158-001-0160-4.

[32] Kim BS, Lee Y Bin, Choi DH. Comparison study on the accuracy of metamodeling technique for non-convex functions. J Mech Sci Technol 2009;23:1175–81. doi:10.1007/s12206-008-1201-3.

[33] Li YF, Ng SH, Xie M, Goh TN. A systematic comparison of metamodeling techniques for simulation optimization in Decision Support Systems. Appl Soft Comput J 2010;10:1257–73. doi:10.1016/j.asoc.2009.11.034.

[34] Villa-Vialaneix N, Follador M, Ratto M, Leip A. A comparison of eight metamodeling techniques for the simulation of N2O fluxes and N leaching from corn crops. Environ Model Softw 2012;34:51–66. doi:10.1016/j.envsoft.2011.05.003.

[35] Ju Y, Zhang C, Ma L. Artificial intelligence metamodel comparison and application to wind turbine airfoil uncertainty analysis. Adv Mech Eng 2016;8:1–14.

doi:10.1177/1687814016647317.

[36] Jekabsons G. ARESLab - Adaptive Regression Splines toolbox for Matlab/Octave 2016:33.

[37] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning Data Mining, Inference,and Prediction. 2nd ed. Springer; 2009. doi:10.1007/978-0-387-84858-7.

[38] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. 1st ed. New York: 2013. doi:10.1007/978-1-4614-7138-7.

[39] Friedman JH. Multivariate Adaptive Regression Splines. Ann Stat 1991;19:1–67.

[40] Alpaydin E. Introduction to Machine Learning. Third edit. The MIT Press; 2014.

[41] Rasmussen CE, Williams CKI. Gaussian processes for machine learning. vol. 14. Cambridge: MIT Press; 2006.

[42] Shan S, Wang GG. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. Struct Multidiscip Optim 2010;41:219–41. doi:10.1007/s00158-009-0420-2.

[43] Danish Building Research Institute. BSim 2017. http://sbi.dk/bsim/Pages/Start.aspx (accessed June 12, 2017).

[44] Aggerholm S, Grau K, Wittchen KB. SBi-anvisning 213: Bygningers energibehov. Hørsholm: 2014.

[45] Bhattacharya M. Meta Model Based EA for Complex Optimization. Int J Inf Math Sci 2008;4:36–45.

[46] AI-Roomi. Power Systems and Evolutionary Algorithms n.d. http://al-roomi.org/ (accessed April 12, 2017).

[47] Gavana A. Global Optimization Benchmarks and AMPGO n.d. http://infinity77.net/global_optimization (accessed April 12, 2017).

[48] Saltelli A, Chan K, Scott EM. Sensitivity Analysis. 2000. doi:10.1002/047134608X.W2510.

[49] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. Global sensitivity analysis: the primer. Chichester, England: John Wiley & Sons Ltd.; 2008.

[50] Wang GG, Shan S. Review of Metamodeling Techniques in Support of Engineering Design Optimization. J Mech Des 2007;129:370. doi:10.1115/1.2429697.

[51] Sobol' IM, Shukman BV. Random and quasirandom sequences: Numerical estimates of uniformity of distribution. Math Comput Model 1993;18:39–45. doi:10.1016/0895-7177(93)90160-Z.

[52] Kucherenko S, Albrecht D, Saltelli A. Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. 8th IMACS Semin Monte Carlo Methods 2015:1–32. doi:10.1016/j.ress.2017.04.003.

[53] Razavi S, Tolson BA, Burn DH. Review of surrogate modeling in water resources. Water Resour Res 2012;48. doi:10.1029/2011WR011527.

[54] Østergård T, Jensen RL, Maagaard SE. Interactive building design space exploration using regionalized sensitivity analysis. Proc. Build. Simul., San Francisco: 2017.

3.3.2 CONCLUDING REMARKS

Paper D has shown that accurate and rapid metamodels can be generated for diverse building performance indicators. Though, we note that a considerable amount of building simulations is needed to train the metamodels. For instance, hundreds of simulations are required to obtain reasonable accurate models for the “generic” office modelled with BSim. Another drawback is that metamodels are only valid in the domain used to generate the models. Thus, new metamodels must be generated if the choice of design parameters changes of if the ranges of variation are extended.

To round off, we have a few remarks to bridge the gap between Paper C and Paper D in terms of metamodeling. The observant reader may have noticed that we did not include the SDP (state-dependent parameter regression) method in Paper D even though it showed promising accuracy in Paper C. The reason is that SDP did not frequently occur in the literature review [24]. Moreover, we knew from experience that SDP metamodeling easily takes hours with a few thousand training sets. In addition, the available Matlab code was limited to include only first and second order effects.

Another remark relates to the issue of filter criteria, which can have large impact on the histograms used to guide decision-makers during Monte Carlo Filtering [23]. In Paper C, we argued that the specific criteria can be adjusted to obtain better matching histograms. As a result, the linear regression models could yield similar histograms as obtained when filtering the original simulations (see Figure 11 in Paper C). This worked despite the inaccurate regression model for the output *Overtemperature* for which the R^2-value was only 0.41. It was hypothesized that with a more accurate metamodel the inconvenient filter adjustments would be unnecessary. Here, we extend the analysis with accurate GPR metamodels. The GPR models are trained with 1.000 simulations which result in R^2 values larger than 0.98 for test data. Figure 3-4Figure shows input distributions of the “behavioral” simulations obtained with the original model, Be10, and the metamodel methods, SRC, SDP, and GPR.

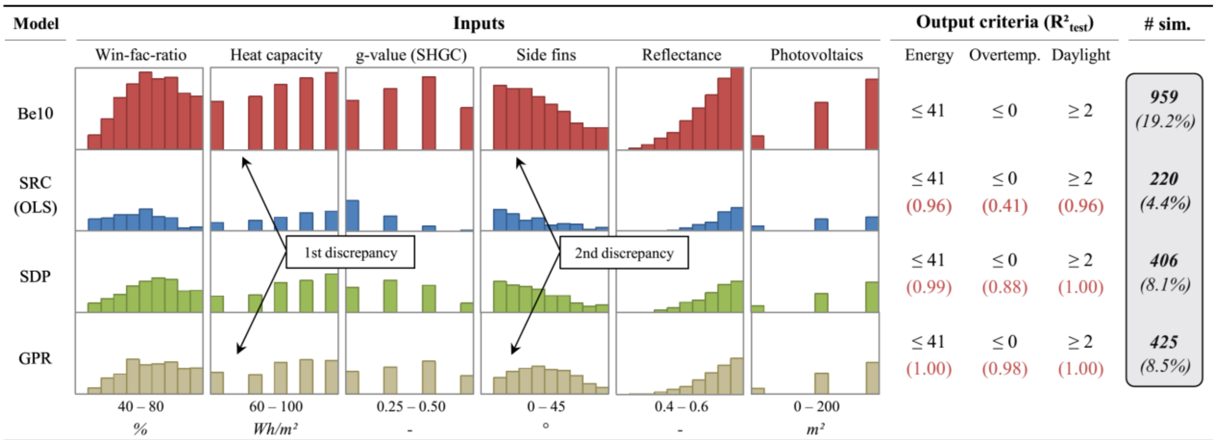


Figure 3-4: Selected input distributions when using the same filter criteria (extension of the Figure 11 in [23]*)*

From Figure 3-4Figure , we notice that the more accurate models from GPR provide better approximation of the g-value histogram as compared to SRC and SDP. Despite the use of accurate metamodels, we still observe discrepancies between histograms related to the original model, Be10, and GPR. Such discrepancies are observed in the low ranges for Heat capacity and Side fins. Moreover, we notice that the number of “behavioral” simulations for GPR (425) is still much lower than for the original model (959). To understand why these differences occur despite the use of highly accurate metamodels, we look at the distributions for *Overtemperature* on Figure 3-5Figure . The tall bin for Be10 is caused by 4269 simulations that yield a value of exactly zero. In contrast, the regression models provide continuous outputs. As consequence, roughly half of these “zero” simulations are slightly less than zero and the others are slightly larger than zero for SDP and GPR.

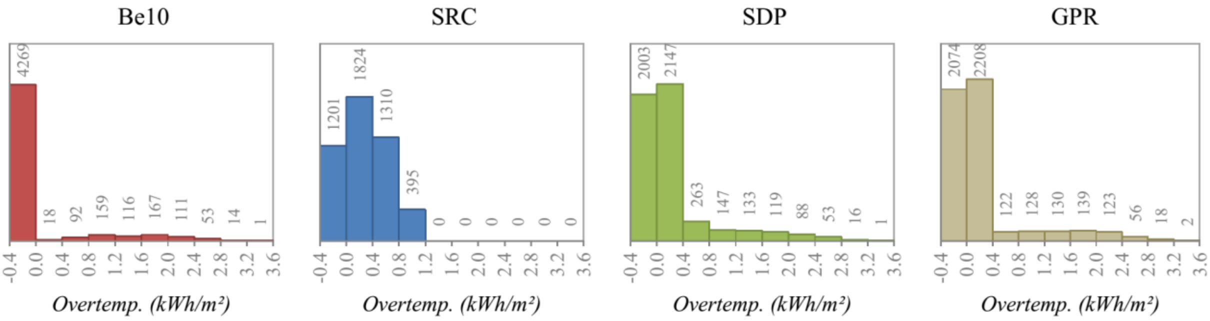


Figure 3-5: Distributions for "Overtemperature" for the original model, Be10, and three metamodels.

In conclusion, awareness is required when building code criteria corresponds to a discrete value that dominates the distribution. Fortunately, this rare issue is easy to observe from the output distributions and it may be circumvented by adding a small "buffer" to the filter value. Thus, metamodeling is still considered a necessity when exploring multidimensional building performance simulations. In paper D, we mentioned how metamodeling may be implemented in three stages, which we elaborate on in the next chapter.

4 NEXT LOGICAL STEPS

To follow up on the developments presented in the previous chapters, we round off with a brief description of current and upcoming work related to holistic, multi-dimensional building simulations. In Paper D, which mainly involves metamodeling, we briefly describe how metamodeling may be applied in a building performance context in three separate stages or "development steps". In this chapter, we elaborate on the logic and novelty of these with emphasis on the first step, which relates to our current, non-published work. Again, the work is influenced by the needs of the host company, MOE, and the common approach to building simulations in Denmark.

4.1 NEXT STEP – GENERIC PRE-SIMULATED ROOMS

The first step is the development of a design method that combines "pre-simulated" generic rooms with the project-specific, whole-building simulations. This approach resembles the method described in Paper F, in which pre-calculated Monte Carlo simulations describe the variability and thermal performance of "critical" rooms. That method, however, was limited to address thermal comfort in Danish residential buildings. Here, the idea of "pre-simulated" spaces is extended by encompassing diverse building functions and addressing multiple performance objectives.

We aim to define and pre-simulate "generic" rooms that represent building spaces that are often "critical" and therefore affect building design. Examples are office spaces, meeting rooms, and lecture rooms, whereas secondary spaces, such as toilets and corridors are omitted. Each "generic" room is described by variable design parameters, which lead to a room-specific, multidimensional design space. To assess the performance of all room configurations, we perform Monte Carlo simulations after which metamodels are created for each performance indicator. For the office example used in Paper D, metamodels were constructed for a few, diverse performance indicators; total energy demand, thermal comfort, daylight factor, and air quality. Actually, other common performance indicators have also been assessed and stored in a database along with the design inputs for all generic rooms. The design team may then choose to include only the performance indicators that suit the given project. The diversity of performance indictors can exemplified by the many contributors to energy demand (e.g. lighting, heating, and cooling) or the different aggregation of outputs (e.g. h>26 °C, h>27 °C, $CO2$>900 ppm, $CO2$>1.000 ppm, etc.).

The intended use of generic rooms is to combine them with whole-building simulations during early design phases. The existence of a BIM model is not necessary. The design team identifies potential "critical" or "representative" rooms based on building shape and a rough disposition of room functions. Often, budget and time constraints limit the number of rooms to consider. For most projects in MOE, this number of rooms is between three and six. The idea is to analyze these rooms using parallel coordinate plots for the corresponding generic rooms. At the same time, a PCP for the whole-building simulations is added to the same graphical interface as shown on Figure 4-1. The design team can then simultaneously test various designs for each room and the whole building.

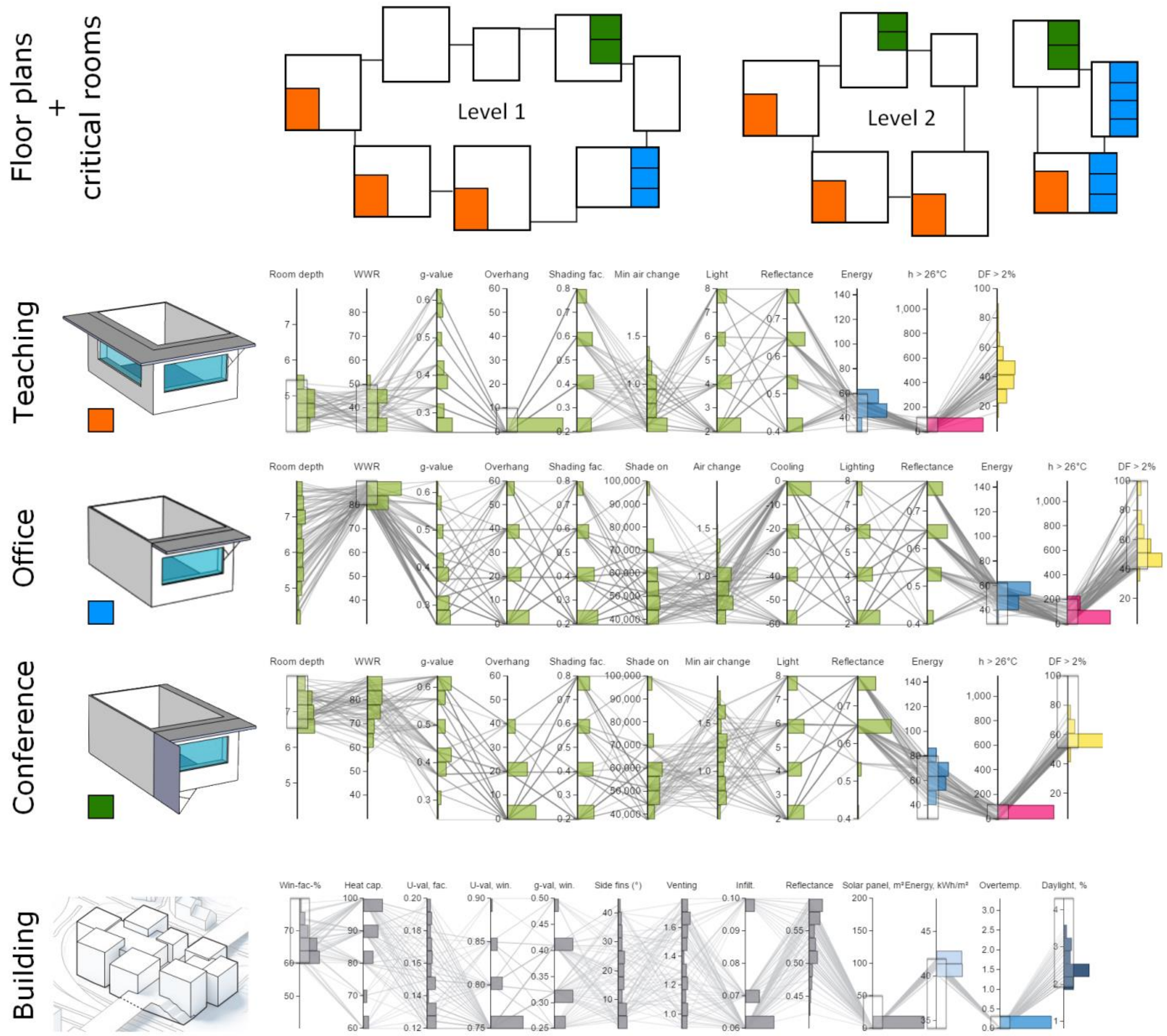


Figure 4-1: Possible setup of graphical user interface which include floor plans with highlights of critical rooms, PCPs for generic rooms, and a PCP for the whole building.

The ideas presented here are still untested and much work remains. This includes defining the rooms and their variability, performing the simulations, constructing the metamodels, and setting up the database and user interface. We also need to consider how to connect correlated design variables and how to model the building such that it fits the variability of the generic rooms.

4.2 FURTHER STEPS

The second step of the planned developments is a refinement of the above approach. Here, the generic rooms are substituted by project-specific spaces which provide better match to the actual geometry, shading obstacles, schedules, internal loads, etc. In absence of pre-defined rooms and direct CAD-BPS interoperability, the modeller must manually set up a baseline model in BSim (or equivalent), perform Monte Carlo simulations, and construct metamodels. The presentation and analysis of designs will be the same as above. This approach is assumed to be applicable in both early and detailed design stages. However, it requires further development of the multivariate BSim framework initiated in a aforementioned student project [11]. Alternatively, OpenStudio may be an option with its parametric and cloud capabilities even though OpenStudio has limited penetration in Denmark.

Finally, we strive to reduce manual labour by automating the global simulations of selected rooms directly in the BIM model. Making this third step relies on positive experience and feedback from the above developments. In terms of software development, it requires a BPS API which is linked directly into Revit (or similar BIM software). Such efforts are supposed to streamline the process of Monte Carlo simulations and metamodeling for project-specific, critical rooms.

5 CONCLUSION

With this industrial PhD project, we set out to address various challenges experienced by consulting engineers specialized in energy demand, indoor climate, and sustainability in buildings. The challenges relate to the guidance of multi-actor design teams and then handling of the enormous design space in the early stages of holistic building design. To accommodate such issues, we advocate the use of Monte Carlo simulations aided by statistical methods and interactive visualizations. This allows the design team to explore an enormous, multivariate design space represented by thousands or millions of building performance simulations (BPS). The interactive parallel coordinate plot (PCP) has shown useful for real-time exploration of such comprehensive design space. Using the PCP, multiple stakeholders apply *filters* to investigate specific parts of the design space and immediately observe consequences for both design parameters and performance indicators.

Sensitivity analysis plays a vital role in the proposed simulation framework. Popular methods, such as Morris' and variance decomposition, help explain model behavior and reveal both important and insignificant design parameters. The Monte Carlo *filtering* used in the PCP relates to regionalized sensitivity analysis. To address multiple objectives and to aid real-time decision-making, we have developed two novel sensitivity methods denoted TOM and TOR, respectively. The former is used to prioritize design parameters according to multiple outputs, which directs attention to the parameters that matter the most in holistic design. The latter method, TOR, improves the use of the PCP by highlighting the parameters affected by the user-defined filtering. Another valuable field from statistics relates to metamodeling, which has proven crucial to obtain immediate feedback and to overcome the computational issues of Monte Carlo simulations. In a BPS context, the best candidates for metamodeling are Gaussian process regression and neural networks based on a comprehensive comparison study.

The PhD study has been completed with the Danish-based company, MOE, as the primary stakeholder. The proposed simulation framework and interactive exploration is now possible for the normative Be15 software, and it has been applied in several projects by MOE engineers. In addition, a design tool for assessing thermal comfort in Danish residential buildings is freely available on the website buildingdesign.moe.dk. Same site contains supplementary information about the PhD project including case studies, interactive plots, and related projects. Future developments and tools will be presented and made available on this site – to some extent.

A concluding remark is that the proposed methods are considered applicable for other research areas and industries, which involve model-driven decision making under uncertainty and with respect to multiple objectives and stakeholders.

LITERATURE LIST

[1] MOE A/S, “MOE Consulting | Engineers.” Available: moe.dk [Accessed: 05-May-2017].

[2] Byggefakta, “Trends for byggebranchen 2012-2016,” (in Danish), Valby, 2016.

[3] R. J. Cole and M. Jose Valdebenito, “The importation of building environmental certification systems: international usages of BREEAM and LEED,” *Building Research & Information*, vol 41(6), pp. 662–676, Jun. 2013. doi: 10.1080/09613218.2013.802115

[4] T. Østergård, S. E. Maagaard, and R. L. Jensen, “A stochastic and holistic method to support decision-making in early building design,” in *Proceedings of Building Simulation*, 2015, pp. 1885–1892.

[5] Danish Building Research Institute, “Beregningsprogrammet Be15.” (in Danish), Available: http://sbi.dk/beregningsprogrammet/Pages/Start.aspx. [Accessed: 14-Jun-2017].

[6] CEN, “ISO 13790:2008 Energy performance of buildings -- Calculation of energy use for space heating and cooling,” Geneva, Switzerland, 2008.

[7] Danish Building Research Institute, “BSim.” (in Danish), Available: http://sbi.dk/bsim. [Accessed: 14-Jun-2017].

[8] T. Østergård, R. L. Jensen, and S. E. Maagaard, “Building simulations supporting decision making in early design – A review,” *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 187–201, 2016. doi: 10.1016/j.rser.2016.03.045

[9] M. J. Sørensen and S. H. Myhre, “Occupancy and weather: How these influence robustness and building design,” Master’s thesis, Aalborg University, 2017.

[10] K. K. Hansen and M. H. Silkjær, “Stochastic analysis of building energy modelling tools,” Master’s thesis, Aalborg University, 2017.

[11] E. G. Bonde, M. Buhrkal-Donau, and F. S. Mikkelsen, “Automation of stochastic modelling in BSim - streamlining the design process,” (in Danish), Master’s thesis, Aalborg University, 2016.

[12] E. Petrova, P. L. Johansen, R. L. Jensen, S. E. Maagaard, and K. Svidt, “Automation of geometry input for building code compliance check,” in *Lean & Computing in Construction Congress*, 2017.

[13] National Renewable Energy Laboratory, “Building Component Library,” 2015. Available: https://bcl.nrel.gov/. [Accessed: 14-May-2017].

[14] A. K. Tangevold, “Uncertainties and changes in the design process,” Master’s thesis, Aalborg University, 2017.

[15] N. Ø. Hansen and T. H. Broholt, “Analyse af energiforbrug,” (in Danish), Bachelor project, Aarhus University, 2015.

[16] M. Bostock, “D3.js - Data-Driven Documents,” 2015. Available: https://d3js.org/. [Accessed: 11-May-2017].

[17] MOE A/S, “Demonstration of Proactive Building Simulations,” 2016. Available: https://buildingdesign.moe.dk/PhD-Project/Demonstration-of-Proactive-Building-Simulations. [Accessed: 09-May-2017].

[18] S. Attia, J. L. M. Hensen, L. Beltrán, and A. De Herde, “Selection criteria for building performance simulation tools: contrasting architects’ and engineers’ needs,” *Journal of Building Performance Simulation*, vol. 5(3). January 2015, pp. 155–169, 2012. doi: 10.1080/19401493.2010.549573

[19] T. Østergård, R. L. Jensen, and S. E. Maagaard, “Interactive building design space exploration using regionalized sensitivity analysis,” in *Proceedings of Building Simulation*, 2017.

[20] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference,and Prediction*, 2nd ed. Springer, 2009.

[21] T. Østergård, S. E. Maagaard, and R. L. Jensen, “Thermal Comfort in Residential Buildings by the Millions - Early Design Support from Stochastic Simulations,” *CLIMA 2016 - proceedings of the 12th REHVA World Congress*, vol. 6, 2016.

[22] MOE A/S, “Sommerkomfort Tool,” 2016. Available: https://buildingdesign.moe.dk/Arch-Engineering/Tools/Sommerkomfort-Tool. [Accessed: 18-May-2017].

[23] T. Østergård, R. L. Jensen, and S. E. Maagaard, “Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis,” *Energy and Buildings*, vol. 142, pp. 8–22, May 2017. doi: 10.1016/j.enbuild.2017.02.059

[24] T. Østergård, R. L. Jensen, and S. E. Maagaard, “A comparison of six metamodeling techniques applied to building performance simulations,” Submitted for *Environmental Modeling & Software*, May 2017.

APPENDICES

Appendix A. Paper E

Interactive building design space exploration
using regionalized sensitivity analysis

Torben Østergård, Rasmus L. Jensen, and Steffen E. Maagaard

Proceedings of Building Simulation 2017, August 2017

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

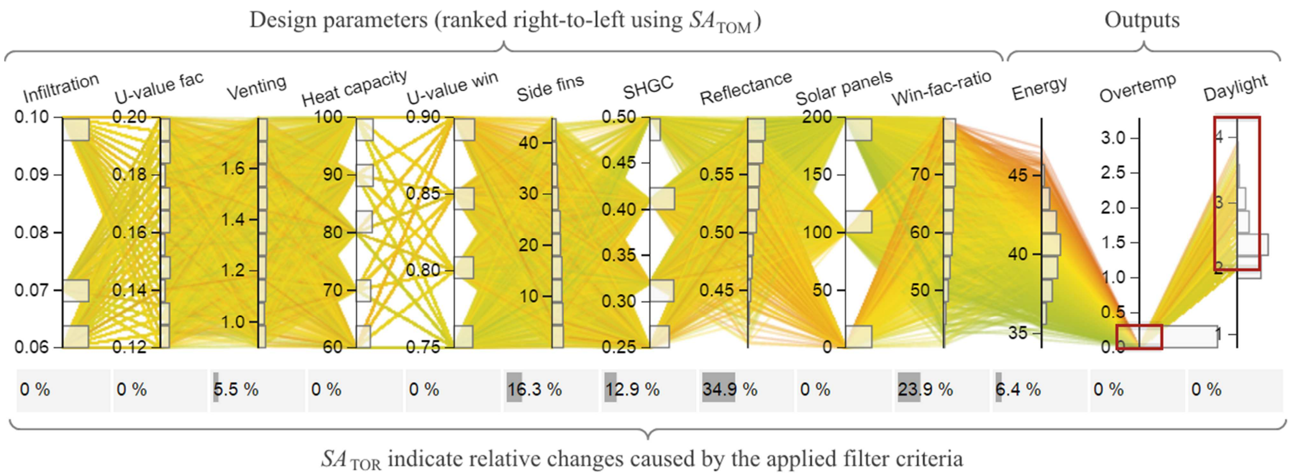


Figure 1: Parallel coordinate plot (PCP) with histograms showing distribution of the simulations, which remain after filtering. The bar plots how much the distributions have been affected by the filters (red rectangles). Each line in the PCP represents one simulation and is coloured according to its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

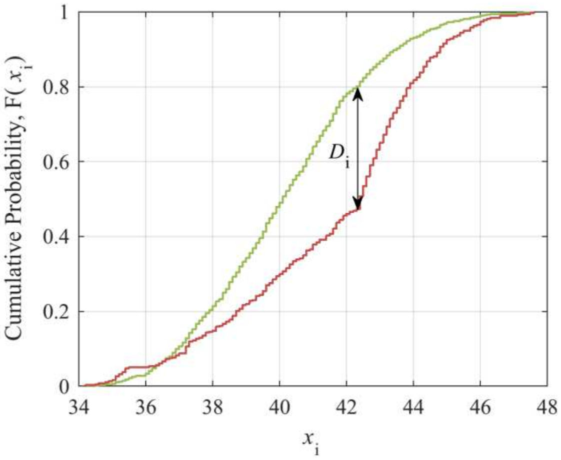


Figure 2: The maximum distance D between two cumulative distributions.

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.

$$SA_{KS2,i} = \frac{D_i}{\sum_i D_i} \qquad (1)$$

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,

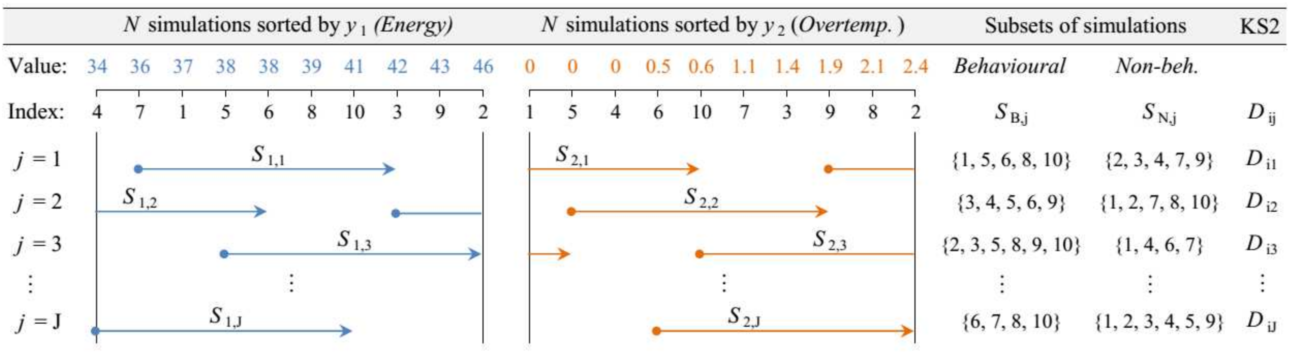


Figure 3: 10 simulations sorted for two outputs. Subsets S_1,j and S_2,j are randomly selected for each repetition, j. The subsets are illustrated with arrows, which have random starting points but same length (Q simulations).

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.

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

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).

$$Q = N \cdot 0.5^{1/m} \tag{3}$$

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).

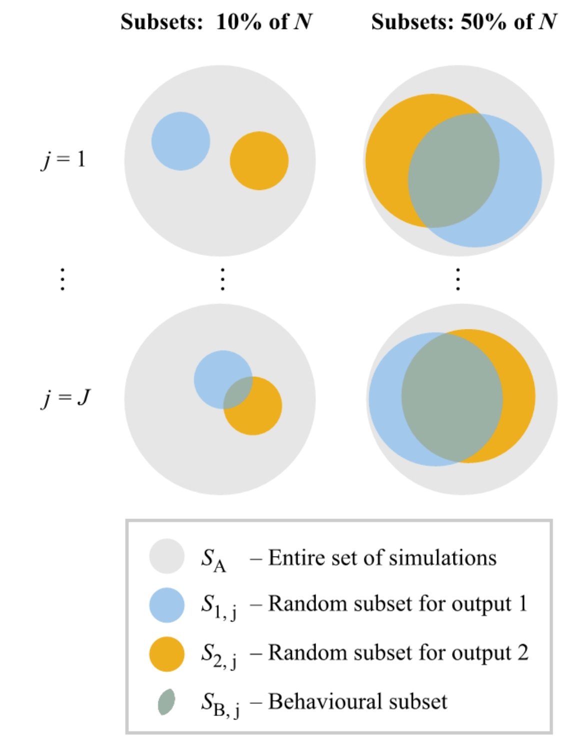


Figure 4: Conceptual illustration of how intersections of randomly selected subsets change. Smaller subsets may have no intersection and thus no behavioral simulations

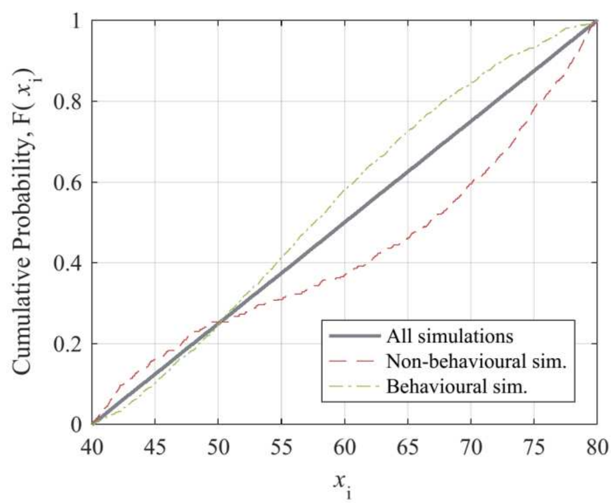


Figure 5: Cumulative distributions for S_A, S_B, and S_N for a uniformly distributed input, x_i.

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

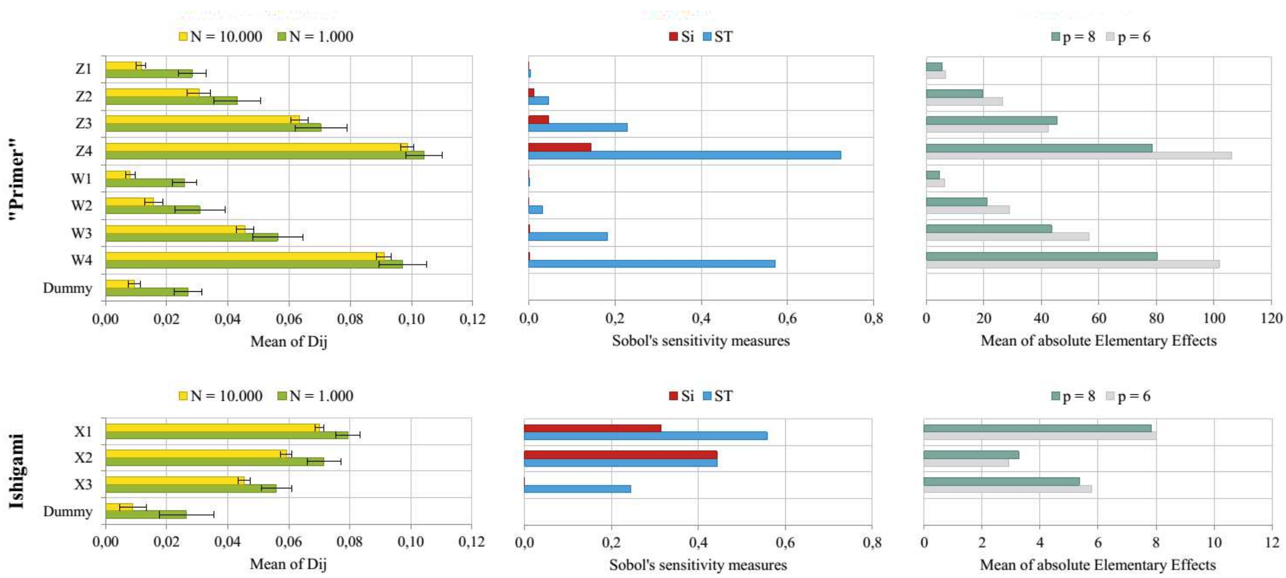


Figure 6: Results of sensitivity analysis for two benchmark models, using the TOM, Sobol', and Morris method.

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"

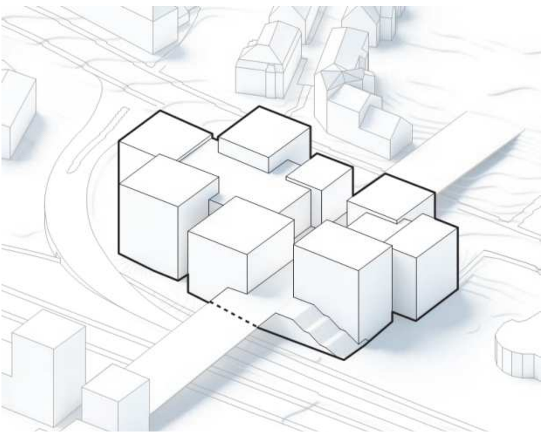


Figure 7: Early design draft of the educational institution. Illustration: EFFEKT Architects.

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.

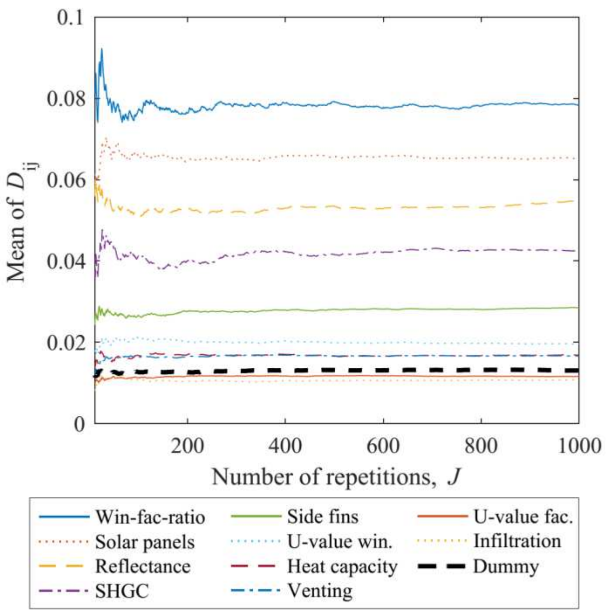


Figure 8: Convergence of the mean values of D_{ij}.

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).

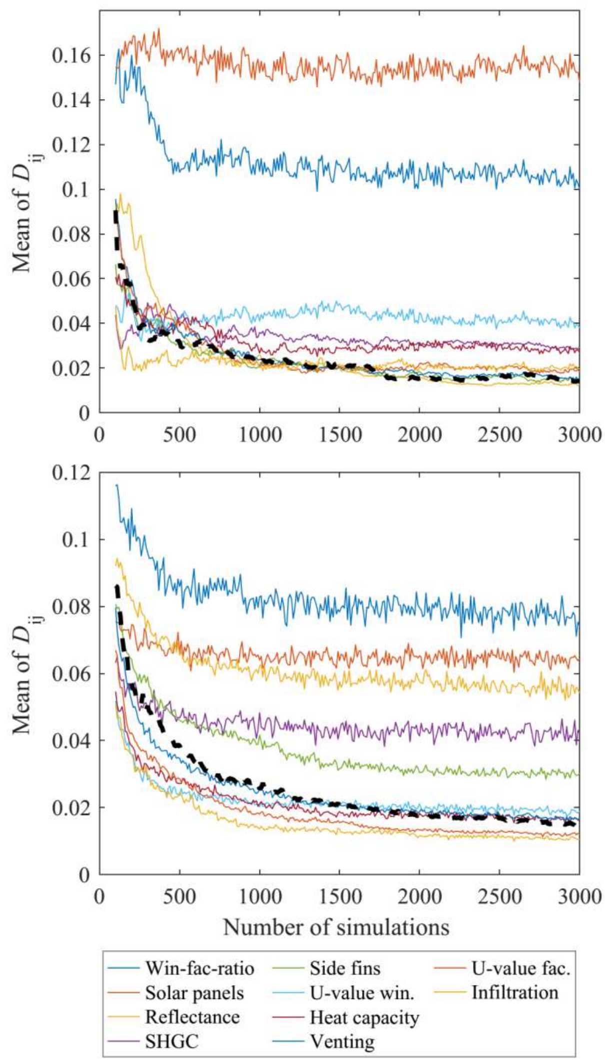


Figure 9: Mean of D_{ij} in steps of 10 simulations with respect one output, Energy demand (top), and all three outputs (bottom).

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

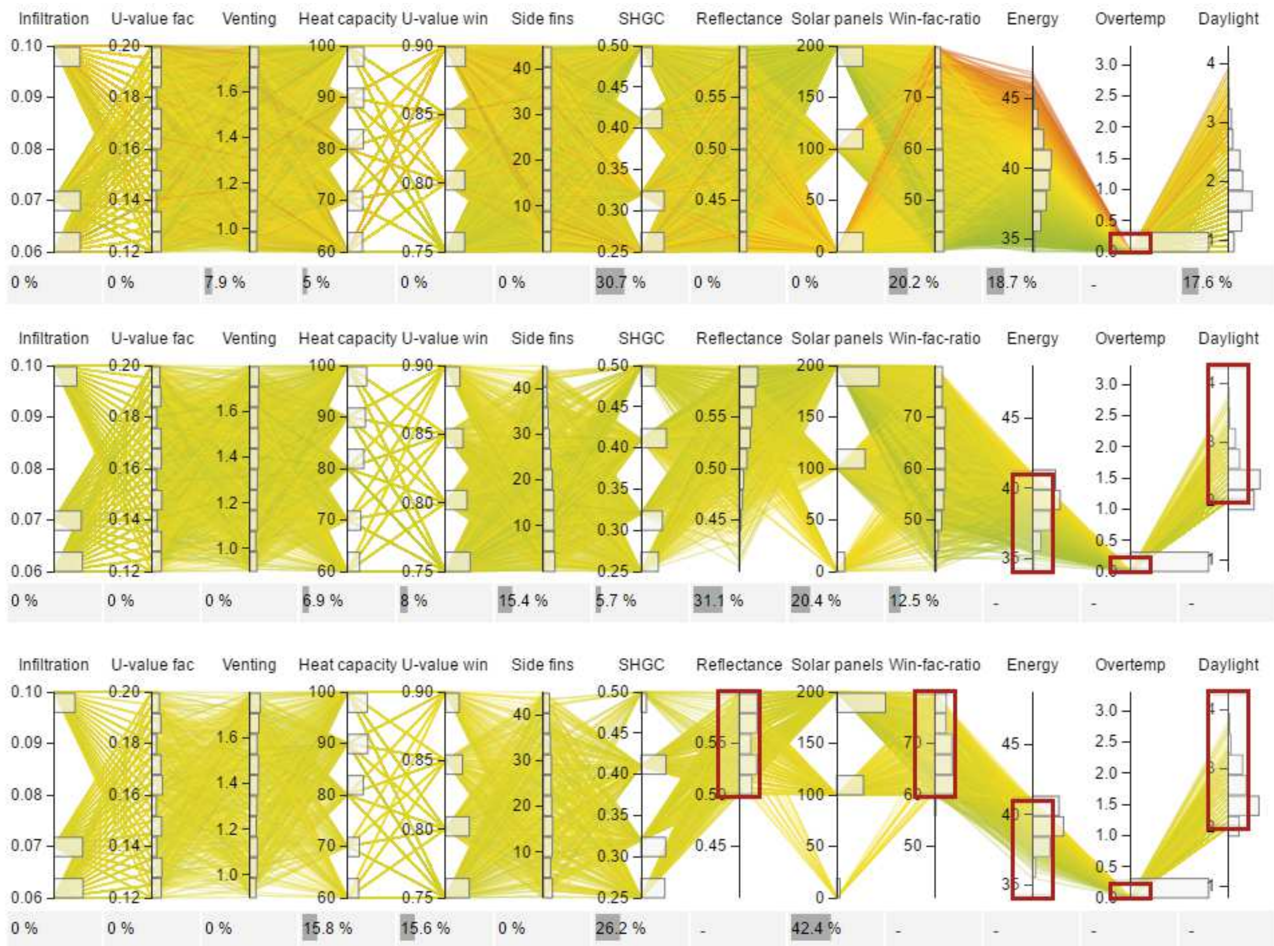


Figure 10: PCP's with user-defined filters illustrated with red rectangles. Based on TOR, bar plots indicate which parameters have been affected the most by the filtering. The inputs have been ranked right-to-left using TOM.

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).*

Appendix B. Paper F

Thermal comfort in residential buildings by the millions
- early design support from stochastic simulations

Torben Østergård, Steffen E. Maagaard, and Rasmus L. Jensen

CLIMA 2016 - proceedings of the 12th REHVA World Congress: volume 6, May 2016

Thermal Comfort in Residential Buildings by the Millions - Early Design Support from Stochastic Simulations

Torben Østergård[#*1], Steffen E. Maagaard[*2], Rasmus L. Jensen[#3]

[#] *Department of Civil Engineering, Aalborg University, Denmark*

[1]to@civil.aau.dk

[3]rlj@civil.aau.dk

[*]*MOE, Consulting Engineers, Aarhus, Denmark*

[2]sem@moe.dk

Abstract

In Danish building code and many design briefings, criteria regarding thermal comfort are defined for "critical" rooms in residential buildings. Identifying the critical room is both difficult and time-consuming for large, multistory buildings. To reduce costs and time, such requirement often causes other less critical rooms to be designed with the same constraints as the critical one. In this paper, we propose a method to overcome the difficulty of identifying critical rooms and exploit the design potential of other rooms. First we have defined a set of typical room variations present in most residential buildings. For each room variation, we perform 100.000 simulations while varying important design inputs such as window-floor-ratio, ventilation rates, glazing properties, and shading properties. Prior to this, the Morris method was used to identify and fixate insignificant inputs. A simulation engine based on the hourly version of ISO 13790 is used to calculate the number of hours with unacceptable operative temperature. As a result, the design team can assess a large number of room variations and input configurations by filtering millions of pre-calculated simulations accessible through a web service. An interactive parallel coordinate plot helps the design team to filter the many simulations. Such analysis reveals favorable input spans and assists the design team to quickly assess various design choices.

Keywords – probabilistic simulations; sensitivity analysis; interactive visualization.

1. Introduction

The design and construction of low-energy buildings have received much attention in recent years. Though, emphasis on reducing energy demand has sometimes come at the expense of thermal comfort in residential buildings [1][2]. In temperate climates, passive strategies include large south-facing windows to increase solar gain and small north-facing windows to reduce heat loss. The former may lead to overheating during summer if shading and venting are insufficient. Prolonged periods with overheating arise when thermally heavy constructions, designed to keep the building cooled, cannot release the absorbed heat during nighttime due to a lack of ventilation. To address this issue of overheating, Danish building code requires documentation of thermal comfort in dwellings from July 2016 [3]. Since 2010, this requirement has been mandatory only for the voluntary low-energy class 2015. An idealized model based on ISO 13790 [4] was developed for code compliance [5]. Despite the simplicity of the model and the need to evaluate only the “critical room” this requirement becomes time-consuming and challenging in the design of multi-story buildings. We elaborate on this in the following.

Building design is an iterative, multi-collaborator process in which the design team seeks to optimize on many, conflicting objectives. When designing multi-story residential buildings, architects and building owners often want to maximize view and daylight under the constraints of thermal comfort, energy demand, and building costs. When considering thermal comfort, the notion of a “critical room” implies that other rooms are less exposed with a potential for larger windows. To demonstrate the extensive work load related to an iterative, optimizing design approach, we will look at number of possible critical rooms in two different buildings.

First, we consider a simple building with a high degree of repetition, straight lines, and a plain geometry. Fig. 1 shows a section of residential building with five floors. For this case, at least seven rooms may become “critical” due to different ventilation rates, floor areas, and window variations (note that loads and schedules are fixed due to regulations). If actions are needed to meet the requirements for the upper floors, the corresponding rooms on the lower floors with shadowing balconies must also be addressed. Additional degrees of freedom, such as variable *g*-values, ventilation rates, and window sizes, will complicate the design process even further. To highlight the challenge of optimizing on room level, we show a prestigious and complicated building project on Fig. 2. For this building the number of room variations exceeds 100. This is due to its skewed angles, terraced roofs, and irregular balconies, while *g*-values are allowed to vary on the individual facades. Finally, the total number, of rooms to evaluate, increases rapidly if we take into account the changing design proposals.

Fig. 1 Section of a multi-story residential building with rectangles indicating the possible “critical” rooms caused by different room geometries, ventilation rates, windows, and overhang. The room enclosed by a blue rectangle is used for the case study below (Illustration: AART architects).

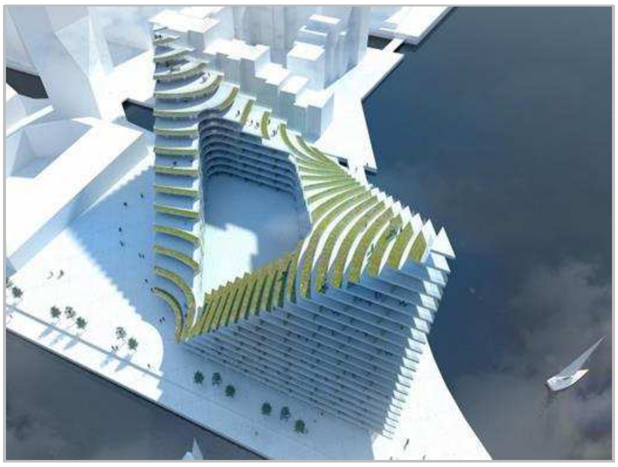


Fig. 2 Complex multi-story residential building with a trapezoidal floor plan and terraced roofs (Illustration: B.I.G architects).

In this paper, we propose a novel method in which millions of "pre-calculated" simulations provide guidance to decision-makers. Sensitivity analysis has been applied to reduce the design problem. In addition, interactive visualization facilitates "real-time" analysis with multiple stakeholders present. Hypothetically, the method helps to: a) reduce the number of design iterations, b) reduce modeling time, c) optimize on room level, and d) rapidly identify favorable input spans or actions needed to reach compliance.

The scenario described above involves various challenges related to building design: time-consuming modeling, optimization of multi-variate problems, and iterative, multi-actor decision-making. Covering these diverse topics is outside the scope of this paper. Though, this work is part of a PhD project evolving the challenges of decision-making in early building design and more information is available on *buildingdesign.moe.dk*. Here, we mention references that influenced this paper.

Attia et al. (2012) used the phrase "pre-design informative" when evaluating building performance simulation tools [6]. The logic behind is to be pro-active and to guide building design rather than evaluate individual design proposals. Stochastic modelling enables the design team to explore a large, global design space prior to decision-making. In this regard, sensitivity analysis helps identify inputs that matter the most. The application of sensitivity analysis in relation to building simulation is covered in depth by Tian [7] while the fundamental techniques are described in e.g. "Global Sensitivity Analysis: The Primer" [8]. Since stochastic modeling often involves thousands of simulations, the results cannot be analyzed and visualized in the same way as the common comparison of a few deterministic simulations. One efficient approach is to adopt the interactive parallel coordinate plot which help narrow down the results and test different designs [9][10]. In the following, we propose to combine it all and put it to use.

2. Method

In this section, we describe the logic behind the proposed methodology of using millions of pre-calculated simulations to guide the early building design. First, we define a limited number of "typical" rooms that presumably will cover the vast majority of rooms in in multi-story residential buildings. The scope of the design problem is reduced further by applying sensitivity analysis to reduce the number design variables. Next, we suggest a sampling strategy to be used for the selected rooms and reduced set of variables. Finally, we present a way to analyze and visualize the millions of simulations.

Defining Typical Rooms

For this work, we apply an idealized simulation model with few inputs which make it possible to define typical rooms and calculate (almost) all possible configurations for this model. The hourly-based model was developed by the Danish Building Research Institute to assess thermal comfort of the critical room in residential buildings [5]. The model, based on ISO 13790, evaluates the number of hours during the year in which the operative temperature exceeds 26 °C and 27 °C, respectively. Since the model is used for code compliance, some important inputs are kept fixed and cannot be changed. I.e. the combined internal loads from occupants and equipment are fixed at 5 W/m² gross floor area and the room is assumed in use all year. The room layout is defined only by floor area while solar gains depend on windows' sizes and orientation but not position and shape. Shading objects are described by the variables "horizon", "overhang", and "side fins" which are measured in degrees from windows' center points. Ventilation rates are defined from guidelines in Danish building code which only consider opening areas, opening type, and whether the room has single-sided ventilation or cross ventilation. Thus, the idealized model does not take into consideration wind pressure coefficients and thermal driving forces.

When using the idealized model for multi-story residential building, we postulate that four room types will cover the vast majority possibly critical rooms:

- Windows in one facade
- Windows in opposing facades
- Windows in two facades in a building corner
- Windows in one facade with shading side fin(s)

In the following, we consider the simplest case with windows in only one facade.

Applying Sensitivity Analysis to Reduce the Design Problem

We wish to perform an exhaustive investigation of a global design space. First try to consider a model with 20 inputs – each discretized into 5 possible values. Evaluating all combinations would then require $5^{20} \sim 10^{14}$ simulations! Fortunately, for many models, each output is mainly driven by a few inputs, e.g. the 5 to 10 most sensitive inputs [11]. We will therefore fixate variables that have negligible influence on thermal comfort (overheating only). To identify the most important inputs, we perform sensitivity analysis using the extended Morris Method [12][13]. Thus, we sample distributions of the so-called elementary effects, EE's, from a global input space in

Table 1. Distributions for 15 inputs used in the evaluation of thermal comfort.

	Parameter	Unit	Discrete values					Min.	Max.
1	Orientation	-	W	SW	S	SE	E		
2	*UA*, envelope	W/K (m^2)						0.1	0.3
3	*U*, windows	W/m^2 K						0.7	1.1
4	Recess	%						0	15
5	Ventilation, day	l/s m^2						0.9	5
6	Ventilation, night	l/s m^2						0	3
7	Ventilation, winter	l/s m^2						0	3
8	Glass-floor-ratio	%						10	40
9	*g*-value	-	0.23	0.3	0.35	0.42	0.5		
10	Heat capacity	Wh/K m^2	60	80	100	120	140		
11	F_c (shading)	-	0.2	0.4	0.6	0.8	1		
12	Overhang	°	0	20	40	60			
13	Horizon	°	10	25	40				
14	Fins, left	°						0	30
15	Fins, right	°						0	30

which each input is discretized into *p* levels. The distributions are created by following a number of trajectories, *r*, where only one factor is changed at-a-time. Finally, we obtain two sensitivity measures for each input. The mean of the absolute values of EE_i's ($\mu*$) estimates the i^{th} input's total influence on the output. The standard deviation (σ) of the EE_i's is a measure of the interaction with other inputs or non-linear effects. To perform the sensitivity analysis, we first need to assign probability distributions for all inputs.

The probability distributions shown on table 1 reflect the possible variations of inputs that may produce a critical room. With aid from practitioners, limits have been defined for typical low-energy buildings. For example, heat loss from building envelope, "*UA*, envelope", primarily depends on geometry and insulation. By comparing five low-energy, multi-story buildings, we estimated the variation of heat loss to vary between 0.1 to 0.3 W/K per square meter floor area.

Since the Morris analyses involve discretization into *p* levels, continuous uniform functions are preferable for the sensitivity analysis. In contrast, discrete uniform functions are used to create the final sets of "pre-calculated" simulations. For some variables, discrete values represent actual options better and they help the design team to the narrow the solutions. Repeated runs of the Morris analyses showed that a large number of trajectories were necessary before the ranking of the parameters' importance were consistent. This is due to the irregular, aggregated output "hours above 26 °C". For example, evaluation of a cold room with no heat loads will result in zero hours, which also is a possible result for a moderately warm room. To confirm this non-linear behavior, we applied multi-linear regression to 1.000 simulations using quasi-random

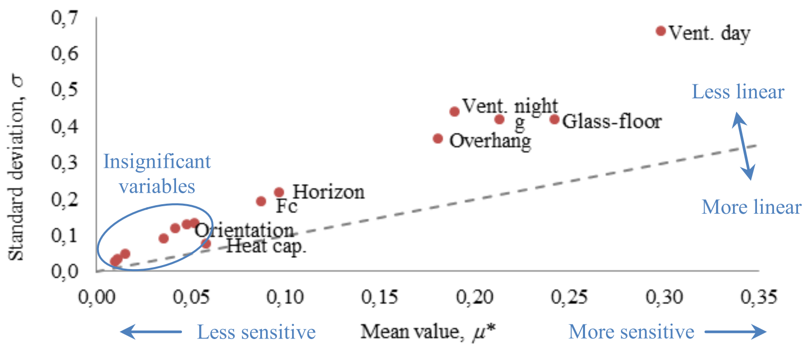


Fig. 3 Estimated means and standard deviations of the distributions of EE's in relation to the simulation output "h > 26 °C". Number of levels, p, is 8 and number of trajectories, r, is 500. The dashed line corresponds to standard-error-of-mean.

sampling (Sobol's LP_τ). The resulting standardized regression coefficients had a very low coefficient of determination, $R^2 = 0.28$, which emphasizes the need for a sensitivity technique that can handle non-linearity and interaction effects. Fig. 3 shows a plot of the sensitivity measures, μ^* and σ, for $r = 500$ and $p = 8$. The most influential variables are the ventilation rates, glass-to-floor-ratio, *g*-value, and overhang. The encircled variables close to the origin have negligible influence. This includes the averaged heat loss, "*UA*, envelope", which means that the proposed method should be valid for any geometry of well-insulated, multi-story buildings. All of the insignificant variables will be kept fixed during the upcoming stochastic simulations.

Sampling and Visualizing

The idea of simulating all design combinations requires discrete inputs and factorial sampling. Table 2 shows how the number of simulations increases when applying factorial sampling for an increasing number of discrete variables. The number of choices for each input can be interpreted as the "design resolution" for that input.

Alternative to factorial sampling, the modeler may apply low-discrepancy sequences such as Sobol's LP_τ sequences [14]. Such sampling allows for continuous distributions and produces the same factorial simulations if all inputs are discrete. The benefit of continuous distributions is that they are easier to visualize and interpret. For example, this applies to the parallel coordinate plot in which limits and line density becomes more apparent (see Fig. 4).

Table 2. Accumulated number of factorial samples when successively adding inputs.

Rank	Parameter	Unit	Steps	Factorial sim.
1	Ventilation, day	*l/s m²*	10	10
2	Glass-floor-ratio	*%*	10	100
3	*g*-value	-	5	500
4	Ventilation, night	*l/s m²*	10	5.000
5	Overhang	°	4	20.000
6	Horizon	°	3	60.000
7	F_c (shading)	-	5	300.000
8	Heat capacity	*Wh/K m²*	5	1.500.000
9	Orientation	-	5	7.500.000

To visualize and explore the many simulations, we implement the interactive parallel coordinate plot shown on Fig. 4. Each line represents the input and output values of a single simulation. The design team may interactively apply filters to the output coordinates to remove simulations not meeting the requirements. Afterwards, the team may explore different designs by adding more filters to the varying inputs. A computational limit of the applied interactive plot is roughly 100.000. In order to manage millions of simulations, we separate the simulations such that “heat capacity” and “orientation” are chosen before rendering the plot. The logic behind this is that “heat capacity” is usually fixed for a given project whereas the “orientation” is fixed for a given room. By removing these, the total number factorial simulations in table 2 would be “only” 300.000 which are closer to the computational limit of the plot. The ordering of the remaining coordinates is made from a combination of the sensitivity indices and an intuitive work flow. To appreciate the strength of the interactive plot, the reader is encouraged to get a “hands-on experience” on *buildingdesign.moe.dk*.

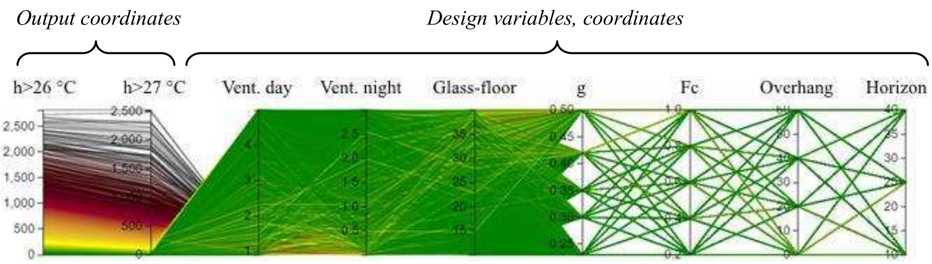


Fig. 4 Interactive parallel coordinate plot showing inputs and outputs for 100.000 simulations. Coloring is used to highlight simulations that meet the requirements (green), exceed the limits slightly (yellow), and exceed the limits extremely (red/brown).

3. Case study

To illustrate the use of the proposed method, we consider again the multi-story residential building shown on Fig. 1. The building has a rectangular shape with a floor area of 4.776 m² divided into five stories. The potentially "critical rooms" are located at the south-faced, elongated facade that has almost no shading obstructions. All of rooms may be represented by the four proposed room types within the limits described on table 2. The following examples are based on an 11 m² bedroom with openings in one wall (indicated by a blue rectangle on Fig. 1). The heat capacity is estimated to 100 Wh/K m² based on the combination of wooden floors, concrete ceiling, and exposed concrete in the facade. First, we demonstrate a "forward" approach suitable for the early design stage in which the design team seeks limits or wants to test different design paths. Secondly, we illustrate a "backward" approach showing how to find possible solutions for a design not meeting the requirements.

In the "forward" approach, we assume that the designer prioritize daylight and seeks a glass-floor-ratio of at least 25 % (corresponding to 2.75 m² glass and ~3.1 m² windows). The engineer estimates a maximum ventilation rate of 3 l/s m² and half of that during nights due to the risk of draught, noise, and burglary. As shown on Fig. 5 (top), we can now filter the simulations based on these constraints; the thermal requirements, and the fixed values (orientation, heat capacity, horizon, and overhang). Despite, these limitations more than 200 simulations still remain. These include full variability of either *g*-value or shading factor, F_c. A next step might be to set a lower limit of the *g*-value to 0.42 as a way to reduce heating demand. Consequently, the remaining simulations indicate an upper limit of 0.8 for the shading factor, F_c. In Fig. 5, histograms show that most remaining simulations are located near the lower limits of *g*-value and F_c.

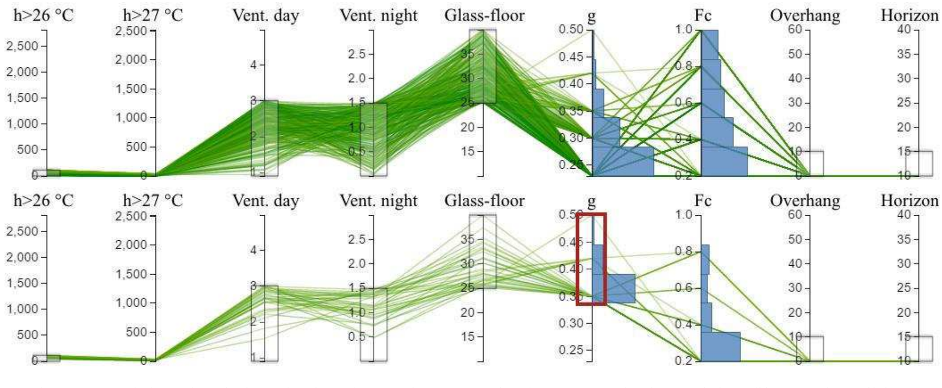


Fig. 5 Remaining simulations in the "forward" approach after successively applying filters using sliders (black rectangles). The red rectangle highlights a changed filter. Histograms show distributions of possible solutions for the non-fixed variables.

In the "backward" example, we assume that a design proposal is already given. First, an openable French door and fixed window correspond to a glass-to-floor ratio of

25.4 %. To reduce energy demand, the mechanical ventilation is turned off during summer. From these assumptions, the maximum value for the ventilation level is estimated to 1.1 l/s m². There is no shading and the *g*-value is preferable 0.5 in order to reduce heating demand and to match other rooms. As shown on Fig. 6 (top), this setup results in exceedance of the comfort criteria with 2.200 – 2.600 hours above 26 °C. To remedy this, we interactively adjust the filters to find limits for ventilation and solar shading that meet the thermal criteria. Fig. 6 (bottom) shows one feasible scenario in which the *g*-value is reduced and higher ventilation rates are achieved by making the window openable and by turning on mechanical ventilation.

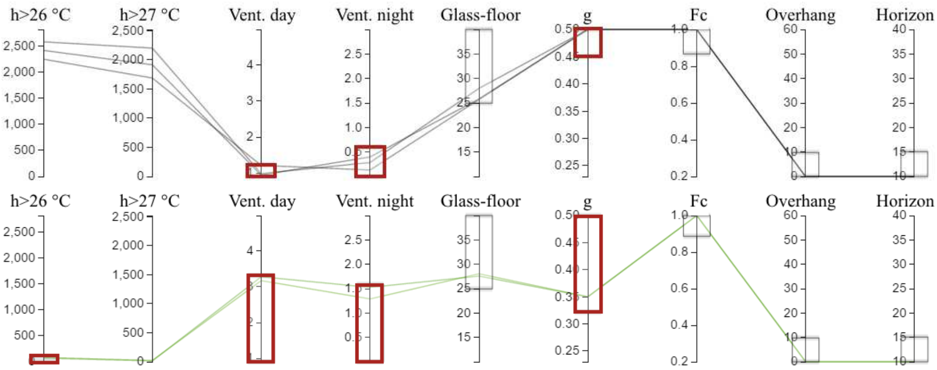


Fig. 6 Top: A few simulations corresponding to a design proposal not meeting the requirements. Bottom: Possible solutions when expanding the initial limits and constraining the output.

4. Discussion

The proposed method is not meant to document thermal comfort for the final critical room. Instead, it facilitates quick identification of approximate limits and helps test various design paths for multiple rooms. Feedback from practitioners will tell if the number of simulations and room variations are sufficient.

An exhaustive representation of the design space requires fewer simulations when using an idealized model with few inputs. Though, the method may be improved by using a detailed model that addresses energy, illuminance, glare, etc. The increased level of detail will require more inputs and thus more simulations are needed. In that case, sensitivity analysis can again be applied to reduce the number of variable inputs and to investigate the influence of varying room layouts, window positions, loads, and more.

5.Conclusion

A novel method was proposed to support building design in relation to thermal comfort in residential buildings. Initially, we performed sensitivity analysis, using the method of Morris, to identify the most influential inputs in an idealized model used for code compliance. This analysis helped reduce the number of design variables from roughly 15 to 7. While varying this reduced set of inputs, 100.000 simulations were performed for each combination of four typical room types, five orientations, and five discrete values of building heat capacity. An interactive parallel coordinate plot enabled rapid exploration of the vast amount of simulations. The large dataset and the interactive plot will make it possible to test many different designs during meetings between building owner, architects, and engineers. Presumably, this will reduce the number of time-consuming and costly design iterations. Moreover, the large dataset helps to quickly identify critical rooms and enable optimization of non-critical rooms.

A multi-story residential building was used to show the challenge of evaluating thermal comfort in the critical room and to demonstrate implementation of the proposed method. The method may be further improved by using sophisticated simulation software to quantitatively assess energy, daylight, and other performance objectives.

Acknowledgements

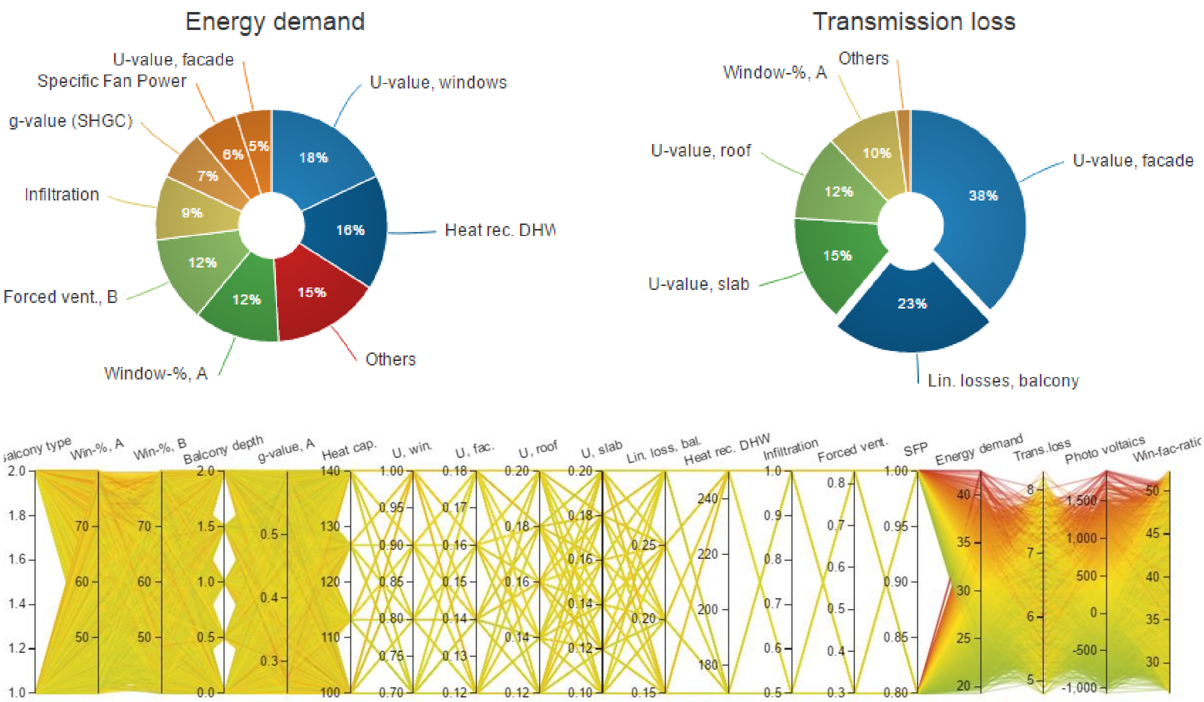
Innovation Fund Denmark and MOE A/S provided funding. The work was part of an industrial doctorate program with Aalborg University and consultancy company MOE A/S.

References

[1] T. S. Larsen, Vurdering af indeklimaet i hidtidigt lavenergibyggeri - med henblik på forbedringer i fremtidens lavenergibyggeri, DCE Contract Report no. 100, (2011).
[2] H. N. Knudsen, K. E. Thomsen, and O. Mørck, Occupant Experiences and Satisfaction with New Low-Energy Houses, In: 11th International Conference CLIMA 2013, Prague, Czech Republic, 16-19 June 2013.
[3] Energistyrelsen, Danish Building Regulations 2015, In: http://bygningsreglementet.dk/, 2016.
[4] CEN, ISO 13790:2008 Energy performance of buildings -- Calculation of energy use for space heating and cooling, Geneva, Switzerland, 2008.
[5] L. H. Mortensen and S. Aggerholm, Simplified hourly method to calculate summer temperatures in dwellings, In: 33rd AIVC and 2nd TightVent Conference, Copenhagen, 10-11 October, 2012.
[6] S. Attia, E. Gratia, A. De Herde, and J. L. M. Hensen, Simulation-based decision support tool for early stages of zero-energy building design, Energy and Buildings, vol. 49, (2012) 2–15.
[7] W. Tian, A review of sensitivity analysis methods in building energy analysis, Renewable and Sustainable Energy Reviews, vol. 20, (2013), 411–419.
[8] A. Saltelli, et al., Global sensitivity analysis: The Primer, Wiley & Sons, 2008.
[9] F. Ritter, Simulation-based Decision-making in Early Design Stages, In: 32nd CIB W78 Conference, Eindhoven, The Netherlands, 27-29 October, 2015.
[10] D. L. Macumber, B. L. Ball, and N. L. Long, A graphical tool for cloud-based building energy simulation, In: 2014 ASHRAE/IBPSA-USA Building Simulation Conf., Atlanta, USA, 10-12 Sept. 2014.
[11] B. Eisenhower, Z. O. Neill, V. A. Fonoberov, and I. Mezi, Uncertainty and sensitivity decomposition of building energy models, Journal of Building Performance Simulation, 5:3, (2012), 171–184.
[12] M. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics, 33:2, (1991) 161–174.
[13] F. Campolongo, J. Cariboni, and A. Saltelli, An effective screening design for sensitivity analysis of large models, Environmental Modelling & Software, 22:10, (2007), 1509–1518.
[14] I. M. Sobol' and B. V. Shukman, Random and quasirandom sequences: Numerical estimates of uniformity of distribution, Mathematical and Computer Modelling, 18:8, (1993), 39–45.

Appendix C. Feedback from case study

Figure C-1: Illustrations from early and detailed stages of the second case study in Paper C (BIG architects).



Balcony ty	Win-%, A	Win-%, B	Balcony de	g-value, A	Heat cap.	U, win.	U, fac.	U, roof	U, slab	Lin. loss, b	Heat rec. D	Infiltration	Forced ver	SFP	Energy der	Trans.loss	Photo volta	Win-fac-rat
2	60	60	1	0.425	126.6667	0.9	0.16	0.16	0.16	0.25	250	1	0.83	1	35.3	6.9	978.4	39.39384
2	50	70	0.5	0.5125	113.3333	1	0.14	0.18	0.18	0.2	250	0.5	0.83	0.8	31.1	6.4	440.8	35.04661
2	70	50	1.5	0.3375	140	0.8	0.18	0.14	0.12	0.3	170	1	0.3	1	29.7	7.3	261.6	43.74108
2	45	65	2	0.55625	126.6667	0.7	0.14	0.14	0.2	0.25	250	1	0.83	0.8	29.2	6.4	197.6	31.76379
2	65	45	0.5	0.38125	100	0.9	0.18	0.2	0.14	0.15	170	0.5	0.3	1	27.6	7	-7.2	40.45826
2	55	55	1.5	0.29375	140	1	0.12	0.18	0.1	0.3	170	1	0.3	0.8	29.5	6	236	36.11102
2	75	75	0	0.46875	113.3333	0.8	0.16	0.12	0.16	0.2	250	0.5	0.83	1	35.6	6.6	1016.8	49.24231
2	42.5	77.5	1.5	0.359375	100	0.7	0.14	0.16	0.18	0.25	170	0.5	0.3	1	23.6	6.4	-519.2	31.78618
2	62.5	57.5	0	0.534375	126.6667	0.9	0.18	0.12	0.12	0.15	250	1	0.83	0.8	31	6.4	428	40.48065
2	52.5	47.5	2	0.446875	113.3333	1	0.12	0.14	0.1	0.3	250	0.5	0.83	1	33.1	5.8	696.8	33.91501

Figure C-2: Interactive plots used for the multi-collaborator design meeting. The labels have been translated to English and the color and size of the parallel coordinate plot (bottom) has been adjusted. The pie charts (above) show relative parameter importance based to sensitivity analysis.

Table C-1: Summary of feedback from five design team members. Four did not respond and two MOE engineers were "disqualified" since they contributed to the development of the method and the plots.

1) Profession

	Building owner	Architect	Engineer	Contractor
What is your profession?	1	1	2	1

2) Introduction

Which of the following do you need more information about	# of respondents
• Explanation of the calculation method (exploration of global design space, variable inputs, SA)	3
• Motivation - Why should we use this method	1
• Prerequisites - Which assumptions and sofware (Be10, Bsim, etc.) have been used	2
• Variable inputs - What is the reasoning behind the chosen variable and distributions?	2
• Sensitivity analysis - Explanation of inputs that matter most	2
• Explanation of the applied plots - What do they show and how can they used?	1

3) Need for simulation based support for decision-making in early design

Building simulations are often used to asses energy demand, daylight, indoor climate, environmental impact, building costs, etc. To what extent do you agree with following in a Danish context?	Strongly disagree	Partly disagree	Indifferent	Partly agree	Strongly agree	Don't know
• Building simulations er is valuable tool for code compliance				1	4	
• Building simulations typically provide valuable decision support in early design stages					5	
• Building simulations are required but have no reel value	3	1		1		

4) Most important property of building simulation tool

What is most important to you? (mostly for architects and engineers)	1 (least important)	2	3	4	5 (Most important)
• Intelligence - the ability to aid decision making				1	4
• User friendly	1	1	2	1	
• Interoperability between CAD and simulation software	1	1		2	
• Accuracy of calculations		3	2		
• Holistic - ability to simultaneously assess energy, indoor climate, costs, environmental impact	2		1	1	1

5) Feedback of the proposed method

What are you thoughts about the "method" when comparing with common practice in Denmark	Strongly disagree	Partly disagree	Indifferent	Partly agree	Strongly agree	Don't know
A. The method could be used to investigate the design options you needed?			2		3	
This is important?					5	
B. The method enabled you to make good design decisions?				3	1	1
This is important?					3	
C. The method support collaboration between building owner, architect, and engineer?				1	3	
This is important?					4	1
D. The method provide more information about consequences instead of just evaluating designs?				1	2	2
This is important?				2	2	1
E. The sensitivity analysis highlighted the most important parameters?					5	
This is important?					5	

6) What would you prefer to obtain more information about?

Which performance criteria should we try to include in our following worh - what is most important to you?	1 Unimportant	2	3	4	5 Super important
Energy demand					5
Building costs (life cycle cost)				4	
Daylight			3		2
Thermal comfort			1	1	3
Indoor air quality			3	1	
Environmental impact (life cycle assessment)		2	1	1	
Acoustics (e.g. reverberation time)		3	1	1	
Other		1			1

7) Conclusion

All in all (including future expansions), what do you think about the potential of the proposed method and form of collaboration?	Irrelevant	Small potential	Some potential	Great potential	Don't know
Improved collaboration and communication between building owner, architect, and engineer?				3	1
The potential for providing wise decisions and exploring a vast design space?				3	1

SUMMARY

Simulations are commonly used to assess building performance with respect to energy demand and indoor environment. However, the use of performance simulations is limited during the early stages characterized by large uncertainties. This industrial Ph.D. study presents a novel simulation approach that relies on thousands of simulations representing the multidimensional design space. Interactive visualizations enable decision-makers to explore, in real-time, the vast design space and identify favorable solutions which satisfy the needs of different stakeholders. Sensitivity analysis helps reveal important design parameters that require the most attention when seeking to improve building performance. Fast metamodels facilitate immediate feedback on design changes and reduce time-consumption related to performance assessment. Ultimately, the work described in this thesis and on buildingdesign.moe.dk facilitates proactive guidance and supports collaboration between building owners, architects, engineers, and contractors. This helps the design team to create buildings with high performance and minimum costs

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-091-3

AALBORG **UNIVERSITY PRESS**