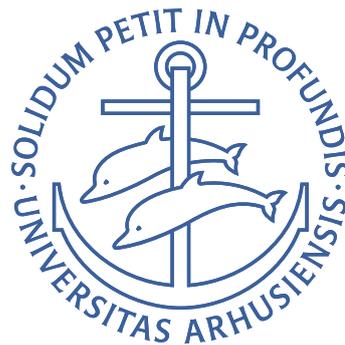


Martin Heine Kristensen

Urban building energy modelling for retrofit analysis under uncertainty

PhD Thesis



Graduate School of Science and Technology
Aarhus University

November 2018



ISBN: 978-87-7507-441-9
DOI: 10.7146/aui.301.208

**Urban building energy modelling for
retrofit analysis under uncertainty**

Urban building energy modelling for retrofit analysis under uncertainty

PhD thesis

This thesis is submitted in partial fulfilment of the requirements for the degree of PhD at the Graduate School of Science and Technology at Aarhus University. It was successfully defended on 31 January 2018 at the Department of Engineering, Aarhus University.

Copyright © 2018 by Martin Heine Kristensen.

All right reserved. The content of this thesis is freely available, but publication (with reference) may only be pursued due to agreement with the author. Except where otherwise indicated, the content is produced by the author.

Supervisor:

Steffen Petersen, Associate Professor, Aarhus University, Denmark

Assessment Committee:

Yeonsook Heo, Associate Professor, Korea University, Korea

Igor Sartori, Senior Research Scientist, SINTEF, Norway

Søren Peder Madsen (Chair), Associate Professor, Aarhus University, Denmark

Aarhus University

Graduate School of Science and Technology

Nordre Ringgade 1

DK-8000 Aarhus C

<http://phd.au.dk/graduate-schools/scienceandtechnology/>

Abstract

Urban building energy modelling (UBEM) is a growing research field that seeks to expand conventional building energy modelling to the realm of neighbourhoods, cities or even entire building stocks. The aim is to establish frameworks for analysing combined urban effects rather than those of individual buildings, which city governments, utilities and other energy policy stakeholders can use to assess the *current* environmental impact of our buildings, and, maybe more importantly, the *future* effects that energy renovation programmes and energy supply infrastructure changes might have. However, the task of creating reliable models of new or existing urban areas is difficult, as it requires an enormous amount of detailed input data – data which is rarely available. A solution to this problem is the introduction of archetype modelling, which is used to break down the building stock into a manageable subset of semantic building archetypes, for which, it is possible to characterize their parameters. It is the focus of this thesis to explore and develop new methods for stochastic archetype characterization that can enable archetype-based UBEM to be used for accurate urban-scale time series analysis.

The thesis is divided into three parts. The first part acts as an introduction to case study data of the residential building stock of detached single-family houses (SFHs) in Aarhus, Denmark, which is used throughout the thesis for demonstration purposes.

The second part concerns the development of methods for archetype modelling. Bayesian methods for archetype parameter calibration are presented that incorporates the variability of the underlying cluster of buildings, and correlation between parameters, to enable informed predictions of unseen buildings from the archetype under uncertainty. The capabilities of archetype-based UBEM are further widened through the introduction of dynamic building energy modelling that allows for time series analysis.

The third part of the thesis is devoted to demonstrating the usefulness of the proposed archetype formulation as a building block for urban-scale applications. An exhaustive test scheme is employed to validate the predictive performance of the framework before establishing a city-scale UBEM of approx. 23,000 SFHs in Aarhus. It is used to forecast citywide heating energy use from 2017 up until 2050 under uncertainty of energy renovations and climate change.

Overall, the proposed archetype-based UBEM framework promises very useful for fast, flexible and reliable urban-scale time series analysis, including forecasting the effects of energy renovation or city densification, to establish an informed basis for energy policy decision-making.

Keywords: urban building energy modelling; UBEM; single-family houses; building archetypes; hierarchical modelling; archetype heterogeneity; propagation of uncertainty; Bayesian calibration; time series data; district heating; energy renovation; climate change

Resumé (Abstract in Danish)

Bygningers energiforbrug modelleres traditionelt for at dokumentere deres tekniske formåen til opnåelse af byggetilladelse, eller for at sandsynliggøre gevinsterne ved en kommende energirenovering. I de senere år er der imidlertid opstået et behov for at udvide dette felt til, også at omfatte modellering af grupper af bygninger, byområder, eller endda den samlede bygningsmasse. Hermed er målet; at etablere simuleringsbaserede platforme til analyse af bygningsmassens resulterende interaktive og dynamiske effekter, i modsætning til de meget enkeltstående effekter af konventionel bygningsmodellering. Disse dynamiske platforme kan bruges af kommuner, forsyningsselskaber og andre energipolitiske interessenter til at evaluere det nuværende energiforbrug, og evt. afledte miljømæssige konsekvenser heraf, samt, måske endnu vigtigere; de fremtidige konsekvenser som energirenovering, og infrastruktur-mæssige ændringer i fx fjernvarmesystemer, måtte have. Etableringen af sådanne storskalamodeller er dog stærk udfordret af de enorme mængder af inputdata, som er nødvendig, for at beskrive de underliggende termodynamiske forhold. En løsning til dette problem er anvendelsen af arketypermodellering, hvilket bruges til at nedbryde bygningsmassen til en overskuelig mængde af arketyperiske bygningskategorier. Det er denne artikel-baserede afhandlings fokus at udvikle nye metoder for stokastisk identificering af værdierne til sådanne arketyper underliggende inputparametre, samt at bruge disse til at etablere storskalamodeller for timeopløst analyse af energiforbruget.

Afhandlingen består af tre dele. Den første del fungerer som en introduktion til case data af danske parcelhuse i Aarhus Kommune, hvilket bruges gennem hele afhandlingen.

Anden del omhandler udviklingen af metoder for arketypermodellering, herunder præsenteres nye Bayesianiske metoder for kalibrering af arketyperparametre, som tager højde for forskelligheden (heterogeniteten) i den underliggende bygningsmasse, samt korrelationen mellem de forskellige parametres værdier. Endvidere udbygges de prædikative egenskaber af arketyperbaseret modellering med indførelsen af en dynamisk beregningskerne, som tillader timeopløst analyse og fremskrivning af energiforbruget.

Tredje del er tilegnet demonstreringen af de foreslåede arketypermetoder som byggesten i konstruktionen af byskalamodeller. Den foreslåede metode for modellering af arketyper valideres grundigt, før en samlet model etableres for næsten 23.000 parcelhuse i Aarhus. Denne bruges til at fremskrive fjernvarmeforbruget fra 2017 til 2050 under forskellige scenarier af energirenovering, samt under usikkerhed af klimaforandringer.

Afhandlingen præsenterer samlet set en arketyperbaseret bygningsmodel for hurtig, fleksibel og pålidelig analyse af byområders timeopløste energiforbrug, herunder evnen til at fremskrive effekterne af fx energirenovering og byfortætning, til etableringen af informerede beslutningsgrundlag for energipolitiske aftaler og konsekvensanalyser.

Nøgleord: by-skala; arketyper; parcelhuse; bygningsmasse; hierarkisk modellering; arketyper heterogenitet; sandsynlighed; Bayesianisk kalibrering; tidsserie; fjernvarme; klimaforandringer; renovering

Preface

This thesis comprises the academic output of a 3-year PhD Fellowship at the Graduate School of Science and Technology (GSST), Aarhus University, where I was enrolled in the Engineering programme between August 2015 and November 2018, under the supervision of Associate Professor Steffen Petersen. The study was conducted as part of the READY project (<http://www.smartcity-ready.eu/>), financed by the 7th EU Framework Programme (FP7) for research and innovation funding (FP7-Energy project ID: 609127).

The thesis is submitted for the Danish degree of PhD in partial fulfilment and conformity with the requirements of the GSST Rules and Regulations and the Ministerial Order on the PhD Degree Programme. It is constructed around a collection of papers with additional descriptions of case data and some of my newest research that has not yet been published. A [List of publications](#) is to be found following this preface. Except where explicit reference is made to the work of others, this thesis is the result of my own work. It has not, nor has any similar thesis, been submitted for a degree or any qualification at this or any other university.

I would like to thank my supervisor, Associate Professor Steffen Petersen, for considering me for the PhD project, and for his guidance and always optimistic mindset throughout the project. Thanks also go to the staff at AffaldVarme Aarhus, in particular IT manager Adam Brun, for supplying most of the district heating data that make up an indispensable part of this thesis. I also owe many thanks to Reader in Architectural Engineering, Ruchi Choudhary, and her research group at the Department of Engineering, University of Cambridge, for hosting me during visits at their department. They saved me from many sleepless nights over the troubles of Bayesian statistics. A special thanks goes to my great colleagues and fellow PhD students, especially Rasmus Elbæk Hedegaard, with whom I have enjoyed many coffee-fuelled discussions on Bayesian statistics and building energy modelling in general. Finally, I would like express my heartfelt gratitude to my girlfriend, Tove, who has been very supportive at all times, and had to do without me during my stays abroad and long working hours in the final months of the PhD. Thanks to you all.

Martin Heine Kristensen
Aarhus University, 13 November 2018

List of publications

Throughout the PhD project, I have contributed to a total of five journal papers and five conference papers of relevance to the PhD project that has been published, or is being considered for publication, in high-ranked international scientific journals or conference proceedings. These papers are divided into publications of primary and secondary interest to the research objectives of the PhD thesis. They are listed below in order of appearance in the thesis.

Primary publications

The primary publications comprise work of which I am the main contributor and first author. They make up the primary scientific output of the PhD project and thus form the foundation of this thesis. They are all presented within the thesis with the layout of the published articles.

- [P1] **Predicting Danish residential heating energy use from publicly available building characteristics** by Martin Heine Kristensen, Adam Brun and Steffen Petersen. Published in *Energy and Buildings* ([Kristensen, Brun and Petersen, 2018](#)).
- [P2] **Bayesian calibration of residential building clusters using a single geometric building representation** by Martin Heine Kristensen, Ruchi Choudhary, Rasmus Høst Pedersen and Steffen Petersen. Published in *Proceedings of Building Simulation 2017: The 15th IBPSA Conference* ([Kristensen, Choudhary, Pedersen and Petersen, 2017](#)).
- [P3] **Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution** by Martin Heine Kristensen, Ruchi Choudhary and Steffen Petersen. Published in *Energy Procedia, Proceedings of the CISBAT 2017 International Conference – Future Buildings & Districts* ([Kristensen, Choudhary and Petersen, 2017](#)).
- [P4] **Hierarchical calibration of archetypes for urban building energy modeling** by Martin Heine Kristensen, Rasmus Elbæk Hedegaard and Steffen Petersen. Published in *Energy and Buildings* ([Kristensen, Hedegaard and Petersen, 2018a](#)).
- [P5] **Urban-scale dynamic building energy modeling and prediction using hierarchical archetypes: A case study of two Danish towns** by Martin Heine Kristensen, Rasmus Elbæk Hedegaard and Steffen Petersen. Published in *Proceedings of BSO 2018: The 4th Building Simulation and Optimization Conference* ([Kristensen, Hedegaard and Petersen, 2018b](#)).

Secondary publications

Research reported in the secondary publications comprises work that can be considered important stepping stones towards the primary publications, or applications of my work in other research fields. Although these publications are considered secondary results of the PhD project, they should be seen as a product of the process and exploration of the research objectives. I have contributed to these publications in terms of either conceiving the research questions, establishing the methodology, creating results and performing analysis, and/or writing the manuscripts. They are not presented in the main text directly, but rather discussed in the context of the primary publications. The full-length articles are attached to the thesis as appendices.

- [S1] **Choosing the appropriate sensitivity analysis method for building energy model-based investigations** by Martin Heine Kristensen and Steffen Petersen. Published in *Energy and Buildings* ([Kristensen and Petersen, 2016](#)).
- [S2] **Prerequisites for reliable sensitivity analysis of a high fidelity building energy model** by Steffen Petersen, Martin Heine Kristensen and Michael Dahl Knudsen. Published in *Energy and Buildings* ([Petersen et al., 2019](#)).
- [S3] **Contrasting the capabilities of three different sensitivity analysis methods for building energy model-based investigations** by Martin Heine Kristensen and Steffen Petersen. Published in *Proceedings of BSO 2018: The 4th Building Simulation and Optimization Conference* ([Kristensen and Petersen, 2018](#)).
- [S4] **Explaining variability in metered energy use for similar buildings using Bayesian inference** by Martin Heine Kristensen and Steffen Petersen. Published in *Energy Procedia, Proceedings of NSB 2017: The 11th Nordic Symposium on Building Physics* ([Kristensen and Petersen, 2017](#)).
- [S5] **Bottom-up modelling methodology for urban-scale analysis of residential space heating demand response** by Rasmus Elbæk Hedegaard, Martin Heine Kristensen, Theis Heidmann Pedersen, Adam Brun and Steffen Petersen. Submitted to *Applied Energy*, November 2018.

Abbreviations

AVA	AffaldVarme Aarhus (municipal district heating company in Aarhus)
BDR	Building and Dwelling Register
BEM	Building energy modelling (<i>or</i> building energy model)
CI	Confidence interval
CVRMSE	Coefficient of variation of the root mean squared error
DH	District heating
DHW	Domestic hot water
ECM	Energy conservation measure
EED	Energy Efficiency Directive
EPBD	Energy Performance of Buildings Directive
EPC	Energy performance certificate
EU	European Union
EUI	Energy use intensity
GHG	Greenhouse gas
GP	Gaussian process
GPR	Gaussian process regression
HPD	Highest posterior density
KOH	Kennedy & O’Hagan
LR	Linear regression
MAP	Maximum a posteriori
MAPE	Mean absolute error
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MFH	Multi-family house
ML	Maximum likelihood
MLE	Maximum likelihood estimate
NMBE	Normalized mean bias error
NN	Neural network (<i>synonymous with</i> artificial neural network, ANN)
nZEB	Nearly zero-energy building
PDF	Probability density function
PSRF	Potential scale reduction factor
RC	Resistance-capacitance
RCP	Representative concentration pathway
RF	Random forest
ROM	Reduced-order modelling (<i>or</i> reduced-order model)
SA	Sensitivity analysis
SFH	Single-family house
SVM	Support vector machines
UBEM	Urban building energy modelling (<i>or</i> urban building energy model)
UN	United Nations
UNFCCC	United Nations Framework Convention on Climate Change

Table of Contents

Abstract	v
Resumé (Abstract in Danish)	vii
Preface	ix
List of publications	xi
Abbreviations	xiii
1 Introduction	1
1.1 The role of buildings in an energy-efficient future	1
1.2 Urban building energy modelling	2
1.3 Building archetypes	5
1.4 Uncertainty modelling	8
1.5 Summary of challenges	15
1.6 Research objectives	15
1.7 Thesis outline	16
Part 1 Case data presentation and preliminary statistical analysis	17
2 Urban case study: The city of Aarhus	19
2.1 The building stock of Aarhus	19
2.2 Aarhus district heating system	21
3 Statistical analysis of annual building energy use	25
3.1 Motivation	25
3.2 Paper [P1]: "Predicting Danish residential heating energy use from publicly available building characteristics"	25
3.2 Epilogue	36

Part 2	Physics-based archetype modelling	37
4	A regression-based approach using annual data	39
4.1	Motivation	39
4.2	Paper [P2]: "Bayesian calibration of residential building clusters using a single geometric building representation"	39
4.2	Epilogue	50
5	The effect of temporal resolution on calibration	51
5.1	Motivation	51
5.2	Paper [P3]: "Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution"	51
5.2	Epilogue	58
6	A hierarchical approach using time series data	59
6.1	Motivation	59
6.2	Paper [P4]: "Hierarchical calibration of archetypes for urban building energy modeling"	59
6.2	Epilogue	76
Part 3	Urban building energy modelling for retrofit analysis under uncertainty	77
7	Archetype-based UBEM	79
7.1	Motivation	79
7.2	Paper [P5]: "Urban-scale dynamic building energy modeling and prediction using hierarchical archetypes: A case study of two Danish towns"	79
7.2	Epilogue	82
8	Citywide retrofit analysis under uncertainty	83
8.1	Motivation	83
8.2	Data cleansing and archetype segmentation	84
8.3	Training and validation of hierarchical archetypes	84
8.4	Forecasting citywide energy demand from 2017-2050	92
8.5	Discussion	97

9 Thesis conclusion	101
9.1 Main contributions	101
9.2 Limitations and future work	102
References	105
Appendix – Secondary publications	117
Paper [S1]	117
Paper [S2]	129
Paper [S3]	145
Paper [S4]	153
Paper [S5]	159

Chapter 1

Introduction

1.1 The role of buildings in an energy-efficient future

The 21st Conference of the Parties (COP 21) of the United Nations Framework Convention on Climate Change (UNFCCC) was held in Paris, France, in December 2015 where 197 State Parties to the UNFCCC came together for the first time for a common cause known as the Paris Agreement to reduce global warming ([United Nations, 2015](#)). As of July 2018, 195 Parties have ratified the Paris Agreement's long-term goal to keep the increase in global average temperature to well below 2 °C, and to pursue efforts to limit the increase to 1.5 °C only.

Cutting the energy consumption of buildings plays a key role in reaching these goals. In the European Union, as of 2016, the residential and commercial building sectors together account for 40 % of final energy consumption and about 36 % of final EU greenhouse gas (GHG) emissions ([Sajn, 2016](#)). The residential building sector is particularly important, accounting for as much as 25.7 % of the total energy consumption in the EU. Similar numbers are seen for the U.S.¹, where 39 % of the total U.S. energy consumption was consumed by the residential and commercial sectors in 2017, of which residential buildings accounted for 20.4 % ([EIA, 2018](#)).

Special focus is therefore brought towards increasing the energy efficiency of buildings, and residential buildings in particular. The EU initiated an ambitious legislative framework for increasing the energy efficiency of its Member States' buildings in 2002 with the adoption of the Energy Performance of Buildings Directive (EPBD – Directive 2002/91/EC), which introduced calculation methodologies for building energy use, energy performance certificates (EPCs) and minimum requirements for building energy use. The EPBD was later amended with the 2010 EPBD recast (Directive 2010/31/EU), which introduced the concept of nearly zero-energy buildings (nZEB) as a mandatory EU standard for all new buildings by 31 December 2020. However, focusing on new-build alone is simply not sufficient; the current rate of approx. 1% (depending on the country) at which the existing EU building stock is being energy renovated each year has to be increased ([European Commission, 2017](#)). In 2012, the EU adopted the Energy Efficiency Directive (EED – Directive 2012/27/EU), a new legally binding framework of measures for the promotion of energy efficiency at all stages of the energy chain. The EED obligates EU Member States to carry out energy efficient renovations on at least 3 % (by floor area) of its own governmental buildings every year from 2014 onwards, and with the newest 2018

¹Even though U.S. President Donald Trump withdrew the United States from the Paris Agreement in June 2017, U.S. numbers are shown for the sake of comparison.

amendments of the EPBD and the EED (Directive 2018/844/EU), EU Member States are now required to draft strategies for a cost-effective transformation of their existing buildings into fully de-carbonized national building stocks by 2050, and to set 2030 and 2040 milestones to evaluate this progress; this includes measurable progress indicators for renovation rates and the overall energy efficiency of buildings.

While goals of reducing GHG emissions and increasing the energy efficiency of the building stock are often set at the national or international level, major actions have to be taken at the city scale (Li et al., 2017). In fact, many municipalities in Europe have set goals of reducing GHG emissions that are more aggressive than those mandated by their national government or the EU. For instance, more than 7,700 cities and local authorities in 53 countries have now signed up the EU Covenant of Mayors for Climate & Energy², the world's largest movement for local climate and energy actions. It is a voluntary commitment initiated by the European Commission in 2008 to establish ambitious local energy action plans that has later expanded to now also include plans for adapting to climate change. Signatories are committed to developing action plans describing the necessary steps towards reaching its 2020 or 2030 targets; all signatories should at minimum endorse the EU 40% GHG reduction target by 2030.

Evaluating the progress and effects of energy-efficient renovations and other initiatives to reduce building-related GHG emissions is very difficult, and it is further complicated by the absence of data. Considerable work is still needed to map the energy consumption of the building stock in time and space, and to analyse the energy renovation potential of existing buildings, both on the national scale and city scale. These issues and societal challenges constitute the overall thematic background of my thesis work; to investigate and further develop methodologies for modelling building energy use on larger urban scales, which can be used to analyse and evaluate energy conservation measures (ECMs) in an effort to help meet the goal of a de-carbonized building stock. The remainder of the thesis introduction is organized as follows: A comprehensive review of current practice and state-of-the-art building stock modelling is provided in Sections 1.2-1.4, before summing up challenges and knowledge gaps in the field in Section 1.5. These identified challenges serve as the basis for establishing the research objectives of the thesis in Section 1.6.

1.2 Urban building energy modelling

Statistical modelling approaches have traditionally been used to assess the effects of energy policy using macro-level (aggregated) data and various regression analysis techniques. An example herof is the dynamic housing stock model by Sandberg et al. (2014a,b) that uses statistical data on population size, number of persons per dwelling, the demand for dwellings, etc., to estimate the demolition and renovation activity of the Norwegian building stock towards year 2050. The model has later been expanded with the ability to estimate the aggregated stock energy use and energy savings potential per m² (Sandberg, Sartori, Vestrum and Brattebø, 2016; Sandberg et al., 2017), and has also been applied to the residential building stock of 11 European countries (Sandberg, Sartori, Heidrich, Dawson, Dascalaki, Dimitriou, Vimm-r, Filippidou, Stegnar, Zavrl and Brattebø, 2016) for similar purposes.

However, even though useful for the aggregated analysis of entire national building stocks, these top-down methods fall short when it comes to evaluating the effects of specific building technologies, for instance, energy retrofits, or when the aim is to forecast such

²Covenant of Mayors for Climate & Energy homepage: <https://www.covenantofmayors.eu/>

effects into the future (Kavgic et al., 2010). Moreover, top-down models will fail when discontinuity is encountered as they have no inherent capability to model discontinuous changes in technology. These limitations make top-down models limited in their ability to predict the impact of retrofit interventions, which is the theme of this thesis. For this reason, focus is solely on bottom-up modelling. For the readers interested in top-down modelling, I refer to the reviews of Swan and Ugursal (2009), Oladokun and Odesola (2015), and Li et al. (2017).

Urban building energy modelling (UBEM) is a relatively new but rapidly growing research field that seeks to facilitate analyses on the building stock by combining the effects of individual bottom-up building energy modelling (BEM) into an aggregated urban-scale model (Reinhart and Davila, 2016). Contrary to the purely statistical approaches of top-down modelling, bottom-up UBEM relies on the use of either physical laws for representing the thermodynamic mechanisms of buildings, or a statistical treatment on a set of explanatory variables (Figure 1.1). These two sub-approaches to bottom-up UBEM are outlined in the following sections.

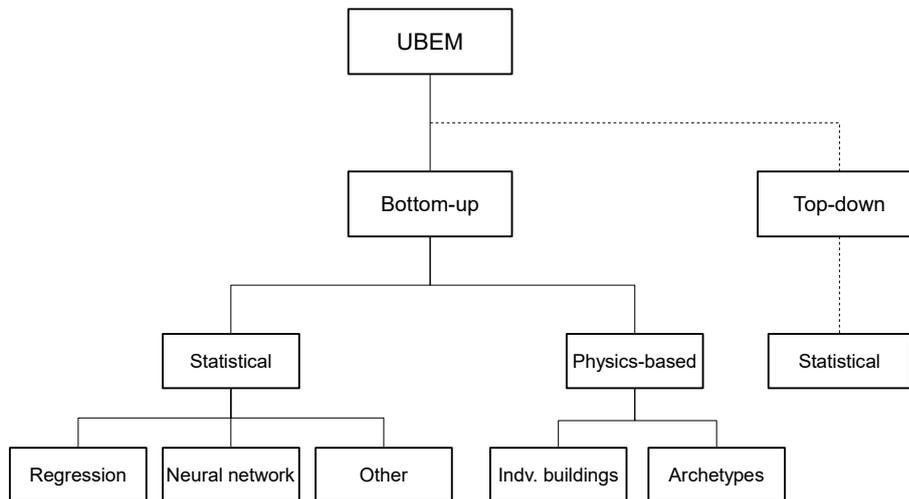


Figure 1.1. Modelling approaches to urban building energy modelling (UBEM). The top-down approaches are not reviewed in this thesis.

1.2.1 Statistical bottom-up

Building energy models based on empirical data are often constructed around simple low-fidelity statistical or black-box methods (Lee et al., 2015). These models make use of historical data gathered from building meters, building management systems, weather stations, supplier billing data, or from less accurate public databases depending on the modelling scale. Statistical models can be applied for highly accurate predictions of building energy use, both on the aggregated and disaggregated scale, and will often work with even very limited input data. However, purely statistical bottom-up UBEM does not provide the same interpretability as physics-based modelling, and often suffers from a lack of transparency. In fact, it often turns out to be very difficult to validate and explain unexpected results or model behaviour, as no causal relationships are defined. It is essential when performing statistical analysis to keep in mind that correlation does not imply causation; even the most accurate and robust statistical model can never be used to infer causal relationships between its variables. As a consequence, statistical models often

fail to accurately describe building behaviour and energy consumption for new unseen boundary conditions that has not been used for training its model parameters.

Frequently used techniques for statistical bottom-up modelling include ordinary-least-squares (OLS) or linear regression (LR), random forest (RF) regression, artificial neural network (NN) and other machine learning approaches, such as support vector machines (SVM) and Bayesian networks. Several of these techniques are reviewed in depth in e.g. [Tso and Yau \(2007\)](#), [Aydinalp-Koksal and Ugursal \(2008\)](#), [Swan and Ugursal \(2009\)](#), and [Oladokun and Odesola \(2015\)](#). An overview of the most recent statistical bottom-up UBEMs and building stock models in the literature is given in [Table 1.1](#).

1.2.2 Physics-based bottom-up

Contrary to the black-box approach applied in statistical methods, physics-based models are white-box models explaining the causal relationship between variables, and therefore do not rely on historical data ([Swan and Ugursal, 2009](#)). These models rely on the laws of physics, which makes this approach the most flexible when it comes to integrating new technologies. However, occupant behaviour must be assumed, which is one of the biggest weaknesses of the physics-based approach. The level of disaggregation, complexity, and assumptions made in these models varies significantly. One can generally divide the physics-based bottom-up approach into models of high fidelity (complex models), and models of low fidelity (simplified models).

High-fidelity physics-based models rely on advanced and often complex building energy modelling based on non-linear hygrothermal processes and physical conservation laws. Many of the developed physics-based models and retrofit tool kits available are based on highly available calculation engines, e.g. EnergyPlus and IDA ICE ([Table 1.1](#)), enabling transient calculations typically in hourly time steps or lower. The potentially unlimited resolution and level of detail of high-fidelity models is perhaps their biggest advantage, but also one of their biggest weaknesses: UBEM founded on a sheer building-physical description suffers from model complexity, which leads to long data inputs, an extensive level of explicit information needed for describing the buildings in the urban setting, and long simulation run time ([Lee et al., 2015](#)).

Low-fidelity physics-based models are based on reduced-order modelling (ROM), where several physical model parameters are lumped together to form less complex model structures. Such models are based on quasi-steady-state heat balance equations or simple dynamic methods using resistance-capacitance (RC) formulations, as used in, for example, the international standard ISO 13790:2008. State-space formulation is often used to conveniently represent ROMs in a compact mathematical vector-matrix notation as input, output and state variables related by first-order differential equations ([Madsen, 2008](#)).

1.2.3 Modelling approach for UBEM

The need of interpretability, and a physical meaningful model structure, is a strong argument for choosing a bottom-up physics-based approach to UBEM when analysing the effect of implementing ECMs on an existing building stock. The abilities of a physics-based approach enable one to carry out retrofit analysis, changing boundary conditions, and to forecast urban-scale energy use under previously unseen conditions. In practice, however, physics-based UBEMs are often calibrated using measured performance data, where uncertain model parameters are tuned to fit an observed training dataset, leading to what could be defined as a hybrid approach between physics-based and statistical bottom-

Table 1.1. Bottom-up UBE M studies.

Bottom-up approach	Scale of application	BEM tool	Archetype modelling	Calibration of model	Reported error ^{a)}	Reference
Physics-based	Country	IDA ICE	✓		N/A	Tuominen et al. (2014)
	Country	ROM	✓	✓	13%	Reynders et al. (2014)
	Country	ROM	✓		44%	Balaras et al. (2016)
	Country	ROM			2-7.5%	Buffat et al. (2017)
	City	EnergyPlus	✓		4%	Caputo et al. (2013)
	City	EnergyPlus	✓		40% [†]	Davila et al. (2016)
	City	EnergyPlus			15%	Dogan and Reinhart (2017)
	City	EnergyPlus			49%	Lin et al. (2017)
	City	IDA ICE	✓		1-61%	Nageler et al. (2017)
	City	IDA ICE	✓	✓	N/A	Dominković et al. (2018)
	City	eQuest	✓	✓	10-13%	Heiple and Sailor (2008)
	City	ROM		✓	2-7.5%	Zhao et al. (2016)
	City	N/A	✓		18%	Shimoda et al. (2004)
	Neighbourhood	EnergyPlus			N/A	Zucker et al. (2016)
	Neighbourhood	EnergyPlus	✓	✓	44-47%	Sokol et al. (2017)
	Neighbourhood	IDA ICE	✓	✓	0% [‡]	Nageler et al. (2018)
	Neighbourhood	ROM		✓	5-25%	Nouvel et al. (2015)
	Neighbourhood	ROM	✓		3%	Österbring et al. (2016)
	Neighbourhood	ROM	✓		6%	Remmen et al. (2017)
	Neighbourhood	ROM			11% [†]	Panão and Brito (2018)
Statistical	Country	RA;ROM		✓	N/A	Brögger et al. (2018)
	City	LR;RF;SVM		✓	N/A	Kontokosta and Tull (2017)
	City	RA		✓	4%	Booth et al. (2013)
	City	RA		✓	N/A	Hsu (2014)
	City	RA		✓	9-13%	Mastrucci et al. (2014)
	City	RA		✓	N/A	Braulio-Gonzalo et al. (2016)
	City	RA		✓	2%	Kristensen, Brun and Petersen (2018)*
	City	RA	✓	✓	N/A	Moghadam et al. (2018)

^{a)} Mean error on aggregated scale if available. Studies where another error measure is used are marked with (†). Studies where the error assessment is performed on calibration data are marked with (‡). Studies where no error assessment is reported are marked with (N/A).

* Study presented in this PhD thesis.

up approaches (grey-box models). Stochastic calibration techniques have in recent years become the standard approach, not relying on a single deterministic values for uncertain quantities, but incorporating and exploring a distribution of possible values; especially the Bayesian approach to probabilistic modelling has become popular (Section 1.4.2).

Regardless of the modelling approach, and to some degree the calibration approach, the overall challenge of UBE M remains that of collecting and assigning all the necessary data inputs to establish a sufficiently detailed model of the buildings in an area of interest, and to actually carry out the required simulations. Even though some studies seek to establish all-inclusive models based on large data surveys (e.g. Dogan and Reinhart, 2017), most state-of-the-art research within UBE M and building stock modelling incorporates some sort of clustering or archetype segmentation of the buildings to reduce the burden of data gathering, model calibration and simulation time. In the following section, I will outline the concept of *archetypes* and their use in UBE M.

1.3 Building archetypes

In order to set up and carry out UBE M on a larger urban area, it can be necessary to segment the building stock into a reduced number of manageable building archetypes (Reinhart and Davila, 2016). In fact, most studies on bottom-up UBE M found in the current literature make use of some sort of archetype simplification of its building stock to simplify the analysis (Table 1.1); however, archetypes can be many different things. The research and development of building archetypes for urban-scale modelling is still in

its infancy and to the best of my knowledge, no exhaustive definition of what a building archetype really is and how to establish it exists. The process of defining archetypes has been described in recent work by researchers from MIT (e.g. [Cerezo et al., 2015](#); [Reinhart and Davila, 2016](#); [Cerezo et al., 2017](#); [Sokol et al., 2017](#)) to constitute two to three steps before being applied in UBEM simulations:

1. Classification of buildings into archetypes,
2. Characterization of archetype parameters, and potentially;
3. Calibration and validation of uncertain archetype parameters.

More detailed descriptions of these steps are provided in the following sections.

1.3.1 Classification

The segmentation of buildings into a much smaller subset of archetypes, that represent groups of typologically identical buildings, is an important first step in the definition of representative building typologies. Buildings are usually grouped by one or more readily available building classifiers such as *usage type*, *construction year*, *technical systems*, *geometry*, and *climatic conditions*. Depending on the available information about the buildings, the segmentation can be either coarse, detailed, or anything in between, thereby resulting in a few general archetypes, or many highly specific and homogeneous archetypes. [Theodoridou et al. \(2011\)](#) is an example of a coarse classification where approx. 2,500,000 multi-family houses (MFHs) from the Hellenic building stock were grouped into only five archetypes based on their construction year and usage type (all being MFHs), while [Fracastoro and Serraino \(2011\)](#) applied a much more detailed segmentation scheme to group approx. 900,000 dwellings from the Italian building stock into 3,168 archetypes based on their construction year, geometry, heating systems and climatic conditions. A summary of publications where the archetype approach is applied to segment a building stock is given in [Table 1.2](#).

The level of resolution applied in the classification step is obviously of paramount importance for the resulting predictive performance of the UBEM, as well as its ability to accurately incorporate ECMs and other manipulations of the model ([Cerezo et al.](#),

Table 1.2. Classification, characterization and calibration approaches used in archetype-based studies. All studies apply the synthetic characterization method.

No. of buildings	No. of archetypes	Classification parameters	Characterization approach	Calibration approach	Reference
N/A (mixed)	30	Type;Age;System	Deterministic	None	Heiple and Sailor (2008)
877,144 (SFH)	3,168	Type;Age;System; Geometry;Climate	Deterministic	None	Fracastoro and Serraino (2011)
2,514,161 (MFH)	5	Type;Age	Deterministic	None	Theodoridou et al. (2011)
35 (flat)	1	Type	Probabilistic	Bayesian	Booth et al. (2012)
1,320 (MFH)	7	Type;Age	Deterministic	LR	Dall'O' et al. (2012)
35 (mixed)	13	Type;Size;System	Deterministic	Clustering	Famuyibo et al. (2012)
N/A (School)	1	Type	Probabilistic	Bayesian	Tian and Choudhary (2012)
450 (SFH)	1	Type;Age	Probabilistic	Bayesian	Kristensen, Choudhary, Pedersen and Petersen (2017)*
336 (SFH) ^{a)}	1 (case A)	Type	Deterministic	None	Cerezo et al. (2017)
	4 (case B)	Type;Age	Deterministic	None	
	4 (case C)	Type;Age	Probabilistic	None	
	4 (case D)	Type;Age	Probabilistic	Bayesian	
399 (SFH) ^{b)}	N/A	Type;Age;System	Probabilistic	Bayesian	Sokol et al. (2017)
54 (SFH)	6	Type;Age	Deterministic	N/A	Dominković et al. (2018)
50 (SFH)	1	Type;Age	Probabilistic	Bayesian	Kristensen, Hedegaard and Petersen (2018a)*
30 (SFH)	8	Type;Age	Deterministic	N/A	Nageler et al. (2018)

^{a)} Four different combinations of archetype classification, characterization and calibration were assessed (case A to D).

^{b)} Each building was defined by more than one archetype. All archetypes were calibrated simultaneously.

^{c)} Studies presented in this PhD thesis.

2015). Surprisingly, little work has been done to thoroughly investigate the effect of granularity in the archetype segmentation. Cerezo et al. (2017) and Sokol et al. (2017) both argue that this is a consequence of UBEM modellers not having access to measured energy demand data for individual buildings, which in turn makes it impossible to validate the effectiveness of a chosen classification scheme. An example of a study that actually performs this investigation is the work by Aksoezen et al. (2015), where *building age* was studied as an indicator of building energy use in the city of Basel, Switzerland, using the classification tree methodology. Another example is the work by Famuyibo et al. (2012), who used clustering analysis on Irish dwellings to derive 13 archetypes representing 65% of the building stock. In fact, clustering analysis has been used in several studies to classify and extract features from building data. Hsu (2015) used and compared different clustering methods in an effort to segment a dataset of annual metered energy use from approx. 3,900 MFHs in New York City using more than 250 classifiers. do Carmo and Christensen (2016) applied similar cluster-wise regression and classification methods to decompose hourly district heating time series of 139 Danish SFHs into daily archetypical heating profiles using 16 building and household characteristics. Gianniou et al. (2018) repeated this exercise using a much larger dataset of hourly district heating time series from 8,293 Danish SFHs, resulting in the discovery of nine daily archetypical behavioural profiles, but only five different profiles when segmented with regards to the energy use intensity. Using regression analysis, they found *building age* and *building area* to be highly significant profile classifiers, while *number of adults*, *teenagers* and *children* were deemed insignificant.

1.3.2 Characterization

After having classified the building stock into archetype typologies, they must undergo a characterization process to define their functionality and BEM input parameters. This is often a difficult task. From a general point of view, archetypes are characterized and used in two fundamentally different ways throughout research literature; either purely deterministically as *real example buildings* or as *average synthetic buildings*, where the latter may be done either deterministically or probabilistically. Both approaches are outlined below with emphasis on the most-used average synthetic approach.

Real example buildings are characterized based on the characteristics of an audited real exemplary building (Cerezo et al., 2017), i.e. they are carefully selected whole building representations of the building stock. An example of this approach is the trans-European IEE project TABULA (performed from 2009-2012), in which residential archetype typologies were established for 20 EU countries (Loga et al., 2016) using *building usage type*, *age* and *size* as classifiers for the segmentation³. The TABULA building typologies have been used in many studies in recent years, e.g. on the Italian building stock (Ballarini et al., 2014) and the Hellenic building stock (Balaras et al., 2016).

Average synthetic building archetypes, also known as the *average approach*, are characterized based on a statistical analysis of a large sample of audited real buildings or literature data, such as national building surveys, EPC data, and building codes and standards. Often, modellers collect parameter data from a mixture of these information sources in an effort to draw a holistic picture of the archetype; however, the availability of data is often very specific to the city and country in question. Moreover, data acquisition easily becomes a delicate compromise as data access is often limited due to legal constraints and privacy considerations; especially in terms of measured consumption

³TABULA WebTool: www.webtool.building-typology.eu

data. A particularly difficult task is the description of occupant-related parameters that influence many aspects of building operation and, in extension, the associated energy use (Huebner et al., 2015). One can thus seldomly expect to have data that are sufficiently elaborate to obtain a purely deterministic characterization of the archetype parameters, i.e. a single value assigned each parameter Cerezo et al. (2017). Quantifying, modelling and propagating this uncertainty is one of the biggest challenges in archetype modelling, which could be the reason why many studies proceed with a deterministic characterization of the archetype parameters (Table 1.2); this is often in terms of average values of survey data or EPC data. In recent years, however, an increasing amount of work has been conducted on probabilistic treatment of model uncertainty, including archetype modelling.

The final step in the process of defining archetypes is step 3, calibration and validation of uncertain archetype model parameters; however, before going into this, it is necessary to establish a terminology and conceptual understanding of what uncertainty entails in the context of this thesis.

1.4 Uncertainty modelling

Uncertainty is virtually ubiquitous in building energy modelling, as well as in any other data-driven modelling field. Inaccuracies in BEM simulations may be attributed to the presence of several uncertainties that tend to affect any modelling process. Examples are: the simplifications made in the used model structure, uncertainties in the characterization of its parameter values, and numerical approximations. Moreover, in assessing the error between model output and measurements hereof, one has to take into account the presence of noisy observations. These phenomena are all well-known sources of error in mathematic models; however, working with archetypes for urban building energy modelling introduces yet another layer of uncertainty; variability due to the archetype segmentation itself, which could be denoted *archetype heterogeneity*. Heo (2011) and Booth et al. (2012) have already addressed several of these issues in their work on uncertainty quantification in BEM and housing stock models. I here take the freedom to reformulate, expand and elaborate on their definitions to make them more consistent with the terminology applied in this thesis work and the current state-of-the-art:

Structural uncertainty Variability that occurs due to the inadequacy of the BEM in describing the true energy consumption process of the building. This is also known as model bias or model discrepancy and includes both modelling of thermodynamic processes as well as building operation and occupant behaviour.

Parameter uncertainty Building-level variability that occurs due to insufficient knowledge about BEM input parameters. This also includes variability due to human behaviour in terms of occupancy, operation of appliances, heating and cooling set point preferences, etc.

Archetype heterogeneity Archetype-level variability that occurs due to differences in building characteristics across the sampled buildings. Archetype heterogeneity is strongly affected by the granularity of the archetype segmentation. This uncertainty can be difficult to distinguish from that of ordinary parameter uncertainty.

Numerical uncertainty Algorithmic variability that occurs due to numerical approximations, small sample sizes, insufficient convergence of calibration, extrapolation, etc.

Observation error Variability in measurements of the true process, e.g. noisy observation of annually aggregated or hourly time series of building energy use, to be used for verification and validation of simulations, or as training data for model calibration. Repeated measurements can help to quantify this variability.

Urban building energy modelling – and archetype parameter characterization in particular – is a natural example of where a probabilistic view on data variability is beneficial (Cerezo et al., 2017). By assigning probability density functions (PDFs) to model input parameters (archetype parameters) to reflect their uncertainty of estimation instead of just a fixed best-guess or average value, we are able to incorporate this variability and propagate its effects throughout the model structure to the output by applying probabilistic sampling techniques, e.g. Monte Carlo (MC) or Latin hypercube sampling (LHS). This process of quantifying uncertainty is commonly known as uncertainty analysis (UA) and has many applications in building energy assessment (see review by Tian et al., 2018) and retrofit decision making (e.g. Booth and Choudhary, 2013); however, relatively few studies on archetype modelling apply probabilistic characterization of the parameters (Table 1.2). Sensitivity analysis (SA) is a probabilistic analysis technique closely related to UA, but instead of quantifying output variability, SA focuses on quantifying the effects of individual model parameters and their interactions with each other through the model structure (see reviews by Tian, 2013; Menberg et al., 2016). This is useful for selecting a reduced set of important parameters to concentrate on in, for example, early stages of building design (e.g. Heiselberg et al., 2009), or when selecting parameters for model calibration.

Modellers have traditionally employed a strict frequentist approach to UBEM and archetype parameter characterization where deterministic parameter estimates are derived from limited samples of empirical data (Table 1.2). This type of analysis somewhat ignores the abovementioned uncertainties, especially variability between parameters of seemingly identical buildings risk being ignored, which could result in parameter uncertainty and archetype heterogeneity being underestimated. No model is perfect, and as the data available for archetype characterization and model validation are often limited both in terms of quantity (only few data points) and quality (low temporal resolution), a purely deterministic approach to data analysis may easily lead us to arrive at conclusions and decisions from our modelling work which ignore other plausible solutions; solutions that may yield better results. If we instead embrace uncertainty, we can move from the notion of having a single deterministic model to that of keeping a potentially infinite collection of models and combining them to make informed decisions under uncertainty. This is one of the fundamental ideas behind probabilistic modelling and Bayesian inference in particular (Frigola-Alcalde, 2015).

The following section contrasts the two leading schools of probability in order to have a proper foundation for selecting the most appropriate school for the work of this thesis.

1.4.1 Frequentism vs. Bayesianism

There are, in general, two broad categories of probability interpretations, which are often denoted the *objective* and the *subjective* schools of probability, referring to the frequentist and Bayesian views, respectively.

In a pure frequentist view, probabilities are only meaningful as frequencies, or proportions, of occurrences from physical phenomena, e.g. dice rolling. Probabilities are interpreted objectively and as a constant, for example, there is a 1/6 probability of a

die yielding a six. We might not get a six if we roll the die only six times, but a $1/6$ of all rolls will yield a six when the number of rolls approaches the limit of infinity. This also means that in a strict frequentist view, it is meaningless to talk about the probability of e.g. the U-value of a building's external wall; the U-value is (by definition) a single *fixed value*, and to talk about a frequency distribution for a fixed value is nonsense. If repeating 50 measurements of the U-value, a frequentist would take the mean value of the sample as an estimate of the true U-value and derive a confidence interval (CI) around the mean, which includes the true parameter with, for instance, 95% probability.

In a Bayesian view, on the other hand, the concept of probability is extended to cover degrees of certainty about statements. Probabilities are interpreted subjectively and as a dynamic changing phenomena. As a Bayesian, we would claim to have measured the U-value with some probability based on the sample of measurements and our own knowledge about the parameter. This means that we can meaningfully talk about the true U-value as the *probability of values* in a given range without relying on it to be a specific value with some uncertainty specification around to account for its estimation. That probability is the product of our prior beliefs of what the measurement result will be and the probability of the sampled data. Given the extreme theoretical case of a totally unbiased (non-informative) prior understanding of the U-value where we believe all possible values to be equally plausible, we would estimate the exact same value as in a frequentist view, if we take the most probable value – the maximum a-posteriori probability (MAP) estimate – as a point estimate of its value. However, in a strict Bayesian view, one would never reduce uncertainty to point estimates such as the MAP estimate. Instead, a Bayesian might report the highest posterior density (HPD) interval, i.e. the shortest possible interval for the parameter with a fixed probability, say 95%. He would claim to be 95% sure that the parameter is in this interval, whose width yields the highest possible accuracy. If using a flat (uniform) prior, the frequentist CI and the Bayesian HPD may numerically agree.

Bayesian inference techniques rest on the application of Bayes's theorem (Eq. 1.1) to describe the conditional probability – our subjective *degree of belief* – of an event, or an unknown parameter as in the case of the U-value, based on prior knowledge of the event:

$$p(\theta | \mathcal{D}) = \frac{p(\mathcal{D} | \theta) p(\theta)}{p(\mathcal{D})} \tag{1.1}$$

where:

$p(\theta | \mathcal{D})$ = Posterior

$p(\mathcal{D} | \theta)$ = Likelihood

$p(\theta)$ = Prior

$p(\mathcal{D})$ = Evidence

Bayes's theorem links the *a priori* degree of belief of the parameter before accounting for evidence, $p(\theta)$, and the *a posteriori* degree of belief of the parameter after accounting for evidence, $p(\theta | \mathcal{D})$, where the evidence is represented in terms of available data $p(\mathcal{D})$. The term $p(\mathcal{D} | \theta)$ is the likelihood of the data occurring given that the parameter θ is true. The likelihood is a statistical model for the observed data \mathcal{D} as a function of the parameter θ . Since the evidence term, $p(\mathcal{D})$, is going to be the same for all values of θ , it can be ignored as a constant in practical inference problems reducing the posterior $p(\theta | \mathcal{D})$ to be proportional to the product of the likelihood $p(\mathcal{D} | \theta)$ and the prior $p(\theta)$:

$$p(\theta | \mathcal{D}) \propto p(\mathcal{D} | \theta)p(\theta) \quad (1.2)$$

Example: Estimate the mean of a population

To exemplify the differences between frequentist and Bayesian inference, we return to our example with the estimation of the U-value of an external wall. Consider an experiment in which we repeat $i = 1, 2, \dots, N$ measurements $\mathcal{D} = \{y_i\}$ of the U-value with a known measurement error variance, σ^2 , which we take to be an independent and identical sequence of Gaussian distributed measurement noise:

$$y_i \sim \mathcal{N}(\theta, \sigma^2), \text{ for } i = 1, 2, \dots, N \quad (1.3)$$

We are interested in estimating the parameter θ as the underlying value of the U-value. The data likelihood is then the Gaussian probability density function over all the measurements $p(\mathcal{D} | \theta) = \prod_{i=1}^N (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (y_i - \theta)^2\right\}$. We can rewrite this probability model to express the likelihood of a given parameter value θ as a function of all the observed data \mathcal{D} ; an expression known as the likelihood function:

$$\mathcal{L}(\theta | \mathcal{D}) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \theta)^2\right\} \quad (1.4)$$

In a frequentist view, we now apply the method of maximum likelihood estimation (MLE) to find the parameter value $\hat{\theta}_n$ that maximizes the probability of the data. In this simple case, the ML estimator is equal to the empirical mean of the data sample:

$$\hat{\theta}_n = \frac{1}{N} \sum_{i=1}^N y_i \quad (1.5)$$

In a Bayesian view, we start by specifying our prior beliefs about the parameter $p(\theta)$. Selecting a *conjugated* prior model for the likelihood function results in a convenient analytical solution where the posterior has the same form as the prior. The conjugated prior $p(\theta)$ to a Gaussian likelihood is also Gaussian with prior hyperparameters θ_0, σ_0 :

$$p(\theta) = \mathcal{N}(\theta_0, \sigma_0^2) \quad (1.6)$$

The resulting posterior distribution $p(\theta | \mathcal{D})$ cf. Bayes's theorem (Eq. 1.1) is then the product of two Gaussians, which conveniently results in another Gaussian with posterior hyperparameters θ_n, σ_n :

$$p(\theta | \mathcal{D}) = \mathcal{N}(\theta_n, \sigma_n^2) \quad (1.7a)$$

$$\theta_n = \sigma_n^2 \left(\frac{\theta_0}{\sigma_0^2} + \frac{N\bar{y}}{\sigma^2} \right) \quad (1.7b)$$

$$\sigma_n^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{2}{\sigma_0^2}} \quad (1.7c)$$

Figure 1.2 is an example with the U-value measured with a known observation error variance $\sigma^2 = 0.1^2$ and six different combinations of sample size $N = \{2, 10\}$ and prior variance $\sigma_0^2 = \{0.01^2, 0.1^2, 1^2\}$. The underlying, but unknown, value of θ used for sampling the data in this experiment is set to $\theta_{\text{true}} = 0.55$. The prior hyperparameter mean is $\theta_0 = 0.45$ for all six cases.

In case of a non-informative flat prior with $\sigma_0^2 = 1^2$ (first row of Figure 1.2), it is seen how the Bayesian posterior distribution coincides with the likelihood, which, when maximized, gives the frequentist ML estimate. However, introducing more prior knowledge in terms of $\sigma_0^2 = 0.1^2$ and $\sigma_0^2 = 0.01^2$ (row two and row three of Figure 1.2, respectively) biases the posterior distribution away from the data likelihood distribution towards the prior, while this has no effect on the MLE. The true U-value is shown in the figure, but in reality, this value is unknown and can never be estimated precisely; thus, we would have little chance of accurately validating our estimates. In this case study it is seen how the MLE overestimates the true value (most pronounced in the case of a small data sample $N = 2$), while the Bayesian posterior overestimates and underestimates the value using a non-informative ($N = 2$, $\sigma_0^2 = 1^2$) and a very informative ($N = 2$, $\sigma_0^2 = 0.01^2$) prior, respectively. Applying moderate prior knowledge ($N = 2$, $\sigma_0^2 = 0.1^2$) results in the best estimate of the true value in this case study.

1.4.2 Bayesian inference in building energy modelling

The Bayesian school of probability and its inherent probabilistic approach to uncertainty modelling seem attractive from an engineering perspective (Carstens et al., 2018). In the case of archetype characterization where little or no data are available to describe archetype features, a strict frequentist characterization using, for example, simple average values from survey data may not only result in error-prone point estimates, it is also not capable of addressing and propagating the underlying uncertainty of the estimation. Using Bayesian methods, we are fully allowed to incorporate any prior knowledge that could help in the inference of unknown parameters; knowledge which is often available from literature data, historical records or as educated guesses by the modeller. Bayesian inference may therefore be better suited for UBEM applications (Cerezo et al., 2017).

A direct characterization of archetype parameters from available data (mean value estimation) is, however, rarely sufficient to fully account for model discrepancy, archetype heterogeneity, observation error and numerical uncertainty, which may easily cause large inaccuracies when archetype-based models are applied for urban-scale predictions. For that reason, it can be necessary to employ probabilistic inference techniques to fine-tune parameter distributions in an effort to match model output to measured training data; a process known as model calibration, or simply system identification.

Calibration of building energy models (Coakley et al., 2014) has traditionally been carried out using deterministic trial-and-error methodology or frequentist optimization methodology; however, probabilistic calibration is becoming increasingly popular in the scientific literature with many new studies being published on Bayesian methods, in particular, in the last few years. An overview of relevant studies on Bayesian calibration is given in Table 1.3 based on literature reviews by Lim and Zhai (2017b), Lim and Zhai (2018) and Chong and Menberg (2018), supplemented with additions from the newest literature.

Hundreds or thousands of model evaluations are often needed to calibrate a high-dimensional input space of uncertain parameters in a probabilistic setting using, for

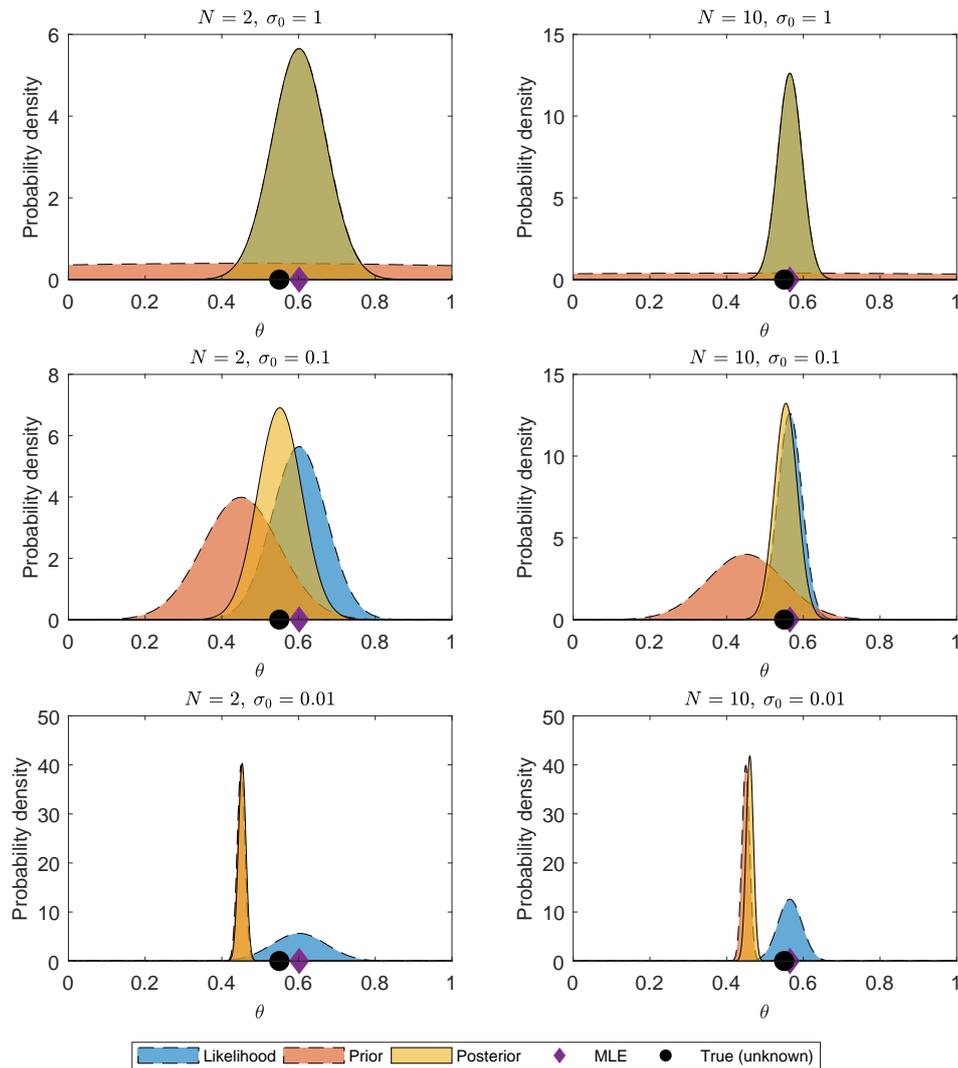


Figure 1.2. Frequentist maximum likelihood estimate (MLE) and Bayesian posterior distribution of θ for different combinations of sample size N and prior variance σ_0^2 . The first row of plots corresponds to a case with weak or no prior knowledge ($\sigma_0^2 = 1^2$). The second row of plots corresponds to a case with moderate prior knowledge ($\sigma_0^2 = 0.1^2$). The third row of plots corresponds to a case with strong and very informative prior knowledge ($\sigma_0^2 = 0.01^2$). The left and right columns correspond to cases with a small ($N = 2$) and large ($N = 10$) sample size, respectively, to establish the likelihood distribution.

example, Monte Carlo sampling. The calibration process can therefore easily become infeasible for high-fidelity UBEM or archetype models relying on tools as EnergyPlus or IDA ICE. As a consequence, most studies employing Bayesian calibration to tune bottom-up physics-based BEMs therefore rely on the application of emulators in place of the BEM tool itself to generate these model evaluations (Table 1.3). An emulator is a statistical meta-model, often some sort of regression model, of the physical BEM that allows a much faster simulation of the physical process using the uncertain parameters as the only explanatory variables (Lim and Zhai, 2017a). However, as the emulator is trained on a limited number of simulation data (input/output) – perhaps 10 to 200 evaluations – another layer of numerical uncertainty is inevitably introduced in the calibration due to insufficient coverage of structural model behaviour.

Table 1.3. Stochastic inference techniques used for calibrating bottom-up physics-based building energy models (BEMs).

Scale of application	No. of buildings	BEM tool	Emulator of BEM tool ^{a)}	Resolution of calibration data	Reference
Stock (Offices)	765	ROM	LR	Annual	Zhao et al. (2016)
Archetype (School)	N/A	EnergyPlus	SRC;MARS	Annual	Tian and Choudhary (2012)
Archetype (Flat)	35	ROM	GPR	Daily	Booth et al. (2012)
Archetype (SFH) ^{b)}	336	EnergyPlus	None	Annual	Cerezo et al. (2017)
Archetype (SFH) ^{c)}	399	EnergyPlus	PolyR	Monthly	Sokol et al. (2017)
Archetype (SFH)	450	ROM	GPR	Annual	Kristensen, Choudhary, Pedersen and Petersen (2017)*
Archetype (SFH)	50	ROM	None	3-hourly	Kristensen, Hedegaard and Petersen (2018a)*
Building (Office)	1	EnergyPlus	GPR;LR	Monthly	Li et al. (2016)
Building (Office)	1	EnergyPlus	GPR	Hourly	Kim and Park (2016)
Building (Office)	1	ROM	GPR	Monthly	Heo et al. (2012)
Building (Office)	1	ROM	GPR	Monthly	Kim et al. (2013)
Building (Office)	1	ROM	GPR; PW-LR;None	Hourly	Manfren et al. (2013)
Building (Office)	2	ROM	GPR	Monthly	Heo et al. (2015)
Building (SFH)	1	ROM	GPR	6-hourly;Daily; Weekly;Monthly	Kristensen, Choudhary and Petersen (2017)*
Building (Office)	1	eQuest	GPR	Monthly	Kang and Krarti (2016)
Building (Mixed) ^{d)}	3	EnergyPlus	RF;NN	Monthly	Nagpal et al. (2018)
Chiller ^{e)}	2	EnergyPlus; TRNSYS	GPR	Hourly	Chong et al. (2017)

^{a)} (GPR): Gaussian process regression; (LR): Linear regression; (MARS): Multivariate adaptive regression; (NN): Neural network; (PolyR): Polynomial regression; (PW-LR): Piecewise linear regression; (RF): Random forest regression; (SCR): Standardized regression coefficient; (None): No emulator was applied, i.e. the BEM was calibrated directly.

^{b)} Four different archetypes were calibrated.

^{c)} Each building was defined by more than one archetype. All archetypes were calibrated simultaneously.

^{d)} Three different buildings were calibrated: a residential building, an office building and a laboratory building.

^{e)} Two different water-cooled chiller systems, in two different buildings, were calibrated.

* Studies presented in this PhD thesis.

By far the most used, and most complex, emulator is the Gaussian process regression (GPR), which is a stochastic process that can be seen as a normal distribution of infinite dimensionality over functions of the data with a continuous domain, e.g. time and/or space (Rasmussen and Williams, 2006). Many studies have successfully applied the methodology described in *Bayesian calibration of computer models* by Kennedy and O’Hagan (2001) or variations thereof (Higdon et al., 2004), which relies on the use of GPR emulation to tune model parameters and simultaneously fit a bias function to account for residual variability that cannot be captured by the model physics (model discrepancy). The benefit of the Kennedy & O’Hagan (KOH) approach is that modellers may calibrate very complex model structures using only a limited sample of measurement data and computer simulations from the actual BEM.

To the best of my knowledge, no study employing the KOH framework or any other Bayesian technique to tune models of building energy use applies time series data for the calibration; i.e. they do not use *time* as the explanatory domain to identify the temporal dynamics. Most studies apply calibration data on an aggregated temporal scale, e.g. monthly or annually aggregated data (Table 1.3), which makes it difficult to identify dynamic parameters such as the capacity of the thermal mass. Such dynamic parameters have to be discovered with data on a much smaller timescale, e.g. hourly or sub-hourly (Madsen et al., 2015). Calibration data of an hourly resolution has been applied in only a few recent studies on building energy modelling by Manfren et al. (2013), Kim and Park (2016) and Chong et al. (2017), but it remains to be shown how to apply hourly time series data in Bayesian calibration of BEMs and archetype models, for accurate predictions of out-of-sample time series of hourly building energy use.

In addition to the issue of temporal resolution of building models, UBEM validation is most often conducted on an aggregated spatial scale, e.g. on the scale of neighbourhoods, zip codes, cities, or even entire national building stocks, using aggregated temporal

performance data, e.g. monthly or even annually aggregated consumption data, where building-level errors tend to average out (the cancellation effect), resulting in very low errors (Table 1.1). Future work on UBEM and archetype modelling therefore ought to focus on modelling, calibration and validation on a higher temporal and spatial scale to ensure an accurate representation of individual buildings. Brøgger and Wittchen (2018) support this argument in their review of methodologies for building stock modelling where they find that modellers rarely validate the central assumptions made in their models in any meaningful way, including the representativeness of the chosen archetype classification scheme.

1.5 Summary of challenges

In summing up the challenges of state-of-the-art bottom-up archetype-based UBEM from the current scientific literature, it is found that it contains little or no work on:

- how probabilistic archetype-based UBEMs can be characterized and calibrated without relying on the application of statistical emulators, which could eliminate a layer of numerical uncertainty;
- archetype-based time series modelling and calibration that can cope with system dynamics satisfactorily; and
- how the presence of archetype heterogeneity and parameter correlation affects performance and what could be done to incorporate this phenomenon in archetype characterization and calibration.

1.6 Research objectives

This thesis aims to meet the challenges of bottom-up archetype-based urban energy modelling outlined in Section 1.5 through the exploration of two research objectives:

1. To develop a bottom-up archetype modelling and calibration methodology formulated in a Bayesian framework that takes into account and propagates all types of error and uncertainty throughout the calibration process, including *archetype heterogeneity* and *parameter correlation*. The methodology should incorporate time series modelling to account for building dynamics.
2. To establish a framework for analysing citywide residential retrofit potential year-by-year until year 2050 under uncertainty of the rate of energy renovations and climate change using bottom-up archetype-based UBEM.

The thesis is delimited to focus on the Danish residential building stock of detached single-family houses (SFHs) using case data from the municipality of Aarhus, Denmark. Data used throughout the thesis consist of publically available building characteristics data from the Danish Building and Dwelling Register (BDR) in combination with confidential data on district heating energy use of individual buildings of both annual and hourly temporal resolution. Data access is constrained to be used for research purposes only, and as a consequence, only anonymized and aggregated results are shown in the thesis.

1.7 Thesis outline

A number of peer-reviewed scientific journal papers and papers in conference proceedings have been produced during the PhD project, each covering different aspects and approaches towards resolving the research objectives, and they are in themselves considered the largest contribution of this paper-based thesis (see [List of publications](#), p. xi). However, I take the opportunity that the thesis format offers to build on the publications and extend the presentation of the material. The main text of the thesis (Chapters 2-8) is divided into three parts, each considering aspects of the research objectives through a motivation and presentation of relevant publications followed by an epilogue on the implications of the research. The three parts are structured as follows:

Part 1 is dedicated to the introduction and preliminary statistical analysis of the case data on the residential building stock of SFHs in the municipality of Aarhus, Denmark, which are used continuously throughout the thesis. Chapter 2 contains the presentation of the case data whereas Chapter 3 presents journal paper [P1] where the data are explored using statistical methods.

Part 2 addresses research objective 1 by providing insights into the exploration and development of new archetype-based modelling and Bayesian calibration methodologies, and how the lack of model input information and presence of uncertainty in UBEM can be handled. Chapter 4 presents conference paper [P2] on the inclusion of archetype heterogeneity using annual data, Chapter 5 presents conference paper [P3] on the effect of temporal resolution in calibration data, and Chapter 6 presents journal paper [P4] where a novel hierarchical (multi-level) archetype framework is proposed incorporating the criteria defined in research objective 1.

Part 3 addresses thesis objective 2 by employing the hierarchical archetype approach from Part 2 for urban building energy modelling. Chapter 7 presents conference paper [P5] where the archetype framework is applied to construct a UBEM of two suburban towns. Chapter 8 presents new, previously unpublished, research where the archetype framework is applied to set up a combined UBEM of the SFHs in Aarhus, Denmark, and subsequently used to forecast citywide heating energy use up until the year 2050 under uncertainty of energy renovation and climate change.

The thesis is rounded off with a unifying summary of main contributions and suggestions for future work in Chapter 9.

Part 1

Case data presentation and preliminary statistical analysis

Chapter 2

Urban case study: The city of Aarhus

2.1 The building stock of Aarhus

Aarhus is the second largest city of Denmark with a population of approx. 341,000 people in the municipality (approx. 273,000 people within the city) as of October 2018 ([Aarhus Kommune, 2018](#)). The population is projected to increase almost linearly by approx. 20% over the next 27 years to reach 405,000 people in 2045 ([Aarhus Kommune, 2018](#)). This expansion adds a considerable load to the existing city infrastructure in terms of energy supply, mobility and housing.

Aarhus is characterized by a high level of low-rise, some mid-rise and only few high-rise buildings. It is dominated by its large share of residential buildings, which make up approx. 84% of all buildings in the municipality and constitute approx. 45% of the entire heated floor area (Figure 2.1). The residential building stock is primarily composed of detached single-family houses (SFHs) in terms of building count, but the much larger mid-rise and high-rise multi-family houses (MFHs) constitute the largest share based on the overall heated floor area. Terraced houses (THs) only comprise a smaller share of the stock. Farm houses (FHs), dormitories and 24-hour day care centres together comprise around 1% of the building stock. All research presented from this point in the thesis is delimited to the residential stock of SFHs only.

The stock of SFHs in the municipality of Aarhus consists of approx. 27,800 buildings as of 2015, which make up the data used for research presented in this thesis. Basic building characteristics and property data on these buildings have been extracted from the Danish Building and Dwelling Register (BDR) – a database originally intended for public administration purposes – including the year of construction, building footprints, number of floors, utilized areas in basements and attics, materials of the building envelope, roof construction principles, etc. The distribution of the construction year of SFHs built from 1850 onwards is shown in Figure 2.2A along with heated floor area vs. construction year in Figure 2.2B. Buildings older than 1850 are not included, which corresponds to 0.3% (92 buildings) of all SFHs. The oldest building in the dataset is registered as being from 1600.

The interquartile range, i.e. the middle 50%, of the SFHs were constructed between 1951 and 1977, with the median being 1967 (Figure 2.2A). The construction of SFHs began to accelerate in the beginning of the 20th century up until the beginning of World

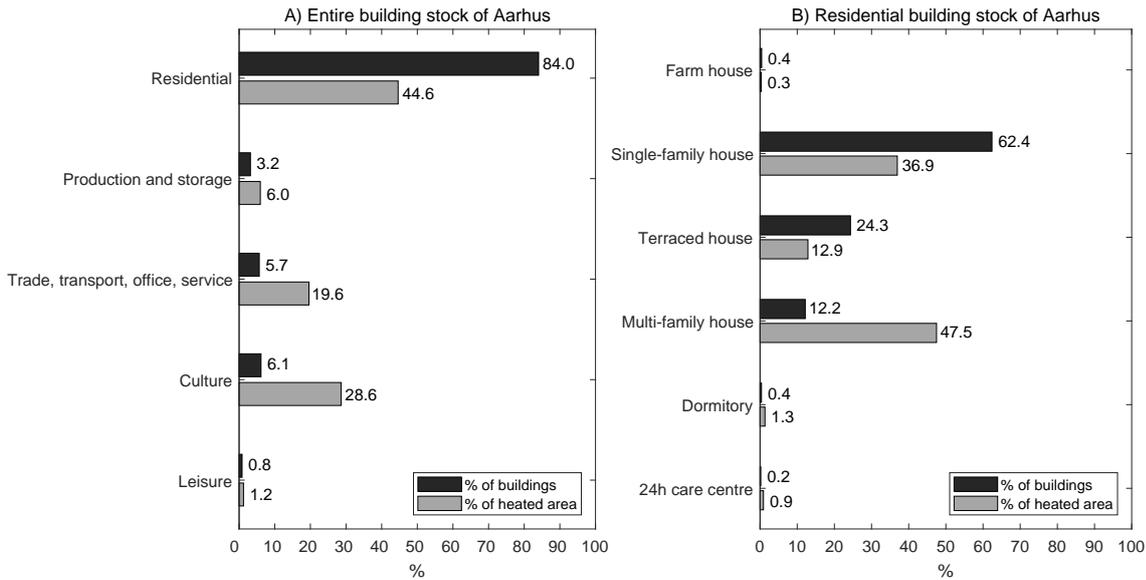


Figure 2.1. Building stock of the municipality of Aarhus as of 2016 (53,008 buildings) grouped into building usage categories. The percentage of buildings and heated floor area in each usage category, respectively, is given as horizontal bars. A) depicts the entire building stock divided into overall usage categories. B) depicts the residential building stock only subdivided into the different dwelling types. Data source: Danish Building and Dwelling Register (BDR).

War 2 in 1939, where it more or less stopped. Houses in this period were generally smaller than those of previous eras, and somewhat smaller than the overall average of the stock. With the end of the war in 1945, the construction of SFHs quickly increased to higher levels than before and continued this way at an almost exponential pace throughout the 1950s and 1960s when the country experienced a financial boom. This also means that houses slowly began to grow in size. The good times stopped all of a sudden with the energy crises in 1973-1974 and again in 1979-1980; thereafter the construction of SFHs quickly decelerated, and houses became slightly smaller again, until the early 1990s, where the rate of construction in Aarhus was as low as during the second world war. However, from the middle of the 1990s, it started to increase to a moderate plateau on which it has stayed since. This has not stopped the size of SFHs from increasing however; newly built SFHs are now larger than ever (Figure 2.2B).

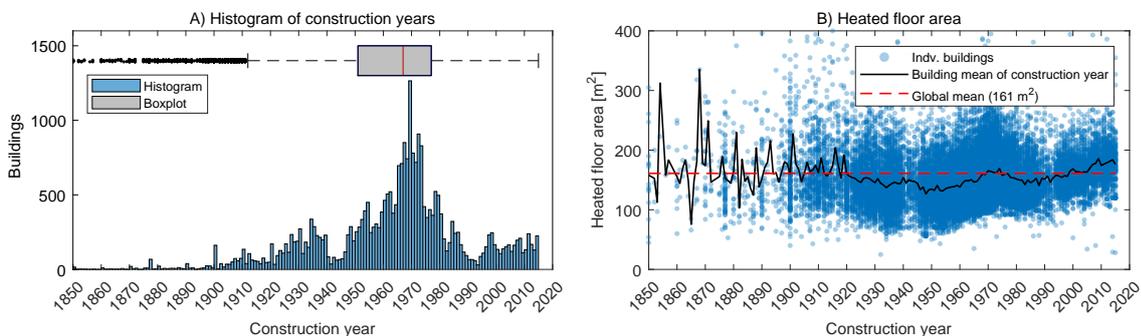


Figure 2.2. Descriptive statistics of approx. 27,800 SFHs constructed from 1850 to 2015 in the municipality of Aarhus, Denmark. A) depicts the frequency histogram and distribution boxplot of the construction years. B) depicts a scatter of heated floor area vs. construction year with the mean of each year and global mean of the entire stock overlain. Buildings with a floor area above 400 m² are not shown. Data source: Danish Building and Dwelling Register (BDR).

2.2 Aarhus district heating system

Approx. 95% of all residents in the municipality of Aarhus are supplied with district heating (DH) for the purpose of space heating and on-site domestic hot water (DHW) preparation from the municipal district heating utility, AffaldVarme Aarhus (AVA). This makes the Aarhus district heating system the second largest in Denmark. The heat has traditionally been produced based on fossil fuels, but it has been produced (nearly) fossil free since 2016, where the main plant was converted into burning CO₂-neutral wood pellets (biomass) instead of coal. Today, the majority of the heat is produced from waste incineration and biomass (primarily wood pellets, straw, wood chips and biogas) with supplementation of electricity (electric boiler) and excess heat from industry. However, a few oil boilers are still available and used for peak load situations ([AffaldVarme Aarhus, 2018](#)).

District heating is delivered to the costumers through a large network of underground pipes, heat exchangers, and booster stations from distributed productions plants around the municipality (Figure 2.3). Hot pressurized water is delivered from the production plants to a large-diameter transmission grid with typical supply conditions around 105 °C and return conditions around 45 °C. These transmission lines transport heat through more than 136 km pipe across the municipality to be exchanged to a lower energy level (approx. 75 °C supply and 45 °C return) in about 50 heat exchangers located in strategic places around the municipality. From here, a distribution grid of more than 2000 km of piping in local networks delivers the heat to the customers ([Dahl, 2018](#)).

2.2.1 Heating data

The production, supply and billing of district heating to the consumers in Aarhus is handled by AVA, but the consumption data itself are private and owned by the consumers. Access to consumption data is therefore restricted in accordance with the EU General Data Protection Regulation (GDPR) directive. End-use DH energy consumption has traditionally been metered by AVA with energy meters read off manually once per year. However, since 2013, AVA has begun the installation of digital smart meters with hourly readings in all consumer installations in its network – a process which was finalized around 2016. For the purpose of this research, access was initially granted to the annually aggregated energy use data of individual buildings, and later to the hourly time series.

In Figure 2.4A, the mean DH energy use of three years of annual metered data (2013, 2014 and 2015) is shown for all SFHs in Aarhus as function of their construction year. In Figure 2.4B, the area-normalized energy use – the energy use intensity – is shown. Both scatters are overlain with the mean consumption of individual construction years and the overall global mean of all SFHs across construction years. The annual DH energy use of SFHs from 1850 to the 1970s is characterized by a fairly constant mean consumption of approx. 20 MWh/year, but with significant variation within individual vintages (Figure 2.4A). However, it is evident that buildings constructed from around the late 1970s and onwards are more energy efficient as they consume significantly less heat – a trend which can likely be attributed to the energy crises of the 1970s, leading to the considerable tightening of the requirements for heat insulation in the Danish Building Regulations of January 1977. In addition to the overall drop in consumption after the 1970s, the data also suggests that the increased focus on energy efficiency also reduced the variability between individual buildings.

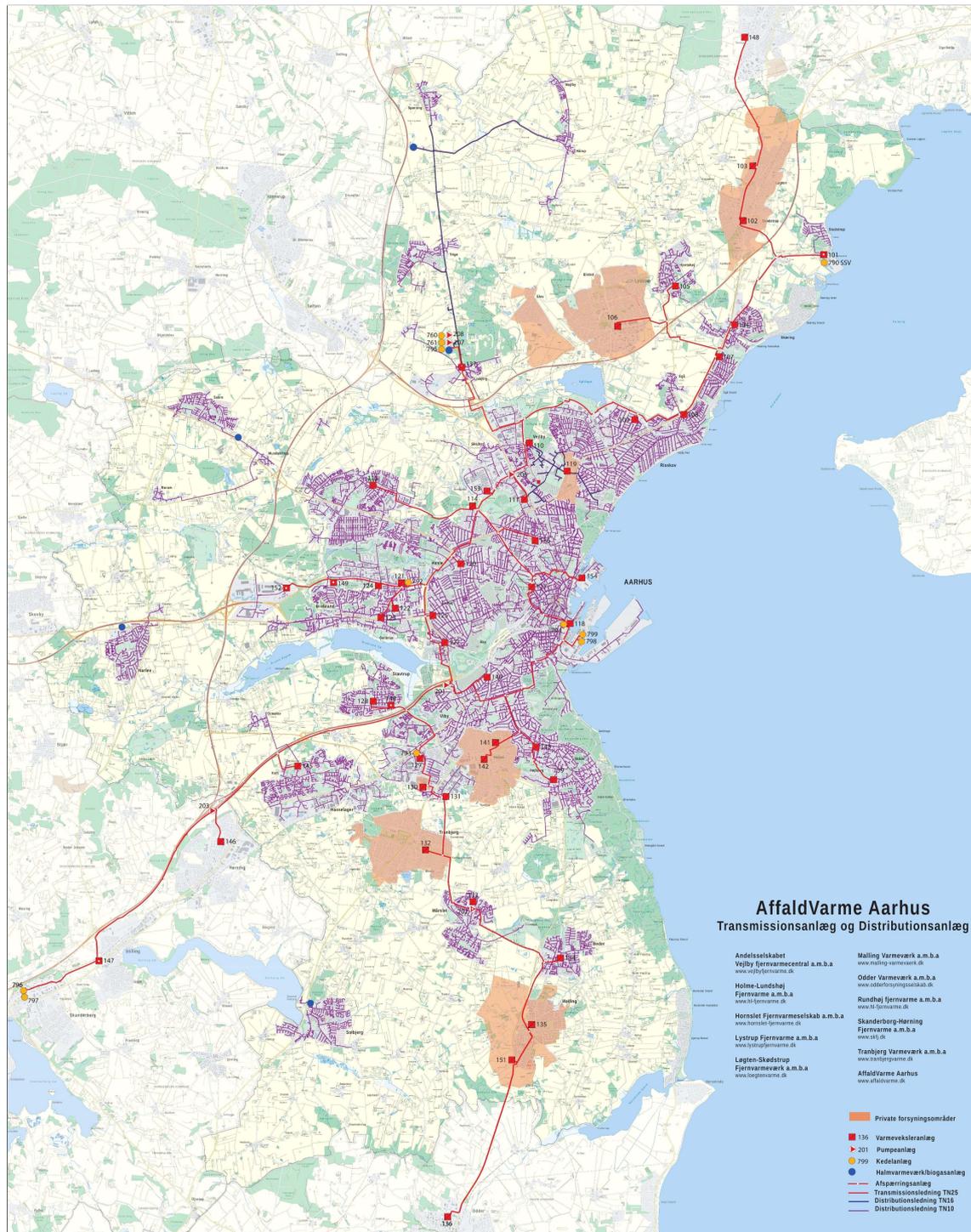


Figure 2.3. Aarhus district heating system. Yellow dots denote boiler plants and blue dots denote plants using biomass. Red boxes denote heat exchangers. Red lines denote transmission grid. Purple lines denote distribution grid. Source: AffaldVarme Aarhus (AVA).

By taking the size of the buildings into account, we can obtain the energy use intensity as shown in Figure 2.4B, which can be interpreted as a measure of energy efficiency. The trend is more or less the same as with the absolute DH energy use; however, due to the smaller houses constructed from 1920-1970, which have a relatively larger surface

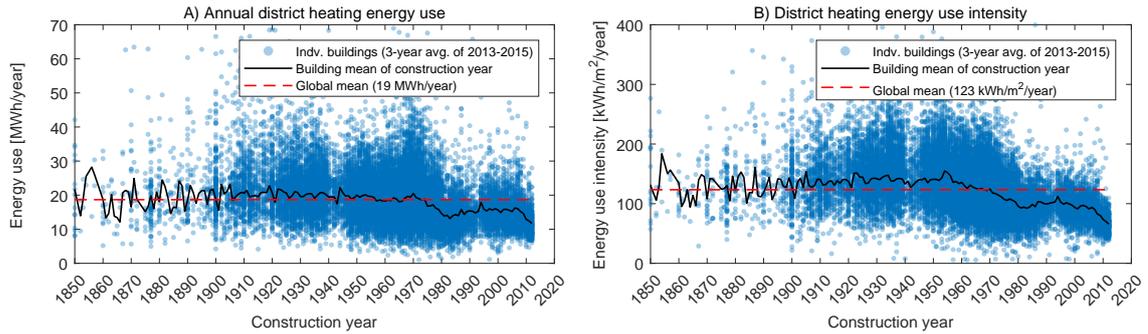


Figure 2.4. District heating energy use of approx. 27,800 SFHs constructed from 1850 to 2015 in the municipality of Aarhus, Denmark. A) depicts a scatter of the 3-year mean DH energy use of buildings vs. construction year with the mean of each year and the global mean of the entire stock overlain. Buildings with an energy use above 70 MWh/year are not shown. B) depicts the same data normalized by the heated floor area of the buildings. Buildings with an energy use intensity above 400 kWh/m²/year are not shown. Source: Danish Building and Dwelling Register (BDR) and AffaldVarme Aarhus (AVA).

area, the energy-efficiency of these buildings is smaller than average of the building stock (Figure 2.2B). Danish building regulations have, since 1985, placed requirements on the energy use intensity of buildings and this metric is currently the most demanding requirement to fulfil for newly built houses.

Chapter 3

Statistical analysis of annual building energy use

3.1 Motivation

Engineers are faced with steadily increasing amounts of sensor data from building management systems, building and property data from public databases, utility data, and other types of logged data that could contain information about building energy use. This *big data* is often a chaotic mix of information of varying quality and quantity, and composed of data on different aggregation levels, e.g. system level, building level, zip code level, or even county or national level data, which makes it difficult to apply it for traditional physics-based modelling. Information processing, data mining, feature selection, clustering and other techniques in the field of machine learning are therefore becoming increasingly popular for UBEM applications in these years (Table 1.1). With the increasing amount of information and data being produced in our buildings and in society in general, it is expected that such data-driven methods are going to play an important role in future UBEM.

This understanding of the capabilities of data-driven inference techniques made me curious to know, whether a statistical treatment of the features contained in the case data applied in this PhD project – the BDR database and annually metered building energy use – could be used to train a UBEM capable of predicting building energy use of entire urban areas.

The following article (journal paper [P1]), published in Elsevier journal *Energy and Buildings*, explores how features from the Danish BDR database can be used as explanatory variables in estimating annual heating energy use in SFHs. The study is based on a bottom-up statistical UBEM formulation using Bayesian multi-level log-normal regression analysis on three years of annual data for a sub-sample of 10,000 randomly selected SFHs from the Aarhus dataset.



Contents lists available at ScienceDirect

Energy & Buildings

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

Predicting Danish residential heating energy use from publicly available building characteristics

Martin Heine Kristensen^{a,*}, Adam Brun^b, Steffen Petersen^a

^a Department of Engineering, Aarhus University, Aarhus C DK-8000, Denmark

^b AffaldVarme Aarhus, Teknik og Miljø, Aarhus Kommune, Bautavej 1, Aarhus V DK-8210, Denmark



ARTICLE INFO

Article history:

Received 9 January 2018

Revised 7 March 2018

Accepted 5 May 2018

Available online 9 May 2018

Keywords:

Residential building stock
Building and Dwelling Register
Urban building energy modeling
Hierarchical modeling
Multiple linear regression
District heating data

ABSTRACT

Urban building energy modeling (UBEM) is a valuable tool for analyzing the building stock. Many different model approaches have been proposed in recent years suggesting various ways of dealing with the challenges of UBEM; however, central for all modeling approaches is the need for informative input data about the building stock. The availability of data for urban-scale modeling is both country-specific, time-consuming to aggregate, and data access is often limited due to privacy constraints. In this paper, we present a hierarchical bottom-up model of the Danish residential building stock using public building data for predicting the annual heating energy consumption. For more than 10,000 randomly selected single-family dwellings, the annual energy consumption is modeled and validated for the city of Aarhus, Denmark. We found that approx. 50% of the energy use is explained using only four widely available building characteristics, which enables building-scale predictions with a mean absolute error of approx. 25%. In addition, for city-scale predictions, the regression-based model enables aggregated predictions with a mean bias error of less than $\pm 2\%$. Even though building-scale predictions are only somewhat accurate, the performance remains comparable to state-of-the-art high-fidelity models in the literature.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The increasing need for energy efficient building operation leads to a need for tools that enable city governments, energy policy-makers and supply companies to plan and predict the effect of energy efficiency strategies on different urban scales. The emerging research field of urban building energy modeling (UBEM) seeks to facilitate such analysis by generating models based on aggregated effects of individual building energy models (BEM) or archetype models to the scale of neighborhoods, cities or ultimately a complete building stock. As such, one can use UBEM to establish an overview of the current state and environmental footprint of an existing urban area, and/or to estimate the effects of investments in energy conserving measures (ECM). The challenge of UBEM, however, remains to collect and assign the necessary data inputs to model the urban energy use reliably. These input data are often compiled from a mixture of different sources of information with the availability of data being very much project and country-specific. In addition, data access is often limited due to legal constraints and privacy considerations making data acquisition a delicate compromise. One can thus seldom expect to have

data that is elaborate enough to obtain a purely deterministic description of the building physics. Instead, modelers may have to resort to 'guessing' either by means of educated guesses, some kind of analysis of historical data or the like, or through a stochastic treatment of uncertain data. Consequently, statistical modeling approaches are currently a popular choice for many modelers – either in a hybrid form that includes a building-physical description or as a sheer data-driven approach.

1.1. Previous work

Urban building energy modeling is a growing research field with new modeling approaches being proposed regularly for various analytical purposes. Several review articles have in recent years summarized and defined the concept of UBEM; however, the prevailing UBEM approaches all offspring from a paradigm of two opposite modeling approaches; top-down or bottom-up, as suggested by Swan and Ugursal [1] and Kavgić et al. [2]. The top-down approach applies a data-driven statistical approach for modeling energy consumption while the bottom-up applies either statistics or building physics to setup the relationship between parameters. For the remaining part of this paper, we focus only on the bottom-up approach. Even though Reinhart and Cerezo [3] in a recent review of urban-scale modeling limit UBEM to encompass only techniques categorized as building-

* Corresponding author.

E-mail address: mhk@eng.au.dk (M.H. Kristensen).

<https://doi.org/10.1016/j.enbuild.2018.05.011>

0378-7788/© 2018 Elsevier B.V. All rights reserved.

Nomenclature	
<i>Variables</i>	
$y_{b,i}$	Metered building energy use for heating for building b , observation i (data) [MWh/year]
\hat{y}_b	Sample mean of metered building energy use for heating for building b (data) [MWh/year]
\bar{y}	Sample mean of metered building energy use for heating for all buildings (data) [MWh/year]
$\hat{\sigma}_b$	Sample standard deviation of metered building energy use for heating (data) [MWh/year]
x	Explanatory variables (data)
μ_b	Expected building energy use for heating for building b (fitted) [MWh/year]
σ_b	Standard deviation of building energy use for heating for building b (fitted) [MWh/year]
γ	Scale parameter of half-Cauchy distribution for standard deviations (fitted) [MWh/year]
τ	Standard deviation of regression level error (fitted) [MWh/year]
β	Regression coefficients (fitted)
N	Number of buildings in sample
n_b	Number of observations for building b in sample (between 2 and 3 observations per building)
<i>Indices</i>	
b	Indexing buildings in sample from 1 to N
i	Indexing observations in building b from 1 to n_b
$train$	Indexing training sample
$valid$	Indexing validation sample



physics, current literature holds suggestions for both statistical and physics-based bottom-up approaches for various UBEM purposes (Table 1).

For the majority of the bottom-up UBEMs in current literature, advanced modeling and a wealth of data inputs is applied to reconstruct the energy use of an urban area. In many cases, however, these high-fidelity models fail to validate the energy performance sufficiently reliably. Only five out of 15 recently published papers treating bottom-up UBEM (Table 1) report an error less than 10% on the aggregated urban level, and only a few of the papers even validate the performance on the disaggregated level of individual buildings. In fact, some papers do not even validate the performance of the proposed UBEM at all. For some studies, this might be because predictive performance is not the scope of the analysis, or because of insufficient validation data. There is no evidence to suggest that the more complex models relying on a building-physical description yields better predictive performance than the simpler data-driven statistical models.

1.2. Contribution

We present in this paper a simple yet efficient approach to modeling the energy consumption of buildings in an urban setting. In particular, we model the annual building energy use of Danish single-family dwellings using a hierarchical multiple log-linear regression model on data from the publicly available Danish Building and Dwelling Register (BDR), and metered annual district heating energy use of a large sample of the Danish residential build-

Table 1
Modeling approach, data sources and the reported performance for 15 bottom-up UBEM studies.

Modeling approach	Tools/techniques	Scale of application ^a	Data sources ^b	Performance data used for calibration	Performance validation				Ref.
					Dataset used for validation	Data points for validation	Scale of error assessment	Error on aggregated scale	
Building-physics	EnergyPlus	City (Boston)	A; B; C; E; I	–	Validation data	23 zip codes	Zip code	40%	[4]
Building-physics	EnergyPlus; Radiance	Neighborhood (3 cities, USA)	E; I	–	Statistical data	121 buildings	Neighborhood	<15%	[5]
Building-physics	EnergyPlus	Neighborhood (Cambridge)	A; B; C; E; F; G	+	Validation data	2263 buildings	Neighborhood	44%–47%	[6]
Building-physics	EnergyPlus	City (Milan)	B; E; F	–	Statistical data	N/A	City	4%	[7]
Building-physics	EnergyPlus	City (Kaohsiung, Tainan)	B; F; I	–	Validation data	683 grid areas	Grid area	49%	[8]
Building-physics	Dynamic reduced order model	Neighborhood (Gothenburg)	B; D; E; F	–	Validation data	433 buildings	Neighborhood	3%	[9]
Building-physics	Dynamic reduced order model	Neighborhood (Germany)	B; F; H; I	–	Validation data	200 buildings	Neighborhood	6%	[10]
Building-physics	Unknown software	City (Osaka)	E; F; I;	–	Statistical data	1,128,000 buildings	City	18%	[11]
Building-physics	eQuest	City (Houston)	B; E; F; I	+	Statistical data	N/A	City	10%–13%	[12]
Statistical	Regression; RF; SVM	City (New York)	C; F	+	Training data	176 zip codes	Zip code	N/A	[13]
Statistical	Regression	City (New York)	C; F	+	N/A	N/A	N/A	N/A	[14]
Statistical	Regression	City (Salford)	B; H	+	Training data	10 city districts	City	4%	[15]
Statistical	Regression	City (Rotterdam)	B; E; F	+	Training data	993 zip codes	Zip code	9%–13%	[16]
Statistical	Regression; EnergyPlus	City district (Castellón de la Plana)	N/A	+(simulated)	N/A	N/A	N/A	N/A	[17]
Statistical	Regression	City (Aarhus)	C	+	Validation data	10,000 buildings	City; Building	<2%; 25%	This study

Notes:

^a Application scales applied: “Building”, “Neighborhood”, “Zip code”, “City district”, or “City”.

^b Data originates from (A): weather stations; (B): GIS maps and/or 2.5D/ 3D city models; (C): property and tax registers; (D): EPC databases; (E): building codes and/or standards; (F): statistical and/or historical databases; (G): utility companies; (H): project specific observations; (I): user definitions/ unknown source.

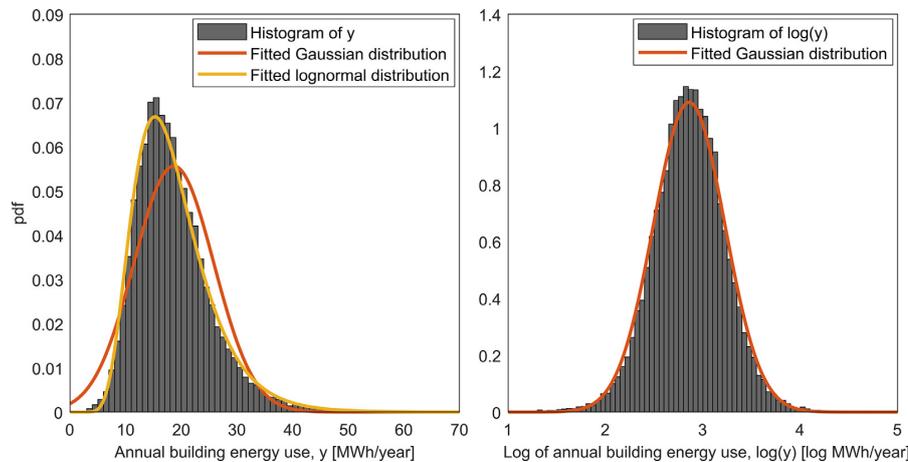


Fig. 1. Histogram of annual building energy use (sample mean of three years) for 27,000 buildings on the original scale (left) and logarithmic scale (right).

ing stock. Even though a statistical UBEM like this does not provide the same transparency and level of detail as models based on a sheer building physics approach, it greatly benefits from its simplicity. We demonstrate that even a few explanatory building characteristics from a database like the BDR can be used for accurately predicting annual building energy use on an aggregated urban level, and to a wide extent on the disaggregated level of individual buildings. Such models are of great value for public authorities to estimate the current environmental footprint of urban areas or even the entire building stock using only widely available and simple input data.

2. Method

2.1. Data

The annual metered district heating energy use for hydronic space heating and on-site domestic hot water (DHW) preparation for approx. 27,000 detached single-family dwellings in Aarhus, Denmark, was collected from the local district heating energy supplier, AffaldVarme Aarhus. For each building, the annual energy use was collected for three years (2013–2015) yielding a total of approx. 81,000 data point. The data did not allow any distinction between how much energy was spent on space heating and DHW, respectively. Fig. 1 displays histograms of the sample mean annual energy use of the 27,000 buildings on the original scale and the logarithmic scale. Due to the strictly positive nature of the consumption data, a lognormal distribution fits reasonably well (Shapiro–Wilk test of normality on the logarithm of the data, $p=0.012$).

Four different building characteristics were collected from the Danish Building and Dwelling Register (BDR) – a publicly available national property database¹ that is primarily used for building census, public administration, and property valuation and taxation – to be assessed as potential predictors of annual building energy use (Fig. 2). They were: 1) the construction year of the building, 2) the heated floor area, 3) the refurbishment status, and 4) the presence of supplementary heating installations in addition to district

heating, e.g. heat pumps, electric radiators, fireplaces or wood-burning stoves. The predictor ‘Refurbishment status’ comprise an unknown degree of refurbishment or retrofitting of the house. We assume in this study that the ‘Refurbishment status’ in average reflects a combination of energy upgrades and an expansion of the heated floor area. It is noted that it is the responsibility of the individual house owners to keep information in the BDR database up to date, which is why we also expect that many refurbishments or retrofits are unregistered; this could potentially disturb the inference.

For ‘Supplementary heating installations’, the predominant level is ‘Stove/Fireplace’ (15.6% of data), besides of course ‘No supplementary heating installation’ (83.5% of data). The three other levels only make up a negligible part (approx. 1% of data). Thus, for the remaining part of the paper, all supplementary heating installations are therefore collapsed, effectually making the categorical variable binary (Suppl. heating installations: Yes/No).

The correlation between the four selected building characteristics and the logarithm of the annual building energy use for each building is shown in Fig. 3. The heated floor area is treated on the logarithmic scale, as it is a strictly positive quantity. For all four predictors, a least squares regression line is overlain to highlight the linear relationship with the log energy use. This potential linear relationship on the logarithmic scale leads us to expect the predictors to have multiplicative effects in an exponential model on the original scale. In addition, one might expect each of the four building characteristics to engage in interaction effects with the others; for example, as the need of refurbishments and energy upgrades depends on the age of the building, we expect the refurbishment status to interact with both the construction year, the heated floor area, and potentially the presence of supplementary heating installations. No multicollinearity is present between the four predictors (variance inflation factors < 1.2).

The combined dataset was initially cleaned for obvious outliers and wrong data points; mainly faulty entries in the BDR database such as negative building areas. In addition, outliers were removed due to energy use data being either unreasonable low (below 10 kWh/m²) or high (above 500 kWh/m²). Moreover, if the energy use of one year deviated more than 100% from that of the other two years of the same building, it was also considered an outlier and removed from the dataset. Approximately 2% of the data was initially discarded as outliers and thus omitted from the analysis.

¹ Only manual one-building-at-a-time look-ups are allowed using the publicly available version of the BDR (link: <https://www.ois.dk/>). A non-public version of the database was at our disposal for the purpose of this study.

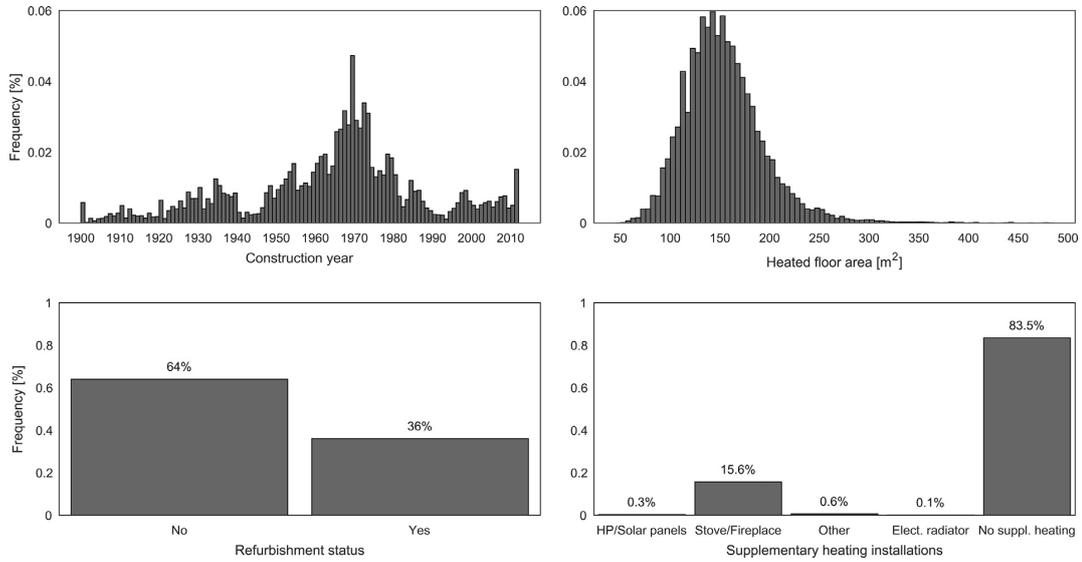


Fig. 2. Histograms of four building characteristics from the Danish BDR database (27,000 buildings).

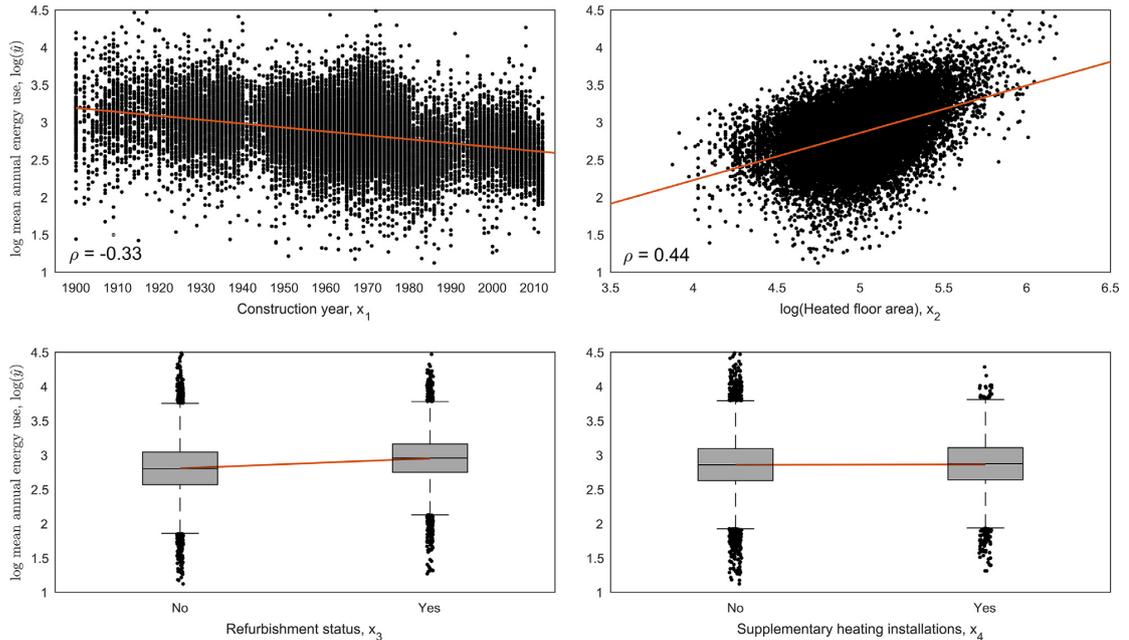


Fig. 3. Relationship between logarithm of annual building energy use (sample mean of three years) and four building characteristics from the Danish BDR database (27,000 buildings). A least squares regression line is overlain each plot to highlight linearity. The linear correlation coefficient is given for the two continuous variables.

2.2. Hierarchical regression model

A hierarchical multiple log-linear regression model with fixed and random effects was fitted to the data using Eqs. (1)–(3). The logarithm of the i th observation ($i = 1, 2, \dots, n_b$) of annual building energy use from the b th building ($b = 1, 2, \dots, N$) was modeled according to Eq. (1) as an independent and identical distributed Gaussian random sample $\log(y_{b,i})$ with unknown true mean annual building energy use $\mu_b \in \mathbb{R}$, and unknown standard deviation

$$\sigma_b \in \mathbb{R} | \sigma \geq 0.$$

$$\log(y_{b,i}) \sim N(\mu_b, \sigma_b^2). \quad (1)$$

The building-specific standard deviation, σ_b , accounts for within-building variation, i.e. variability between observations from the same building. As it is difficult to estimate the standard deviation for small sample sizes (a maximum of three observations per building is available), the standard deviations from all buildings were hierarchically constrained to follow a common half-Cauchy distribution according to Eq. (2) with scale parame-

Table 2
 Proposed regression models.

Model	Description	Number of regression terms, p
1	Main effects only	5
2	Main and pairwise interaction effects	11

ter $\gamma \in \mathbb{R} | \gamma > 0$ as recommended by Gelman [18]. The half-Cauchy distribution is considered a weakly informative prior distribution that to some degree helps draw σ_b towards zero when the data likelihood is weak. This attractive feature reduces the possibility of the standard deviations to overshoot for the small sample sizes.

$$\sigma_b \sim \text{HalfCauchy}(0, \gamma). \quad (2)$$

The hierarchical modeling introduced with Eq. (2) further allows buildings with less information embedded in their observations to borrow strength through this shared prior distribution, hereby constituting a compromise between two conflicting models: a ‘no-pooling’ model and a ‘complete pooling’ model. The no-pooling model is the limiting case that occurs when $\gamma \rightarrow \infty$ (the half-Cauchy distribution becomes a uniform distribution from zero to infinity), meaning that separate estimates of σ_b are given each building, i.e. $\text{Var}(\log(y_{i,b})) = \sigma_b^2$. The opposite case – the complete pooling model – occurs when $\gamma \rightarrow 0$ meaning that no information is available to distinguish the within-building variation of individual buildings from each other and thus a single shared estimate of σ is given all buildings, i.e. $\text{var}(\log(y_{i,b})) = \sigma^2 \forall b$. In the hierarchical model, Eq. (2), γ is estimated from the data hence constituting a partially pooled meta-model compromise between the no-pooling and complete pooling cases.

The true mean annual energy use μ_b of the b th building was subsequently modeled as a multiple linear regression on the four predictors of the Danish BDR database (Fig. 3) with independent and identical distributed Gaussian noise to allow for buildings with identical BDR information to have different energy use and thus account for residual error:

$$\mu_b \sim N(\boldsymbol{\beta} \cdot \mathbf{x}_b, \tau^2) \quad (3)$$

In this linear regression, $\boldsymbol{\beta}$ is the p -dimensional vector of regression coefficients to be inferred, and \mathbf{x}_b is the p -dimensional design vector of building b with the independent variables themselves. Two different models were proposed (Table 2): 1) a model with only the linear first order terms of the four predictors (main

effects) in addition to an intercept (constant), and 2) a model with main effects and pairwise interaction effects between the predictors (pairwise products of main effects) in addition to the intercept. The two models have five and 11 independent variables, respectively.

The regression error of Eq. (3) – represented by the standard deviation $\tau \in \mathbb{R} | \tau \geq 0$ – accounts for residual variability between the mean annual building energy use μ_b of ‘identical’ buildings, i.e. between-building variability that cannot be captured by the linear model of the four building characteristics of the BDR database. Given a perfect fit to data, this regression-level variability will approximate zero indicating that annual building energy use is completely predictable from the four investigated predictors, which obviously is not the case.

2.3. Training and validation scheme

Standardization of the predictors using $x' = (x - \mu_x) / \sigma_x$ was initially applied to improve convergence speed of the regression problem. This standardization makes the values of each predictor have zero-mean and unit-variance.

A subsample of $N_{\text{train}} = 1000$ randomly selected buildings (approx. 4% of the buildings in the city) was used for training the model. Using residual diagnostics, i.e. Cook’s distance and various plots of the studentized residuals, a few highly influential data points were detected within the training sample. A Cook’s distance higher than four times the mean was found for approx. 1% of the training data. These data points were deemed outliers and removed from the model before it was refitted. None of the remaining data points had a Cook’s distance higher than one.

An overall citywide validation of the model using a holdout subsample of $N_{\text{valid}} = 10,000$ randomly selected buildings (approx. 40% of the buildings in the city) was performed to test the generalizability of the results within the city. In addition to this general validation, the model was furthermore tested on a neighborhood-scale test area consisting of $N_{\text{test}} = 547$ dwellings (Fig. 4), all located in a well-defined urban area (approx. 2% of the buildings in the city). Using a validation scheme with different urban aggregation levels allows for the quantification of effects that links to the urban context. None of the validation data was used for training the model.

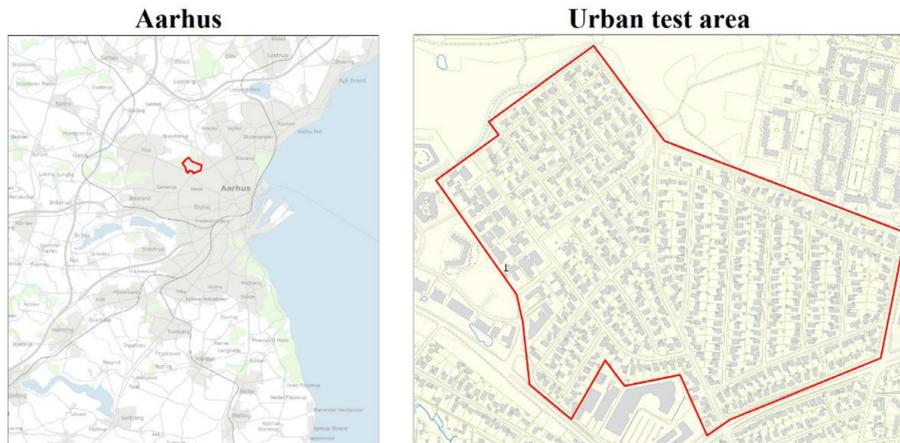


Fig. 4. Map of Aarhus, Denmark (left) and neighborhood test area (right). The test area contains $N_{\text{test}} = 547$ dwellings.

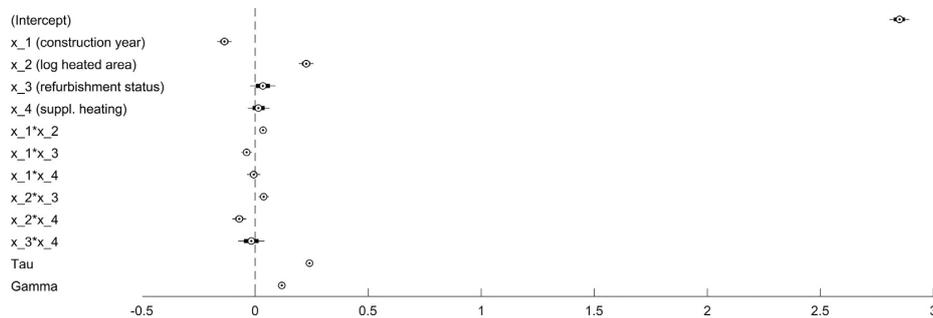


Fig. 5. Estimated standardized parameters of Model 2 on the logarithmic scale using a total of 2983 training observations of annual building energy use from 1000 buildings. Point estimates and error bars show posterior medians (circle), 50% intervals (thick bar), and 95% intervals (thin line). Parameters including zero are not necessarily statistically insignificant in a Bayesian framework.

2.4. Bayesian inference

A fully Bayesian inference using Hamiltonian Markov Chain Monte Carlo sampling was conducted using STAN – a statistical programming platform – with an interface for MATLAB (MatlabStan). All free parameters were given non-informative uniform prior distributions, i.e. $p(\tau, \gamma, \beta) \propto 1$ to allow for a data-driven inference. For both models (Table 2), six chains were run in parallel each consisting of 5000 samples with a 1000-sample warm-up period to draw posterior samples of the model parameters. Convergence was assessed using the potential scale reduction factor [19] for which, all parameters were below $\hat{R} < 1.1$ for both models and hence deemed converged.

2.5. Measures for model selection

Model selection was based on the ability of the two models in Table 2 to fit the data by means of their expected predictive accuracy. For this purpose, we used the Watanabe-Akaike information criterion (WAIC), a fully Bayesian measure of model fit [20]. Compared to non-Bayesian measures like AIC [21] and BIC [22], and the somewhat Bayesian DIC [23], WAIC has the desirable property of averaging over the entire posterior distribution rather than conditioning on point estimates, making WAIC a fully Bayesian approach for model selection.

In addition to the WAIC measure for assessing the likelihood of accurately fitting the data in the entire hierarchical structure, the regression-level predictive performance (prediction of μ_b) was assessed using more traditional deterministic goodness-of-fit measures. Assuming the sample means of annual building energy use \hat{y}_b to represent the true data means, and by using the posterior median estimate of μ_b as point estimate for the regression output, the proportion of explained variance by the regression model was assessed using the coefficient of determination (R^2). The overall accuracy of the median posterior estimates of μ_b was evaluated using the mean absolute percentage error (MAPE) yielding the expected level of error for single-building predictions:

$$\text{MAPE} = \frac{\sum_{b=1}^N |\hat{y}_b - \mu_b|}{N} \times 100. \quad (4)$$

The concept of accuracy consists of two elements in conjunction: bias (trueness) and precision (variance). Model bias is the closeness of agreement between the average model prediction and the average data \bar{y} . The bias was evaluated as the normalized mean bias error (NMBE) in compliance with ASHRAE Guideline 14–2014 [24]:

$$\text{NMBE} = \frac{\sum_{b=1}^N (\hat{y}_b - \mu_b)}{N} / \bar{y} \times 100, \quad (5)$$

which hereby expresses the expected percentage error for large-sample aggregated predictions of μ_b , e.g. that of an urban area. As the NMBE measure is prone to the cancellation effect, it ought to approximate zero (unbiased predictions). If model predictions are unbiased, but to some extent still inaccurate on the individual building level, it is due to a lack of model precision as precision is the closeness of agreement between independent model predictions of μ_b under the same conditions. Model precision was evaluated as the coefficient of variation of the root mean squared error (CVRMSE) in compliance with ASHRAE Guideline 14–2014 [24]:

$$\text{CVRMSE} = \sqrt{\frac{\sum_{b=1}^N (\hat{y}_b - \mu_b)^2}{N}} / \bar{y} \times 100, \quad (6)$$

which is the root mean squared error (RMSE) normalized by the mean of the data, hereby becoming a measure of percentage variability in predictions of mean annual building energy use with identical BDR information.

3. Results

The WAIC score for the two models in Table 2 were almost identical; -2837 (Model 1) and -2850 (Model 2), respectively. The smaller score of Model 2 indicates that it is marginally better in terms of representing the data, and hence yields the best predictive performance. However, with a difference in explained variance (R^2) of only 1%, the interaction effects are not influential. One could thus regard Model 1 (main effects only) as the proper parsimonious model choice; however, Model 2 was selected for further analysis as we believe the interaction effects are important to ensure a trustworthy interpretation of the regression parameters.

3.1. Posterior model parameters

Estimated standardized parameters of Model 2 are shown in Fig. 5. The linear model is transformed to a multiplicative exponential model on the original arithmetic scale of the observations as the observations of annual building energy use are modeled on the logarithmic scale. Hence, the estimated effects of the individual building characteristics on the annual building energy use is not immediately clear from the estimated regression coefficients given in Fig. 5. In addition, due to the presence of interaction effects, and because the heated floor area (x_2) is log-transformed, the marginal effects are even less obvious. For the sake of clarity, we have therefore visualized the marginal effects of changes in the explanatory variables by plotting the expected annual heating energy use μ_b as function of the two continuous predictors, namely construction year and heated floor area (Fig. 6).

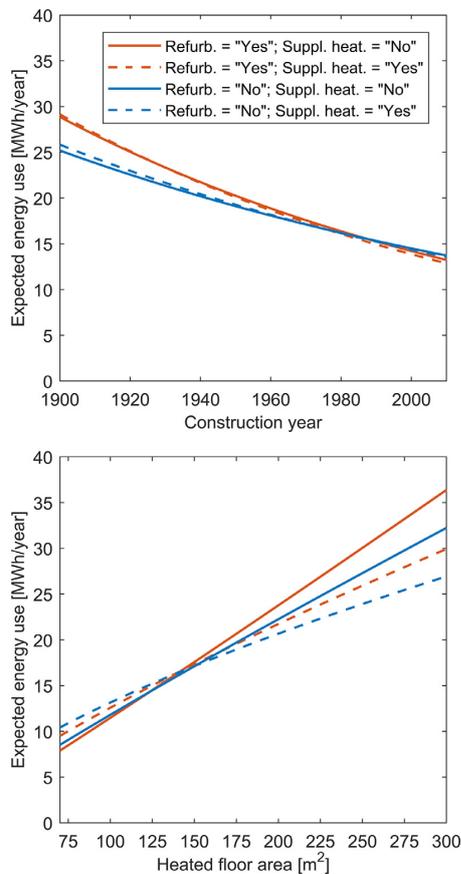


Fig. 6. Expected annual heating energy use μ as function of BDR predictors. Top: 150 m² house with varying construction year. Bottom: House constructed in 1970 with varying heated floor area. Only posterior median estimates are plotted. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

In the top plot, the expected annual energy use of a 150 m² house is shown as function of the construction year, while the energy use of a house erected in 1970 is shown as function of the heated floor area in the bottom plot. The interactive effects of the refurbishment status and the presence of supplementary heating installations are illustrated as the differences in the plotted graphs. The results imply that an old refurbished house (red curves in Fig. 6) in average consumes more energy than a non-refurbished house (blue curves in Fig. 6), but also that this effect varies with the construction year such that the difference diminishes for newer buildings. In addition, the size of the heated floor area seems to interact with the effect of a building being labeled as ‘refurbished’. We suspect the refurbishment of especially the older refurbished buildings to mainly constitute extensions to the original house and hence an extension of the heated floor area, which subsequently increases the energy use. This would explain why the effect of refurbishments is ‘negative’ for old and now large buildings and ‘positive’ for new and/or small buildings.

3.2. Predictive capabilities

The fitted multiple linear regression model can be used for predictions of the expected annual heating energy use μ_b using

Table 3

Goodness-of-fit measures for predicting expected annual heating energy use μ . The posterior predictive probability of μ is reduced to point estimates using the posterior medians. Sample expectations are the direct sample means of the observed data points.

Dataset	N (building data points)	MAPE	NMBE	CVRMSE	R ²
Training	1000	21%	-0.5%	25%	53%
Validation	10,000	25%	-1.5%	29%	49%
Urban test area	547	24%	0.3%	32%	35%

the four building characteristics from the Danish BDR database. Fig. 8(left) illustrates that a good fit was obtained for within-sample predictions (training sample) as posterior estimates of μ_b overlap the observed sample means \hat{y}_b . This was due to separate estimates of the true mean μ_b and standard deviation σ_b for each building. For out-of-sample predictions (validation sample), μ_b was not fitted on the observations of annual energy use y but inferred using the regression model. Using point estimates of μ_b (median posterior predictions), conventional goodness-of-fit measures was calculated for agreement with the sample expectations, i.e. the direct sample means of y for each building (Table 3). However, these measures – simple and familiar as they are – do not represent a fair evaluation of the performance of the regression model. The reason is that the regression model was trained on the ‘true’ means by accounting for the small sample size (2–3 samples of annual energy use per building) through σ_b and thus predicts the ‘true’ mean annual energy use. Consequently, the goodness-of-fit measures were very good for the training data. However, the goodness-of-fit measures for the validation data remains a pessimistic evaluation as we do not possess a true and unbiased estimate of the means of the validation data to test the regression against besides the direct sample means.

As a feature of the regression-based modeling, the mean bias error approximates zero (NMBE < 2%) meaning that the aggregated (summarized) energy consumption predicted for a larger area is very accurate in average. However, relatively large deviations emerge on the scale of individual building predictions (MAPE < 25%) due to the presence of unexplained variance ($R^2 \approx 50\%$) between individual buildings. Plots of the fit (on the original arithmetic scale) is shown in Fig. 7 with predicted expectations vs. sample expectations (top), a histogram of the residual error approximating that of a Gaussian distribution (middle), and a scatter of the residual error vs. data points confirming the desired presence of homoscedastic variance (bottom).

When applied to the urban test case, the model was not able to capture as much of the data variance ($R^2 \approx 35\%$) as in the more general setting with randomly selected buildings ($R^2 \approx 50\%$). This indicates that context-specific effects that were not modeled influenced the predictions. These context-specific effects, however, have no significant influence on the aggregated error (NMBE $\approx 0\%$), nor the error of predicting individual buildings (MAPE < 25%), as these measures remain the same as in the more general setting with randomly selected buildings from all over the city.

3.3. Effect of building operation

The total data variance has been partitioned into within-building variability, σ_b , using partial pooling through Eq. (2), and between-building variability, τ . For each building b , we have estimated the proportion of total data variance explained by the building itself as the intraclass correlation coefficient (ICC). The ICC measures the influence of phenomena occurring due to building

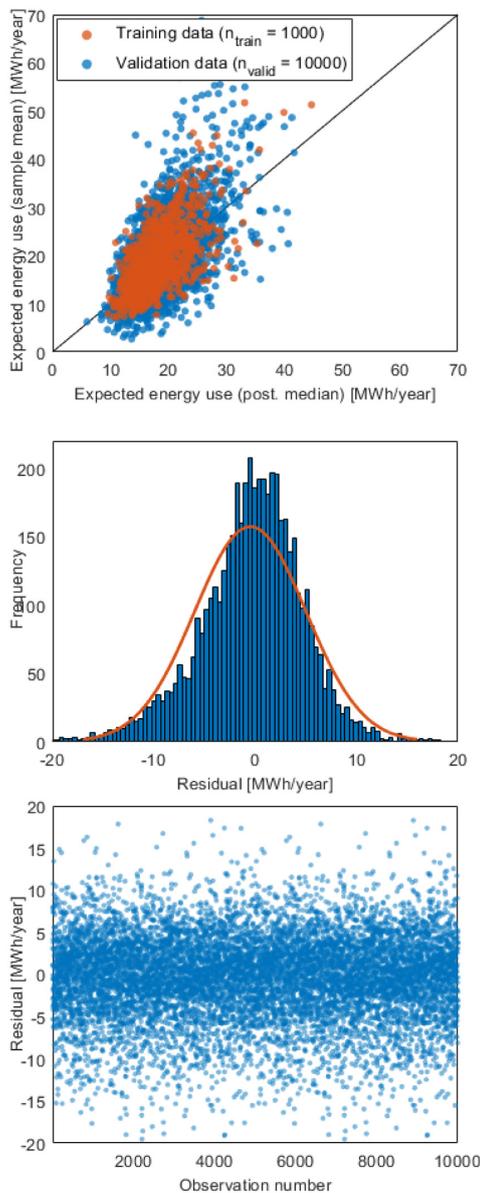


Fig. 7. Accuracy of regression fit. Top: Sample expectations vs. predicted expectations for training and validation samples. Middle: Histogram of validation sample residuals with fitted Gaussian distribution. Bottom: Scatter of validation sample residuals vs. data points.

operation in relation to the overall data variation.

$$ICC_b = \frac{\sigma_b^2}{\sigma_b^2 + \tau^2} \quad (7)$$

In Fig. 8 (middle plot), the sample standard deviations and the posterior estimates of σ_b are shown for the first 20 buildings in the training sample. With only three observations available for each building, the sample standard deviations $\hat{\sigma}_b$ are very uncertain and thus not reliable. Using the hierarchical pooling, uncertainty about the individual building sample standard deviations $\hat{\sigma}_b$ are jointly taken into account by borrowing strength from one an-

other to infer estimates of σ_b with reduced posterior spread. Still, however, large differences in the standard deviations persists between individual buildings indicating a large spread in the effect of how individual buildings are operated. The ICC of the first 20 buildings, i.e. the proportional effect of individual buildings, is shown in Fig. 8 (right plot). The average ICC coefficient across the building sample was 0.33–0.35 (95% highest posterior density) indicating that phenomena within individual buildings, e.g. occupant behavior and settings for building operation, in average explains approx. 34% of the total data variance. The remaining data variance – in average approx. 66% – is caused by uncertainty in fixing the expected annual building energy consumption μ_b ; that is, residual error in the linear regression.

With the estimated distribution of the scale parameter γ , we can draw posterior predictive samples of the within-building standard deviation $\sigma_b|\gamma$ for new unseen buildings (Fig. 9). This reflects the expected standard deviation of a random detached single-family dwelling in the city.

Although the inherent uncertainty of building operation obviously does not allow us to infer precise estimates of σ_b for new unseen buildings, the hierarchical pooling has ensured that it can be predicted with some degree of confidence.

4. Discussion

The predictive capabilities of many UBEM studies documented in the literature is low (see Table 1) mainly due to the quality of input and validation data for performance testing. Overall, their predictive performance appears unsatisfactory considering the high level of detail and high model complexity that characterize many of these UBEMS. In this study, we presented a much more parsimonious approach to the UBEM challenge using a simple regression-based framework on public building data. Even though the original objective of this study was to detect predictors of building energy use in the Danish Building and Dwelling Register, the resulting linear model – banal as it is – proves to be equally accurate compared to high fidelity models in terms of the aggregated predictive performance (city-scale) and to some degree the disaggregated performance (building-scale). We must emphasize though, that the case study-based nature of many UBEM studies make it difficult to carry out a consistent and unbiased comparison. Thus, we do not necessarily expect the regression model presented in this study to perform equally well under all circumstances. On the other hand, we find no arguments for not expecting similar predictive performance using building inventory databases of other countries.

4.1. Significant predictors of building energy use

The main contribution of this study lies in the documentation of how simple building characteristics from a public database like the Danish BDR contains significant information to be used for prediction of building energy use. We found that all four BDR parameters considered – construction year, heated floor area, refurbishment status and the presence of supplementary heating installations – contain information about annual building heating energy use. In total, 10 transformations of the four parameters – including both main effects and pairwise interaction effects – contribute in describing annual building energy use. This is a very common phenomenon of large datasets as even the smallest of effects becomes visible. This does not mean that those terms contribute a lot in explaining data variance, i.e. increasing the R^2 -value but only that their effects are indisputable. In total, the linear model explains approx. 53% of the training data variance, approx. 49% of the variance in the holdout validation sample, and only approx. 35% of the holdout neighborhood-scale test area, leaving a lot of information to be explained by other variables than those investigated. The lower

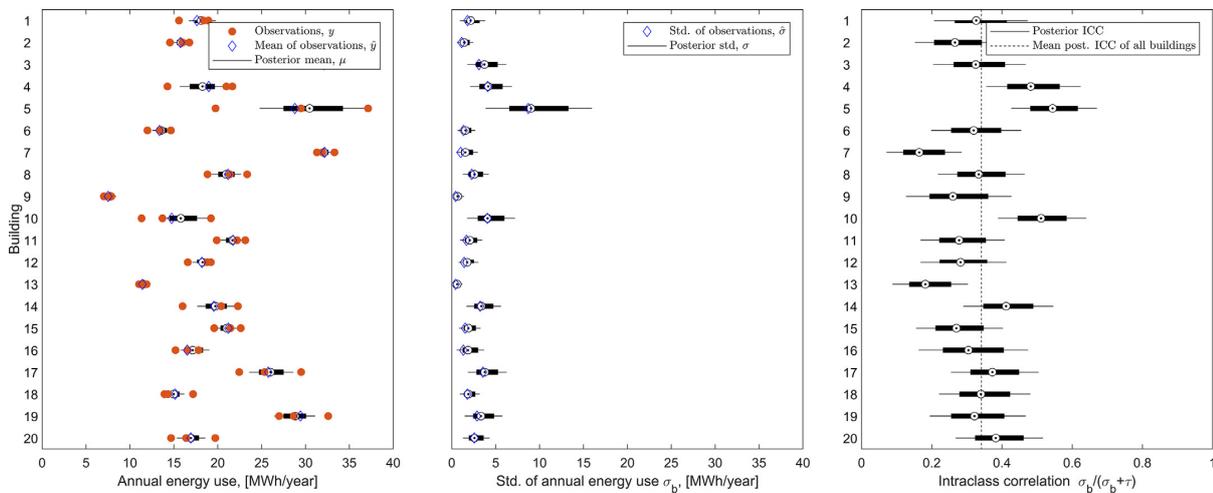


Fig. 8. Fit to training data for the first 20 buildings. Left: Observations, sample mean of observations and posterior estimates of true means. Middle: Sample standard deviation of observations and posterior estimates of standard deviations. Right: Posterior estimates of ICC and mean ICC estimate, approx. 0.34, of entire training sample.

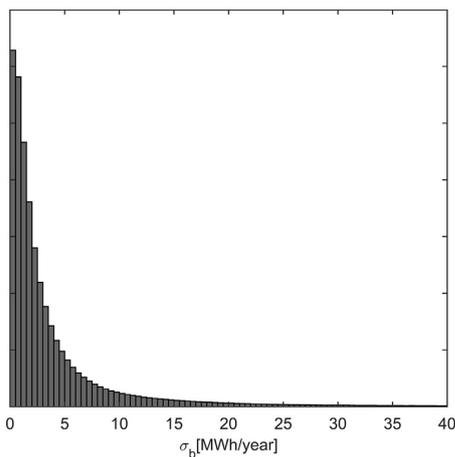


Fig. 9. Posterior predictive distribution of $\sigma_b|\gamma$.

R^2 -value found for the test area suggests the presence of context-specific effects that are averaged out on the larger aggregated scale as also found in previous studies [25]. Additional explanatory variables may thus be necessary to perform accurate predictions of individual buildings or smaller urban areas; however, if one is willing to accept a mean absolute error in single-building predictions of approx. 25%, the four widely available building characteristics investigated in this study is sufficiently informative for such purposes.

The BDR contains many other variables not considered in this study that might serve as additional predictors of building energy use. However, many of these variables are related to legislative and regulatory classifications, or additional non-heated buildings on the property. BDR data that might contain information about building energy use, but not included in this study, is utilized basement and attic areas. Although these features were not explicitly treated, heated basement and attic area is already included in the total heated floor area that we included.

Although the amount of data variance explained by the four building characteristics may seem low, our findings are in line with

previous research on the topic. In a thorough statistical analysis of 924 English households, Huebner et al. [26] found that 11 building variables in total, including floor area, dwelling type and dwelling age, contribute in explaining approx. 39% of the variability in heating energy use. In another analysis of 15,000 dwellings across the Netherlands, Santin et al. [27] found that the type of dwelling only accounts for approx. 4% of the variability in heating energy use, while another 38% is explained by building variables, including floor area and construction year. Accounting for the fact that these studies cover several different dwelling types and different heating systems (electrical and gas systems), we find no arguments for not trusting our findings.

4.2. Explaining residual variability

What could explain the remaining 50%–75% of the data variance? The role of occupants, i.e. socio-demographics, comfort preferences, behaviors and attitudes are often ascribed significant effect [26–29,25]; however, these phenomena are very complex and not easy to separate from one another although many have tried. In the study by Huebner et al. [26], they found that 43% of domestic heating energy use could be ascribed such occupant-related effects, while results by Santin et al. [27] show that occupants and behavior characteristics only account for 4% of heating energy use variation. Other studies have found values around 18%–20% [30,31].

Even though many obvious outliers were removed during the initial data cleaning process, we still suspect some degree of error to remain in the BDR data that might explain parts of the residual data variability. As an example, many buildings are listed as constructed in year 1900; however, it seems unreasonable that this specific year would have a larger rate of construction than e.g. year 1901. This phenomenon might be explained by a general practice of entering year 1900 whenever the actual year of construction is unknown. One should thus be careful trusting single data points from public databases like the BDR.

5. Conclusions

The annual heating energy use observed over three successive years for 27,000 Danish detached single-family dwellings was analyzed using hierarchical regression modeling on four building char-

acteristics from the Danish Building and Dwelling register. The analysis suggests that knowing these four specific building characteristics (construction year, heated floor area, refurbishment status, and presence of supplementary heating installations) will in average explain approx. 50% of the variability in the expected annual heating energy use of randomly selected buildings. For aggregated predictions on the scale of neighborhoods or cities, the proposed regression model can be used to approximate the true annual energy consumption (NMBE < 2%), but due to the presence of unexplained variance between individual buildings, single-building predictions remain somewhat imprecise (MAPE < 25%). This performance is comparable to current state-of-the art high fidelity UBEMs, underlining a need for further work on how to apply UBEM for more than just aggregated analysis. Thus, if the purpose of a given urban modeling task is to calculate the aggregated energy use on an urban scale – where the influence of occupant behavior is negligible – we recommend a simple regression-based approach like the one proposed in this study relying on much less input data compared to high fidelity physics-based models. However, we should emphasize that the statistical model is not able to assess the effect of various energy conserving measures in building retrofit scenarios.

Acknowledgments

The research was conducted as part of the 'Resource Efficient Cities Implementing Advanced Smart City Solutions' (READY) project, work package 3, financed by the 7th EU Framework Programme (FP7-Energy project reference: 609127). Furthermore, the authors would like to thank the district heating company in Aarhus, AffaldVarme Aarhus, for supplying the building energy data that forms the basis of the study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.enbuild.2018.05.011](https://doi.org/10.1016/j.enbuild.2018.05.011).

References

- [1] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, *Renewable Sustainable Energy Rev.* 13 (2009) 1819–1835.
- [2] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, *Build. Environ.* 45 (2010) 1683–1697.
- [3] C.F. Reinhart, C. Cerezo, Urban building energy modeling – A review of a nascent field, *Build. Environ.* 97 (2016) 196–202.
- [4] C.C. Davila, C.F. Reinhart, J.L. Bemis, Modeling Boston: A workflow for the efficient generation and maintenance of urban energy models from existing geospatial datasets, *Energy* 117 (2016) 237–250.
- [5] T. Dogan, C. Reinhart, Shoeboxer: an algorithm for abstracted rapid multi-zone urban building energy model generation and simulation, *Energy Build.* 140 (2017) 140–153.
- [6] J. Sokol, C.D. Cerezo, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, *Energy Build.* 134 (2017) 11–24.
- [7] P. Caputo, G. Costa, S. Ferrari, A supporting method for defining energy strategies in the building sector at urban scale, *Energy Policy* 55 (2013) 261–270.
- [8] F.-Y. Lin, T.-P. Lin, R.-L. Hwang, Using geospatial information and building energy simulation to construct urban residential energy use map with high resolution for Taiwan cities, *Energy Build.* 157 (2017) 166–175.
- [9] M. Österbring, É. Mata, L. Thuvander, M. Mangold, F. Johnsson, H. Wallbaum, A differentiated description of building-stocks for georeferenced bottom-up building-stock model, *Energy Build.* 120 (2016) 78–84.
- [10] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, D. Müller, TEASER: an open tool for urban energy modelling of building stocks, *J. Building Perform. Simul.* (2017) 1–15.
- [11] Y. Shimoda, T. Fujii, T. Morikawa, M. Mizuno, Residential end-use energy simulation at city scale, *Build. Environ.* 39 (2004) 959–967.
- [12] S. Heiple, D.J. Sailor, Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles, *Energy Build.* 40 (2008) 1426–1436.
- [13] C.E. Kontokosta, C. Tull, A data-driven predictive model of city-scale energy use in buildings, *Appl. Energy* 197 (2017) 303–317.
- [14] D. Hsu, How much information disclosure of building energy performance is necessary? *Energy Policy* 64 (2014) 263–272.
- [15] A.T. Booth, R. Choudhary, D.J. Spiegelhalter, A hierarchical Bayesian framework for calibrating micro-level models with macro-level data, *J. Building Perform. Simul.* 4 (6) (2013) 293–318.
- [16] A. Mastrucci, O. Baume, F. Stazi, U. Leopold, Estimating energy savings for the residential building stock of an entire city: a GIS-based statistical downscaling approach applied to Rotterdam, *Energy Build.* 75 (2014) 358–367.
- [17] M. Braulio-Gonzalo, P. Juan, M.D. Bovea, M.J. Ruá, Modelling energy efficiency performance of residential building stocks based on Bayesian statistical inference, *Environ. Modell. Softw.* 83 (2016) 198–211.
- [18] A. Gelman, Prior distributions for variance parameters in hierarchical models, *Bayesian Anal.* 3 (1) (2006) 515–534.
- [19] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, *Bayesian Data Analysis*, third ed., CRC Press, Boca Raton, 2014.
- [20] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, Cambridge University Press, Cambridge, UK, 2009.
- [21] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: *Proceedings of the Second International Symposium on Information Theory*, Budapest, 1973.
- [22] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* 6 (2) (1978) 461–464.
- [23] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, A. van der Linde, Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B* 64 (4) (2002) 583–639.
- [24] ASHRAE, ASHRAE Guideline 14-2014 – Measurement of Energy, Demand, and Water Savings, ASHRAE, 2014.
- [25] M.H. Kristensen, S. Petersen, Explaining variability in metered energy use for similar buildings using Bayesian inference, *Energy Procedia* 132 (2017) 897–902.
- [26] G.M. Huebner, I. Hamilton, D. Shipworth, T. Oreszczyn, Explaining domestic energy consumption – the comparative contribution of building factors, socio-demographics, behaviours and attitudes, *Appl. Energy* 159 (2015) 589–600.
- [27] O.G. Santin, L. Itard, H. Visscher, The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock, *Energy Build.* 41 (2009) 1223–1232.
- [28] W. Abrahamse, L. Steg, How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *J. Econ. Psychol.* 30 (2009) 711–720.
- [29] S. Yang, M. Shipworth, G. Huebner, His, hers or both's? The role of male and female's attitudes in explaining their home energy use behaviours, *Energy Build.* 96 (2015) 140–148.
- [30] R.C. Sonderegger, Movers and stayers: the resident's contribution to variation across houses in energy consumption for space heating, *Energy Build.* 3 (1) (1978) 313–324.
- [31] K. Steemers, G.Y. Yun, Household energy consumption: a study of the role of occupants, *Build. Res. Inf.* 5-6 (37) (2009) 625–637.

3.2 Epilogue

It is found that approx. 50% of the variability in annual heating energy use of SFHs across the city can be explained using four widely available building characteristics: construction year, heated floor area, refurbishment status (yes/no), and the presence of supplementary heating installations in addition to district heating (yes/no). However, the explanatory power of building characteristics falls to around 35% when applied in a specific urban context; a neighbourhood test area of very similar buildings. This suggests the presence of additional area-specific effects that were not accounted for. One of these effects – the effect of zip code – was later investigated in another study using a sample of 350 SFHs from the same construction year and located in 37 city districts in the municipality of Aarhus. The study is reported in conference paper [S4] that is appended to the thesis.

The statistical analysis on the BDR database presented in this chapter can be used for estimating the aggregated or averaged building energy use of large urban areas where features that make individual buildings differ from the group mean are levelled out (the cancellation effect). This can be very useful as an initial and rough assessment of the energy efficiency of an urban area, or larger part of the building stock. However, it lacks the physical meaningfulness that is necessary to evaluate ECMs and other retrofit interventions because it is founded on a purely statistical treatment of variables. Therefore, it cannot stand alone in accurately predicting residential building energy use; additional information and modelling work is needed for this purpose. Our focus shall therefore be on bottom-up physics-based modelling alone in the remaining part of the thesis to remedy these shortcomings.

Part 2

Physics-based archetype modelling

Chapter 4

A regression-based approach using annual data

4.1 Motivation

The current literature on archetype modelling and calibration only contains a few examples of studies in which stochastic methods are applied (Table 1.3). These studies address the diversity and heterogeneity of the underlying building stock by either: 1) average training data prior to calibrating a single archetype distribution (Booth et al., 2012), or 2) calibrate parameter estimates of individual training buildings and subsequently combine them into an average archetype distribution (Cerezo et al., 2017). Even though stochastic in nature, these approaches simplify the true archetype heterogeneity of the training data through the averaging process. This leaves a need for research on methods to include this source of error.

The following paper (conference paper [P2]), presented at the 15th conference of the International Building Performance Simulation Association (IBPSA), *Building Simulation 2017*, demonstrates an archetype-based application of the Bayesian calibration methodology by Kennedy and O’Hagan (2001) that was briefly reviewed in Section 1.4.2. It shows how to identify posterior distributions of seven unknown input parameters to an hourly dynamic physics-based BEM of a SFH archetype using annual DH energy use data of 450 training buildings. The study constitutes a novel effort to expand the KOH framework to include the inference of shared archetype parameters, and not just parameters of individual buildings, as has been current practice.

The paper makes use of sensitivity analysis (SA) as a method to decide which BEM input parameters to calibrate. The process of setting up and carrying out this SA lead to a considerable amount of comparative research on the appropriateness of SA methods for BEM-based analysis, which was conducted concurrently with the research on archetype-based inference. This research resulted in the publication of journal papers [S1] and [S2], and conference paper [S3]. These secondary publications are appended to the thesis.



Bayesian Calibration Of Residential Building Clusters Using A Single Geometric Building Representation

Martin Heine Kristensen^{1*}, Ruchi Choudhary², Rasmus Høst Pedersen³, Steffen Petersen¹

¹Department of Engineering, Aarhus University, 8000 Aarhus C, DK

²Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

³AffaldVarme Aarhus, Bautavej 1, 8210 Aarhus V, DK

*Corresponding author (mhk@eng.au.dk)

Abstract

For a homogeneous cluster of single-family dwellings, an archetype model incorporating simple scalable geometry and an hourly dynamic building energy model was set-up to represent its energy performance. Using metered annual energy use for a random sample of 450 buildings in the cluster, the archetype model was calibrated in a Bayesian regression framework using the floor area as common scale for regression of the physics-based input to the hourly dynamic energy model. In this process, posterior estimates of seven selected building parameters shared by buildings within the cluster were inferred. The calibrated archetype model was used to make predictions of annual building energy use with a normalized mean bias error (NMBE) of 2.3% and a coefficient of variation of the root mean squared error (CVRMSE) of 26.5%.

Introduction

Building energy modeling (BEM) often relies on physics-based principles for representing the thermodynamic mechanisms of buildings (see Kavgic et al. (2010) for a review on physics-based approaches). To have better estimates of the inputs to such building models, it is often advantageous to calibrate them using appropriate field data. In the case where a vast amount of buildings is to be modeled for analysis, e.g. the housing stock of a larger city, it might be practical to model a group of similar buildings as a homogeneous cluster, also known as archetype modeling. However, as emphasized by Reinhart and Davila (2016), the inevitable presence of heterogeneity among buildings within an archetype can make it difficult to identify a deterministic best fit of physical model parameters in the calibration process. Using instead probabilistic calibration, one may incorporate and propagate parameter uncertainty more appropriately including both aleatory uncertainty (i.e. heterogeneity across a building cluster) and epistemic uncertainty (i.e. uncertainty about the true value of parameters).

Booth et al. (2012) used a monthly average quasi-steady-state energy model in combination with the Bayesian calibration framework by Kennedy and

O'Hagan (2001) to set up a probabilistic model and calibrate uncertain model parameters for a group of 35 identical flats using energy performance certificates data and metered energy use. In Booth et al. (2013), this framework was expanded to the scale of housing stock models using synthetic/regressed calibration data from a macro scale and GIS-recorded building geometry. In a recent study by Sokol et al. (2017), a residential building stock was subdivided into archetypes and calibrated using the Bayesian framework and individual EnergyPlus BEMs to set up an aggregated urban building energy model (UBEM). They used monthly and annually metered calibration data, in combination with simple building-specific property information from tax assessments and GIS-based estimates of the geometry, to achieve a mean error of 44%-47% and a CVRMSE of 58%-66% for a validation sample of 2263 buildings.

In all of these studies, information about the geometry of the individual buildings of the cluster was available (or inferred); however, in many cases, this information may be sparse or difficult to collect. In this study, we therefore investigate whether a single geometric and scalable building representation – an archetype model – modeled using an hourly dynamic BEM can adequately represent the variations within a cluster of similar residential buildings. Using the Kennedy and O'Hagan (2001) calibration methodology that relies on Bayesian propagation of uncertainties, we train the model using metered district heating consumption data, and information about the construction year and building footprint area from 450 buildings belonging to the archetype, and validate it against an unseen sample of 150 buildings to test the reliability in the inferred results. As a result, both the aleatory and epistemic uncertainties in model parameters are quantified for the cluster to perform highly accurate archetype model predictions.

Method

Characteristics of the building cluster

A building cluster consisting of detached single-family dwellings from the Danish building stock was iden-

tified based on the typological characterizations defined by the European research project TABULA (Wittchen and Kragh, 2012; Loga et al., 2016). The cluster represents buildings constructed between 1979-1998 in Denmark, a period with rather uniform building traditions and non-varying building codes. The buildings within the cluster were therefore assumed to exhibit similar building-physical properties. In fact, the value and state of unknown technical properties, whether it be the U-values or the air tightness of the envelope, etc., were assumed independent and identically distributed samples drawn from the same population. Pictures of typical buildings represented by the cluster are shown in Figure 1.



Figure 1: Examples of typical buildings of the cluster.

A total sample of $N = 2000$ existing buildings situated in the district heating grid of Aarhus, Denmark, were sampled from the cluster. Information about the construction year, building footprint area, number of floors and envelope materials for each building i within the cluster sample was collected from the Danish Building and Dwelling Register (BDR), which is publicly available (Table 1). The aggregated annual district heating energy use, E^{DH} , for the last three years (2013-2015) was collected for all N buildings from the local district heating supplier (AffaldVarme Aarhus). As indicated in (1), E^{DH} consists of energy use for hydronic space heating, E^{SH} , and energy use for on-site domestic hot water preparation, E^{DHW} . The data does not allow any distinction between how much energy is spent on space heating and DHW, respectively.

$$E_i^{DH} = E_i^{SH} + E_i^{DHW}. \quad (1)$$

The expected mean annual energy use per building, $\overline{E^{DH}}$, was computed as the direct average of the metered consumption over the last three years (Table 1) in accordance with (2). It was assumed that this three-year average represented the true mean.

$$\overline{E_i^{DH}} = \frac{1}{3} \sum_{y=1}^3 E_{i,y}^{DH}. \quad (2)$$

Geometric building representation

The actual geometry of the individual buildings in the sample was unknown. The individual building geometry was instead estimated using a rectangular box (Figure 2) with dimensions based on simple rules applied to known information about the floor area, A_{floor} , and number of floors, n_{floors} , (Table 1), and

our a-priori beliefs about the length-to-width-ratio, ρ_{LW} , and floor height, h_{floors} , (Table 2). In the case of two or three-storey buildings with unequal floor areas, A_{floor} was averaged over the floors to fit the geometric model. The four facades were assumed to face the four cardinal directions (North, East, South, West) while the unknown window area in each facade was allowed to vary for each building (Table 2).

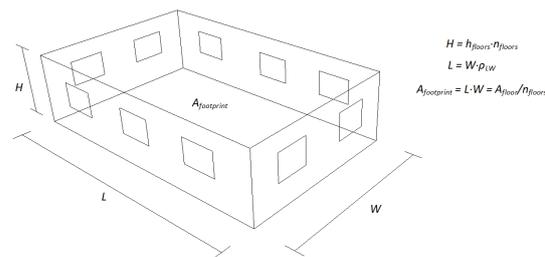


Figure 2: Geometric model of archetype building.

Building energy model for space heating and domestic hot water

The energy use for space heating, E_{SH} , was modeled using the simple hourly dynamic model of ISO 13790:2008 (International Organization for Standardization, 2008), treating the building as a single thermal zone. The Danish building code energy requirements in force at the time of construction (BR77, BR82 and BR85), historical surveys of the Danish building stock (Bøhm et al., 2009; Bergsøe, 2015), and our educated guesses were used to estimate the majority of the technical and occupation-related inputs required to run the model (Table 2). Ventilation was assumed a mix of infiltration and opening of windows. Based on studies by Rijal et al. (2007), the airflow through windows was modeled hourly as a percentage of maximum design airflow using a logistic transformation of a linear regression on the outdoor temperature (3) where T_{out} is the outdoor temperature, and a and b are empirical regression coefficients (Table 2).

$$p = \frac{\exp(aT_{out} + b)}{1 + \exp(aT_{out} + b)}. \quad (3)$$

Internal heat loads from occupants were modeled as a scalable day profile for the activity level constructed with a variable period without any presence called Away time (Figure 3).

Energy use for DHW, E_{DHW} , was modeled using a simple linear model proportional to the consumed amount of hot water under the assumption that the rate of consumption was approximately constant

Table 1: Data quantiles of known cluster characteristics ($N = 2000$).

Name	Unit	Quantiles				
		2.5%	25%	Median	75%	97.5%
Construction year	-	1979	1982	1987	1995	1998
Number of floors*, n_{floors}	-	1	1	1	1	1
Floor area, A_{floor}	m^2	98	130	149	170	238
Energy use, E^{DH}	MWh/year	7.20	11.85	14.65	17.75	28.60
Energy use intensity**, EUI	$kWh/m^2/year$	54.7	83.4	97.9	112.8	155.3

* 97.8% of the buildings had one floor, 2.15% had two floors, and 0.05% had three floors.

** Calculated quantity, $EUI = E^{DH}/A_{floor}$.

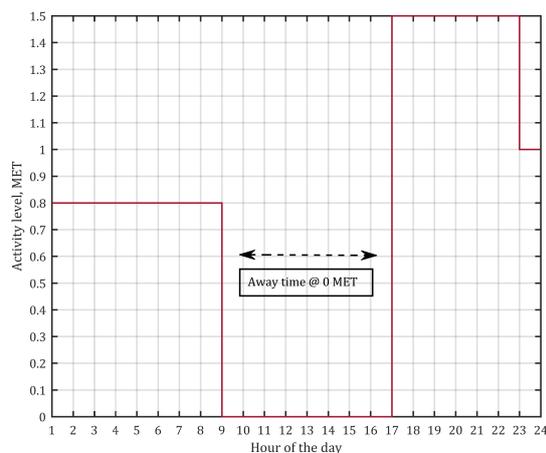


Figure 3: Day profile for occupant heat loads with variable period without presence.

throughout the year:

$$E_{DHW} = \rho c V_{occ} n_{occ} (T_{flow} - T_{mains}) + \sum_{t=1}^L H_{circ}^{(t)} (T_{flow} - T_{air}^{(t)}). \quad (4)$$

The room air temperature, T_{air} , was modeled dynamically in the ISO 13790 model, and the volumetric heat capacity of the water, ρc , was given as $4140 \text{ kJ}/(\text{m}^3\text{K})$. The remaining model parameters were unknown, i.e. V_{occ} (hot water consumption per occupant per year), n_{occ} (number of occupants), T_{flow} (flow temperature), T_{mains} (mains supply temperature), and H_{circ} (specific heat loss in the circulation system). An additional time varying term was added to account for residual hot water circulation loss to the building, assuming that the temperature difference between hot water and room air was proportional. This loss was also added as an internal heat gain in the dynamic model for calculation of energy for space heating.

The simulated district heating energy use should represent the true mean energy use, hence average weather conditions of the outdoor temperature and

global solar irradiation were needed as boundary conditions. For this end, the Danish design reference year (DRY) was applied; a set of hourly weather conditions composed from historical measurements in Denmark used for designing and testing buildings (Jensen and Lund, 1995).

Sensitivity analysis for selecting calibration parameters

Prior to calibration of the archetype model, 32 input parameters were unknown. For each parameter, probability density functions (PDFs) were assigned to reflect our beliefs about their values; distribution types *Uniform*(min, max), *Beta*($shape1, shape2$), *Gamma*($shape, scale$), and *Normal*($mean, variance$) were applied (Table 2). As these distributions rely on a-priori information, they are referred to as *a-priori distributions* or simply *priors*.

A probabilistic sensitivity analysis using the method of Sobol' (Sobol', 1993) was conducted to rank the 32 unknown model input parameters, given their prior distributions, in descending order of importance (Table 2). Based on recommendations by Kristensen and Petersen (2016), who analyzed the performance of three different sensitivity analysis methods on the ISO 13790 energy calculation models, the Sobol' total effects index, ST_i , was applied as a measure of explaining individual parameters' combined effect on model output variability. A total of 15 000 Monte Carlo iterations were performed to obtain convergence of the Sobol' algorithm, $\sum(ST_i) \approx const$. No correlations between model parameters were taken into account.

Ideally, one ought to calibrate all uncertain model parameters; however, as this is computationally infeasible, a limited number of parameters must be chosen. We selected the seven highest ranked parameters from the sensitivity analysis, which in total accounted for approx. 75% of the model output variability (Figure 4). In descending order of importance, the selected parameters were: the heating set point, U-value (windows), appliances heat load, infiltration rate @50 Pa, window-wall ratio (North), window-wall ratio (South), and U-value (walls). The remaining 25 parameters were left uncalibrated with their prior uncertainty specification being propagated through the

model.

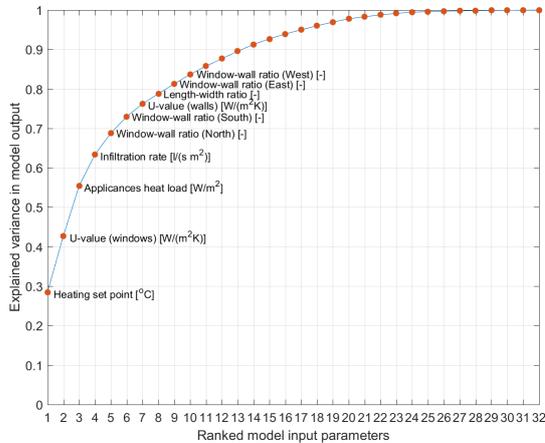


Figure 4: Cumulative sum of explained variance in model output (ST_i) by ranked input parameters.

Calibration framework

The Bayesian approach used in this study followed the approach proposed by Kennedy and O'Hagan (2001). It explicitly incorporates uncertainty in model inputs, uncertainty due to limited numbers of simulation runs, and discrepancy between the model outputs and the actual energy-consuming mechanism of the buildings. We represented the measured energy use data (Table 1) statistically as

$$y_i^{meas} = \zeta(\mathbf{x}_i^{meas}) + e_i^{meas} \quad i = 1, \dots, n, \quad (5)$$

where y_i^{meas} is the field data for the i -th building, i.e. measured annual energy use in kWh/year, each observed by a p -dimensional vector of known explanatory design points, $\mathbf{x} = (x_1, x_2, \dots, x_p)$, for which $x \in \mathbb{R}^p$. In this study, we applied the known heated floor area as the only explanatory variable; hence $p = 1$. The field data, y_i^{meas} , was assumed to be realizations from the true energy-consuming process, $\zeta(\mathbf{x}_i)$, at the observed design points, \mathbf{x}_i^{meas} , with an unknown observation error, e_i^{meas} . We modeled the true process as:

$$y_i^{meas} = \eta(\mathbf{x}_i^{meas}, \boldsymbol{\theta}) + \delta(\mathbf{x}_i^{meas}) + e_i^{meas} \quad i = 1, \dots, n, \quad (6)$$

using the building energy model as an emulator, $\eta(\mathbf{x}_i, \boldsymbol{\theta})$, evaluated at the $(p + q)$ -dimensional input vector ($\mathbf{x}_i^{meas}, \boldsymbol{\theta}$) comprising the vector of known design points, \mathbf{x}^{meas} , and additional calibrated parameters, $\boldsymbol{\theta} \in \mathbb{R}^q$, represented by the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)$. For this study, we took the top-seven ranked model parameters (Figure 4) as the calibration parameters; hence $q = 7$. As no model is perfect, a stochastic model bias was introduced through

$\delta(\cdot)$ independently of $\eta(\cdot, \cdot)$ to account for discrepancy between the model and the true physical process.

The inclusion of a noise-term, e_i^{meas} , allowed for different observations of y at identical settings of \mathbf{x} . In addition to measurement error, the true archetype energy-consuming process may be inherently unpredictable and stochastic due to e.g. occupant behavior. Thus, strictly speaking, e_i^{meas} includes any residual variability that cannot be decomposed in addition to measurement error. This also includes violations of the assumption of cluster homogeneity. We denote these different sources of uncertainty (archetype heterogeneity, measurement error, and chance variability) *aleatory uncertainty*.

The *true* or *best achievable* distributions of the calibration parameters, $\boldsymbol{\theta}$, are to be inferred in the calibration process, and are thus unknown. They are referred to as *a-posteriori* distributions or simply *posteriors*. The posterior distributions represent the *epistemic parameter* uncertainty, that is uncertainty due to modeling assumptions of the BEM, our estimates of the prior distributions, and the amount of calibration data available for a precise identification of the true parameter values. We represented the calibration parameters in the simulations in terms of their priors, $\mathbf{t} = (t_1, t_2, \dots, t_q)$, given in Table 2. Evaluating the BEM at randomly selected design points, \mathbf{x}_j^{sim} , and at random samples of the prior distributions, \mathbf{t}_j , yields simulated annual energy use data denoted y_j^{sim} .

$$y_j^{sim} = \eta(\mathbf{x}_j^{sim}, \mathbf{t}_j) + e_j^{sim} \quad j = 1, \dots, m. \quad (7)$$

As proposed by Higdon et al. (2004), we added a small numerical white noise error-term, e_j^{sim} , in the statistical representation (7) to secure the positive definiteness of the later covariance matrix (16) necessary for the Gaussian process (GP) regression. We modeled both e_i^{meas} and e_j^{sim} as i.i.d. Gaussian distributed noise:

$$e_i^{meas} \sim N(0, \sigma_{meas}^2), \quad (8)$$

$$e_j^{sim} \sim N(0, \sigma_{sim}^2). \quad (9)$$

For the evaluation of the model, $\eta(\cdot, \cdot)$, the remaining 25 non-calibrated input parameters were sampled at random for each simulation run j . As the value of the functions $\eta(\cdot, \cdot)$ and $\delta(\cdot)$ are known only at the applied design points, they were modeled as Gaussian processes (GP) represented in terms of multivariate Gaussian distributions (10)-(11), allowing an infinite number of different regressions to be fit. Other formulations are available though, e.g. multiple linear regression as proposed by Li et al. (2016). A key property of the applied GP regression model is that it obtains a perfect fit to the training data. However, due to the inclusion of the observation error, e_i^{meas} ,



Table 2: Prior distributions and ranking of 32 unknown model parameters using the Sobol' sensitivity analysis method.

Parameter	Unit	PDF	Quantiles					Sensitivity Analysis	
			2.5%	25%	Median	75%	97.5%	STi	Rank
Geometry									
Length-width ratio	-	Unif(0.3;3.0)	0.4	1.0	1.7	2.3	2.9	2.6%	8
Room height	m	Unif(2.1;3.1)	2.12	2.35	2.60	2.85	3.08	1.2%	-
Window-wall ratio (North)	-	Beta(3;3)	0.10	0.30	0.40	0.50	0.75	5.4%	5
Window-wall ratio (East)	-	Beta(3;3)	0.10	0.30	0.40	0.50	0.75	2.5%	9
Window-wall ratio (South)	-	Beta(3;3)	0.10	0.30	0.40	0.50	0.75	4.2%	6
Window-wall ratio (West)	-	Beta(3;3)	0.10	0.30	0.40	0.50	0.75	2.4%	10
Window frame fraction	-	Beta(8;24)	0.12	0.20	0.24	0.30	0.41	0.5%	-
Transmission									
Temp. adjustment factor (ground)	-	Beta(10;1.1)	0.78	0.87	0.91	0.94	0.98	0.1%	-
U-value (floors)	W/(m ² K)	Gam(7;0.029)	0.08	0.15	0.19	0.24	0.37	1.7%	-
U-value (walls)	W/(m ² K)	Gam(7;0.043)	0.12	0.22	0.29	0.37	0.56	3.3%	7
U-value (basement)	W/(m ² K)	Gam(4;0.088)	0.10	0.22	0.32	0.45	0.77	1.4%	-
U-value (roofs)	W/(m ² K)	Gam(7;0.029)	0.08	0.15	0.19	0.24	0.37	2.1%	-
U-value (windows)	W/(m ² K)	Gam(50;0.05)	1.86	2.25	2.48	2.73	3.24	14.2%	2
SHGC (windows)	-	Beta(45;32.6)	0.49	0.56	0.60	0.64	0.71	0.4%	-
Internal shading coef.	-	Unif(0.2;0.9)	0.22	0.38	0.55	0.73	0.88	0.1%	-
Internal heat capacity	kJ/(m ² K)	Gam(10;17)	59	138	225	312	391	1.0%	-
Effective mass area	m ² /m ²	Gam(70;0.04)	2.2	2.6	2.8	3.0	3.5	<0.1%	-
Heat conduction (mass)	W/(m ² K)	Gam(350;0.026)	7.1	8.3	9.1	9.8	11.3	<0.1%	-
Heat transfer coef. (surf.-air)	W/(m ² K)	Gam(350;0.01)	3.1	3.3	3.4	3.6	3.8	<0.1%	-
Ventilation									
Infiltration airflow @ 50 Pa	l/(s m ²)	Gam(5;0.50)	0.97	2.02	2.80	3.76	6.15	7.9%	4
Design airflow (windows)	l/(s m ²)	Gam(10;0.04)	0.19	0.31	0.39	0.48	0.68	0.9%	-
Regression coef. <i>a</i>	-	Unif(0.1;0.4)	0.11	0.17	0.25	0.33	0.39	0.3%	-
Regression coef. <i>b</i>	-	Unif(-4;-1)	-3.92	-3.25	-2.50	-1.75	-1.07	1.9%	-
Occupation									
Occupant density	m ² /pers	Gam(10;5.18)	25	40	50	62	88	1.9%	8
Occupant heat load	W/(MET pers)	Gam(50;1.60)	60	70	80	90	100	<0.1%	-
Appliances heat load	W/m ²	Gam(40;0.75)	21	27	30	33	40	12.8%	3
Away time @ 0MET	h/day	Unif(0;12)	0.3	3.0	6.0	9.0	11.7	0.5%	-
Room heating set point	°C	Norm(20.5;22)	16.6	19.2	20.5	21.8	24.4	28.5%	1
Domestic hot water									
DHW flow temperature	°C	Gam(200;0.263)	42.7	48.9	52.3	56.0	63.3	0.9%	-
Mains temperature	°C	Gam(100;0.125)	10.2	11.6	12.5	13.3	15.1	<0.1%	-
Specific circulation loss	W/K	Gam(25;0.32)	5	7	8	9	11	<0.1%	-
Hot water consumption	m ³ /(pers year)	Gam(25;0.75)	9.7	12.9	14.8	16.9	21.4	1.1%	-

* Sobol' total effects index, $0 \leq ST_i \leq 1$

the training data was treated as noisy measurements letting instead the GP model identify and fit the *true* mean energy use at each value of \mathbf{x} by attributing fluctuations from the archetype mean energy use to the noise-term.

$$\eta(\cdot, \cdot) \sim N(0, \Sigma_\eta). \quad (10)$$

$$\delta(\cdot) \sim N(0, \Sigma_\delta). \quad (11)$$

By standardizing the data ($y, \mathbf{x}, \mathbf{t}, \boldsymbol{\theta}$) to a range [0; 1], the two GPs were specified with constant mean functions of zero and squared exponential covariance functions (12)-(13), allowing a smooth and stationary representation of the process less prone to numerical problems:

$$\Sigma_\eta = \sigma_\eta^2 \exp \left(- \sum_{d=1}^p \beta_{\eta,d} |x_{i,d} - x'_{i,d}|^2 \right) \cdot \exp \left(- \sum_{d'=1}^q \beta_{\eta,p+d'} |t_{i,d} - t'_{i,d}|^2 \right), \quad (12)$$

$$\Sigma_\delta = \sigma_\delta^2 \exp \left(- \sum_{d=1}^p \beta_{\delta,d} |x_{i,d} - x'_{i,d}|^2 \right). \quad (13)$$

In (12)-(13), σ_η^2 and σ_δ^2 control the marginal function variance of the energy-consuming process, i.e. variation that is explained by the GP regression function, while β_η and β_δ are weighting parameters for the d -th dimension of the input space controlling the strength of each model parameter as a predictor. The larger β is, the more dimension d is weighted in the summation of the Euclidean distances. We refer to these variance and weighting parameters as hyperparameters of the GPs. They were unknown and hence to be inferred in the calibration process. The total number of



unknown model parameters were thus comprised by calibration parameters θ , variance parameters σ_{meas}^2 , σ_{sim}^2 , σ_η^2 and σ_δ^2 , and weighting parameters β_η and β_δ .

Defining a $(n + m)$ observation vector, $\mathbf{z} = (\mathbf{y}^{meas,T}, \mathbf{y}^{sim,T})^T$, containing n field observations $\mathbf{y}^{meas} = (y_1^{meas}, \dots, y_n^{meas})^T$, and m simulation runs $\mathbf{y}^{sim} = (y_1^{sim}, \dots, y_m^{sim})^T$, the regression model becomes the following by combining (5) and (6):

$$\mathbf{z} = \begin{bmatrix} y_1^{meas} \\ \vdots \\ y_n^{meas} \\ y_1^{sim} \\ \vdots \\ y_m^{sim} \end{bmatrix} = \begin{bmatrix} \eta(\mathbf{x}_1^{meas}, \theta) + \delta(\mathbf{x}_1^{meas}) + e_1^{meas} \\ \vdots \\ \eta(\mathbf{x}_n^{meas}, \theta) + \delta(\mathbf{x}_n^{meas}) + e_n^{meas} \\ \eta(\mathbf{x}_1^{sim}, \mathbf{t}_1) + e_1^{sim} \\ \vdots \\ \eta(\mathbf{x}_m^{sim}, \mathbf{t}_m) + e_m^{sim} \end{bmatrix}. \quad (14)$$

Applying Bayes theorem, the joint posterior distribution of the parameters, conditional on the augmented observation vector \mathbf{z} , is obtained:

$$p(\theta, \sigma_{meas}^2, \sigma_{sim}^2, \sigma_\eta^2, \sigma_\delta^2, \beta_\eta, \beta_\delta | \mathbf{z}) \propto L(\mathbf{z} | \theta, \sigma_{meas}^2, \sigma_{sim}^2, \sigma_\eta^2, \sigma_\delta^2, \beta_\eta, \beta_\delta) p(\theta) p(\sigma_{meas}^2) p(\sigma_{sim}^2) p(\sigma_\eta^2) p(\sigma_\delta^2) p(\beta_\eta) p(\beta_\delta). \quad (15)$$

The likelihood of \mathbf{z} conditional on the parameters, $L(\cdot | \cdot)$, is computed from a multivariate Gaussian distribution with zero mean function and a combined $(n + m) \cdot (n + m)$ covariance matrix Σ_z :

$$\Sigma_z = \Sigma_\eta + \begin{bmatrix} \Sigma_\delta & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{I}_n \sigma_{meas}^2 & 0 \\ 0 & \mathbf{I}_m \sigma_{sim}^2 \end{bmatrix}, \quad (16)$$

where \mathbf{I}_n and \mathbf{I}_m are the $n \cdot n$ and $m \cdot m$ identity matrices, respectively, to put the Gaussian noise-terms on the diagonal.

For this study, we performed 1000 simulations, \mathbf{y}^{sim} , using draws from the prior distributions ($m = 1000$), and applied a sub-sample of 450 randomly selected field data points, \mathbf{y}^{meas} , for calibration ($n = 450$) from the $N = 2000$ cluster sample.

Prior distributions

Prior information is a very important aspect of the Bayesian framework as it influences the calibration output through a weighted fitting of the prior information and data at hand. For the calibration parameters, θ , we applied the distributions defined

in Table 2 to reflect our a-priori beliefs about their values. The standardized variance hyperparameters, σ_{meas}^2 , σ_{sim}^2 , σ_η^2 and σ_δ^2 , were parametrized as precision parameters, $\lambda = 1/\sigma^2$, for which Gamma distributions were applied constraining the values to $\lambda \in \mathbb{R} | 0 < \lambda < \infty$. The weighting parameters, β_η and β_δ , were parametrized as $\rho = \exp(-\beta/4)$, constraining the values to $\rho \in \mathbb{R} | 0 < \rho < 1$, making it convenient to apply Beta distributions (Table 3).

For the parametrised weighting parameters, ρ , most of the prior mass is placed on values of ρ near 1 indicating vague parameter strength.

MCMC algorithm and convergence for posterior inference

The multi-dimensional joint posterior distribution cannot be obtained analytically. Therefore, we employed a random walk Markov Chain Monte Carlo (MCMC) algorithm – the Metropolis-Hastings algorithm (Gelman et al., 2014) – whose equilibrium distribution is indeed an approximation of the joint posterior distribution. Four chains were run in parallel with randomly dispersed starting points in the parameter space to draw samples from the posterior distribution. For each chain, 14000 MCMC samples were drawn with the first 7000 samples of the chains being considered cool, meaning that information about the starting point might still prevail. Samples from this cold period were thus discarded leaving only the warm part of the chains for analysis.

Convergence in the warm chains was monitored in terms of the potential scale reduction factor, \hat{R} , for which $\hat{R} \in \mathbb{R} | 1 < \hat{R} < \infty$. It is an estimate of the scale with which the variations in the inferred parameter distributions might be reduced if the simulations were continued in the limit $n \rightarrow \infty$ ($\lim_{n \rightarrow \infty} \hat{R} \rightarrow 1$) (Gelman et al., 2014). \hat{R} accounts for the within-chain and between-chain variance in the warm chains, simultaneously evaluating both the mixing and the stationarity of it. For $\hat{R} < 1.1$, a stable and converged estimation was considered for each parameter, respectively.

Results

The MCMC algorithm is considered converged for both the calibration parameters and the hyperparameters ($\hat{R} < 1.1$ for all parameters). The calibration parameters are shown in Figure 5 in terms of their prior and posterior (calibrated) distributions.

Using the posterior parameter distributions, 1000 draws from the GP regression model are shown in Figure 6 with posterior uncertainty given as 95% uncertainty bands. In Figure 6A, predictions from the combined regression model, $y = \eta(\mathbf{x}, \theta) + \delta(\mathbf{x}) + \epsilon$, are shown; in Figure 6B, isolated predictions of the physics-based ISO 13790 emulator, $\eta(\mathbf{x}, \theta)$, are shown; in Figure 6C, isolated predictions of the model bias term, $\delta(\mathbf{x})$, explaining the inadequacy of the physical model are shown; and in Figure 6D, the pos-

Table 3: Standardized prior distributions for hyperparameters of covariance functions.

Hyperparameter	PDF	Constraints		Quantiles				
		Lower	Upper	2.5%	25%	Median	75%	97.5%
λ_η	Gam (10,0.10)	0.1	∞	0.5	0.8	1.0	1.2	1.7
λ_δ	Gam (10,3.33)	0	∞	16.0	25.7	32.2	39.7	56.9
λ_{meas}	Gam (10,20.0)	0	∞	95.9	154.6	193.4	238.3	342.0
λ_{sim}	Gam (10,1000)	100	2e5	4799	7721	9662	11907	17054
ρ_η	Beta(1.0;0.5)	0	1	0.05	0.44	0.75	0.94	1.0
ρ_δ	Beta(1.0;0.4)	0	1	0.06	0.51	0.82	0.97	1.0

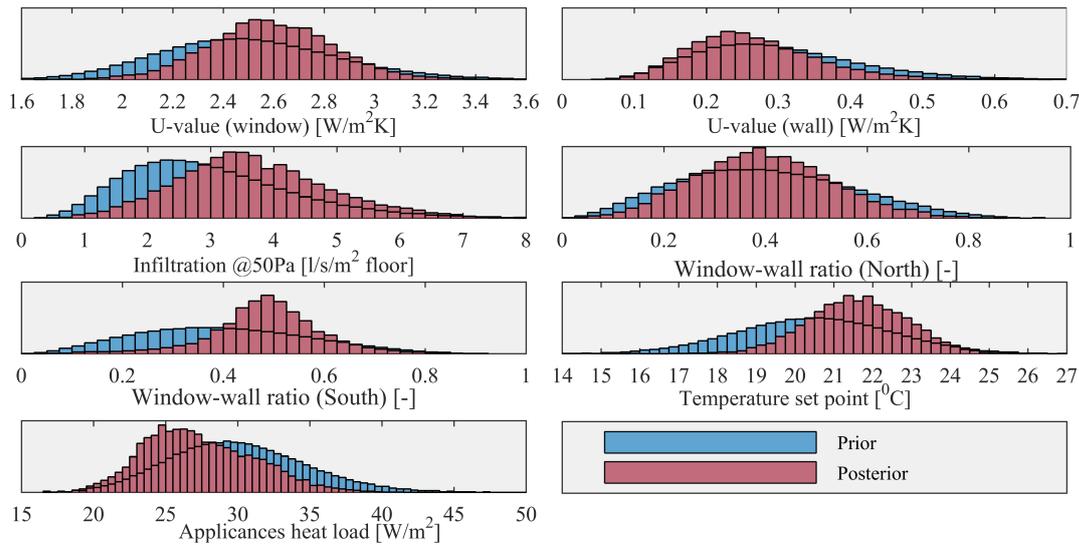


Figure 5: Prior PDFs (Table 2) and posterior PDFs (1000 draws) for calibration parameters.

terior distribution of the standard deviation of the observation error, σ_{meas} , is shown. It should be noted that the bias term only makes up a negligible part of the overall model predictions (approx. zero across the entire range of x), leaving predictions of the true energy-consuming process of the archetype building, $\zeta(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x})$ (noise-free predictions), to be based mainly on the physics-based term of the model. Following ASHRAE Guideline 14 (ASHRAE, 2014), the normalized mean bias error (NMBE), the coefficient of variation of the root mean squared error (CVRMSE), and the coefficient of determination (R^2) are used to assess model fit. These are supplemented with the mean absolute percentage error (MAPE) and shown in Table 4.

Discussion

The application of a common scalable geometric building model to fit all buildings of the cluster proves to be appropriate for the given field data. The regression model ascribes no significant value to the model bias term, $\delta(x)$, indicating that the field data conforms to the physics-based model (geometrical model in combination with hourly dynamic ISO 13790 model and DHW model). Given the constraints of the

archetype model, the seven calibrated parameters are thus regarded true for the cluster.

As opposed to previous studies applying monthly quasi-steady-state housing stock modeling (Booth et al., 2012, 2013), the application of hourly dynamic modeling allows a more meaningful interpretation of the calibration parameters by permitting a better inclusion of time-varying parameters. However, additional work is required to improve the representation of individual occupant schedules, user pattern, and other highly stochastic and time-varying mechanisms. As long as the model does not fully and accurately capture the energy-consuming phenomena, the calibrated parameters can only be considered true when applied in connection with the applied model.

Goodness of fit

Regressing average annual building energy use as a function of heated floor area leaves approx. 60% unexplained variance ($R^2 = 42\%$). This residual variability the model has successfully identified and captured as i.i.d. Gaussian distributed observation error, e^{meas} , leaving only the true mean building energy use at a given floor area, $\zeta(x)$, to be fitted. Hence, in average, the regression model predicts the energy use of

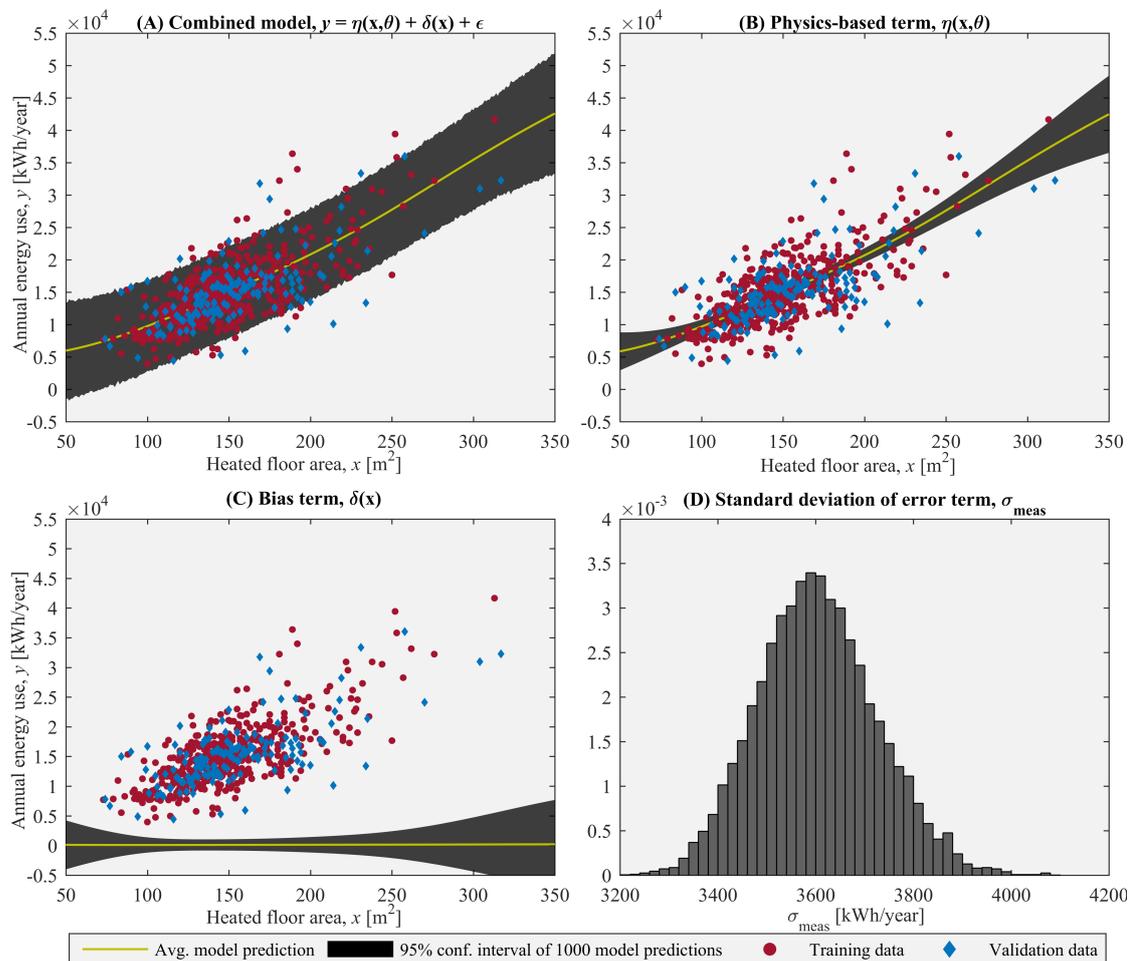


Figure 6: Archetype model predictions using 1000 draws from posterior distributions. (A) Combined regression model. (B) Physics-based term (ISO 13790 emulator). (C) Bias term. (D) Histogram of posterior distribution of error term standard deviation.

Table 4: Measures of model predictive accuracy; NMBE (normalized mean bias error), MAPE (mean absolute percentage error), CVRMSE (coefficient of variation of the RMSE), and R^2 (coefficient of determination). All measures were computed using posterior mean predictions of the true energy-consuming process, $\zeta(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x})$ (noise-free predictions).

Dataset	n	NMBE \pm 95% conf. interval	MAPE \pm 95% conf. interval	CVRMSE	R^2
Training set	450	-0.3% \pm 2.2%	20.0% \pm 1.9%	24.1%	61.6%
Validation set	150	+2.3% \pm 4.3%	21.9% \pm 4.5%	26.5%	41.9%

the archetype within the ranges of the validation data (50-350 m²) to an NMBE of approx. 2.3 % for the validation set. The NMBE measure holds for the average of large-sample predictions, e.g. aggregated predictions on the urban scale, but as the NMBE measure is subject to cancellation errors, the model predictive accuracy of single-building predictions may be more intuitively interpreted in terms of the MAPE, a measure of absolute error, resulting in approx. 22 %.

This means that while the aggregated prediction error of a large sample of the archetype buildings is approx. 2.3 % in average, single-building predictions will be approx. 22 % uncertain in average. The model precision, i.e. the consistency of the model predictive accuracy, is assessed in terms of the CVRMSE, a normalized measure of variability between field data and model predictions across the ranges of the validation data (50-350 m²). A CVRMSE score of 26.5 % seems



satisfactory given the fact that measurements originate from different buildings, and hence should not be confused with the ASHRAE Guideline 14 (ASHRAE, 2014) requirements for monthly calibrations (15%) and hourly calibrations (30%) as such requirements only cover single-building calibrations, where data is not subjected to the same level of heterogeneity and hence ought to be calibrated more easily.

Sources of error

As elaborated upon in the section *Calibration framework*, the observation error, e^{meas} , includes a mixture of uncertainties: geometrical heterogeneity between buildings of the archetype (even though the floor area is similar, the surface-to-volume ratio might not be), heterogeneity of technical properties between buildings (U-values, air tightness, etc.), variations due to occupant-related building operation (heating set point temperature, schedules, venting, domestic hot water, etc.), measurement error, and any residual variability that cannot be decomposed and ascribed to the parameters of the Gaussian process models. We assume the majority of this error to be driven by chance variability among buildings, predominantly occupant-related variation. However, also the assumption of archetype homogeneity is suspected to contribute significantly.

In order to fully understand the effect of these individual sources of uncertainty and ultimately secure a better and more reliable parameter calibration, additional work has to be performed to address the sources of error individually. In this context, the homogeneity assumption could be investigated in a comparison of single-building calibrations against the entire archetype calibration.

To further reduce model uncertainty, one could include additional explanatory input dimensions in \mathbf{x} , i.e. other design points known for each observation besides the heated floor area, e.g. number of occupants registered on the address, number of heating and/or cooling degree days in the year, level of refurbishment, etc. The application of Gaussian processes to model such diverse data spaces enables a very flexible framework that will encompass both linear and non-linear phenomena without relying on a specific model structure and, at the same time, allow the modeler to embed his prior knowledge if he wants to. This feature in particular is highly beneficial for sparsely distributed input spaces. However, we have to acknowledge that modeling the two-dimensional data presented here might have been done using a more parsimonious model, e.g. a linear model, without any noticeable difference in model predictive accuracy.

Scaling archetype predictions to the urban level

A natural next step would be to set up similar archetype models for the remaining building stock

and use these for composing urban-scale building energy models to assess the aggregated energy use and subsequently the retrofit potential of such case studies. However, as advocated for by Reinhart and Davila (2016), the approach of scaling archetype models to the urban level by multiplying each individual archetype by the number of buildings in the urban area ignores the urban context, i.e. the presence of shading, local wind patterns, etc. Hence, one has to find a way to reliably compose aggregated models and, at the same time, account for the presence of such context-specific uncertainties, e.g. by inclusion of GIS-data.

Assessing retrofit potential

The probabilistic nature of the archetype model makes it suitable for assessing the averaged effect of building retrofits under uncertainty. This could be done by setting priors on selected physical parameters to reflect the anticipated, but uncertain, value of the parameters in question after implementation of energy conserving measures (ECM). Using this framework, the uncertainty of the ECM would then be propagated through the model to reflect its effect on the annual energy use of an average building represented by the archetype. The details of such an analysis, and how to actually conduct it, we leave for future work; however, inspiration for a similar study can be found in Booth and Choudhary (2013) who performed a retrofit analysis for the UK housing stock under uncertainty.

Conclusion

Using a Bayesian calibration technique it was demonstrated how a homogeneous cluster of single-family dwellings could be fitted to a single archetype geometry and hourly dynamic building energy model. The probabilistic treatment of the model allowed a quantification of the epistemic uncertainties embedded in the calibration parameters and the aleatory uncertainty in the cluster homogeneity assumption. The application of a single archetype model to fit the cluster showed to be adequate as all weight was attributed to the physics-based model, leaving only insignificant influence to the statistical model bias term. The calibrated archetype model can be used to make predictions of annual building energy use with a NMBE of 2.3% and a CVRMSE of 26.5%.

Acknowledgements

The research was conducted as part of the 'Resource Efficient Cities Implementing Advanced Smart City Solutions' (READY) project, work package 3, financed by the 7th EU Framework Programme (FP7-Energy project reference: 609127). Furthermore, the authors would like to thank the district heating company in Aarhus, AffaldVarme Aarhus, for supplying the building energy data that forms the basis of the study. Especially business development manager



Adam Brun deserves our gratitude, without whose help and cooperation this study would not have been possible.

References

- ASHRAE (2014). *ASHRAE Guideline 14-2014 – Measurement of Energy, Demand, and Water Savings*. ASHRAE.
- Bergsøe, N. C. (2015). *SBi 2015:25 Tæthed i eksisterende bygninger – Analyse af målte værdier (in Danish)*. Danish Building Research Institute, Aalborg University.
- Bøhm, B., F. Schröder, and N. C. Bergsøe (2009). *SBi 2009:10 Varmt Brugsvand – Måling af forbrug og varmetab fra cirkulationsledninger (in Danish)*. Danish Building Research Institute, Aalborg University.
- Booth, A. T. and R. Choudhary (2013). Decision making under uncertainty in the retrofit analysis of the uk housing stock: Implications for the green deal. *Energy and Buildings* 64, 292–308.
- Booth, A. T., R. Choudhary, and D. J. Spiegelhalter (2012). Handling uncertainty in housing stock models. *Building and Environment* 48, 35–47.
- Booth, A. T., R. Choudhary, and D. J. Spiegelhalter (2013). A hierarchical bayesian framework for calibrating micro-level models with macro-level data. *Journal of Building Performance Simulation* 6(4), 293–318.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2014). *Bayesian Data Analysis* (3 ed.). CRC Press.
- Higdon, D., M. Kennedy, J. C. Cavendish, J. A. Cafo, and R. D. Ryne (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing* 26(2), 448–466.
- International Organization for Standardization (2008). *ISO 13790:2008 - Energy performance of buildings - Calculation of energy use for space heating and cooling* (2 ed.).
- Jensen, J. M. and H. Lund (1995). *Design Reference Year, DRY - Et nyt dansk referenceår (in Danish)*. Technical University of Denmark. Announcement number 281.
- Kavgic, M., A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment* 45, 1683–1697.
- Kennedy, M. C. and A. O'Hagan (2001). Bayesian calibration of computer models. *Journal of Royal Statistical Society. Series B (Statistical Methodology)* 63(3), 425–464.
- Kristensen, M. H. and S. Petersen (2016). Choosing the appropriate sensitivity analysis method for building energy model-based investigations. *Energy and Buildings* 130, 166–176.
- Li, Q., G. Augenbroe, and J. Brown (2016). Assessment of linear emulators in lightweight bayesian calibration of dynamic building energy models for parameter estimation and performance prediction. *Energy and Buildings* 124, 194–202.
- Loga, T., B. Stein, and N. Diefenbach (2016). TABULA building typologies in 20 european countries – making energy-related features of residential building stocks comparable. *Energy and Buildings* 132, 4–12.
- Reinhart, C. F. and C. C. Davila (2016). Urban building energy modeling – a review of a nascent field. *Building and Environment* 97, 196–202.
- Rijal, H. B., P. Tuohy, M. A. Humphreys, J. F. Nicol, A. Samuel, and J. Clarke (2007). Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. *Energy and Buildings* 39(7), 823–836.
- Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment* 1(4), 407–414.
- Sokol, J., C. C. Davila, and C. F. Reinhart (2017). Validation of a bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings* 134, 11–24.
- Wittchen, K. B. and J. Kragh (2012). *SBi 2012:01 Danish building typologies – Participation in the TABULA project*. Danish Building Research Institute, Aalborg University.

4.2 Epilogue

An archetype model (denoted *building cluster* in the paper) for simulating annual heating energy use of SFHs constructed from 1978-1998 was calibrated using the KOH framework. The employment of the GP-based calibration framework in this study demonstrates a way to deal with the issue of archetype heterogeneity, by taking into account the heated floor area of the buildings when fitting the GP regression line that is used to tune the BEM input parameters. However, the heterogeneity assessment becomes one-dimensional when only taking into account the heated floor area as an explanatory variable, i.e. only heterogeneity originating from differences in building size is inferred. This one-dimensional assessment is obviously a crude simplification of the true variability between the 450 SFHs of the study; nonetheless, the analysis demonstrates a useful first take on how to approach the issue of archetype heterogeneity. It would be interesting in future work to extend the approach to include multiple regression dimensions to extract a larger portion of variability (heterogeneity) in the training buildings, which I believe would lead to more informed and accurate posterior distributions of the calibration parameters.

The calibration was conducted using training and validation data of annual resolution. It may seem strange to apply data of annual resolution in this study when the availability and granularity of building energy data to be used for modelling and calibration of UBEMs is increasing in general; however, the reality is that many modellers are still faced with annual or monthly data for the wider building stock without having access to the dynamic information contained in time series data of higher resolution (Table 1.3). Nonetheless, there is no doubt that such time series data is going to become generally available in the near future for even entire building stocks. A next step is therefore to incorporate time series data of a higher resolution to adequately address system dynamics and archetype heterogeneity. This issue is addressed in the following two chapters.

Chapter 5

The effect of temporal resolution on calibration

5.1 Motivation

The previous chapter focused on the presence of heterogeneity in archetype models using annual calibration data. In this chapter, the focus is on the temporal resolution in calibration data and how it influences the inferred posterior distributions of the calibration parameters, and the predictive accuracy of the tuned model.

The following paper (conference paper [P3]) was presented at the *CISBAT 2017 International Conference – Future Buildings & Districts* and later published in Elsevier journal *Energy Procedia*. It continues with the emulator-based KOH framework to assess the effects of varying training data resolution. Four different levels of temporal aggregation of hourly data are investigated in the calibration of six unknown BEM input parameters of a case study building: 1) 6-hourly, 2) daily, 3) weekly, and 4) monthly data. The same data period (one year of hourly DH meter readings) is used in the four scenarios, only the aggregation of the data is different. As argued for in the paper, the lowest temporal resolution applied is 6-hourly due to the high reading uncertainty (observation error) of the hourly values.



Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 122 (2017) 277–282

Energy

Procedia

www.elsevier.com/locate/procedia

CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale, CISBAT 2017 6-8 September 2017, Lausanne, Switzerland

Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution

Martin Heine Kristensen^{a,*}, Ruchi Choudhary^b, Steffen Petersen^a

^a*Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus, Denmark*

^b*Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom*

Abstract

Modern smart meters in heating systems offer building energy data of high temporal resolution. Compared to the annually aggregated readings used for conventional billing, the continuous information flow from these smart meters can be made available as time series data containing monthly, daily or even hourly aggregated values. In this paper, the effect of different temporal aggregation levels of commercial smart meter data on building energy model (BEM) calibration is investigated. Four different aggregation levels of a training data set were applied for calibration of six BEM input parameters to set up a Gaussian process emulator of the physical system. The performance of the emulator was subsequently tested on an unseen validation data set. Results reveal a systematic pattern of increasing predictive accuracy as a function of increasing training data resolution.

© 2017 The Authors. Published by Elsevier Ltd.

Peer-review under responsibility of the scientific committee of the CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale

Keywords: Bayesian calibration; building energy modeling; temporal data resolution; smart meter; detached single-family house

* Corresponding author.

E-mail address: mhk@eng.au.dk

1876-6102 © 2017 The Authors. Published by Elsevier Ltd.

Peer-review under responsibility of the scientific committee of the CISBAT 2017 International Conference – Future Buildings & Districts – Energy Efficiency from Nano to Urban Scale

10.1016/j.egypro.2017.07.322

1. Introduction

Setting up a valid building energy model (BEM) is often a difficult task, e.g. in the case of modeling an already existing building stock where one, to some extent, have to resort to conjecturing about construction details, type of materials and their state of condition. The uncertainty embedded in such BEMs may be dealt with by means of calibration [1, 2] where parameters are fitted to metered energy use data; however, depending on the level of uncertainty in the model, it can be difficult to find a deterministic *best fit* of model parameters. Probabilistic calibration is suitable for incorporating and quantifying this error, as one does not have to rely on single-value estimates, but can allow noisy data and the uncertainty of unidentifiable parameters to be propagated through the model.

In recent years, a plethora of Bayesian-based calibration techniques have been proposed and demonstrated, e.g. [3, 4, 5, 6, 7, 8, 9]. These references all offspring from the original emulator-based framework proposed by Kennedy and O'Hagan [10] and Higdon et al. [11] utilizing Gaussian process (GP) regression to match BEM evaluations with observed data by fitting calibration parameters. The emulator-based Bayesian approach enables probabilistic calibration of uncertain inputs for a given BEM using a limited number of evaluations from the building model. Implementing prior information about uncertain input parameters further enables the modeler to bias or even constrain the posterior inference – an option that can be reasonable to use when a limited amount of observed data is available for calibration. The calibration efficacy of the Bayesian framework has previously been studied under different levels of uncertainty [6] and different training set sizes [8]. However, it remains unclear how the temporal data resolution of the observed training set affects the posterior parameter inference and overall model accuracy.

In this paper, we investigate the issue of training data resolution by presenting a study on how different temporal resolution of metered district heating (DH) data affects the predictive accuracy of a BEM of a detached single-family house. We compare the posterior estimates of the calibration parameters and the predictive capabilities of the posterior model as measures of this effect. As such, this paper advances our understanding of how the temporal resolution of currently available DH smart meter read data affects calibration quality. This knowledge is valuable in many situations, for instance when modeling existing buildings for retrofit decision making under uncertainty, and for the future design of building energy management systems.

2. Methods

In the following subsection, we shortly outline the Kennedy and O'Hagan calibration formulation [10] in the context of BEM and point out changes made for the purpose of this study.

2.1. Emulator-based Bayesian calibration framework

The building-physical system used to generate $i = 1, 2, \dots, n$ observations of building energy use y_i at observed settings \mathbf{x}_i and unknown observation error $\varepsilon_{obs,i}$ is represented as

$$y_i = \zeta(\mathbf{x}_i) + \varepsilon_{obs,i} \quad i = 1, \dots, n, \quad (1)$$

where $\zeta(\mathbf{x}_i)$ denotes the true energy-consuming process. The observable setting \mathbf{x}_i consists of a p -dimensional vector of explanatory design points $\mathbf{x} \in \mathbb{R}^p$. In this study, we took $p = 2$ by letting x_1 index the outdoor air temperature and x_2 index the insolation. The inclusion of a Gaussian distributed noise-term $\varepsilon_{obs,i} \sim N(0, \sigma_{obs}^2)$ allowed for different observations of y at identical settings of \mathbf{x} , hereby accounting for the very stochastic nature of the energy-consuming process, e.g. occupant behaviour.

Using a BEM to represent the energy-consuming process, the observations were statistically modeled as

$$y_i = \eta(\mathbf{x}_i, \boldsymbol{\theta}) + \delta(\mathbf{x}_i) + \varepsilon_{obs,i} \quad i = 1, \dots, n, \quad (2)$$

where $\eta(\mathbf{x}_i, \boldsymbol{\theta})$ denotes evaluations of the BEM at the $p+q$ -dimensional input vector $(\mathbf{x}_i, \boldsymbol{\theta})$ comprising observed design points \mathbf{x}_i and additional calibrated parameters $\boldsymbol{\theta} \in \mathbb{R}^q$. In this study, we selected $q = 6$ BEM input parameters for calibration based on a Sobol sensitivity analysis on the model as demonstrated by Kristensen and Petersen [12]. They were *U-value (windows)*, *Infiltration@50Pa*, *Thermal mass*, *Heating temperature set point*, *Occupant density*

and *Appliances heat load*. Because no model is perfect, a stochastic model bias was introduced through $\delta(\mathbf{x}_i)$, independently of $\eta(\mathbf{x}_i, \boldsymbol{\theta})$, to account for discrepancy between the BEM and the true energy-consuming process $\zeta(\mathbf{x}_i)$. The *true* or *best achievable* calibration parameters $\boldsymbol{\theta}$ were inferred in the calibration process and thus unknown. We represented them in the simulations in terms of our prior estimates $\mathbf{t} \in \mathbb{R}^q$.

For a fixed set of $j = 1, 2, \dots, m$ BEM simulations, we obtained simulated data $y_j^* = \eta(\mathbf{x}_j^*, \mathbf{t}_j^*)$. Using the combined data set \mathcal{D} of n observed data points y , and m simulated data points y^* , a statistical surrogate model of η was constructed to emulate the BEM at untried input settings of (\mathbf{x}, \mathbf{t}) . For this purpose, a GP specified by a constant mean function at zero and a squared exponential covariance function was used. The bias term was likewise modeled as a GP to account for discrepancies between the emulator and the observed data. Formally, the two GPs were trained by simultaneously learning the posterior densities of their hyperparameters $\boldsymbol{\varphi}$ and the calibration parameters $\boldsymbol{\theta}$ that map the GPs to the observed data through the following regression formula:

$$\mathcal{D} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \\ y_1^* \\ \vdots \\ y_m^* \end{pmatrix} = \begin{pmatrix} \eta(\mathbf{x}_1, \boldsymbol{\theta}) + \delta(\mathbf{x}_1) + \varepsilon_{obs,1} \\ \vdots \\ \eta(\mathbf{x}_n, \boldsymbol{\theta}) + \delta(\mathbf{x}_n) + \varepsilon_{obs,n} \\ \eta(\mathbf{x}_1^*, \mathbf{t}_1^*) + \varepsilon_{num,1} \\ \vdots \\ \eta(\mathbf{x}_m^*, \mathbf{t}_m^*) + \varepsilon_{num,m} \end{pmatrix}. \quad (3)$$

To ensure numerical stability of the model, a small Gaussian distributed white-noise error-term $\varepsilon_{num,j} \sim N(0, \sigma_{num}^2)$ was added to the emulator. The multivariate Gaussian likelihood model of the data \mathcal{D} then became

$$\mathcal{L}(\mathcal{D} | \boldsymbol{\theta}, \sigma, \sigma_{num}, \boldsymbol{\varphi}) \propto \Sigma_{\mathcal{D}}^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathcal{D} - \mu \mathbf{I}_{n+m})^T \Sigma_{\mathcal{D}}^{-1} (\mathcal{D} - \mu \mathbf{I}_{n+m}) \right\}, \quad (4)$$

where the $(n+m) \times (n+m)$ covariance matrix $\Sigma_{\mathcal{D}}$ equals

$$\Sigma_{\mathcal{D}} = \Sigma_{\eta} + \begin{pmatrix} \Sigma_{\delta} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{I}_n \sigma_{obs}^2 & 0 \\ 0 & \mathbf{I}_m \sigma_{num}^2 \end{pmatrix}. \quad (5)$$

Applying Bayes theorem, the joint posterior distribution, conditional on the data, became

$$p(\boldsymbol{\theta}, \sigma, \sigma_{num}, \boldsymbol{\varphi} | \mathcal{D}) \propto \mathcal{L}(\mathcal{D} | \boldsymbol{\theta}, \sigma_{obs}, \sigma_{num}, \boldsymbol{\varphi}) \times p(\boldsymbol{\theta}) p(\sigma_{obs}) p(\sigma_{num}) p(\boldsymbol{\varphi}), \quad (6)$$

which was found using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. We ran four auto-tuning MCMC chains in parallel with randomly selected starting points until all parameters, and the combined log-likelihood, converged using a warm-up period of 50% of the chain length. Convergence was assessed using the potential scale reduction factor (PSRF) [13] that accounts for within-chain and between-chain variance, hereby simultaneously evaluating both the mixing and stationarity of the chains (convergence criterion: PSRF < 1.1).

2.2. Observed data

A randomly selected detached single-family dwelling constructed in 1992 in an urban area in Aarhus, Denmark, was used as case building. The total floor area is 173 m² divided between two floors. Of these, 92 m² make up the ground floor and the remaining 81 m² make up the first floor. The house has no basement or heated attic, it is made of brick walls with cement stone roofing, and uses district heating (DH) for space heating and domestic hot water (DHW) preparation. All known information about the building was obtained from the publically available Danish Building and Dwelling Register.

DH energy use for space heating and DHW preparation was gauged using a Kamstrup MULTICAL[®] smart heat meter with a manufacturer-calculated accuracy, cf. European standard EN 1434 of maximum $\pm 5\%$. The meter value displaying accumulated energy use in kWh was digitally read off once every hour (reading resolution = 1 kWh). Climate data consisting of the outdoor air temperature and the insolation was measured from a local weather station located within a 5 km radius from the building.

The hourly data was logged from January 1, 2015, until June 31, 2016. The entire year of 2015 was used as training data, and the first five months of 2016 as validation data. Four different aggregated temporal resolutions of the data

were considered: 1) six-hourly, 2) daily, 3) weekly and 4) monthly. The lowest resolution investigated was *six-hourly* and not *hourly* due to the low reading resolution of the data.

2.3. Simulated data

The building was modeled as a single-zone rectangular box with dimensions based on simple rules applied to known information about the floor area and number of floors, and our prior beliefs about the length-to-width-ratio and floor height (see [9] for further information). The four façades were assumed to face the four cardinal directions (North, East, South, West) while the unknown window area in each façade was allowed to vary based on our prior beliefs.

Energy use for space heating was modeled using the simple hourly dynamic model of ISO 13790:2008. The fully prescribed calculation method incorporates the thermal inertia of the building mass in hourly time steps by modeling the thermal resistances and capacitances, as well as the internal and solar heat gains, in an equivalent three-node resistance capacitance network (5RIC model). Energy requirements of the Danish Building Regulations in force at the time of construction (BR85), historical surveys of the Danish building stock and our educated guesses were used to set up prior estimates of the technical and occupation-related inputs required to run the model. Internal heat loads from occupants were modeled as a scalable day profile for the activity level constructed with a variable period without any presence, and ventilation was assumed a mix of infiltration and opening of windows. Airflow through the windows was modeled hourly as a percentage of maximum design airflow using a logistic transformation of a linear regression on the outdoor temperature [14]. Energy use for DHW was modeled using a simple linear model proportional to the consumed amount of hot water under the assumption that the rate of consumption was approximately constant throughout the year. For each of the four scenarios, the BEM was simulated $m = 500$ times.

3. Results and discussion

The posterior estimates of the six calibration parameters θ generally demonstrate the same behavior across the four scenarios with some variation present (Figure 1).

The largest overall deviation seems to be between the posteriors of the scenarios with daily and weekly values.

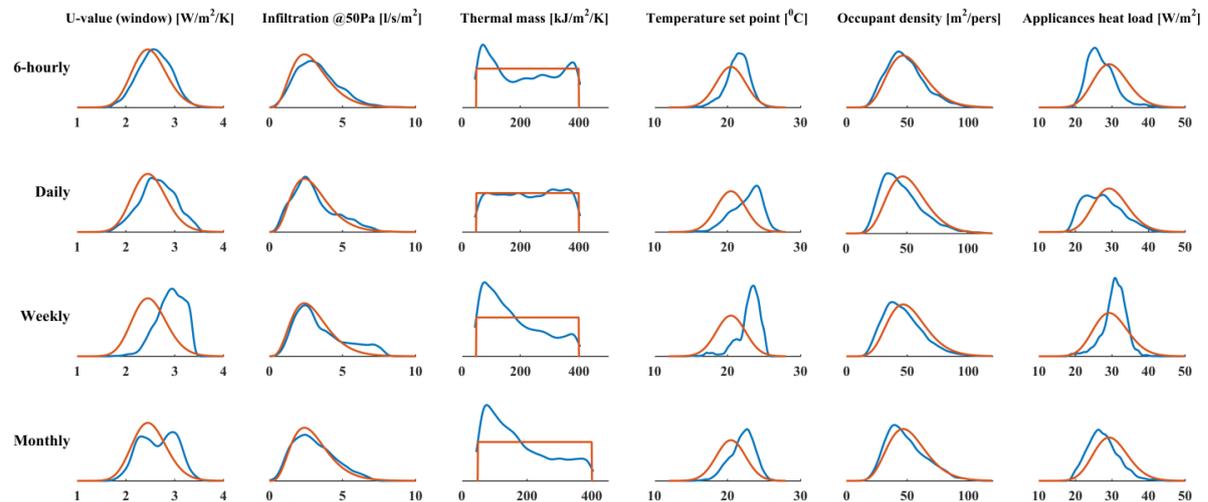


Figure 1. Prior (orange) and posterior (blue) probability densities of the six calibration parameters θ .

This *shift* in the posteriors is most obvious for the *Thermal mass* parameter, which we find more skewed for the two scenarios with lowest temporal resolution, but also parameters *U-value (window)*, *Temperature set point* and *Appliances heat load* seem to be affected. This effect may be explained by the fact that these two scenarios represent

the largest temporal difference (seven days per one week) compared to e.g. the difference between the scenarios with 6-hour values and daily values (three 6-hour values per one day).

As the posterior estimates of the hyperparameters ϕ – including the parameters governing the stochastic model bias $\delta(x_i)$ – reveal no noteworthy difference ($E[\delta(x_i)] \approx 0 \forall x_i$), we have omitted them from the paper. However, we must emphasise that the hyperparameters contain important information about the behavior of the GP emulator and hence influence the trueness of the six physics-based parameter estimates θ to some unknown degree.

The model fit (Table 1) of the GP emulators was assessed using four goodness-of-fit (GOF) measures: 1) the coefficient of determination (R^2), 2) the mean absolute percentage error (MAPE) and, following ASHRAE Guideline 14 [15], these were supplemented with 3) the normalized mean bias error (NMBE) and 4) the coefficient of variation of the root mean squared error (CVRMSE). For the training set, a scenario with data of monthly resolution ($n_{train} = 12$) obviously ensures a better fit than a scenario with 6-hourly resolution ($n_{train} = 1460$) as the model has to fit a much smaller data sample. However, in order to assess model fit justly and ensure comparability between the four calibration scenarios, GOF measures should be based on the predictive performance assessed from an unseen validation set of the same temporal resolution – here in the form of 6-hourly data from the first five months of 2016 ($n_{validation} = 500$). GOF measures of the validation set reveal a systematic pattern of increasing predictive accuracy as a function of increasing temporal resolution of the training data (Figure 2). The model trained using data of monthly resolution yields an NMBE of more than 10% in average, while the model trained using the same data, but with a temporal resolution of 6-hour values, yields an NMBE of less than 2.5% in average. Similar trends are seen for the other GOF measures. The results hereby indicate that the applicability of a model to perform out-of-sample predictions is highly influenced by the temporal resolution of the training data available. These results are obviously influenced by the case data at hand – especially dynamic phenomena such as occupant behavior and weather conditions. However, we find no arguments for not expecting similar trends and effects when using other data sets.

Table 1. Goodness-of-fit measures of the mean posterior emulator estimates.

Training set (January - December, 2015)						Validation set (January - June, 2016)					
Resolution	n	R ²	MAPE	NMBE	CVRMSE	Resolution	n	R ²	MAPE	NMBE	CVRMSE
6-hourly	1460	71.8 %	25.8 %	0.0 %	25.4 %	6-hourly	500	54.0 %	16.3 %	2.4 %	19.8 %
Daily	365	82.2 %	18.4 %	0.0 %	19.4 %	6-hourly	500	45.5 %	19.0 %	2.5 %	21.5 %
Weekly	53	91.1 %	11.0 %	0.4 %	13.0 %	6-hourly	500	29.7 %	19.4 %	9.6 %	24.5 %
Monthly	12	99.7 %	2.4 %	0.1 %	2.4 %	6-hourly	500	15.0 %	20.9 %	10.5 %	26.9 %

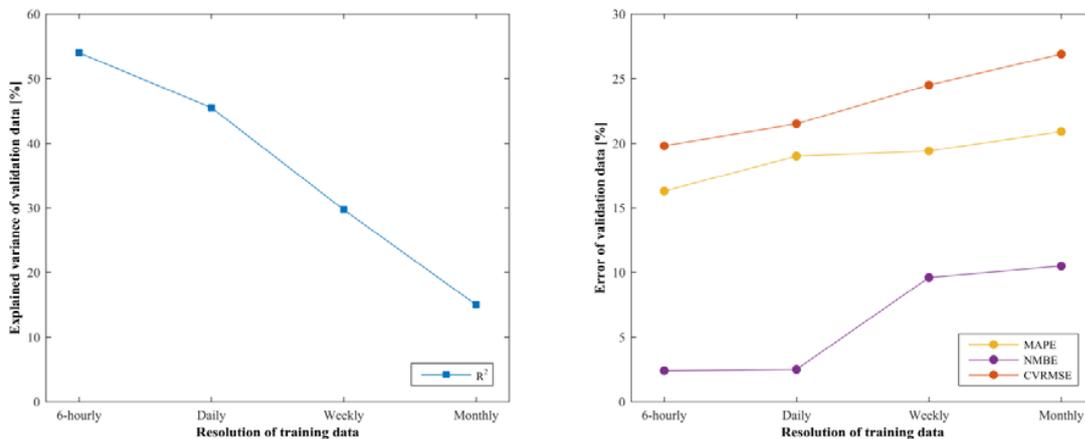


Figure 2. Goodness-of-fit measures of the four models using the validation set (6-hourly resolution).

4. Conclusion

A Bayesian emulator-based calibration of the ISO 13790 hourly dynamic building energy model was conducted using training data from four different temporal resolutions. Results suggest that the temporal aggregation level of training data has a significant influence on the calibration quality in terms of both the estimated parameter values and the predictive performance of the calibrated model. Information about important thermodynamic processes seems to be leveled out or even lost with decreasing temporal resolution of the training data. These findings support the promotion of commercial utility data meters with high temporal resolution (≤ 1 hour) and high reading resolution (0.1 kWh) for more accurate model calibration and parameter inference.

Acknowledgements

The research was conducted as part of the “Resource Efficient Cities Implementing Advanced Smart City Solutions” (READY) project, work package 3, financed by the 7th EU Framework Programme (FP7-Energy project reference: 609127). Furthermore, the authors would like to thank the district heating company in Aarhus, AffaldVarme Aarhus, for supplying the building energy data that forms the basis of the study.

References

- [1] D. Coakley, P. Raftery and M. Keane. A review of methods to match building energy simulation models. *Renew. Sustain. Energy Rev.* 37 (2014) 123-141.
- [2] E. Fabrizio and V. Monetti. Methodologies and Advancements in the Calibration of Building Energy Models. *Energies* 8 (2015) 2548-2574.
- [3] Y. Heo, R. Choudhary and G. A. Augenbroe. Calibration of building energy models for retrofit analysis under uncertainty. *Energy Build.* 47 (2012) 550-560.
- [4] A. T. Booth, R. Choudhary and D. J. Spiegelhalter. Handling uncertainty in housing stock models. *Build. Environ.* 48 (2012) 35-47.
- [5] A. T. Booth, R. Choudhary and D. J. Spiegelhalter. A hierarchical Bayesian framework for calibrating micro-level models with macro-level data. *J. Build. Perform. Simul.* 4 (6) (2013) 293-318.
- [6] Y. Heo, D. J. Graziano, L. Guzowski and R. T. Muehleisen. Evaluation of calibration efficacy under different levels of uncertainty. *J. Build. Perform. Simul.* 8 (3) (2015) 135-144.
- [7] Y. Heo, G. Augenbroe, D. Graziano, R. T. Muehleisen and L. Guzowski. Scalable methodology for large scale building energy improvement: Relevance of calibration in model-based retrofit analysis. *Build. and Environ.* 87 (2015) 342-350.
- [8] Y. Kang and M. Krarti. Bayesian-Emulator based parameter identification for calibrating energy models for existing buildings. *Build. Simul.* 9 (2016) 411-428.
- [9] M. H. Kristensen, R. Choudhary, R. H. Pedersen and S. Petersen. Bayesian calibration of residential building clusters using a single geometric building representation. In *Proceedings of Building Simulation 2017*, San Francisco, August, 2017.
- [10] M. C. Kennedy and A. O'Hagan. Bayesian Calibration of Computer Models. *J. Roy. Stat. Soc. Stat. Met. B* 63 (3) (2001) 425-464.
- [11] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo and R. D. Ryne. Combining field data and computer simulations for calibration and prediction. *SIAM J. Sci. Comput.* 26 (2) (2004) 448-466.
- [12] M. H. Kristensen and S. Petersen. Choosing the appropriate sensitivity analysis method for building energy model-based investigations. *Energy Build.* 130 (2016) 166-176.
- [13] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin. *Bayesian Data Analysis*. 3 ed., CRC Press, 2014.
- [14] H. B. Rijal, P. Tuohy, M. A. Humphreys, J. F. Nicol, A. Samuel and J. Clarke. Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. *Energy Build.* 39 (7) (2007) 823-836.
- [15] ASHRAE Guideline 14-2014 -- Measurement of Energy, Demand, and Water Savings. ASHRAE, 2014.

5.2 Epilogue

The paper suggests that the applicability of a model to perform out-of-sample predictions is highly influenced by the temporal resolution of the training data available. Moreover, it is seen that the inferred posterior distributions of the calibration parameters vary significantly with the temporal resolution applied for calibration; in particular, the thermal capacity of the ISO 13790 model is influenced, which is the only heat capacity in the model and thus controls all of its dynamic behaviour. A truthful identification of this parameter is therefore particularly important in order to perform accurate hourly predictions from the model. However, it is not possible to obtain such identification of the capacity using low-resolution data, for instance monthly metered consumption values, as the data lack the excitation that reveals dynamic behaviour.

With this realization in mind and to accommodate the need for data containing dynamic information in the analysis of archetype heterogeneity, we shall focus our attention on the application of hourly smart meter time series data in the following chapter, in contrast to the annually aggregated billing data applied in this and previous chapters. However, the calibration method applied so far relies on the GP-based calibration framework by KOH, including the inversion of covariance matrices on the input space to obtain posterior distributions of the calibration parameters. Dealing with model outputs that are time series extends the input space with an additional term for the time t ; however, this quickly increases the size of the covariance matrix to a point where it becomes computationally intractable, even for relatively short time series ([Higdon et al., 2008](#)). For the remaining part of the thesis, we shall therefore focus on other calibration methods, which apply a more direct evaluation of the time series output. However, it should be noted that alternative methods do exist for the inclusion of time series outputs in GP-based calibration of dynamic models that is fully capable of addressing the benefits of time series data. The interested reader may want to see the works by [Liu and West \(2009\)](#) and [Frigola-Alcalde \(2015\)](#).

Chapter 6

A hierarchical approach using time series data

6.1 Motivation

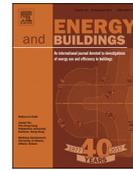
The previous two chapters on Bayesian calibration of archetypes applied the KOH emulator-based approach. An initial attempt to account for archetype heterogeneity was presented in Chapter 4 using annual training data and one-dimensional data variance. It was later shown in Chapter 5 that the temporal resolution of the training data has a significant impact on both posterior estimates of the calibration parameters, and on the predictive accuracy of the tuned model. Both studies rely on the application of the ISO 13790 hourly dynamic building energy model for calibration; however, to the best of my knowledge, it remains to be shown how to calibrate an archetype model using the ISO 13790 hourly dynamic model, or other dynamic model structures, with time series data. This would be needed to adequately account for dynamic behaviour, and at the same time account for difference in training buildings (archetype heterogeneity).

In the following paper (journal paper [P4]), published in Elsevier journal *Energy and Buildings*, a novel hierarchical archetype calibration methodology is proposed and demonstrated using case data from Aarhus. It employs a direct tuning of the building-physical archetype model itself by evaluating the differences between measured and simulated time series data. The application of direct tuning, in combination with a ISO 13790 building energy model structure, renders the use of the KOH emulator-based approach – and emulators is general – redundant. The direct calibration hereby eliminates a layer of numerical uncertainty in the calibration process. However, the novelty of the work primarily consists of the hierarchical structure itself, in which training buildings are set up to form a statistical archetype without relying on the training buildings being completely pooled together, i.e. archetype parameters are not formed by a direct average of individual building parameters. Instead, the hierarchical archetype model makes use of a statistical phenomenon known as *shrinkage*. Shrinkage is a process that occurs when individual building parameters are partially pooled to form a statistical compromise based on the assumption that they are exchangeable samples from a multivariate Gaussian distribution, in which the mean parameters are those of the archetype. This enables the inclusion of parameter covariance, i.e. correlation between input parameters, and variance across training buildings (archetype heterogeneity).



Contents lists available at ScienceDirect

Energy & Buildings

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

Hierarchical calibration of archetypes for urban building energy modeling

Martin Heine Kristensen*, Rasmus Elbæk Hedegaard, Steffen Petersen

Department of Engineering, Aarhus University, 8000 Aarhus, Denmark



ARTICLE INFO

Article history:

Received 23 April 2018

Revised 3 July 2018

Accepted 6 July 2018

Available online 17 July 2018

Keywords:

Archetypes

Building energy use

Hierarchical modeling

Multilevel modeling

Bayesian calibration

Prediction

Archetype homogeneity

Smart meter

ABSTRACT

The application of building archetypes is a widespread approach used in urban building energy modeling. Working with archetypes has a range of benefits, but it is important that modelers avoid using oversimplified approaches when establishing the archetype as they lead to loss of uncertainty and, consequently, to models with inferior predictive capabilities. In this paper, we propose a multilevel take on the challenge of establishing archetypes. A simultaneous modeling and calibration framework is formulated using Bayesian inference techniques – a technique that allows for the propagation of uncertainty throughout the calibration process. By means of hierarchical modeling, information from training buildings is partially pooled together to form an optimal solution between separate building energy models and a completely pooled model. This enables the inference of uncertain archetype parameters that are less prone to building outliers than what is achieved using ordinary aggregation of individual building estimates. The proposed framework incorporates dynamic building energy modeling of arbitrary temporal resolution where uncertain parameters are fitted for individual building models and the archetype model simultaneously. The application of the framework is demonstrated using case-study data from the Danish residential building stock, containing 3-hourly measurements of energy use for 50 training buildings. The model is tested for the prediction of 100 out-of-sample test buildings' aggregated energy use time series on a holdout validation period. With a prediction error of only $NMBE = 2.9\%$ and $CVRMSE = 7.8\%$, the archetype framework promises well for urban modeling applications.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

City governments, utility companies, and other energy policy stakeholders work on the urban scale of neighborhoods, cities, or even entire building stocks when planning and predicting the effect of various energy efficiency and production strategies. They are in need of tools and platforms that enable the analysis of aggregated effects rather than individual building-level effects.

Urban building energy modeling (UBEM) is a growing research field that seeks to facilitate such analyses by combining the effects of individual buildings into an aggregated urban model. The modeling approach of UBEM is either to model buildings independently and then aggregate their simulated energy use, or to model buildings collectively in an all-inclusive urban model with context-specific boundary conditions and interactive effects. Regardless of the modeling approach, the overall challenge of UBEM is to collect and assign all the necessary data inputs for establishing sufficiently detailed building energy models of all buildings in the

urban area without introducing too many assumptions and simplifications [1]. Because of this, the establishment of an accurate all-inclusive physics-based UBEM persists to be an extremely difficult task. However, one can make use of different techniques for reasonable tradeoffs between feasibility and accuracy to overcome this; of these techniques, the application of archetype models seems to offer an attractive solution.

1.1. Archetype modeling

The archetype approach seeks to reduce the number of buildings in a given building stock or urban area to a much smaller subset of homogeneous archetypes that represent groups of typologically identical buildings where information that would allow further differentiation is typically not available. This approach inevitably obscures the natural variability of occupant behavior and construction elements, but in turn reduces requirements for data acquisition and computational load.

The definition and use of building archetypes for urban-scale modeling have undergone a lot of work in recent years. In general, the literature describes the process of defining archetypes as consisting of three steps before simulation: (1) classification of build-

* Corresponding author.

E-mail address: mhk@eng.au.dk (M.H. Kristensen).

<https://doi.org/10.1016/j.enbuild.2018.07.030>

0378-7788/© 2018 Elsevier B.V. All rights reserved.

Nomenclature	
<i>Variables and parameters</i>	
A_m	effective area of thermal mass [m^2/m^2]
b_{ground}	temperature adjustment factor for building elements facing the ground [-]
C_m	capacity of thermal mass [$kJ/(m^2K)$]
h_{room}	room height [m]
H	heat loss coefficient (subscripts are used to define the element) [W/K]
f_{frame}	window frame fraction [-]
LWR	length-width ratio of building geometry [-]
n_b	number of buildings
n_s	number of simulations from the building energy model
n_t	number of simulation time steps
n_θ	number of calibration parameters
p	number of levels for Morris sensitivity analysis
$q_{in@50Pa}$	infiltration airflow at 50 Pa pressure difference [$l/(s m^2)$]
q_{vent}	ventilation airflow [$l/(s m^2)$]
r	number of trajectories for Morris sensitivity analysis
\hat{R}	potential scale reduction factor (abbrev.: PSRF) [-]
$SHGC$	solar heat gain coefficient [-]
T	temperature (index defines which temperature) [$^\circ C$]
U	heat transfer coefficient (index defines which element) [W/(m^2K)]
V_{occ}	hot water consumption of occupants [$m^3/(\text{pers. year})$]
\mathbf{W}	matrix of weather parameter inputs for building energy model (subscripts are used)
WFR	window-floor-ratio of building geometry [-]
\mathbf{X}	matrix of fixed parameter inputs for building energy model (subscripts are used)
\mathbf{y}	vector of measured time series energy use (subscripts are used) [kW]
\mathbf{y}^*	vector of simulated time series energy use (subscripts are used) [kW]
γ	archetype-level scale parameter in half-Cauchy distribution
ε	building-level error between measured and simulated energy use ($y-y^*$) [kW]
θ	vector of building-level calibration parameters (subscripts are used)
κ_0	a-priori number of "observations" of archetype-level means μ on the Σ scale
Λ_0	a-priori scale-matrix of the archetype-level covariance matrix Σ
μ	vector of means of the building-level parameters θ
μ_0	a-priori vector of mean values for archetype-level means μ
ν_0	a-priori degrees of freedom of the archetype-level covariance matrix Σ
Σ	covariance matrix of the building-level parameters θ
σ	standard deviation of building-level error ε [kW]
τ	vector of standard deviations of the building-level parameters θ
Φ_{app}	appliances, equipment and lighting heat load [W]
Φ_{DH}	district heating energy use ($\Phi_{DHW} + \Phi_{SH}$) [W]
Φ_{DHW}	energy use for domestic hot water [W]
Φ_{occ}	occupant heat load [W]
Φ_{SH}	energy use for space heating [W]
Φ_{sol}	solar radiation [W]
ω_0	a-priori upper boundary on archetype-level scale-parameter γ
<i>Indices and subscripts</i>	
b	indexing buildings $b = 1, 2, \dots, n_b$
i	indexing parameters for Morris sensitivity analysis
s	indexing simulations $s = 1, 2, \dots, n_s$
t	indexing time steps $t = 1, 2, \dots, n_t$
out-of-sample	quantity based on out-of-sample buildings
post	posterior quantity
pred	predictive quantity
train	quantity based on training data period
valid	quantity based on validation data period
within-sample	quantity based on within-sample buildings
<i>Abbreviations</i>	
BEM	building energy model (or "modeling")
BDR	building and dwelling register (Danish building and property database)
CVRMSE	coefficient of variation of the root mean squared error
DHW	domestic hot water
EPC	energy performance certificate
GIS	geographic information system
GPR	Gaussian process regression
MAP	maximum a-posteriori probability
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
NMBE	normalized mean bias error
PDF	probability density function
PSRF	potential scale reduction factor (symbol: \hat{R})
UBEM	urban building energy model (or "modeling")

ings into archetypes, (2) characterization of archetype parameters, and (3) calibration and validation of uncertain archetype parameters [2–6]. In many studies, the application of classifiers such as *usage type*, *construction year*, and *geometry* (e.g. surface-volume ratio) serves as segmentation parameters for clustering buildings into archetypes [2,7–9]. These simple segmentation parameters are often readily available from public databases such as geographic information systems (GIS) and tax and property registers. Parameter characterization, on the other hand, is often more difficult. Modelers can compile parameter data from a mixture of different information sources in an effort to draw a holistic picture of the archetype; however, the availability of data is very specific to the city, region or country in question. Moreover, as data access is often limited due to legal constraints and privacy considerations, data acquisition easily becomes a delicate compromise. A particularly difficult task is the description of occupant-related parameters that influence many aspects of building operation and thereby energy use. One can thus seldom expect to have data that is elaborate enough to obtain a purely deterministic description of the archetype parameters. Instead, modelers often have to resort to 'guessing', either by means of educated guesses, some kind of analysis of historical data, or through a stochastic treatment of uncertain data. For that reason, it is necessary to apply calibration to uncertain archetype parameters [4]. The literature holds a growing body of work on calibration methodologies for building energy models [10,11] of which the probabilistic calibration approaches, e.g. approaches based on Bayesian inference, have become increasingly popular in recent years [3,4,6,12–20]. As the whole concept of archetypes rests on a stochastic treatment of building data, it would then only seem natural to expand the probabilistic calibra-

tion methodologies to the scale of archetypes and building stocks. Nonetheless, there are only few attempts to do so [3,4,6,12,13].

1.2. Bayesian calibration of archetypes

In one of the earliest attempts at probabilistic calibration of archetypes by Booth et al. [12], a Bayesian framework was proposed for the calibration of a bottom-up physics-based archetype based on earlier work by Kennedy and O'Hagan [21]. They applied Gaussian process regression (GPR) for fitting four uncertain parameters of a quasi-steady-state building energy model (BEM) – the archetype model – to the daily building-averaged energy consumption data from 35 similar buildings, matching the archetype classification, over 61 winter days. Booth et al. [12] list several sources of uncertainty related to building stock modeling, which can be summarized to the following four types:

- *Parameter uncertainty*: Building-level variability due to insufficient knowledge about BEM input parameters. This also includes variability due to human behavior in terms of occupancy, operation of appliances, heating and cooling set point preferences, etc.
- *Structural uncertainty*: Variability due to the inadequacy of the BEM in describing the true energy consumption process of the building. This is also known as model bias or model discrepancy.
- *Archetype heterogeneity*: Archetype-level variability due to differences in building characteristics across the sampled buildings.
- *Numerical uncertainty*: Algorithmic variability due to numerical approximations, too small sample sizes, insufficient convergence of calibration, etc.

Booth et al. [12] explicitly addressed and incorporated *parameter uncertainty* by using the GPR technique to fit uncertain parameter distributions from a-priori uncertainty specifications, and to some degree *structural* and *numerical uncertainty* by including a statistical bias-correcting term. However, by fitting the calibration regression line to averaged building data, they did not account for *archetype heterogeneity*.

Kristensen et al. [22] also used the GPR technique, but on the annual heating energy use from a cluster of 450 similar residential buildings to estimate seven shared archetype parameters of a dynamic BEM. They fitted the calibration regression line to the non-averaged, building-specific training data, whereby they acknowledged the uncertainty due to differences in the sampled archetype buildings, i.e. *archetype heterogeneity*. This introduction of a disaggregated building-level likelihood assessment was important as archetype-aggregated data tends to average out much of the variability of the building stock, resulting in less informed posterior estimates.

In studies by Cerezo et al. [3,4,6], a new semi-Bayesian approach was proposed relying on an iterative error-analysis between dynamic BEM simulations and annually or monthly aggregated data, respectively. They used an upper limit for the simulated errors as a binary likelihood function to filter building-specific distributions for the calibration parameters for each building independently. The inferred building-specific parameter estimates were subsequently merged together into joint archetype-estimates to be used for prediction. By evaluating the likelihood of each building independently before combining the data, they implicitly accounted for archetype heterogeneity.

In agreement with the abovementioned sources of uncertainty, previous studies on the calibration efficacy of individual buildings have shown that the Bayesian framework is affected by the level of uncertainty in fixed and uncalibrated model parameters [23], the amount of training data [17], and the level of temporal aggregation

of calibration data [23,24]. Even though these findings apply to the calibration of individual buildings, they presumably also hold for the calibration of archetypes; however, this remains unaccounted for. A natural next step would thus be to investigate the effect of applying calibration data of high temporal resolution, e.g. hourly measurements of energy use. The application of high-resolution data, in combination with a detailed dynamic BEM, will most likely allow for a better estimation of dynamical parameters [23]. However, a binary likelihood function as proposed by Cerezo et al. [3,4,6] could prove to be too simple to fully exploit the information embedded in high-resolution data; for this end, a fully Bayesian continuous likelihood, e.g. Gaussian-distributed errors, would probably serve as a better 'filter'. Furthermore, a specific feature of estimating archetype parameters is the potential correlation between calibrated parameters. To our knowledge, this correlation has not previously been addressed in the literature of BEM calibration despite its importance for making accurate out-of-sample predictions of new buildings subscribing to the same archetype.

1.3. Contributions of this paper

In this paper, we present a new probabilistic archetype modeling and calibration framework where we use data from a number of observed training buildings to calibrate uncertain archetype parameters in a hierarchical setting. This feature allows the archetype calibration to draw strength from all training building datasets simultaneously, hereby exploring the true diversity of the archetype. In addition to the propagation of uncertainty throughout the calibration process due to the abovementioned sources of uncertainty, the proposed framework introduces the inclusion of four key features:

1. Dynamic physics-based building energy modeling;
2. Time series data and model outputs of arbitrary temporal resolution, e.g. hourly resolution;
3. Hierarchical/multilevel likelihood assessment of parameter proposals (both on building level and archetype level);
4. Calibration of both building-level and archetype-level parameters including their correlation.

We have organized the paper with an initial presentation of the proposed hierarchical framework in Section 2. All assumptions and necessary statistical definitions are laid out openly allowing readers to implement the model using their own data, if wanted. In Section 3, we demonstrate the application of the framework using an archetype case study from the Danish residential building stock. It is demonstrated how to draw inference about archetype parameters even though parameter values of the individual training buildings are not themselves observed, and how to perform out-of-sample predictions of unseen buildings matching the archetype definition. In Section 4, we provide a discussion on the applicability, limitations and possible future work for further optimization of the framework before we draw conclusions in Section 5.

2. Proposed archetype framework

We propose a hierarchical archetype modeling and calibration framework using a statistical formulation to describe the correlation between buildings that share the same archetype classification, as depicted in Fig. 1. At the archetype level, it is assumed that the buildings in the archetype exhibit exchangeable properties, meaning they can be seen as a sequence of exchangeable random variables sharing an underlying distribution function – i.e. a shared archetype description. Uncertain parameters at both building level and archetype level are then calibrated using measured energy use time series from training buildings at the building level

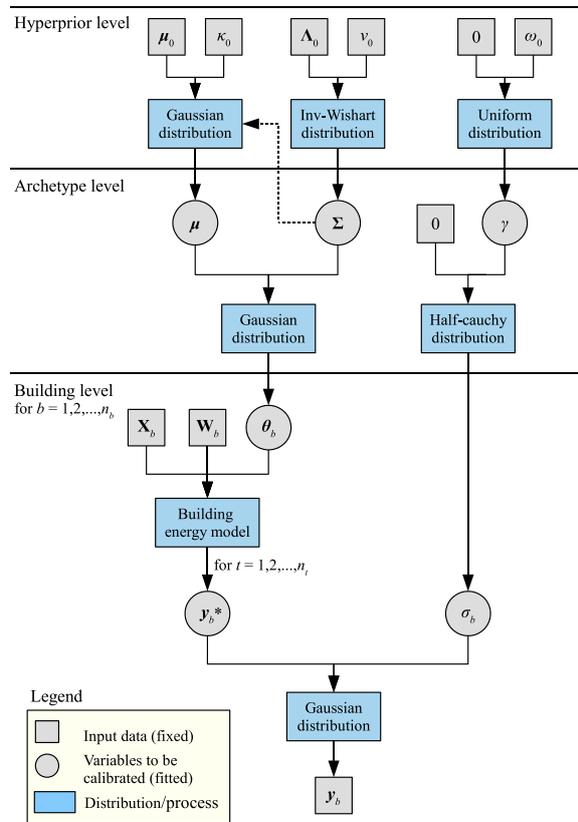


Fig. 1. Statistical representation of hierarchical framework.

in a Bayesian setting. This allows for user-specified prior information about their values to be taken into account at hyperprior level. The following sections describe the procedures of the framework in detail.

2.1. Building-level formulation

First, consider a building for which we observe a time series $\mathbf{y} = [y_1, y_2, \dots, y_{n_t}]^T$ of its energy use. We model the building using a physics-based building energy model (BEM), which we use to create a vector of n_t simulation outputs $\mathbf{y}^* = [y_1^*, y_2^*, \dots, y_{n_t}^*]^T$ matching the vector of observed data \mathbf{y} . The relation between observed and simulated energy use can be described as:

$$\mathbf{y} = \mathbf{y}^* + \boldsymbol{\varepsilon}, \quad (1a)$$

$$\mathbf{y}^* = \mathcal{M}(\mathbf{X}, \mathbf{W}, \boldsymbol{\theta}), \quad (1b)$$

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (1c)$$

where $\mathcal{M}(\mathbf{X}, \mathbf{W}, \boldsymbol{\theta})$ denotes the BEM evaluated on a matrix $\mathbf{X} \in \mathbb{R}^{n_t \times n_x}$ of observed and/or fixed building-specific input parameters, a matrix $\mathbf{W} \in \mathbb{R}^{n_t \times n_w}$ of relevant weather measurements, and a vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_{n_\theta}]^T \in \mathbb{R}^{n_\theta \times 1}$ of unknown parameters that we want to tune. The error-term $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n_t}]^T \in \mathbb{R}^{n_t \times 1}$ holds any residual variation between observations and simulations that cannot be decomposed as well as observation errors. We assume the errors to be independent and identical (i.i.d.) Gaussian

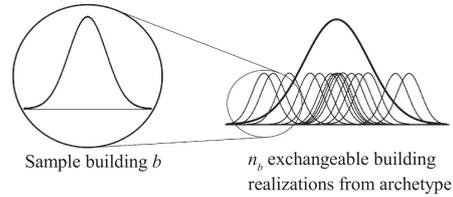


Fig. 2. Conceptual idea of archetype building realizations.

distributed and to exhibit homoscedasticity across time t such that $\text{var}(\varepsilon_t) = \sigma^2 \forall t$.

2.2. Archetype-level formulation

Now, consider $b = 1, 2, \dots, n_b$ buildings as described in Section 2.1, each of which we assume to be i.i.d. realizations from a single archetype building (Fig. 2). Once again, we underline that the process of segmenting building stocks and defining archetypes is not in focus here; we expect the modeler to have gone through this classification process already. For the archetype, we do not presume the sample buildings to be indistinguishable from each other, but whatever difference there is, it is due to random variation around a central archetype building.

For these n_b buildings, we employ a two-stage hierarchical formulation on the relationship between the exchangeable building datasets (Fig. 1). Formally, the first stage of the hierarchy – the building level – reparametrizes the abovementioned data model (1a)–(1c) using b to index individual buildings:

$$\mathbf{y}_b \sim \mathcal{N}(\mathbf{y}_b^*, \sigma_b^2 \mathbf{I}), \quad b = 1, 2, \dots, n_b. \quad (2)$$

We model the observed data \mathbf{y}_b of the b th building as a sample from the corresponding simulated data \mathbf{y}_b^* with i.i.d. random noise across time to account for residual error. The second stage of the hierarchy – the archetype level – defines the link between the n_b sample buildings, i.e. the interconnectivity of the unknown building-level calibration parameters $\boldsymbol{\theta}_b$:

$$\boldsymbol{\theta}_b \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad b = 1, 2, \dots, n_b. \quad (3)$$

In (3), the vector of n_θ calibration parameters $\boldsymbol{\theta}_b$ of the b th building is taken to be an i.i.d. sample from a multivariate Gaussian distribution with an unknown but shared vector of archetype mean parameters $\boldsymbol{\mu} \in \mathbb{R}^{1 \times n_\theta}$ and unknown covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n_\theta \times n_\theta}$ describing the variance of the calibration parameters across buildings and their internal correlation. It can be necessary to constrain or augment the calibration parameters $\boldsymbol{\theta}$ to ensure a realistic sampling distribution; for example, U-values, ventilation rates, thermal masses, etc. are all constrained to the positive domain only. By initially log transforming such calibration parameters, physical boundedness is ensured.

Moreover, in the second stage of the hierarchy, we employ an archetype-level formulation for the unknown standard deviation σ_b of the building-specific error vector $\boldsymbol{\varepsilon}_b$ to help pool the individual building estimates towards a common shared estimate. This is particularly useful for buildings with less well-behaved datasets, i.e. for datasets where unusual occupancy patterns, among other things, would otherwise result in larger errors and thereby hinder inference about the calibration parameters. The application of the half-Cauchy⁺ distribution (4) is a weakly informative choice for the distribution of the σ_b s that is centered on zero with a heavy tail towards infinity governed by the scale parameter γ [25]:

$$\sigma_b \sim \text{halfCauchy}^+(0, \gamma), \quad b = 1, 2, \dots, n_b. \quad (4)$$

For individual building datasets obtained using the same data collection methods and under similar general conditions, the employment of a hierarchical pooling of the error standard deviations

σ_b seems reasonable. However, in many cases, the unpredictable and stochastic nature of occupancy may preclude the assumption of exchangeable error-terms across buildings. In such cases, we might abandon the hierarchical pooling (4) in favor of separate building estimates.

2.3. Calibrating archetype parameters

The hierarchical structure describes a combined sampling distribution for the data $\mathbf{y}_{1:n_b}$. However, for the purpose of parameter calibration, we are more interested in viewing the model as a function of the data. The *likelihood* formally sums up all the data as a function of the free, i.e. unknown, parameters of the statistical model and hence describes the plausibility (probability density) of proposed parameter values given the data. The likelihood of the b th building dataset containing $t = 1, 2, \dots, n_t$ data points is a Gaussian probability density function due to the assumption of Gaussian distributed errors (1c):

$$p(\mathbf{y}_b | \boldsymbol{\theta}_b, \sigma_b) = \frac{1}{(2\pi\sigma_b)^{n_t/2}} e^{\left\{ -\frac{1}{2\sigma_b^2} \sum_{t=1}^{n_t} (y_{b,t} - y_{b,t}^e)^2 \right\}}, \quad (5a)$$

The simulated data \mathbf{y}_b^e is obtained through the BEM (1b), which is only a function of $\boldsymbol{\theta}_b$ since we consider the observed input parameters of \mathbf{X} and \mathbf{W} to be fixed. Implementing the full hierarchical structure to fit the entire dataset of $b = 1, 2, \dots, n_b$ buildings, the combined data likelihood, conditional on the hierarchical model, becomes the product of n_b Gaussian likelihoods:

$$p(\mathbf{y}_{1:n_b} | \boldsymbol{\theta}_{1:n_b}, \sigma_{1:n_b}) = \prod_{b=1}^{n_b} \frac{1}{(2\pi\sigma_b)^{n_t/2}} e^{\left\{ -\frac{1}{2\sigma_b^2} \sum_{t=1}^{n_t} (y_{b,t} - y_{b,t}^e)^2 \right\}}. \quad (5b)$$

The likelihood function only describes the dependence of the data on the immediate data-level parameters. However, as we are not specifically interested in these quantities, but more in the archetype-level parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, γ that indirectly affect the data through the hierarchical structure, we thus expand the likelihood function by multiplying the immediate data-level likelihood (5b) with the probability density of the building-level parameters ((3)–(4)) to obtain what we could call the joint likelihood function:

$$p(\mathbf{y}_{1:n_b} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma) = p(\mathbf{y}_{1:n_b} | \boldsymbol{\theta}_{1:n_b}, \sigma_{1:n_b}) p(\boldsymbol{\theta}_{1:n_b} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\sigma_{1:n_b} | \gamma). \quad (6)$$

From the left-hand-side of (6), it is now evident that we do not need to consider the values of individual building-level parameters to draw an inference about the archetype-level parameters as the conditional dependency is accounted for. To infer the values of the unknown archetype-level parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, γ , we could apply the method of maximum likelihood (ML) estimation to approximate the *most likely* values (point estimates) given the data. However, as the ML estimate may be seen as a special case of the Bayesian maximum a-posteriori probability (MAP) estimate that allows for a more thorough treatment of uncertainties, we apply the more general Bayesian approach here [26]. In a Bayesian context, the *true*, i.e. calibrated, parameter distributions *after* seeing the data are referred to as posterior distributions following Bayes theorem. The posterior probability density of the parameters in the context of this model is:

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma | \mathbf{y}_{1:n_b}) \propto p(\mathbf{y}_{1:n_b} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma) p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma), \quad (7)$$

where $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma | \mathbf{y}_{1:n_b})$ is the joint posterior density of the archetype-level parameters conditional on the data, $p(\mathbf{y}_{1:n_b} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma)$ is the joint data likelihood conditional on the model and parameters as given in (6), and $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \gamma)$ is the

joint prior density of the archetype-level parameters. To fulfill the hierarchical model formulation in a Bayesian context, we thus need to specify prior PDFs that reflect our subjective beliefs about the unknown archetype-level parameters *before* seeing the data – illustrated as the hyperprior level in Fig. 1. Different options of priors are available for the mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ of the multivariate normal distribution of the $\boldsymbol{\theta}_b$ s (3); we make the convenient choice of using the conjugated prior – the normal-inverse-Wishart distribution [26] – that ensures the posterior to be multivariate normal as well:

$$\boldsymbol{\Sigma} \sim \text{InvWishart}(\boldsymbol{\Lambda}_0^{-1}, \nu_0), \quad (8a)$$

$$\boldsymbol{\mu} | \boldsymbol{\Sigma} \sim \text{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma} / \kappa_0), \quad (8b)$$

$$\gamma \sim \text{U}(0, \omega_0), \quad (8c)$$

where the parameters ν_0 and $\boldsymbol{\Lambda}_0$ represent the degrees of freedom and the scale matrix, respectively, of the inverse-Wishart distribution on $\boldsymbol{\Sigma}$, and the parameters $\boldsymbol{\mu}_0$ and κ_0 represent the prior mean vector and the number of prior observations on the $\boldsymbol{\Sigma}$ scale. Setting $\nu_0 = n_\theta + 1$ and $\boldsymbol{\Lambda}_0 = \mathbf{I}$ (identity matrix) results in a weakly informative prior where each correlation parameter in $\boldsymbol{\Sigma}$ has a marginal uniform prior distribution. Likewise, setting κ_0 to a low number, e.g. 1–10 depending on the number of building datasets n_b being modeled, results in less weight being given to the chosen prior mean values $\boldsymbol{\mu}_0$. For the scale γ of the half-Cauchy distribution (4) that controls the level of pooling of the building-specific error standard deviations σ_b , we apply a uniform prior distribution constrained to the positive domain (8c). Setting ω_0 to a high number relative to the scale of the data ensures a data-driven inference [25].

Employing a straightforward Markov chain Monte Carlo (MCMC) algorithm, e.g. the Metropolis-Hastings algorithm [25], can simulate the joint posterior distribution of the model, and subsequently, the marginal posterior distributions of the individual parameters. We can address the inferred values of calibrated parameters in terms of either the individual MAP estimates, or by using the full posterior probability distribution, effectively retaining all model uncertainty.

Employing an MCMC algorithm to make inferences about the uncertain parameters (inferring their posterior distribution) requires the evaluation of the joint data likelihood (6) and prior distributions (8a)–(8c) thousands of times. Evaluating the likelihood of a given parameter proposal easily becomes computationally inconvenient and even practically impossible for larger datasets, i.e. many buildings and/or many time series data points. This is due to the small probabilities that arise, which often cause numerical instability. The widespread approach of using log-probabilities is therefore strongly encouraged to improve both numerical stability and accuracy (stability is ensured for small probabilities), as well as computational speed (addition in the log-domain is less expensive than multiplication in the arithmetic domain).

2.4. Predictive performance

As is typical in Bayesian data analysis, predictions/forecasts from the hierarchical model are based on the posterior predictive distribution of the parameter of interest, e.g. the predictive distribution of an energy use time series $p_{\text{pred}}(\mathbf{y})$. We hereby aim to report inferences about future predictions of building energy use in such a way that the full uncertainty over \mathbf{y} from all layers of the hierarchical model is accounted for throughout the analysis.

We can easily forecast energy use time series from the existing $b = 1, 2, \dots, n_b$ buildings used to calibrate the archetype by generating random draws from the posterior distributions of the trained

Table 1
Archetype classification applied in demonstration case study.

Segmentation parameter	Value
Usage/Type	Detached single-family house
Construction period	1950–1959
Location	Aarhus, Denmark
Number of stories above ground	1
Basement	No
Attic utilized for living	No
Heating source	District heating (space heating and DHW)
Suppl. heating installations	No

building-specific calibration parameters $p_{\text{post}}(\theta_b)$ and subsequently apply these in the BEM (1b) with existing or forecasted weather conditions to generate $y_b | X_b, W_{\text{new}}, \theta_b$. In general, however, we are more interested in predicting the consumption of other buildings than those used to calibrate the archetype, i.e. we want to predict the performance y_{new} for a new set of building-specific parameters X_{new} . In this case, we have no posterior distribution of the calibration parameters. Instead, we sample the parameters from their predictive distribution, i.e. the archetype posterior distributions of the calibration parameters:

$$p_{\text{pred}}(\theta_{\text{new}}) = N(\mu_{\text{post}}, \Sigma_{\text{post}}). \quad (9)$$

We then apply the BEM (1b) to generate predictions $y_{\text{new}} | \theta_{\text{new}}, W, X_{\text{new}}$ for new buildings using random draws from $p_{\text{pred}}(\theta_{\text{new}})$ in combination with the fixed building-specific parameters X_{new} and existing or forecasted weather conditions as input. Repeating this sampling many times allows us to construct an empirical distribution of the posterior predictive space of time series predictions $p_{\text{pred}}(y_{\text{new}} | \theta_{\text{new}}, W, X_{\text{new}})$ for that specific building and those weather conditions.

3. Demonstration: Danish detached single-family dwellings from the 1950s

In this section, we demonstrate how energy use time series data from a limited number of clustered training buildings can be used for identifying a shared archetype model using the archetype calibration framework described in Section 2. We do not focus on the clustering process itself, i.e. the archetype definition and classification, but rather on quantifying the embedded variability (homogeneity) in an arbitrary archetype definition selected for the purpose of this case study demonstration, and how this archetype model may be used for prediction. The predictive capabilities of the calibrated archetype are validated against each individual training building on a new holdout validation period (within-sample prediction), and against new unseen test buildings that have not been used for training the archetype (out-of-sample prediction).

3.1. Archetype classification and case data

An archetype was defined covering Danish one-storied detached single-family dwellings constructed in the 1950s (full archetype classification given in Table 1).

3.1.1. Building data

A publicly available database containing building and property characteristics for the Danish building stock – the Building and Dwelling Register (BDR) – was used to identify buildings matching the archetype definition. We filtered the BDR information on the entire residential building stock of Aarhus, Denmark (approx. 80,000 buildings) using the segmentation parameters in Table 1. A pool of 2,775 buildings matching the archetype description was obtained. Of these 2,775 potential archetype buildings, we randomly selected 50 training buildings for the purpose of archetype

calibration, and an additional 100 test buildings for archetype performance testing (out-of-sample validation). Besides the segmentation parameters listed in Table 1, the only additional parameter from the BDR database used for setting up the building energy model was the “heated floor area”.

3.1.2. Time series of building energy use

The local district heating supplier, AffaldVarme Aarhus, supplies most of the city of Aarhus with heat for hydronic space heating and/or on-site domestic hot water (DHW) preparation through its underground distribution system. They began the replacement of their old heat meters with new modern smart meters (Kampstrup Multical® heat meters) in all of its consumer units throughout the city in 2015 and finished in 2017. The accumulated district heating consumption of each building is now digitally read off its smart meter once every hour using a remote reading system and logged by the utility in truncated kWh (an actual consumption of 6,529.999 kWh is reported as 6,529 kWh). No energy is “lost” during the meter reading process though; only the precise temporal fixation of the decimals on individual hourly readings remains unaccounted for. Consequently, minor parts of the energy consumption may be shifted one or two hours forward in time. We were given access to these hourly logged time series readings from the 150 archetype buildings for the purpose of this study. All data was subsequently anonymized.

The reading uncertainty was negligible for the large accumulated meter values, but it was substantial for the hourly differences that were in the order of 0 kWh to 5 kWh (difference between two successive accumulated hourly readings). The reading uncertainty (0 kWh \leq reading uncertainty < 1 kWh) could potentially amount to as much as 100% of the hourly values in hours with low or no consumption and thus hinder an efficient identification of uncertain building parameters. The effect of reading uncertainty is reduced by reducing the temporal resolution of the data from hourly values into e.g. 2-hourly, 3-hourly, 6-hourly, or 12-hourly values, etc. but at the cost of data resolution, and consequently, the ability to account for model dynamics [24]. We made a compromise and aggregated all hourly values into 3-hourly values (difference between every third accumulated meter reading).

3.1.3. Weather data

A weather file was compiled from hourly measured values of air temperature and global horizontal irradiance from a local weather station in the city of Aarhus located within a 15 km radius from all training buildings.

3.1.4. Training and validation data periods

To identify the thermodynamic behavior and characteristics of a building, one needs response data captured during transient conditions that excite the dynamics of the building sufficiently [27]. Although we selected training and validation periods (Table 2) in which external loads varied substantially ($-9^\circ\text{C} <$ outdoor air temperature $< +9^\circ\text{C}$), we had no prior knowledge about variations in internal loads, nor the actual heating set points. Therefore, estimates of the dynamic properties remain subject to a high degree of uncertainty.

All building datasets were initially scrutinized for missing energy use data in the training and validation periods. Buildings with missing data in this two-month period were not accepted in the analysis, but instead replaced with a new, randomly sampled building from the city with associated district heating readings in accordance with the archetype classification (Table 1). This was also the case for buildings with erroneous BDR data, i.e. a “negative” or “zero” heated area, and for buildings missing construction year.

Table 2
Size and temporal resolution of data.

	Training buildings (within-sample)	Test buildings (out-of-sample)
Sample size, n_b	50 buildings	100 buildings
Training period	1.1.2017–31.1.2017 ($n_t = 248$ 3-hourly values)	Out-of-sample buildings are not trained
Validation period	1.2.2017–29.2.2017 ($n_t = 248$ 3-hourly values)	1.2.2017–29.2.2017 ($n_t = 224$ 3-hourly values)

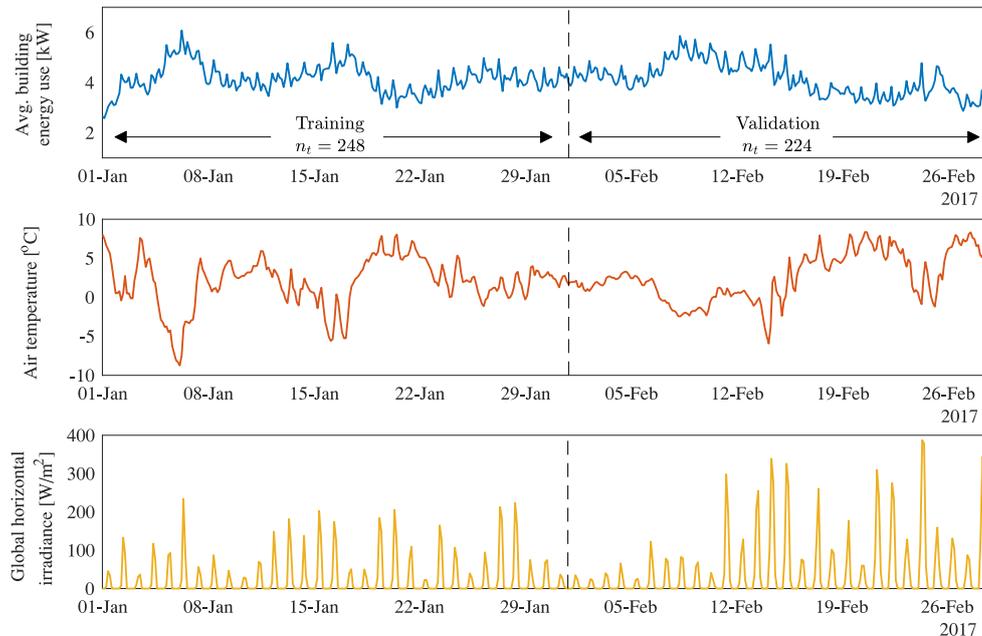


Fig. 3. Measured time series data. Top: Averaged building energy use for all $n_b = 50$ within-sample training buildings. Middle: Outdoor air temperature. Bottom: Global horizontal irradiance. Only data from the training period was used for calibration.

The averaged energy use time series from all $n_b = 50$ training buildings along with corresponding weather data (air temperature and global horizontal irradiance) are shown in Fig. 3 for the training and validation periods.

3.2. Building energy model

The hourly measured district heating energy use (Φ_{DH}) was modeled using two separate models; one for space heating (Φ_{SH}) and one for DHW (Φ_{DHW}), which were subsequently added so that $\Phi_{DH} = \Phi_{SH} + \Phi_{DHW}$. The following sections explain the models for Φ_{SH} and Φ_{DHW} in detail.

3.2.1. Building geometry

The actual and detailed geometric layout of the individual buildings was unknown. We therefore applied a general and scalable geometric representation to be used for all buildings, similar to what was used in Kristensen et al. [22]. The geometric layout consisted of a rectangular box (Fig. 4) with dimensions based on simple rules applied to known information about the floor area (information from BDR), our a-priori beliefs about the length-to-width-ratio (LWR), and the floor height of a typical SFH (Table 3).

The building facades were assumed to face the four cardinal directions. As the total window area was unknown, and because all buildings varied in size, we modeled the total window area as a scalable proportion of the floor area (window-to-floor ratio, WFR). Moreover, we fixed the partitioning of the total window area on the four facades following Danish standard calculation pro-

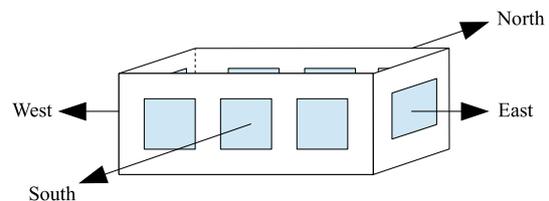


Fig. 4. Geometric model of archetype building.

cedures (North = 26%; South = 41%; East/West = 16.5%) for single-family dwellings [28].

3.2.2. Space heating

Energy use for space heating Φ_{SH} was modeled using a slightly modified version of the hourly dynamic calculation method described in ISO 13790:2008 (Fig. 5). The building was treated as a single thermal zone and the thermal inertia of the building was accounted for by modeling thermal resistances and the effective thermal capacity, as well as the internal and solar heat gains in an equivalent three-node resistance capacitance network (5R1C). The supply air temperature of the ventilation was assumed identical to the external air temperature as the ventilation principle of Danish dwellings from the construction period of the archetype usually is natural ventilation, i.e. infiltration, manual opening of windows, window valves.

Table 3

List of user-specified values for 20 uncertain model input parameters. The top 10 most influential parameters were identified using the Morris sensitivity screening method [29], assuming a uniformly distributed prior data range (max/min).

Uncertain model parameter	Sensitivity analysis			Selected value
	Min. value	Max. value	Result (top 10) *	
<i>Geometry</i>				
Length-width ratio, LWR [-]	0.10	1.00	X	0.50
Room height, h_{room} [m]	2.30	3.00		2.60
Window-floor ratio, WFR [-]	0.10	0.50	X	Calibrated
Window frame fraction, f_{frame} [-]	0.10	0.50		0.25
<i>Transmission</i>				
Temp. adjustment factor (ground), b_{ground} [-]	0.50	1.00		0.70
U -value (floors) [W/(m ² K)]	0.10	0.50	X	0.30
U -value (walls/roof) [W/(m ² K)]	0.10	0.50	X	Calibrated
U -value (windows) [W/(m ² K)]	1.00	5.00	X	1.60
Solar heat gain coef., $SHGC$ [-]	0.50	0.70		0.60
Capacity of thermal mass, C_m [kJ/(m ² K)]	50	600	X	Calibrated
Effective mass area, A_m [m ² /m ²]	3.00	5.00		2.5–3.5**
Heat conduction (mass) [W/(m ² K)]	8.50	10.00		9.10**
Heat transfer coef. (surf.-air) [W/(m ² K)]	2.00	5.00		3.45**
<i>Ventilation</i>				
Infiltration airflow, $q_{inf@50 Pa}$ [l/s/m ²]	0.10	8.00	X	Calibrated
Design airflow (nat. ventilation), q_{vent} [l/s/m ²]	0.10	2.00	X	0.30
<i>Occupation</i>				
Occupant density, $Occ.Density$ [m ² /pers.]	10	150	X	Calibrated
Room heating set point, $T_{set,H}$ [°C]	18.0	25.0	X	21.5
<i>Domestic hot water</i>				
DHW flow temperature, T_{DHW} [°C]	40.0	60.0		55.0
Mains temperature, T_{mains} [°C]	5.0	15.0		10.0
Hot water consumption, V_{occ} [m ³ /pers./year]	10	20		15

*Top 10 most influential parameters according to the Morris method [29].

**Parameter values are defined in ISO 13790:2008.

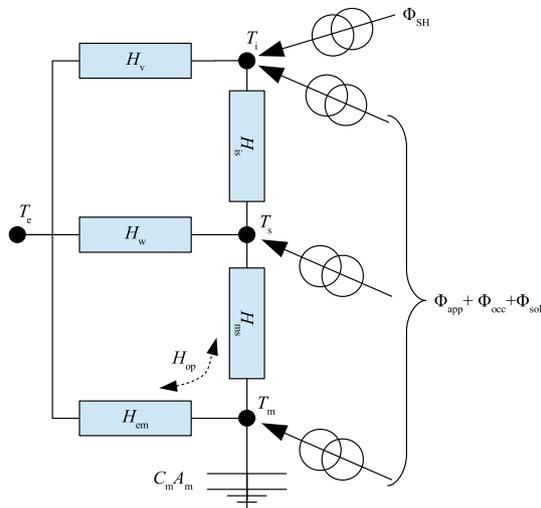


Fig. 5. Space heating model. Modified version of the network presented in ISO 13790:2008. The area of the building elements A is used to transform resistances R into heat transfer coefficients $H = A/R$.

The space heating model (Fig. 5) contains three internal temperature nodes: room air temperature T_i , surface temperature of the thermal mass T_s , and internal temperature of the thermal mass T_m . The nodes are interconnected by three heat transfer coefficients describing the building elements (opaque envelope ele-

ments H_{op} , windows H_w , and ventilation H_v) and two coefficients describing the heat transfer between the indoor air, indoor surfaces, and thermal mass (H_{is} and H_{ms}). Finally, the thermal mass of the building is governed by the capacity C_m . Several of the parameters, e.g. T_s , T_m , and H_{em} , only exist as provisional values in the algorithm of the ISO 13790:2008 calculation method and thus did not need any prescribed value. Other parameters, mainly heat transfer coefficients between building elements, are defined in the standard. User-specified inputs were necessary for the remaining building-specific parameters. As no information was available about U -values, $SHGC$, ventilation airflows, etc., their values were based on the Danish building code in force at the time of construction, historical surveys of the Danish building stock, and our a-priori beliefs (Table 3). Ventilation was assumed a mix of infiltration and opening of windows; no mechanical ventilation was modeled. Based on studies by Rijal et al. [30], the airflow through windows was modeled hourly as a percentage of maximum design airflow using a logistic regression on the outdoor temperature T_e :

$$P_{airflow} = \frac{e^{(0.25T_e - 0.25)}}{1 + e^{(0.25T_e - 0.25)}} \quad (10)$$

The RC network (Fig. 5) is exposed to external boundary conditions in terms of the outdoor air temperature T_e , solar radiation Φ_{sol} , internal heat loads from equipment/lighting Φ_{app} , occupant heat Φ_{occ} , and space heating Φ_{SH} delivered from the district heating system. Outdoor air temperature T_e and solar irradiation Φ_{sol} were specified using measured weather conditions (Section 0; Fig. 3). The hourly average of Φ_{app} was estimated using a regression model based on statistical data of annual electricity consump-

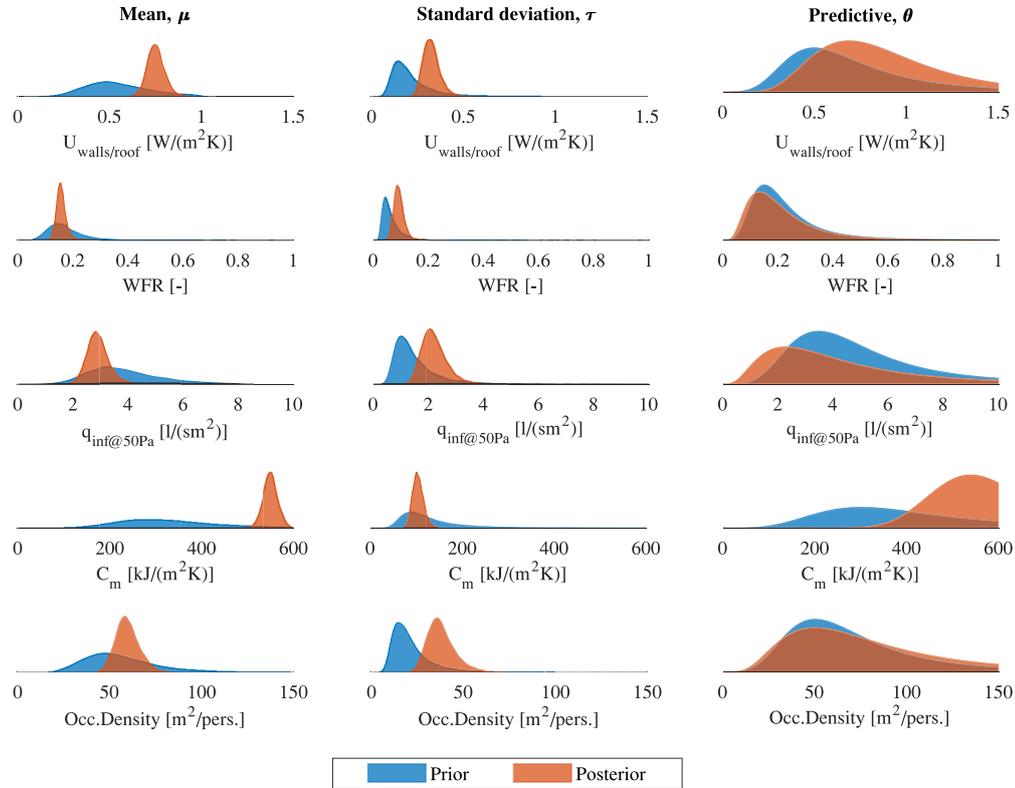


Fig. 7. Marginal prior and posterior distributions of the five archetype-level calibration parameters. The calibration parameter mean values μ are shown in the left-hand column, their decoupled standard deviations τ are shown in the middle column, and their predictive distributions are shown in the right-hand column.

$\hat{R} \in \mathbb{R} | 1 \leq \hat{R} < \infty$ [26]. The PSRF measures how much the scale of the variations in the inferred parameter distributions might have been reduced if the number of draws simulated by the MCMC-algorithm approached infinity, $\lim_{n \rightarrow \infty} (\hat{R} \rightarrow 1)$. By accounting for the within-chain to between-chain variance in the warm chains, the PSRF evaluates both the mixing and stationarity of the chains simultaneously. A stable and converged solution was considered for any given parameter estimation when $\hat{R} < 1.1$.

As it is infeasible to visualize the posterior distributions of the five calibration parameters θ for all 50 training buildings (first level of the hierarchy), we only show the shared archetype-level parameters μ , Σ , γ here (second level of the hierarchy). Furthermore, we have decomposed the covariance matrix Σ into parameter specific variance $\tau^2 = \text{var}(\theta)$ and correlation coefficients $\rho = \text{corr}(\theta)$ for the purpose of visualization. The prior and posterior distributions of the five archetype mean values μ are displayed in Fig. 7 (left-hand column), while their corresponding standard deviations τ are shown in Fig. 7 (middle column). In the right-hand column in Fig. 7, the mean and variances have been applied to draw the predictive distributions $p_{\text{pred}}(\theta_{1:n_\theta}) = N(\mu_{1:n_\theta}, \tau_{1:n_\theta}^2)$ of the five calibration parameters themselves. However, it is important to note that these predictive distributions are marginal and thus do not account for the modeled correlation ρ (Fig. 8) between the parameters, which must be taken into account through the covariance matrix Σ when predicting coherent values of the parameters $p_{\text{pred}}(\theta) = N(\mu, \Sigma)$ for new buildings.

Setting off from broad and weakly informative prior distributions, the data has successfully focused the posteriors of the means $p_{\text{post}}(\mu)$ and to some extent the standard devia-

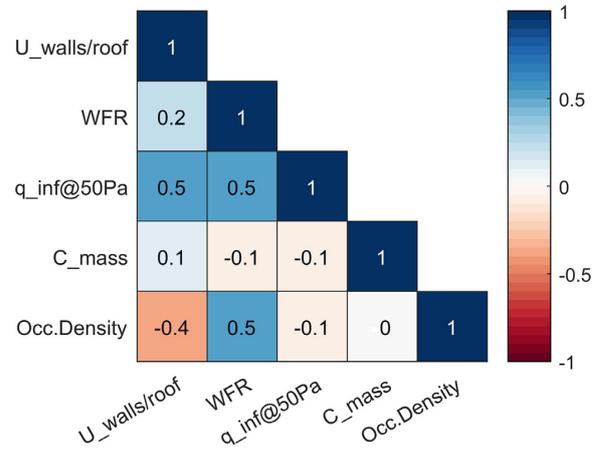


Fig. 8. Expected posterior correlation matrix, $E_{\text{post}}[\text{corr}(\theta)]$, of the five calibration parameters.

tions $p_{\text{post}}(\tau)$. The posteriors of μ_{WFR} , $\mu_{\text{qinf@50Pa}}$ and $\mu_{\text{Occ.Density}}$ are focused near the expected value of their respective priors, whereas for $\mu_{\text{Uwalls/roof}}$ and μ_{Cmass} , the calibration has drawn their posteriors out into the tails towards their maximum values (Table 3). Common for all five posteriors of the standard deviations $p_{\text{post}}(\tau)$ is that the distributions are weakly informative and

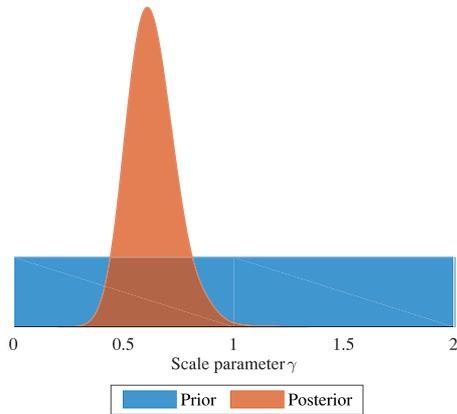


Fig. 9. Prior and posterior distribution of the archetype-level scale parameter γ that controls scale of the error-term standard deviations σ .

located relatively far away from zero. This is especially true for $\tau_{U_{\text{walls/roof}}}$, τ_{WFR} , $\tau_{q_{\text{inf@50Pa}}}$, and $\tau_{\text{Occ.Density}}$ where the amount of uncertainty in the posteriors remains very high and overlaps with the posterior mean values $p_{\text{post}}(\mu)$. The posterior uncertainty has only been reduced for one of the standard deviation parameters: the capacity of the thermal mass $\tau_{C_{\text{mass}}}$. However, even though the uncertainty of the archetype-level parameters μ , τ may have been reduced, it is the resulting predictive distribution of the building-level parameters θ that is of most interest. The marginal predictive distributions of the θ s are shown in the right-hand column in Fig. 7 using the prior and posterior archetype-level parameters, respectively. From the plot of the predictive distributions, the effect of the large standard deviations τ is immediately evident. For several of the parameters, the priors and posteriors look alike; however, for the capacity of the thermal mass, the high posterior mean value $\mu_{C_{\text{mass}}}$ has drawn the posterior predictive distribution out to the boundary of 600 kJ/(m²K).

The expected posterior correlation matrix $E_{\text{post}}[\text{corr}(\theta)]$ is shown in Fig. 8. No strong correlations are found between the parameters, but all of the parameters are found to correlate moderately with each other (max. correlation strength = 0.5), except for the capacity of the thermal mass, which does not exhibit any significant correlation with the other parameters. These correlations are inevitable as the parameters interact through the overall heat balance in every time step; e.g., increasing occupant density [m²/pers.] leads to a smaller heat gain from occupants and therefore brings the heat balance out of equilibrium. This imbalance might be corrected by decreasing the U-value (walls/roof) [W/(m²K)] and thus the heat loss through the envelope (correlation = -0.4). Another possibility could be to increase the WFR and thus the overall window area, whereby more sunlight enters the building. The correlation obviously depends on other factors also, such as overall floor area, U-value (windows), SHGC, window frame fraction, etc.

The posterior distribution for the archetype-level half-Cauchy scale parameter $p_{\text{post}}(\gamma)$ (see Eq. (4)) is displayed in Fig. 9. Since a non-informative prior distribution (uniform) was used, the posterior inference is driven by data alone.

3.5. Performance of calibrated archetype model

The performance and overall quality of the calibrated archetype model were analyzed based on its predictive capabilities on two different hierarchical levels; (1) the within-sample predictive per-

Table 5

ASHRAE Guideline 14-2014 compliance requirements.

Measure	Monthly data	Hourly data
NMBE	< ± 5%	< ± 10%
CVRMSE	< 15%	< 30%

formance for the 50 sampled training buildings, and (2) the out-of-sample predictive performance for another 100 unseen test buildings. Sample details are given in Table 2. The predictive performance was assessed on both the scale of individual buildings and on the aggregated urban scale.

Following definitions by the ASHRAE Guideline 14-2014, we employed two measures to assess the predictive performance of the calibrated building simulations; the normalized mean bias error (NMBE) and the coefficient of variation of the root mean squared error (CVRMSE):

$$\text{NMBE} = \frac{\sum_{t=1}^{n_t} (y_t - y_t^*)}{n_t} \bigg/ \bar{y} \times 100, \quad (14)$$

$$\text{CVRMSE} = \sqrt{\frac{\sum_{t=1}^{n_t} (y_t - y_t^*)^2}{n_t}} \bigg/ \bar{y} \times 100. \quad (15)$$

The ASHRAE Guideline 14-2014 compliance requirements are given in Table 5 for monthly and hourly data, respectively. As we employed 3-hourly data in this study, it seems reasonable to accept levels above the monthly requirements but somewhat below the ones for hourly data.

3.5.1. Within-sample predictive performance

The within-sample predictive performance, i.e. the ability of the model to forecast building energy use time series for the 50 known training buildings, was used to evaluate the immediate quality of fit for the sampled buildings independently and the archetype as a whole. This internal assessment of the inferred building-level parameters θ was carried out using measured and simulated data from the training period and the holdout validation period (Table 2).

Measured and simulated data is shown in Fig. 10A for training building $b=6$. The figure displays $n_s = 1000$ time series predictions $y_6^* | X_6, W_{\text{train/valid}}, \theta_6^{1:1000}$ from the BEM, each simulated using independent draws from the posterior distribution of the building-level parameters $p_{\text{post}}(\theta_6)$, fixed building parameters X_6 , and the corresponding weather data, either W_{train} or W_{valid} . The fully acceptable performance of the training period generalizes well into the validation period with new weather conditions (Fig. 10A). This suggests that a representative set of parameters was estimated for the building. The performance measures (NMBE and CVRMSE) of building $b=6$ are shown together with the remaining 49 training buildings in Fig. 11 for both the training and validation period.

Values of NMBE and CVRMSE were calculated for all 50 training buildings individually for both the training and validation period (Fig. 11). Some of the buildings do not fulfill the ASHRAE Guideline 14-2014 compliance requirements for monthly data in the training period when using the expected values of NMBE and CVRMSE as points of reference, but most comply with the requirements for hourly data (one building exceeds NMBE = 10% and five buildings exceed CVRMSE = 30% in the training period). We find fairly large variations in especially the NMBE values for many buildings (error bars in Fig. 11 give the 95% confidence interval of individual measures around the mean), but for most buildings, their values reach across NMBE = 0% with 95% confidence, and all buildings have a 95% CI with values from within the NMBE < ± 5% band. One building ($b=15$) stands out as the one with the highest bias by far

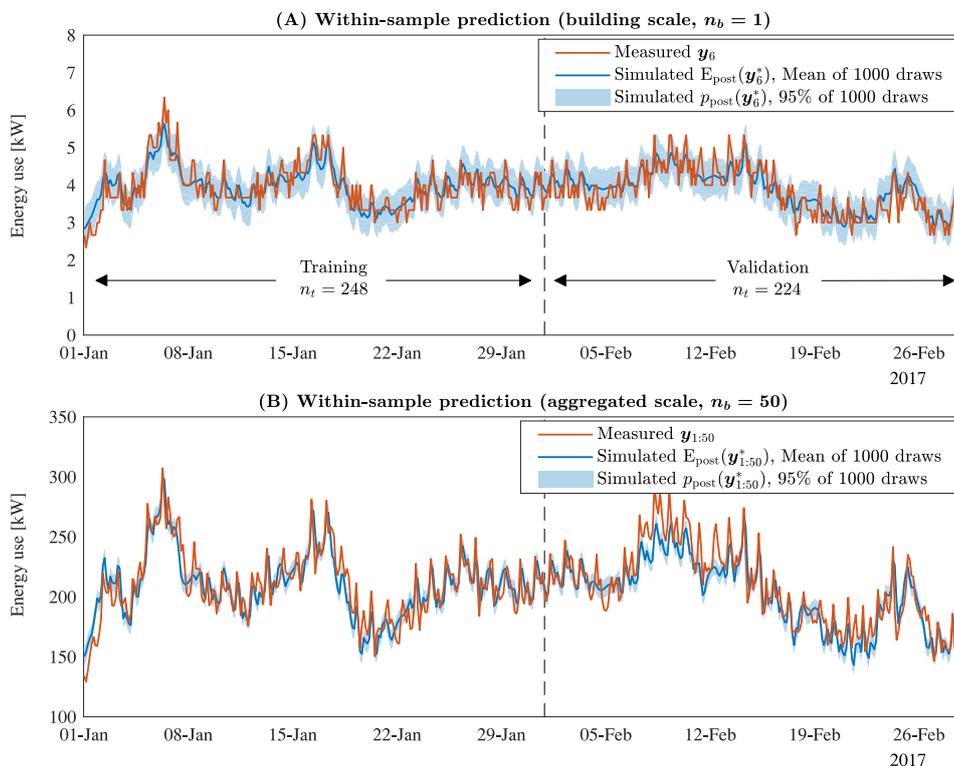


Fig. 10. Predictive performance of within-sample prediction. (A): Energy use predictions of building $b = 6$ against measured data. (B): Aggregated energy use predictions of buildings $b = 1:50$ against aggregated measured data. Simulated data was generated using 1000 draws from the posterior distribution of the calibration parameters $p_{\text{post}}(\theta_b)$, fixed parameters \mathbf{X}_b , and weather conditions for the training period $\mathbf{W}_{\text{train}}$ and validation period $\mathbf{W}_{\text{valid}}$, respectively.

in the training period: $\text{NMBE} = 7.4\% \pm 14.3\%$ (mean \pm 95% CI). Even though the building's bias is higher than that of the other buildings, we chose to not deem it an outlier as there was no reason for doing so, e.g. extraordinary oscillations in the training data, besides the relatively bad fit.

Predictions from the validation period generally perform worse than predictions from the training period as would be expected (the NMBE is often more than three times higher in the validation period for many buildings). This is most likely caused by the unpredictable nature and presence of occupants as their daily movements and presence remain unaccounted for and hence were not modeled explicitly (schedules of occupancy were assumed flat and fixed across time). This simplification create a challenge for the prediction of the validation period (February 2017) as it contains the weeklong Danish winter holiday (public schools are closed) which, in the municipality of Aarhus, took place from 10 February 2017–19 February 2017. From the aggregated time series predictions shown in Fig. 10B, it is clear that much of the bias in the validation period is located within the winter holiday period. As school holidays, bank holidays, etc. are spread across the year, we expect similar predictive performance in the remaining periods of the year. Summary statistics of the mean values of within-sample validation period performance are given in Table 6.

In addition to the building-scale performance, the within-sample predictive performance was also evaluated on the aggregated scale to represent the expected performance in an urban setting (Fig. 10B). Aggregating all 50 within-sample buildings into a single model naturally obscures much of the data variability as an aggregated model pays no regard to the fit of individual

buildings besides their contribution to the summarized consumption pattern. With an $\text{NMBE} = 3.0\% \pm 0.5\%$ (mean \pm 95% CI) and a $\text{CVRMSE} = 7.2\% \pm 0.3\%$ (mean \pm 95% CI) for the validation period, the aggregated performance is good and, as expected, much better than for most of the individual buildings (Table 6).

3.5.2. Out-of-sample predictive performance

The out-of-sample predictive performance, i.e. the ability of the model to forecast time series of energy use from unseen test buildings $\mathbf{X}_{\text{out-of-sample}}$ not used for the calibration of the archetype model, is used to evaluate the external robustness and homogeneity of the archetype calibration. However, no direct posterior estimate exists for the uncertain parameters $\theta_{\text{out-of-sample}}$ of the 100 out-of-sample test buildings to be used for simulation; instead, samples of the parameters are drawn from their posterior predictive distribution (9) that is generated using draws from the posterior distributions of the archetype-level parameters $p_{\text{post}}(\mu, \Sigma)$. As the archetype-level parameters were inferred from the combined information contained by all 50 within-sample training buildings, posterior predictions of $\theta_{\text{out-of-sample}}$ represent a generalized archetype estimate with the inherent variability and heterogeneity of the sampled training buildings.

Aggregated predictions of the 100 out-of-sample test buildings are shown in Fig. 12 against the aggregated measured data for both the training and validation periods. An $\text{NMBE} = 2.9\% \pm 6.2\%$ (mean \pm 95% CI) and a $\text{CVRMSE} = 7.8\% \pm 2.9\%$ (mean \pm 95% CI) are found for the aggregated predictions in the validation period (Table 6).

When comparing the performance of out-of-sample aggregated predictions (Fig. 12) against the performance of within-sample

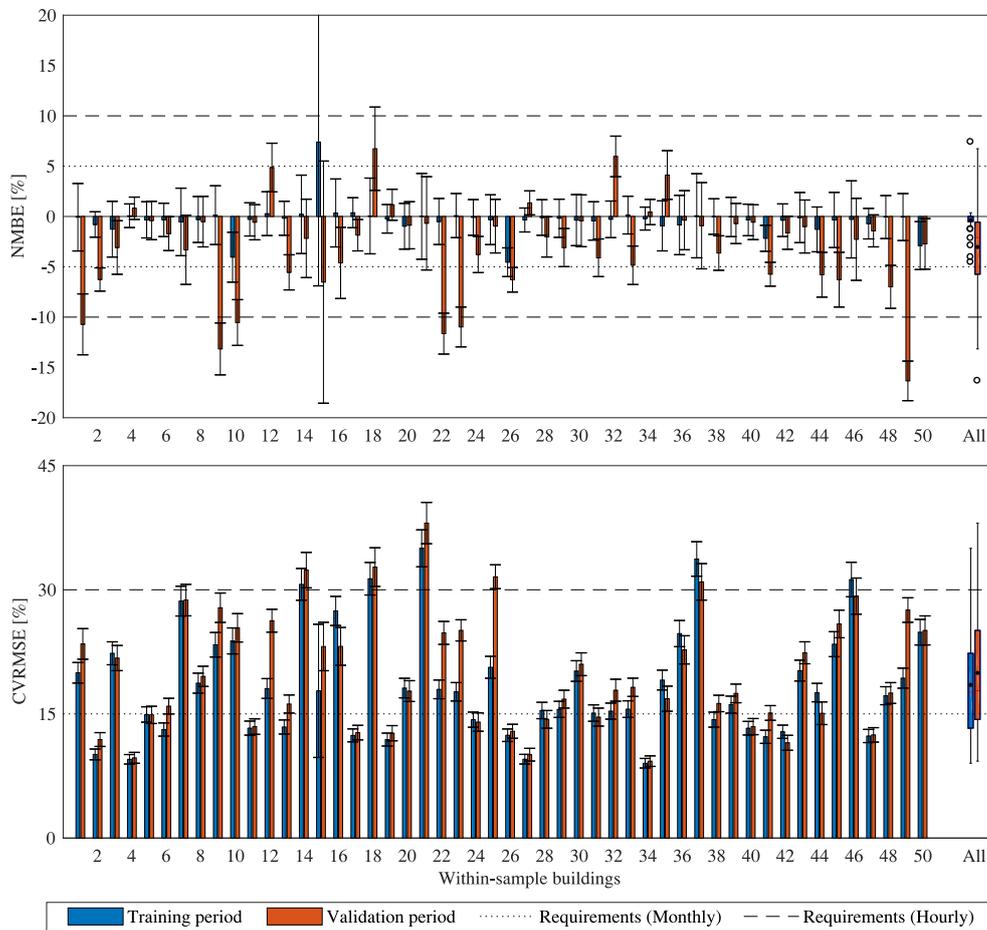


Fig. 11. Normalized mean bias error (NMBE) and coefficient of variation of the root mean squared error (CVRMSE) for the 50 within-sample buildings individually. Measures are calculated using 1000 simulations per building, each using a random draw of the calibrated parameters from their posterior distribution $p_{\text{post}}(\theta_b)$ against measured data y_b . The height of the bars gives the mean value with error bars around covering the 95% CI. The boxplots at x-value "All" comprise the distribution of mean values for all 50 buildings. ASHRAE Guideline 14-2014 compliance requirements are given for both measures.

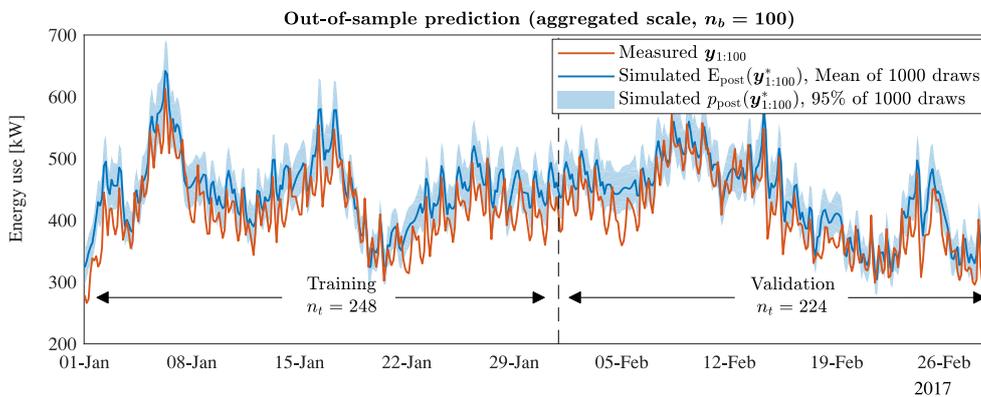


Fig. 12. Aggregated predictive performance of out-of-sample buildings $b = 51:150$ against aggregated measured data. Simulations were generated using draws from the posterior predictive distribution of calibration parameters $p_{\text{post}}(\theta_{\text{out-of-sample}}) = N(\mu, \Sigma)$.

aggregated predictions (Fig. 10B), it is evident that the aggregated out-of-sample predictions suffer from a positive bias compared to the negative bias of the within-sample predictions (2.9% vs. –3.0%). Measures of fit (mean values of NMBE and CVRMSE) for within-sample and out-of-sample predictions are summarized in Table 6 on the disaggregated scale of individual buildings and on the aggregated scale, respectively.

Out-of-sample predictions are, as expected, inferior to within-sample predictions. An insufficient ability to match measured data is seen for most out-of-sample predictions on the scale of individual buildings (95% interval of NMBEs = [–38.7% 112.9%] and CVRMSEs = [28.7% 120.1%]). This mismatch is, however, smeared out on the scale of aggregated predictions where the cancellation effect effectively reduces the differences to an acceptable level (NMBE = 2.9% and CVRMSE = 7.8%) almost identical to that of the within-sample aggregated predictions (NMBE = –3.0% and CVRMSE = 7.2%). Thus, the established archetype model can hardly be used for single-building predictions of high-resolution energy use time series from unseen buildings without considerable uncertainty. However, the results promise good performance for aggregated predictions on an urban level.

4. Discussion

4.1. Why hierarchical modeling?

The proposed archetype framework takes advantage of a hierarchical (multilevel) link between data from individual buildings. It is a generalization of the ordinary Bayesian calibration framework in which the building-specific calibration parameters θ are themselves given an archetype model – in this case the multivariate Gaussian distribution – whose parameters μ , Σ are also estimated from the data. This statistical setup has several appealing features:

- 1 Calibrating uncertain parameters without a hierarchical formulation corresponds to either inferring separate building-specific estimates [3,4,6], or a single pooled archetype estimate [12]. The more general hierarchical model allows for a reasonable compromise between these two extremes.
- 2 It allows for the modeling of the correlation within calibration parameters of the archetype. This is essential for any application that relies on drawing archetype parameter-sets of new unseen buildings.
- 3 Predictions of new unseen buildings draw on a much larger dataset that explores the variability of the archetype and should therefore be more representative.

Fitting separate building-independent estimates of the calibration parameters θ corresponds to assuming that all buildings are unique without sharing any similarities at all. Alternatively, averaging the data and fitting only a single pooled estimate correspond to the opposite case of assuming that all buildings are the same, whereby all differences are ignored. Both cases are unsatisfactory as the reality lies somewhere in between. Applying a hierarchical model allows us to fit in uncertain archetype parameters

μ using the individual building datasets while accounting for archetype heterogeneity among the buildings (represented through the variance in the covariance matrix Σ). Through the process of “shrinking”, the hierarchical model pools the individual building estimates θ towards a common archetype mean μ as a result of the archetype distribution (estimates far away from the archetype mean have very low probability under the normality assumption). Using a non-hierarchical model based on averaging of separately fitted building estimates results in a model more prone to outliers [26].

The explicit modeling of the correlation and variance between calibration parameters at the archetype level, represented through the covariance matrix Σ , allows for informed predictions of unseen buildings under the archetype. This acknowledges not only the heterogeneity of the archetype estimates (variance), but also the potential correlation $\text{corr}(\theta)$ that would otherwise be ignored.

4.2. Within-sample vs. out-of-sample predictions

The lower accuracy of individual out-of-sample building predictions (building-scale) compared to predictions of individual within-sample buildings (Table 6) is most likely a consequence of the uncertainties related to (1) archetype heterogeneity, (2) building sample size n_b , and (3) human behavior and preferences. Within-sample training buildings were all fitted building-specific values of the calibration parameters $p_{\text{post}}(\theta_{\text{within-sample}})$, whereas for out-of-sample buildings, these values were predicted from the posterior predictive distribution $p_{\text{pred}}(\theta_{\text{out-of-sample}}) = N(\mu, \Sigma)$. The posterior archetype estimates $p_{\text{post}}(\mu, \Sigma)$ can be viewed as a compromise between the characteristics of the training buildings, which inevitably will be characterized by some degree of heterogeneity. As such, the archetype model gives us an estimate of the average building, as well as an estimate of the variability that characterizes the archetype. Therefore, evaluating the performance of the archetype model's mean prediction on individual buildings will inevitably indicate a poor performance for some buildings, even if the pool of buildings used for training contained practically identical buildings. The only way of improving out-of-sample prediction performance is to eliminate the heterogeneity of the buildings belonging to the archetype by significantly tightening up the archetype definition. However, as a very tight archetype definition defeats the whole idea of simplifying the building stock into a “few” archetypes, a fairly large degree of error has to be accepted for out-of-sample predictions of single buildings. Finally, even if a very tight and homogeneous archetype definition is applied, human behavior remains a very uncertain factor with a significant effect on building energy use [36–40], which cannot be removed by tightening the archetype definition. Incorporating a stochastic model of occupancy, e.g. using an agent-based [41,42] or a Markov chain-based [43,44] approach, as opposed to the fixed and non-varying flat schedules of the demonstration case could potentially reduce much of the variability in out-of-sample predictions and help reduce archetype heterogeneity to that originating from differences in technical parameters. However, we foresee sev-

Table 6

Summary statistics of expected predictive performance for sampled training buildings (within-sample) and unseen test buildings (out-of-sample). Measures in square brackets cover the [2.5% | 50% (median) | 97.5%] percentiles, i.e. the 95% central interval of the distributions of n_b individual building means. On the aggregated scale, there is only one mean value.

Sample	Period	Buildings in sample, n_b	Predictive performance	
			Building scale	Aggregated scale
Within-sample	Validation	50	[–14.0% –2.1% 6.1%] (NMBE)	–3.0% (NMBE)
			[9.6% 17.8% 34.1%] (CVRMSE)	7.2% (CVRMSE)
Out-of-sample	Validation	100	[–38.7% 0.6% 112.9%] (NMBE)	2.9% (NMBE)
			[28.7% 38.8% 120.1%] (CVRMSE)	7.8% (CVRMSE)

eral problems with doing this, e.g. issues of over-parameterization and identifiability, but it remains to be investigated in future work.

The performance of aggregated predictions is not affected by archetype heterogeneity to the same extent as individual building-scale predictions, but is mainly controlled by uncertainty in estimating the mean parameters μ that represent an average archetype realization. The law of large numbers states that the estimated mean of a sample approaches the true mean as the number of samples goes to infinity. Here, obtaining an archetype model can be considered analogous to estimating the mean of a sample, where the archetype model would become more representative of out-of-sample buildings belonging to the archetype as the number of buildings used for training increases. The relatively small difference in aggregated performance between within-sample and out-of-sample buildings (NMBE_{within-sample} = -3.0% vs. NMBE_{out-of-sample} = 2.9%) suggests that the number of buildings n_b needed for estimating the archetype may not be much larger than the sample size used in this case study demonstration.

4.3. Parameter identification

In the demonstration of the framework, we applied a first-order 5R1C building energy model for simulating hourly heating energy use, and subsequently, for calibrating five model parameters. This model was selected for the purpose of demonstration and because of its fast computational speed. However, other model formulations may easily be applied in future applications of the framework, e.g. higher-order RC formulations or high-fidelity BEM tools such as EnergyPlus. In fact, reduced-order models may be too simple for identifying the true estimates of building characteristics [32,34,35]; nevertheless, while the obtained parameter estimates may not compare directly to reality, they give meaningful information about a building's characteristics in the context of the model. Reynders et al. [35] argue that models of at least fourth order (four thermal capacitances) are needed for a reliable identification of building parameters, and that additional boundary conditions are needed in terms of indoor air temperature and heat fluxes through envelope walls. Hedegaard and Petersen [32] addressed similar issues in a simulation-based study of four RC-models with the most complex being a third-order model; however, they infer that a second-order model allows for a fit to data that is fully compatible to that of a more complex third-order model. They further conclude that none of these low-order models are capable of providing a reliable partitioning of the overall heat transfer coefficient into heat transfer by transmission and ventilation, respectively. One should therefore be very cautious about trusting the exact value of calibration parameters θ of individual buildings when using low-order RC models. Whether these issues of parameter interpretation and identifiability persist in applications with high-fidelity BEM tools such as EnergyPlus remains unclear and ought to be studied in future work. However, Heo et al. [23] show, in a case study-based investigation of Bayesian calibration efficacy under different levels of uncertainty in model input data for EnergyPlus models, that the amount of certainty in input data highly affects the posterior estimates of calibration parameters and thus the identifiability of the model. We therefore expect the identifiability of the calibration parameters θ , and the predictive performance of the proposed archetype framework to increase as more concrete and specific data about the training buildings is applied. Such data could originate from detailed surveys and audits in carefully selected training buildings. Another less intrusive option could be to incorporate energy performance certificate (EPC) data if available. Finally, as demonstrated in recent state-of-the-art UBEM studies [5,45], GIS may be employed for establishing more accurate geometric models than those used in this study.

5. Conclusions and outlook

A hierarchical multilevel framework for the calibration of archetype physics-based BEM parameters was proposed. The framework relies on the statistical assumption of exchangeability among archetype buildings, i.e. the buildings represent a homogeneous sample from the archetype. Using Bayesian inference, information available in independent time series datasets is used collectively to pool individual estimates of building parameters towards a common archetype description of the calibration parameters, as well as their variability and correlation. This results in the archetype model being less vulnerable to the presence of outliers in the building stock used to train the archetype model, thus enabling it to make out-of-sample predictions that are more robust than predictions made with models founded on ordinary aggregation of individual building estimates.

The application and performance of the framework were subsequently demonstrated in a case study of Danish residential single-family dwellings from the 1950s. The case demonstrates how the framework can be used to identify a set of shared archetype parameters, and how the inevitable presence of heterogeneity in the buildings used for training is manifested into the variance parameters and the prediction of parameters for new unseen buildings.

The proposed archetype framework is most suited as the central calibration engine for urban-scale building energy modeling where archetype models are used as "puzzle pieces" for modeling urban areas which would otherwise be too complex to model and calibrate using independent building energy models. However, the suitability of the proposed archetype framework for such applications remains to be proven in future work.

Funding

The European Union's Research and Innovation Funding Programme (7th EU Framework Programme) supported this work through the FP7-Energy project "Resource Efficient Cities Implementing Advanced Smart City Solutions" (READY), work package 3 [project reference 609127].

Acknowledgments

The authors would like to thank the district heating company in Aarhus, AffaldVarme Aarhus, for supplying the case data that serves to demonstrate the application of the proposed calibration framework.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.enbuild.2018.07.030](https://doi.org/10.1016/j.enbuild.2018.07.030).

References

- [1] C.F. Reinhart, C.C. Davila, Urban building energy modeling – a review of a nascent field, *Build. Environ.* 97 (2016) 196–202, doi:10.1016/j.buildenv.2015.12.001.
- [2] E. Mata, A. Sasic Kalagasidis, F. Johnsson, Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK, *Build. Environ.* 81 (2014) 270–282, doi:10.1016/j.buildenv.2014.06.013.
- [3] C. Cerezo, J. Sokol, C. Reinhart, A. Al-Mumin, Three methods for characterizing building archetypes in urban energy simulation. A case study in Kuwait City, in: Proceedings of the Fourteenth Conference of International Building Performance Simulation Association Building Simulation, IBPSA, Hyderabad, India, 2015, pp. 2873–2880. URL: <http://www.ibpsa.org/proceedings/BS2015/p2435.pdf>.
- [4] C. Cerezo, J. Sokol, S. AlKhaled, C. Reinhart, A. Al-Mumin, A. Hajiah, Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): a residential case study in Kuwait City, *Energy Build.* 154 (2017) 321–334, doi:10.1016/j.enbuild.2017.08.029.

- [5] T. Dogan, C. Reinhart, Shoeboxer: an algorithm for abstracted rapid multi-zone urban building energy model generation and simulation, *Energy Build.* 140 (2017) 140–153, doi:[10.1016/j.enbuild.2017.01.030](https://doi.org/10.1016/j.enbuild.2017.01.030).
- [6] J. Sokol, C. Cerezo Davila, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, *Energy Build.* 134 (2017) 11–24, doi:[10.1016/j.enbuild.2016.10.050](https://doi.org/10.1016/j.enbuild.2016.10.050).
- [7] G. Dall'O, A. Galante, M. Torri, A methodology for the energy performance classification of residential building stock on an urban scale, *Energy Build.* 48 (2012) 211–219, doi:[10.1016/j.enbuild.2012.01.034](https://doi.org/10.1016/j.enbuild.2012.01.034).
- [8] L. Filogamo, G. Peri, G. Rizzo, A. Giaccone, On the classification of large residential buildings stocks by sample typologies for energy planning purposes, *Appl. Energy* 135 (2014) 825–835, doi:[10.1016/j.apenergy.2014.04.002](https://doi.org/10.1016/j.apenergy.2014.04.002).
- [9] M. Aksoezen, M. Daniel, U. Hassler, N. Kohler, Building age as an indicator for energy consumption, *Energy Build.* 87 (2015) 74–86, doi:[10.1016/j.enbuild.2014.10.074](https://doi.org/10.1016/j.enbuild.2014.10.074).
- [10] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, *Renew. Sustain. Energy Rev.* 37 (2014) 123–141, doi:[10.1016/j.rser.2014.05.007](https://doi.org/10.1016/j.rser.2014.05.007).
- [11] H. Lim, Z.J. Zhai, Review on stochastic modeling methods for building stock energy prediction, *Build. Simul.* 10 (2017) 607–624, doi:[10.1007/s12273-017-0383-y](https://doi.org/10.1007/s12273-017-0383-y).
- [12] A.T. Booth, R. Choudhary, D.J. Spiegelhalter, Handling uncertainty in housing stock models, *Build. Environ.* 48 (2012) 35–47, doi:[10.1016/j.buildenv.2011.08.016](https://doi.org/10.1016/j.buildenv.2011.08.016).
- [13] A.T. Booth, R. Choudhary, D.J. Spiegelhalter, A hierarchical Bayesian framework for calibrating micro-level models with macro-level data, *J. Build. Perform. Simul.* 6 (2013) 293–318, doi:[10.1080/19401493.2012.723750](https://doi.org/10.1080/19401493.2012.723750).
- [14] Q. Li, G. Augenbroe, J. Brown, Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction, *Energy Build.* 124 (2016) 194–202, doi:[10.1016/j.enbuild.2016.04.025](https://doi.org/10.1016/j.enbuild.2016.04.025).
- [15] Y. Heo, R. Choudhary, G.A. Augenbroe, Calibration of building energy models for retrofit analysis under uncertainty, *Energy Build.* 47 (2012) 550–560, doi:[10.1016/j.enbuild.2011.12.029](https://doi.org/10.1016/j.enbuild.2011.12.029).
- [16] Y. Heo, G. Augenbroe, D. Graziano, R. T. Muehleisen, L. Guzowski, Scalable methodology for large scale building energy improvement: relevance of calibration in model-based retrofit analysis, *Build. Environ.* 87 (2015) 342–350, doi:[10.1016/j.buildenv.2014.12.016](https://doi.org/10.1016/j.buildenv.2014.12.016).
- [17] Y. Kang, M. Krarti, Bayesian-emulator based parameter identification for calibrating energy models for existing buildings, *Build. Simul.* 9 (2016) 411–428, doi:[10.1007/s12273-016-0291-6](https://doi.org/10.1007/s12273-016-0291-6).
- [18] V.J. Kim, C.S. Park, Stepwise deterministic and stochastic calibration of an energy simulation model for an existing building, *Energy Build.* 133 (2016) 455–468, doi:[10.1016/j.enbuild.2016.10.009](https://doi.org/10.1016/j.enbuild.2016.10.009).
- [19] A. Chong, K.P. Lam, M. Pozzi, J. Yang, Bayesian calibration of building energy models with large datasets, *Energy Build.* 154 (2017) 343–355, doi:[10.1016/j.enbuild.2017.08.069](https://doi.org/10.1016/j.enbuild.2017.08.069).
- [20] W. Tian, R. Choudhary, A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London, *Energy Build.* 54 (2012) 1–11, doi:[10.1016/j.enbuild.2012.06.031](https://doi.org/10.1016/j.enbuild.2012.06.031).
- [21] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 63 (3) (2001) 425–464, doi:[10.1111/1467-9868.00294](https://doi.org/10.1111/1467-9868.00294).
- [22] M.H. Kristensen, R. Choudhary, R.H. Pedersen, S. Petersen, Bayesian calibration of residential building clusters using a single geometric building representation, in: *Proceedings of the Fifteenth Conference of International Building Performance Simulation Association Building Simulation, IBPSA, San Francisco, August 2017*, pp. 1294–1303. URL: http://www.ibpsa.org/proceedings/BS2017/BS2017_330.pdf.
- [23] Y. Heo, D.J. Graziano, L. Guzowski, R.T. Muehleisen, Evaluation of calibration efficacy under different levels of uncertainty, *J. Build. Perform. Simul.* 8 (3) (2014) 135–144, doi:[10.1080/19401493.2014.896947](https://doi.org/10.1080/19401493.2014.896947).
- [24] M.H. Kristensen, R. Choudhary, S. Petersen, Bayesian calibration of building energy models: comparison of predictive accuracy using metered utility data of different temporal resolution, *Energy Procedia.* 122 (2017) 277–282, doi:[10.1016/j.egypro.2017.07.322](https://doi.org/10.1016/j.egypro.2017.07.322).
- [25] A. Gelman, Prior distributions for variance parameters in hierarchical models, *Bayesian Anal.* 1 (3) (2006) 515–534, doi:[10.1214/06-ba117a](https://doi.org/10.1214/06-ba117a).
- [26] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, *Bayesian Data Analysis*, third ed., Chapman and Hall/CRC, 2013. ISBN 1439840954.
- [27] H. Madsen, P. Bacher, G. Bauwens, A.-H. Deconinck, G. Reynders, S. Roels, E. Himpe, G. Leth, Thermal Performance Characterization Using Time Series Data – IEA EBC Annex 58 Guidelines, Technical University of Denmark (DTU), Lyngby, Denmark, 2015 'Technical report' URL: http://orbit.dtu.dk/files/127709087/guidelines_IEA58_statisticalModeling_2016_11_28.pdfhttp://orbit.dtu.dk/files/127709087/guidelines_IEA58_statisticalModeling_2016_11_28.pdf.
- [28] K. Duer, S. Svendsen, M.M. Mogensen, J.B. Laustsen, Energy labeling of glazings and windows in Denmark: calculated and measured values, *Solar Energy* 73 (1) (2002) 23–31, doi:[10.1016/S0038-092X\(02\)00031-2](https://doi.org/10.1016/S0038-092X(02)00031-2).
- [29] M.D. Morris, Factorial sampling plans for preliminary computational experiments, *Technometrics* 33 (2) (1991) 161–174 ISSN 0040-1706, doi:[10.1080/00401706.1991.10484804](https://doi.org/10.1080/00401706.1991.10484804).
- [30] H.B. Rijal, P. Tuohy, M.A. Humphreys, J.F. Nicol, A. Samuel, J. Clarke, Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings, *Energy Build.* 39 (7) (2007) 823–836 ISSN 0378-7788, doi:[10.1016/j.enbuild.2007.02.003](https://doi.org/10.1016/j.enbuild.2007.02.003).
- [31] K. Gram-Hanssen, Husholdningers elforbrug - hvem bruger hvor meget, til hvad og hvorfor? SBI 2005:12. Danish Building Research Institute, 2005. ISBN 87-563-1235-0. URL <https://sbi.dk/Assets/Husholdningers-elforbrug-hvem-bruger-hvor-megget-til-hvad-og-hvorfor/2006-01-12-0481751096.pdf>. Language: Danish.
- [32] R.E. Hedegaard, S. Petersen, Evaluation of grey-box model parameter estimates intended for thermal characterization of buildings, *Energy Procedia.* 132 (2017) 982–987, doi:[10.1016/j.egypro.2017.09.692](https://doi.org/10.1016/j.egypro.2017.09.692).
- [33] M.H. Kristensen, S. Petersen, Choosing the appropriate sensitivity analysis method for building energy model-based investigations, *Energy Build.* 130 (2016) 166–176, doi:[10.1016/j.enbuild.2016.08.038](https://doi.org/10.1016/j.enbuild.2016.08.038).
- [34] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, *Energy Build.* 43 (7) (2011) 1511–1522 ISSN 0378-7788, doi:[10.1016/j.enbuild.2011.02.005](https://doi.org/10.1016/j.enbuild.2011.02.005).
- [35] G. Reynders, J. Diriken, D. Saelens, Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals, *Energy Build.* 82 (2014) 263–274, doi:[10.1016/j.enbuild.2014.07.025](https://doi.org/10.1016/j.enbuild.2014.07.025).
- [36] W. Abrahamse, L. Steg, How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings, *J. Econ. Psychol.* 30 (5) (2009) 711–720, doi:[10.1016/j.joep.2009.05.006](https://doi.org/10.1016/j.joep.2009.05.006).
- [37] O.G. Santin, L. Itard, H. Visscher, The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock, *Energy Build.* 41 (11) (2009) 1223–1232, doi:[10.1016/j.enbuild.2009.07.002](https://doi.org/10.1016/j.enbuild.2009.07.002).
- [38] G.M. Huebner, I. Hamilton, Z. Chalabi, D. Shipworth, T. Oreszczyn, Explaining domestic energy consumption – the comparative contribution of building factors, socio-demographics, behaviours and attitudes, *Appl. Energy* 159 (2015) 589–600, doi:[10.1016/j.apenergy.2015.09.028](https://doi.org/10.1016/j.apenergy.2015.09.028).
- [39] S. Yang, M. Shipworth, G. Huebner, His, hers or both's? The role of male and female's attitudes in explaining their home energy use behaviours, *Energy Build.* 96 (2015) 140–148, doi:[10.1016/j.enbuild.2015.03.009](https://doi.org/10.1016/j.enbuild.2015.03.009).
- [40] M.H. Kristensen, S. Petersen, Explaining variability in metered energy use for similar buildings using Bayesian inference, *Energy Procedia* 132 (2017) 897–902, doi:[10.1016/j.egypro.2017.09.709](https://doi.org/10.1016/j.egypro.2017.09.709).
- [41] Y.S. Lee, A. M. Malkawi, Simulating multiple occupant behaviors in buildings: An agent-based modeling approach, *Energy Build.* 69 (2014) 407–416, doi:[10.1016/j.enbuild.2013.11.020](https://doi.org/10.1016/j.enbuild.2013.11.020).
- [42] J. Langevin, J. Wen, P.L. Gurian, Simulating the human-building interaction: development and validation of an agent-based model of office occupant behaviors, *Build. Environ.* 88 (2015) 27–45, doi:[10.1016/j.buildenv.2014.11.037](https://doi.org/10.1016/j.buildenv.2014.11.037).
- [43] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, F. Descamps, A method for the identification and modeling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, *Build. Environ.* 75 (2014) 67–78, doi:[10.1016/j.buildenv.2014.01.021](https://doi.org/10.1016/j.buildenv.2014.01.021).
- [44] P.D. Andersen, A. Iversen, H. Madsen, C. Rode, Dynamic modeling of presence of occupants using inhomogeneous Markov chains, *Energy Build.* 69 (2014) 213–223, doi:[10.1016/j.enbuild.2013.10.001](https://doi.org/10.1016/j.enbuild.2013.10.001).
- [45] C.C. Davila, C.F. Reinhart, J.L. Bemis, Modeling Boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, *Energy* 117 (2016) 237–250, doi:[10.1016/j.energy.2016.10.057](https://doi.org/10.1016/j.energy.2016.10.057).

6.2 Epilogue

This chapter formally ends Part 2 of the thesis and the treatment of research objective 1 (Section 1.6). It has been shown how to incorporate both dynamic time series modelling, archetype heterogeneity and parameter correlation in the characterization of uncertain archetype parameters, without relying on the use of statistical emulators. The framework shows a considerable promise for UBEM applications, as is demonstrated in Part 3 of the thesis, but it also contains simplifications and choices that ought to be further investigated in future work, as is also partially discussed in the paper [P4].

One thing that I suspect to have a major impact on the archetype model is the underlying geometric model; in this case a simple scalable rectangular box with a fixed non-varying orientation, and fixed standardized location of the windows and external shadowing. With a relatively simple upgrade of the model it should be possible to include information from building-specific EPC schemes and urban GIS models, if available, which contains a great deal of geometric and technical data. The application of the CityGML modelling standard for establishing 3D massing models, in particular, offers considerable promise (Chen et al., 2017; Nageler et al., 2017; Jaeger et al., 2018). It would allow for a much more detailed assessment of transmission losses and solar heat gains, thus securing a more accurate identification of uncertain archetype parameters in the calibration process. Another important assumption is the modelling of occupancy behaviour. This is notoriously difficult in UBEM settings, but recent stochastic developments may be worth investigating in future work (Happle et al., 2018).

In journal paper manuscript [S5] that is appended to this thesis and which is currently under review for publication in Elsevier journal *Applied Energy*, we applied a non-archetype approach to the establishment of a neighbourhood-scale UBEM in the city of Aarhus, i.e. all buildings were characterized independently. This paper is worth mentioning in relation to the discussion on occupant behaviour and usage schedules because we herein applied Bayesian, i.e. stochastic, methods to identify building-specific 24-hour occupant schedules for weekdays and weekends. It would be interesting to extend this usage pattern identification to the scale of archetypes using the hierarchical set-up already in place in [P4] to investigate whether such profiles exist for building archetypes. Data clustering studies on time series energy use data have previously shown that, in fact, there appears to exist a semantic subset of usage profiles, as described in Section 1.3.1 on the classification of archetypes. However, it remains doubtful whether these correlate with a classical segmentation of the building stock based on the construction year.

A less complex approach would be to explicitly model known calendar and holiday data, as demonstrated by Dahl et al. (2018) in the forecast of aggregated DH energy loads. The authors find that including calendar data (discrimination between month, weekday and hour of day) may significantly increase performance, while including local holiday data (observances, national holidays and school holidays) provides only a slight performance increment.

The improvements outlined above should be seen as suggestions for future work, as they are not included in this thesis. A unifying discussion on and suggestions for future work can be found in Section 9.2. The last part of the thesis – Part 3 – concerns the treatment of research objective 2 and the application of the hierarchical archetype approach for setting up and simulating urban building energy models under uncertainty. To do so, the entire stock of SFHs in Aarhus is classified into archetypes and characterized using the hierarchical archetype calibration framework presented in this chapter.

Part 3

Urban building energy modelling for retrofit analysis under uncertainty

Chapter 7

Archetype-based UBEM

7.1 Motivation

The hierarchical archetype calibration framework presented in Chapter 6 was demonstrated on the stock of SFHs from the 1950s using one month of 3-hourly training data from 50 training buildings, and validated on its predictive capabilities using a different month. In addition, the ability of the archetype to accurately characterize 100 new unseen buildings belonging to the archetype was tested based on the predictive capability of these 100 buildings in the validation month.

The following two-page extended abstract (conference paper [P5]) was presented at the 4th IBPSA England conference, *Building Simulation & Optimization 2018*. The hierarchical archetype calibration approach of Chapter 6 is herein applied on the entire stock of SFHs in the municipality of Aarhus (1850-2015), and subsequently used to populate UBEMs of two suburban towns within the municipality. The two UBEMs account for 2,153 and 690 SFHs, respectively. Each of these buildings was simulated independently (1,000 simulations per building to explore model uncertainty) with a subsequent aggregation of their time series to form aggregated UBEM predictions. The paper shows 3-hourly time series predictions from the two UBEMs against measured data and discusses the potential applications of such archetype-based urban modelling.



Urban-scale dynamic building energy modeling and prediction using hierarchical archetypes: A case study of two Danish towns

Martin Heine Kristensen*, Rasmus Elbæk Hedegaard, Steffen Petersen
 Department of Engineering, Aarhus University, DK-8000 Aarhus, Denmark
 * Corresponding author (mhk@eng.au.dk)

Introduction

It remains practically infeasible to gather all the required data inputs for physics-based urban building-by-building energy modelling (Reinhart and Davila, 2016). Simplifications may therefore be necessary, e.g. through archetype segregation of the building stock to reduce the task of data acquisition and calibration of uncertain parameters. The authors of this extended abstract recently proposed a novel hierarchical archetype calibration methodology that allows a robust probabilistic inference of unknown archetype input parameters for unseen buildings belonging to an archetype (Kristensen et al., 2018). The methodology has been proven fast and accurate for urban-scale predictions of aggregated building energy use under uncertainty.

In this contribution we demonstrate how hierarchically calibrated archetype models of Danish detached single-family houses (SFH's) can accurately predict the urban district heating energy use of unseen buildings in two different suburban towns. We end up by discussing the various practical applications of such urban models.

Methods

Data

Three-hourly district heating (DH) energy consumption data for 27,000 SFH's located in the municipality of Aarhus, Denmark, were coupled with six building-specific data fields from the Danish Building and Dwelling Register (BDR): 1) usage type, 2) construction year, 3) footprint, 4) number of floors, 5) basement and 6) attic area utilized for living/heated. The DH time series consisted of combined energy use for space heating and on-site domestic hot water (DHW) preparation.

Archetype segregation

The building stock was partitioned into nine SFH archetypes following the building stock segregation performed Danish Building Research institute as part of the European research project TABULA (Wittchen and Kragh, 2012). Only the construction year was used for segregation. The nine archetype age groups are shown in Table 1.

Building energy modeling

The DH energy use time series of each building was modeled using the hourly dynamic resistance-capacitance model of ISO 13790:2008 that treats each building as a single thermal zone, in combination with a

simple DHW consumption model. The only available and known data inputs were the footprint, number of floors, and the heated basement and attic area (besides basic climate data logged from a nearby weather station). All other inputs necessary to simulate the BEM were unknown at the level of individual buildings. A-priori probability density functions (PDF's) or fixed values were therefore given each uncertain input parameter at the level of archetypes to reflect historical data and educated guesses.

Hierarchical archetype calibration

The archetype calibration methodology proposed by (Kristensen et al., 2018) was applied to infer a-posteriori PDF's for six of the uncertain parameters per archetype: 1) window-floor ratio, 2) U-value of ext. walls/roof, 3) capacity of thermal mass, 4) infiltration airflow@50pa, 5) occupant density, and 6) room heating set point temperature. The methodology applies a Bayesian hierarchical formulation that binds training buildings together around a shared archetype estimate whereby the inference draw strength from all training building datasets simultaneously. The methodology allows training buildings that are very "likely" to dominate the inference of uncertain parameters, while outlying/unlikely buildings are given less weight – a process known as "shrinkage". Each archetype was trained on a sample of 35 randomly selected SFH's from the dataset, each with time series of three-hourly DH energy use of January 2017 (248 data points).

Urban case towns for prediction

Two suburban case towns were selected for validation of the urban-scale predictive capabilities of the archetype framework: 1) "DK-8250 Egå" and 2) "DK-8330 Beder" (Table 1). February 2017 (224 data points) was used for validation.

Table 1

Classification of case town buildings into nine archetypes.

Archetype partitioning	DK-8250 Egå	DK-8330 Beder
Arch. 1 (1851-1930)	105 (4.9%)	56 (8.1%)
Arch. 2 (1931-1950)	37 (1.7%)	43 (6.2%)
Arch. 3 (1951-1960)	74 (3.4%)	12 (1.7%)
Arch. 4 (1961-1972)	1166 (54.2%)	302 (43.8%)
Arch. 5 (1973-1978)	369 (17.1%)	83 (12.0%)
Arch. 6 (1979-1998)	226 (10.5%)	149 (21.6%)
Arch. 7 (1999-2006)	129 (6.0%)	37 (5.4%)
Arch. 8 (2007-2010)	21 (1.0%)	4 (0.6%)
Arch. 9 (2011-2015)	26 (1.2%)	4 (0.6%)
Total, n_p	2153 (100%)	690 (100%)

Results

The measured and simulated DH energy use of the two case towns is shown in Fig. 1 and measures of predictive performance are given in Table 2.

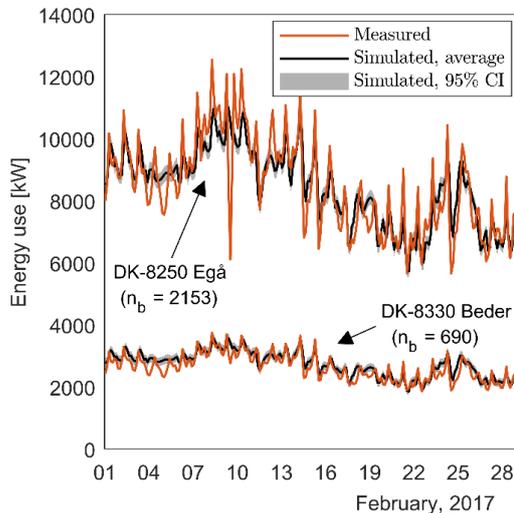


Fig. 1. Measured and simulated aggregated DH energy use (three-hourly) of two suburban towns, respectively. Simulated energy use consist of the 95% central posterior density region derived from aggregating 1000 random simulations from each of the individual building energy models. The average of the 1000 aggregated simulations is highlighted in black.

Table 2
Measures of predictive performance.

Metric	DK-8250 Egå	DK-8330 Beder
NMBE	0.4%	4.2%
MAPE	6.8%	7.2%
CVRMSE	9.0%	8.5%
R ²	74.0%	78.4%

The simulated energy use fit the measured energy use very well. The entire energy consumption of the validation period was predicted within a 4.2% error margin for both towns when measured with the normalized mean bias error (NMBE) metric. The accuracy of which individual data points (three-hourly values) were predicted was measured with the mean absolute percentage error (MAPE) metric to be within 7.2%. The coefficient of variation of the root mean squared error (CVRMSE) and the coefficient of determination (R²) measures the variability of the residuals and thus the explanatory power of the predictions. The two urban models explained approx. 74%-78% (R²-values) of the variability in the measured time series.

Discussion

Urban-scale models of this kind may allow city governments, utility companies, and other energy policy stakeholders that work on the urban scale of neighborhoods, cities, or even entire building stocks, to plan and predict

the effect of various energy efficiency measures and production strategies. The application of simple and publicly available building and property information as the only need-to-have input data about the buildings to be predicted (besides measured energy use datasets from a subsample of buildings for the initial archetype calibration) provides a flexible platform that can easily be expanded or further developed. Because the model structure is based on thermodynamic principles, it may also have use in investigating urban-scale effects on e.g. peak loads and overall energy use due to various interventions in the building stock, e.g. retrofitting, city densification or expansion, and building technologies for facilitating demand response programs.

The application of archetypes to represent the building stock is obviously a crude simplification of its true diversity. However, applying a probabilistic representation and calibration of the uncertain archetype parameters on the level of individual buildings through a hierarchical structure like in this study preserves much of the natural heterogeneity that defines the variability within archetypes. This preservation of heterogeneity is crucial for accurate predictions of new and unseen buildings from the archetypes.

The hierarchical archetype framework proves capable of predicting the aggregated energy use of buildings in larger urban areas with high accuracy as demonstrated for two suburban towns in this study. Although the framework remains to be implemented for other buildings than detached single-family dwellings in order to fully represent a true urban area with many different building types, we do not believe this to pose any difficulties for the framework. The temporal resolution of the predictions is in no way limited to three-hourly data points, but solely defined by the underlying physical model structure and quality of calibration data. Urban models of increasing temporal resolution will therefore be possible in the near future as the distribution of smart energy meters proceed to penetrate the market.

Acknowledgements

The authors would like to thank the district heating company in Aarhus, AffaldVarme Aarhus, for supplying the case data that serves to demonstrate the application of the proposed calibration framework. The work was funded by the 7th EU Framework Programme (FP7-Energy project "READY", project reference: 609127).

References

- Kristensen, M. H., Hedegaard, R. E., and Petersen, S. (2018). Hierarchical calibration of archetypes for urban building energy modeling. *Energy and Buildings [In Press, Available online]*.
- Reinhart, C. F. and Davila, C. C. (2016). Urban building energy modeling – A review of a nascent field. *Building and Environment* 97, 196–202.
- Wittchen, K. B. and Kragh, J. (2012). SBi 2012:01 Danish building typologies – Participation in the TABULA project. Technical report, Danish Building Research Institute, Aalborg University.

7.2 Epilogue

The extended abstract documents preliminary results from the work on resolving research objective 2. The end-goal is to use the hierarchical archetype approach to assemble a single large UBEM of all SFHs in the municipality of Aarhus, and to validate its ability to accurately simulate the energy use for all hours of the year. This effort is continued in the following chapter, the final chapter of Part 3.

Chapter 8

Citywide retrofit analysis under uncertainty

This chapter comprises work that has not yet been submitted for publication, but is expected to be in the near future. The research presented in the following sections is therefore not formatted as a self-contained paper manuscript with its own introduction, but rather builds on the published papers of the thesis and findings presented in previous chapters. In particular, the reader might want to familiarize him- or herself with the work on hierarchical archetype calibration outlined in journal paper [P4] that was presented in Chapter 6, before delving into this final chapter of the thesis.

8.1 Motivation

An application of the hierarchical archetype modelling and calibration framework was demonstrated in Chapter 7, where two UBEMs of smaller towns in the municipality of Aarhus were established using nine archetypes of SFHs from the Danish building stock. These archetypes were identified using training data from 35 buildings per archetype, each with 3-hourly time series data from one month. The predictive performance of the UBEMs was subsequently tested on the following month. However, it remains to be demonstrated how to apply the hierarchical archetype approach to meet the goals of research objective 2 (Section 1.6, p. 15); this endeavour is the focal point of the current chapter.

It is the intention of this chapter to illustrate an application of the hierarchical archetype framework where the energy use of an entire city is forecast far into the future under uncertainty of both energy renovations and climate change. As such, it has a three-fold objective:

1. To classify the building stock of SFHs into a reduced number of archetypes.
2. To further refine the archetype training and validation scheme applied in [P4] and demonstrated on two UBEMs in [P5].
3. To set up a single UBEM covering all SFHs in the municipality of Aarhus, and to apply it to forecast citywide SFH energy demand from 2017-2050 under uncertainty of energy renovations and climate change.

The details of the investigation are outlined below in Sections 8.2-8.4.

8.2 Data cleansing and archetype segmentation

The stock of SFHs in the municipality of Aarhus consists of approx. 27,800 buildings (Figure 2.2; Figure 2.4). Most of these buildings – approx. 23,000 (83%) – have no secondary heating installed in addition to district heating (DH) and thus rely on DH for both space heating and on-site domestic hot water (DHW) preparation. The remaining 17% of the SFHs are equipped with at least one heating installation in addition to DH, e.g. heat pumps, electric radiators, oil-fired boilers, wood-burning stoves, etc. Only the approx. 23,000 SFHs without any additional heating installations were considered in this study to reduce heterogeneity.

The dataset was cleaned for buildings with erroneous data entries, as well as buildings for which consumption levels differed significantly from the majority of the building stock, further treatment. Buildings with a "negative" or "zero" heated area, missing construction year, missing energy use data, and buildings with either unreasonably low (below 10 kWh/m²/year) or high (above 500 kWh/m²/year) energy use intensity were removed – in total 115 buildings. The BDR building data and DH energy use data of the final 22,914 SFHs are shown in the three left plots of Figure 8.1A, where the buildings are classified by construction years.

The largest share of buildings (69%) was constructed between 1940 and 1990, with 1969 being the year with by far the most newly built SFHs (5% of the stock). The heated floor area is on average 159 m² for the whole dataset¹ with the smallest houses being those constructed from 1930 to 1960, on average 141 m², and the largest houses being those constructed from 2007 to 2015, on average 178 m². The annual DH energy consumption per heated floor area (mean consumption of 2013, 2014 and 2015) is on average around 145 kWh/m²/year for buildings construction up until the 1950s. For buildings from the late 1950s onwards, the energy use intensity decreases steadily until it reaches a somewhat constant plateau of approx. 100 kWh/m²/year for buildings from the 1980s and 1990s. Hereafter, the trend of the mean exhibits the steepest decline so far reaching approx. 50 kWh/m²/year on average for buildings from 2015.

The SFHs were partitioned into nine archetypes (Table 8.1) following the segmentation performed by the Danish Building Research Institute (Wittchen and Kragh, 2012) as part of the EU's IEE projects TABULA (2009-2012) and EPISCOPE (2013-2016), in which residential building archetypes were developed for 20 European countries (Loga et al., 2016). Besides from all buildings being SFHs, only the construction year of the buildings was used for the segmentation. A visual interpretation of the building data after being classified into archetypes is shown in the three right plots of Figure 8.1B.

8.3 Training and validation of hierarchical archetypes

The following subsections provide details on the training and validation of nine hierarchical archetypes, which together are used to set up the UBEM.

8.3.1 Building energy modelling

Building energy models (BEMs) were established for the SFHs to simulate their space heating and DHW energy consumption using the physics-based 5R1C hourly dynamic BEM of ISO13790:2008 in combination with a simple scalable building geometry similar

¹The global average of all SFHs in the municipality of Aarhus before cleansing is 161 m² cf. Figure 2.2.

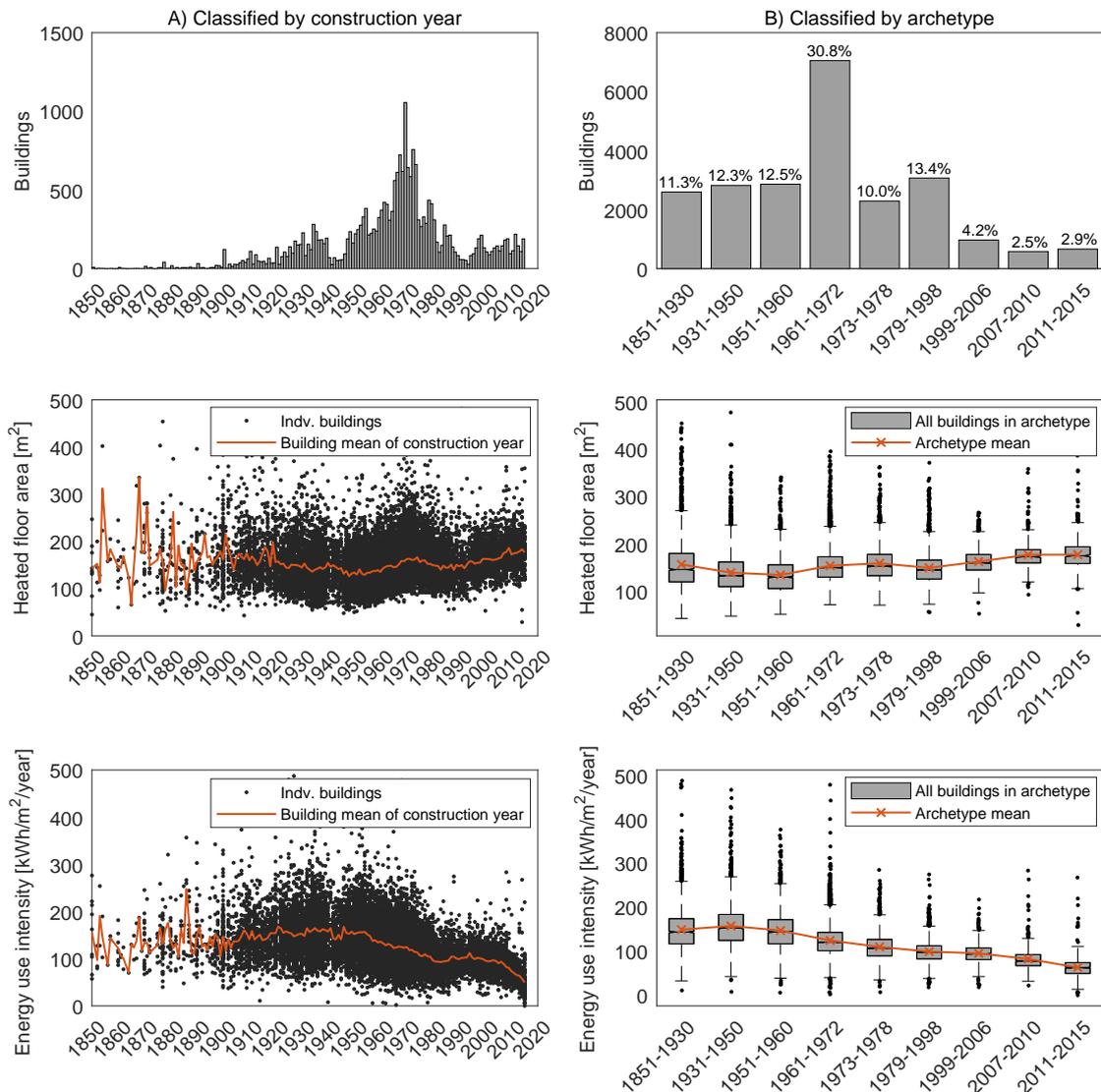


Figure 8.1. Archetype segmentation of 22,914 SFHs in the municipality of Aarhus from 1850-2015. A) The three left plots depict the building stock classified by their construction year. B) The three right plots depict the building stock classified in nine archetypes. Data source: Danish Building and Dwelling Register (BDR) and AffaldVarme Aarhus (AVA).

to the models presented in [P2], [P3] and [P4]. The only data that were available and thus known about the buildings were the building characteristics gathered from the Danish national BDR database. Four of these parameters – the building footprint area, the number of floors, the utilized attic area, and the utilized basement area – were used to set up the building geometry. The remaining data inputs that were necessary for establishing the BEMs were unknown at the level of individual buildings. A database of BEM input data was therefore constructed at the level of archetypes using available historical data and educated guesses (Table 8.2).

Table 8.1. Partitioning of Danish SFHs constructed from 1851-2015 into nine archetypes based on the construction year. Archetype periods are defined by shifts in building traditions and in energy requirements given by the Danish Building Regulations (BR). Pictures and segmentation arguments are partly reproduced from "Danish building typologies – Participation in the TABULA project" (Wittchen and Kragh, 2012).

Archetype	Example	Building period	Segmentation argument
Archetype 1		1851-1930	Shift in building tradition
Archetype 2		1931-1950	Cavity walls introduced
Archetype 3		1951-1960	Insulated cavity walls introduced
Archetype 4		1961-1972	First energy requirements (BR61)
Archetype 5		1973-1978	Tightened energy requirements (BR72)
Archetype 6		1979-1998	Tightened energy requirements (BR78)
Archetype 7		1999-2006	Tightened energy requirements (BR98)
Archetype 8		2007-2010	Tightened energy requirements (BR06/BR08)
Archetype 9		2011-2015	Tightened energy requirements (BR10)

8.3.2 Hierarchical archetype calibration scheme

Since the majority of necessary BEM input parameters were unknown at the level of individual buildings, estimates of their values were established at the level of archetypes based on educated guesses (Table 8.2). These guessed archetype-level BEM input parameters were tuned using the Bayesian archetype calibration framework proposed in journal paper [P4] that was presented in Chapter 6. However, to enable the construction a UBEM from these nine archetypes that can validly facilitate urban-scale retrofit analysis (objective 1 of this chapter), it proved necessary to further enhance the archetype training and validation scheme proposed in [P4], and later demonstrated in [P5]. This enhancement encompasses four specific upgrades:

1. The resolution of the time series data is increased from 3-hour values to 1-hour values to allow for a more nuanced identification of system dynamics.
2. The number of training buildings per archetype n_b is increased to 75 to allow for

Table 8.2. Uncertain archetype-level BEM input parameter values. Five parameters marked with (★) were calibrated within the designated prior range. Three parameters marked with (★★) have their values prescribed in the ISO 13790:2008 international standard.

Uncertain model parameter	Unit	Prior range		Archetype								
		Min.	Max.	1	2	3	4	5	6	7	8	9
<i>Geometry</i>												
Length-width-ratio	[-]	0.10	1.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Room height	[m]	2.30	3.00	2.30	2.40	2.40	2.40	2.50	2.50	2.60	2.70	2.80
Window-floor-ratio★	[m ² /m ²]	0.10	0.50									
Frame-window-ratio	[m ² /m ²]	0.10	0.50	0.30	0.30	0.25	0.25	0.25	0.25	0.20	0.15	0.15
<i>Transmission</i>												
Temp. factor ground	[-]	0.50	1.00	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
U-value (floor)	[W/(m ² K)]	0.10	0.50	0.50	0.38	0.36	0.30	0.30	0.20	0.12	0.12	0.12
U-value (basement)	[W/(m ² K)]	0.10	1.20	1.00	1.00	1.00	0.65	0.40	0.35	0.30	0.20	0.18
U-value (walls)★	[W/(m ² K)]	0.10	0.50									
U-value (roof)★	[W/(m ² K)]	0.10	0.50									
U-value (windows)★	[W/(m ² K)]	0.70	5.00									
Solar heat gain coef.	[-]	0.50	0.70	0.60	0.60	0.60	0.60	0.60	0.60	0.50	0.50	0.50
Thermal capacity (mass)★	[kJ/(m ² K)]	50	600									
Effective area (mass)★★	[m ² /m ²]	2.5	3.5	Building specific, see the ISO 1379:2008 standard								
Heat conduction (mass)★★	[W/(m ² K)]	9.10	9.10	9.10	9.10	9.10	9.10	9.10	9.10	9.10	9.10	9.10
Heat transfer coef. (surf.-air)★★	[W/(m ² K)]	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45
<i>Ventilation</i>												
Infiltration airflow★	[l/(sm ²)]	0.10	10.0									
Mechanical ventilation	[-]	No	Yes	No	No	No	No	No	No	Yes	Yes	Yes
Heat recovery efficiency	[-]	0.50	0.90	N/A	N/A	N/A	N/A	N/A	N/A	0.60	0.70	0.85
Design airflow	[l/(sm ²)]	0.10	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
<i>Occupancy</i>												
Occupant density★	[m ² /pers.]	10	150									
Heating set-point temp.★	[°C]	17.0	25.0									
<i>Domestic hot water</i>												
Flow temperature	[°C]	40.0	60.0	55.0	55.0	55.0	55.0	55.0	55.0	55.0	55.0	55.0
Mains supply temperature	[°C]	5.0	15.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
Circulation pipe heat loss★	[W/K]	0.00	20.0									
Hot water consumption	[m ³ /(pers.yr)]	5	25	15	15	15	15	15	15	15	15	15

a more informed identification of the archetypes and their heterogeneity, which, in turn, allows for more accurate predictions of parameters θ_{new} for new out-of-sample buildings.

3. The training period is extended from 1 month (January) to 3 months (January, March and May), and the validation period is similarly extended from 1 month (February) to 3 months (February, April and June). New out-of-sample buildings are tested on the entire 6 months from January up to and including June (4,344 hourly values). This enables a more reliable validation of the predictive performance that takes into account the changing seasons from winter to summer. Ideally, the performance would be tested on hourly data for the entire 12 months of the year, but such data were not available at the time of the study. However, as the seasonal conditions of the second half-year (from July up to and including December) more or less mirror those of the first half-year, it is assumed that the buildings' behaviour and performance equal that of the first half-year. This assumption allows for predictions of annual building energy use, which is necessary in the retrofit analysis (Section 8.4).
4. The out-of-sample predictive performance is tested on the level of individual archetypes (75 out-of-sample buildings per archetype) and on the level of three different urban areas, to account for urban-specific effects.

8.3.3 Archetype-level performance validation

BEMs for 75 out-of-sample buildings were established for each of the nine archetypes using their known input data (data from the BDR supplemented with data from Table 8.2). The values of the calibrated input parameters were drawn at random from the archetype-specific posterior predictive distributions of θ_{new} . This process was repeated $N_{\theta} = 1,000$ times per building using Monte Carlo (MC) sampling to cover all areas of the 5-dimensional distribution (five calibration parameters). By the central limit theorem, the sampling error approximates:

$$\varepsilon_{\text{sampling}} \approx \frac{1}{\sqrt{N}}, \quad (8.1)$$

which for $N_{\theta} = 1,000$ yields a 3% sampling error. Quadrupling the number of draws from θ_{new} , and thus the number of building simulations, halves the error, regardless of the number of dimensions in θ (calibration parameters).

These 1,000 different time series simulations for each of the 75 out-of-sample buildings per archetype were subsequently merged to form an aggregated distribution of 1,000 time series simulations per archetype. Posterior predictive simulations are shown in Figure 8.2 for Archetype 1 and 9, respectively, against aggregated measured data for the 75 out-of-sample buildings.

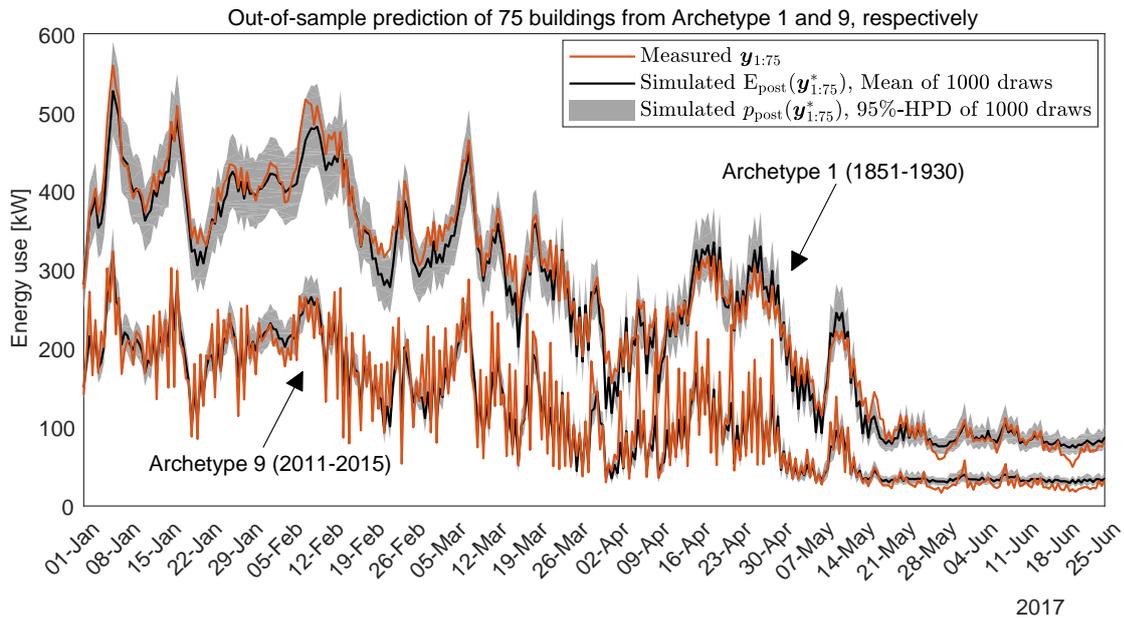


Figure 8.2. Posterior predictions of Archetype 1 and Archetype 9, respectively. Aggregated time series data from 75 new out-of-sample buildings from each of the two archetypes are shown together with 1,000 posterior MC simulations of the buildings, each applying a random draw from the archetype-specific predictive distribution of θ_{new} . The temporal resolution of the graphs is three hours for the purpose of visualization.

The resulting predictive performance of the nine calibrated archetypes is shown in Figure 8.3 in terms of the normalized mean bias error (NMBE) and coefficient of the root mean squared error (CVRMSE) metrics. The distribution of metrics for each archetype

were derived based on the aggregated time series of all of the 1,000 simulations of each of the 75 out-of-sample buildings.

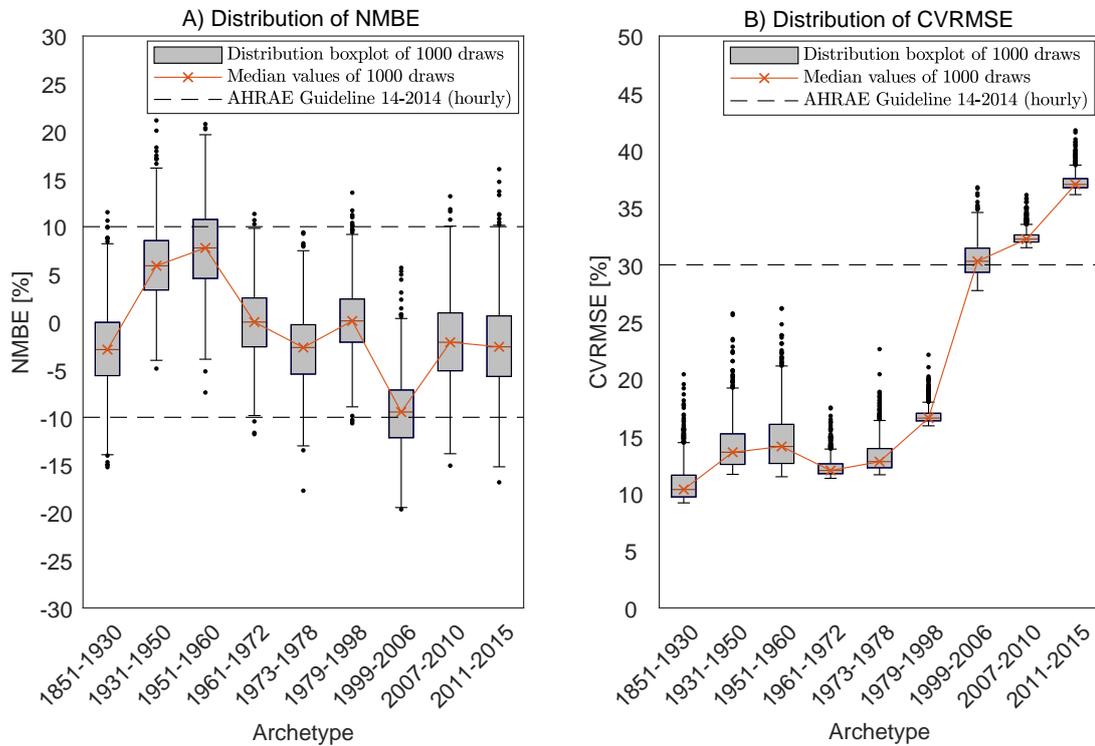


Figure 8.3. Distributions of NMBE and CVRMSE for the aggregated prediction of the time series energy use of 75 out-of-sample buildings from 1 January 2017 to 30 June 2017. The error distributions comprise the outcome of 1,000 repeated predictions per archetype.

All archetypes are found to perform satisfactorily considering the fact that the 75 buildings used for evaluation constitute new unseen buildings that were not used for training, and that the time series predictions consist of 4,344 hourly values across the changing seasons from winter to summer. The median NMBE of all archetypes complies with the $\text{NMBE} < 10\%$ requirement of ASHRAE Guideline 14-2014 for hourly data, meaning that the summarized energy use across the period is less than 10% wrong; however, the box-plots of Figure 8.3A bear witness of a large uncertainty in the predictions (the NMBE distributions' tails cover a large range of values). The 50% most probable outcome space – the interquartile range of the distributions – is illustrated by the boxes themselves in grey. Several of the NMBE distributions include values within $-5\% < \text{NMBE} < 5\%$ and even 0%, with 50% certainty.

A different picture is drawn when looking at the CVRMSE metric (Figure 8.3B), which measures the scale of variations in the simulations compared to the measurements. The CVRMSE of the first six archetypes fall well below the ASHRAE requirement of $\text{CVRMSE} < 30\%$ for hourly data time series; however, for the three newest archetypes (1999-2006, 2007-2010 and 2011-2015), the CVRMSE distributions cross or go beyond the recommended upper bound of $\text{CVRMSE} = 30\%$. This trend of increasing CVRMSE (worse fit) with the construction years of the archetypes is suspected to be an effect of the inadequacy of the DHW and occupant models in describing individual SFHs' comfort preferences, behaviours and attitudes. The preparation of DHW requires a large amount of heating power in a short time period, e.g. 10-15 min for a shower. This becomes more

important in newer energy-efficient houses as they experience an increasing share of overall heating energy use spent on DHW preparation compared to aged houses. In addition, older houses rely on the use of DHW tanks, which charge slowly, to a larger extent than newer houses, which almost exclusively make use of plate heat exchangers. Nevertheless, the impact of the three newest archetypes' inability to accurately match hourly consumption peaks is deemed less important in the prediction of urban-scale energy use, as they only make up less than 10% of the building stock of SFHs. Moreover, a low CVRMSE score is not deemed particularly important for the application presented in this chapter, i.e. to forecast annually aggregated urban-scale building energy use.

8.3.4 Urban-level performance validation

In addition to the testing of the individual archetypes' predictive performance in general, their interacting performance in real urban settings was also tested, where other factors such as demography and local architectural design might influence performance. Different sub-samples of Aarhus were selected for this task, as it was computationally infeasible to validate the archetypes' urban-scale performance on the entire municipality (simply too much validation data to gather and process).

Three self-contained urban areas in the context of residential suburbs were considered as case study towns for prediction (Table 8.3). These three suburban areas provided suitable cases for testing the effect of urban context on archetype predictive performance due to their status as independent towns with local communities. Moreover, the SFHs of the towns offer somewhat different combinations of the nine archetypes and hence represent three different technical compositions. For example, approx. 50% of all the SFHs in Harlev are from Archetype 4 (1961-1972), while this is only true for 20% of the SFHs in Højbjerg. On the other hand, 32% of the SFHs in Højbjerg are from Archetype 3 (1951-1960), while this is true of 2% in Beder.

Table 8.3. Partitioning of SFHs in three urban test areas (towns) into archetypes.

Urban test area	Buildings included	Partitioning of buildings into archetypes								
		1	2	3	4	5	6	7	8	9
DK-8270 Højbjerg	3191 (57%)	10.4%	17.6%	31.7%	19.8%	7.6%	8.5%	1.8%	0.8%	1.9%
DK-8330 Beder	645 (90%)	8.7%	6.2%	1.9%	44.0%	12.9%	20.2%	5.0%	0.6%	0.6%
DK-8462 Harlev	800 (81%)	11.6%	6.4%	6.5%	47.9%	10.8%	7.9%	3.8%	4.6%	0.6%

All SFHs located in the three towns that have correct and complete hourly datasets throughout the period from 1 January 2017 up until and including 30 June 2017, were included in the analysis. Unfortunately, many of the SFHs did not have smart meters installed and running in January 2017, which is why only a reduced subset of the buildings in the three towns was analysed (see Table 8.3 for the number of included test buildings). On average, 63% of the buildings across the three towns were included the analysis, meaning that 37% were left out due to insufficient validation data. However, there is no reason to expect the omitted buildings to perform any different than those included in the analysis.

The predicted time series of Højbjerg and Harlev are shown in Figure 8.4 against the aggregated measured data of their buildings, while the overall performance of the three towns is shown in Figure 8.5 in terms of the NMBE and the CVRMSE metrics. The presentation of time series predictions from Beder is omitted from Figure 8.4 because it is located very close to the time series of Harlev, and thus would obscure the visualization.

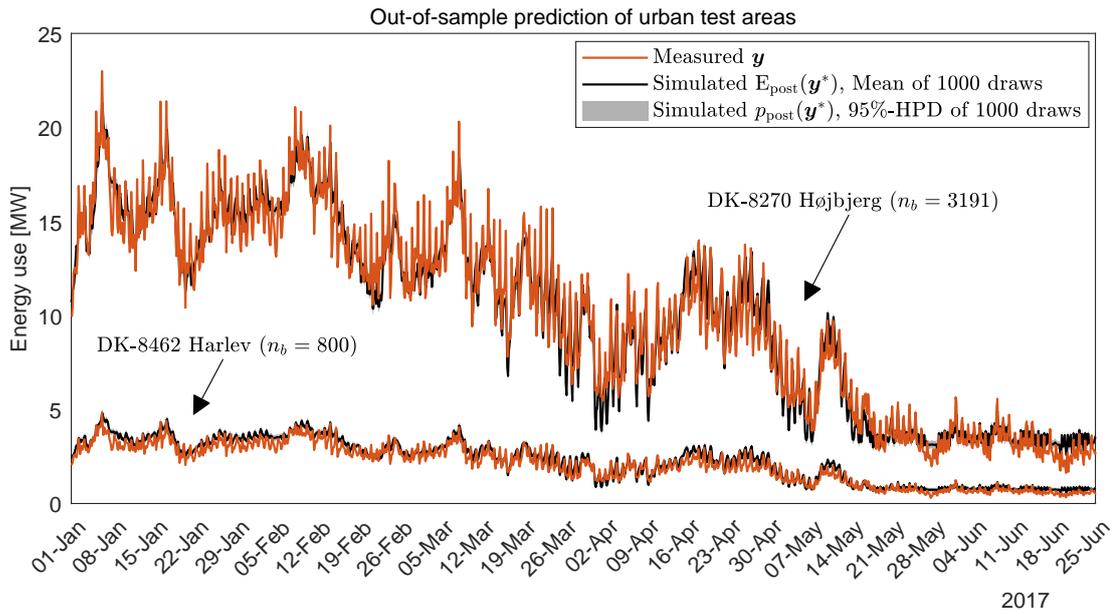


Figure 8.4. Posterior predictions of urban test areas DK-8270 Højbjerg and DK-8462 Harlev, respectively. Aggregated time series data from their buildings are shown together with 1,000 posterior MC simulations of the buildings, each applying a random draw from the archetype-specific predictive distribution of θ_{new} . The temporal resolution of the graphs is three hours for the purpose of visualization.

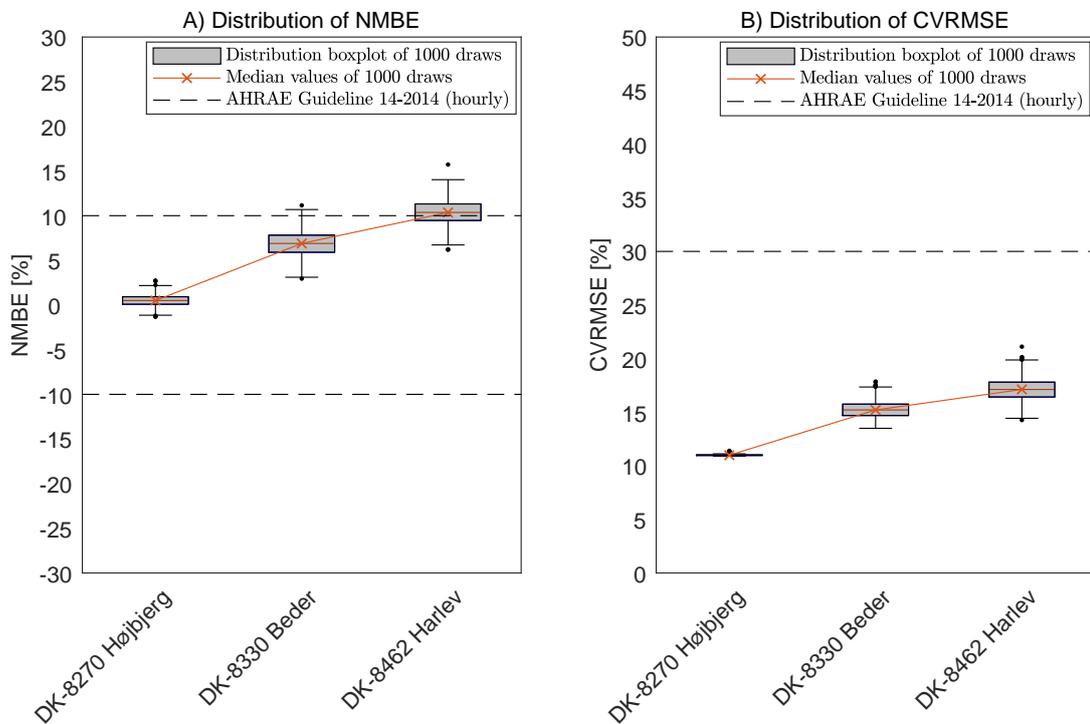


Figure 8.5. Distributions of NMBE and CVRMSE for the aggregated prediction of the time series energy use from 1 January 2017 to 30 June 2017, for the three urban test areas. The error distributions comprise the outcome of 1,000 repeated predictions for each urban area.

The urban-scale predictions are, in general, very accurate, with Højbjerg being the most accurately predicted town and Harlev the least accurately predicted town. However, all three towns perform satisfactorily; only Harlev has some of its NMBE distribution (approx. 50%) above the $\text{NMBE} = 10\%$ line, indicating a tendency to overshoot the energy consumption (Figure 8.5A). The CVRMSE measures are in the area of 10-20%, indicating a good ability of the UBEMs to accurately capture urban-scale fluctuations in consumption.

8.4 Forecasting citywide energy demand from 2017-2050

8.4.1 Urban building energy model of Aarhus

The nine archetypes were applied to populate BEMs of the 22,914 SFHs supplied with district heating in the municipality of Aarhus (Figure 8.1). Aggregated predictions of their hourly energy use are shown in Figure 8.6 for the entire year of 2017 using $N_\theta = 100$ MC draws from the archetype-specific distributions of θ_{new} per building ($\varepsilon_{\text{sampling}} \approx 10\%$), and locally measured weather data.

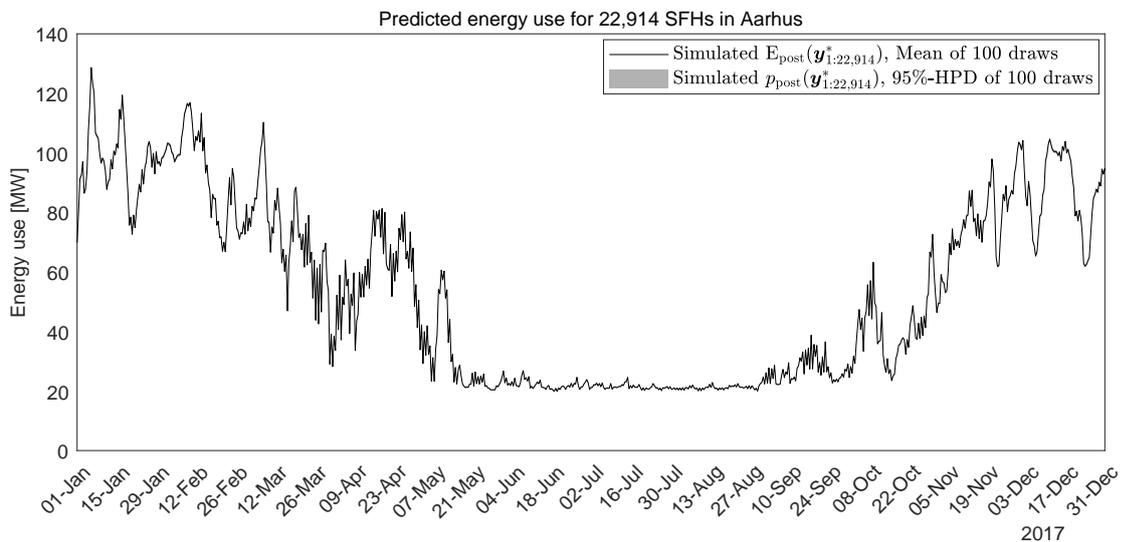


Figure 8.6. Posterior prediction of aggregated district heating energy use of 22,914 SFHs in Aarhus in 2017, simulated using $N = 100$ individual MC draws per building.

A very low variability is present in the predictions of the aggregated energy use due to a large number of individual building energy use time series being aggregated (the cancellation effect). The predicted energy use intensity of the buildings in 2017 is, on average, 142-143 kWh/m²/yr (95%-HPD of 100 draws per building). This is approx. 15% higher than the global mean intensity found from 3 years (2013, 2014 and 2015) of measured DH data for approx. 27,800 SFHs in Aarhus (Figure 2.4). It is difficult to say whether this deviation is caused by: 1) natural between-year variations in the weather, 2) inaccuracies in predicting the second half-year of 2017, for which no time series data were available for validation (Section 8.3.2), or by 3) numerical error due to insufficient coverage of θ_{new} ; probably a mixture of all three. A similar predictive bias of approx. 10% was found for the time series validation of the urban test area of Harlev (Figure 8.5).

8.4.2 Building retrofit scenarios

Two different building energy renovation rate scenarios were analyzed in addition to a no-renovation scenario:

- A) A business-as-usual scenario with 1% of the building stock being renovated per year.
- B) An aggressive scenario with 5% of the building stock being renovated per year.

These two scenarios were selected to represent the lower and upper boundaries of what can be achieved. With a renovation rate of 1% (Scenario A), only 32% of the building stock will have undergone energy retrofit by the end of year 2049, while a renovation rate of 5% (Scenario B) means that all buildings are retrofitted by the end of year 2037. The reality may be that the achievable renovation rate is somewhere in between. Article 5 of the EED (Section 1.1) stipulates that a least 3% of all buildings owned and occupied by EU member state governments should be energy renovated each year to meet minimum energy performance requirements. No official data exist to assess whether this target is met; however, some studies reveal that the actual renovation rate may be even lower than 1% per year in practice (Filippidou et al., 2017).

Many different definitions exist regarding what can be considered to be energy renovation. In the EU there is a general understanding that a deep renovation should improve building energy use by at least 75% and/or have a primary energy consumption after renovation of less than 60 kWh/m²/yr (Shnapp et al., 2013). Such comprehensive renovations require a complete reconstruction of most energy systems and envelope elements that affect building energy use, and that is rare (Filippidou et al., 2017). A less ambitious but perhaps more realistically achievable level of retrofit was applied in this study with only three elements of the building envelope in focus: external walls, roof and windows.

All buildings were sorted in order of their heat loss coefficient (HLC) from highest to lowest after each annual simulation, starting from 2017 up until year 2049. The worst performing buildings in terms of highest HLC – e.g. the worst 1% in Scenario A – then had their three building elements retrofitted to a new standard if the current value of the building elements exceeded a given retrofit threshold value (Table 8.4). If one, or all three, of the building elements exceeded the retrofit threshold value, the value of the building element was replaced with the retrofit value corresponding to an energy upgrade/or replacement of that building element. Accordingly, only the worst-performing buildings in terms of HLC were energy renovated, and to various degrees depending on their current state. Some buildings only had one or two building elements retrofitted, while others would have all three building elements retrofitted. No building could be energy renovated more than once.

Table 8.4. Replacement values for the three building elements when engaged in retrofit. Only if the current building element value exceeds the retrofit threshold value is it replaced with the new standard value.

Building element	Retrofit threshold	New standard
Ext. walls	U-value >0.5 [W/(m ² K)]	U-value = 0.5 [W/(m ² K)]
Roof	U-value >0.5 [W/(m ² K)]	U-value = 0.3 [W/(m ² K)]
Windows	U-value >2.5 [W/(m ² K)]	U-value = 1.5 [W/(m ² K)]

The chain of 32 annual simulations of the period 2017-2050 were repeated $N_{\theta} = 200$ times for both renovation rate scenario A and B to account for the stochastic nature of the uncertain parameters assigned to individual buildings (i.e. the parameters consist of distributions of values and not fixed deterministic values). Each of the chains of simulations were started in year 2017 by randomly assigning a value from the archetype-specific posterior predictive distribution of θ_{new} to each building. Buildings were thereby exposed to a different characterization of its uncertain archetype parameters in each of the 200 chains, which, subsequently, also lead to 200 different sequences of buildings being energy renovated. These random Monte Carlo draws made it possible to propagate the uncertainty throughout the simulations and thereby explore the space of possible outcomes ($\epsilon_{\text{sampling}} \approx 7\%$).

8.4.3 Projected weather conditions

The internationally agreed upon RCPs – RCP2.6 (van Vuuren et al., 2011), RCP4.5 (Thomson et al., 2011), RCP6.0 and RCP8.5 (Riahi et al., 2011) – consist of four independent greenhouse gas concentration trajectories used for climate modelling. They are named after their radiative forcing level projections² in the year 2100 relative to pre-industrial age values (+2.6, +4.5, +6.0, and +8.5 W/m², respectively). As such, they describe four different future climate scenarios, all of which are considered potential outcomes depending on global greenhouse gas emissions in the forthcoming years. The RCPs are used in many scientific studies, e.g. in the climate change assessments by the Intergovernmental Panel on Climate Change (IPCC) under the United Nations.

Two of the RCPs were applied in this study, namely RCP2.6 and RCP8.5, which correspond to the most optimistic and most pessimistic scenario, respectively, in terms of global temperature increase in the year 2100. Necessary weather conditions for building energy modelling of the 22,914 SFHs in Aarhus (hourly values of air temperature and global horizontal irradiance) from 2017 up until 2050 were extracted from the ICHEC-EC-EARTH HIRHAM5 climate model (Christensen et al., 2007; Hazeleger et al., 2012) using projections generated on the basis of RCP2.6 and RCP8.5. Projections of air temperature and global horizontal irradiance is shown as annually averaged values from 2010 up until 2100 in Figure 8.7. However, only the values in the period 2017-2050 were used in the simulations of this study.

The mean annual temperature projections of the two RCP scenarios increase steadily with approx. 1°C in the simulation period from 2017-2050 (Figure 8.7A). Only subtle differences are present between the two projections to distinguish between them in the simulation period. Thereafter, the RCP2.6 projection flattens out, whereas the RCP8.5 projection continues its incline up until 2100. The projections of mean annual irradiance remain somewhat unchanged throughout the entire period from 2010-2100 (Figure 8.7B).

8.4.4 Citywide heating energy use forecasts from 2017-2050

The annually aggregated forecasts of district heating energy use for the 22,914 SFHs are shown in Figure 8.8A as the energy use intensity (aggregated energy use normalized by the aggregated heated floor area), using projected weather data from 2017 up until the end of 2049 for both climate scenarios (RCP2.6 and RCP8.5). The forecasts of how the building

²Radiative forcing is the difference in energy flux between insolation (sunlight) absorbed by the Earth and energy radiated back to space. A positive radiative forcing results in near-surface warming (Köhler et al., 2010).

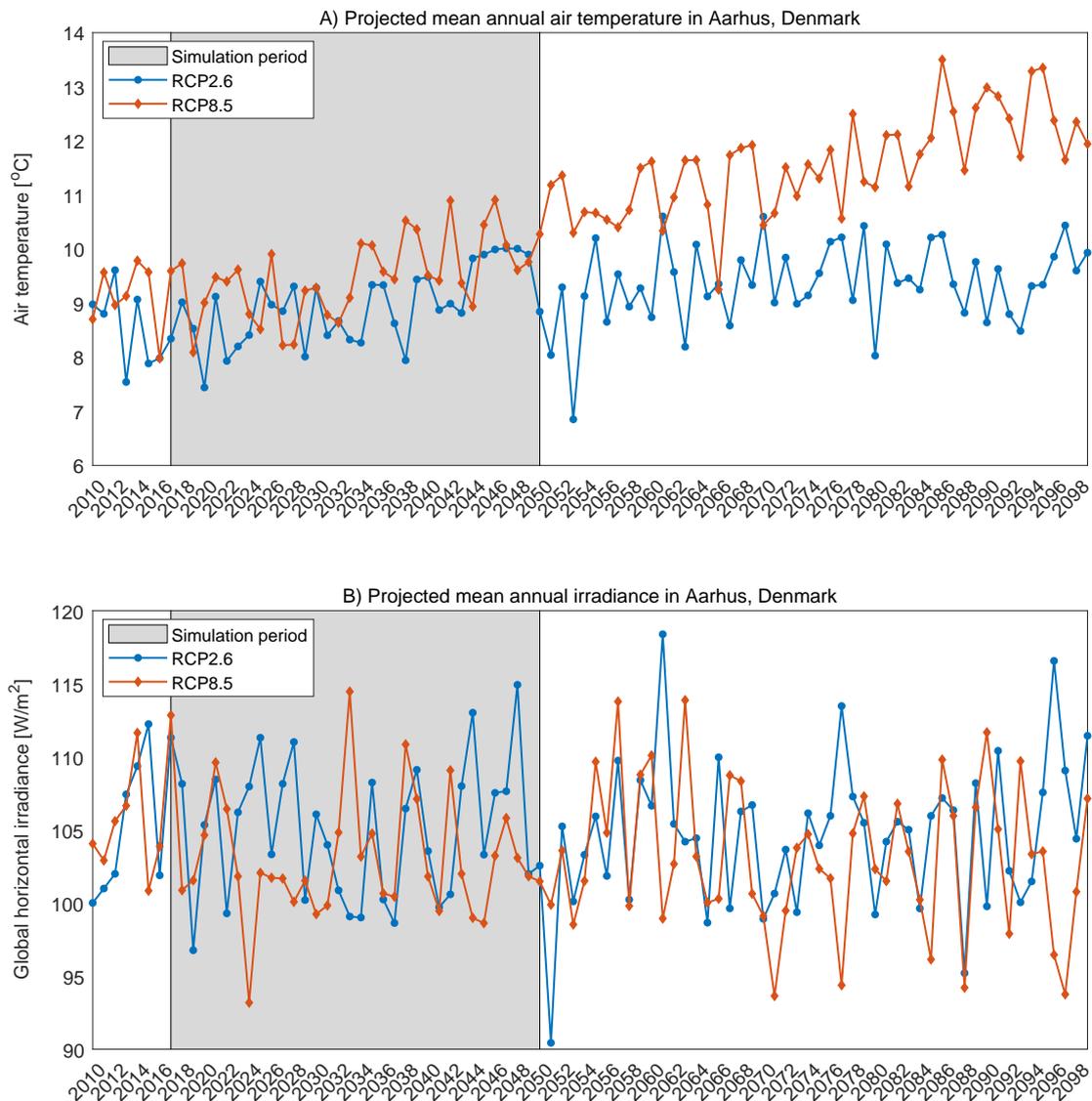


Figure 8.7. Projections of mean annual air temperature and global horizontal irradiance in Aarhus, Denmark, from 2010 up until 2100 using representative concentration pathways RCP2.6 and RCP8.5, respectively. Only the period shaded in grey (2017-2050) is used for simulation.

stock energy use evolves over the years are plotted for both energy renovation scenarios (1% and 5%) in addition to a no-renovation scenario (0%) for the sake of comparison. Each of the six trajectories comprise 200 chains of consecutive and interconnected MC simulations from the UBEM to cover the predictive probability distribution of the forecast. In Figure 8.8B is shown the relative energy use savings potential of the renovation rate scenarios (1% and 5%) compared to the no-renovation scenario (0%).

It is obvious from the forecasts (Figure 8.8A) that the weather has a large influence on the future energy use. Large fluctuations are present from year-to-year, even within the same RCP scenario; in fact, these variations are at the same level as the variations between the two RCP scenarios for the same energy renovation rate.

With a 5% energy renovation rate, the energy use quickly drops across the building stock to a new plateau around year 2033 where it fattens out relative to the no-renovation

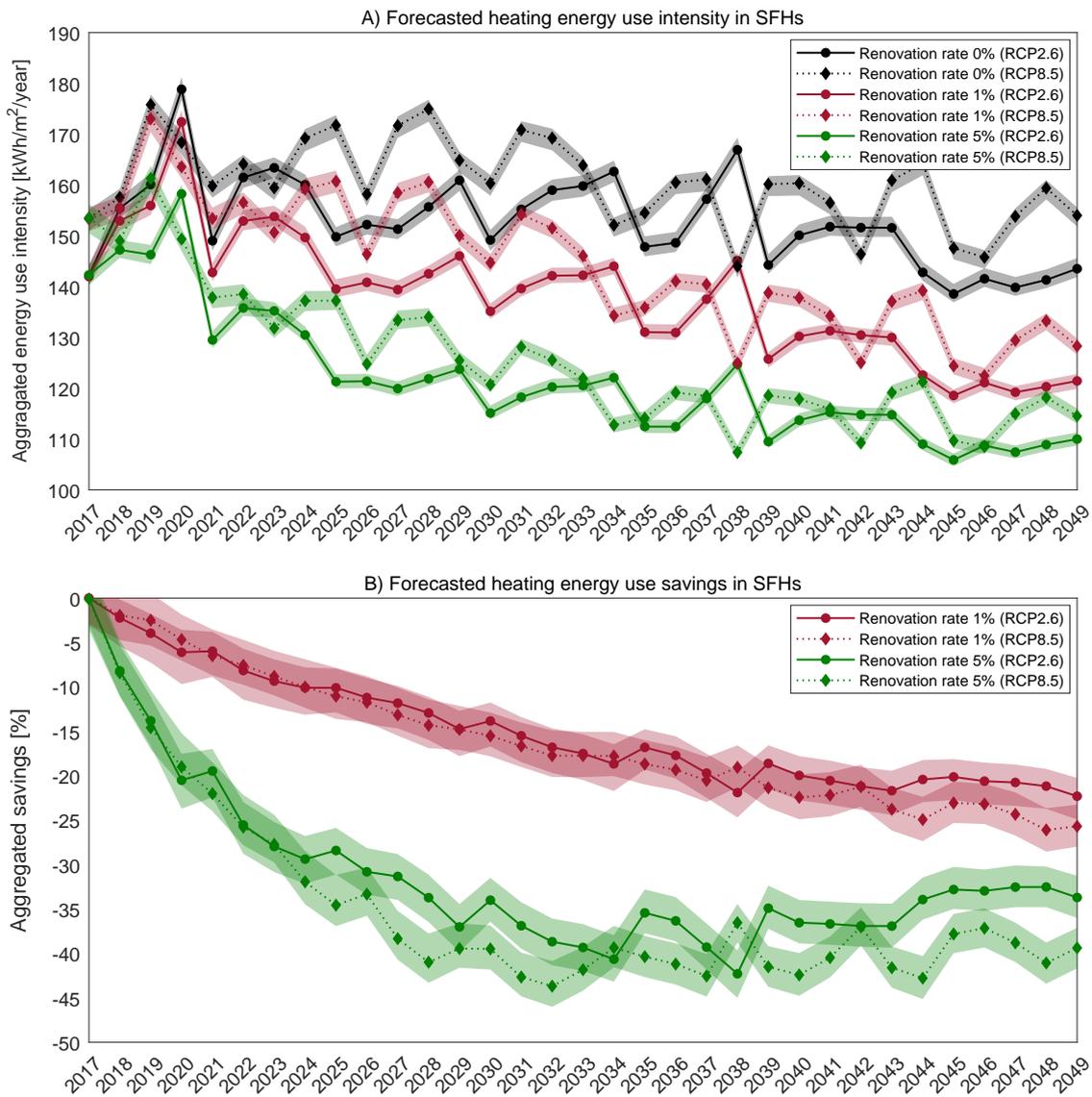


Figure 8.8. Forecasts of the aggregated energy use of 22,914 SFHs from 2017 up until 2050. A) Aggregated energy use intensity of the buildings for renovation rate scenario 0%, 1% and 5%, all simulated using both RCP-scenarios, respectively. Each of the trajectories constitute the mean of $N_{\theta} = 200$ chains of consecutive and interconnected simulations with the shaded bands visualizing the 95%-HPD of the simulations. B) Aggregated savings potential of a renovation rate scenario of 1% and 5%, respectively, relative to a 0% scenario (no renovation).

scenario (Figure 8.8B), which is when all buildings are simulated to have undergone retrofit to some degree. An aggregated savings potential of approx. 35-45% is forecasted compared to the no-renovation scenario, dependent on the RCP scenario in question; however, due to annual variations in the weather and climate change, which affects retrofitted and non-retrofitted buildings differently, the relative savings potential varies from year to year, even after all buildings have undergone energy renovation in 2033.

For the 1% scenario on the other hand, only 33% of the buildings are simulated to have undergone retrofit in 2050, which is why the two 1% trajectories decrease throughout the whole period without reaching the plateau of the 5% trajectories. This would theoretically

happen after 100 years in year 2117 if no further energy renovation were to be initiated. An aggregated savings potential of approx. 20-25% is forecasted compared to the no-renovation scenario in the end of year 2049.

8.5 Discussion

It has been shown how the hierarchical archetype model and calibration framework presented in Chapter 6 can be used to fit archetype models of an entire building stock; nine archetypes of SFHs in this case. In this chapter, it was applied using 3 months of hourly series training data from 75 training buildings per archetype, distributed across the first half-year of 2017, in an effort to further refine the capabilities and reliability of the originally proposed training scheme. The predictive capabilities were tested on the level of archetypes and urban test areas before being applied to populate an UBEM of Aarhus, and subsequently used to forecast the effects of energy renovation from 2017-2050.

8.5.1 Archetype modelling

From the results of testing the archetypes' predictive capabilities it is evident that not all archetypes perform equally satisfactorily in terms of imitating the peaks of the consumption profiles; in particular, the three newest archetypes are affected, which together cover buildings constructed from 1999 onwards. This effect could be caused by several phenomena, of which the following ones are considered most important:

- Newer SFHs are more diverse and unique in their geometry than older buildings. They also have larger window areas, and a larger proportion of the window area facing south, which violates the assumption of a fixed distribution of the windows. The scalable geometric model may thus not fit newer houses that well. Resulting from a larger window area facing south, newer houses may experience a larger amount of solar heat gain that could lead to fluctuations in the indoor air temperature.
- Newer buildings have faster reacting heating systems which might create larger peaks than the slower reacting systems of older buildings.
- The proportion of the DH energy use that is spent on the preparation of DHW is larger in newer houses than older houses. Newer houses furthermore draw DHW instantaneously using plate heat exchangers ($P \approx 30$ kW), which generate large peaks, whereas old houses to a larger extent are expected to use hot water tanks ($P \approx 5$ -15 kW), which charge slowly and thus may make the DHW energy use look like space heating. This phenomena is considered the largest and most influential difference between the older and newer archetype models.

To remedy these shortcomings, future work ought to focus on further rectifying the inadequacy of the model structure, especially the part that concerns geometry, DHW and occupancy in general. Such improvements would ensure a higher level of model representativeness, and identifiability of uncertain model parameters. This is particularly important in retrofit applications where these parameters are later altered without the ability to validate the resulting performance of the model. Further discussion on these matters is to be found in the discussion section of journal paper [P4] and the epilogue provided in Section 6.2.

Testing the predictive capabilities of the archetypes on the aggregated scale of three urban areas indicated a high out-of-sample predictive performance. This might be a consequence of the fact that most buildings belong to the pre-1999 archetypes, which perform best. Only one of the three urban test areas, the town of Harlev, had problems with biased predictions. Whether this was an effect of Harlev having a slightly newer building stock than the two other test areas (Table 8.3), or because of underlying demographic effects (occupant behaviour, attitudes and preferences), remains uncertain. However, we have previously suggested the existence of such effects across the city of Aarhus in a statistical study of annual district heating energy use, which is to be found in conference paper [S4] that is appended to this thesis.

8.5.2 Modelling energy renovations

Two energy renovation rates were incorporated in the analysis: 1% and 5% of the buildings per year, respectively. These rates encircle a space of potential outcome; the actual renovation rate is most likely going to be somewhere in between. However, assessing energy renovation on an aggregated scale is more complex than that. A relatively simple retrofit model for the renovation of individual buildings was applied in this study: Buildings were sorted for each year in order of their overall heat loss – the theoretically derived HLC – and the worst-performing buildings, 1% and 5%, respectively, were allowed to have their external walls, roofs, and windows retrofitted to a fixed new standard, independently, if their current standard were worse than some fixed threshold value (Table 8.4). This retrofit decision model is obviously a simplified representation of reality, but contains some degree of randomness and variation across buildings, which is desired, to account for the stochastic nature of occupants. However, future work should focus on further refining the decision model on a range of areas:

- The extent of the renovations considered; e.g. single building elements, several independent elements, or packages of coherent ECMs with their derived effects taken into account. An example of such a package could be the retrofit of envelope walls with the addition of external facade isolation and a simultaneous replacement and movement of windows out in the facade. This would further result in a reduced infiltration heat loss from increased air tightness, and lower linear thermal transmittance in window junctions.
- The magnitude of the individual ECMs; e.g. a simple replacement or refurbishment of worn building elements, e.g. windows, with new elements of minimum energy efficiency, or a much more ambitious retrofit that seeks to drastically reduce energy use by selecting low-energy solutions.
- The driver of building energy renovation; is it because the overall energy efficiency of the building is low, because building elements need to be refurbished, or due to politically driven factors, such as support by government subvention? Some buildings may never be energy renovated, even though they require it. Others may be retrofitted several times throughout the lifetime of the building.

Although simple, the energy renovation model applied in this study demonstrates a valuable application of the archetype framework, which can easily be expanded to accommodate other requirements and political targets.

8.5.3 Climate scenarios

Future climate data were projected using two of the four RCPs; the most optimistic (RCP2.6) and the most pessimistic (RCP8.5) trajectory, respectively, in terms of global radiative forcing levels in year 2100. However, a warmer climate is not necessarily a bad thing if looking at heating energy use in buildings from an isolated perspective. One might thus expect the overall warmer projections from RCP8.5 to result in the lowest energy use forecasts, but the results are not unambiguous in that respect; forecasts using the RCP2.6 and RCP8.5 climate projections exhibit a considerable degree of overlap (Figure 8.8). This could be due to the fact that the difference between the two RCPs are small in the simulation period from 2017-2050 (Figure 8.7). If the simulation period were to be extended further into the future, the forecasts would probably deviate significantly more. This suggest that the effects of climate change on building energy use are complex and cannot be accounted for by investigating individual projections of future climate. Future research using climate projections to assess building energy use therefore ought to include a larger ensemble of climate data, e.g. all four RCPs, to map the uncertainty of the climate.

Chapter 9

Thesis conclusion

9.1 Main contributions

The thesis set out to meet the challenges of bottom-up archetype-based urban energy modelling (Section 1.5), identified from a review of the current state-of-the-art, through the exploration of two research objectives (Section 1.6). A number of contributions toward resolving these objectives have been offered through a presentation and discussion of both published and unpublished work with the main contributions listed below.

Bayesian view on uncertainty: All work presented in this thesis is based on a Bayesian view on probability to facilitate a fully stochastic treatment and propagation of all sources of error throughout model inference, and to allow for prior knowledge about uncertain archetype parameters to be included, if available. This propagation of uncertainty has made it possible to quantify the uncertainty of predictions made with archetype models, which is useful to improve, for example, retrofit decision-making and the impact of energy policy. However, modelling and propagating uncertainty in a stochastic framework imposes an extra computational burden that the non-probabilistic counterparts do not have to bear.

Dynamic archetype modelling: Calibration data of different temporal resolution have been investigated in the thesis (Chapter 5). It is found that the application of dynamic building energy modelling, which offers the opportunity to learn archetype parameters using, for instance, hourly time series data, gives the best predictive performance. Moreover, the use of dynamic modelling has the ability to facilitate analysis on an hourly basis, or higher if needed, which is required in the analysis of energy systems, e.g. district heating system operation. It is found that the application of hourly time series data affects the inference of uncertain archetype parameters, as such data contain information about system dynamics that is not available with data of monthly or annual resolution.

Hierarchical archetype modelling: Archetype modelling is generally based on the assumption that a given building stock can be segmented into a semantic subset of homogeneous building archetypes. Quantifying and propagating the error of this assumption – the level of archetype heterogeneity – is vital for the successful implementation of archetypes in urban building energy modelling. A hierarchical (multi-level) archetype modelling framework is introduced in this thesis (Chapter 6), which incorporates the heterogeneity of a given archetype classification scheme and propagates this throughout the calibration of archetype parameters.

Fast simulation without statistical emulators: The application of statistical emulators – surrogate models – is widespread in the literature on Bayesian archetype parameter calibration and Bayesian model calibration in general. The Bayesian calibration methodology offered in this thesis is based on Markov Chain Monte Carlo tuning of uncertain parameters from a direct evaluation of the errors between simulated and measured time series outputs whereby the use of statistical emulators is rendered redundant. This approach ensures a more transparent calibration process and eliminates a layer of numerical error in obtaining posteriors over uncertain parameters. This direct calibration approach is viable because the underlying building energy model is based on a fast and simple 1st-order ROM formulation (ISO13790:2008 hourly dynamic model).

Urban-scale forecasts under uncertainty: A framework is offered to forecast urban building energy use under uncertainty of energy renovations and climate change (Chapter 8). This was demonstrated for the stock of approx. 23,000 district heating-supplied SFHs in the municipality of Aarhus, Denmark, from 2017 up until 2050, using the archetype-based modelling approach. The projection of citywide heating energy use was conducted for two different energy renovation rates and two different climate change scenarios to explore the future outcome. Such analyses are valuable in the context of energy policy decision-making, and planning and operation of energy supply systems. It would be relatively easy to implement other retrofit decision models, and to expand the framework to include other building types.

9.2 Limitations and future work

The models and methods introduced in the thesis have a number of limitations and areas that might be further developed to better fit the needs of specific applications. Several of these issues are already discussed in part in the thesis and relevant papers. They are recapitulated below together with a number of interesting future research directions resulting from the work.

Model fidelity: The underlying BEM relies on a simple 1st-order ROM formulation. The simplicity of the BEM structure allows for very fast simulations of hourly building energy use, which makes it possible to do without statistical emulators. However, it comes at a price of reduced fidelity and representation of system dynamics compared to more complex alternatives such as EnergyPlus, IDA ICE, etc. This also means that there is a limit to what can be included in the model and the analysis that can be performed. It may therefore be necessary to resort to more complex BEM formulations, possibly in connection with a statistical emulator to reduce computational time, in studies where the application requires enhanced control of system dynamics on an hourly basis or more. An example of this is the study reported in journal paper [S5] where we expanded the 1st-order ROM to a 2nd-order variant. This was done to accommodate an application where model predictive control (MPC) schemes were used to exploit the thermal mass of buildings in a neighbourhood-scale UBEM study – an analysis that requires a more detailed representation of the dynamic properties of buildings. However, we continued to work without emulating the BEM, as this 2nd-order ROM formulation remained relatively simple and very fast to evaluate.

Modelling occupant behaviour: The presence, behaviour and attitude of occupants is

assumed to be fixed and time-invariant. This is a limitation of the framework that reduces the analysis to be performed only on the aggregated scale where the effects of occupants are levelled out. Future work ought to focus on how to improve the representation of occupants on the building level.

Known input data: The implementation of additional input data, in addition to the public building characteristics data of the BDR database, is considered to be a rather straightforward way to improve model representativeness. Many recent UBEM studies have shown how the inclusion of GIS data can be used to extract geometrical building shape data for the generation of extruded massing models (e.g. [Davila et al., 2016](#); [Chen et al., 2017](#); [Jaeger et al., 2018](#)). In this context, state-of-the-art seems to revolve around the usage of the CityGML standard for establishing citywide GIS-based models. Moreover, building survey data, such as EPC schemes, are becoming readily available in many countries; data which contain a great deal of technical information to be used for modelling.

Archetype training and identification of model parameters: The calibration of uncertain model parameters is a central feature of the archetype framework. These tuned parameter estimates are used both to secure accurate model predictions, and in the subsequent retrofit analysis of the building stock. However, it is a fact that the 1st-order model structure applied in this thesis together with measurements of only external loads to the system, i.e. building energy use and weather conditions, encounters problems with the partitioning of heat loss into that of transmission and ventilation, as discussed in journal paper [P4] of Chapter 6. Several efforts have been proposed to remedy some of these shortcomings in Chapter 8, e.g. extending the training and testing period to 3 and 6 months of hourly data, respectively, and extending the number of training buildings from 50 to 75. However, it remains uncertain to which extent the inferred parameter estimates are consistent with those of the actual buildings; this would need to be tested in an controlled experimental set-up of a test-building.

Archetype classification and heterogeneity: The segmentation of building stocks into archetypes has never been the scope of this thesis; focus has solely been on the characterization and calibration of archetypes. Instead, the classification scheme offered by the TABULA project was applied in Chapter 8 with a small modification to account for newer buildings. However, optimizing the classification scheme to better fit the diversity of the building stock could potentially result in a large performance improvements of the subsequent archetype models, as the heterogeneity would be reduced. In this context, it would be interesting to further investigate the possibility of different classification schemes depending on the parameters and phenomena in question. For example: one archetype segmentation based on the construction year of the buildings to account for technical aspects, building traditions and architecture, and another segmentation based on, for instance, demography or location of the buildings to account for occupant-related aspects.

Other building categories: All research presented in this thesis has revolved around the application of case data from detached SFHs in Aarhus, Denmark. These buildings are relatively homogeneous in nature; stand-alone buildings with no thermal connection to adjacent buildings, single-family occupants, and only small variations in architecture and geometrical layout. However, in order to simulate all buildings in a city, it would be necessary to extend the archetype framework to cover all residential building categories, as well as commercial buildings, cultural

buildings, etc. (Chapter 2, Figure 2.1). For several of these building categories, e.g. industrial buildings or high-rise buildings of mixed-use, this could prove very challenging. Some buildings are so unique that they in themselves form an *archetype*, e.g. large cultural buildings or old city landmarks. One should thus not expect the same level of performance and homogeneity for the remaining building categories as obtained in this thesis using SFHs.

Decision model for energy renovation: The model applied to forecast energy renovation of the building stock from 2017-2050 relies on a relatively simple set of decision rules for the initiation of individual retrofit measures. This makes the energy renovation analysis both transparent and fast to integrate in the existing forecasting set up; however, it is suggested for future work to further expand on this approach to distinguish between simple and deep retrofit solutions, as well as how and when energy renovation is initiated in individual buildings.

Other urban-scale applications: Finally, the archetype-based UBEM framework could be applied in applications other than estimating the overall environmental footprint of the building stock and forecasting the effects of energy renovation, as demonstrated in this thesis. Utility companies, such as suppliers of district heating, may be interested in supply-side management applications for: i) forecasting daily load profiles to optimize district heating operation and supply, ii) forecasting the effects of running city densification (increasing load) in combination with running energy renovation of the ageing building stock (reducing load) to evaluate maintenance and expansion plans of existing distribution grids underneath the city, and iii) planning and dimensioning of new distribution grids in suburban areas subdivided for newly built houses.

References

- Aarhus Kommune (2018), ‘Aarhus i tal’, Interactive web tool [in Danish]. Accessed: 15.10.2018.
URL: <https://ledelsesinformation.aarhuskommune.dk/aarhus-i-tal/default.aspx?doc=vfs://Global/Befolkning-antal.xview> (Cited on p. 19.)
- AffaldVarme Aarhus (2018), ‘Aarhus i tal’, Online [in Danish]. Accessed: 17.10.2018.
URL: <https://affaldvarme.aarhus.dk/om-os/varmeforsyning-i-aarhus-omraadet/varmeproduktion/> (Cited on p. 21.)
- Aksoezen, M., Daniel, M., Hassler, U. and Kohler, N. (2015), ‘Building age as an indicator for energy consumption’, *Energy and Buildings* **87**, 74–86.
URL: <https://doi.org/10.1016/j.enbuild.2014.10.074> (Cited on p. 7.)
- Aydinalp-Koksal, M. and Ugursal, V. I. (2008), ‘Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector’, *Applied Energy* **85**(4), 271–296.
URL: <https://doi.org/10.1016/j.apenergy.2006.09.012> (Cited on p. 4.)
- Balaras, C. A., Dascalaki, E. G., Droutsa, K. G. and Kontoyiannidis, S. (2016), ‘Empirical assessment of calculated and actual heating energy use in Hellenic residential buildings’, *Applied Energy* **164**, 115–132.
URL: <https://doi.org/10.1016/j.apenergy.2015.11.027> (Cited on pp. 5 and 7.)
- Ballarini, I., Corgnati, S. P. and Corrado, V. (2014), ‘Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project’, *Energy Policy* **68**, 273–284.
URL: <https://doi.org/10.1016/j.enpol.2014.01.027> (Cited on p. 7.)
- Booth, A. and Choudhary, R. (2013), ‘Decision making under uncertainty in the retrofit analysis of the UK housing stock: Implications for the Green Deal’, *Energy and Buildings* **64**, 292–308.
URL: <https://doi.org/10.1016/j.enbuild.2013.05.014> (Cited on p. 9.)
- Booth, A. T., Choudhary, R. and Spiegelhalter, D. J. (2012), ‘Handling uncertainty in housing stock models’, *Building and Environment* **48**, 35–47.
URL: <https://doi.org/10.1016/j.buildenv.2011.08.016> (Cited on pp. 6, 8, 14, and 39.)
- Booth, A. T., Choudhary, R. and Spiegelhalter, D. J. (2013), ‘A hierarchical Bayesian framework for calibrating micro-level models with macro-level data’, *Journal of Building Performance Simulation* **6**, 293–318.
URL: <https://doi.org/10.1080/19401493.2012.723750> (Cited on p. 5.)
- Braulio-Gonzalo, M., Juan, P., Bovea, M. D. and Ruá, M. J. (2016), ‘Modelling energy efficiency performance of residential building stocks based on Bayesian statistical inference’,

- Environmental Modelling & Software* **83**, 198–211.
URL: <https://doi.org/10.1016/j.envsoft.2016.05.018> (Cited on p. 5.)
- Brøgger, M., Bacher, P., Madsen, H. and Wittchen, K. B. (2018), ‘Estimating the influence of rebound effects on the energy-saving potential in building stocks’, *Energy and Buildings*. [Accepted Manuscript, In Press]. (Cited on p. 5.)
- Brøgger, M. and Wittchen, K. B. (2018), ‘Estimating the energy-saving potential in national building stocks – A methodology review’, *Renewable and Sustainable Energy Reviews* **82**, 1489–1496.
URL: <https://doi.org/10.1016/j.rser.2017.05.239> (Cited on p. 15.)
- Buffat, R., Froemelt, A., Heeren, N., Raubal, M. and Hellweg, S. (2017), ‘Big data GIS analysis for novel approaches in building stock modelling’, *Applied Energy* **208**, 277–290.
URL: <https://doi.org/10.1016/j.apenergy.2017.10.041> (Cited on p. 5.)
- Caputo, P., Costa, G. and Ferrari, S. (2013), ‘A supporting method for defining energy strategies in the building sector at urban scale’, *Energy Policy* **55**, 261–270.
URL: <https://doi.org/10.1016/j.enpol.2012.12.006> (Cited on p. 5.)
- Carstens, H., Xia, X. and Yadavalli, S. (2018), ‘Bayesian Energy Measurement and Verification Analysis’, *Energies* **11**(2), 380.
URL: <https://doi.org/10.3390/en11020380> (Cited on p. 12.)
- Cerezo, C., Sokol, J., AlKhaled, S., Reinhart, C., Al-Mumin, A. and Hajiah, A. (2017), ‘Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City’, *Energy and Buildings* **154**, 321–334.
URL: <https://doi.org/10.1016/j.enbuild.2017.08.029> (Cited on pp. 6, 7, 8, 9, 12, 14, and 39.)
- Cerezo, C., Sokol, J., Reinhart, C. and Al-Mumin, A. (2015), Three methods for characterizing building archetypes in urban energy simulation. A case study in Kuwait City, in ‘Building Simulation 2015: 14th Conference of International Building Performance Simulation Association’, IBPSA, Hyderabad, India, pp. 2873–2880.
URL: <http://www.ibpsa.org/proceedings/BS2015/p2435.pdf> (Cited on p. 6.)
- Chen, Y., Hong, T. and Piette, M. A. (2017), ‘Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis’, *Applied Energy* **205**, 323–335.
URL: <https://doi.org/10.1016/j.apenergy.2017.07.128> (Cited on pp. 76 and 103.)
- Chong, A., Lam, K. P., Pozzi, M. and Yang, J. (2017), ‘Bayesian calibration of building energy models with large datasets’, *Energy and Buildings* **154**, 343–355.
URL: <https://doi.org/10.1016/j.enbuild.2017.08.069> (Cited on p. 14.)
- Chong, A. and Menberg, K. (2018), ‘Guidelines for the Bayesian calibration of building energy models’, *Energy and Buildings* **174**, 527–547.
URL: <https://doi.org/10.1016/j.enbuild.2018.06.028> (Cited on p. 12.)
- Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I. and Rinke, A. (2007), The HIRHAM Regional Climate Model. Version 5 (beta), Technical report, Danish Climate Centre, Danish Meteorological Institute. (Cited on p. 94.)

- Coakley, D., Raftery, P. and Keane, M. (2014), ‘A review of methods to match building energy simulation models to measured data’, *Renewable and Sustainable Energy Reviews* **37**, 123–141.
URL: <https://doi.org/10.1016/j.rser.2014.05.007> (Cited on p. 12.)
- Dahl, M. (2018), Production planning of energy systems: Cost and risk assessment for district heating, phdthesis, Aarhus University.
URL: <https://doi.org/10.7146/aul.270.188> (Cited on p. 21.)
- Dahl, M., Brun, A., Kirsebom, O. and Andresen, G. (2018), ‘Improving short-term heat load forecasts with calendar and holiday data’, *Energies* **11**(7), 1678.
URL: <https://doi.org/10.3390/en11071678> (Cited on p. 76.)
- Dall’O’, G., Galante, A. and Torri, M. (2012), ‘A methodology for the energy performance classification of residential building stock on an urban scale’, *Energy and Buildings* **48**, 211–219.
URL: <https://doi.org/10.1016/j.enbuild.2012.01.034> (Cited on p. 6.)
- Davila, C. C., Reinhart, C. F. and Bemis, J. L. (2016), ‘Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets’, *Energy* **117**, 237–250.
URL: <https://doi.org/10.1016/j.energy.2016.10.057> (Cited on pp. 5 and 103.)
- do Carmo, C. M. R. and Christensen, T. H. (2016), ‘Cluster analysis of residential heat load profiles and the role of technical and household characteristics’, *Energy and Buildings* **125**, 171–180.
URL: <https://doi.org/10.1016/j.enbuild.2016.04.079> (Cited on p. 7.)
- Dogan, T. and Reinhart, C. (2017), ‘Shoemaker: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation’, *Energy and Buildings* **140**, 140–153.
URL: <https://doi.org/10.1016/j.enbuild.2017.01.030> (Cited on p. 5.)
- Dominković, D., Gianniou, P., Münster, M., Heller, A. and Rode, C. (2018), ‘Utilizing thermal building mass for storage in district heating systems: Combined building level simulations and system level optimization’, *Energy* **153**, 949–966.
URL: <https://doi.org/10.1016/j.energy.2018.04.093> (Cited on pp. 5 and 6.)
- EIA (2018), Monthly Energy Review – August 2018, Technical report, U.S. Energy Information Administration (EIA). A monthly updated publication of recent and historical energy statistics.
URL: <https://www.eia.gov/totalenergy/data/monthly/> (Cited on p. 1.)
- European Commission (2017), ‘European Commission – Fact Sheet’, Online. Accessed: 9.11.2018.
URL: http://europa.eu/rapid/press-release_MEMO-17-162_en.htm (Cited on p. 1.)
- Famuyibo, A. A., Duffy, A. and Strachan, P. (2012), ‘Developing archetypes for domestic dwellings— An Irish case study’, *Energy and Buildings* **50**, 150–157.
URL: <https://doi.org/10.1016/j.enbuild.2012.03.033> (Cited on pp. 6 and 7.)
- Filippidou, F., Nieboer, N. and Visscher, H. (2017), ‘Are we moving fast enough? The energy renovation rate of the Dutch non-profit housing using the national energy labelling database’, *Energy Policy* **109**, 488–498.
URL: <https://doi.org/10.1016/j.enpol.2017.07.025> (Cited on p. 93.)
- Fracastoro, G. V. and Serraino, M. (2011), ‘A methodology for assessing the energy performance of large scale building stocks and possible applications’, *Energy and Buildings* **43**(4), 844–852.
URL: <https://doi.org/10.1016/j.enbuild.2010.12.004> (Cited on p. 6.)

- Frigola-Alcalde, R. (2015), Bayesian Time Series Learning with Gaussain Processes, phdthesis, Department of Engineering, University of Cambridge. (Cited on pp. 9 and 58.)
- Gianniou, P., Liu, X., Heller, A., Nielsen, P. S. and Rode, C. (2018), ‘Clustering-based analysis for residential district heating data’, *Energy Conversion and Management* **165**, 840–850.
URL: <https://doi.org/10.1016/j.enconman.2018.03.015> (Cited on p. 7.)
- Happle, G., Fonseca, J. A. and Schlueter, A. (2018), ‘A review on occupant behavior in urban building energy models’, *Energy and Buildings* **174**, 276–292.
URL: <https://doi.org/10.1016/j.enbuild.2018.06.030> (Cited on p. 76.)
- Hazeleger, W., Wang, X., Severijns, C., Ștefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B. et al. (2012), ‘EC-Earth V2.2: description and validation of a new seamless earth system prediction model’, *Climate dynamics* **39**(11), 2611–2629. (Cited on p. 94.)
- Heiple, S. and Sailor, D. J. (2008), ‘Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles’, *Energy and Buildings* **40**(8), 1426–1436.
URL: <https://doi.org/10.1016/j.enbuild.2008.01.005> (Cited on pp. 5 and 6.)
- Heiselberg, P., Brohus, H., Hesselholt, A., Rasmussen, H., Seinre, E. and Thomas, S. (2009), ‘Application of sensitivity analysis in design of sustainable buildings’, *Renewable Energy* **34**(9), 2030–2036.
URL: <https://doi.org/10.1016/j.renene.2009.02.016> (Cited on p. 9.)
- Heo, Y. (2011), Bayesian calibration of building energy models for energy retrofit decision-making under uncertainty, phdthesis, College of Architecture, Georgia Institute of Technology. (Cited on p. 8.)
- Heo, Y., Augenbroe, G., Graziano, D., Muehleisen, R. T. and Guzowski, L. (2015), ‘Scalable methodology for large scale building energy improvement: Relevance of calibration in model-based retrofit analysis’, *Building and Environment* **87**, 342–350.
URL: <https://doi.org/10.1016/j.buildenv.2014.12.016> (Cited on p. 14.)
- Heo, Y., Choudhary, R. and Augenbroe, G. A. (2012), ‘Calibration of building energy models for retrofit analysis under uncertainty’, *Energy and Buildings* **47**, 550–560.
URL: <https://doi.org/10.1016/j.enbuild.2011.12.029> (Cited on p. 14.)
- Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008), ‘Computer model calibration using high-dimensional output’, *Journal of the American Statistical Association* **103**(482), 570–583.
URL: <https://doi.org/10.1198/016214507000000888> (Cited on p. 58.)
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A. and Ryne, R. D. (2004), ‘Combining field data and computer simulations for calibration and prediction’, *SIAM Journal on Scientific Computing* **26**(2), 448–466.
URL: <https://doi.org/10.1137/s1064827503426693> (Cited on p. 14.)
- Hsu, D. (2014), ‘How much information disclosure of building energy performance is necessary?’, *Energy Policy* **64**, 263–272.
URL: <https://doi.org/10.1016/j.enpol.2013.08.094> (Cited on p. 5.)
- Hsu, D. (2015), ‘Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data’, *Applied Energy* **160**, 153–163.
URL: <https://doi.org/10.1016/j.apenergy.2015.08.126> (Cited on p. 7.)

- Huebner, G. M., Hamilton, I., Chalabi, Z., Shipworth, D. and Oreszczyn, T. (2015), ‘Explaining domestic energy consumption – The comparative contribution of building factors, socio-demographics, behaviours and attitudes’, *Applied Energy* **159**, 589–600.
URL: <https://doi.org/10.1016/j.apenergy.2015.09.028> (Cited on p. 8.)
- Jaeger, I. D., Reynders, G., Ma, Y. and Saelens, D. (2018), ‘Impact of building geometry description within district energy simulations’, *Energy* **158**, 1060–1069.
URL: <https://doi.org/10.1016/j.energy.2018.06.098> (Cited on pp. 76 and 103.)
- Kang, Y. and Krarti, M. (2016), ‘Bayesian-Emulator based parameter identification for calibrating energy models for existing buildings’, *Building Simulation* **9**, 411–428.
URL: <https://doi.org/10.1007/s12273-016-0291-6> (Cited on p. 14.)
- Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z. and Djurovic-Petrovic, M. (2010), ‘A review of bottom-up building stock models for energy consumption in the residential sector’, *Building and Environment* **45**(7), 1683–1697.
URL: <https://doi.org/10.1016/j.buildenv.2010.01.021> (Cited on p. 3.)
- Kennedy, M. C. and O’Hagan, A. (2001), ‘Bayesian calibration of computer models’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(3), 425–464. (Cited on pp. 14 and 39.)
- Kim, Y.-J. and Park, C.-S. (2016), ‘Stepwise deterministic and stochastic calibration of an energy simulation model for an existing building’, *Energy and Buildings* **133**, 455–468.
URL: <https://doi.org/10.1016/j.enbuild.2016.10.009> (Cited on p. 14.)
- Kim, Y.-J., Yoon, S.-H. and Park, C.-S. (2013), ‘Stochastic comparison between simplified energy calculation and dynamic simulation’, *Energy and Buildings* **64**, 332–342.
URL: <https://doi.org/10.1016/j.enbuild.2013.05.026> (Cited on p. 14.)
- Köhler, P., Bintanja, R., Fischer, H., Joos, F., Knutti, R., Lohmann, G. and Masson-Delmotte, V. (2010), ‘What caused Earth’s temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints on climate sensitivity’, *Quaternary Science Reviews* **29**(1-2), 129–145.
URL: <https://doi.org/10.1016/j.quascirev.2009.09.026> (Cited on p. 94.)
- Kontokosta, C. E. and Tull, C. (2017), ‘A data-driven predictive model of city-scale energy use in buildings’, *Applied Energy* **197**, 303–317.
URL: <https://doi.org/10.1016/j.apenergy.2017.04.005> (Cited on p. 5.)
- Kristensen, M. H., Brun, A. and Petersen, S. (2018), ‘Predicting Danish residential heating energy use from publicly available building characteristics’, *Energy and Buildings* **173**, 28–37.
URL: <https://doi.org/10.1016/j.enbuild.2018.05.011> (Cited on pp. xi and 5.)
- Kristensen, M. H., Choudhary, R., Pedersen, R. H. and Petersen, S. (2017), Bayesian calibration of residential building clusters using a single geometric building representation, in ‘Proceedings of Building Simulation 2017: The 15th IBPSA Conference’, San Francisco, USA, pp. 1294–1303.
URL: http://www.ibpsa.org/proceedings/BS2017/BS2017_330.pdf (Cited on pp. xi, 6, and 14.)
- Kristensen, M. H., Choudhary, R. and Petersen, S. (2017), ‘Bayesian calibration of building energy models: Comparison of predictive accuracy using metered utility data of different temporal resolution’, *Energy Procedia* **122**, 277–282. Proceedings of the CISBAT International Conference – Future Buildings & Districts.
URL: <https://doi.org/10.1016/j.egypro.2017.07.322> (Cited on pp. xi and 14.)

- Kristensen, M. H., Hedegaard, R. E. and Petersen, S. (2018a), ‘Hierarchical calibration of archetypes for urban building energy modeling’, *Energy and Buildings* **175**, 219–234.
URL: <https://doi.org/10.1016/j.enbuild.2018.07.030> (Cited on pp. xi, 6, and 14.)
- Kristensen, M. H., Hedegaard, R. E. and Petersen, S. (2018b), Urban-scale dynamic building energy modeling and prediction using hierarchical archetypes: A case study of two Danish towns, in ‘Proceedings of BSO 2018: The 4th Building Simulation and Optimization Conference’, IBPSA-England, Cambridge, UK, pp. 348–349. (Cited on p. xi.)
- Kristensen, M. H. and Petersen, S. (2016), ‘Choosing the appropriate sensitivity analysis method for building energy model-based investigations’, *Energy and Buildings* **130**, 166–176.
URL: <https://doi.org/10.1016/j.enbuild.2016.08.038> (Cited on p. xii.)
- Kristensen, M. H. and Petersen, S. (2017), ‘Explaining variability in metered energy use for similar buildings using Bayesian inference’, *Energy Procedia* **132**, 897–902. Proceedings of NSB 2017: The 11th Nordic Symposium on Building Physics.
URL: <https://doi.org/10.1016/j.egypro.2017.09.709> (Cited on p. xii.)
- Kristensen, M. H. and Petersen, S. (2018), Contrasting the capabilities of three different sensitivity analysis methods for building energy model-based investigations, in ‘Proceedings of BSO 2018: The 4th Building Simulation and Optimization Conference’, IBPSA-England, Cambridge, UK, pp. 560–566. (Cited on p. xii.)
- Lee, S. H., Hong, T., Piette, M. A. and Taylor-Lange, S. C. (2015), ‘Energy retrofit analysis toolkits for commercial buildings: A review’, *Energy* **89**, 1087–1100.
URL: <https://doi.org/10.1016/j.energy.2015.06.112> (Cited on pp. 3 and 4.)
- Li, Q., Augenbroe, G. and Brown, J. (2016), ‘Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction’, *Energy and Buildings* **124**, 194–202.
URL: <https://doi.org/10.1016/j.enbuild.2016.04.025> (Cited on p. 14.)
- Li, W., Zhou, Y., Cetin, K., Eom, J., Wang, Y., Chen, G. and Zhang, X. (2017), ‘Modeling urban building energy use: A review of modeling approaches and procedures’, *Energy* **141**, 2445–2457.
URL: <https://doi.org/10.1016/j.energy.2017.11.071> (Cited on pp. 2 and 3.)
- Lim, H. and Zhai, Z. J. (2017a), ‘Comprehensive evaluation of the influence of meta-models on Bayesian calibration’, *Energy and Buildings* **155**, 66–75.
URL: <https://doi.org/10.1016/j.enbuild.2017.09.009> (Cited on p. 13.)
- Lim, H. and Zhai, Z. J. (2017b), ‘Review on stochastic modeling methods for building stock energy prediction’, *Building Simulation* **10**, 607–624.
URL: <https://doi.org/10.1007/s12273-017-0383-y> (Cited on p. 12.)
- Lim, H. and Zhai, Z. J. (2018), ‘Influences of energy data on Bayesian calibration of building energy model’, *Applied Energy* **231**, 686–698.
URL: <https://doi.org/10.1016/j.apenergy.2018.09.156> (Cited on p. 12.)
- Lin, F.-Y., Lin, T.-P. and Hwang, R.-L. (2017), ‘Using geospatial information and building energy simulation to construct urban residential energy use map with high resolution for Taiwan cities’, *Energy and Buildings* **157**, 166–175.
URL: <https://doi.org/10.1016/j.enbuild.2017.01.040> (Cited on p. 5.)
- Liu, F. and West, M. (2009), ‘A dynamic modelling strategy for Bayesian computer model emulation’, *Bayesian Analysis* **4**(2), 393–411.
URL: <https://doi.org/10.1214/09-ba415> (Cited on p. 58.)

- Loga, T., Stein, B. and Diefenbach, N. (2016), ‘TABULA building typologies in 20 European countries – Making energy-related features of residential building stocks comparable’, *Energy and Buildings* **132**, 4–12.
URL: <https://doi.org/10.1016/j.enbuild.2016.06.094> (Cited on pp. 7 and 84.)
- Madsen, H. (2008), *Time Series Analysis*, Chapman & Hall / CRC Press. (Cited on p. 4.)
- Madsen, H., Bacher, P., Bauwens, G., Deconinck, A.-H., Reynders, G., Roels, S., Himpe, E. and Lethe, G. (2015), Thermal performance characterization using time series data – IEA EBC Annex 58 Guidelines, techreport, International Energy Agency, Lyngby, Denmark. (Cited on p. 14.)
- Manfredi, M., Aste, N. and Moshksar, R. (2013), ‘Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation’, *Applied Energy* **103**, 627–641.
URL: <https://doi.org/10.1016/j.apenergy.2012.10.031> (Cited on p. 14.)
- Mastrucci, A., Baume, O., Stazi, F. and Leopold, U. (2014), ‘Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam’, *Energy and Buildings* **75**, 358–367.
URL: <https://doi.org/10.1016/j.enbuild.2014.02.032> (Cited on p. 5.)
- Menberg, K., Heo, Y. and Choudhary, R. (2016), ‘Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information’, *Energy and Buildings* **133**, 433–445.
URL: <https://doi.org/10.1016/j.enbuild.2016.10.005> (Cited on p. 9.)
- Moghadam, S. T., Toniolo, J., Mutani, G. and Lombardi, P. (2018), ‘A GIS-statistical approach for assessing built environment energy use at urban scale’, *Sustainable Cities and Society* **37**, 70–84.
URL: <https://doi.org/10.1016/j.scs.2017.10.002> (Cited on p. 5.)
- Nageler, P., Koch, A., Mauthner, F., Leusbrock, I., Mach, T., Hochenauer, C. and Heimrath, R. (2018), ‘Comparison of dynamic urban building energy models (UBEM): Sigmoid energy signature and physical modelling approach’, *Energy and Buildings* . [In Press, Accepted manuscript.
URL: <https://doi.org/10.1016/j.enbuild.2018.09.034> (Cited on pp. 5 and 6.)
- Nageler, P., Zahrer, G., Heimrath, R., Mach, T., Mauthner, F., Leusbrock, I., Schranzhofer, H. and Hochenauer, C. (2017), ‘Novel validated method for GIS based automated dynamic urban building energy simulations’, *Energy* **139**, 142–154.
URL: <https://doi.org/10.1016/j.energy.2017.07.151> (Cited on pp. 5 and 76.)
- Nagpal, S., Mueller, C., Aijazi, A. and Reinhart, C. F. (2018), ‘A methodology for auto-calibrating urban building energy models using surrogate modeling techniques’, *Journal of Building Performance Simulation* pp. 1–16.
URL: <https://doi.org/10.1080/19401493.2018.1457722> (Cited on p. 14.)
- Nouvel, R., Mastrucci, A., Leopold, U., Baume, O., Coors, V. and Eicker, U. (2015), ‘Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support’, *Energy and Buildings* **107**, 204–212.
URL: <https://doi.org/10.1016/j.enbuild.2015.08.021> (Cited on p. 5.)
- Oladokun, M. G. and Odesola, I. A. (2015), ‘Household energy consumption and carbon emissions for sustainable cities – A critical review of modelling approaches’, *International Journal of Sustainable Built Environment* **4**(2), 231–247.
URL: <https://doi.org/10.1016/j.ijbsbe.2015.07.005> (Cited on pp. 3 and 4.)

- Österbring, M., Mata, É., Thuvander, L., Mangold, M., Johnsson, F. and Wallbaum, H. (2016), ‘A differentiated description of building-stocks for a georeferenced urban bottom-up building-stock model’, *Energy and Buildings* **120**, 78–84.
URL: <https://doi.org/10.1016/j.enbuild.2016.03.060> (Cited on p. 5.)
- Panão, M. J. O. and Brito, M. C. (2018), ‘Modelling aggregate hourly electricity consumption based on bottom-up building stock’, *Energy and Buildings* **170**, 170–182.
URL: <https://doi.org/10.1016/j.enbuild.2018.04.010> (Cited on p. 5.)
- Petersen, S., Kristensen, M. H. and Knudsen, M. D. (2019), ‘Prerequisites for reliable sensitivity analysis of a high fidelity building energy model’, *Energy and Buildings* **183**, 1–16.
URL: <https://doi.org/10.1016/j.enbuild.2018.10.035> (Cited on p. xii.)
- Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, second edn, The MIT Press. (Cited on p. 14.)
- Reinhart, C. F. and Davila, C. C. (2016), ‘Urban building energy modeling – A review of a nascent field’, *Building and Environment* **97**, 196–202.
URL: <https://doi.org/10.1016/j.buildenv.2015.12.001> (Cited on pp. 3, 5, and 6.)
- Remmen, P., Lauster, M., Mans, M., Fuchs, M., Osterhage, T. and Müller, D. (2017), ‘TEASER: an open tool for urban energy modelling of building stocks’, *Journal of Building Performance Simulation* **11**(1), 84–98.
URL: <https://doi.org/10.1080/19401493.2017.1283539> (Cited on p. 5.)
- Reynders, G., Diriken, J. and Saelens, D. (2014), Bottom-up modeling of the Belgian residential building stock: influence of model complexity, in ‘Proceedings of SSB 2014: The 9th International Conference on System Simulation in Buildings’, Liege, Belgium. (Cited on p. 5.)
- Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N. and Rafaj, P. (2011), ‘RCP 8.5 – A scenario of comparatively high greenhouse gas emissions’, *Climatic Change* **109**(1-2), 33.
URL: <https://doi.org/10.1007/s10584-011-0149-y> (Cited on p. 94.)
- Sajn, N. (2016), Energy efficiency of buildings – A nearly-zero energy future?, Technical report, European Parliamentary Research Service (EPRS). Members’ Research Service PE 582.022.
URL: [http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/582022/EPRS_BRI\(2016\)582022_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/582022/EPRS_BRI(2016)582022_EN.pdf) (Cited on p. 1.)
- Sandberg, N. H., Sartori, I. and Brattebø, H. (2014a), ‘Sensitivity analysis in long-term dynamic building stock modeling — Exploring the importance of uncertainty of input parameters in Norwegian segmented dwelling stock model’, *Energy and Buildings* **85**, 136–144.
URL: <https://doi.org/10.1016/j.enbuild.2014.09.028> (Cited on p. 2.)
- Sandberg, N. H., Sartori, I. and Brattebø, H. (2014b), ‘Using a dynamic segmented model to examine future renovation activities in the Norwegian dwelling stock’, *Energy and Buildings* **82**, 287–295.
URL: <https://doi.org/10.1016/j.enbuild.2014.07.005> (Cited on p. 2.)
- Sandberg, N. H., Sartori, I., Heidrich, O., Dawson, R., Dascalaki, E., Dimitriou, S., Vimm-r, T., Filippidou, F., Stegnar, G., Zavrl, M. Š. and Brattebø, H. (2016), ‘Dynamic building stock modelling: Application to 11 European countries to support the energy efficiency and retrofit ambitions of the EU’, *Energy and Buildings* **132**, 26–38.
URL: <https://doi.org/10.1016/j.enbuild.2016.05.100> (Cited on p. 2.)

- Sandberg, N. H., Sartori, I., Vestrum, M. I. and Brattebø, H. (2016), ‘Explaining the historical energy use in dwelling stocks with a segmented dynamic model: Case study of Norway 1960–2015’, *Energy and Buildings* **132**, 141–153.
URL: <https://doi.org/10.1016/j.enbuild.2016.05.099> (Cited on p. 2.)
- Sandberg, N. H., Sartori, I., Vestrum, M. I. and Brattebø, H. (2017), ‘Using a segmented dynamic dwelling stock model for scenario analysis of future energy demand: The dwelling stock of Norway 2016–2050’, *Energy and Buildings* **146**, 220–232.
URL: <https://doi.org/10.1016/j.enbuild.2017.04.016> (Cited on p. 2.)
- Shimoda, Y., Fujii, T., Morikawa, T. and Mizuno, M. (2004), ‘Residential end-use energy simulation at city scale’, *Building and Environment* **39**(8), 959–967.
URL: <https://doi.org/10.1016/j.buildenv.2004.01.020> (Cited on p. 5.)
- Shnapp, S., Sitja, R. and Laustsen, J. (2013), What is a deep renovation definition?, techreport, Global Buildings Performance Network (GBPN). (Cited on p. 93.)
- Sokol, J., Davila, C. C. and Reinhart, C. F. (2017), ‘Validation of a Bayesian-based method for defining residential archetypes in urban building energy models’, *Energy and Buildings* **134**, 11–24.
URL: <https://doi.org/10.1016/j.enbuild.2016.10.050> (Cited on pp. 5, 6, 7, and 14.)
- Swan, L. G. and Ugursal, V. I. (2009), ‘Modeling of end-use energy consumption in the residential sector: A review of modeling techniques’, *Renewable and Sustainable Energy Reviews* **13**(8), 1819–1835.
URL: <https://doi.org/10.1016/j.rser.2008.09.033> (Cited on pp. 3 and 4.)
- Theodoridou, I., Papadopoulos, A. M. and Hegger, M. (2011), ‘A typological classification of the Greek residential building stock’, *Energy and Buildings* **43**(10), 2779–2787.
URL: <https://doi.org/10.1016/j.enbuild.2011.06.036> (Cited on p. 6.)
- Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., Delgado-Arias, S., Bond-Lamberty, B., Wise, M. A., Clarke, L. E. et al. (2011), ‘RCP4.5: a pathway for stabilization of radiative forcing by 2100’, *Climatic change* **109**(1-2), 77.
URL: <https://doi.org/10.1007/s10584-011-0151-4> (Cited on p. 94.)
- Tian, W. (2013), ‘A review of sensitivity analysis methods in building energy analysis’, *Renewable and Sustainable Energy Reviews* **20**, 411–419.
URL: <https://doi.org/10.1016/j.rser.2012.12.014> (Cited on p. 9.)
- Tian, W. and Choudhary, R. (2012), ‘A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London’, *Energy and Buildings* **54**, 1–11.
URL: <https://doi.org/10.1016/j.enbuild.2012.06.031> (Cited on pp. 6 and 14.)
- Tian, W., Heo, Y., de Wilde, P., Li, Z., Yan, D., Park, C. S., Feng, X. and Augenbroe, G. (2018), ‘A review of uncertainty analysis in building energy assessment’, *Renewable and Sustainable Energy Reviews* **93**, 285–301.
URL: <https://doi.org/10.1016/j.rser.2018.05.029> (Cited on p. 9.)
- Tso, G. K. and Yau, K. K. (2007), ‘Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks’, *Energy* **32**(9), 1761–1768.
URL: <https://doi.org/10.1016/j.energy.2006.11.010> (Cited on p. 4.)

- Tuominen, P., Holopainen, R., Eskola, L., Jokisalo, J. and Airaksinen, M. (2014), ‘Calculation method and tool for assessing energy consumption in the building stock’, *Building and Environment* **75**, 153–160.
URL: <https://doi.org/10.1016/j.buildenv.2014.02.001> (Cited on p. 5.)
- United Nations (2015), ‘Paris Agreement’. Agreement resulting from the 21st Conference of the Parties (COP 21) of the United Nations Framework Convention on Climate Change (UNFCCC).
URL: https://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf (Cited on p. 1.)
- van Vuuren, D. P., Stehfest, E., den Elzen, M. G. J., Kram, T., van Vliet, J., Deetman, S., Isaac, M., Goldewijk, K. K., Hof, A., Beltran, A. M., Oostenrijk, R. and van Ruijven, B. (2011), ‘RCP2.6: exploring the possibility to keep global mean temperature increase below 2°C’, *Climatic Change* **109**(1-2), 95–116.
URL: <https://doi.org/10.1007/s10584-011-0152-3> (Cited on p. 94.)
- Wittchen, K. B. and Kragh, J. (2012), SBi 2012:01 Danish building typologies – Participation in the TABULA project, Technical report, Danish Building Research Institute, Aalborg University, Dr. Neergaards Vej 15, DK-2970 Hørsholm.
URL: <https://sbi.dk/Assets/Danish-building-typologies/sbi-2012-01-pdf.pdf> (Cited on pp. 84 and 86.)
- Zhao, F., Lee, S. H. and Augenbroe, G. (2016), ‘Reconstructing building stock to replicate energy consumption data’, *Energy and Buildings* **117**, 301–312.
URL: <https://doi.org/10.1016/j.enbuild.2015.10.001> (Cited on pp. 5 and 14.)
- Zucker, G., Judex, F., Blöchle, M., Köstl, M., Widl, E., Hauer, S., Bres, A. and Zeilinger, J. (2016), ‘A new method for optimizing operation of large neighborhoods of buildings using thermal simulation’, *Energy and Buildings* **125**, 153–160.
URL: <https://doi.org/10.1016/j.enbuild.2016.04.081> (Cited on p. 5.)

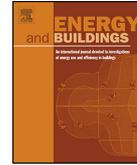
Appendix

Secondary publications



Contents lists available at ScienceDirect

Energy and Buildings

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

Choosing the appropriate sensitivity analysis method for building energy model-based investigations



Martin Heine Kristensen*, Steffen Petersen

Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus, Denmark

ARTICLE INFO

Article history:

Received 17 June 2016

Received in revised form 27 July 2016

Accepted 12 August 2016

Available online 18 August 2016

Keywords:

Sensitivity analysis

Morris method

Sobol method

Building energy modelling

Input information

Probability density functions

ABSTRACT

Current literature holds various examples of investigations that make use of a building energy model (BEM) combined with a sensitivity analysis (SA) technique to identify and rank the BEM input parameters that the model output is most sensitive to. However, a sound argumentation that vouches for the reliability, validity and necessary complexity of the chosen SA method for the specific purpose of the BEM-based analysis, is rare. This paper reports on an investigation of how two different levels of a-priori information about input parameters, applied to three different SA methods (Local, Morris and Sobol), influenced the identification and ranking of the input parameters that the annual energy need output of a quasi-steady-state BEM using monthly time steps, and a simple dynamic BEM using hourly time steps, is most sensitive to. It was found that the three SA methods, to a great extent, were able to identify the same cluster of most sensitive input parameters, independent of the level of a-priori input parameter information and BEM. However, the ranking of most sensitive input parameters varied with the applied SA method, BEM, and level of a-priori input parameter information. From a practical point of view, the choice of appropriate SA method is concluded to depend on the purpose of the SA analysis.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In the context of using building energy modelling (BEM) for performance predictions, it is often valuable to conduct a sensitivity analysis (SA) to explore the model behaviour and to identify which input parameters that drive the majority of the model output variation. Sensitivity analysis is thus a generic description of different techniques for quantification of how variability in model output can be apportioned to the variability and uncertainty of the model input parameters. SA methods are often categorised as either local sensitivity analysis (LSA) or global sensitivity analysis (GSA) [1].

1.1. Local sensitivity analysis

LSA methods rely on an OAT-methodology (one-parameter-at-a-time) where the effect of the variation of a single input parameter to a BEM tool is valued at discrete points of the input space while all other input parameters are held constant at their reference value. The nature and behaviour of the input parameters are not taken into account, i.e. all values have an equal probability of

occurrence without considering the effect of range and shape of the probability density function (PDF). Consequently, LSA methods do not consider any effects from correlated input parameters or any non-linear and non-additive model behaviour.

LSA methods have previously been used for various types of BEM-based analysis [2–5]. Petersen and Svendsen [2] used LSA together with a simple hourly dynamic BEM tool to provide building designers with an overview of the consequences of adjusting a performance-decisive parameter, in terms of energy performance and indoor environment, prior to any actual design decision. In a study of macro-parameters for net zero energy building design by Sun [3], LSA was used to quantify the impact of design parameters on the outcome of BEM tool TRNSYS. Besides this direct application, LSA has also been used to reduce large sets of parameters to smaller manageable sets before applying more complicated and computationally demanding GSA methods [4,5].

1.2. Global sensitivity analysis

GSA is a generic description of methods that evaluates the effect of an input parameter on the output by varying not only the parameter in question, but all other input parameters chosen for analysis as well. GSA uses a probabilistic framework where the effect of range and shape of input PDFs are incorporated. The

* Corresponding author.

E-mail address: mhk@eng.au.dk (M.H. Kristensen).

assignment of an individual PDF for each input parameter is an important and often difficult task; however, in most cases one can narrow down the range of variation and chose an appropriate PDF describing the variation [4]. The probabilistic framework of GSA requires that the model output be evaluated multiple times on randomly selected input samples from the entire input space. This requires a large number of Monte Carlo-based (MC) evaluations of the model. The accuracy of a MC analysis strongly depends on the sampling technique that must ensure good coverage of the input space [6]. Several different sampling techniques are available, but specific GSA methods often require a specific type of sampling. The prevailing sampling techniques includes simple random sampling (e.g. Monte Carlo sampling), quasi-random low-discrepancy sampling (e.g. Sobol' sequences [7,8]) and stratified sampling (e.g. Latin hypercube sampling (LHS) [9]). Quasi-random sampling using Sobol' sequences and LHS are known to outperform crude Monte Carlo random sampling [7,9]. Moreover, Helton et al. [10] argues that LHS is a good choice for computationally demanding models, because its efficient stratification properties allow a broad coverage of the entire probability distribution, at a relatively low computational cost.

A commonly used group of GSA methods for BEM-based analysis are the so-called screening-based methods. These methods seek to identify the least important input parameters that can be fixed at any given value without considerably reducing output variance. In this way, these methods are capable of ranking input variables by their importance in descending order using only a relatively small number of model evaluations [11]. Screening-based methods are efficient for computational heavy models, and/or models with a large number of input parameters, as e.g. building energy models [4]. They are, however, most reliable when the number of important input parameters is low, but as this is often the case, they are widely applicable [11]. For a detailed review on screening-based methods, see Saltelli et al. [12]. The screening-based method by Morris [13] has been widely used for BEM-based analysis, e.g. in an identification of important design parameters in sustainable buildings [4], in an uncertainty study of retrofits for residential buildings [14], in an investigation of important parameters for the performance of different active cooling systems [15], in an evaluation of how geometry effects building energy use [16], and for reducing the number of uncertain parameters in early building design [17,18].

Another group of GSA methods is the so-called variance-based methods. They rely on a decomposition of the model output variance; thus, they are known as ANOVA (analysis of variance) methods. They are regarded as model independent black-box methods making them suitable for complex non-linear and non-additive models. These advantages come, however, at the price of a high computational cost compared to the screening-based methods. A popular variance-based GSA method is the method of Sobol' [19], which, based on a decomposition of the output variance, is capable of estimating sensitivity indices describing the contribution of first-order effects for each input parameter alone, second-order interaction effects between two parameters, third-order effects and so on. The original method has since evolved as described in Borgonovo and Plischke [1] and Saltelli et al. [7]. Another popular variance-based method is the FAST (Fourier Amplitude Sensitivity Test) [20], and the extended FAST [21], which both uses a Fourier decomposition of the model output to estimate first-order and total-order effects. The main difference between Sobol' and FAST lies in the numerical estimation of the multidimensional integrals of the model necessary for the computation of the variances. The method by Sobol' applies Monte Carlo integration loops while FAST applies a sinusoidal function for pattern search. Variance-based methods are not used as widely for BEM-based analysis as screening-based methods; the few examples, to the knowledge of the authors, are the studies by Mechri et al. [22]

who used the FAST method to assess important parameters in office energy design, Shen and Tzempelikos [23] who applied a FAST sensitivity analysis of daylighting and energy performance of offices, Mara and Tarantola [6] who used the Sobol' method in an investigation of the thermal behaviour of an experimental test cell, and Spitz et al. [5] who used the Sobol' method in an experimental setup where measured parameter uncertainty was used as input data for the sensitivity analysis.

1.3. Information level in input parameters

The information level of input parameters to a BEM is important to the outcome of the sensitivity analysis; therefore, one has to specify an input parameter space of interest to be examined. The input parameter space for the SA can be defined by a range and a distribution of the likelihood for each value within this range. If the purpose is to explore the effect of equally possible parameter values, for example in a design situation, this space can be assumed uniformly distributed across a defined range meaning that the likelihood of different input parameter values are given equal weight. This uniform distribution is called a *non-informative distribution* as no information can be extracted from it besides the range of variation. One can also explore the output variability of a non-uniform distribution. A non-uniform distribution is called an *informative distribution* because it is based on a-priori information about the variation of the parameters, which could be available from e.g. expert judgements, historical data, or sample measurements. All SA methods are able to handle non-informative distributions, but not all SA methods are able to handle informative distributions.

1.4. Aim of this paper

The previous sections mention several examples where SA has been applied for BEM-based analysis. However, a sound argumentation that vouches for the reliability, validity and necessary complexity of the chosen SA method in the context it is being used, namely for BEM-based analysis, is rare.

The aim of this paper is to provide future BEM-based research with an argumentation for choosing an appropriate SA method, when using SA for identifying and ranking the most important model parameters, given a certain a-priori information level about the parameter input space. Three different but commonly used SA methods were therefore applied to a building test zone, representing an existing residential building stock, using two different BEM models and two different degrees of a-priori input parameter information (a uniform and a non-uniform distribution within the same range). The outputs from these analyses were then compared to identify whether the choice of SA method, BEM, and/or information level in input parameters, affected the identification and ranking of the input parameters most sensitive to the BEM output.

2. Methods

The methodology used to investigate the research statement is illustrated in Fig. 1. First, PDFs were assigned to each input parameter describing the a-priori beliefs of their shape and range of variation (see section 2.2 for details). For input scenario A, the information about shape of the individual input parameters were ignored by only considering the range of the PDFs using the 0.01 and 0.99 quantiles as boundaries. Hereby parameter variation is treated as uniformly distributed between the boundaries, which would correspond to the level of information available in the early building design process where the likelihood of the actual parameter value is uniformly distributed. For input scenario B, all information about the shape of the PDFs was taken into account, which could be the situation in the case of uncertainty analysis where the likelihood of

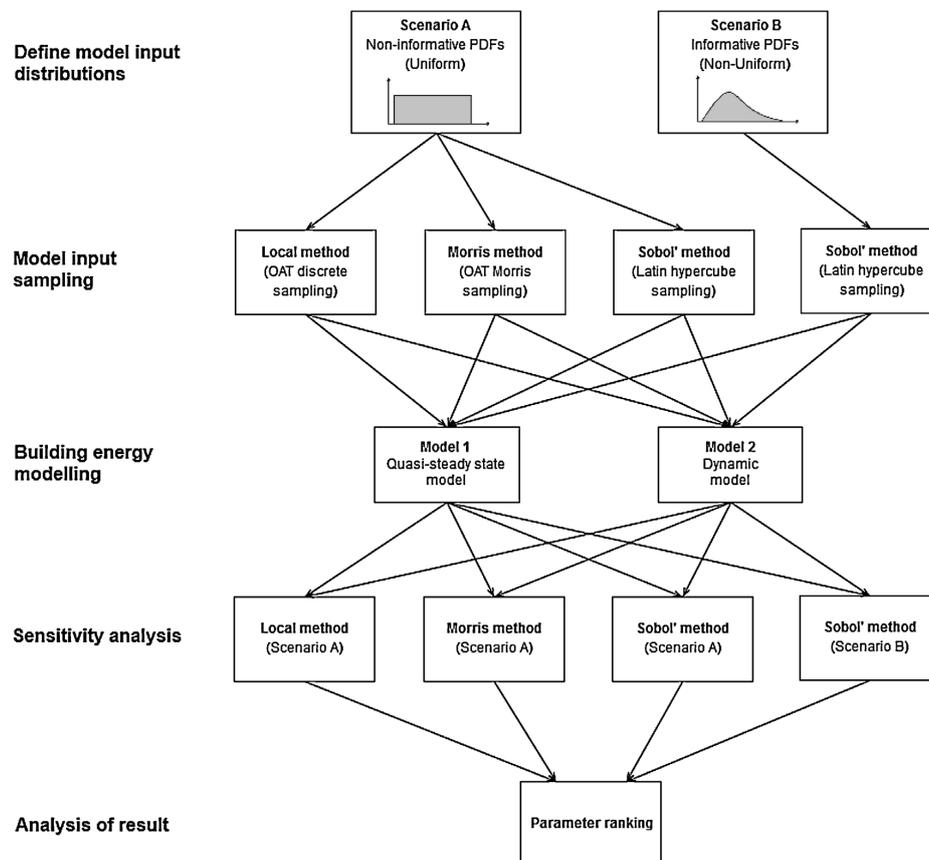


Fig. 1. Methodical approach.

the actual value of a parameter in some way is distributed around the most likely value (e.g. late in the planning of a new building or when working with existing buildings). Two different BEMs (see section 2.1 for details) were then run to calculate the annual energy need for space heating and cooling, using three different SA methods to sample from the input distributions (see section 2.3 for details). In total, four different sensitivity analyses outcomes for each of the two BEMs were generated as two of the SA methods are restricted to sample from input scenario A, while the third is capable of sampling from both scenario A and B, respectively, which has been utilised to compare the effect of input parameter distributions. Finally, the importance of the input parameters for each SA method in terms of describing model output uncertainty was quantified by ranking them in descending order (see section 3 for results).

2.1. Building energy models

The two different building energy models (BEMs) used in this study are fully prescribed in ISO 13790 (Energy performance of buildings - Calculation of energy use for space heating and cooling) [24] as the

- Monthly quasi-steady-state calculation method
- Simple hourly dynamic calculation method

Table 1
Time-in-use schedule for internal loads and ventilation.

Hours	Internal heat loads	Ventilation
23:00–07:00	25%	10%
07:00–16:00	50%	0%
16:00–23:00	100%	100%

In the monthly quasi-steady-state method, the energy balance is calculated over sufficiently long time to ignore heat stored and released; however dynamic effects are taken into account by introducing correlation factors in terms of empirical gain/loss utilisation factors. In the simple hourly dynamic method, the energy balance is calculated with short time steps taking into account the thermal inertia of the building mass by modelling the thermal resistances and capacitances, as well as the internal and solar heat gains, in an equivalent three node resistance capacitance network (5R1C model).

Both methods were applied as models to calculate the annual energy need for space heating and cooling (kWh/year) of a building zone test case where geometric and technical properties were treated as stochastic model input parameters using the PDFs given in Table 2. Time-in-use schedules for internal heat loads (occupancy and appliances) and ventilation are given in Table 1.

Weather conditions were modelled using the Danish design reference year (DRY) dataset, containing hourly values of the necessary weather parameters (air temperature, normal solar radiation

Table 2
Probability density functions assigned to model input parameters.

Input parameters	Unit	Building energy model		Quantiles					
		Quasi-steady-state	Simple dynamic	Distribution	1%	25%	median	75%	99%
Floor area	m ²	X	X	lnN[5.3;0.222]	120	173	200	233	333
Room height	m	X	X	lnN[1.1;0.062]	2.6	2.9	3.0	3.1	3.5
U-value (walls)	W/(m ² K)	X	X	lnN[-1.4;0.942]	0.03	0.14	0.26	0.48	2.28
U-value (roof)	W/(m ² K)	X	X	lnN[-2.1;1.332]	0.01	0.05	0.12	0.30	2.77
U-value (floor)	W/(m ² K)	X	X	lnN[-0.9;0.692]	0.08	0.25	0.40	0.63	1.96
Adjustment factor (ground)	-	X	X	Beta[30;3.33]	0.75	0.87	0.91	0.94	0.98
Window-%, (N,E,S,W) ^a	-	X	X	Beta[5;20]	0.06	0.14	0.19	0.25	0.41
Window frame fraction	-	X	X	Beta[15;60]	0.11	0.17	0.20	0.23	0.32
U-value (window)	W/(m ² K)	X	X	lnN[0.7;0.222]	1.16	1.68	1.95	2.27	3.29
SHGC	-	X	X	Beta[20;13.33]	0.40	0.54	0.60	0.66	0.78
Internal solar shading (N,E,S,W)	-	X	X	Beta[30;3.33]	0.75	0.87	0.91	0.94	0.98
Ventilation rate	l/(s m ²)	X	X	lnN[-0.03;0.222]	0.58	0.84	0.98	1.13	1.64
Infiltration rate	l/(s m ²)	X	X	lnN[-0.43;0.802]	0.10	0.38	0.65	1.12	4.18
Heat recovery	-	X	X	Beta[50;16.67]	0.62	0.72	0.75	0.79	0.86
Internal heat loads (occupancy + appliances)	W/m ²	X	X	lnN[0.87;0.312]	1.2	1.9	2.4	2.9	4.9
Thermal mass	kJ/(m ² K)	X	X	lnN[11.85;0.372]	60	110	141	180	330
Area of exposed thermal mass	m ² /m ²	-	X	lnN[1.09;0.112]	2.3	2.8	3.0	3.2	3.8
Reference numerical parameter ^b	-	X	-	lnN[-0.01;0.152]	0.7	0.9	1.0	1.1	1.4
Reference time constant ^c	[h]	X	-	lnN[2.64;0.372]	6	11	14	18	33
Heating set point	[°C]	X	X	N[22;12]	19.7	21.3	22.0	22.7	24.3
Cooling set point	[°C]	X	X	N[27;12]	24.7	26.3	27.0	27.7	29.3

^a Window-% was changed independently for each of the four cardinal directions (N,E,S,W), using the same parameter distribution.

^b Reference numerical parameter for heating ($a_{H,0}$) and cooling ($a_{C,0}$). In both cases, ISO 13790 gives an empirical value of 1.0.

^c Reference time constant for heating ($\tau_{H,0}$) and cooling ($\tau_{C,0}$). In both cases, ISO 13790 gives an empirical value of 15 h.

and diffuse solar radiation) [25]. Total solar radiation perpendicular to the building facades was calculated in both BEM models using the solar algorithm described in [26].

2.2. Model input parameters

A PDF was assigned to each of the 24 building geometrical and technical input parameters defining the test zone (Table 2) to reflect the uncertainty of an existing residential building stock; as such, these PDFs serves as an example of the a-priori uncertainty that could be contained in such parameters. Three different distributions were applied: the normal distribution $N(\mu, \sigma^2)$, the log-normal distribution $\ln N(\mu, \sigma^2)$ and the beta distribution $Beta(\alpha, \beta)$. The normal distribution was used to represent the variation of the continuously defined parameters, e.g. temperature set points, whereas the log-normal distribution was used to represent those continuously defined parameters that only exist in the positive domain, e.g. U-values. The beta distribution was assigned to factors only defined in the range $\{0 - 1\}$, e.g. the solar heat gain coefficient (SHGC). Correlations between parameters, e.g. window U-values and SHGC, has not been taken into account.

2.3. Sensitivity analysis methods

Three SA methods were investigated: a local partial derivative-based method (Appendix A), the OAT global screening-based method of Morris (Appendix B) and the global variance-based method of Sobol' (Appendix C). These three SA methods apply very different schemes for assessing the parameter influence on model output variability. Their capabilities are summarised in Table 3 based on four properties: (1) ability to take into account the scale and shape of input parameter distributions, (2) ability to take into account multi-dimensional parameter influence where all parameters are varied simultaneously, (3) being model independent, i.e. embracing both non-linear and non-additive effects where parameter interactions are taken into account, and (4) the relative computational time [11].

Table 3
Properties of sensitivity analyses methods.

Property	Sensitivity analysis method		
	Local	Morris	Sobol'
Scale and shape of input parameter PDF	No	No	Yes
Multi-dimensional averaging	No	No/Yes	Yes
Model independent	No	Yes	Yes
Relative computational time	Low	Medium	High

The only SA method to encompass all three properties is the global variance-based method of Sobol'. The Sobol' method utilises a comprehensive representation of the model behaviour through a complete decomposition of the output variance by searching across the entire input space while simultaneously taking into account range and shape of parameter distributions and correlated effects. The total-order sensitivity indices (C.4) are bounded to sum to one, making them physically meaningful measures of parameter influence. A numerical error is introduced though, as the Monte Carlo framework only approximates the sensitivity indices. The convergence of the numerical approximations can be assessed as the variation of the total-order indices and the internal ranking of the parameters.

A similar direct interpretation can be made with the local sensitivity index; however, where the Sobol' indices sum to one for each parameter, the local method do not feature this boundedness. Instead, the local method applies a single SI-measure (A.1) using the extreme parameter values to calculate the point elasticity around the mean. This would be sufficient for identifying the sensitivity of a linear model, but it might be insufficient for non-linear models. Another limitation of the local method is that it does not allow any utilisation of knowledge about the shape of the parameter distributions, meaning that all values within the input range are treated as equally possible (uniformly distributed); moreover, as the local method is an OAT method, it does not allow any quantification of correlated parameter effects.

The Morris method applies the absolute mean of a population of elementary effects (B.3) to quantify the influence of a given parameter. Even though this methodology is regarded partly global by

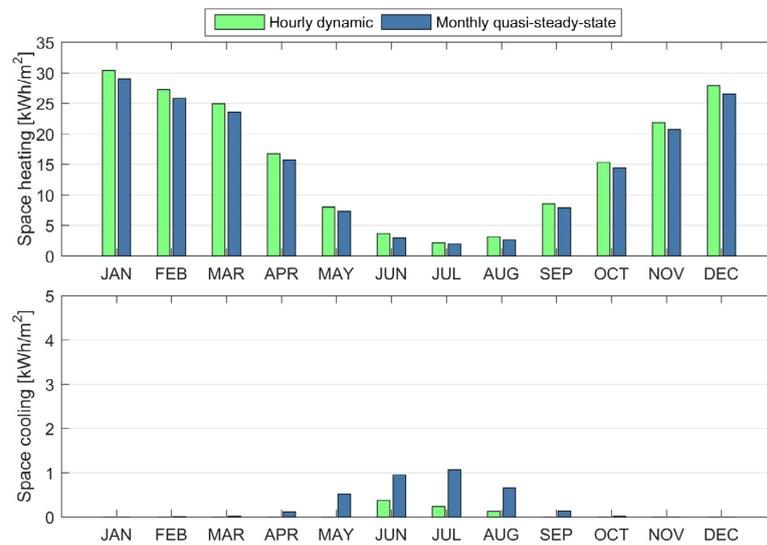


Fig. 2. Energy need for space heating and cooling calculated with the hourly dynamic model and the monthly quasi-steady-state model, respectively, using the median value of input distributions in Table 2.

averaging across multiple local effects at different points in the input space, each parameter is still varied one-at-a-time (OAT), hereby potentially neglecting correlated effects, hence the ‘No/Yes’ in Table 3. Another disadvantage of the Morris method is the dubious interpretation of the mean elementary effect as a measure of sensitivity. Just like the local derivative-based method, the Morris method is also based on derivatives thus lacking the boundedness of the Sobol method. One should be careful with interpreting a large absolute value of the mean elementary effect as a sign of great parameter influence as such values will vary from one model to another. Only the internal ranking of the means can be used to quantify the influence of the parameters and sort them in clusters of importance. The complexity of using the different SA methods

in BEM practice are obvious: the Local method requires less effort and time than the Morris method, which requires less effort and time than the method of Sobol’. From a practical point of view, it is therefore interesting to investigate whether the differences in SA capabilities has any influence on the ranking of input parameters in terms of their impact on the BEM output.

3. Results

3.1. Comparison of energy models

The calculated thermal energy need for space heating and cooling, using the two ISO 13790 energy models, is shown in Fig. 2

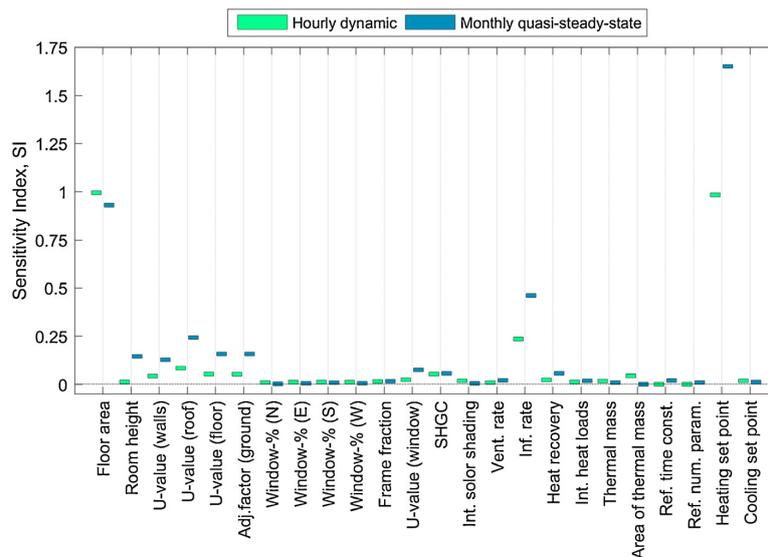


Fig. 3. Local method sensitivity measure (sensitivity indices).

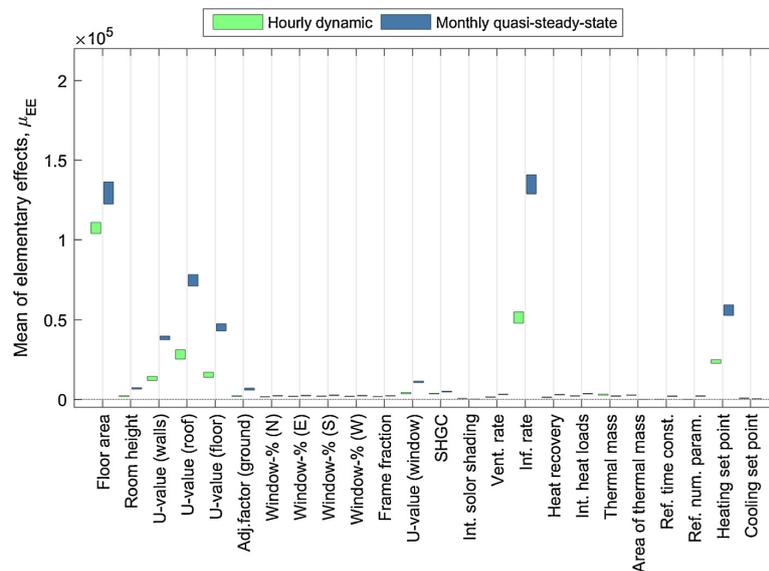


Fig. 4. Morris method sensitivity measure (mean of elementary effects). The boxes indicate the 95% confidence bounds.

using the reference value from all input distributions (median value in Table 2). In general, the quasi-steady-state model predicts a lower monthly energy need for space heating and higher monthly energy need for cooling compared to the simple hourly dynamic model. The monthly differences even out in the total annual energy need to 191 kWh/(m² year) and 182 kWh/(m² year) for the hourly dynamic model and monthly quasi-steady-state model, respectively, yielding an annual relative deviation of approximately 4.4%.

3.2. Sensitivity indices

The sensitivity indices of the output (annual energy need for space heating and cooling) of each of the two BEMs, towards changes in the model input parameters when applying the local SA method (Appendix A), are shown in Fig. 3; the mean values of the elementary effects from the Morris method (Appendix B) are shown in Fig. 4; the total-order effects from the Sobol' method (Appendix C) are shown in Figs. 5a and 5b for the non-informative and the informative cases, respectively. Ranking of the top-ten parameters for each analysis is given in Table 4.

3.2.1. Effect of parameter information level

The effect of changing input parameter information level can be analysed by comparing the Sobol' method ranking results from the use of non-informative and informative PDFs, respectively (Table 4). According to the Sobol' total-order sensitivity indices with non-informative parameter distributions (Fig. 5a), approximately 87% to 92% of the output variability in the two BEMs, respectively, can be ascribed to the top five ranked parameters; floor area, infiltration rate, U-value of the roof, heating set point, and U-value of the floor. These five parameters are therefore considered the most influential parameters when using uniformly distributed inputs in both BEMs. Applying informative distributions (Fig. 5b), the ranking of the top five parameters is slightly different depending on the applied model. For both BEMs, application of informative distributions results in the heating set point parameter being replaced with the U-value of the walls in the top five ranking,

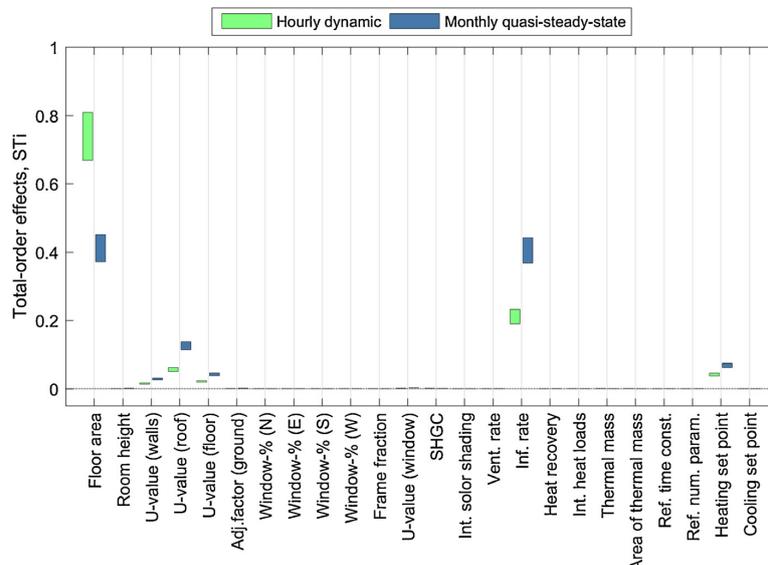
and the amount of variability accounted for in top five parameters makes up approximately 97% for BEMs.

3.2.2. Effect of sensitivity analysis method

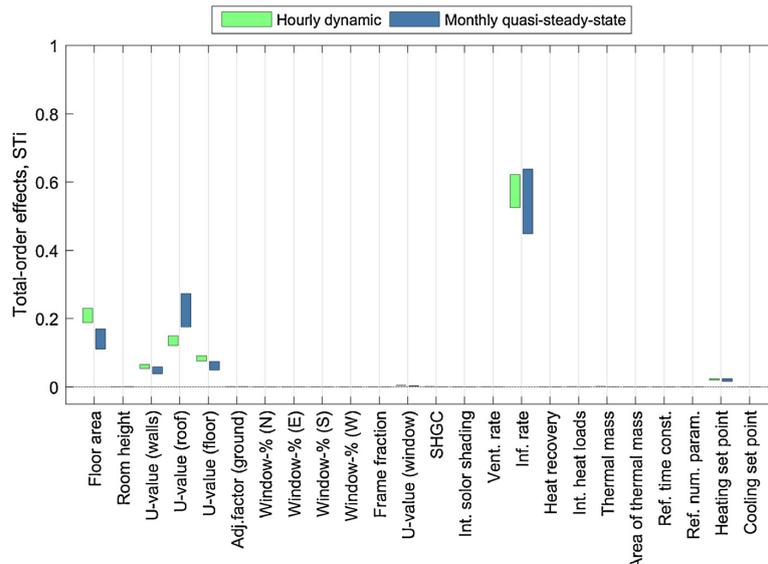
To make a comparison of the performance of the different SA methods, the ranking of the top five and top ten most influential parameters identified with the Local method and the Morris method is compared to the ranking from the Sobol' method using non-informative PDFs (Sobol' (A)) and informative PDFs (Sobol' (B)), respectively (Table 4).

The local SA method applied to the hourly dynamic BEM as well as the monthly quasi-steady-state BEM identified four out of five most influential parameters (heating set point, floor area, U-value of the roof, and infiltration rate) when compared to Sobol' (A), but ranks them in a different order. When compared to Sobol' (B), the Local method identified three out of the five most influential parameters (floor area, U-value of the roof, and infiltration rate) when compared to Sobol' (B) for both BEMs, but ranks them in a different order. The top five parameters identified with the Local method makes up approximately 86% and 89% of the sensitivity for Sobol' (B) in the hourly dynamic and the monthly quasi-steady-state BEM, respectively. The Local SA method identified nine out of the top ten most influential parameters for both the hourly simple dynamic and monthly quasi-steady-state model when compared to Sobol' (A) and (B), but the ranking was different in all cases.

The global screening-based method of Morris identified the same ten most influential parameters as in Sobol' (A) for both BEMs and ranks them in the exact same way with the exception of the top two parameters which changed places in the monthly quasi-steady-state BEM. When compared to Sobol' (B), the Morris method identifies four out of the five most influential parameters when applied to both BEMs (floor area, U-value of the roof, U-value of the floor, and infiltration rate), but with a slightly different ranking in both cases. The top five parameters identified with the Morris method makes up approximately 94% and 95% of the sensitivity for Sobol' (B) in the hourly dynamic and the monthly quasi-steady-state BEM, respectively. The Morris method identified nine out of the top ten most influential parameters in the hourly simple



(a) Non-informative PDFs (uniform distribution).



(b) Informative PDFs (non-uniform distribution).

Fig. 5. Sobol' method sensitivity measure (total-order sensitivity index). The boxes indicate the 95% confidence bounds. Variation within the ranking stopped after approximately 6000 iterations for both scenarios. For (a): the sum of total-order indices is 1.05 and 1.06 for the hourly simple dynamic and monthly quasi-steady-state BEM, respectively. For (b): the sum of total-order indices is 1.02 and 1.03 for the hourly simple dynamic and monthly quasi-steady-state BEM, respectively.

dynamic BEM model when compared to Sobol' (B), but with a quite different ranking, and ten out of ten for the monthly quasi-steady-state BEM, but with a somewhat different ranking.

It is noted that it is only for the hourly dynamic model where a parameter regarding thermal mass, 'area of exposed thermal mass', comes into top ten (except for Sobol' (B)).

4. Discussion

When compared to the Sobol' method using non-informative uniform PDFs, it has been demonstrated that the Local method

and the Morris method may be applicable as methods for identifying the most important cluster of parameters. However, it was only the Morris method that was able to rank the parameters in the same order as the Sobol' method. When applying informative PDFs, the performance of both the Local method and the Morris method decreased when compared to Sobol' method as none of the two methods are capable of addressing information about parameter shape. However, it is possible to include this information as proposed by Campolongo and Braddock [27] who expanded the Morris method to consider also higher order effects (e.g. interactions between any pair of input parameters). Such a second-order

Table 4
Ranking of top ten most influencing input parameters on the two BEMs for each sensitivity method.

Input	Hourly dynamic model				Monthly quasi-steady-state model			
	Local	Morris	Sobol' (A) ^a	Sobol' (B) ^b	Local	Morris	Sobol' (A) ^a	Sobol' (B) ^b
Floor area	1	1	1	2	2	2	1	3
Room height	–	–	–	–	7	9	9	9
U-value (walls)	9	6	6	5	8	6	6	5
U-value (roof)	4	3	3	3	4	3	3	2
U-value (floor)	6	5	5	4	6	5	5	4
Adjustment factor (ground)	7	–	–	10	5	8	8	8
Window-%, (N,E,S,W)	–	–	–	–	–	–	–	–
Window frame fraction	–	–	–	–	–	–	–	–
U-value (window)	10	7	7	7	9	7	7	7
SHGC	5	8	8	9	–	10	10	10
Internal solar shading (N,E,S,W)	–	–	–	–	–	–	–	–
Ventilation rate	–	–	–	–	–	–	–	–
Infiltration rate	3	2	2	1	3	1	2	1
Heat recovery	–	–	–	–	10	–	–	–
Internal heat loads	–	–	–	–	–	–	–	–
Thermal mass	–	9	9	8	–	–	–	–
Area of exposed thermal mass	8	10	10	–	–	–	–	–
Reference numerical parameter	–	–	–	–	–	–	–	–
Reference time constant	–	–	–	–	–	–	–	–
Heating set point	2	4	4	6	1	4	4	6
Cooling set point	–	–	–	–	–	–	–	–

^a Sobol' (A): non-informative PDF (uniform distribution).

^b Sobol' (B): informative PDF (non-uniform distribution).

design has been compared to the original first-order design by Morris in a sensitivity study of an apartment building [28] where it was found that the implementation of second-order elementary effects could be helpful to sort variables and to specify their interaction in pairs. Various versions and extensions of the Local method do also exist making it possible to investigate a larger domain of the input space [29].

Only the shape of the PDFs was assessed in the investigation of the effect of a-priori parameter information. The outcome of SA could also be affected by the range of the PDFs if there are nonlinear interactions in the BEM. However, the range will have no influence for the outcome if the BEM is strictly linear. The behaviour of the BEMs investigated in this paper are predominantly linear and additive of nature, and different PDF ranges is therefore most likely not influencing the identification and ranking of most important parameters in any of the SA methods.

The two ISO 13790 BEMs used in the analysis presented in this paper are both based on simplified representation of the building physics. There are several BEM tools available, which rely on more accurate models of the building physics. The application of such a tool could have influenced the results of this study; however, other studies have demonstrated that the ISO 13790 models are in good compliance with results from more sophisticated BEM tools like EnergyPlus [30] and BSim [31]. Furthermore, Kim et al. [30] found that the ten out of top ten most influential input parameters on energy need for space heating, and nine out of ten concerning space cooling, were the same when applying the original Morris method to the monthly quasi-steady-state BEM of ISO 13790 and the sophisticated BEM tool EnergyPlus; however, the ranking of the parameters were slightly different in both cases. This could imply that the overall model behaviour of EnergyPlus has a similar effect on parameter identification and ranking as the monthly quasi-steady-state model when modelling the energy need for space heating and cooling. However, this may not hold for more dynamic and detailed systems, e.g. advanced facade systems. In a future study it would be interesting to apply the Sobol' method to a more complex and sophisticated BEM to investigate whether it identifies and ranks input parameters differently due to any non-linear and non-additive behaviour in the sophisticated BEM model that the Morris method cannot detect.

5. Conclusions

The purpose of the study presented in this paper was to elucidate how the choice of sensitivity analysis technique, and the level of a-priori parameter information one applies in the analysis, affects the identification and ranking of input parameters to which the output from a given BEM was most sensitive. For this purpose, three different SA methods of increasing complexity and capability (Local, Morris, and Sobol') have been applied to the annual energy need output from an hourly simple dynamic and a monthly quasi-steady-state building energy model.

Using the Sobol' method it has been demonstrated that the level of parameter input information, i.e. the shape of the parameter distribution in a certain range, has an effect on the result of the sensitivity analysis outcome. The effect of distribution shape is likely to be both model-specific as well as affected by any potential parameter correlations. The use of the Sobol' method is therefore recommended for cases where a continuous or discrete range and shape of the parameter distributions is known, as the Sobol' method was the only of the three investigated methods that was able to account for the aggregated effect of variability in specific parameter distributions. However, the use of the less computational demanding Morris method (no. model evaluations: Local = 49; Morris = 6,250; Sobol' = 260,000) is recommended when the variation of input parameters are uniformly distributed between chosen boundaries, e.g. when designing a new building, as it seems to lead to the same identification and ranking as the Sobol' method.

The study also demonstrated that the Local method was capable of identifying practically the same cluster of most important parameters as the more sophisticated Morris and Sobol' method, but it was not capable of identifying the same ranking of parameters in terms of importance to the model output. Consequently, the use of the simple Local method is only recommended for situations where identification of a cluster of most sensible input parameters is of interest and not the actual ranking.

The study furthermore demonstrates that the choice of BEM has an effect of the outcome of all three applied SA methods mainly in terms of ranking. Overall, the Sobol' method seems to be slightly more robust towards the choice of BEM than the Local and the Morris method.

The above conclusions are drawn based on simulations of an existing residential building stock in a temperate climate, which has a significant heating and a limited cooling need. Conclusions might be different for another type of building and/or climate.

Acknowledgements

The research was conducted as part of the 'Resource Efficient Cities Implementing Advanced Smart City Solutions' (READY) project, work package 3, financed by the 7th EU Framework Programme (FP7-Energy project reference: 609127).

Appendix A. Local method

The simplest SA method included in the investigation was a local method based on partial derivatives [29]. In the Local method the model input space, Ω , is made up of k number of input parameters x_i , for which a dimensionless sensitivity index, Sl_i , is calculated OAT based on the partial derivative of the model output, y :

$$Sl_i = \frac{\% \Delta y}{\% \Delta x_i} = \left| \frac{\frac{y^{(2)} - y^{(1)}}{\bar{y}}}{\frac{x_i^{(2)} - x_i^{(1)}}{\bar{x}_i}} \right| \quad (\text{A.1})$$

The sensitivity index, Sl_i , expresses the output elasticity of variation around the mean output value as percentage change in output, $\% \Delta y$, per percentage change in input, $\% \Delta x_i$. The superscripts 1 and 2 refer to values of input, x_i , and corresponding output, y , respectively, while the overlined parameters, \bar{x}_i and \bar{y} , refer to the mean value of those parameters. In this study, the inputs $x_i^{(1)}$ and $x_i^{(2)}$ were sampled from the input distributions as the 0.01 and 0.99 quantiles transforming the sensitivity index to an expression of the midpoint arc elasticity. All input assumptions are summarised in Table A.5.

Appendix B. Morris method

The global OAT screening-based method by Morris [13] that, as opposed to the local method, is capable of addressing non-linear and interaction effects, was included in the investigation as a simple global method. In the Morris method, the input space, Ω , is made up of k number of input parameters, x_i , each defined within a discrete range from a specified minimum, x_i^- , to a maximum, x_i^+ , normalised to a scale $\{0 - 1\}$. The range of each parameter is subdivided in p number of points, denoted *levels*, with a distance Δ between them. Hereby x_i assumes integer values in the range:

$$x_i = \left\{ 0, \frac{1}{p-1}, \frac{2}{p-1}, \dots, 1 \right\} \quad (\text{B.1})$$

The input space, Ω , thus becomes a k -dimensional p -level grid with p^k points. The Morris method is usually applied with uniform input distributions which means that the levels, p , are obtained by dividing the interval in which each factor varies (from x_i^- to x_i^+) into equally large parts. It is however possible to let input parameters follow non-uniform distributions by selecting the levels e.g. as the quantiles of the distribution [4,11]. In this study, we have used the 0.01 and 0.99 quantiles from the input distributions as the delimiting range of variation in the discrete input range and subdivided it in $p=5$ levels (Table A.5). This way, an equal distance between each level was maintained.

From the input space, each parameter is assigned a random base value in the discretized grid forming the base input vector, \mathbf{X}^* , corresponding to a point in Ω . From here, the model is evaluated at different values of \mathbf{X} , which is obtained by moving randomly around in Ω , increasing or decreasing x_i by Δ while keeping all other parameters constant. This way a trajectory is 'followed' through Ω with the index i being selected at random from $\{1, 2, \dots, k\}$. Each

realisation of the input vector \mathbf{X} should differ from previous realisations. The algorithm continues until all parameters have been evaluated at two different values creating $k+1$ model evaluations in the trajectory using input vector $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(k+1)}$.

Morris suggested evaluating the influence of the individual input parameters using a so-called elementary effect (EE) for each parameter. The elementary effect, which can be seen as a global extension of a local derivative-based method, is defined as:

$$EE_i = \frac{y(x_1, \dots, x_{i-1}, x_i + \Delta, \dots, x_k) - y(x_1, \dots, x_k)}{\Delta} \quad (\text{B.2})$$

By following r different trajectories (indexed t), providing r estimates of the elementary effects for each input parameter, the individual parameter influence is evaluated by the mean value, μ_i , and standard deviation, σ_i , of the population of elementary effects for each parameter:

$$\mu_i = \frac{1}{r} \sum_{t=1}^r |EE_{i,t}| \quad (\text{B.3})$$

$$\sigma_i = \sqrt{\frac{1}{r-1} \sum_{t=1}^r |EE_{i,t} - \mu_i|^2} \quad (\text{B.4})$$

The mean value, μ_i , is an estimation of the influence of the i -th input parameter, x_i , while the standard deviation, σ_i , is a measure of the interactions with other parameters and any non-linear effects that the parameter takes part in. The mean value is used for ranking the parameters in order of importance.

The expressions in ((B.3)-(B.4)) are revisions by Saltelli et al. [11] of the original work by Morris, as Morris did not make use of the absolute values, which can lead to underestimations of the parameter influence as values of opposite sign may cancel out each other. Using absolute values as in ((B.3)-(B.4)) is sometimes referred to as the Revised Morris method.

At a combined computation cost of $r \cdot (k+1)$ simulations, the Morris method is practically inexpensive compared to the local method (Table A.5). In Heiselberg et al. [4] it is advised to use a minimum sampling number $r=4$ to reasonably cover the parameters and $r=10$ to obtain very reliable results. In this study, however, the elementary effects were approximated using $r=250$ random trajectories through the input space resulting in 6,250 model evaluations in order to obtain stable mean values of the elementary effects, μ_i , without changing the ranking of the parameters. 95% confidence bounds of μ_i were derived using 500 bootstrapping samples. All input assumptions are summarised in Table A.5.

Appendix C. Sobol' method

The global variance-based method by Sobol' [19] was the most complex SA method included in the investigation. The governing idea of this method is that the total unconditional variance of the model output, $V(Y)$, can be written as a sum of terms describing the variance contribution of first-order effects for each input parameter alone, V_i , second-order interaction effects between two parameters, V_{ij} , third-order effects, and so on [11]:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + V_{1,2,\dots,k} \quad (\text{C.1})$$

The total model variance is thus attributed to individual input parameters and interactions between parameters. The expression in (C.1) can be written as:

$$V(Y) = V_{X_i} (E_{\mathbf{X}_{-i}}(Y | X_i)) + E_{X_i} (V_{\mathbf{X}_{-i}}(Y | X_i)) \quad (\text{C.2})$$

The total unconditional model variance, $V(Y)$, is divided into variance explained by fixation of input parameter X_i (first term

Table A.5
Input assumptions for the different sensitivity analysis methods.

	Sensitivity analysis method		
	Local	Morris	Sobol'
Points used from PDF	0.01 ($x_i^{(1)}$) and 0.99 ($x_i^{(2)}$)	0.01 (x_i^-) and 0.99 (x_i^+), $p = 5$ (levels)	Entire PDF
Sample size	1	$r = 250$ (trajectories)	$N = 10,000$ (LHS)
No. of model evaluations	$2k + 1 = 49$	$r(k + 1) = 6,250$	$N(k + 2) = 260,000$
Convergence measure	N/A	$\mu_i \approx \text{constant}^a$	$ST_i \approx 1$

^a Parameter ranking does not change with increasing sample size.

in (C.2)) and residual variance (last term in (C.2)). The first term, $V_{X_i}(E_{\mathbf{X}_{-i}}(Y | X_i))$, express model variance conditional on X_i and thus represents the first-order (e.g. additive) effect of X_i , which is a measure of sensitivity varying between zero and $V(Y)$. The inner expectation operator $E_{\mathbf{X}_{-i}}$ is the mean of Y taken over all possible values of the input matrix \mathbf{X} while keeping X_i fixed. The outer variance V_{X_i} is taken over all possible values of X_i . By normalising with the total unconditional variance $V(Y)$, the following first-order sensitivity index is established for the i -th input parameter [7]:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{-i}}(Y | X_i))}{V(Y)} \quad (C.3)$$

Likewise, it is possible to establish expressions from (C.1) for the higher order effects, where input parameters interact with each other. However, it is often enough to account for the combined interaction effects of each input parameter by assessing the so-called total-order effect sensitivity index [7]:

$$ST_i = 1 - \frac{V_{\mathbf{X}_{-i}}(E_{X_i}(Y | \mathbf{X}_{-i}))}{V(Y)} \quad (C.4)$$

where $V_{\mathbf{X}_{-i}}(E_{X_i}(Y | \mathbf{X}_{-i}))$ is the expected variance reduction that would be obtained if all parameters but X_i could be fixed, i.e. the model variance is attributable to the combined variation of all other input parameters than X_i . Subtracting this value from 1 leaves only the total variance contribution from X_i . The sum of the total-order effects of all input parameters equals one.

In this study, the Sobol' sensitivity indices S_i and ST_i have been estimated according to the methods formulated by Sobol' [19] and Homma and Saltelli [32] as summarised in Saltelli et al. [11]. A Monte Carlo (MC) framework was used for the computational scheme in which the estimators for S_i and ST_i were approximated in the same set of simulations. From the probability distributions of each input parameter, two independent global input matrices \mathbf{A} (sample matrix) and \mathbf{B} (re-sample matrix) were generated, each containing N samples of the k input distributions. For this, we applied LHS using $N = 10,000$.

By evaluating the sample matrix, \mathbf{A} , a distribution of N model outputs f was obtained with a total unconditional expectation, f_0 , and variance, $V(Y)$:

$$f_0 = E(Y) = \frac{1}{N} \sum_{j=1}^N f(\mathbf{A}_j) \quad (C.5)$$

$$V(Y) = \frac{1}{N} \sum_{j=1}^N f^2(\mathbf{A}_j) - f_0^2 \quad (C.6)$$

Two new matrices $\mathbf{A}_B^{(i)}$ and $\mathbf{B}_A^{(i)}$ were introduced where in $\mathbf{A}_B^{(i)}$ all columns were from the sample matrix, \mathbf{A} , except the i -th, which was from the re-sample matrix, \mathbf{B} . This was opposite in $\mathbf{B}_A^{(i)}$.

The conditional variances necessary for the computation of the sensitivity indices were then estimated in the same MC loop:

$$V_{X_i}(E_{\mathbf{X}_{-i}}(Y | X_i)) = \frac{1}{N} \sum_{j=1}^N f(\mathbf{A}_j) f(\mathbf{B}_A^{(i)})_j - f_0^2 \quad (C.7)$$

$$V_{\mathbf{X}_{-i}}(E_{X_i}(Y | \mathbf{X}_{-i})) = \frac{1}{N} \sum_{j=1}^N f(\mathbf{A}_j) f(\mathbf{A}_B^{(i)})_j - f_0^2 \quad (C.8)$$

This scheme allows the computation of the first-order and total-order indices using a total of $N \cdot (k + 2)$ model evaluations with N being the number of samples used in the MC framework and k being the number of input parameters. In this study, the total-order effects were approximated using $N = 10,000$ Latin hypercube samples at which point convergence was ensured, resulting in 350,000 model evaluations (Table A.5). 95% confidence bounds of the total-order effects were derived using 500 bootstrapping samples.

References

- [1] E. Borgonovo, E. Plischke, Sensitivity analysis: a review of recent advances, Eur. J. Oper. Res. 248 (2016) 869–887, <http://dx.doi.org/10.1016/j.ejor.2015.06.032>.
- [2] S. Petersen, S. Svendsen, Method and simulation program informed decisions in the early stages of building design, Energy Build. 42 (2010) 1113–1119, <http://dx.doi.org/10.1016/j.enbuild.2010.02.002>.
- [3] Y. Sun, Sensitivity analysis of macro-parameters in the system design of net zero energy building, Energy Build. 86 (2015) 464–477, <http://dx.doi.org/10.1016/j.enbuild.2014.10.031>.
- [4] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinare, S. Thomas, Application of sensitivity analysis in design of sustainable buildings, Renew. Energy 34 (2009) 2030–2036, <http://dx.doi.org/10.1016/j.renene.2009.02.016>.
- [5] C. Spitz, L. Mora, E. Wurtz, A. Jay, Practical application of uncertainty analysis and sensitivity analysis on an experimental house, Energy Build. 55 (2012) 459–470, <http://dx.doi.org/10.1016/j.enbuild.2012.08.013>.
- [6] T. Mara, S. Tarantola, Application of global sensitivity analysis of model output to building thermal simulations, Build. Simul. 1 (4) (2008) 290–302, <http://dx.doi.org/10.1007/s12273-008-8129-5>.
- [7] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output design and estimator for the total sensitivity index, Comput. Phys. Commun. 181 (2010) 259–270, <http://dx.doi.org/10.1016/j.cpc.2009.09.018>.
- [8] I.M. Sobol', S.S. Kucherenko, On global sensitivity analysis of quasi-Monte Carlo algorithms, Monte Carlo Methods Appl. 11 (1) (2005) 83–92, <http://dx.doi.org/10.1515/1569396054027274>.
- [9] C.J. Sallaberry, J.C. Helton, S.C. Hora, Extension of Latin hypercube samples with correlated variables, Reliab. Eng. Syst. Saf. 93 (2008) 1047–1059, <http://dx.doi.org/10.1016/j.res.2007.04.005>.
- [10] J.C. Helton, C.J. Sallaberry, C.B. Storlie, Survey of sampling-based methods for uncertainty and sensitivity analysis, Reliab. Eng. Syst. Saf. 91 (2006) 1075–1209, <http://dx.doi.org/10.1016/j.res.2005.11.017>.
- [11] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, John Wiley and Sons, Ltd., The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2004, ISBN 0-470-87093-1.
- [12] A. Saltelli, S. Tarantola, F. Campolongo, Sensitivity analysis as an ingredient of modeling, Stat. Sci. 15 (4) (2000) 377–395 <http://www.jstor.org/stable/2676831>.
- [13] M.D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (2) (1991) 161–174, <http://dx.doi.org/10.2307/1269043>.
- [14] A.T. Booth, R. Choudhary, D.J. Spiegelhalter, Handling uncertainty in housing stock models, Build. Environ. 48 (2012) 35–47, <http://dx.doi.org/10.1016/j.buildenv.2011.08.016>.

- [15] J. Le Dréau, P. Heiselberg, Sensitivity analysis of the thermal performance of radiant and convective terminals for cooling buildings, *Energy Build.* 82 (2014) 482–491, <http://dx.doi.org/10.1016/j.enbuild.2014.07.002>.
- [16] T.L. Hemsath, K.A. Bandhosseini, Sensitivity analysis evaluating basic building geometry's effect on energy use, *Renew. Energy* 76 (2015) 526–538, <http://dx.doi.org/10.1016/j.renene.2014.11.044>.
- [17] T. Østergård, S.E. Maagaard, R.L. Jensen, A stochastic and holistic method to support decision-making in early building design, in: *Proceedings of BS 2015, the 14th International Conference of the International Building Performance Simulation Association*, Hyderabad, India, 2015.
- [18] T. Østergård, S.E. Maagaard, R.L. Jensen, Thermal comfort in residential buildings by the millions – early design support from stochastic simulations, in: *Proceedings of CLIMA 2016, the 12th REHVA World Congress*, vol. 6, Aalborg, Denmark, 2016.
- [19] I.M. Sobol', Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comput. Exp.* 1 (4) (1993) 407–414.
- [20] R.I. Cukier, C.M. Fortuin, K.E. Shuler, A.G. Petschek, J.H. Schaibly, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory, *J. Chem. Phys.* 59 (8) (1973) 38–73, <http://dx.doi.org/10.1063/1.1680571>.
- [21] A. Saltelli, S. Tarantola, K.P.-S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics* 41 (1) (1999) 39–56, <http://dx.doi.org/10.2307/1270993>.
- [22] H.E. Mechri, A. Capozzoli, V. Corrado, Use of the ANOVA approach for sensitive building energy design, *Appl. Energy* 87 (2010) 3073–3083, <http://dx.doi.org/10.1016/j.apenergy.2010.04.001>.
- [23] H. Shen, A. Tzempelikos, Sensitivity analysis on daylighting and energy performance of perimeter offices with automated shading, *Build. Environ.* 59 (2013) 303–314, <http://dx.doi.org/10.1016/j.buildenv.2012.08.028>.
- [24] ISO 13790:2008, *Energy Performance of Buildings – Calculation of Energy Use for Space Heating and Cooling*, 2nd ed., 2008.
- [25] J.M. Jensen, H. Lund, *Design Reference Year, DRY – Et nyt dansk referencår*, Technical University of Denmark, 1995 October (in Danish) announcement number 281.
- [26] B. Bourges, *Climatic Data Handbook for Europe: Climatic Data for the Design of Solar Energy Systems*, 1st ed., Springer Netherlands, 1992, ISBN 978-0-7923-1716-6.
- [27] F. Campolongo, R. Braddock, The use of graph theory in the sensitivity analysis of the model output: a second order screening method, *Reliab. Eng. Syst. Saf.* 64 (1) (1996) 1–12, [http://dx.doi.org/10.1016/S0951-8320\(98\)00008-8](http://dx.doi.org/10.1016/S0951-8320(98)00008-8).
- [28] D.G. Sanchez, B. Lacarrière, M. Musy, B. Bourges, Application of sensitivity analysis in building energy simulations: combining first- and second-order elementary effects methods, *Energy Build.* 68 (2014) 741–750, <http://dx.doi.org/10.1016/j.enbuild.2012.08.048>.
- [29] J.C. Lam, S.C.M. Hui, Sensitivity analysis of energy performance of office buildings, *Build. Environ.* 31 (1) (1996) 27–39, [http://dx.doi.org/10.1016/0360-1323\(95\)00031-3](http://dx.doi.org/10.1016/0360-1323(95)00031-3).
- [30] Y.-J. Kim, S.-H. Yoon, C.-S. Park, Stochastic comparison between simplified energy calculation and dynamic simulation, *Energy Build.* 64 (2013) 332–342, <http://dx.doi.org/10.1016/j.enbuild.2013.05.026>.
- [31] M. Liu, K.B. Wittchen, P.K. Heiselberg, F.V. Winther, Development and sensitivity study of a simplified and dynamic method for double glazing facade and verified by a full-scale facade element, *Energy Build.* 68 (2014) 432–443, <http://dx.doi.org/10.1016/j.enbuild.2013.03.056>.
- [32] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Saf.* 52 (1996) 1–17, [http://dx.doi.org/10.1016/0951-8320\(96\)00002-6](http://dx.doi.org/10.1016/0951-8320(96)00002-6).



Contents lists available at ScienceDirect

Energy & Buildings

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

Prerequisites for reliable sensitivity analysis of a high fidelity building energy model

Steffen Petersen*, Martin Heine Kristensen, Michael Dahl Knudsen

Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, DK-8000 Aarhus C, Denmark

ARTICLE INFO

Article history:

Received 22 May 2018
 Revised 14 September 2018
 Accepted 25 October 2018
 Available online 5 November 2018

Keywords:

Sensitivity analysis
 Morris method
 Sobol method
 Building energy modelling

ABSTRACT

Sensitivity analysis (SA) can be applied to building energy models (BEM) to identify which input parameters that drive the majority of the model output variation. The screening-based Morris method is often applied for this purpose; however, consideration regarding the effect of the user-defined number of levels (p) and trajectories (r) on the obtained results are rare. This paper investigates how the choice of p and r affects the outcome of a SA using the Morris method on a high fidelity BEM. The results indicates that the Morris method was not able to replicate the ranking from the variance-based Sobol' method no matter the choice of r and p . It was, however, able to identify groups of input parameters (parameter clusters) most sensitive to the model output variability, but it required significantly more r than usually applied in studies featuring the Morris method. The reason is that marginal differences in absolute values of elementary effects (the sensitivity indices of the Morris method) for some input parameters may lead to a change in ranking position several times as the number of r increases. Users of the Morris method must therefore not be predetermined on the size of the parameter cluster; instead, they must make a visual assessment of the convergence of the parameter ranking to qualitatively determine the appropriate size of parameter cluster. The final recommendation for future studies deploying the Morris method for SA applied to a high fidelity BEM is to choose $p \geq 4$ as it seems to lead the analysis towards a more truthful ranking, and then run simulations in steps of $r = 100$ when making the visual assessment to determine convergence and the size of parameter cluster. The identified need for more r questions the general notion that the Morris method is a computationally efficient screening method in terms of absolute time use. However, the Morris method is still much more computational efficient than a Sobol'-based analysis if the purpose of the SA is to identify a cluster of input parameters most sensitive to the model output variability.

© 2018 Published by Elsevier B.V.

1. Introduction

Building designers may find it informative to employ a sensitivity analysis (SA) to a building energy model (BEM) to identify which design variables that drive the majority of the model output variation in terms of indoor climate and energy use. SA methods for this purpose can in general be categorised as either local sensitivity analysis (LSA) or global sensitivity analysis (GSA) [1].

LSA methods rely on a one-parameter-at-a-time (OAT) technique where all parameter values have equal probability of occurrence. The OAT technique means that LSA methods do not account for any effects from correlated input parameters. However, LSA methods are easy to implement and fast to conduct as they require only few model evaluations.

The GSA category covers a range of methods applying different techniques. Common for the methods in the GSA category is that they evaluate the effect of an input parameter on the output by varying not only the parameter in question, but all other input parameters chosen for analysis as well. GSA methods are therefore able to include effects from correlated input parameters as well as non-linear and non-additive model behavior. The outcome of a GSA may therefore be more reliable than the outcome from a LSA but GSA methods are more complicated to implement and are significantly slower to conduct as they require many model evaluations.

A specific group of GSA methods are the so-called screening methods [2]. Screening-based SA methods are often considered useful for qualitative identification of design variables to which the model output variability is most sensitive, whereas more advanced GSA methods, such as a variance-based method, must be applied if a quantitative ranking of parameters is desirable. The screening method initially described by Morris [3], and since refined and expanded by different authors [4,5], seems to be widely

* Corresponding author.

E-mail address: stp@eng.au.dk (S. Petersen).

Table 1

List of studies reported in literature using the Morris method for BEM-based analysis. Year: year of publication, Purpose: the purpose of using the Morris method, BEM: the building energy model(s) used, Levels and Trajectories: the reported settings used, Arguments: the arguments used for choosing p and r (respectively). N/A means “no argument”.

Reference	Year	Purpose	Building energy model (BEM)	Levels (p)	Trajectories (r)	Arguments	
de Wit and Augenbroe [17]	2002	Ranking input parameters	ESP-r and BFEP	N/A	5	p : N/A, r : a crude uncertainty analysis was desired.	
Corrado and Mechri [18]	2009		ISO 13790 (quasi-steady state)	N/A	N/A	p : N/A, r : N/A	
Heiselberg et al. [10]	2009		ISO 13790 (quasi-steady state)	4	4	p : N/A, r : minimum according to literature [31].	
Sanchez et al. [19]	2014		ESP-r	10	10	p : N/A, r : N/A	
Hemsath and Bandhosseini [20]	2015		N/A	4	16	p : To enable building orientation in steps of 45° in the sampling, r : N/A	
Østergaard et al. [21]	2015		ISO 13790 (quasi-steady state)	8	10 (100)*	p : N/A, r : N/A	
Yang et al. [22]	2016	Identifying influential input parameters for more detailed analysis	EnergyPlus	N/A	5	p : N/A, r : N/A	
Faggianelli et al. [29]	2017		EnergyPlus	4	20 (200)*	Greater than the recommendations by Saltelli et al. [9]	
de Wit [23]	1997		BFEP	2	3	p : N/A, r : N/A	
Brohus et al. [24]	2009		ISO 13790 (quasi-steady state)	N/A	N/A	p : N/A, r : N/A	
Booth et al. [25]	2012		ISO 13790 (quasi-steady state)	N/A	N/A	p : N/A, r : N/A	
Kim et al. [16]	2013		ISO 13790 (quasi-steady state) and EnergyPlus	4	4	p and r : minimum values according to literature [14,16].	
Le Drau and Heiselberg [26]	2014		Bottom-up heat balance	8	90	p : N/A, r : N/A	
Yang and Becerik-Gerber [27]	2015		EnergyPlus	N/A	5	p : N/A, r : to make the analysis efficient.	
Østergaard et al. [28]	2015		ISO 13790 (hourly, but simplified)	8	500	p : N/A, r : to obtain consistent ranking.	
Nguyen and Reiter [13]	2015		How ranking of input parameters is affected by SA method	EnergyPlus	4/6/8	49/70	Investigates the effect of combinations of level and samples on BEM outcome.
Kristensen and Petersen [7]	2016		ISO 13790 (quasi-steady state) ISO 13790 (hourly)	5	250	p : N/A, r : to obtain consistent ranking	
Menberg et al. [11]	2016		TRNSYS	N/A	150	p : N/A, r : to obtain consistent ranking	
Kristensen and Petersen [30]	2018	ISO 13790 (quasi-steady state) ISO 13790 (hourly)	4	300	p : N/A, r : N/A		

* This study used the modified sampling method proposed by Campolongo et al. [5] where an initial set of trajectories is reduced prior to simulation.

used for BEM-based analysis; see Table 1 for an overview of BEM-based studies featuring the Morris method. A compelling argument for applying the Morris method, instead of a more comprehensive variance-based GSA method, is that it is a computational efficient alternative if only a rough ranking of the parameters is desired [6]. However, the analysis provided by Kristensen and Petersen [7] suggested that the Morris method is able to come up with an identical ranking of the input parameters most sensitive to the output of simplified BEMs, as the variance-based GSA method of Sobol' [8] using less computational time, but only when the probability density functions (PDF) of the input parameters are uniformly distributed. The Morris method could therefore be a computational efficient alternative to a global SA method e.g. in the early design stage. However, there are still issues to be investigated to fully understand the possibilities and limitations of the Morris method when used for SA of BEM. To explain these issues, the following sections provides a short description of the Morris method (Section 1.1), a literature review on the use of the Morris method for BEM-based analysis (Section 1.2), before finally outlining the specific contribution of this paper to the existing knowledge base (Section 1.3).

1.1. The Morris method

This section provides a short description of the Morris method, which is intended to serve as background for the motivation of the

investigation presented in this paper; see ref. [3–5] for more detailed descriptions of the method.

The user of the Morris method needs to define a model input space (Ω) of interest to be explored by the SA. This k -dimensional input space Ω is comprised by user-defined input parameters x_i for $i = 1, 2, \dots, k$, where k is the number of chosen input parameters to be investigated. For each parameter x_i , the user must define a range of possible values, i.e. a minimum value (x_i^-) and a maximum value (x_i^+), to be explored by the SA. The ranges are subdivided in a user-defined p number of points, denoted *levels*, with a distance Δ between them. Using uniform input distributions, Δ is obtained by dividing the interval in which each input parameter varies (i.e. from x_i^- to x_i^+) into equally large parts. However, the user can also let input parameters follow non-uniform distributions by manually selecting the levels e.g. as the quantiles of the distribution [9]. Once the user has defined Ω , the Morris method employs a random one-at-a-time (OAT) sampling procedure to generate trajectories through Ω with each trajectory comprising $k+1$ random model realisations from Ω . This sampling procedure is repeated r times, each with randomly dispersed starting points for the trajectories, creating a global set of $r \cdot (k+1)$ building energy models to be simulated. The so-called elementary effect (EE_i) for each input parameter x_i is then calculated from the BEM output for every r set of $k+1$ models consequently providing r independent and identically distributed estimates of the EEs for each input parameter (Eq. 1).

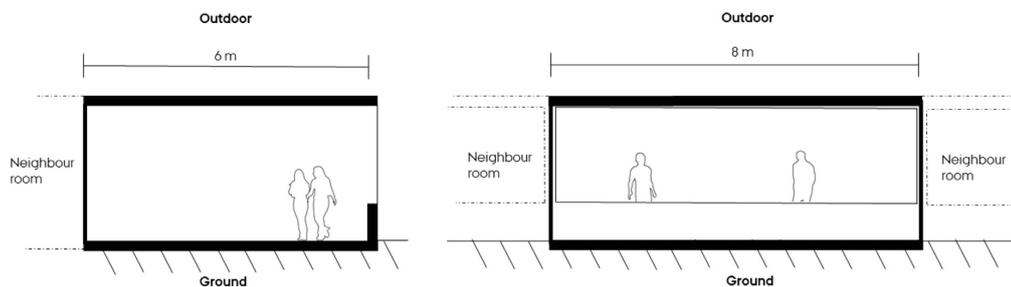


Fig. 1. Internal dimensions and boundary conditions for the office building zone. Left: vertical section of the zone. Right: front elevation of the façade. The walls facing neighbour rooms are assumed adiabatic.

$$EE_i(x) = \frac{Y(x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - Y(x_1, \dots, x_k)}{\Delta} \quad (1)$$

A central assumption in the Morris method is that the distribution of EE_i is Gaussian for each input parameter independently, $EE_i \sim N(\mu_i, \sigma_i^2)$. The model output sensitivity to the input parameters can be assessed using the mean of the absolute value of the elementary effects, μ_i^* , for an r set of trajectories (indexed $t = 1, 2, \dots, r$) is used for ranking the parameters in order of importance (Eq. 2),¹ while the standard deviation σ_i is used as a measure of the interactions with other parameters and any non-linear effects that the parameter takes part in (Eq. 3).

$$\mu_i^* = \frac{1}{r} \sum_{t=1}^r |EE_{i,t}| \quad (2)$$

$$\sigma_i = \sqrt{\frac{1}{(r-1)} \sum_{t=1}^r (EE_{i,t} - \mu_i)^2} \quad (3)$$

1.2. Literature review

As described in the previous section, the Morris method relies on the ranking of μ^* and σ as measures of input parameter sensitivity. This immediately raised the question: what is the sensitivity of the value of μ^* and σ to the user-defined number of levels (p) and trajectories (r)? A review of the literature where the Morris method has been applied for BEM-based analysis indicated that this concern seems to be rare in existing studies using the Morris method for BEM-based analysis. Table 1 indicates that only very few have arguments for choosing values for p and r ; in fact, some authors do not even state the value for p and r used in their analysis. Especially information about p is absent. Heiselberg et al. [10] state that literature recommends a minimum value of $r=4$ to make sure that the region of variation is reasonably covered for all input parameters, while a value of $r=10$ is recommended to obtain very reliable results. No considerations regarding the choice of p are provided. More recent studies includes much larger values of r (> 100) to a fixed value of p arguing that this is necessary to gain a consistent parameter ranking [7,11,12]. Nguyen and Reiter [13] was the only study found that investigated the sensitivity of their results to different values for p and r . They found that different sets of p and r could result in different parameter rankings, possibly because the random sampling sometimes led to an uneven distribution of input vectors on the designed levels of input parameters. A similar issue related to the random sampling was reported by Menberg et al. [11] who found that the parameter ranking can be biased by the occurrence (or absence) of outliers

¹ Eq. (2) is a revision of the original expressions by Morris [3], see Saltelli et al. [9] for details.

in individual Morris method runs as a consequence of the low number of r in combination with a comparably large parameter space (number of p is unknown). The findings of Nguyen and Reiter [13] and Menberg et al. [11] is aligned with Saltelli et al. [9] who from a general point of view note that the choice of p is strictly linked to the choice of r . More specifically, Saltelli et al. [9] state that an increase of r increases the probability that all levels are explored at least once, and that while a high value of p only appears to augment the accuracy of the sampling, it must be coupled with the choice of a high value of r ; otherwise, many possible levels will remain unexplored. They indirectly suggests the use of $p=4$ and $r=10$ as it has produced good results in previous experiments involving chemical and environmental models [2,14–15]. BEM-based studies like Heiselberg et al. [10] and Kim et al. [16] refers to the experiences of such studies as argument for choosing p and r for BEM-based analysis. However, the study by Kristensen and Petersen [7] indicates that the ranking of input parameters using the Morris method can be influenced by the choice of BEM. This suggests that the appropriate choice of p and r may also depend on the model to which the Morris method is applied.

1.3. Scope of this paper

Based on the findings from the review of current literature, we found it necessary to conduct a study on how the choice of p and r affects the outcome of a SA using the Morris method on a high fidelity BEM.² The intention of the study is to provide a guideline for future studies to select the minimum values of r for a certain p needed for the Morris method to consistently rank input parameters according to their influence on the model output variability of a high fidelity BEM.

2. Method

A one-storey office building illustrated in Fig. 1 was modelled as one thermal zone in EnergyPlus (EP) [32] using the inputs listed in Table 2. This model was then subject to a SA using the Sobol' method [8], the Morris method [3], and a One-At-the-Time (OAT) method [33], respectively. The quantities of interest for the SA was the energy use per year (kWh/year) for heating, cooling and mechanical ventilation, respectively. The annual simulations were performed using the Danish design reference year [37] and with a simulation time step of two minutes. Execution of the multiple

² The term 'high fidelity BEM' is used to differentiate tools that attempts to model physical behavior with a high level of detail (e.g. EnergyPlus and TRNSYS) from tools relying on more simplified representations of the physics (e.g. linearised hourly models and monthly quasi-steady state models).

Table 2

Description of input data to the EP model. Model parameters are defined by fixed and/or variable inputs where variable inputs are subject to SA. The numeric values of the variable inputs are provided in Table 3. The far right column explains the EP modelling assumptions for the model parameter. The heat transfer algorithm used in all simulations is "conduction transfer functions" and a time step of 2 minutes.

	Model parameter	Fixed input	Variable input	Modelling assumptions
Building	North axis	-	Orientation	-
	Area	-	Window-to-façade area ratio	-
	Frame	-	U-value; area fraction of window hole	"Simpleglazing" is used in favour of a more detailed option. It is noted that this method can lead to a minor error in absolute energy use compared to the full spectral method [34]. The total U-value of windows is calculated according to EN ISO 10077-1 [35].
	Glazing	LT is assumed to be a function of glazing SHGC; $LT = 0.5 \cdot SHGC + 0.45$	U-value; glazing SHGC; frame width; Ψ -value of spacer profile	
	Overhang	-	Length perpendicular out from the façade	The overhang always has the same width as the window, and is always flush with the top of the window.
Constructions	External façade	Bricks: 0.108 m; $\lambda = 0.034$ W/(mK); $\rho = 1600$ kg/m ³ ; $c = 840$ J/(kg K) Mineral wool: $\lambda = 0.037$ W/(mK); $\rho = 200$ kg/m ³ ; $c = 700$ J/(kg K) Concrete: 0.1 m	- Thickness of insulation layer Heat conduction (λ); density (ρ); specific heat capacity (c)	- Density ρ is a variable; $\lambda = 0.0015\rho - 1.5$ and $c = 0.043\rho + 830$ (derived from DS 418 [36]).
	Floor construction	Mineral wool: $\lambda = 0.034$ W/(m K); $\rho = 200$ kg/m ³ ; $c = 700$ J/(kg K) Concrete: 0.1 m; $\lambda = 2.1$ W/(m K); $\rho = 2400$ kg/m ³ ; $c = 1000$ J/(kg K) Wood floor: 0.02 m; $\lambda = 0.12$ W/(m K); $\rho = 850$ kg/m ³ ; $c = 800$ J/(kg K)	Thickness of insulation layer -	Ground modelled as ground-coupled slab model (GroundDomain). -
	Roof	Mineral wool: $\lambda = 0.034$ W/(m K); $\rho = 200$ kg/m ³ ; $c = 700$ J/(kg K) Concrete: 0.1 m; $\lambda = 2.1$ W/(m K); $\rho = 2400$ kg/m ³ ; $c = 1000$ J/(kg K) Air gap: $R = 0.17$ (m ² K)/K Gypsum: 0.013 m; $\lambda = 0.16$ W/(m K); $\rho = 800$ kg/m ³ ; $c = 1090$ J/(kg K)	Thickness of insulation layer - - -	- - -
	Internal walls	Concrete: 0.1 m	Heat conduction (λ); density (ρ); specific heat capacity (c)	Density ρ is a variable; $\lambda = 0.0015\rho - 1.5$ and $c = 0.043\rho + 830$ (derived from DS418 [36]).
	Ventilation	Infiltration	Coefficients: $A = 0.606$; $B = 0.03636$; $C = 0.117$; $D = 0$	Infiltration rate
	Ventilation rate, in-use	-	Ventilation rate (m ³ /s)	Constant air volume with a constant inlet air temperature of 18 °C.
	Ventilation rate, out-of-use	-	Ventilation rate (m ³ /s)	Only available in the months Jun-Aug (summer).
	Heat recovery rate	-	Sensible heat recovery effectiveness	-
	Heating set point, in-use	-	Set point	Added directly to the zone air (radiator).
	Cooling set point, in-use	-	Set point	Removed directly from the zone air (chilled beam).
	Heating set point, out-of-use	-	Set point	Added directly to the zone air (radiator).
	Cooling set point, out-of-use	-	-	Cooling not available in out-of-use periods.
	Specific fan power, ventilation	Motor efficiency = 0.9; Motor in air stream fraction = 0; Fan total efficiency = 0.7	Pressure rise (Pa)	The pressure rise is varied to express how different SFP (kJ/m ³) affects inlet temperature and thereby heating and cooling load. We derive the pressure rise from the sampled SFP (see Table 3); SFP is fan total efficiency (-) divided by pressure rise (Pa).
Schedule	COP, mechanical cooling	-	COP (-)	-
	In-use, out-of-use periods	In-use: 8:00–17:00, weekdays Out-of-use: remaining hours Danish Design reference year [37]	-	-
Weather data	People load	Six persons	-	-
	Heat load from appliances, time-in-use	-	W/m ²	Auto-calculated sensible heat fraction
	Standby heat load from appliances, out-of-use	-	W/m ²	-
	Lighting (daylight controlled)	-	-	The heat load from daylight controlled lighting systems is omitted from the analysis. However, a fraction of the 'heat load from appliances' could be regarded as a simple representation of this heat load.

Table 3
The PDFs of the 24 variable input parameters in the EP model.

Input parameters	Unit	Uniform PDF [Min:Max]
Building orientation	Degrees	[0;360]
Room height	m	[2.5;3.5]
Insulation thickness, external walls	m	[0.1;0.35]
Insulation thickness, roof	m	[0.1;0.45]
Insulation thickness, floor	m	[0.1;0.35]
Window-to-façade area ratio	–	[0.15;0.95]
Glazing U-value	W/(m ² K)	[0.5;1.0]
Glazing SHGC	–	[0.15;0.6]
Glazing linear loss (Ψ)	W/(m K)	[0.03;0.2]
Window frame U-value	W/(m ² K)	[0.8;2.0]
Window frame fraction	–	[0.05;0.25]
Overhang ^a	m	[0;1]
Ventilation rate, in-use	l/s/person	[4;10]
Infiltration rate	l/s/m ² @50 Pa	[0.5;1.5]
Heat recovery rate	–	[0.65;0.9]
Heat load from appliances, in-use	W/m ²	[2;10]
Standby heat load from appliances, out-of-use ^a	–	[0.05;1]
Thermal capacity, inner layer of walls	KJ/m ² K	[1200;2400]
Heating set point, in-use	°C	[20;24]
Cooling set point, in-use	°C	[25;27]
Heating set point, out-of-use	°C	[16;20]
Night ventilation rate, summer	l/s m ²	[0;2]
Specific fan power (ventilation)	kJ/m ³	[0.5;2]
COP (mechanical cooling)	–	[1;5]

^a Fraction of 'Heat load from appliances, in-use'.

model evaluations in EP needed for the SA analyses was handled using the 'multidirrun' file provided in the EP program folder. The following sections provides further details on these methods and why they are applied in the study.

2.1. Sobol' analysis

The purpose of conducting an SA using Sobol' method was to establish a benchmark for assessing the minimum values of trajectories r for a certain level p needed for the Morris method to consistently rank input parameters according to their sensitivity to the model output variability. The SA method by Sobol' [8] is a global variance-based method which is able to attribute the total model variance to individual input parameters. The contribution of each parameter in explaining total model variance is often assessed using the so-called first-order effects (S_i) that describe the immediate effect of variations of the parameters independently, and so-called total-order effects (S_{Ti}) that take into account all possible interactions and non-linear effects that the parameters take part in. In this study, the Sobol' sensitivity indices S_i and S_{Ti} was obtained the same way as described by Kristensen and Petersen [7]. We used S_{Ti} to rank input parameters because SA methods which includes higher order interactions in complex models is known to alter parameter rankings based on S_i or μ^*_i [38]. It therefore also seems reasonable to use ranking according to S_{Ti} as benchmark for the performance of the Morris method.

A significant benefit from using Sobol' method for SA is its ability to take into account non-uniform distributions – a feature that the standard Morris method is incapable of by definition. We therefore make use of uniformly distributed PDFs in the Sobol' method to make a fair benchmark for the Morris method. The PDFs for the input parameters are listed in Table 3. A total of $N \cdot (k + 2)$ model evaluations in EP needs to be calculated where N is the number of samples and k being the number of input parameters. The appropriate number of N relies on user-defined convergence criteria for S_i and S_{Ti} . We found it difficult to formulate a suitable convergence criteria for the quantities of interest in this study (energy use), which is why we decided to generate an immediate large quantity of models using $N = 10,000$ Latin hypercube samples

from the PDFs of the $k = 24$ input parameter listed in Table 3 resulting in 260,000 model evaluations, and then make a qualitative assessment of the convergence issue. The 95% confidence bounds of S_{Ti} were derived using 2000 bootstrapping samples.

2.2. Morris analysis

The principle of the Morris method has been described in Section 1.1; this section describes the assumptions used for the Morris analysis in this study. The model input space Ω for the Morris method was defined by the uniformly distributed PDFs of the 24 input parameters listed in Table 3. The original Morris sampling method [3] was applied using $r = 1000$ for six different values of p (2;4;6;8;10;12), resulting in 25,000 building zone models per level (150,000 simulations in total) for evaluation in EP. This way we are able to assess how an incrementally increasing number of r affects parameter ranking (i.e. μ^*_i listed in descending order) and, consequently, determine the minimum number of r needed for a consistent parameter ranking for different p . The reason repeating the Morris SA for various p is to investigate whether the choice of p affects the outcome of a Morris SA; the chosen p adds $p = 12$ to the range of typical values of p applied in previous studies (see Table 1). It is noted that μ^*_i cannot be used for quantification of the magnitude of parameter influence as one can do based on the S_{Ti} obtained in the Sobol' method.

2.3. One-At-the-Time method

The purpose of conducting an OAT analysis was to investigate whether this much less computationally demanding method compared to Sobol' and Morris is able to come up with the same ranking. If so, the use of OAT analysis for SA would be preferable as it is much more computationally efficient; only 49 simulations are required for the case used in this paper. The OAT method used for the analysis reported in this paper was based on partial derivatives where parameters are ranked according to a dimensionless sensitivity index S_i ; see e.g. refs. [7,32] for further details regarding this OAT method.

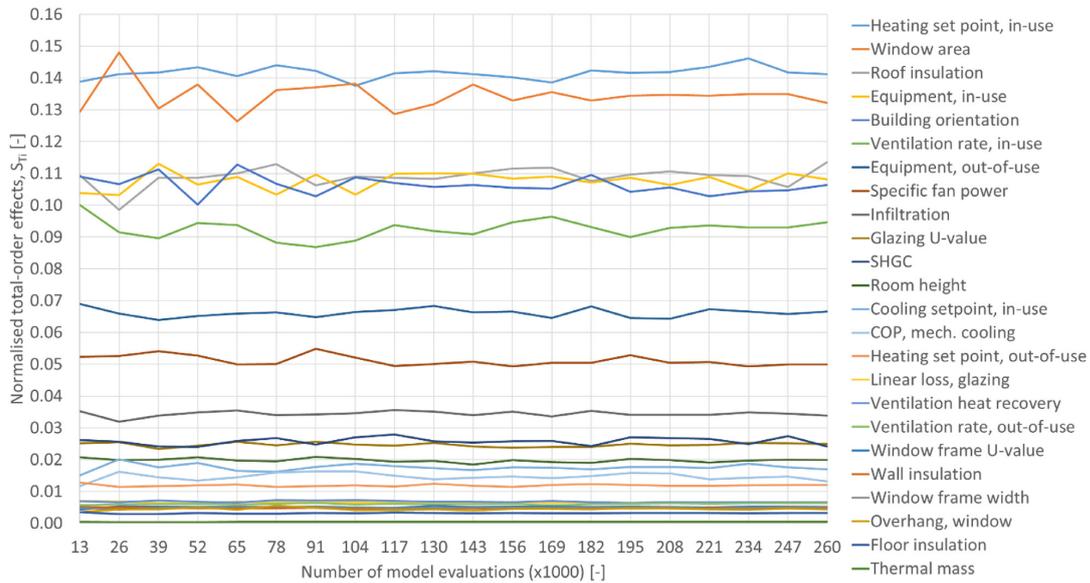


Fig. 2. Normalised total-order effects (S_{Ti}) of the total energy use as a function of model evaluations in steps of 13,000. The order of the legend corresponds to the order of the lines in the graph.

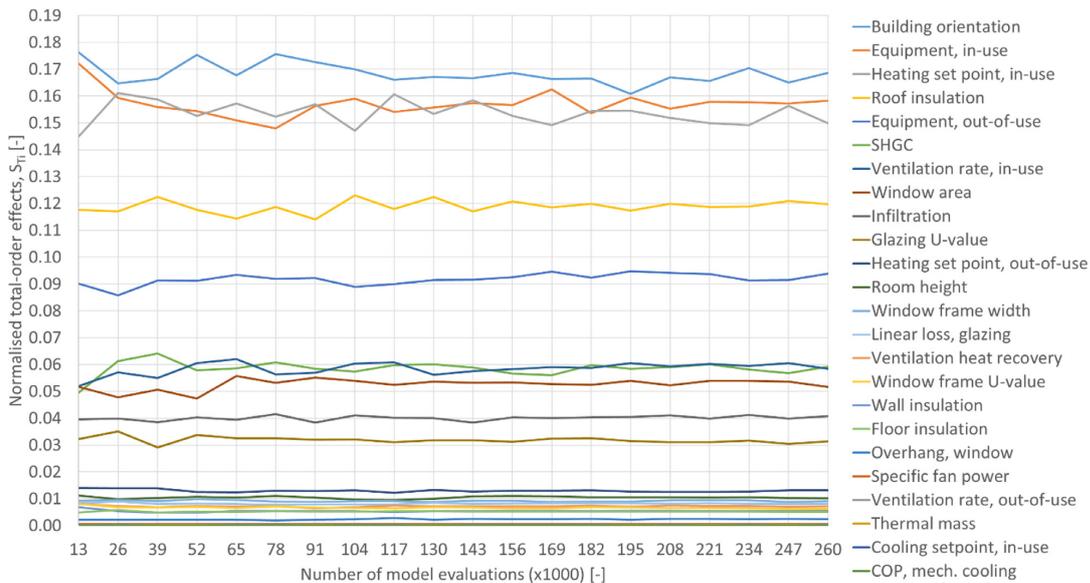


Fig. 3. Normalised total-order effects (S_{Ti}) of the heating energy use only as a function of model evaluations in steps of 13,000. The order of the legend corresponds to the order of the lines in the graph.

3. Results

For the Sobol' method, Figs. 2–4 depict the normalised S_{Ti} for the total energy use, heating only, and cooling only, respectively, as a function of model evaluations in steps of 13,000 (see Appendix A for further details). The figures show no clear sign of convergence even after 260,000 model evaluations. The reason is that the absolute difference between normalised S_{Ti} of some parameters is marginal (e.g. between the input parameters “Insula-

tion, roof” and “Appliances heat load, in-use” in Fig. 2) leading to many shifts in relative ranking as a function of model evaluations.

For the Morris method, Figs. 5–7 show the ranking of the 24 input parameters according to their μ^* , after $r = 1000$ for all investigated p alongside the ranking obtained using the Sobol' method for the total energy use, heating only and cooling only, respectively (see Appendix B for further details). From these figures it is evident that no matter the value of p , the Morris method using $r = 1000$ was never able to rank the parameters according to the ranking

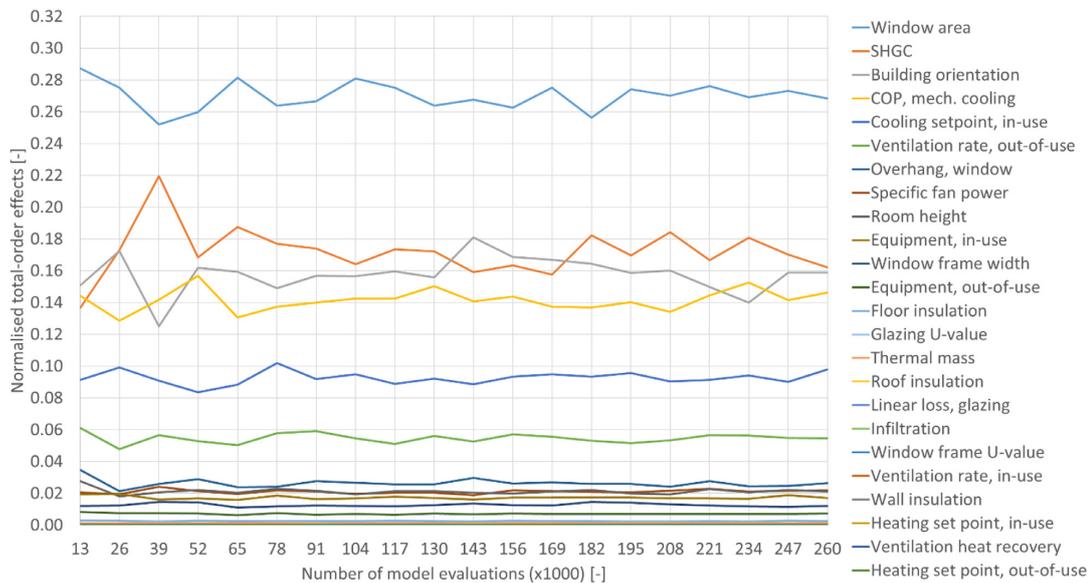


Fig. 4. Normalised total-order effects (S_{Ti}) of the cooling energy use only as a function of model evaluations in steps of 13,000. The order of the legend corresponds to the order of the lines in the graph.

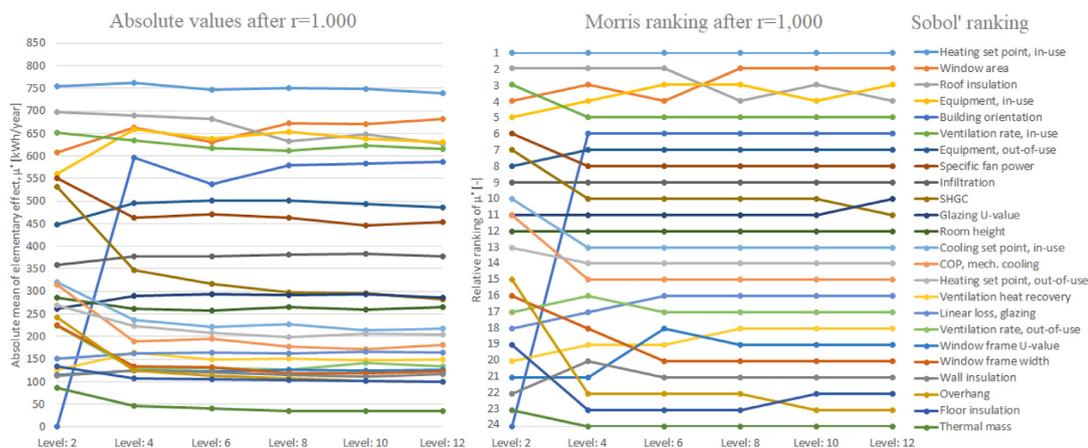


Fig. 5. Values of μ_i^* and relative ranking of the 24 input parameters according to their μ_i^* with respect to the total energy use after $r = 1000$ for all investigated p alongside the ranking obtained using S_{Ti} from the Sobol' method after 260,000 model evaluations. The order of the legend corresponds to the order of the lines in the graph.

based on S_{Ti} from the Sobol' method. However, it seems like the Morris method at some point during increasing r was able to consistently identify similar clusters of input parameters most influential to the variability of the model output as the Sobol' method. Our definition of such a 'cluster' is when the Morris method has identified the same group of input parameters as most influential to the variability of the model output as the Sobol' method but not ranked them in the same order. It is also noted that there seems to be a significant rearrangement of the parameter ranking when going from $p = 2$ to $p \geq 4$.

Table 4 shows the minimum values of trajectories r for all levels p needed for the Morris method to identify the same cluster of x_i that the Sobol' method identified as most influential to the variability of the model output. The influence of p on the num-

ber of r needed for a consistent identification varies depending on the type of energy consumption (total, heating, or cooling), and the number parameters included in the cluster of parameters most influential to the variability of the model output (top 1–11). Special for $p = 2$ is that μ_i^* of the parameter 'orientation' was always zero, and consequently never appeared in the top 11 parameters. The reason is that $p = 2$ leads to no actual variation of the orientation as the two levels are the minimum and maximum parameter values, 0° or 360° , respectively, which is both due south by definition. All calculations with $p = 2$ are therefore ignored in further interpretations of the results in Table 4. There is no clear tendency that the number of p has any effect on the needed number of r for a consistent identification of clusters of parameters most sensitive to the output variability in terms of the total energy use. For heat-

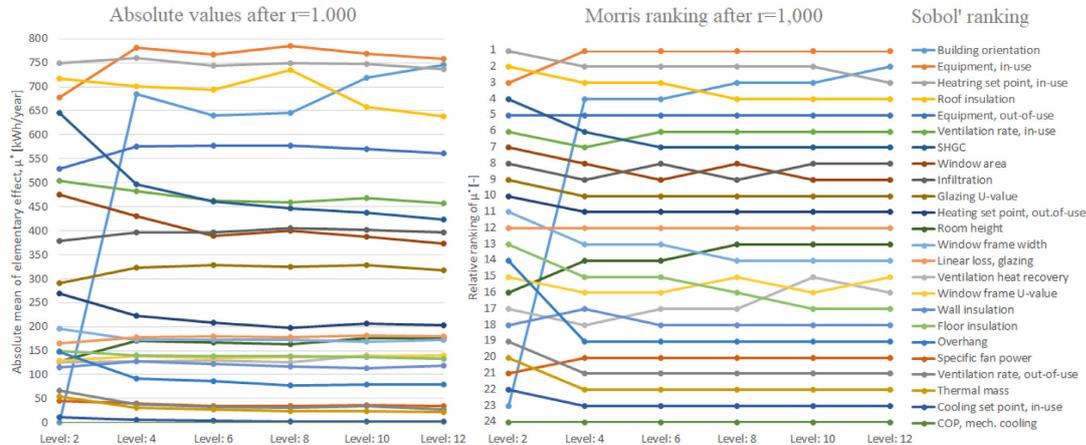


Fig. 6. Values of μ^*_i and relative ranking of the 24 input parameters according to their μ^*_i with respect to the heating energy use only after $r = 1000$ for all investigated p alongside the ranking obtained using S_{T1} from the Sobol' method after 260,000 model evaluations. The order of the legend corresponds to the order of the lines in the graph.

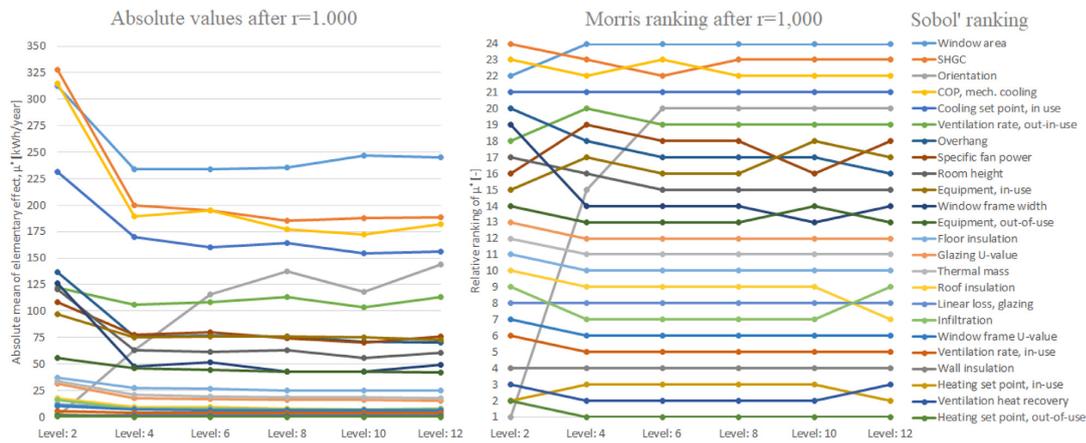


Fig. 7. Values of μ^*_i and relative ranking of the 24 input parameters according to their μ^*_i with respect to the cooling energy use only after $r = 1000$ for all investigated p alongside the ranking obtained using S_{T1} from the Sobol' method after 260,000 model evaluations. The order of the legend corresponds to the order of the lines in the graph.

ing energy only, there is a tendency that the identification of top 3 input parameters became better with increasing number of p . For cooling energy only, there seems to be a slight benefit from choosing $p > 4$. This is probably because the cooling system has non-linear behavior. For all three energy consumptions, the number of r needed to consistently identify top 1–11 input parameters for each p depends on the absolute difference between μ^*_i for all x_i . For example, the reason that the number of r for $p = 4$ for total energy use (see Table 4) increases from nine to 476 when screening for top 6 and top 7, respectively, is that the value of μ^*_i of the seventh and eighth parameter in top 8 are only marginally different until approx. $r = 476$. Prior to $r = 476$, the two parameters changes ranking position several times, and thereby the content of the top 7 cluster (see Appendix B).

The result of the OAT analysis is also listed in Table 4 and shows that the OAT method was rarely able to identify the same cluster of input parameters to which the model output variability was most sensitive as the Sobol' method (the full outcome of the OAT method is shown in Appendix C).

4. Discussion

As stated in the introduction, the aim of this study was to provide a guideline for future studies to select the minimum values of r for a certain p needed for the Morris method to consistently rank input parameters that has most influence on the model output variability. This aim was partly based on the findings by Kristensen and Petersen [7], which suggested that the Morris method is able to identify the same ranking of the input parameters most sensitive to the output of simplified BEMs as the Sobol' method provided that the PDFs of the input parameters are uniformly distributed. However, results of this study suggests that this is not true for high fidelity BEMs, but it seems to be able to identify clusters of input parameters to which the model output variability is most sensitive. Some overall guidelines for applying the Morris method to identify clusters of input parameters are provided in the following.³

³ The guidelines are only considered valid for a Morris analysis using uniform PDFs and sophisticated BEMs.

Table 4

The number of trajectories needed for the Morris method to consistently identify the same top 1–11 of x_i that the Sobol' method identified as most influential to the variability of the model output. The term 'never' means that the Morris method was not able to identify the parameters after $r=1000$ trajectories. The number in brackets in the heading of the columns is the cumulative sum of normalised S_{Ti} according to the Sobol' method. The OAT rows indicate whether the OAT method was able to identify the same top 1–11 as the Sobol' method (Yes/No).

Total energy use	Top 1 (14.1%)	Top 2 (27.5%)	Top 3 (38.4%)	Top 4 (49.2%)	Top 5 (59.8%)	Top 6 (69.0%)	Top 7 (75.7%)	Top 8 (80.8%)	Top 9 (84.2%)	Top 10 (86.9%)	Top 11 (89.4%)
$p=2$	21	Never	Never								
$p=4$	108	Never	283	137	Never	9	476	5	44	15	46
$p=6$	22	Never	Never	301	Never	27	155	33	97	24	26
$p=8$	27	286	Never	379	Never	20	29	44	72	11	11
$p=10$	29	46	373	634	Never	39	113	16	13	572	68
$p=12$	70	458	Never	71	Never	56	14	13	9	907	226
OAT	No	No	Yes	No	No	No	No	No	Yes	No	No
Heating energy use	Top 1 (16.9%)	Top 2 (32.6%)	Top 3 (47.8%)	Top 4 (59.7%)	Top 5 (68.9%)	Top 6 (74.8%)	Top 7 (80.6%)	Top 8 (86.0%)	Top 9 (90.0%)	Top 10 (93.1%)	Top 11 (94.4%)
$p=2$	Never	Never									
$p=4$	Never	Never	Never	11	60	Never	8	5	4	6	69
$p=6$	Never	Never	Never	12	25	881	25	Never	48	5	11
$p=8$	Never	Never	12	91	43	538	6	334	91	10	6
$p=10$	Never	Never	113	28	5	31	76	Never	86	6	29
$p=12$	Never	417	128	19	17	13	307	Never	3	4	46
OAT	No	Yes	No	No							
Cooling energy use	Top 1 (26.9%)	Top 2 (43.9%)	Top 3 (59.5%)	Top 4 (74.0%)	Top 5 (83.4%)	Top 6 (88.9%)	Top 7 (91.5%)	Top 8 (93.6%)	Top 9 (95.7%)	Top 10 (97.4%)	Top 11 (98.7%)
$p=2$	Never	Never									
$p=4$	1	726	Never	100	21						
$p=6$	42	Never	Never	Never	11	34	Never	33	Never	57	12
$p=8$	30	25	Never	Never	18	Never	308	Never	Never	30	64
$p=10$	12	101	Never	Never	5	32	Never	Never	Never	Never	Never
$p=12$	21	93	Never	Never	24	5	Never	Never	Never	Never	Never
OAT	Yes	No	No								

First of all, choosing $p \geq 4$ seems to lead the analysis towards a more truthful ranking and, consequently, a more reliable identification of most important parameter clusters – especially if orientation of window areas is included in the same way as in this paper. Note that an even higher value of p (closer to $p=12$) seems to be beneficial if only cooling energy is of interest.

It is difficult to provide an exact recommendation of the number of r needed for an outcome of the Morris analysis similar to the Sobol' method. The reason is that any marginal differences in values of μ^*_i between two parameters means that an excessive number of r is needed for the ranking to converge (see result section). This reason also makes it difficult to predetermine the appropriate size of the cluster of most influential parameters. For example, for total energy use (Table 4) no matter the choice of $p > 2$, it would make sense to have top 6, 8 or 9 in the cluster containing the most important parameters—but e.g. not top 5 as μ^*_i of the sixth parameter is always close to the fifth, and not top 7 as it is difficult to determine which parameters are actually belonging to this cluster. It is noted that this is not only an issue for the Morris method; the ranking according to the Sobol' method is also sensitive to the absolute difference between indices (S_{Ti}). The value of the individual S_{Ti} should therefore also be listed to enable a qualitative assessment on how many of the ranked input parameters that would be appropriate to highlight as most sensitive to the model output variability according to the Sobol' method.

The uncertainty of parameter ranking from the Sobol' method seem to defeat the whole purpose of using the Sobol' method to benchmark the ranking from Morris method. This is why it seems more reasonable to use the Morris method—and even the Sobol' method—to identify a certain clusters of most influential input parameters rather than attempting to obtain a true ranking of the input parameters. Based on the results of this study, the recommended approach for the identification of a cluster of most influ-

ential input parameters using the Morris method is to (1) generate models for $r=1000$ (or more) but start by simulating only a fraction of the models, e.g. for the first $r=100$ models, (2) calculate and plot μ^*_i for the quantity of interest for all x_i as a function of r like in Appendix B, and (3) make a qualitative (visual) assessment of whether the values of μ^*_i have converged to a degree where it seems possible to determine a cluster for the most influential parameters. It is noted that one should not have a pre-determined cluster size for the most influential parameters prior to this qualitative assessment but decide how many parameters to include during the qualitative assessment. If μ^*_i seems not to be converged, then simulate the performance of the next e.g. $r=100$ models, update the plot of μ^*_i for the quantity of interest for all x_i as a function of r , and make a new qualitative assessment. Repeat this process until μ^*_i seems converged. A similar approach could also be used for the Sobol' method to investigate convergence of S_{Ti} and appropriate cluster size.

The above recommendation for the Morris method suggests a step of $r=100$ in the attempt to obtain convergence of μ^*_i . This is far from what is commonly used for similar analyses using high fidelity BEMs (see Table 1); here r between five and ten is often used, which corresponds to the recommendations provided in the fundamental literature describing the Morris method [2,14,15]. This may change the notion of Morris being a computationally efficient method for parameter screening. One model evaluation of the EP model used in this study takes approx. 1 minute, which leads to a total calculation time of approx. 42 hours for a model with 24 variable input parameters and $r=100$ (2500 model evaluations). It is therefore of practical interest to reduce the number of r needed for a reliable Morris SA as this also would reduce computational time. One option that could be investigated in future studies is to rank the input parameters according to the median value of EEs, as findings by Menberg et al. [11] indicated that ranking based on median

values converges after fewer numbers of r compared to ranking based on mean values. Another option could be to use the modified sampling method suggested by Campolongo et al. [5] who claims that this method is always to be preferred over the original Morris sampling method as it reduces the number of model executions needed for a reliable analysis.

We acknowledge that the use of different weather data, a different set of input parameters and/or different PDF ranges may lead to different rankings than the ones observed in this study, but the above recommendations is in this regard considered to be on the safe side. Furthermore, this study made use of EnergyPlus for model evaluations but it seems like many prefer to use more simplified BEMs (Table 1). A future study could be to repeat the study of this paper using simplified BEMs for model evaluations. In this relation, it would also be relevant to analyse the consequence of different BEM approaches on input parameter ranking as the study by Kristensen and Petersen [7] indicates that it can be influenced by the choice of BEM.

5. Conclusion

The intention of this study was to provide a guideline for future studies to select the minimum values of trajectories (r) for a certain level (p) needed for the Morris method to consistently rank input parameters according to their influence on the model output variability of a high fidelity BEM when compared to parameter ranking using the Sobol' method. Results indicate that the Morris method is not able to replicate the ranking from Sobol' method no matter the choice of r and p . The reason is that ranking according to the Morris method as well as the Sobol' method is quite sensitive to marginal absolute differences in the metric used for ranking. Consequently, it is not possible to provide guidelines for future studies with precise values for r and p as intended. However, the study enables us to provide some guidance on how to produce a reliable SA using the Morris method.

The Morris method may not be able to generate a reliable ranking of input parameters but it is able to identify the same cluster of input parameters—i.e. groups of unranked input parameters—most

sensitive to the model output variability as the Sobol' method. However, reliable identification of such clusters seems to require significantly more r than usually applied in studies featuring the Morris method. Furthermore, users must not be predetermined on the size of the parameter cluster prior to the analysis; instead, one must make a visual assessment of the convergence of the parameter ranking to qualitatively determine the size of parameter cluster.

The need for more r may question the general notion that the Morris method is a computationally efficient screening method in terms of absolute time use, but it is still much more computationally efficient than a Sobol'-based analysis. A simple One-At-the-Time method, which can be regarded the best sensitivity analysis method in terms of computational efficiency, was also tested; however, it did not produce clusters comparable to the outcome of the Sobol' or Morris method.

Acknowledgement

The authors gratefully acknowledge associate Professor Christian Anker Hviid at the Technical University of Denmark for lending us the powerful computer cluster used for the numerous EnergyPlus simulations.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.enbuild.2018.10.035](https://doi.org/10.1016/j.enbuild.2018.10.035).

Appendix A. Sobol' method

The Figs. A.1–A.3 below depict the outcome of the Sobol' analysis, i.e. the total-order effects (S_{Ti}) for each input parameter for the total energy need (heating + cooling + ventilation), heating energy only, and cooling energy only, respectively. S_{Ti} for energy for ventilation only is not displayed because it only is linearly affected by the input parameter 'Ventilation, in-use', 'Ventilation, out-of-use', and 'Specific fan power'.

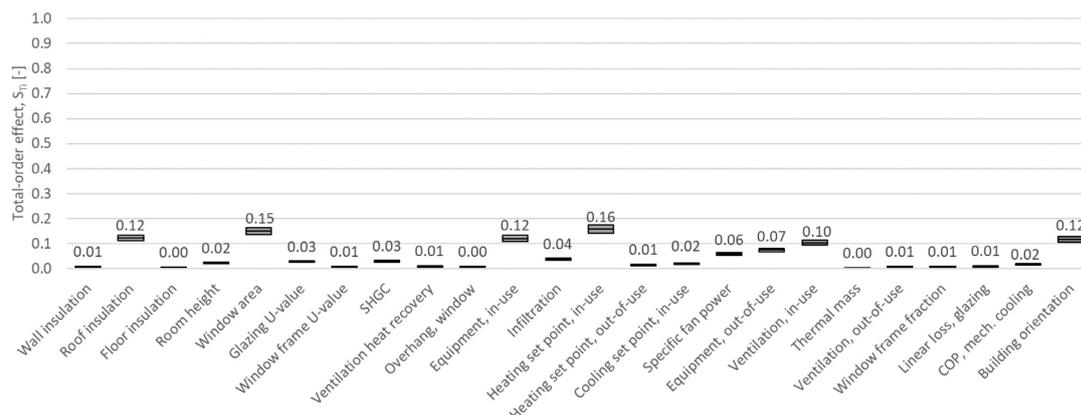


Fig. A.1. Total-order effects (S_{Ti}) for each input parameter for the total energy need (heating + cooling + ventilation) after 260,000 model evaluations. Boxes indicate the 95% confidence intervals around the mean value (black line).

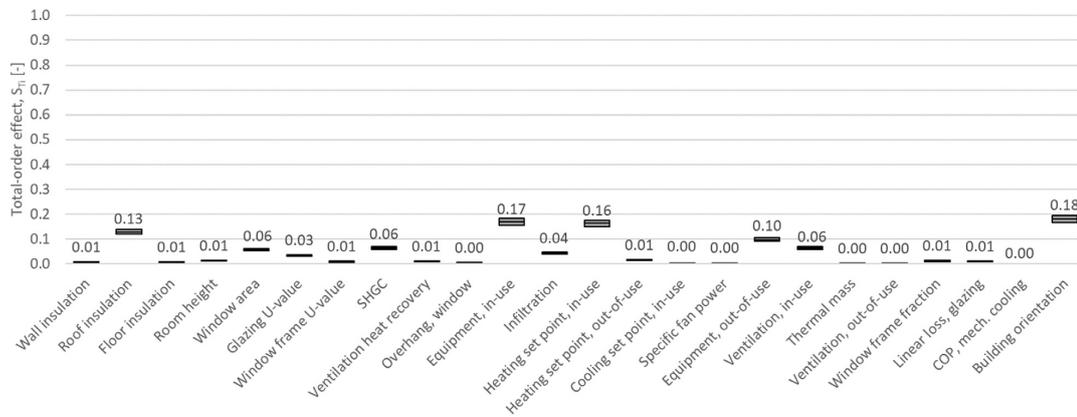


Fig. A.2. Total-order effects (S_T) for each input parameter for the heating energy only after 260,000 model evaluations. Boxes indicate the 95% confidence intervals around the mean value (black line).

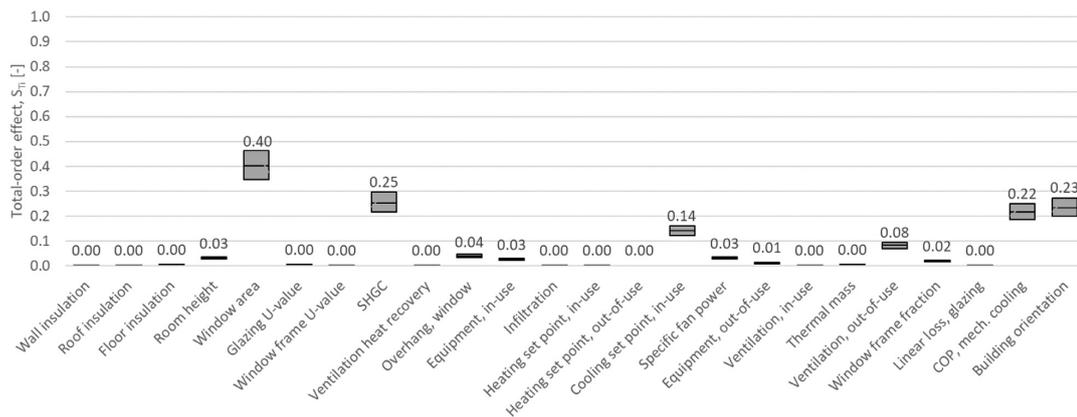


Fig. A.3. Total-order effects (S_T) for each input parameter for the cooling energy only after 260,000 model evaluations. Boxes indicate the 95% confidence intervals around the mean value (black line).

Appendix B. Morris method

Figs. B.1–B.3 illustrate for every p the evolution of μ_i on the total energy need (heating + cooling + ventilation), heating only, and

cooling only, respectively, for all x_i as a function of r . The evolution of μ_i^* on ventilation is not displayed because the value of μ_i^* for all of the 24 input parameters is not affected by $r > 1$.

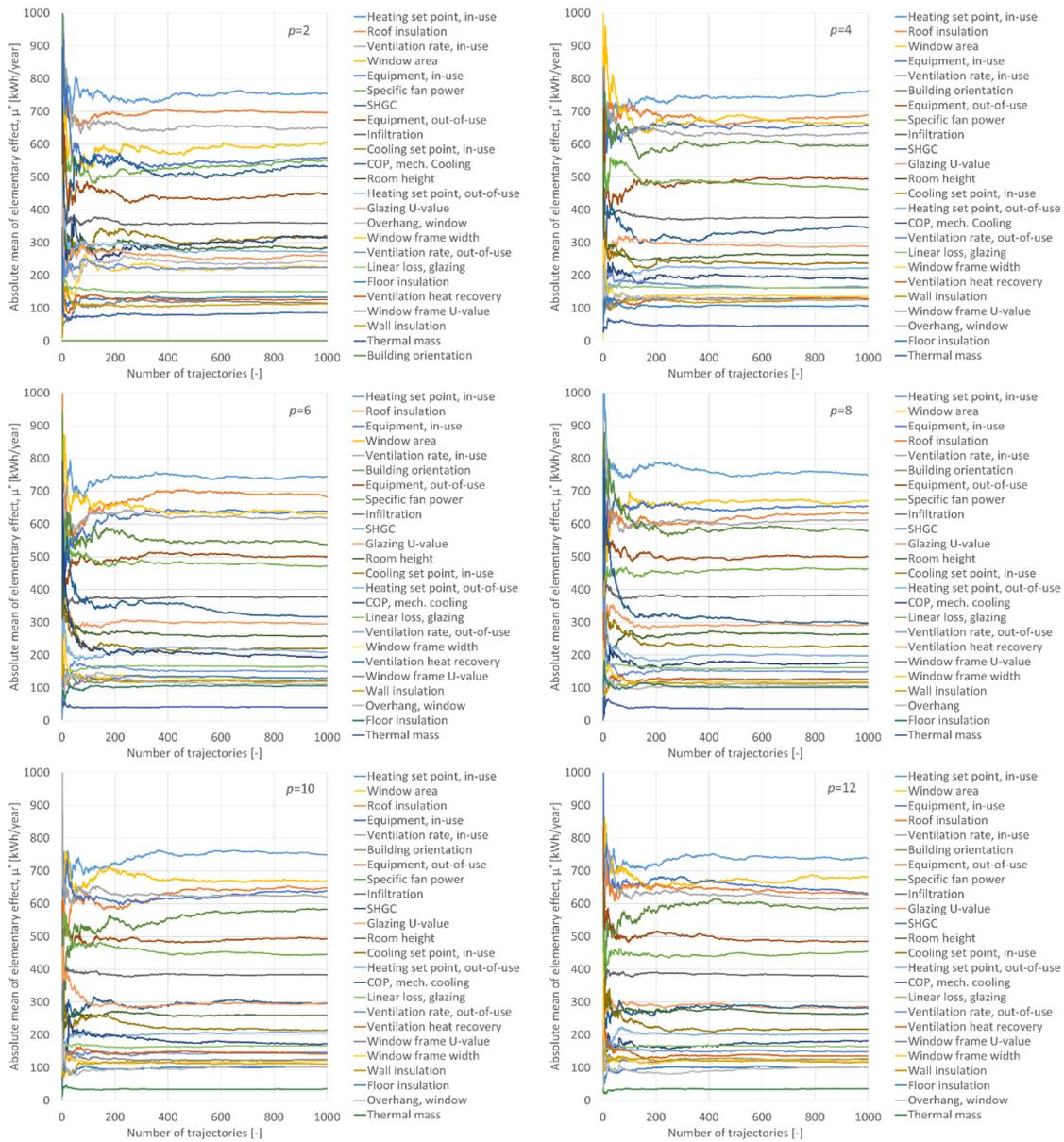


Fig. B.1. The evolution of mean elementary effect (μ_i^*) on the total energy need (heating + cooling + ventilation) for all measures (x_i) as a function of the number of trajectories (r) for six different levels (p). The order of the legend corresponds to the order of the lines in the graph.

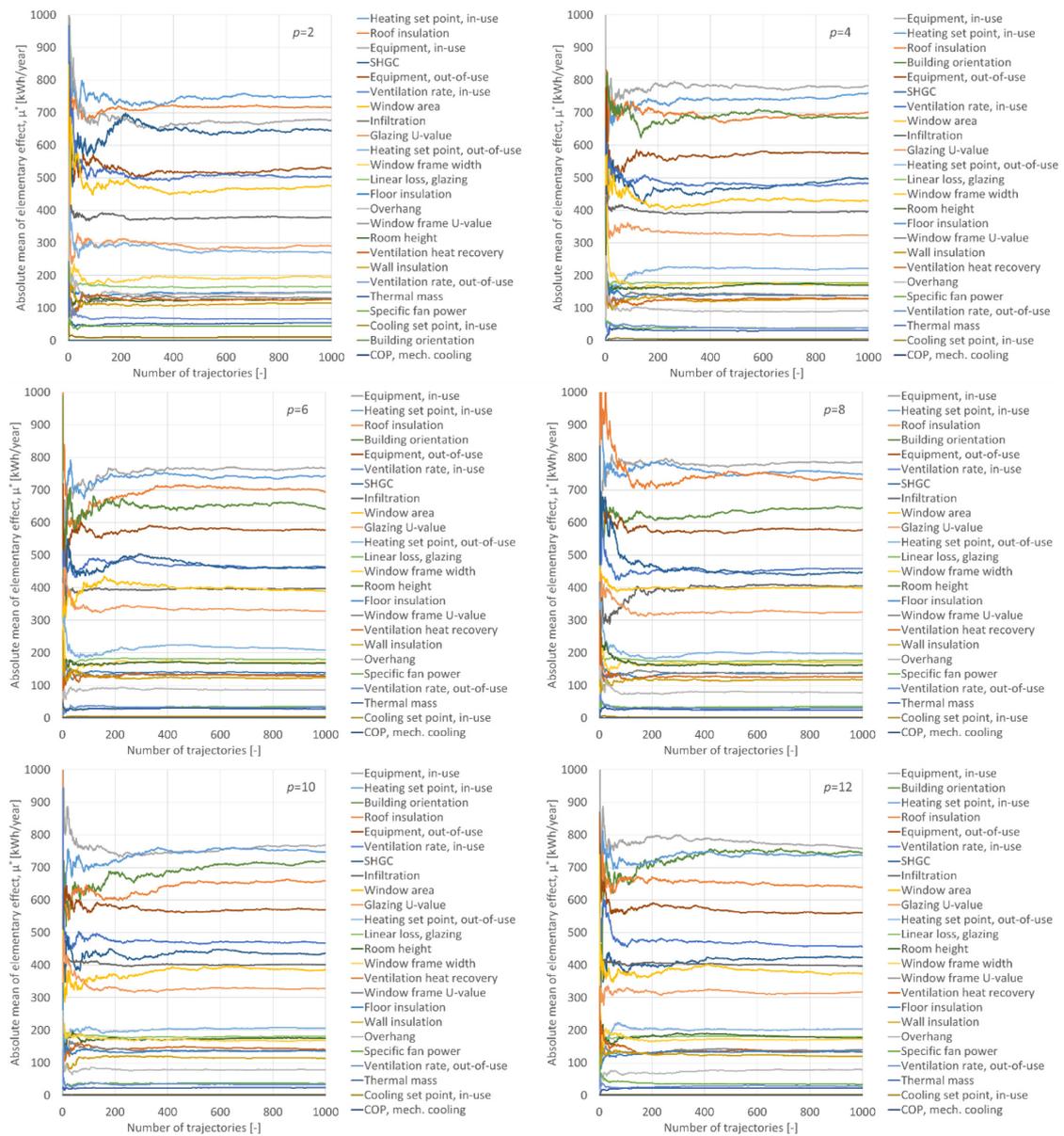


Fig. B.2. The evolution of mean elementary effect (μ^*) on the heating energy need for all measures (x_i) as a function of the number of trajectories (r) for six different levels (p). The order of the legend corresponds to the order of the lines in the graph.

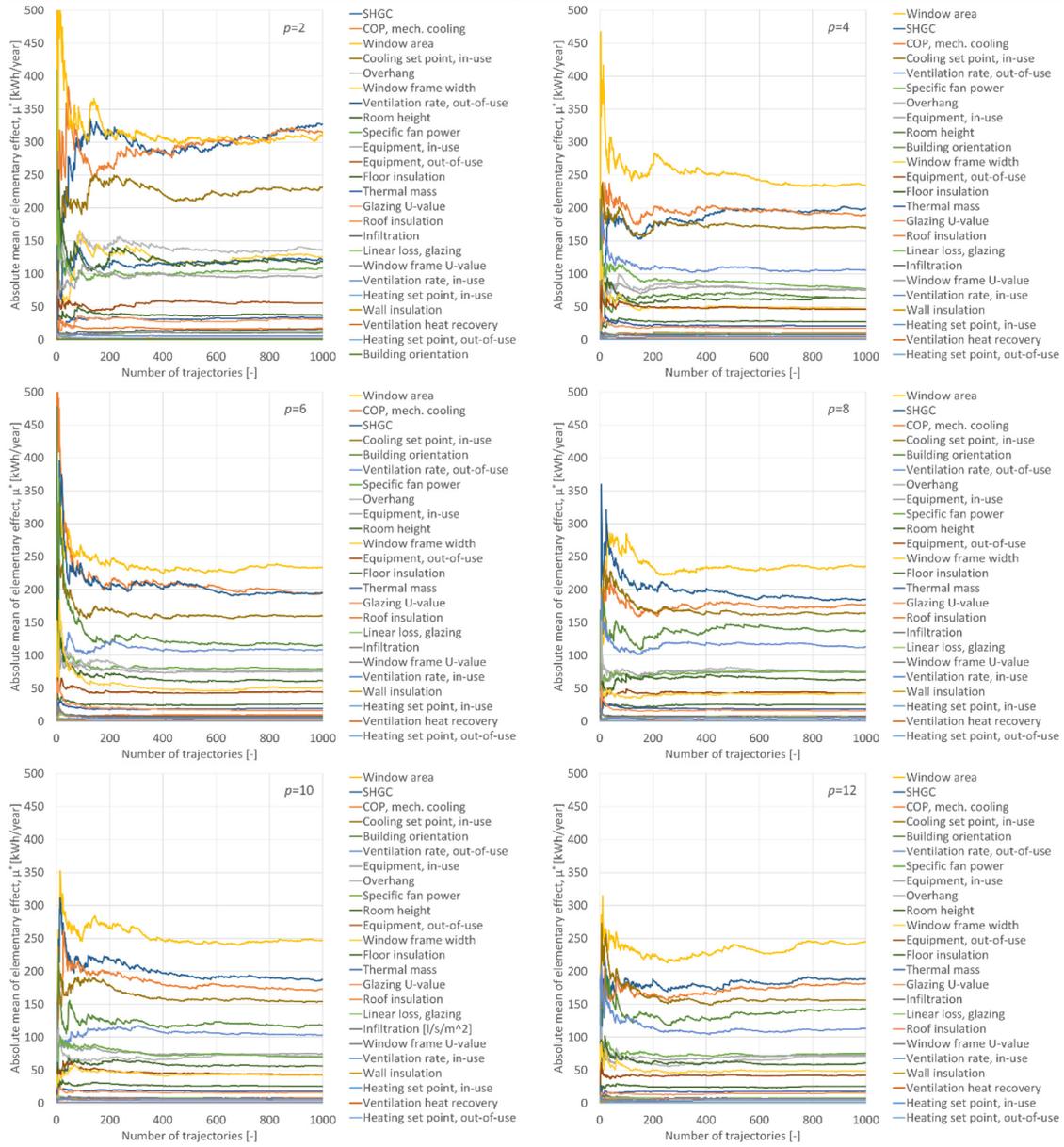


Fig. B.3. The evolution of mean elementary effect (μ^*) on the cooling energy need for all measures (x_i) as a function of the number of trajectories (r) for six different levels (p). The order of the legend corresponds to the order of the lines in the graph.

Table C.1Ranked sensitivity indexes (SI_i) for all input parameters calculated OAT based on the partial derivative of the model output.

Total energy Input parameter	SI_i	Heating energy only Input parameter	SI_i	Cooling energy only Input parameter	SI_i
Window area	1.064	Window area	0.932	Window area	4.127
Roof insulation	0.622	Roof insulation	0.858	Cooling set point, in-use	1.607
Heating set point, in-use	0.553	Heating set point, in-use	0.765	SHGC	1.364
Equipment, out-of-use	0.388	Equipment, out-of-use	0.614	COP, mech. cooling	1.333
Infiltration	0.300	Infiltration	0.428	Ventilation rate, out-of-use	1.333
Specific fan power	0.290	Building orientation	0.413	Overhang	0.628
Building orientation	0.262	Equipment, in-use	0.375	Specific fan power	0.597
Ventilation rate, in-use	0.244	Ventilation rate, in-use	0.238	Equipment, in-use	0.488
Equipment, in-use	0.216	SHGC	0.238	Equipment, out-of-use	0.483
Wall insulation	0.160	Linear loss, glazing	0.226	Building orientation	0.324
Linear loss, glazing	0.152	Wall insulation	0.224	Window frame width	0.315
Cooling set point, in-use	0.136	Glazing U-value	0.191	Floor insulation	0.299
Glazing U-value	0.128	Floor insulation	0.180	Thermal mass	0.228
Floor insulation	0.096	Ventilation heat recovery	0.127	Roof insulation	0.129
Ventilation heat recovery	0.092	Room height	0.119	Linear loss, glazing	0.080
COP, mech. cooling	0.085	Heating set point, out-of-use	0.116	Glazing U-value	0.071
Heating set point, out-of-use	0.084	Window frame width	0.096	Infiltration	0.059
Room height	0.079	Window frame U-value	0.080	Ventilation rate, in-use	0.048
Window frame U-value	0.054	Overhang	0.045	Room height	0.046
Ventilation rate, out-of-use	0.036	Thermal mass	0.026	Window frame U-value	0.030
Window frame width	0.035	Specific fan power	0.020	Wall insulation	0.019
Overhang	0.031	Ventilation rate, out-of-use	0.009	Ventilation heat recovery	0.006
SHGC	0.028	Cooling set point, in-use	0.000	Heating set point, in-use	0.003
Thermal mass	0.002	COP, mech. cooling	0.000	Heating set point, out-of-use	0.000

Appendix C. Local method

Table C.1 lists all input parameters ranked according to the sensitivity index, SI_i , calculated OAT based on the partial derivative of the model output for total energy use (heating + cooling + ventilation), heating only, and cooling only.

References

- [1] E. Borgonovo, E. Plischke, Sensitivity analysis: A review of recent advances, *Eur. J. Op. Res.* 248 (2016) 869–887.
- [2] A. Saltelli, S. Tarantola, F. Campolongo, Sensitivity analysis as an ingredient of modeling, *Stat. Sci.* 15 (4) (2000) 377–395.
- [3] M.D. Morris, Factorial sampling plans for preliminary computational experiments, *Technometrics* 33 (2) (1991) 161–174.
- [4] F. Campolongo, R. Braddock, The use of graph theory in the sensitivity analysis of the model output: A second order screening method, *Reliab. Eng. Syst. Safety* 64 (1999) 1–12.
- [5] F. Campolongo, J. Cariboni, A. Saltelli, An effective screening design for sensitivity analysis of large models, *Environ. Model. Softw.* 22 (2007) 1509–1518.
- [6] W. Tian, A review of sensitivity analysis methods in building energy analysis, *Renew. Sust. Energy Rev.* 20 (2013) 411–419.
- [7] M.H. Kristensen, S. Petersen, Choosing the appropriate sensitivity analysis method for building energy model-based investigations, *Energy Build.* 130 (2016) 166–176.
- [8] I.M. Sobol', Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comput. Exp.* 1 (4) (1993) 407–414.
- [9] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity Analysis in practice: A guide to Assessing Scientific Models, John Wiley & Sons Ltd, West Sussex, England, 2004.
- [10] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seirne, S. Thomas, Application of sensitivity analysis in design of sustainable buildings, *Renew. Energy* 34 (2009) 2030–2036.
- [11] K. Menberg, Y. Heo, R. Choudhary, Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information, *Energy Build.* 133 (2016) 433–445.
- [12] T. Østergaard, S.E. Maagaard, R.L. Jensen, Thermal comfort in residential buildings by the millions – early design support from stochastic simulations, in: Proceedings of CLIMA 2016, the 12th REHVA World Congress: volume 6, Aalborg, Denmark, 2016.
- [13] A.T. Nguyen, S. Reiter, A performance comparison of sensitivity analysis methods for building energy models, *Build. Simu.* 8 (2015) 651–664.
- [14] F. Campolongo, A. Saltelli, Sensitivity analysis of an environmental model; an application of different analysis methods, *Reliab. Eng. Syst. Safety* 57 (1) (1997) 49–69.
- [15] F. Campolongo, S. Tarantola, A. Saltelli, Tackling quantitatively large dimensionality problems, *Comput. Phys. Commun.* 117 (1999) 75–85.
- [16] Y.-J. Kim, S.-H. Yoon, C.-S. Park, Stochastic comparison between simplified energy calculation and dynamic simulation, *Energy Build.* 64 (2013) 332–342.
- [17] S. de Wit, G. Augenbroe, Analysis of uncertainty in building design evaluations and its implications, *Energy Build.* 34 (2002) 951–958.
- [18] V. Corrado, H.E. Mechri, Uncertainty and sensitivity analysis for building energy rating, *J. Build. Phys.* 33 (2) (2009) 125–156.
- [19] D. Garcia Sanchez, B. Lacarri re, M. Musy, B. Bourges, Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods, *Energy Build.* 68 (2014) 741–750.
- [20] T.L. Hemsath, K.A. Bandhosseini, Sensitivity analysis evaluating basic building geometry's effect on energy use, *Renew. Energy* 76 (2015) 526–538.
- [21] T. Østergaard, S.E. Maagaard, R.L. Jensen, A stochastic and holistic method to support decision-making in early building design, in: Proceedings of BS 2015, the 14th International Conference of the International Building Performance Simulation Association, Hyderabad, India, 2015.
- [22] Z. Yang, B. Becerik-Gerber, A model calibration framework for simultaneous multi-level building energy simulation, *Appl. Energy* 149 (2015) 415–431.
- [23] M. de Witt, Identification of the important parameters in thermal building simulation models, *J. Stat. Comput. Simu.* 57 (1997) 305–320.
- [24] H. Brohus, P. Heiselberg, A. Hesselholt, H. Rasmussen, Application of partial safety factors in building energy performance assessment, Eleventh International IBPSA Conference, 2009.
- [25] A.T. Booth, R. Choudhary, D.J. Spiegelhalter, Handling uncertainty in housing stock models, *Build. Environ.* 48 (2012) 35.
- [26] J. Le Drau, P. Heiselberg, Sensitivity analysis of the thermal performance of radiant and convective terminals for cooling buildings, *Energy Build.* 82 (2014) 482–491.
- [27] S. Yang, W. Tian, E. Cubi, Q.-X. Meng, Y.-L. Liu, L. Wei, Comparison of sensitivity analysis methods in building energy assessment, *Procedia Eng.* 146 (2016) 174–181.
- [28] T. Østergaard, S.E. Maagaard, R.L. Jensen, Thermal comfort in residential buildings by the millions – early design support from stochastic simulations, in: Proceedings of CLIMA 2016, the 12th REHVA World Congress: volume 6, Aalborg, Denmark, 2016.
- [29] G.A. Faggianelli, L. Mora, R. Merheb, Uncertainty quantification for energy savings performance contracting: Application to an office building, *Energy Build.* 152 (2017) 61–72.
- [30] M.H. Kristensen, S. Petersen, Contrasting the capabilities of three different sensitivity analysis methods for building energy model-based investigations, in: Proceedings of BSO 2018, 4th Building Simulation and Optimization Conference, Cambridge, UK, 2018.
- [31] JRC, Joint Research Centre of the European Commission, Ispra, Italy, 2004 <http://sensitivity-analysis.jrc.ec.eu.int/>.
- [32] U.S. Department of Energy, EnergyPlus 8.6.0, [Online].
- [33] J.C. Lam, S.C.M. Hui, Sensitivity analysis of energy performance of office buildings, *Build. Environ.* 31 (1) (1996) 27–39.
- [34] P. Lyons, J. Wong, M. Bhandari, A comparison of window modeling methods in EnergyPlus 4.0, in: Proceedings of the 4th National Conference of IBPSA-USA, New York City, USA, 2010.
- [35] EN ISO 10077-1, Thermal performance of windows, doors and shutters – calculation of thermal transmittance – part 1, Simplified method, European Committee for Standardization, 2000.

- [36] DS 418, Calculation of Heat Loss from Buildings, Danish Standard, Copenhagen, Denmark, 2002.
- [37] P.G. Wang, M. Scharling, K.P. Nielsen, K.B. Wittchen, K. Kern-Hansen, Danish Design Reference Year – Reference Climate Dataset for Technical Dimensioning in Building, Construction and other Sectors (Technical report 13-19), Danish Meteorological Institute, Copenhagen, Denmark, 2013.
- [38] J.K. Ravalico, H.R. Maier, G.C. Dandy, J.P. Norton, B.F.W. Croke, A comparison of sensitivity analysis techniques for complex models for environmental management, in: *Proceeding of MODSIM05: International Congress on Modelling and Simulation Advances and Applications for Management and Decision Making Proceedings*, Melbourne, Australia, 2005.

Contrasting the Capabilities of Three Different Sensitivity Analysis Methods for Building Energy Model-based Investigations

Martin Heine Kristensen, Steffen Petersen
Department of Engineering, Aarhus University, Aarhus, Denmark

Abstract

When using building energy models (BEM) for building design, it is often valuable to conduct a sensitivity analysis (SA) to help designers to focus their efforts on design variables that drive the majority of the building performance indicators such as energy use.

In this study, three different SA methods (Local, Morris and Sobol') were applied to two different BEMs (hourly dynamic and monthly quasi-steady-state) for SA analysis in two different stages of the building design process. The finding is that the choice of appropriate SA method depends on the purpose of the SA; whether it is a screening of equally probable design options, or a more thorough quantification and ranking of parameter uncertainty.

Introduction

Sensitivity analysis (SA) can be used to explore the behaviour of building energy models (BEM) and thereby identify which input parameters that drive the majority of the model output variation. Such an analysis is valuable as it enables building designers and contractors to focus their efforts on designing and obtaining the functional requirements of parameters most critical to the energy performance.

There are many examples in the literature on how to apply SA for BEM-based design, e.g. Heiselberg et al. (2009), Mechri et al. (2010), Spitz et al. (2012), and Østergaard et al. (2015) to mention a few. However, a sound argumentation that vouches for the reliability, validity and necessary complexity of the chosen SA is rare. Kristensen and Petersen (2016) used a model of an existing residential building stock in a temperate climate (i.e. energy need was predominantly space heating) as case to demonstrate that the choice of SA method affects the identification and ranking of the input parameters most sensitive to the model output. The overall conclusion was that it is essential not to interpret the outcome of SA in a way that lies beyond the capabilities of the used SA method, as this may lead to suboptimal design decisions and wrong focus areas in the construction phase. Furthermore, the study also showed that the SA outcome – to some extent – is affected by the chosen BEM method; in this case, the simple hourly and monthly methods in ISO 13790:2008. Practitioners must therefore be careful to choose an appropriate combination of SA method and BEM that fits the purpose of the SA.

In practice, SA can be used for various purposes. In the early design stage, SA can help designer to identify which critical design variables to focus on. Prior to initiating the construction phase, SA can be used to identify which functional requirement to have special focus on obtaining during construction. The objective of this paper is to investigate the performance of three different SA methods combined with two different BEMs for the two above-mentioned SA purposes using an office building in a temperate climate as case. The intention is to provide an example to guide building designers in selecting the appropriate SA method depending on the purpose of the analysis and the type of BEM applied.

Method

An office case building was modelled using two different BEMs to calculate the annual energy need for space heating and cooling. Using three different SA methods of increasing capability and complexity, the sensitivity of the two BEMs was investigated and analysed for two different phases in the design of the case building;

1. the early design phase (Case 1) with uncertainty embedded in the free choice of model parameter values, and
2. the detailed design phase (Case 2) where uncertainty is embedded in the fixed parameter values chosen in the early design stage due to e.g. imperfections in building materials, construction errors, and stochastic occupant behaviour.

Description of case

The case consisted of a 24 m² (6x4 m) south-facing two-person office room in a single-story building (Figure 1); the window façade and roof faced the outdoor while the floor faced the ground. The remaining surfaces were assumed adiabatic.

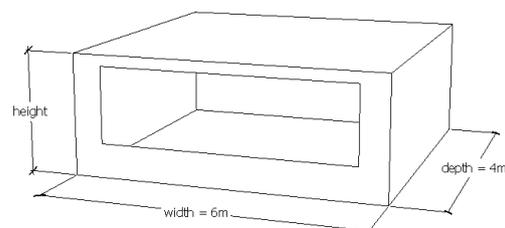


Figure 1: Case office room.

The room was assumed occupied weekdays between 8 am and 5 pm in which period a constant-air-volume system ventilated the room. Heating and cooling was supplied by separate systems. During the weekend, ventilation was turned off and no internal heat loads were assumed. The installed capacity of the heating and cooling systems was assumed able to always meet their respective set points. Internal heat loads, airflow rates and set points are given in Table 1.

In Case 1 (early design phase), seven model parameters were open for design decisions: room height, U-values of external wall, roof and floor, %-window area of the external wall area, and the SHGC of the window. These seven unselected model parameters were all assigned a uniformly distributed probability range (Table 1) to reflect the uncertainty faced by building designers in the early design stage. The purpose of this analysis was to help building designers focus on the model parameters of most influence to the model output (energy need).

In Case 2 (uncertainty related to the realisation of values chosen in the early design stage), all parameter values were assumed fixed at satisfying levels by the building designers. However, due to e.g. imperfections in building materials, construction errors, and stochastic occupant behaviour, these selected values are subject to uncertainty. In Case 2, the sensitivity of the model output to this kind of uncertainty was investigated. This information can be regarded as advice on which parameters that demands special attention in the detailed design and construction phase to reduce the risk of not realising the design intentions.

Building energy modelling

The two BEMs used to calculate the annual energy need for space heating and cooling were 1) the quasi-steady-state calculation method with monthly time steps, and 2) the simple dynamic calculation method based on hourly time steps. Both model are described in ISO 13790:2008.

Only energy need for space heating and cooling were considered; thus, COP coefficients of chillers etc. were not considered. Weather conditions were modelled using the Danish design reference year (DRY) dataset, containing hourly values of the necessary weather parameters (air temperature, normal solar radiation and diffuse solar radiation) (Jensen and Lund, 1995). Total solar radiation perpendicular to the building facades was calculated in both BEMs using the solar algorithm described by Bourges (1992).

Sensitivity analysis methods

Three SA methods were applied: a local partial derivative-based method (Lam and Hui, 1996), the global screening-based method of Morris (Morris, 1991), and the global variance-based method of Sobol' (Sobol', 1993). All three methods and how they were implemented for the analysis in this paper are described in detail in Kristensen and Petersen (2016). Information about the setup of the SA methods is given in Table 2.

Overall, the capabilities of the three SA methods can be contrasted in terms of their ability to take into account

1. the range and shape of input parameter distributions,
2. multi-dimensional parameter influence on the outcome when all input parameters are varied simultaneously, and
3. non-linear and non-additive effects when input parameter interactions are taken into account (model independency).

The only SA method featured in this paper that encompasses all three of the above-mentioned abilities is the global variance-based method of Sobol'. The Sobol' method makes a complete decomposition of the output variance by searching across the entire input space, simultaneously taking into account range and shape of parameter distributions and correlated effects.

Table 1: Probability density functions assigned to model input parameters for Case 1 and Case 2.

Input parameters	Unit	Case 1	Case 2
Room width	[m]	6.0	6.0
Room depth	[m]	4.0	4.0
Room height	[m]	Uniform (3.0;4.0)	3.5
U-value (ext. wall)	[W/m ² K]	Uniform (0.10;0.30)	Lognormal (-1.966;0.020 ²)
U-value (roof)	[W/m ² K]	Uniform (0.08;0.20)	Lognormal (-2.303;0.039 ²)
U-value (floor)	[W/m ² K]	Uniform (0.10;0.20)	Lognormal (-2.121;0.027 ²)
Adjustment factor (ground)	[-]	0.7	Beta (15;3.75)
Window-%	[-]	Uniform (30%; 60%)	40%
Window frame fraction	[-]	20%	20%
U-value (window)	[W/m ² K]	0.8	0.8
SHGC	[-]	Uniform (0.3;0.6)	0.3
Ventilation rate (CAV)	[l/s/m ²]	0.83	Uniform (0.75;0.92)
Infiltration rate @ 50Pa	[l/s/m ²]	0.50	Lognormal (-0.240; 0.693 ²)
Heat recovery efficiency	[-]	0.85	Beta (100;21.95)
Internal heat loads (people, light, appliances)	[W/m ²]	17	Lognormal (6.073; 0.269 ²)
Internal heat capacity	[KJ/m ² K]	Uniform (110;260)	Lognormal (5.102; 0.088 ²)
Heating set point	[°C]	20	Normal (21.5;1 ²)
Cooling set point	[°C]	26	Normal (26;0.4 ²)

Table 2: Input assumptions for the SA methods.

	Local	Morris	Sobol'
Points used from PDF	1% and 99% quantiles	1% and 99% quantiles, $p = 4$ (levels)	Entire PDF
Sample size	1	$r = 300$ (trajectories)	$N = 10,000$ (LHS)
No. model evaluations	$2k+1$	$r(k+1)$	$N(k+2)$
Convergence measure	N/A	$\Sigma(\mu) \approx$ constant	$\Sigma(ST) \approx$ constant

The resulting total-order sensitivity indices are bounded to sum to one which makes it physically meaningful to use them for identifying and ranking input parameters that drive the majority of the model output variation.

The method of Morris applies the absolute mean of a population of local elementary effects to quantify the global influence of a given input parameter. The method is to some extent able to take into account non-linear and non-additive effects (ability 3), but is not able to account for non-uniform distributions of model input parameters (ability 1) using the traditional Morris sampling technique (factorial sampling) applied in this study. Furthermore, the Morris method is potentially neglecting correlated effects (ability 2) because each parameter is varied locally one-at-a-time (OAT). Another disadvantage of the Morris method is the dubious interpretation of the mean elementary effect as a measure of global sensitivity. One should be careful with interpreting a large absolute value of the mean elementary effect from the Morris method as a sign of great parameter influence as such values vary from one model to another. Only the internal ranking of the means can be used to quantify the influence of the parameters and sort them in clusters of importance.

The Local method uses inputs and outputs from OAT parameter variations (one sample) to calculate a dimensionless sensitivity index expressing the elasticity of variation around the mean value as percentage change in output per percentage change in input. This SI-index is then used for identifying and ranking the input parameters most sensitive to the output. The Local method implies a strictly linear model (not fulfilling ability 3), it does not allow any quantification of correlated parameter effects (not fulfilling ability 2), and does not allow any utilisation of knowledge about the shape of the parameter distributions (not fulfilling ability 1). One should therefore be careful interpreting the identification and ranking if the model is not 100% linear, have interacting input parameters, and anything but uniformly distributed input parameters.

Model input parameters

In order to carry out the sensitivity analyses, input spaces had to be specified for the uncertain input parameters in

both cases to reflect the a-priori uncertainty of their value. To do so, different continuous probability density functions (PDFs) were applied to set the probability of a given parameter value over a range of variation (Figure 2).

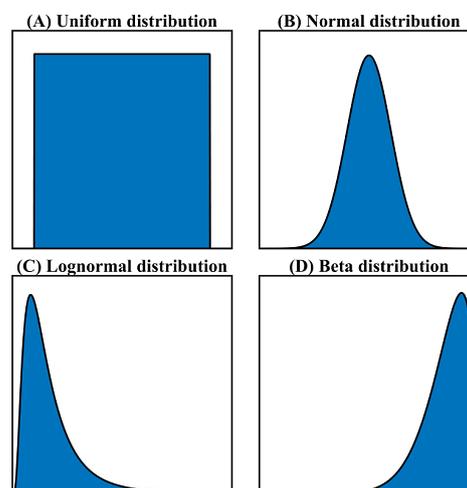


Figure 2: Probability density functions suitable for sensitivity analysis of building energy models.

The uncertainty of a parameter can be uniformly distributed across its defined range of variation $Uniform(A;B)$; doing so, the probability of all values within the parameter range are equal. This was appropriate for Case 1, as the purpose was to explore the effects of equally possible design options in the early design stage prior to any design decisions. The uniform PDF is often referred to as a non-informative PDF as no information can be extracted from it besides the range of variation.

A parameter can also be non-uniformly distributed if a-priori information allows it, e.g. expert judgements, historical data, or measurement error specifications. Such distributions were appropriate for Case 2, as the purpose was to explore the effects of uncertainty of the true value of an already decided parameter, i.e. error related to the practical implementation of the design. For this end, the normal distribution $Normal(\mu; \sigma^2)$ was applied to input parameters with an equally probable chance of variation around a most probable mean value (e.g. set point temperatures). The lognormal distribution $Lognormal(\mu; \sigma^2)$ was applied to positively defined parameters where higher values were more probable than lower values (e.g. U-values and infiltration rate). The beta distribution $Beta(a; b)$ was applied to specify factors defined between 0 and 1 (e.g. heat recovery efficiency). The lognormal distribution is always skewed to the right (positive skew; right-tailed), but the shape will imitate the normal distribution for distributions with large variance. The beta distribution may assume almost any shape and skewness; thus, it is likewise possible to make it imitate the normal distribution if wanted (Figure 2).

Results

Partitioning heating and cooling need

A monthly partitioning of the energy need for space heating and cooling, calculated using the two BEMs respectively, is shown in Figure 3, applying the mean values of the inputs for Case 1 (Table 1).

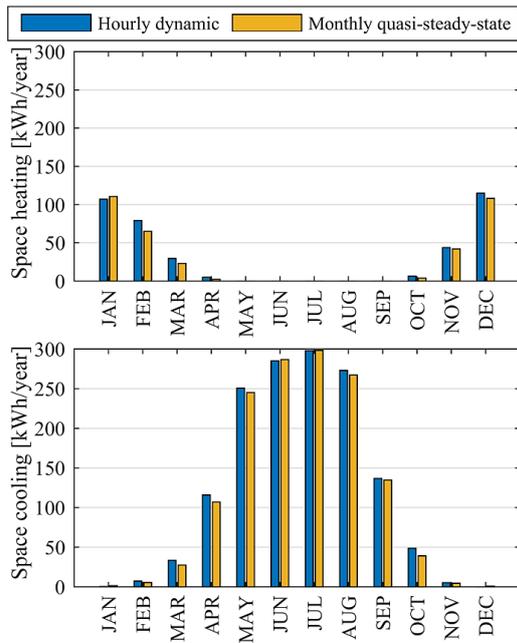


Figure 3: Monthly partitioning of energy need for space heating and cooling. Mean input values for Case 1 was applied.

The annually aggregated energy need for space heating and cooling of the office room is dominated by the need for cooling (heating share of 20%; cooling share of 80%). The deviation between the annually aggregated energy need for space heating and cooling of the hourly dynamic and the monthly quasi-steady-state BEMs is 4% (8% for heating; 2% for cooling). A larger internal deviation is present on the monthly scale.

Case 1: Early design decisions

Given the uniformly defined uncertainty specification for Case 1 (Table 1), the probable outcome of the annual energy need for space heating and cooling of the office room is shown in Figure 4A. Even though the BEMs were not calibrated prior to simulation, and thus were not expected to be consistent, their output distributions exhibit the same variation and shape. Their mean values are 1,890 kWh/year (hourly dynamic) and 1,810 kWh/year (monthly quasi-steady-state), respectively.

The amount of variation caused by each of the seven uncertain parameters is quantified by the three SA methods and depicted in Figure 5, and ranked in order of influence in Table 3.

Table 3: Ranking of input parameters for Case 1 (1 is most influential; 7 is least influential). L = local method; M = Morris method; S = Sobol' method.

Input	Hourly dynamic			Monthly quasi-steady-state		
	L	M	S	L	M	S
Window-%	3	1	1	2	1	1
SHGC	2	2	2	3	2	2
Room height	1	3	3	1	3	3
Internal heat capacity	4	4	4	4	4	4
U-value (roof)	5	5	5	5	5	5
U-value (ext. wall)	7	6	6	6	6	6
U-value (floor)	6	7	7	7	7	7

The parameter ranking based on the Morris and Sobol' analysis is identical for both BEMs, whereas the result of the Local method deviates a bit. The Morris and Sobol' methods identified the window-% as the single most influential parameter (approx. 45%-51% of the output variability in the two BEMs, respectively, can be ascribed the window-% cf. the Sobol' analysis), then SHGC as 2nd and room height as 3rd most influential. The Local method, on the other hand, identified room height to be most influential in both BEMs with the SHGC coming in as 2nd and the window-% as 3rd most influential.

In both BEMs, all three SA methods find the U-values (ext. wall, floor and roof) to be the least influential parameters given the input distributions of Case 1.

Case 2: Uncertainty of practical implementation

Given the mixed uncertainty specification for Case 2 (Table 1), the probable outcome of the annual energy need for space heating of the office room is shown in Figure 4B. The mean values are 1,410 kWh/year (hourly dynamic) and 1,300 kWh/year (monthly quasi-steady-state), respectively. In contrast to the output distribution of Case 1, the distribution of Case 2 has a lower variance (uncertainty).

The amount of variation caused by each of the 11 uncertain parameters is quantified by the three SA methods and shown in Figure 5, and ranked in order of influence in Table 4.

Table 4: Ranking of input parameters for Case 2 (1 is most influential; 11 is least influential). L = local method; M = Morris method; S = Sobol' method.

Input	Hourly dynamic			Monthly quasi-steady-state		
	L	M	S	L	M	S
Internal heat loads	3	2	1	4	2	1
Heating set point	2	1	2	2	1	3
Infiltration rate	5	4	3	6	3	2
Cooling set point	1	3	4	1	4	4
Internal heat capacity	4	5	5	3	5	5
Heat recovery efficiency	10	6	6	5	7	7
Adj. factor (ground)	8	7	7	9	6	6
U-value (roof)	6	8	8	7	8	8
Ventilation rate	11	9	9	11	10	9
U-value (floor)	7	10	10	8	9	10
U-value (ext. wall)	9	11	11	10	11	11

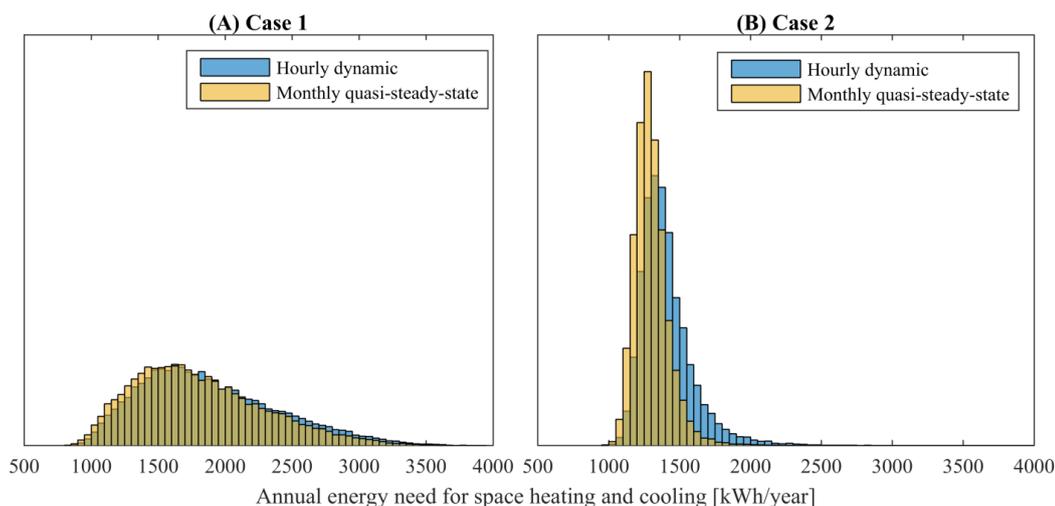


Figure 4: Probability distributions of annual energy need for space heating and cooling for Case 1 and Case 2. Latin Hypercube sampling was used to generate 10,000 simulation runs.

In general, the rankings are different for all three methods; however, the Morris and Sobol' methods tend to agree more with each other than with the Local method. Moreover, as opposed to the parameter ranking of Case 1, larger differences in the ranking are seen between the two BEMs for the same SA method.

According to the variance decomposition of the Sobol' method, the highest ranked parameter is the internal heat loads (people, lighting and appliances) in both BEMs, which accounts for 44%-47% of the output variability in the two BEMs, respectively. In total, the top-3 ranked parameters (internal heat loads, heating set point and infiltration rate) make up approx. 90%-92% of the output variability in both BEMs.

The Morris method identified the same top-3 for the quasi-steady-state BEM as the Sobol' method, but in different order. For the hourly dynamic BEM, the Morris method only identified top-2 from the Sobol' method, also in a different order. The 3rd most influential parameter was found to be the cooling set point instead of infiltration rate.

The Local method deviates from the Sobol' method by identifying the cooling set point as the most influential parameter in both BEMs. The heating set point is 2nd highest ranked in both BEMs while internal heat loads is only in top-3 for the hourly dynamic BEM; the quasi-steady-state BEM has instead the internal heat capacity in top-3. The infiltration rate (which was in top-3 for the Sobol' method in both BEMs) was on 5th and 6th place in the hourly and quasi-steady-state BEM, respectively.

Discussion

For Case 1, i.e. an early design phase with uniformly distributed probabilities, the differences in parameter ranking based on the different SA methods are very limited. The Morris and Sobol' methods tend to agree to

a large extent. Assuming that the Sobol' method is correct, the Local method is in principle leading to a wrong identification of the most important input parameter; however, Figure 5 shows that this is only due to a marginal difference between the top-3 parameters. This is a good example of how the visual presentation of SA results might be more informative to building designers in an early design phase than discrete rankings, as the focus seems to be on parameter screening rather than precise uncertainty quantification.

For Case 2, i.e. the detailed design stage with the option of non-uniform PDFs to represent the modeller's information about the parameter uncertainty, the Sobol' and Morris method by definition outperforms the Local method in terms of parameter ranking. The Local method does not respect the range and shape of the input distributions (ability 1); it assumes the same effect for all possible values of the parameters, which is why it identifies the cooling set point as very important. In reality, the plausible range of the cooling set point is very limited, and thus not that influential in the overall picture. The Morris method does respect the range of the distributions, but not the shape as in the Sobol' method, which makes the result of the Morris method approach the result from the Sobol' method somewhat better than the Local method. From a theoretical point of view, the Sobol' method would thus a-priori be regarded the most appropriate method for Case 2, as it has the ability to take into account the varying and somewhat skewed input distributions that were applied. Nonetheless, the Morris method showed to be appropriate for identifying the unranked cluster of the top-3 most important parameters, which together account for approx. 90% of the uncertainty in the energy need. Thus, if the purpose of the sensitivity analysis is to identify a group of most

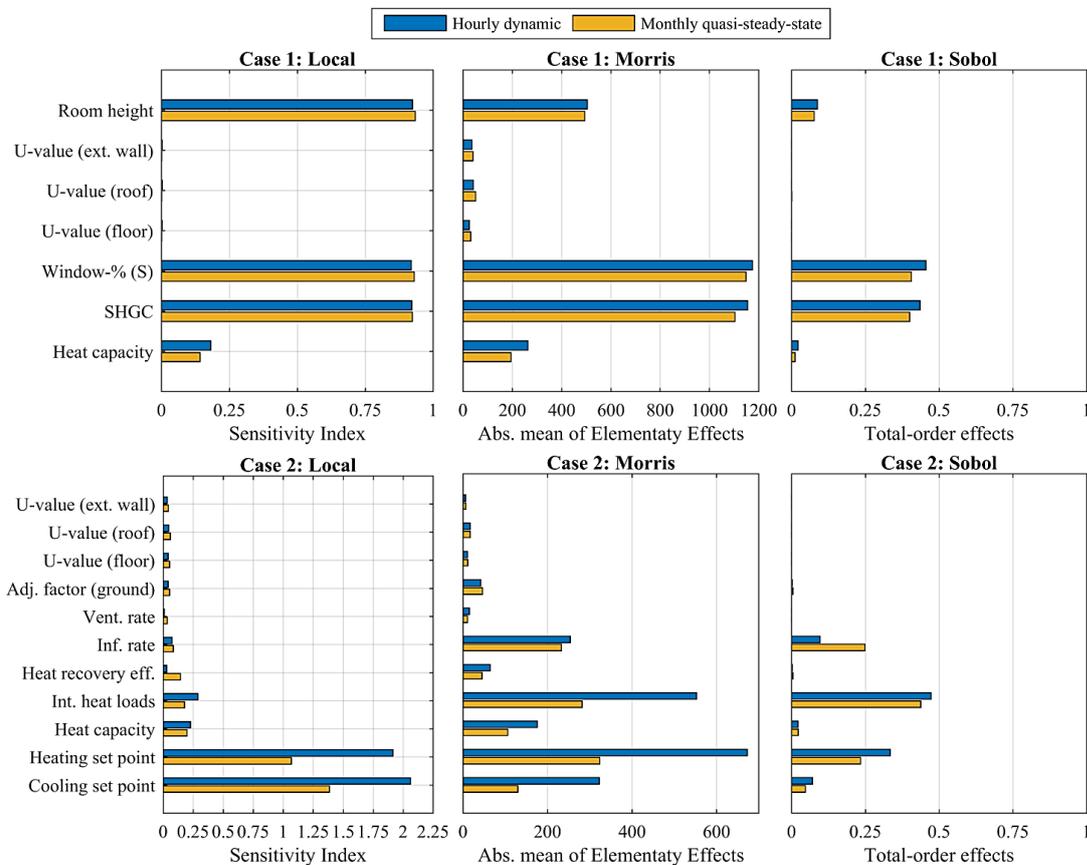


Figure 5: Result of sensitivity analysis of Case 1 and Case 2, using the Local, Morris and Sobol' method, respectively.

important parameters, one might apply the computationally efficient Morris method in favour of the by far more complex and computationally heavy Sobol' method. However, if the purpose is a detailed quantification of the uncertainty contribution of each parameter, and how these interact and affect each other, there is no way around a variance-based method like the one by Sobol'. For practical reasons though, the application of the Sobol' method seems infeasible in ordinary building design. The Local method fell short by only identifying one and two out of top-3 most important parameters, for the monthly quasi-steady-state and hourly dynamic BEMs, respectively; hence, it is most suitable for the simple screening-based analyses in the earliest stages of a building design process.

In general, uncertainty in the inputs was propagated similarly through the two BEMs resulting in approx. equal output distributions. However, some differences in the influence of the inputs were found due differences in model behaviour and dynamics. Thus, one should be careful interpreting the outcome of a sensitivity analysis of one BEM as generally applicable for all BEMs; besides the applied SA method itself, the outcome of a sensitivity analysis is influenced by model behaviour, the exact

selection of input parameters, and their investigated distributions (range and shape).

Conclusion

From the results of this study, it is evident that the applicability of the different SA methods used depends on the purpose of the SA. If the purpose is to identify which parameters – all with uniformly distributed probability – affects the model output, then a simple Local method seems to suffice from a practical point of view. This is especially true if an exact ranking of parameters is of minor importance, and if the building physics is represented using linear equations. However, the Local method can only be used to identify an unranked cluster of maybe the upper half most important parameters if the probability distributions of the input parameters for some reason are non-uniformly distributed. In such cases, the Morris method is preferred as long as the probability distributions are well defined without any long tails; if the majority of the parameters are normal or lognormal distributed with large variance and/or beta distributed, then the Sobol' method is preferred.

In addition to the choice of SA method itself, the results indicate that the ranking of important input parameters – and thus the proper selection of SA method – is influenced

by the applied BEM. This aspect of SA performance is relatively unexplored and ought to be further investigated in future work.

References

- Bourges, B. (1992). Climatic Data Handbook for Europe: Climatic Data for the Design of Solar Energy Systems, 1st Edition, Springer Netherlands, ISBN 978-0-7923-1716-6.
- Heiselberg, P., H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinre, and S. Thomas (2009). Application of sensitivity analysis in design of sustainable buildings. *Renewable Energy* 34, 2030–2036.
- ISO 13790:2008 (2008). Energy performance of buildings – Calculation of energy use for space heating and cooling, 2nd Edition.
- Jensen, J.M. and H. Lund (1995). Design Reference Year, DRY - Et nyt dansk referencer (in Danish), Technical University of Denmark, announcement number 281.
- Kristensen, M.H. and S. Petersen (2016). Choosing the appropriate sensitivity analysis method for building energy model-based investigations. *Energy and Buildings* 130, 166–176.
- Lam, J.C. and S.C.M. Hui (1996). Sensitivity analysis of energy performance of office buildings. *Building and Environment* 31(1), 27–39.
- Mechri, H. E., A. Capozzoli and V. Corrado (2010). Use of the anova approach for sensitive building energy design, *Applied Energy* 87, 3073–3083.
- Morris, M.D. (1991). Factorial sampling plans for preliminary computational experiments, *Technometrics* 33(2) 161–174.
- Sobol, I.M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment* 1(4), 407–414.
- Spitz, C., L. Mora, E. Wurtz and A. Jay (2012). Practical application of uncertainty analysis and sensitivity analysis on an experimental house, *Energy and Buildings* 55, 459–470.
- Østergaard, T., S.E. Maagaard and R.L. Jensen (2015). A stochastic and holistic method to support decision-making in early building design, in: *Procs. of BS 2015, the 14th International Conference of the IBPSA*, Hyderabad, India.



Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 132 (2017) 897–902

Energy

Procedia

www.elsevier.com/locate/procedia

11th Nordic Symposium on Building Physics, NSB2017, 11-14 June 2017, Trondheim, Norway

Explaining variability in metered energy use for similar buildings using Bayesian inference

Martin Heine Kristensen*, Steffen Petersen

Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark

Abstract

Typologically identical buildings may exhibit large differences in energy use due to various stochastic phenomena. In this paper, we present a study on how these phenomena can be explained by analyzing 1,050 observations of metered district heating energy use from a sample of 350 similar detached single-family dwellings located in 37 city districts (urban areas) in the city of Aarhus, Denmark. The results indicate that annual variations within the same buildings due to e.g. weather conditions account for only a minor proportion of the overall data variance (approx. 10%-20%). A larger proportion (approx. 25%-51%) is captured and explained by phenomena between the typologically identical buildings, probably due to the stochastic nature of occupant behavior. The largest proportion (approx. 30%-65%) of the data variance is explained by the district location, which suggests the presence of a socio-economic effect influencing the level of energy use between city districts.

© 2017 The Authors. Published by Elsevier Ltd.

Peer-review under responsibility of the organizing committee of the 11th Nordic Symposium on Building Physics.

Keywords: Metered building energy use; Data variability; Hierarchical modeling; Bayesian inference; Socio-economic effect

1. Introduction

It is rather obvious that actual energy use for building operation varies with the type of building (office, homes, retail, etc.), but even typologically identical buildings that share the same energy ratings exhibit large differences in energy use, as was shown in a recent data visualization video of approx. 28,000 Danish single-family dwellings [1].

* Corresponding author. Tel.: +45 23 73 77 18.
E-mail address: mhk@eng.au.dk

The influence of occupant behavior is often ascribed considerable effect [2], but in reality, we know only very little about the mechanisms and phenomena that drive differences in energy use.

In this paper, we present a study showing how occupant behavior and other stochastic phenomena seem to manifest themselves on different aggregated levels, i.e. variations in energy use within the same buildings and between buildings of similar typology, as well as variations driven by the building location within the same city. The study is based on an analysis of metered district heating energy use for a large sample of similar detached single-family dwellings, all constructed within the same city in a period characterized by uniform building regulations. The data variance is modeled and decomposed in a hierarchical structure whereby the prevalence of the above-mentioned phenomena is analyzed. To do so, we employ a probabilistic approach using Bayesian mixed-effects modeling and regression including information about building energy use, age, location and floor area.

Nomenclature

y	Metered building energy use for heating [kWh/year].
x	Heated building floor area [m ²].
μ	Mean building energy use [kWh/year].
σ	Standard deviation of building energy use [kWh/year].
φ	Mean district energy use [kWh/year].
τ	Standard deviation of district energy use [kWh/year].
α	Intercept (const. effect) of district model [kWh/year].
β	Slope of district model [kWh/m ² /year].
θ	Mean of intercept value [kWh/year].
ω	Standard deviation of intercept value [kWh/year].
i	Indexing the observations from 1 to N_j .
j	Indexing the buildings from 1 to M_k .
k	Indexing the districts from 1 to L .
N_j	Number of metered data points from building j .
M_k	Number of buildings in district k .
L	Number of districts in dataset.

2. Method

2.1. Data

A random sample of 350 detached single-family dwellings, all constructed between 2008 and 2010 in the municipality of Aarhus, Denmark, and supplied by the public district heating network, was selected for analysis. All buildings in this construction period are expected to fulfill identical energy requirements given by the Danish Building Regulation (BR08) in force at the time of construction. The buildings were thus assumed to exhibit similar building physical properties with between-building variations being driven mainly by other factors than the building physical properties.

For each building, information about the construction year, heated floor area and location of the building within the city (urban area code) was collected from the publically available Building and Dwelling Register (BDR) that contains information about the Danish building stock. In total, 37 city districts (urban areas) were represented, each featuring between one and 60 of the 350 sampled buildings. Furthermore, the annual district heating energy use for the last three years was collected for all buildings yielding three observations per building – in total comprising 1050 observations. The district heating energy use consists of energy use for hydronic space heating in e.g. radiators and underfloor heating, and energy use for on-site domestic hot water (DHW) preparation.

2.2. Hierarchical regression model

A three-level mixed-effects model was assumed for the metered annual district heating energy use data, y_{ijk} , by fitting a hierarchical structure to observations $i = 1, \dots, N_j$ for each building $j = 1, \dots, M_k$ within the $k = 1, \dots, L$ city district groups:

Level one (Observations):

$$\log(y_{ijk}) \sim N(\mu_{jk}, \sigma^2). \quad (1)$$

Level two (Buildings):

$$\mu_{jk} \sim N(\varphi_k, \tau^2). \quad (2)$$

$$\varphi_k = \alpha_k + \beta \cdot x_{jk}, \quad (3)$$

Level three (Districts):

$$\alpha_k \sim N(\theta, \omega^2). \quad (4)$$

At level one, the logarithm of the observed annual district heating energy use, $\log(y_{ijk})$, was assumed to be independent and identically distributed (i.i.d.) random samples from a Gaussian distribution with unknown mean building energy use μ_{jk} and standard deviation σ . For simplicity, homoscedasticity was assumed across buildings such that $\text{Var}(\log(y_{ijk})) = \sigma^2 \forall j$. At level two, the building mean energy use, μ_{jk} , was likewise assumed i.i.d. randomly sampled from a Gaussian distribution with unknown district means φ_k and homoscedastic district variations, i.e. $\text{Var}(\mu_{jk}) = \tau^2 \forall k$. By modeling the district means, φ_k , as a linear function (in the logarithmic domain) of the building floor area, x_{jk} , with a constant slope, β , and district-specific intercept, α_k , the effect of the individual districts can be inferred as the posterior variations in α_k . At level three, the intercept, α_k , was i.i.d. sampled at random from a Gaussian distribution with an unknown grand mean θ and standard deviation ω . To complete the hierarchical model and ensure data-driven posterior inference, noninformative Uniform prior distributions were assigned to all the hyperparameters ($\sigma, \beta, \tau, \theta, \omega$) based on recommendations by Gelman et al. [3].

2.3. Model pooling

Given a belief of exchangeability among the district group-level intercept parameters (4), the hierarchical model allows districts with less information to borrow strength from groups with more information through their shared parent distribution, $N(\theta, \omega^2)$, hereby presenting a compromise between two alternative models: a “no pooling” model and a “complete pooling” model. The no pooling model is the limiting case where the between-group variance parameter $\omega = \infty$, i.e. asserting that there is no information hidden in the between-group distribution of α_k , hereby abandoning the hierarchical modeling of α_k and reducing the model to (1)-(3), generating a separate fit for each district. The complete pooling model represents the opposite limiting case where $\omega = 0$, which arises when separation in the district-level is believed to be irrelevant, imposing the restriction that $\alpha_k = \theta \forall k$, i.e. all districts share the same intercept eliminating the effect of the districts in explaining differences in energy use. All three models – the no pooling model, the hierarchical model and the complete pooling model – were fitted to investigate whether information was hidden in the location of the buildings (effect of district grouping).

2.4. MCMC algorithm for posterior inference

The multi-dimensional joint posterior distribution cannot be obtained analytically; hence, a numerical approach was employed based on Hamiltonian Monte Carlo (HMC), a hybrid Markov Chain Monte Carlo (MCMC) algorithm whose equilibrium distribution is indeed an approximation of the joint posterior distribution [4]. Four chains were run in parallel with randomly dispersed starting points in the parameter space to draw samples from the posterior

distribution. For each chain, 2,000 MCMC samples were drawn with the first 1,000 samples being considered cool, meaning that information about the starting point might still prevail. Samples from this cold period were thus discarded leaving only the warm part of the chains for analysis.

Convergence in the warm chains was monitored in terms of the potential scale reduction factor, \hat{R} , for which $\hat{R} \in \mathbb{R} | 1 < \hat{R} < \infty$. It is an estimate of the scale with which the variations in the inferred parameter distributions might be reduced if the simulations were continued in the limit $n \rightarrow \infty$ ($\lim_{n \rightarrow \infty} \hat{R} \rightarrow 1$) [4]. \hat{R} accounts for the within-chain and between-chain variance in the warm chains, simultaneously evaluating both the mixing and stationarity of it. For $\hat{R} < 1.1$, a stable and converged estimation was considered for each parameter, respectively.

3. Results

3.1. Model selection

The ability of the three models to fit the data was assessed by means of their expected predictive accuracy in terms of the Watanabe-Akaike information criterion (WAIC), a state-of-the-art fully Bayesian measure of model fit [5]. Compared to non-Bayesian measures like the AIC [6] and BIC [7], and the somewhat Bayesian measure DIC [8], WAIC has the desirable property of averaging over the posterior distributions rather than conditioning on point estimates, making WAIC a fully Bayesian approach for estimating the out-of-sample expectation [9]. As might be expected, the hierarchical model shows to have the lowest WAIC value and thus constitutes the best fit to data (Table 1).

Table 1. Predictive accuracy of the three fitted models. Lower values of WAIC imply higher predictive accuracy.

	No pooling ($\omega = \infty$)	Hierarchical (ω estimated)	Complete pooling ($\omega = 0$)
WAIC	-487	-495	-486

3.2. Explained variance by groups

The hierarchical structure allowed the unknown data variance to be decomposed and fitted to the standard deviation hyperparameters at the three levels of the hierarchical model (σ, τ, ω). The proportion of variance explained at each level of the model was assessed in terms of the intraclass correlation coefficient (ICC), here shown for the first level (variation between annual measurements of the same building):

$$ICC_{level1} = \frac{\sigma^2}{\sigma^2 + \tau^2 + \omega^2}. \tag{5}$$

The ICC measures (Table 2) indicate that approx. 30%-65% (95% central posterior probability) of the overall data variance can be explained by phenomena between city districts (level three). Adding information about which particular building within a given district the data originates from can explain an additional 25%-51% of the data variance (level two). The remaining data variance is explained by phenomena between the annual measurements of the buildings (level one).

Table 2. Intraclass correlation coefficient (ICC) of groups.

ICC	Posterior quantiles				
	2.5%	25%	Median	75%	97.5%
Within buildings/between years (level one)	0.10	0.13	0.15	0.17	0.20
Within districts/between buildings (level two)	0.25	0.34	0.38	0.43	0.51
Within dataset/between districts (level three)	0.30	0.41	0.47	0.53	0.65

A one-way analysis of variance (ANOVA) was used to test the null-hypothesis that all the district mean energy use distributions, φ_i , were equal ($H_0: \varphi_1 = \varphi_2 = \dots = \varphi_L$) against the alternative hypothesis that at least one of the districts had a different mean energy use ($H_A: \varphi_1 \neq \varphi_2 = \dots = \varphi_L$). The null-hypothesis was rejected ($p < 1e-10$)

indicating the presence of a significant effect of the district location on the annual district heating energy use of buildings.

A model prediction of each district group is shown in Figure 1 using the 37 individual district sub-models (3) of the hierarchical model with individual intercept parameters, α_k . Ignoring the differences in the district groups ($\omega = 0$), the grand mean model is overlaid using the mean intercept, θ .

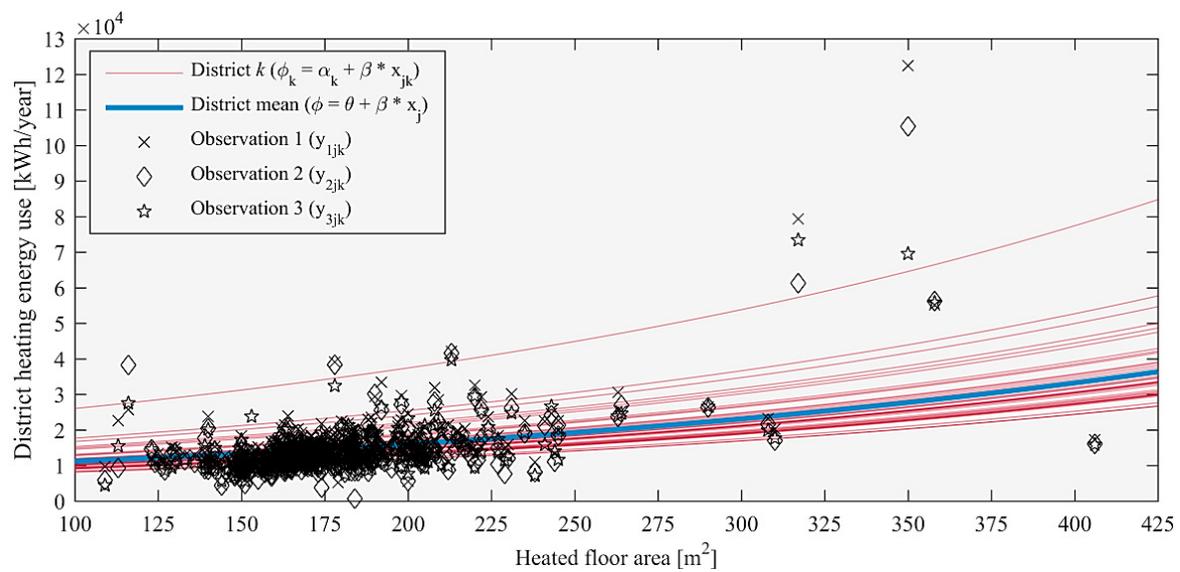


Figure 1. Predictions using the 37 individual district sub-models and the overall mean model.

4. Discussion

4.1. Variability in energy use

The results indicate that energy consuming mechanisms and phenomena between district locations account for 30%-65% (95% central posterior probability) of the overall data variance. This suggests the presence of an additional socio-economic effect causing buildings from different city areas to behave and, to some extent, consume energy differently. Similar conclusions are drawn in another study of the Danish residential building stock [10], where considerable variation in how people consume heat is attributed to various socio-cultural factors. It would be interesting, in future work, to include information about the household income of the different districts to test whether there is indeed a socio-economic correlation between energy use and income. However, district-level variations could also be caused by local regulations on the architecture, environmental exposure, e.g. local wind patterns, shading, etc., and/or because buildings in certain areas are systematically designed as low-energy construction.

The fact that an additional 25%-51% (95% central posterior probability) of the data variance is captured and explained by phenomena between buildings within the individual districts is particularly interesting. Bearing in mind that the investigated buildings were all detached single-family houses constructed during a period of unchanged building regulations, such phenomena can hardly be ascribed to any major building-physical or geometrical differences. Instead, it seems legitimate to expect this variation to be caused mainly by the very stochastic nature of occupant behavior. However, to more accurately account for the influence of occupant behavior, additional information about the buildings would have to be included in the model, e.g. number of occupants living in each building, presence of basements and heated attics, energy label ratings, etc.

The remaining 10%-20% (95% central posterior probability) data variance is attributed to energy consuming mechanisms and phenomena between annual measurements within individual buildings. This annual variation is

probably caused by natural variations in the weather. However, strictly speaking, this number contains any residual variability that cannot be decomposed with the proposed hierarchical structure, e.g. observation error.

4.2. Perspectives on future model expansion

Expanding the hierarchical model to include also other building typologies and vintages, e.g. terraced houses and apartment blocks, would allow for a more exhaustive investigation of the mechanisms behind energy use, and for additional research questions to be asked and answered. Bearing in mind that the hierarchical structure allows information to be borrowed from the different sub-groups, such an extended model could potentially provide a more precise and profound estimate of common effects, e.g. occupant behavior. From a more practical point of view, an expanded model could serve as a city-scale prediction tool for e.g. urban planners and municipality managers and would be beneficial for e.g. assessing the capacity of existing public supply systems when expanding building areas within the city, and for assessing the impact of different retrofit scenarios.

5. Conclusion

Significant information about annual heating energy use is contained in knowing the location within a city from which the measurements are obtained. It is proposed that this phenomenon is caused by a socio-economic effect. The effect is not, however, equally profound for all investigated city districts as the energy use in some districts is more similar than in others. In addition, what seems to be a natural variation in occupant behavior between typologically identical buildings explains the majority of variation within city districts, leaving only a minor variance proportion to be explained by weather phenomena.

Acknowledgements

The research was conducted as part of the “Resource Efficient Cities Implementing Advanced Smart City Solutions” (READY) project, work package 3, financed by the 7th EU Framework Programme (FP7-Energy project reference: 609127). Furthermore, the authors would like to thank the district heating company in Aarhus, AffaldVarme Aarhus, for supplying the building energy data that forms the basis of the study.

References

- [1] M. H. Kristensen, Danskernes fjernvarmeforbrug [in Danish], Aarhus University, 12 12 2016. [Online]. Available: https://www.youtube.com/watch?v=Opd_hJMQBos&feature=youtu.be. [Accessed 28 01 2017].
- [2] R. V. Andersen, J. Toftum, K. K. Andersen and B. W. Olesen, Survey of occupant behaviour and control of indoor environment in Danish dwellings, *Energy and Buildings* 41 (2009) 11–16.
- [3] A. Gelman, Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* 3 (1) (2006) 515-534.
- [4] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin, *Bayesian Data Analysis*, 3 ed., CRC Press, 2014.
- [5] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, Cambridge, UK, Cambridge University Press, 2009.
- [6] H. Akaike, Information theory and an extension of the maximum likelihood principle, in *Proceedings of the Second International Symposium on Information Theory*, Budapest, 1973.
- [7] G. Schwarz, Estimating the dimension of a model, *The Annals of Statistics* 6 (2) (1978) 461-464.
- [8] D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. van der Linde, Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64 (4) (2002) 583-639.
- [9] A. Gelman, J. Hwang and A. Vehtari, Understanding predictive information criteria for Bayesian models, *Statistics and Computing* 24 (6) (2014) 997-1016.
- [10] A. R. Hansen, The social structure of heat consumption in Denmark: New interpretations from quantitative analysis, *Energy Research & Social Science* 11 (2016) 109-118.

Bottom-up modelling methodology for urban-scale analysis of residential space heating demand response

Rasmus Elbæk Hedegaard ^{a*}, Martin Heine Kristensen ^a,
Theis Heidmann Pedersen ^a, Adam Brun ^b, Steffen Petersen ^a

^a *Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark*

^b *AffaldVarme Aarhus (Department of Waste and District Heating), Aarhus Municipality, Bautavej 1, 8210 Aarhus, Denmark*

Abstract

Several studies have indicated a potential to exploit the thermal inertia of individual residential buildings for demand response purposes using model predictive control and time-varying prices. However, studies that investigate the response obtained from applying these techniques to larger groups of buildings, and how this response affects the aggregated load profile, are needed. In this study, we propose a methodology for modelling residential buildings that enables bottom-up modelling of entire urban areas. The methodology is based on the thermal model described in ISO 13790, which was extended to a second order model to improve its capability to describe the thermodynamic behaviour of buildings under dynamic conditions, and a Bayesian statistical framework used for the inference of model parameters. The methodology utilizes three sources of information for model calibration, namely public building registers, weather measurements, and hourly smart-meter consumption data. The methodology was tested through the modelling of a residential neighbourhood consisting of 159 single-family houses in the city of Aarhus, Denmark. The aggregated model was capable of predicting the aggregated district heating consumption in a previously unseen validation period with high accuracy: CVRMSE of 5.58% and NMBE of -1.39%. The model was then used to investigate the effectiveness of a DR scheme with the objective of reducing the daily fluctuations in the district heating consumption due to periods with increased domestic hot water consumption. The results showed that a commonly applied price-based demand response scheme incentivizing consumers through time-of-use energy prices would lead to the formation of new, undesirable peaks. To avoid this, a requirement for a more distributed response from the individual consumers was added to the DR scheme. This significantly improved effectiveness of the DR scheme as the size of two investigated peaks was reduced by 6.3% and 4.3%, respectively, without generating new peaks. This suggests that future research exploring and comparing various DR schemes on their effectiveness and efficiency at addressing various system performance objectives is needed. The methodology presented in this paper seems well-suited for such analysis.

Keywords: Bayesian calibration; Urban scale bottom-up modelling; Demand response; Space heating; Domestic hot water; Smart meter data

* Corresponding author. Tel: +45 61707378

E-mail address: reh@eng.au.dk

1 Introduction

The increasing availability of various high-resolution monitoring data from energy systems in operation leads to new opportunities for maintaining the balance between supply and demand while increasing the efficiency of energy systems as a whole. One of these opportunities is improved demand side management initiatives, i.e. attempts to adapt demand to supply. The conventional notion of demand side management covers both initiatives that seek to reduce the overall demand as well as initiatives that seek to optimize the temporal distribution of demand. The latter is often referred to as demand response (DR); an approach which is increasingly being considered a viable tool for supporting the transition to an energy system based on renewable energy sources (RES) [1, 2], in which an inherent challenge is to establish an instantaneous match between demand and fluctuating energy production. Here, the availability of high-resolution monitoring data is an important prerequisite for enabling DR in practice.

The energy use in residential buildings constitutes a significant potential for DR as they account for 25% of the total energy consumption in the EU, whereof 67% is used for space heating in the North and West regions of EU [3, 4]. Several simulation-based studies have demonstrated that DR schemes for residential space heating may be used to achieve societal objectives such as reductions of peak demand [5, 6], but also cost savings for consumers through strategic consumption [7–9]. These studies exploit the thermal inertia of the buildings to shift the energy consumption used for space heating to achieve economic or societal benefits while ensuring acceptable thermal conditions inside the buildings. While most of these studies focus on the DR potential with respect to the electrical grid, the same type of DR initiatives could in principle also be used for generating benefits for other parts of the energy system such as district heating (DH) networks, e.g. as it was done for production of domestic hot water by Knudsen and Petersen [10]. Furthermore, previous studies have primarily focused on investigating the DR potential of individual buildings or apartments, see e.g. [5–9, 11–14]. O’Connell et al. [15] argue that, while such studies provide great insights on the DR for the specific scenarios considered, such isolated cases may not describe the behaviour of DR on the larger scale. To develop the current body of research on the potential for DR in space heating of residential buildings, it therefore seems reasonable to investigate the DR potential of buildings on an aggregated level using techniques inspired by the emerging field of urban building energy modelling (UBEM) [16]. Research within this field has modelled the aggregated consumption of groups of buildings through a variety of methods, see e.g. [17–21]. Previous studies have applied these methods to investigate various topics, including the performance-gap between theoretical and actual energy consumption [22] and the potential for energy retrofitting on city-scale [23]. Currently, studies on how the aggregated DR from several individual buildings affects operational challenges in urban district heating systems are rare; the only identified studies are those of Dominkovic et al. [24] and Cai et al. [25]. Dominkovic et al. used archetype building models calibrated with data from 54 households to extract performance characteristics of typical DR events, i.e. approaches for shifting consumption. This condensed representation of the building stock and the use of predefined set of example DR events allowed the authors to evaluate the potential for utilizing the thermal mass of buildings for generating flexible consumption in an energy system level

optimization. The authors demonstrated that the flexible demand generated by buildings in their case accounted for 5.5-7.7% of the total demand. Cai et al. used first order models to represent the space heating requirements and hot water tanks of 20 residential apartment buildings as well as a commercial consumer. These models were used in optimization to reduce the operational costs of the supply side through utilization of flexible consumption, while ensuring consumer comfort. The optimization resulted in achieved savings of up to 11% when compared to a baseline.

While both of these studies indicate a potential for utilization of passive thermal storage, both studies used simplified representations of buildings in their analysis, either through the use of archetype models which neglect diversity, or first order models which neglect the thermodynamic phenomena that are involved with exploiting the thermal inertia of buildings for storage purposes. In this paper, we propose a different take on the evaluation of residential DR utilizing passive thermal storage, which rely on statistical calibration of physics-based second order models of individual single-family houses. We then validate the ability of the second order model to describe dynamical conditions in building through experimental and field data, before demonstrating the modelling approach by calibrating a bottom-up model of a neighbourhood consisting of individual building models each with their own thermal characteristics and domestic hot water draw profiles. The model enables detailed investigations on how to design and evaluate the performance of DR schemes which utilizes the flexibility potential in residential space heating to avoid certain operational challenges e.g. in urban district heating systems.

The paper is structured as follows: Section 2 describes the proposed UBEM methodology in terms of the established statistical framework and the physics-based model structure. In section 3, we apply the methodology in a case study where we model the consumption of 159 detached single-family houses. The obtained UBEM model is validated both on the scale of the individual building and on the aggregated level before it is applied in a case study on residential DR potential using model predictive control (MPC). Finally, we draw our final conclusions and outline future work in section 4.

2 Method

The current field of UBEM consists of a variety of methods for modelling of the energy use of groups of buildings. In general, UBEM models can be categorized as either top-down or bottom-up models. Top-down models tend to rely on socio-economic factors such as energy prices, population size and weather conditions for modelling energy use [26]. The use of aggregated data for obtaining top-down models leads to little emphasis on ensuring accurate representation of the energy-consuming processes themselves, thus rendering them ill-suited for evaluating the DR potential of utilizing building thermal mass as passive thermal storage. Bottom-up models, on the other hand, does not suffer from this issue, since they model the physical processes and phenomena of the energy-consuming processes themselves before aggregating the results [27, 28]. These models may be rooted in either statistical methods, physical principles or a combination thereof. Statistical methods include regression analysis, support vector machines and artificial neural networks, whereas physics-based models rely on the first principles to model the energy consuming process [26]. These two modelling paradigms differ significantly in the

prerequisites needed to obtain the model. Statistical methods rely on measurement data describing input-output relationships of the process, whereas physics-based methods rely on knowledge about the process itself. While both modelling approaches have advantages and disadvantages, it is the combination of them that truly makes them useful in practical applications. Combining the two paradigms typically involves setting up an initial model based on thermodynamic principles and any available information regarding the buildings, and then calibrating said model using measured input-output data. This significantly improves the accuracy of the resulting models compared to those derived from the purely physics-based approach, while at the same time lowering the requirements for both the quality and quantity of the data needed in the statistical approach [29, 30]. Examples of studies combining the physics-based and data driven modelling approaches to model individual buildings are plentiful – studies that have modelled buildings on the larger scale include Kristensen et al. [31], who calibrated a UBEM for prediction of the annual energy use in Danish detached single-family-houses, and Gianniou et al. [32], who used ordinary-least-squares linear regression and the degree-day method to derive estimates of the indoor set point temperature and overall heat losses of over 15,000 residential buildings from similar smart-meter data. More recently, Kristensen et al. [33] used smart-meter data in a 3-hourly resolution and a hierarchical modelling approach to construct archetype models capable of predicting the aggregated consumption of out-of-sample groups of buildings.

Taking experiences from previous UBEM studies in literature into account, it seems reasonable to use a modelling approach which relies on bottom-up physics-based models calibrated with measured data to explore the DR potential of flexible space heating consumption through exploitation of the thermal inertia that is inherent to buildings. We therefore propose a modelling methodology which relies on Bayesian calibration methods to derive physics-based models of individual houses. The proposed methodology is unfolded in the following sections. First, the Bayesian statistical framework used for the inference of model parameters is described. Then details on the building modelling including the physics-based model structure; assumptions used for modelling building geometries; and venting and internal loads are provided. Finally, a novel method for separating measurements of the total district heating consumption into its space heating and domestic hot water components is proposed.

2.1 Statistical framework

The statistical calibration of the physical models is based on the Bayesian paradigm, which enables incorporation of a-priori information in an otherwise measurement data-driven model calibration - not in the form of best-guesses, but as arbitrary probability distributions reflecting the uncertainty of the information. This coupling of measurement data and expert knowledge in form of priors is particularly useful in applications where the data alone may not be sufficiently informative to identify the most likely value of a given model parameter. In the context of buildings, priors may therefore be used to guide the inference of parameters which are only vaguely described by the data towards regions of high prior probability. Another benefit of the Bayesian modelling approach is that the resulting models contain full approximations of the parameter posterior distributions, thus allowing all the uncertainty indicated by the inference to

be included in future analysis. These posteriors may be reduced to point estimates such as the maximum a posteriori (MAP) estimate to be used in applications where the full Bayesian model description including the uncertainties are of less relevance.

A key assumption made in the Bayesian inference of the proposed method was that measurements of district heating consumption Φ_{DH} in a residential building can be modelled as the sum of an energy-consuming process, Φ_{sim} , and a stochastic component, ϵ_t . Furthermore, the output of the energy-consuming process was assumed to consist of two components: energy used for space heating (Φ_{h}) and for preparation of domestic hot water (Φ_{DHW}). Time is denoted by t throughout Eq. (1a-1e). In Eq. (1e), hr refers to the hour number within a day while d is a categorical variable denoting the current type of day (“workday” or “weekend day”). n_t denotes the length of a simulation. The relationship between a time series of district heating measurements $\Phi_{\text{DH},t}$ for $t = [1, 2, \dots, n_t]$ and the vector containing the output of the model, $\Phi_{\text{sim},t}$ are described by Eq. (1a-1c). As indicated by the index, $\Phi_{\text{sim},t}$ is the simulated prediction, thus implying that the calibration applies the infinite-step-ahead error criterion.

$$\Phi_{\text{DH},t} = \Phi_{\text{sim},t} + \epsilon_t \quad (1a)$$

$$\Phi_{\text{sim},t} = \Phi_{\text{h},t} + \Phi_{\text{DHW},t} \quad (1b)$$

$$\epsilon_t \sim \text{N}(0, \sigma^2) \quad (1c)$$

$$\Phi_{\text{h},t} = \text{F}(\mathbf{X}, \boldsymbol{\theta}_{\text{B}}, \mathbf{W}, t) \quad (1d)$$

$$\Phi_{\text{DHW},t} = \text{F}(\theta_{\text{occ}}, \boldsymbol{\theta}_{\text{DHW}}, d, hr) \quad (1e)$$

The input arguments used in the individual models for space heating and domestic hot water preparation are listed in Eq. (1d-1e). Here, $\boldsymbol{\theta}_{\text{B}}$ and $\boldsymbol{\theta}_{\text{DHW}}$ are vectors containing the calibrated parameters of the building model and DHW model, respectively, while \mathbf{X} is a vector containing the fixed parameters of the thermal model not subject to calibration, see Appendix 1 of this paper. The matrix \mathbf{W} of height n_t contains columns with measurements of relevant weather conditions. Finally, the scalar $\theta_{\text{occ}} \in \boldsymbol{\theta}_{\text{B}}$ denotes the inferred number of occupants of the building.

The stochastic component ϵ_t in Eq. (1a) governs any unpredictable variation caused by potentially noisy measurements of both weather and district heating consumption, process noise (occupancy) as well as the inevitable mismatch between the chosen model-structure and the true energy-consuming process. Here, a key assumption in the inference is that the residuals of the model output on the measurement data, ϵ_t , are realizations of a normally distributed random variable with zero-mean and homoscedastic variance σ^2 . The likelihood of the data conditional on all free model parameters ($\boldsymbol{\theta}_{\text{B}}$, $\boldsymbol{\theta}_{\text{DHW}}$ and σ) is then:

$$p(\Phi_{\text{DH}} | \boldsymbol{\theta}_{\text{B}}, \boldsymbol{\theta}_{\text{DHW}}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \sum_{t=1}^{n_t} \epsilon_t^2} \quad (1f)$$

Assuming that the model accurately describes the measured phenomenon, Bayes' theorem states that the joint posterior probability of the parameters conditional on the measurement data, $p(\boldsymbol{\theta}_B, \boldsymbol{\theta}_{DHW}, \sigma | \Phi_{DH})$, is proportional to the RHS of Eq. (1g), where $p(\Phi_{DH} | \boldsymbol{\theta}_B, \boldsymbol{\theta}_{DHW}, \sigma)$ is the likelihood of the data while $p(\boldsymbol{\theta}_B)$, $p(\boldsymbol{\theta}_{DHW})$ and $p(\sigma)$ denote the probability of the building model parameters under their respective priors.

$$p(\boldsymbol{\theta}_B, \boldsymbol{\theta}_{DHW}, \sigma | \Phi_{DH}) \propto p(\Phi_{DH} | \boldsymbol{\theta}_B, \boldsymbol{\theta}_{DHW}, \sigma) \cdot p(\boldsymbol{\theta}_B) \cdot p(\boldsymbol{\theta}_{DHW}) \cdot p(\sigma) \quad (1g)$$

The objective of the Bayesian inference was to obtain an approximation of the joint posterior distribution of the parameters of the model chosen for calibration. There are several algorithms that are capable of this through iterative sampling from the joint posterior probability function of Eq. (1g). This study relies on the Metropolis algorithm [34] to do so – further details are given in section 3.1. The following section provides details on the physics-based model structure chosen to describe the energy-consuming process (Φ_{sim}) of Eq. (1a).

2.1 Model structure and assumptions

The physics-based model structure used in the proposed method is a modified version of the “simple hourly method” presented in ISO 13790 [35]. This simple model structure is a so-called reduced order model, which is suited for implementation in model predictive control schemes – the state-of-the-art control method most frequently used for studies on smart buildings [36]. An initial model of a given building in the urban area is established using information available in the publicly available Danish Buildings and Dwellings Register [37] regarding the year of construction, footprint area, number of stories and the areas of any conditioned attic and/or basement. The information on areas were used for establishing a geometric model of the building, while the year of construction was used for the specification of the prior probability of insulation levels used in the Bayesian inference. During the inference, the model was simulated under the assumption of a constant heating temperature set point to obtain the model output, $\Phi_{sim,t}$ for $t = [1, 2, \dots, n_t]$, with n_t depending on the length of the time series used for calibration.

2.1.1 Building energy model structure

The resistance-capacitance model described in ISO 13790 is based on the principles of thermodynamics and consists of five thermal resistances as well as a single thermal capacity describing the inertia of the building's thermal mass (in short: 5R1C). Several reports have documented detailed tests on the ability of the model to describe the energy consumption of buildings accurately – often by comparing the model to more complex modelling approaches such as the EnergyPlus or TRNSYS Type-56 building models [38, 39]. In addition to the state governing the temperature of the thermal mass (C_m), the model features two massless temperature nodes describing the air (T_i) and mean surface temperatures (T_s), respectively.

The fact that the ISO 13790 model is a first order model enables the discretization of the model to rely on the Crank Nicholson scheme while avoiding the need for solving a system of equations. This ensures fast and numerically stable solutions of the model during simulation. However, the

low model order comes at the cost of a rather crude representation of the dynamic behaviour of buildings. While the simplification of representing the indoor environment without thermal inertia does not impact simulations under stationary conditions (i.e. in the intended use of the model), it results in the 5R1C-model being ill-suited to represent buildings under dynamic operational conditions. This issue is especially problematic in the context of using the model for evaluating demand response based on exploiting the thermal characteristics of the building. Since the majority of Danish residential buildings have radiator or convector-based systems for space heating, it is primarily the interior of the building which is acted upon by the heating system. Since the process of storing energy in the heavy building components requires energy to flow through the thermally light interior environment, it is necessary to model both the fast and slower dynamics of the building. This claim is supported by previous studies: Reynders et al. [29] suggested using high-quality data including heat flux measurements from the building envelope to calibrate fourth and fifth order models, but argued that models of lower orders could suffice for control applications. Hedegaard and Petersen [40] compared second and third order models on the physical meaningfulness of their parameters and concluded that the second order model performed well. Harb et al. [41] tested RC-models of orders one to three and concluded that the second order model provided the best compromise between good predictive capabilities and interpretability of model parameters. Finally, Vivian et al. [39] concluded second order models to be preferable over first order models. As the ISO 13790 model is well-suited in terms of all other aspects of the intended application, we opted for a modification of the original model structure instead of replacing it with a different structure. The modification was to model the thermal inertia of the interior (room air, furniture etc.) by turning the previously mass-less air temperature node into a temperature state with an associated thermal capacity (C_i), thereby expanding the previously first order model to a second order model. In order to keep the model simple and relatively robust to the quality of the data used in the calibration, we did not introduce further modifications. A resistance-capacitance network of the proposed 5R2C-model including the modification (in green) is shown in Figure 1. Since the air change in the residential buildings are assumed maintained through natural ventilation, the ventilation supply temperature indicated on the figure (T_{supply}) is assumed to be equal to the exterior air temperature (T_{exterior}).

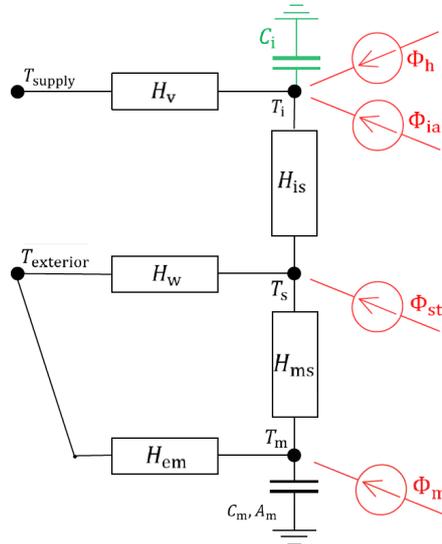


Figure 1 Modified RC-network of the model used in ISO 13790's hourly method for calculating heat consumption. Green: modification, red: thermal loads.

In addition to the improved representation of the fast dynamics of buildings, the introduction of the second state of the model allowed us to remove the feedthrough component of the ISO 13790 model. Feedthrough happens when inputs (in this case heating power Φ_h and solar gains in Φ_{ia}) act directly on the output of the model, and not only indirectly by affecting the states of the model. Removing the feedthrough and adding thermal inertia to the interior (C_i) eliminates unrealistic changes in the temperature output of the model caused by e.g. fluctuations in heating power. This tendency of large fluctuations was also found by Bruno et al. [42], who compared the ISO 13790 5R1C model to the more complex TRNSYS type 56 building model. The full continuous-time state-space representation of the modified model structure is given in Appendix 2 of this paper.

To evaluate the impact of the modification on the ability of the model to describe buildings under dynamic conditions, both the original (5R1C) and modified model (5R2C) structure were initially calibrated using two datasets of hourly time resolution: 1) experimental data from laboratory tests where a pseudo-random binary sequence was used to control a hydronic heating system and impose temperature fluctuations on the test room, and 2) field data measured in an unfurnished terraced house where electrical heaters were used to excite the building by modulating the temperature set point. Both datasets included measurements of the internal air temperature (T_i), heating power (Φ_h) and weather conditions (solar irradiance and air temperature). The calibration of the two model structures relied on an algorithm similar to the one described in section 3.1, which was used for establishing the bottom-up UBEM model.

Five parameters present in both the original (5R1C) and the modified (5R2C) model were calibrated along with the thermal capacity of the interior that is only a part of the modified model structure. Since both datasets were captured under controlled conditions, the assumptions related

to the distribution of windows and the presence of occupants where adjusted to match the actual conditions. Figure 2 shows the model output of both model structures after calibration. The mean absolute percentage error (MAPE) calculated for each model is provided in the legend.

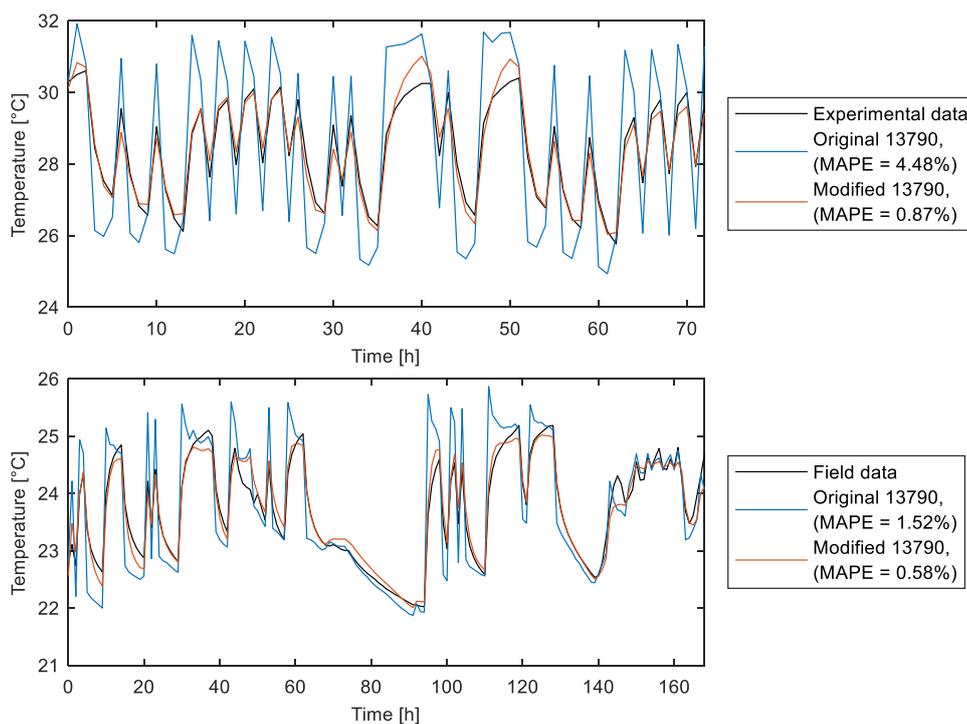


Figure 2 Impact of the modification of the ISO 13790 model. The upper and lower graphs show model performance when calibrated with experimental data and field data, respectively. The depicted models are parameterized with the maximum likelihood-estimates of the calibration parameters.

The modified model structure generally shows better predictive capabilities on both datasets – especially during periods characterized by temperature fluctuations. The calibration was carried out without the use of priors in order to remove any impact of our a-priori beliefs on the result of the model evaluation. As such, the depicted models are parameterized with the maximum likelihood estimates from the calibration.

Despite the indicated benefits of calibrating the thermal capacity of the interior, it is considered infeasible to estimate this model parameter as our measured data does not include measurements of the indoor air temperatures. Therefore, we chose to fix the value of the thermal capacity. By default, EnergyPlus models the thermal capacity of the interior as that of the air volume inside the building alone [43]. We do not consider this an option in the context of low-order resistance-capacitance models used in simulations of hourly temporal resolution. Instead, we find it likely that contributions from interior (e.g. furniture), the heat delivery system itself, and even the innermost layers of construction elements are all lumped into the interior thermal capacity (C_i).

Therefore, it seemed reasonable to draw on the inference of the parameter indicated by the calibration of the unfurnished terraced house (“Field data” in Figure 2). The obtained empirical posterior distribution of the thermal capacity of the interior resembled a Normal distribution with a mean of $56.6 \frac{\text{kJ}}{\text{m}^2\text{K}}$, std. dev. of $2.46 \frac{\text{kJ}}{\text{m}^2\text{K}}$, and a minimum-maximum from approximately 45 to $65 \frac{\text{kJ}}{\text{m}^2\text{K}}$. In the context of exploiting the dynamic behaviour of buildings for realizing flexible consumption, choosing a lower estimate of the available thermal capacity would be on the safe side. Therefore, we chose the minimum value from the obtained posterior, i.e. $C_i = 45 \frac{\text{kJ}}{\text{m}^2\text{K}}$. This value is considered to be conservative, especially as the terraced house was unfurnished. The posteriors of the calibrated parameters obtained in this model validation are depicted in Appendix 3 together with insights related to the identifiability of the other parameters inferred in this model validation.

2.1.2 Geometric model

The geometric model shown in Figure 3 was used to translate the geometrical information from the Buildings and Dwellings Register (see section 2.1) into surface areas of the building envelope of the individual buildings. The approach is similar to the one presented in [31] and [33]. The geometric model used an assumed length-to-width ratio and room height to transform footprint areas to do so. Furthermore, the geometric model assumed each facade of a building to be facing one of the four cardinal directions. Appendix 1 lists the assumed values of parameters both in the geometric model but also the ISO 13790 model. In addition to the assumed relationships related to the overall geometry of the dwellings, a simplified representation of the window-distribution was assumed. This was done primarily to reduce the number of parameters governing the solar heat gains of the model. The assumed distribution was equal to that of the distribution used in the Danish window energy balance calculation methodology (E_{ref} [44]) North = 26%, West = 16.5%, East = 16.5% and South = 41%.

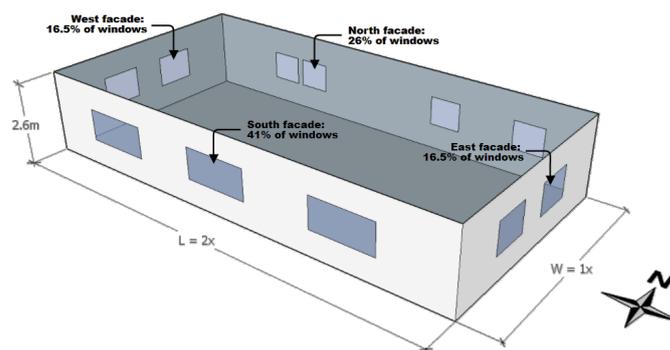


Figure 3 Geometric model used for approximating the areas of the building envelope from the building’s floor area. The model assumes a length/width ratio of 0.5, a room height of 2.6 meters, and distribution of windows as depicted.

2.1.3 Venting and internal loads

The vast majority of existing Danish detached single-family houses are naturally ventilated. The air change in each building was therefore modelled as an intentional component (venting) and infiltration, respectively. Since occupants are expected to vent less in colder periods, a model for the intentional venting reflecting this relationship was introduced. The model for venting was based on Rijal et al. [45] who proposed the use of a logistic model for describing the nonlinear relationship between the external temperature and the fraction of open windows in office buildings based on field surveys. In our study, we adapted the coefficients of this model so that it describes the effective amount of natural ventilation as a fraction of a design flowrate. Figure 4 depicts the relation between the external air temperature and the venting factor b used to scale a base venting rate fixed at $0.4 \frac{l}{s \cdot m^2}$ used in the model. Here, the whole design air change rate is in effect at external temperatures equal to the internal temperature (20 °C), while all venting seized at external temperatures of -20 °C.

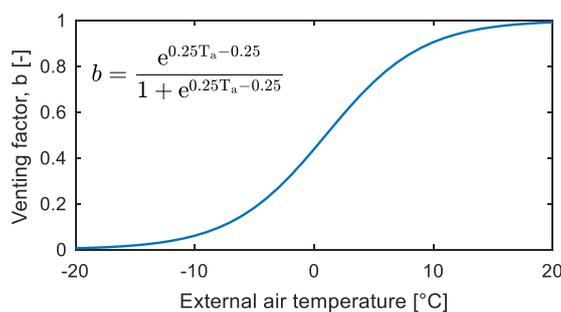


Figure 4 Model for external temperature-dependent manual venting.

The internal heat gains included in the models were gains from household appliances (Φ_{app}) and from the occupants themselves (Φ_{occ}). Each occupants were assumed to each generate 80 W of sensible heat and occupy the dwelling two-thirds of the time, see Eq. (2). The internal heat gains from appliances were modelled as the electricity consumption of a typical Danish household - an assumption considered appropriate in Danish dwellings where cooking is predominantly done using electricity. Gram-Hanssen [46] derived the empirical regression in Eq. (3) for the annual electricity consumption in detached single-family houses (E_{app}) based on data from 8500 dwellings. The regression relies on the area of the building and the number of occupants as predictors for the annual consumption in kWh. Since no information was available on the specific use-times of each dwelling, we assumed flat profiles for all internal heat gains - i.e. Eq. (2) and Eq. (4).

$$\Phi_{\text{occ}} = \frac{2}{3} \cdot 80 \quad (2)$$

$$E_{\text{app}} = 530 + 12 \cdot A_{\text{footprint}} + 690 \cdot \theta_{\text{occ}} \quad (3)$$

$$\Phi_{\text{app}} = \frac{E_{\text{app}} \cdot 1000}{8760} \quad (4)$$

2.1.4 Domestic hot water

The smart meters that measure district heating consumption featured in this study are primarily intended for billing purposes. As such, they are not designed in a way that allows them to separate the consumption related with space heating and preparation of DHW. It is therefore necessary to introduce a model of the daily DHW consumption profiles in order to 1) improve the predictive performance of the model, 2) avoid neglecting variation in the DHW component of the district heating measurements which could potentially affect the estimates of the remaining building-specific parameters, and 3) to distinguish between inflexible and flexible demand. While hot water tanks can be charged in a flexible manner as demonstrated in [10], the DHW production in the dwellings of this study is handled with a flow heat exchanger, thus rendering the DHW consumption inflexible. By separating consumption for DHW and space heating, the model thereby also distinguishes between potentially flexible and inflexible consumption.

Previous studies have proposed various methods for separating the two components. Bacher et al. [47] separated DHW and space heating by using a kernel smoother to identify peaks in 10 minute measurements. However, this method would not work using the hourly smart-meter data available for this study, since the volatility of the DHW consumption is essentially averaged out to an extent where distinguishing between space heating and DHW consumption is infeasible. Burzynski et al. [48] interpreted district heating consumption measured outside the heating season (during summer months) as domestic hot water consumption alone and assumed this consumption pattern to apply throughout the year. While this approach in principle would work despite the low temporal resolution of the smart meter measurements, using data collected during the summer months makes the process of inferring DHW consumption prone to any errors caused by differences in consumption levels between the warmer summer period and the rest of the year – e.g. due to holidays or an increased frequency of showering due to the generally warmer weather.

Because of the above-mentioned limitations of the existing methods, we chose to model the DHW consumption for each building directly and infer it in parallel with the inference of the building-specific parameters. Assuming significant differences in use patterns between weekdays and weekends, two distinct DHW daily profiles of 24 hourly values were inferred for each building; one profile for weekdays and one profile for weekend days, respectively. The model relies on two elements to describe the DHW consumption of a given house on a given day: a normalized shape-profile ($\theta_{\text{DHW},d}$) and a scaling factor (E_{DHW}) denoting the average daily district heating consumption for preparation of domestic hot water. The underlying assumption was that the

average annual DHW consumption per occupant (V_{occ}) amounted to 15 m³ hot water [49], which was assumed evenly distributed across all days of the year. The daily energy consumption for preparation of DHW (E_{DHW}) is given in kWh by Eq. (5).

$$E_{\text{DHW}} = \frac{c_{p,\text{water}}\rho_{\text{water}}V_{\text{occ}}\theta_{\text{occ}}(T_{\text{DHW}} - T_{\text{mains}})}{365} \quad (5)$$

where θ_{occ} denotes the inferred number of occupants in the building, while $c_{p,\text{water}}$ and ρ_{water} denote the thermal capacity and density of water at 30°C, respectively. Finally, a mains water temperature (T_{mains}) of 10 °C and a DHW draw-temperature (T_{DHW}) of 55 °C were assumed. E_{DHW} was distributed onto each hour within a day by the inferred DHW profile ($\theta_{\text{DHW},d}$) associated with that type of day in accordance with Eq. (6):

$$\Phi_{\text{DHW},t} = E_{\text{DHW}}\theta_{\text{DHW},d,hr} \quad (6)$$

where $\theta_{\text{DHW},d,hr}$ describes the share of E_{DHW} within a particular hour (hr) on either a working day or weekend day (d). There are two main factors contributing to the inferred shape of the DHW profiles ($\theta_{\text{DHW},d}$), namely the prior information introduced by us, and the repeated daily patterns observed in the measurement data. This use of repeated patterns may result in other phenomena than DHW consumption being absorbed by the DHW model. Examples of these are routine venting (e.g. each morning) or heating set points which are scheduled through home automation systems. The latter violates one of the core assumptions of the statistical calibration framework (a constant temperature set point), and buildings whose consumption data is characterized by strongly repetitive consumption patterns should either be excluded from the analysis or calibrated under other assumptions. Phenomena such as routine venting, which may be more difficult to identify due to their relatively limited impact, are absorbed by the DHW model in its current implementation. However, this is not considered a critical issue since one of the main purposes of the DHW model is to distinguish potentially flexible and inflexible demand. Since increased space heating consumption due to venting is not considered flexible, it is a desired behaviour of the calibration to assign this consumption to the DHW model.

3 Case study

The urban residential neighbourhood selected for the case-study is depicted in Figure 5. The neighbourhood consists of 206 detached single-family houses in Aarhus, Denmark, all located in a hydraulically well-defined area in the city district heating network, i.e. all consumers are supplied from the same point in the distribution grid located in the lower left corner of the figure.

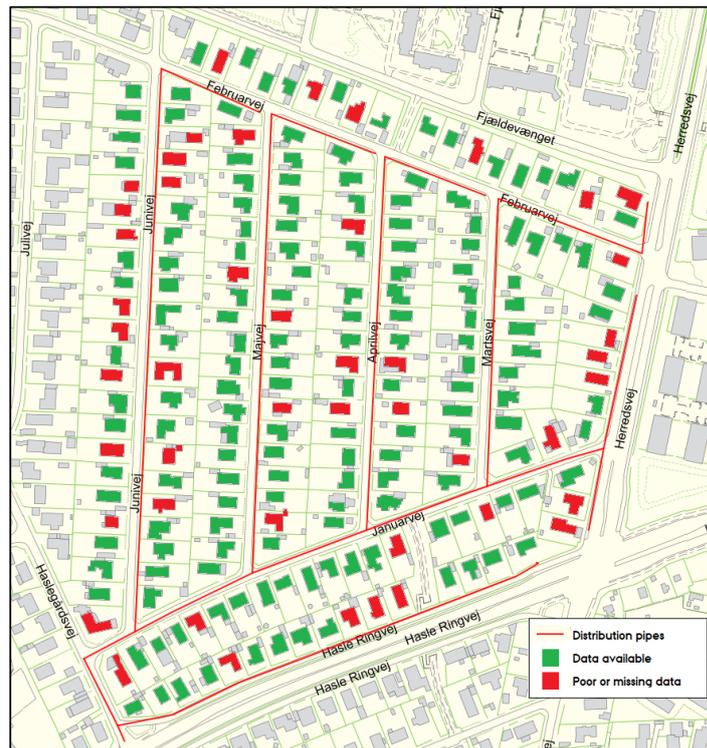


Figure 5 The urban neighbourhood used as case-study. All buildings within the neighbourhood are supplied with district heating from the same point in the grid.

Each building has a smart meter providing measurements of district heating consumption in a truncated hourly kWh-resolution, i.e. in an unrounded state and without decimal points. However, a total of 47 out of the 203 dwellings in the neighbourhood (marked in red in Figure 5) were excluded from the case study due to one of the following three issues: 1) some/all consumption data was missing (12 buildings), 2) data indicated night setback heating control (16 buildings), and finally 3) odd heating patterns perhaps caused by the presence of secondary heating systems (e.g. wood-fired stoves) or frequent occupant intervention of temperature set points (19 buildings). The reason why buildings with the second and third issues was excluded was because they violate the assumption of static set points that is made in the calibration of building models (see section 2.1). Figure 6 depicts the aggregated consumption of the remaining 159 houses for the months of January and February 2017 (along with the weather conditions for the same period) and, as such, serves as an example of the available consumption data used for generating the

UBEM. It is evident that the heating profile is characterized by relatively fast daily fluctuations as well as slower fluctuations exhibiting a large dependency on especially the external air temperature.

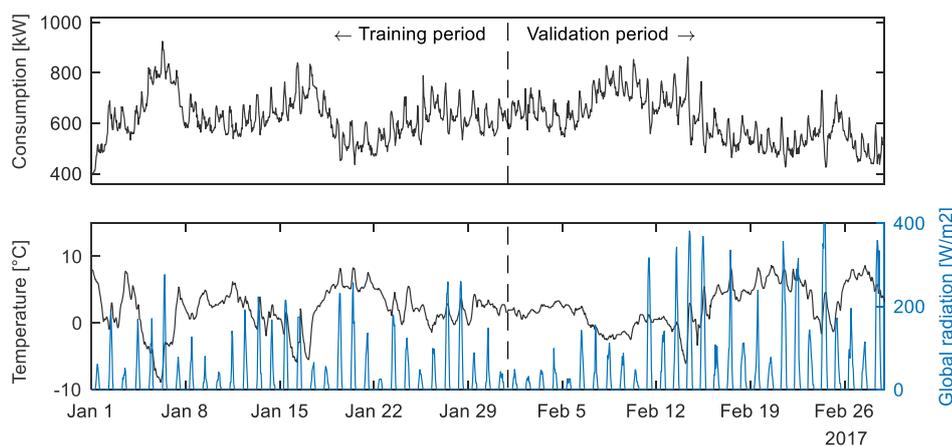


Figure 6 Aggregated DH-consumption of urban district (only the 159 buildings marked in green on Figure 5) and weather data for the training period (January, 2017) and the validation period (February, 2017)

In the following sections we apply the proposed modelling methodology outlined in section 2 to obtain a UBEM model of the urban residential neighbourhood. Section 3.1 presents the choices made in relation to setting up the calibration algorithm (inference), the selection of calibration parameters, and specification of their respective priors. Section 3.2 presents an evaluation of the obtained UBEM model both on individual building level and urban level. Finally, the model is applied in a DR scheme in section 3.3.

3.1 Inference

An approximation of the joint posterior probability distribution for all model parameters given the data can be obtained using one of several Markov Chain Monte Carlo (MCMC) based methods. These methods are characterized by their ability to sample from particularly high-dimensional parameter spaces. The methodology presented in this paper relies on an implementation based on the Metropolis algorithm [34], which from a randomly selected starting point in the parameter space walks randomly through (i.e. samples from) a multi-dimensional Markov chain – a stochastic process satisfying the Markov property – to approximate the joint posterior distribution. When converged, the Metropolis algorithm produces an unbiased mapping of the posterior probability density distribution. The part of the algorithm which allows it to focus on the regions of the parameter space of high probability, and eventually converge to a stationary estimate the posterior density function, is the fact that not all steps of the random walk are successful. Whether a step is rejected or accepted is determined by the ratio between the joint posterior probability of the current draw and the newly proposed draw, respectively. The

computation of probabilities are often carried out in the log-domain to ensure numerical stability. Each proposed step of the random walk in the parameter space is done by drawing from a so-called proposal distribution. For this, the Metropolis algorithm apply marginal normal distribution (or a multivariate normal distribution) centred at the algorithm's current position in the parameter space. It is essential to tune the parameters of the proposal distribution (i.e. the variance for one-dimensional sampling, and the covariance matrix for multi-dimensional sampling) in order to ensure a fast convergence and efficient sampling from the posterior [50].

The building model parameters (θ_B in Eq. (1d)) and the DHW model parameters (θ_{DHW} in Eq. (1e)) were kept separate in the implementation of the Metropolis algorithm, and therefore each had their own proposal distributions. Due to differences in how the building-specific parameters and the DHW-related parameters acted upon the model, the tuning of the two jumping distributions were carried out in slightly different ways. The building model parameters were all proposed from a multivariate normal distribution with zero-mean and covariance matrix Σ_B . The tuning of this jumping distribution included a full covariance adaptation including the correlation between parameters [50]. Since the parameters of the building models may be significantly correlated, the tuning of the parameter-correlation entries of Σ_B ensured a much more efficient sampling and quicker convergence than using a jumping distribution which neglects parameter correlation. In spite of this, tuning of the covariance matrix of the DHW model's jumping distribution (Σ_{DHW}) did not involve tuning of the off-diagonal elements of the matrix. The reason for neglecting the correlation was that the DHW model describes a normalized consumption profile, which after each proposed change required a post-processing of the proposal in the form of a renormalization of the entire profile. This post-processing violated the assumptions made in the covariance tuning used for the building model parameters.

After a number of iterations (jumps) the algorithm converges to a stationary state in the high-probability region of the parameter space. Until this happens the algorithm is considered cold, and all samples are discarded since they may still be influenced by the random point in the parameter space where the algorithm was initialized. The algorithm is considered warm once the marginal Markov chains of all parameters have converged – after which further sampling contributes to the approximation of the posterior distributions. The potential scale reduction factor (PSRF) was used to indicate convergence (further elaborated on later) [51]. For each building, three separate instances of the algorithm were run for 10,000 iterations – out of which the first 8,000 iterations were discarded. The three remaining batches of 2,000 samples were joined and used to form an empirically based posterior distribution. For further details on the Metropolis algorithm we refer to the original work by Metropolis et al. [34]. The following sections present the chosen calibration parameters and their associated prior distributions.

3.1.1 Calibration parameters

Using physics-based building energy models (BEM) to represent performance of an existing building requires calibration of a range of user-defined input parameters. Standard practice for this task is to assume fixed values for input parameters for which prior knowledge are less characterized by uncertainty, while only calibrating values of the most uncertain parameters. This

reduces issues related to the identifiability of the parameters – especially in cases where available data is of limited quantity or quality. Table 1 lists five building-related parameters (θ_B) that were chosen for calibration in this study, while the rest were fixed at values considered suitable for the type of dwelling used in the case study (see Appendix 1 for details). The selection of these specific five parameters for calibration was based on a compromise between the sensitivity of the model output variability to the specific parameter (and thus the identifiability of the parameter), and the relevance of the parameter in light of the intended application of the model (see [33] for further details on parameter selection).

Table 1 Model parameters selected for Bayesian inference (θ_B)

Abbreviation	Description	Unit	
1	WFR	Window-to-floor ratio	[–]
2	q_{inf}	Infiltration rate (at 50Pa)	$\left[\frac{1}{s \cdot m^2}\right]$
3	$U_{envelope}$	U-value for roof and façade walls (assumed equal)	$\left[\frac{W}{K}\right]$
4	Occ	Occupant density	$\left[\frac{m^2}{occupant}\right]$
5	C_m	Thermal capacity of construction elements	$\left[\frac{kJ}{m^2 \cdot K}\right]$

The first three parameters in Table 1 are related to how weather conditions (external temperature and solar radiance) affect the heat balance of the building. The occupant density (used to derive θ_{occ}) determines the impact of occupants on internal heat gains and scales the DHW consumption. Finally, the effect of the thermal mass in the ISO 13790 standard models is governed by two parameters; the effective thermal capacity (C_m) and the effective mass area (A_m). ISO 13790 proposes five classes of thermal mass in buildings ranging from very light to very heavy - each class with its respective values for the thermal capacity and effective mass area. To avoid calibrating both of these parameters, we chose to couple them by a simple piecewise linear relationship for the effective mass area as a function of the thermal mass as depicted on Figure 7, and only calibrate the thermal capacity (C_m).

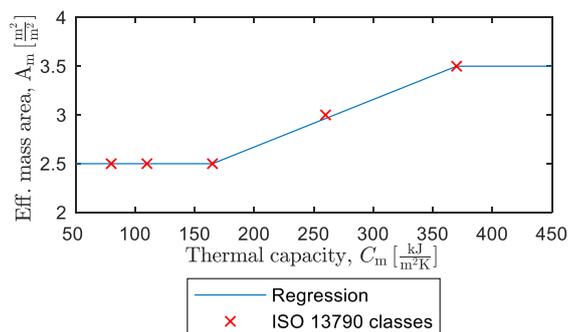


Figure 7 Regression used to couple the thermal capacity (C_m) and the effective mass area (A_m) of ISO 13790.

3.1.2 Specification of priors

The marginal prior for the WFR was specified as a beta-distribution according to Eq. (7), while the priors for the remaining parameters were specified as gamma-distributions according to Eq. (8).

$$WFR \sim \text{Beta}(5,25) \quad (7)$$

$$[q_{\text{inf}}, U_{\text{envelope}}, Occ, C_m, \theta_{\text{DHW}}] \sim \text{Gamma}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (8)$$

In Eq. (8), the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ contain the shape and scale parameters for each parameters prior gamma distribution. The specific values for each parameter prior distribution are listed in Appendix 1 while Figure 8 depicts the distributions for the BEM parameters.

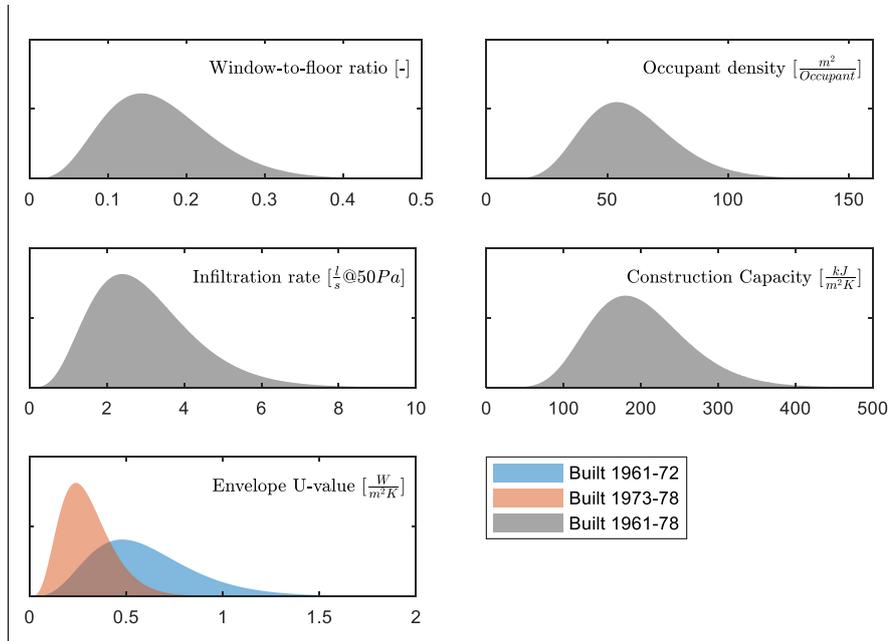


Figure 8 Prior-specification of building-specific parameters (θ_B). The prior of the WFR (top graph) is specified as a Beta distribution, while all other priors were specified as Gamma distributions.

The prior describing the insulation level of a given building to be modelled (envelope U-value) was determined by the construction year according to the Danish archetypes identified in the TABULA-project [52]. The priors for the other parameters were assumed identical across all buildings no matter construction year. The prior for the infiltration rate was based on air tightness measurements of multiple dwellings [53]. It is likely that some of the case buildings to be modelled have undergone minor refurbishments since they were built, but we have no reliable information on this. To account for this uncertainty, we specified fairly broad prior distributions for both the envelope U-values and the airtightness of the building. The prior for the occupant

density was based on statistical estimates [54], while the prior on the thermal capacity was specified to reflect the construction classes of ISO 13790 [35].

The shape of the priors for the two DHW profiles (θ_{DHW}) was based on the average measured DHW consumption for a sample of 107 British dwellings [55]. The prior for each hour was specified as a gamma-distribution parameterized in a way such that the mean value of the distribution coincided with the (normalized) profile from [55]. Since the hourly priors are specified for the normalized profiles, they describe the probability of a given share of the daily consumption falling inside each given hour. Since [55] does not provide separate results for weekends and weekdays, the same prior specification was used for both of the profiles. Figure 9 shows a contour of the prior distributions for each hour of the day, where the intensity of the contour indicates the probability, while the x and y-coordinates indicate the hour number and share of daily consumption, respectively. Here, it should be noted that, despite the continuous appearance of the contours visualizing the distributions, the contour at a given x-coordinate (hour) describes the marginal distribution of that specific hour. The distribution related to a given hour of the day, say $x = 19$ (depicted), is totally separate from the marginal distribution specified for the 18th hour of the day ($x = 18$).

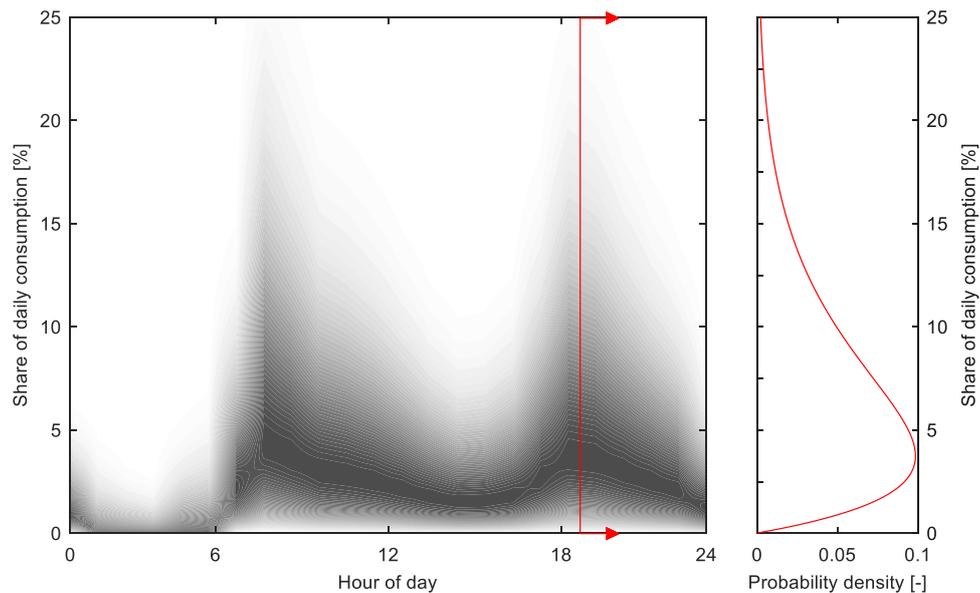


Figure 9 Contours of marginal prior distributions for both DHW profiles (workdays/weekends). The contour at each x-coordinate depicts the marginal distribution for that specific hour of the day – see highlighted hour.

Finally, the last parameter of the statistical framework outlined in section 2.1 is the prior distribution of the standard deviation σ describing the residual errors (ϵ) of Eq. (1a). Here, the prior distribution was chosen as the half-Cauchy distribution with mode 0 and a scale value of 0.25, thus producing a distribution that favours small values of σ .

3.2 Model performance

While the intended application of the obtained UBEM model is analysis on an aggregated level, the phenomena that enables utilization of flexible consumption require the individual BEM to describe the involved energy consuming processes with sufficient accuracy. Therefore, an evaluation of the performance of individual BEMs is a necessary part of interpreting the validity of the bottom-up UBEM model. Section 3.2.1 evaluates the UBEM model on the scale of the individual BEM before the aggregated scale is evaluated in section 3.2.2.

3.2.1 *Individual buildings*

The output of the MCMC-based parameter inference are approximations of the posterior distributions for the calibration parameters. The validity of these posterior distributions rely heavily on the convergence of the MCMC algorithm to a stationary state within the solution space of high probability. A method for quantifying the convergence of MCMC algorithms is to run the algorithm multiple times and use the part of each chain which is assumed to have converged to compute the Potential Scale Reduction Factor (PSRF) [51]. The PSRF estimates the reduction of the scale (uncertainty) of the posterior distribution achievable if the number of MCMC draws were increased to infinity. As such, PSRF values near 1.0 indicate that the algorithm has converged to a stationary distribution. PSRF values below 1.2 can generally be interpreted as an approximate convergence [56]. The PSRF was calculated for all of the obtained marginal distributions for the parameters of each individual BEM. The model parameter posterior estimates achieved PSRF values below 1.1 (highly converged) for all but two BEMs that achieved values below 1.2 (approximate convergence). Convergence alone, however, does not imply that the resulting models have good predictive performance. The predictive performance was therefore evaluated using the normalized mean bias error (NMBE) and the coefficient of variation of the root mean square error (CVRMSE) as proposed by ASHRAE guideline 14 [57]. The NMBE indicates whether the model is accurate on average by describing the bias of the model output as a percentage of the mean value of the measurements. The CVRMSE gives the sample standard deviation of the prediction errors, also normalized by the mean of the dataset, which (contrary to the NMBE) can be interpreted as the ability of the model to correctly predict variation in the data. The CVRMSE assumes only positive values, whereas the NMBE may be both negative and positive. Both metrics indicate higher performance as the value of the metric approaches zero. For calibration data of hourly resolution, the maximum values recommended by ASHRAE [57] are $\pm 10\%$ for the NMBE metric and 30% for the CVRMSE metric. The two metrics were computed by comparing the output of a model parameterized with the MAP parameter estimates to measurements from the training and validation periods indicated on Figure 6. Figure 10 presents the performance of the remaining 159 models to measurements from the training and validation periods indicated on Figure 6. The performance of the five buildings highlighted in red and denoted with letters A–D are further investigated and depicted in Figure 11.

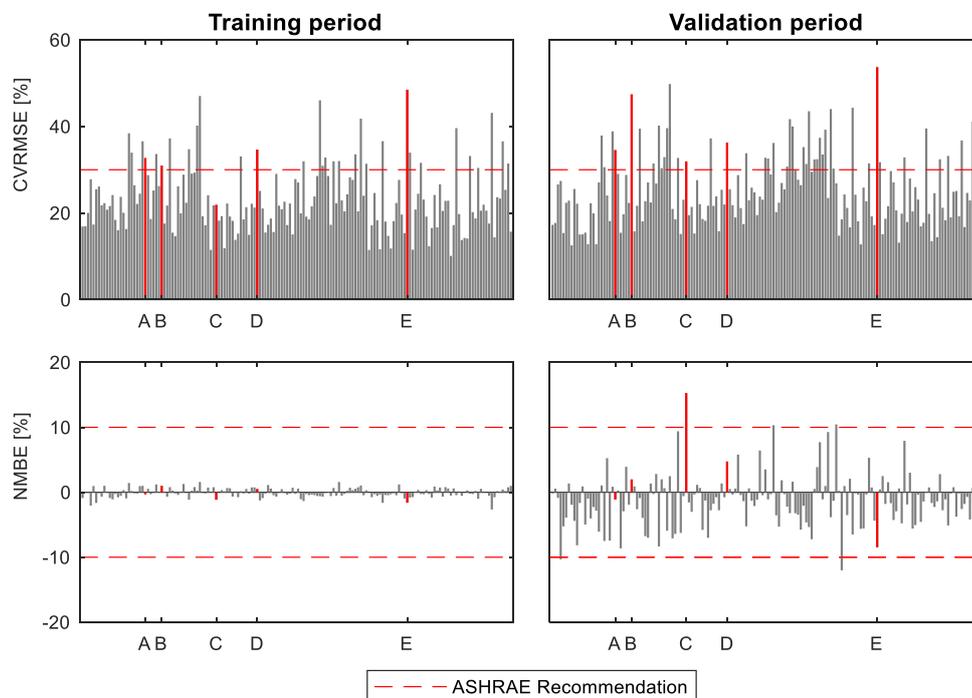


Figure 10 Performance metrics CVRMSE and NMBE for each building model parameterized with the MAP parameter estimates. Time series of the five highlighted buildings (A-E) are presented in Figure 11.

The vast majority of building models achieve an NMBE within the limits suggested by ASHRAE: All models pass on the training period while only five models (3%) were not able to pass on the validation data set. It is noted that a significant number of models actually had higher CVRMSE values than the recommended 30%: 31 models (19%) when evaluating on the training period and 43 models (27%) for the validation period. However, inspection of measurements time series and model predictions revealed that the metrics alone were not suited for deciding whether or not a given model should be discarded. In some cases, high CVRMSE values were caused by extreme consumption peaks that were relatively far-in-between, and thus not represented well by the inferred DHW profiles that describe the average daily DHW-consumption. In other cases it would seem that the building had been vacant for extended periods of time, which were characterized by a lack of DHW consumption and sometimes also a reduced consumption for space heating (lowered temperature set points). Finally, the truncated nature of the smart-meter data used for the comparison is also a cause of discrepancies between measurements and model output.

Figure 11 depicts time series of the five buildings (A-E) highlighted as red bars in Figure 10. These buildings serve as examples of the conditions that in many cases resulted in violation of one or both of the two ASHRAE metrics. While both the calibration and the computation of the two metrics were carried out using hourly values, we present the time series in a 3-hourly resolution for readability. Parts of the time series where the consumption deviated from the

expected level have been highlighted and (even though we do not know the exact causes) labelled with plausible explanations of the observed deviation.

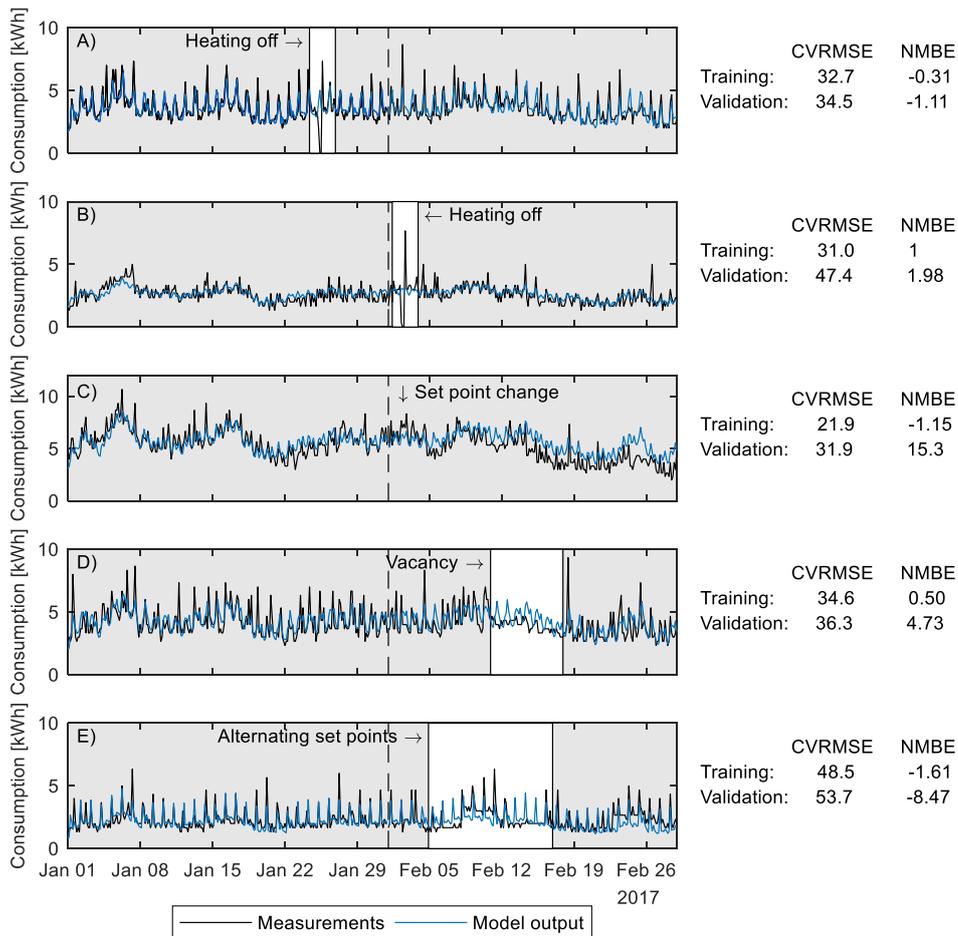


Figure 11 Evaluation of individual building models that performed poorly in terms of CVRMSE/NMBE through comparison of time series. The dashed line separates the training and validation period. Time series of 3-hourly averages were used for readability. The NMBE and CVRMSE metrics are indicated next to the time series.

In addition to the various phenomena indicated in Figure 11, a recurring cause of mismatch between the predicted and measured consumption was peaks that are likely related to DHW consumption. Only a small number of the buildings (e.g. B in Figure 11) did not feature such peaks – a likely explanation being that they are fitted with domestic hot water tanks that spread out this consumption.

Even though separate weekday and weekend DHW profiles were inferred for each dwelling, the DHW model assumes these profiles to apply equally to all of the days for which they were defined (i.e. weekdays or weekend days). Therefore, in cases where large peaks (presumably showering) are spaced randomly with one or multiple days in-between, the inferred DHW profile tend to indicate peaks of average size on all days. While this discrepancy contributes significantly to the value of CVRMSE, we do not consider it critical to the validity of the model in the context of groups of buildings, where the stochastic behaviour of occupants from many different buildings to some degree is averaged out. Therefore, although many building models did fulfil the ASHRAE recommendation, the suggested limits proved perhaps too strict to act as a hard requirement for this application, where the measurements were both truncated and heavily influenced by effects of stochastic occupancy. The inferred DHW profiles for all 159 buildings are shown in Figure 12. Both normalized and absolute profiles are presented as the former is useful for comparing the distribution of DHW consumption between buildings with varying overall consumption, while the absolute profiles describe the actual impact on the grid.

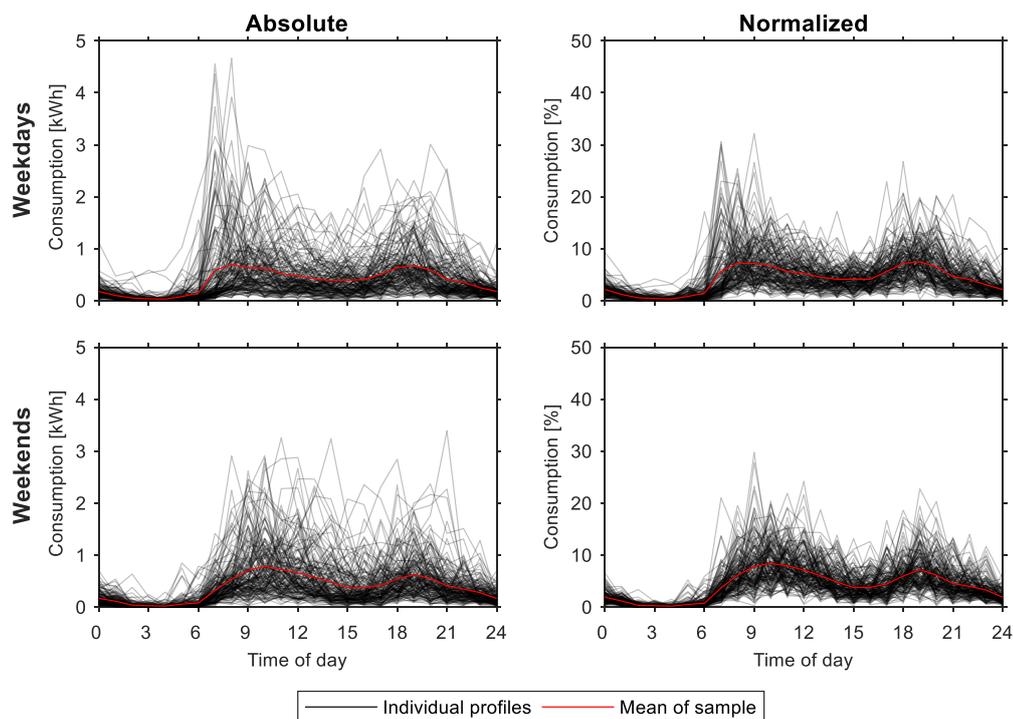


Figure 12 Inferred DHW profiles in absolute and normalized terms along with the average profile of the entire sample. First row presents weekday profiles while the second row presents the profiles of weekend days. In the normalized profiles, each entry represent the share of daily consumption associated with that hour.

Not surprisingly, comparison of the mean profiles indicate that the largest difference between weekdays and weekends is the pickup in consumption in the morning hours: On weekdays, the consumption has a steep incline starting at 06:00 before it peaks in the hour 07:00-08:00. Weekend days are characterized by a smoother pickup in consumption and a peak which happens two hours later than the weekday peak in the 09:00-10:00. Although there currently is no way of evaluating the validity of the inferred DHW-profiles, the profiles are considered plausible since the observed differences (obtained despite the use of identical prior information) between weekend and weekday profiles match both our own expectations and agree with previous research findings [58].

3.2.2 Urban scale performance

The performance of the UBEM of the 159 case-buildings is depicted in Figure 13, where the predicted consumption is compared to measurements. The aggregation of data from the entire pool of buildings improves the readability, thus allowing us to show these results in hourly resolution.

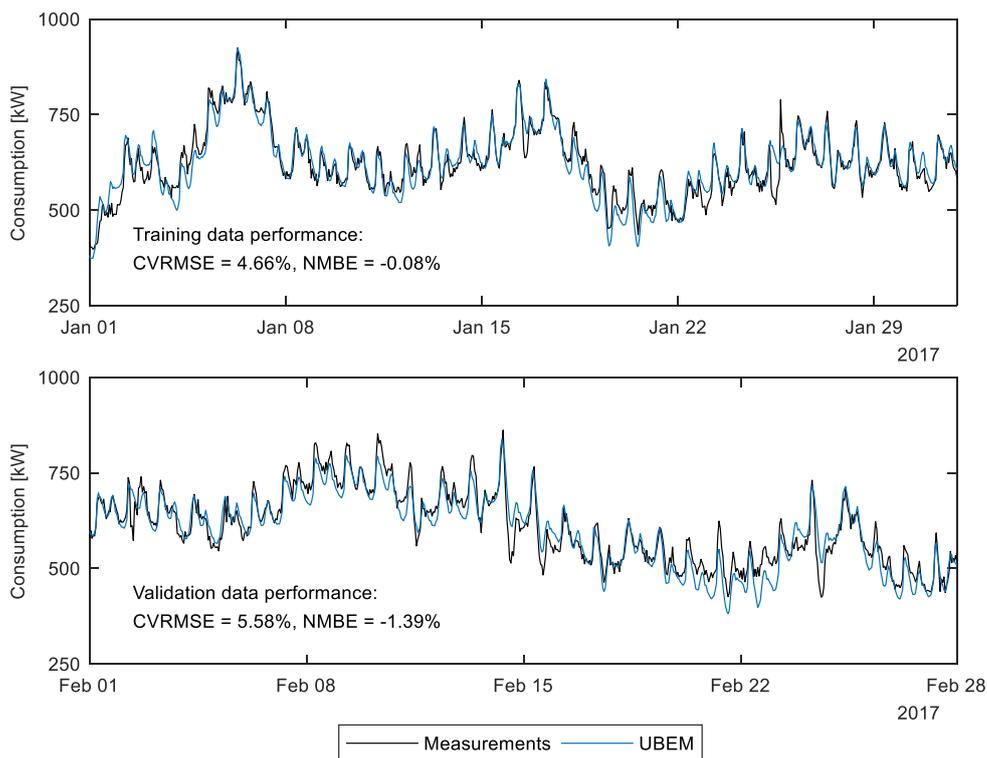


Figure 13 Performance evaluation of the aggregated UBEM (159 buildings) on the dataset used for training (top) and on a previously unseen validation dataset (bottom). Both datasets are in hourly temporal resolution. CVRMSE and NMBE metrics for each period are indicated on each graph.

Visual inspection of the time series suggest an overall good performance of the model in both periods which is supported by the CVRMSE and NMBE metrics. Furthermore, it is clear that the UBEM is able to accurately describe the majority of the daily peaks in consumption, thus indicating that the aggregation of multiple buildings have indeed lowered the issues related to DHW peaks that were seen on the scale of individual building in the previous section. A drop in consumption followed by a pick-up in consumption in the hours afterwards is seen on January 25. This coincides with the highlighted drop in the consumption of “dwelling A” in Figure 11, and may therefore suggest that multiple buildings were taken off the grid temporarily possibly due to pipe maintenance.

While top-down UBEMs may achieve similar performance when calibrated with the aggregated data directly, the advantage of the bottom-up approach is that the diversity of the buildings producing the aggregated consumption profile is preserved. This characteristic of the bottom-up UBEM that allows it to be used as a basis for more detailed analyses e.g. investigating the combined response of the building stock to proposed DR initiatives. The following section provides an example of how the UBEM may be utilized for analysis of DR initiatives on the urban level.

3.3 Utilizing space heating for flexibility

As indicated by previous studies (see the introduction of this paper), the thermal inertia inherent to buildings can be exploited to allow the consumption dedicated towards space heating to be shifted in time without significantly impacting the indoor thermal climate in the buildings. In this case study we evaluate whether the utilisation of this phenomenon can be exploited in a DR scheme tailored to achieve peak reductions in the aggregated consumption profile of an urban neighbourhood. Strategic peak reductions benefit utility companies by allowing them to optimize the daily operation, e.g. by preventing firing up cold boiler plants during peak hours. An even larger potential may be associated with the ability to address congestion issues in distribution networks and reduce the needed standby generation capacity. A way to incentivize building owners to engage in DR is through time-varying energy prices, e.g. by offering cheaper prices at off-peak times than during peak-times. This approach to demand response is generally referred to as *indirect price-based DR* [59] due to the fact that the decision process lies entirely with the consumers themselves while the incentive (i.e. the time-varying prices) is determined by the utility company. Figure 14 presents the principles of price-based DR schemes in the context buildings; A) and B) indicate the fluctuating demand targeted by the DR scheme and the resulting price increase during a specified peak period, respectively, while C) and D) depict the actions made by e.g. a building energy management system in order to minimize the economic expenses of the house owner. Here, the thermal mass of the building is charged by raising the room temperature prior to the peak, thus allowing the building to become autonomous for the whole or part of the duration of the peak.

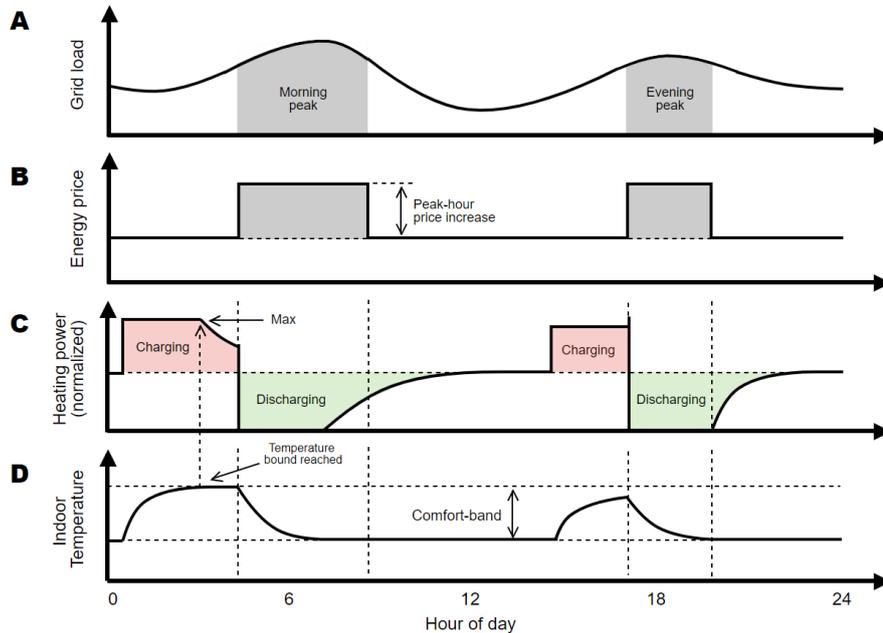


Figure 14 Principle of the price-based demand response scheme evaluated in this study.

One of the approaches for implementing price-based DR is the use of model predictive control, which rely on a model of the building to predict the optimal way of operating the building, given a certain objective function [12]. This objective function could reflect a wish to minimize economical expenses of the home owner, maximize utilization of renewable energy, or minimize the impact of the building on the performance of the grid. The advantage of MPC schemes is that they use a model of the building to identify the exact course of actions that yields the best result in terms of the specified objective (assuming that the model is a sufficiently accurate representation of the building). In practice, MPC schemes do this by solving a control optimization problem which incorporates the dynamic behaviour of the building and relevant external factors such as weather conditions. These control problems may include a variety of constraints which are relevant for the operation of the building such as the allowed range of indoor air temperatures, maximum air temperature rate of change, as well as system-related constraints such as maximum available heating output. An example of a simple MPC optimization problem is presented in Eq. (9a-9e):

$$\underset{u}{\text{minimize}} \quad \sum_{t=1}^{n_t} c_t \cdot u_t \quad (9a)$$

$$\text{subject to} \quad x_{t+1} = \mathbf{A}_t x_t + \mathbf{B} u_t + \mathbf{E}_t d_t \quad (9b)$$

$$T_{i,t} = \mathbf{C} x_t \quad (9c)$$

$$0 \leq u_t \leq P_{\max} \quad (9d)$$

$$T_{\min} \leq T_{i,t} \leq T_{\max} \quad (9e)$$

where Eq. (9a) specifies an economically based objective function describing the objective of minimizing the product of a vector of time-varying energy costs (c_t) and a vector of the energy consumption for each control unit time step (u_t). The latter is the decision variable of the control problem – i.e. the variable that may be manipulated by the control scheme. Since DHW consumption is considered inflexible, u_t corresponds to the consumption aimed towards space heating only. Eq. (9b-9e) are the constraints of the optimization problem solved by the control unit. Here, (9b) describes the dynamic behaviour of the building through the building model and (9c) relates the internal states of the building model (x_t), to the indoor temperature output of the model (T_i). The remaining equations are used to specify constraints relevant for the operation of the building; (9d) describes the maximum heating rate the heating system can deliver and (9e) limits the indoor temperature to be within some user-specified threshold.

In our case, the range of acceptable indoor temperature for the MPC was chosen to be between $T_{\min} = 20^\circ\text{C}$ and $T_{\max} = 24^\circ\text{C}$. The maximum power output achievable for each building P_{\max} was assumed to follow the Danish design conditions, where the heating system should be able to maintain an indoor temperature of $T_i = 20^\circ\text{C}$ at outdoor temperatures of $T_{\text{exterior}} = -12^\circ\text{C}$. This design load was approximated by comparing the indoor-outdoor temperature difference (ΔT) of the coldest hour of the measurement data and then scaling the corresponding (inferred) space heating consumption (disregarding DHW component) for that hour to match the $\Delta T = 32^\circ\text{C}$ temperature difference used for determining the needed size of the heating system in Danish residential buildings. Here, we disregard the discrepancy between the dynamic data used for this extrapolation and the stationary conditions assumed in design calculations. This approach suggests that buildings which have undergone retrofits also have downgraded the capacity of the installed heating system which is considered very unlikely. Both of these factors contribute to a pessimistic estimate of the power output P_{\max} available in the buildings.

A reference scenario was defined in which the buildings were operated in the most energy efficient way, meaning that the indoor temperature was maintained at T_{\min} throughout the simulations to minimize heat loss. Figure 15 (top) depicts the aggregated heating consumption of the 159 houses in the reference scenario for the period 3-8 January with external air temperatures as depicted in Figure 15 (middle). The marked consumption peaks (peak 1 and peak 2) occurring on January 4 and January 6, respectively, were chosen as targets for a DR scheme seeking to lower the daily fluctuation. The two DR events were 48 hours apart from one another to avoid

“spill over” from one DR event to the next, since such interaction complicates the task of determining the prices that yields a suitable response from a larger pool of buildings. The incentive for DR was the time-varying energy prices (c_t) which was increased with different rates during predefined peak periods as depicted in Figure 14B. The duration of the peak periods with increased prices was set to four hours as marked in Figure 15 (bottom). The same part of the figure depicts the response from the building stock to the seven different peak period price scenarios.

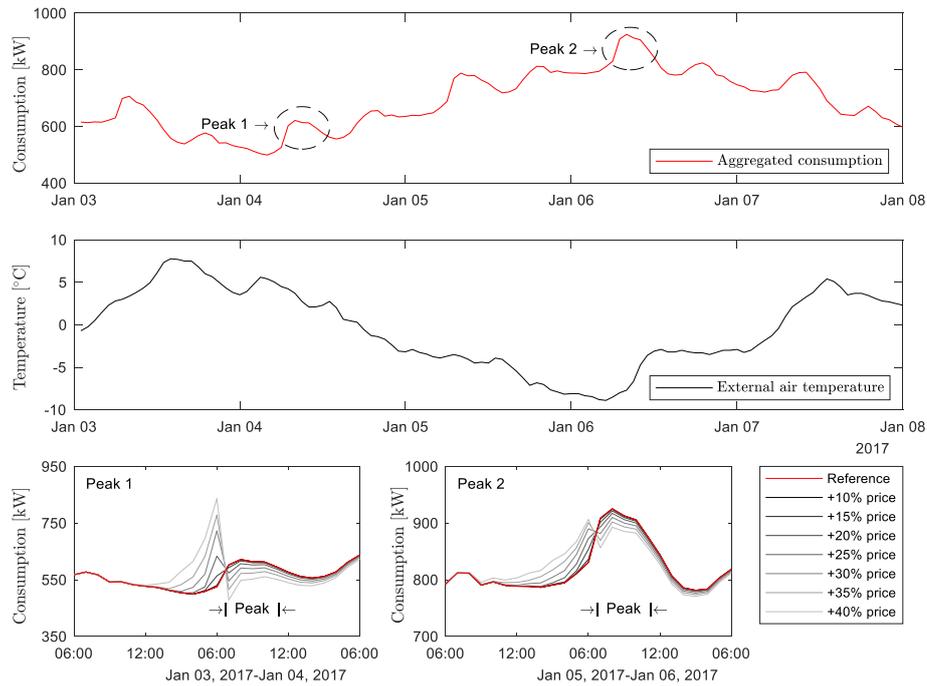


Figure 15 Top: Aggregated heating consumption from the UBEM. Middle: External air temperature for the period. Bottom: The response of the UBEM obtained by solving the optimization problem of Eq. (9a-9e) with different peak prices during the four hours of the peak period.

Figure 15 (bottom) indicates critical issues associated with the charging and discharging as a result of the DR scheme. During “Peak 1”, where external temperatures between 2°C and 5°C meant that over 50% of the heating system capacity in each building was available for charging, the amount of load shifting increased quickly as the incentive (i.e. the energy price during the peak) increased. Since the efficiency associated with load shifting decreases as the length of the load shift increases, the majority of the load shifting occurred in the hour just before the peak of the reference scenario. While the same behaviour would be expected from “Peak 2”, the effect in this DR event was less evident due to colder external temperatures ranging from -8°C to -4°C (the coldest day of the dataset used for model calibration and validation), which meant that less than 10% of the heating system capacity was available for charging. While a high efficiency is

desirable, the simulations indicates that concentrated charging as a consequence of the DR scheme leads to the formation of new peaks in the hours prior to the peaks in the reference scenario. These new peaks increased in size as the peak price increased because it then becomes economically viable for more buildings to engage in the DR scheme. Another critical issue indicated by the figure is that the majority of the stored energy is discharged at an early stage of the reference case peak period, thus resulting in a much lower effect in the last hours of the peak period. This is problematic since the size of the peak seen over the four hour period is then hardly reduced. Finally, the relatively fast charging and discharging of the thermal mass would result in a fast rise and drop in indoor air temperature which may be uncomfortable to the occupants of the buildings.

The critical issues that arises as a consequence of the current DR scheme suggests that modifications are necessary to achieve a more suitable response from the case buildings if the objective is to reduce peaks in the overall DH system. The MPC optimization problem was therefore expanded with the constraints shown in Eq. (10a-10b). This expansion was an attempt to avoid steep indoor air temperature fluctuations while also preventing the creation of new peaks during the charging period. The latter was obtained by ensuring that charging and discharging was evenly distributed throughout their respective time windows.

$$\Phi_{DR,t} = \Phi_{ref,t} + \Delta\Phi_{charge} \quad \forall t \in P_{prepeak} \quad (10a)$$

$$\Phi_{DR,t} = \Phi_{ref,t} + \Delta\Phi_{discharge} \quad \forall t \in P_{onpeak} \quad (10b)$$

In Eq. (10a-10b), $\Phi_{DR,t}$ denotes the consumption during the hour t resulting from DR activities while $\Phi_{ref,t}$ refers to the consumption in the reference scenario without DR. The scalars $\Delta\Phi_{charge}$ and $\Delta\Phi_{discharge}$ denote the pre-peak increase and on-peak decrease in consumption; for instance, a value of $\Delta\Phi_{charge} = 100$ indicates that a given building consumes an extra 100 W intended for storing energy in the thermal mass throughout the designated charging period.

The price which yields the optimal response from the building stock may be determined by solving a series of optimization problems. Here, we introduce the Peak-to-Valley Ratio (PVR) which serves as a performance metric for the DR resulting from a given incentive. The PVR is derived from the consumption of the entire DR event (i.e. both the charging period and the following discharging period, denoted $\Phi_{DR,event}$). The PVR is defined by Eq. (11) and describes the relationship between the peak and the valley of the period and approaches zero as the consumption profile is flattened out through DR.

$$PVR = \frac{\max(\Phi_{DR,event}) - \min(\Phi_{DR,event})}{\min(\Phi_{DR,event})} \cdot 100 \quad (11)$$

In addition to the ability of the DR scheme to reduce the daily fluctuations in demand, it is also relevant to evaluate the costs in terms of the energy losses associated with redistributing demand. Eq. (12) formalizes the load shifting efficiency used in this study, η_{DR} , which was inspired by the formulation in [6] but modified slightly to make it consistent with the economic incentive that can be generated through load shifting. The index label ‘‘peak’’ refers to hours inside the peak period, while the label ‘‘offpeak’’ refer to hours on both sides of the peak. As such, $\Phi_{DR,offpeak}$

includes both the period prior to the peak with increased consumption (charging), and the period following the peak with reduced consumption due to leftover thermal energy still being discharged.

$$\eta_{DR} = \frac{-\sum \Phi_{ref,peak} - \Phi_{DR,peak}}{\sum \Phi_{ref,peak} - \Phi_{DR,offpeak}} \cdot 100 \quad (12)$$

Figure 16 depicts various metrics describing the response of the case buildings as a function of the price-increase imposed on them during the two peaks for new simulations featuring the constraints in Eq. (10a-10b). The response is presented in terms of the two $\Delta\Phi$ variables of Eq. (10a-10b) aggregated across all buildings (top), shifting efficiency (η_{DR}) and the participation rate (middle), and PVR metric (bottom) along with the price that yielded the optimal value of the metric (red crosses). The PVRs to the left of this optimum indicate that prices are still too low to achieve the optimal amount of load shifting (if the objective is to lower the PVR), while values to the right of the optimum indicate the creation of new peaks due to excessive load shifting.

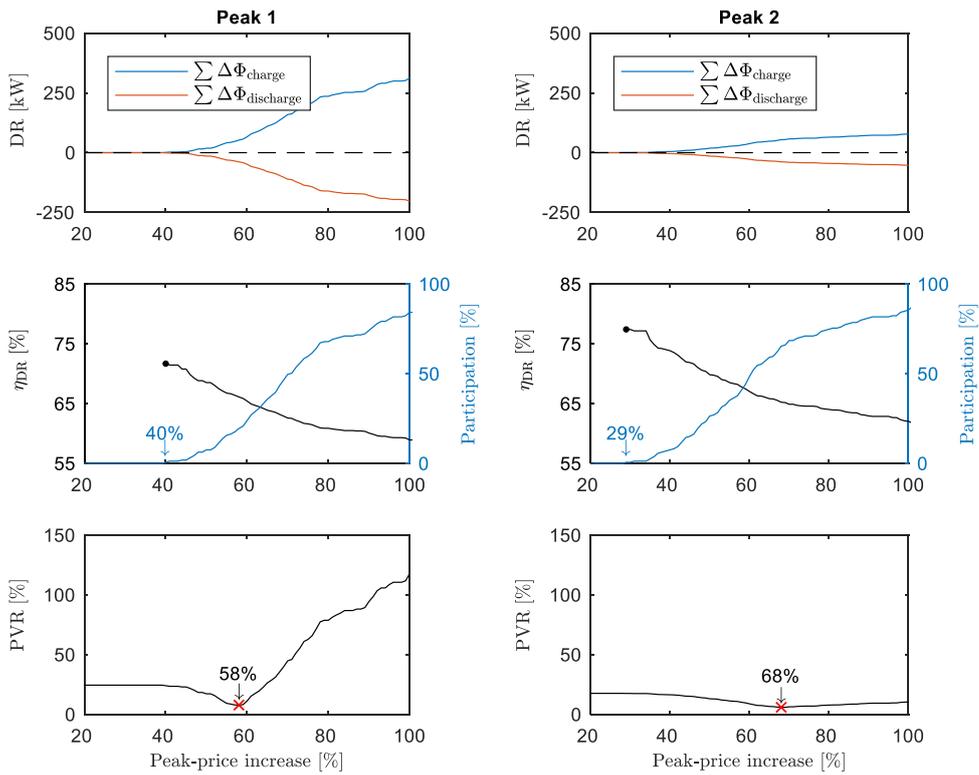


Figure 16 Metrics describing the aggregated response of the buildings of the urban neighbourhood as a function of the increase in peak energy price. Top: Pre-peak increase and on-peak reduction in consumption. Middle: the efficiency and participation rate. Bottom: Peak-to-Valley Ratio (PVR)

Figure 16 indicates several differences in how the case buildings responded to the two DR events. The main driver behind these differences was the significantly colder weather conditions in the period surrounding the second DR event. The aggregated amount of DR during Peak 1, i.e. both the positive ($\Delta\Phi_{\text{charge}}$) and the negative ($\Delta\Phi_{\text{discharge}}$), quickly grows to larger values as the peak price increases. During the second peak, the response of the case buildings was limited due to the lack of reserve heating capacity. Furthermore, the second row of graphs in Figure 16 indicate that the DR event during peak 2 was characterized by a higher efficiency than that of peak 1. This is also reflected in the different prices which caused the first building to engage in DR during each peak; 40% and 29%, respectively. The cause of the increased efficiency during the colder weather is tied to the venting model described in section 2.1.3, which assume increased natural ventilation in periods of milder weather. The higher air change rates associated with increased natural ventilation results in increased energy losses during the charging period, where indoor temperature levels are elevated to enable the flow of heat into the thermal mass of the building.

Finally, the differences between the two peaks indicated on the two upper graphs are also reflected on the PVR metric depicted on the two lower graphs. The left side of the figure indicate that there is a serious risk of generating new peaks during the first DR event if the peak prices imposed on the buildings are too high, whereas the lack of reserve capacity eliminates this risk for the second DR event. Implementing the two prices which yielded optimal PVR values for each respective peak (marked with red crosses in Figure 16) produces the aggregated demand profile depicted in Figure 17.

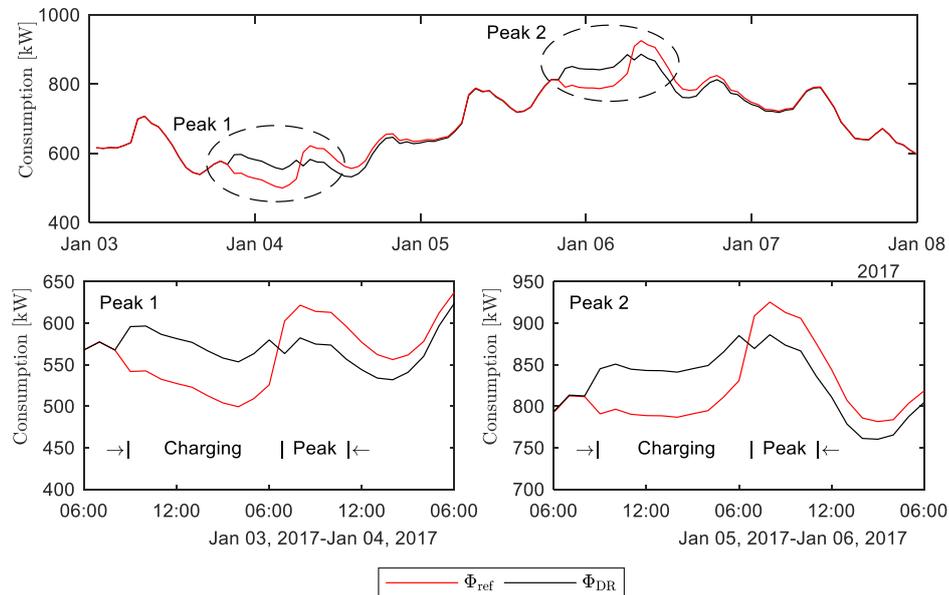


Figure 17 The response of the case buildings obtained by solving the optimization problem of Eq. (9a-9e) including the constraints of Eq. (10a-10b), and imposing the optimal increase in price indicated on Figure 16 during peak periods.

The demand profile shows that the case buildings were able to reduce their aggregated demand during the peak periods by engaging in the DR scheme. For Peak 1, the PVR was reduced from 24.5% (122 kW) to 7.8% (43 kW) while the consumption in the peak hours was reduced by 6.3%. For Peak 2, the PVR was reduced from 17.6% (138 kW) to 5.9% (50 kW), while consumption in the peak period was reduced by 4.3%. The increase in price needed to incentivize the response from the case buildings were relatively high; 58% and 68% for Peak 1 and 2, respectively. This is due to the relatively low efficiencies (η_{DR}) at which load from the two peaks were shifted; 66.1% and 65.2%, respectively, which imply that approximately a third of the energy dedicated towards load shifting is lost in the process. Here, one should note that these percentages are derived taking the reduced consumption following the end of the peak period (see Figure 17) into account. Comparison between the response depicted in Figure 17 and the responses occurring at much lower peak prices depicted in Figure 15 suggests that especially the requirement for evenly distributed DR affected the shifting efficiency and thus the size of the economic incentive needed to ensure economically feasible DR from the consumer point of view. This relationship agrees with the analysis of Reynders [60], who found that increasing the duration of a DR event significantly reduced the storage efficiency.

The need to introduce constraints related to the way that the buildings conduct DR seems to be counter-intuitive to the concept of *indirect* DR. However, such constraints seem to be necessary due to the conflict of interest described by O'Connell et al. [15] where the supply-side has a set of objectives related to ensuring efficient production and distribution of DH, while the objective of consumers is to minimize expenses. If constraints on the DR behaviour of consumers cannot be implemented, this conflict of interest may reduce the effectiveness of price-based DR. Another challenge related to price-based DR is the task of identifying the price signal that yields a suitable response from the consumers. Corradi et al. [61] investigated how models for predicting the response of consumers to time-varying prices in a power-systems context. The authors concluded that the assumption of a linear relationship between price and consumption in the presented models is an aspect in which the models could be improved, as the models currently could not describe saturations in the response of consumer. The differences in the responses seen in the case study of the present paper suggests that such capabilities would indeed be necessary. Furthermore, for thermostatic loads, such models should for obvious reasons account for the weather dependency of the price responsiveness of consumers.

Given the conflicting objectives of the supply and demand sides, a relevant topic of future research is a comparative study of the efficiencies obtained through indirect price-based DR schemes and direct incentive-based DR schemes, respectively. An advantage of the latter is that the behaviour of the consumers is not tied to variation in a broadcasted price signal, but may instead be determined through centralized optimization and coordination of participants to ensure that the DR is carried out with the supply side objectives and efficiency in mind. In this case, the incentive would not be provided through time-varying prices, but e.g. directly through energy bill discounts. Finally, the analysis of this study indicated relatively high losses associated with load shifting, which may render the use of DR unfit for optimization of daily operation of district heating networks. Therefore, another topic of future research is to evaluate the DR potential in

the type of scenarios that are used for sizing of generation and distribution infrastructure, where even modest reductions may be associated with significant investment cost savings.

4 Conclusion

This paper describes a methodology for bottom-up energy modelling of the buildings in urban neighbourhoods with the purpose of investigating the aggregated demand response (DR) potential in model predictive control (MPC) of residential space heating. The methodology uses Bayesian calibration to identify both domestic hot water draw profiles and a second order building model describing the thermodynamics characteristics of each individual building. The methodology was applied for a case study featuring an urban area in the city of Aarhus, Denmark, consisting of 159 single-family houses. The aggregated profile of hourly consumption for the neighbourhood simulated for an unseen one-month validation period indicated a high accuracy: A CVRMSE of 5.58% and an NMBE of -1.39%. The calibrated urban building energy model was then used to investigate the efficiency at which MPC of space heating for larger groups of buildings may be used for DR purposes which are beneficial to the district heating production and distribution. In this case study, the objective was to lower the daily morning peaks in the demand profile caused by domestic hot water consumption. An indirect price-based DR scheme which involved elevated energy prices during peak hours were used to incentivize the buildings to participate in DR events. MPC was used to enable individual buildings to engage in DR by exploiting the inherent thermal mass of the buildings while maintaining the indoor temperature at comfortable levels. The DR scheme was found to be unfit for the peak reduction application since it resulted in the formation of new and larger peaks prior to the original peak. To prevent this, additional constraints were then introduced in the DR scheme, which ensured a more evenly distributed response from the participating buildings. This modified DR scheme allowed the buildings to reduce two investigated peaks by 6.3% and 4.3%, respectively. The introduction of additional constraints related to the distribution of DR efforts significantly affected the efficiency at which the buildings were able to shift consumption away from peak periods. This suggests that future research exploring and comparing various DR schemes on their effectiveness and efficiency at addressing various system performance objectives is needed. The modelling methodology presented in this paper seems well-suited for such analysis.

The authors acknowledge that there are several aspects of the presented methodology which could be further improved upon. One issue to investigate is whether idealized assumptions such as a fully mixed air temperature distribution or an ideal heating system has a significant influence on the model outcome. Another issue is that data used for the calibration of individual buildings was captured under normal operating conditions, and may therefore not contain sufficient information for a trustworthy identification of all input parameters subject to calibrated. This is especially critical to the identification of the thermal capacity of the buildings. Future work is therefore recommended to investigate whether the proposed approach of using priors to reduce this issue yields models that are sufficiently accurate representations of the actual buildings. Finally, it would be interesting to investigate the effect of including indoor temperature measurements in the model calibration as well as the possibility of improving the quality of the calibration data by

imposing subtle excitation of the thermal dynamics of the buildings - e.g. through night setback strategies.

5 Acknowledgement

The authors gratefully acknowledge the support of this work from the project “READY.dk” financed by the Danish energy research and development program ForskEl and the project 'Resource Efficient Cities Implementing Advanced Smart City Solutions' (READY) financed by the 7th EU Framework Program (FP7-Energy project reference: 609127).

6 Nomenclature

Abbreviations			
DR	Demand response	MAPE	Mean absolute percentage error
DH	District heating	MCMC	Markov Chain Monte Carlo
UBEM	Urban building energy modelling	PSRF	Post scale reduction factor
MPC	Model predictive control	NMBE	Normalized mean biased error
DHW	Domestic hot water	CVRMSE	Coefficient of variance of the root mean squared error
MAP	Maximum a posteriori	PVR	Peak-to-Valley ratio
Symbols and variables			
Φ_{DH}	Consumption rate: Vector of district heating measurements	Φ_{occ}	Rate of heat gains from occupant metabolism
Φ_{sim}	Consumption rate: Simulated energy consuming process	Φ_{app}	Rate of heat gains from appliances
Φ_b	Consumption rate: Space heating component	E_{app}	Energy quantity: Annual electricity consumption for appliances
Φ_{DHW}	Consumption rate: Domestic hot water component	$A_{footprint}$	Building footprint area obtain from Building and Dwelling Register (see section 2.2)
ϵ	Stochastic component (assumed normally distributed and i.i.d.)	E_{DHW}	Energy quantity: Annual DHW consumption of a household
σ	Standard deviation of normal distribution	$c_{p,water}$	Specific thermal heat under constant pressure of water
X	Vector containing values for fixed parameters (see Appendix 1)	ρ_{water}	Density of water
θ_b	Vector of calibrated building-specific parameters	V_{occ}	Annual consumption of hot water [m^3]
θ_{DHW}	Vector of calibrated DHW-specific parameters	T_{DHW}	Assumed temperature of tapping water
W	Matrix containing relevant weather measurements	T_{mains}	Assumed temperature of mains water
t	Time (in hours) in simulation	θ_{occ}	Number of occupants of the household (calibrated)
n_t	Duration of simulation (in hours)	Σ_B	Covariance matrix of Metropolis proposal distribution of building-specific parameters
d	Categorical value describing the type of day (workday/weekend day)	Σ_{DHW}	Covariance matrix of Metropolis proposal distribution of DHW parameters
C_i	Thermal capacity of interior	α	Vector of shape parameters for specification of prior Gamma distributions
C_m	Thermal capacity of thermal mass (heavy building elements)	β	Vector of scale parameters for specification of prior Gamma distributions
T_i	Temperature of interior	c	Optimization problem: Vector of energy costs
T_s	Surface temperature	u	Optimization problem: Vector of control inputs (heating power)

T_m	Temperature of thermal mass	d	Optimization problem: Vector of disturbances (weather conditions, internal gains)
T_{exterior}	Outdoor air temperature	x	Optimization problem: States of the state space models
T_{supply}	Temperature of ventilation supply air	A, B, C and E	Matrices of state space representation of the building energy model (see Appendix 2)
Φ_{ia}	Internal gains affecting room air (interior node)	P_{max}	Optimization problem: Maximum available heating power
Φ_{st}	Internal gains affecting internal surfaces	$T_{\text{min}}, T_{\text{max}}$	Optimization problem: Boundaries of allowed temperature interval during DR events
Φ_m	Internal gains affecting thermal mass	Φ_{DR}	Vector of combined consumption rate of case-buildings in the DR scenario
A_m	Effective surface area of thermal mass	Φ_{ref}	Vector of combined consumption rate of case-buildings in the reference scenario
H_v	Ventilation and infiltration heat transfer coefficient (HTF)	$\Delta\Phi_{\text{charge}}$	Vector containing the rates at which each case-building stores energy before peak
H_w	HTF for transmission loss through windows (massless)	$\Delta\Phi_{\text{discharge}}$	Vector containing the reduction in consumption of each building during peak
H_{em}	HTF for transmission loss through opaque building envelope elements	P_{prepeak}	Vector of hours in which charging is allowed of Eq. (10a)
H_{is}	HTF between interior and internal surfaces	P_{onpeak}	Vector of hours in which discharging is allowed of Eq. (10b)
H_{ms}	HTF between thermal mass and internal surfaces	PVR	Peak-to-valley ratio of Eq. 11
b	Venting factor for scaling design air change rate	η_{DR}	Demand response efficiency as defined by Eq. 12

Appendix 1: Parameters values and priors

Table 2 lists the parameters used to parameterize the 5R2C reduced-order model used to describe the energy consumption of residential buildings in this study. Each parameter was either attributed a fixed value or given a prior distribution describing our a-priori beliefs related to the likelihood of the parameter value.

Table 2 Fixed parameter values and priors of the building energy model. * indicates value defined by ISO 13790:2008.

	Model parameter	Selected value	Prior
Geometry			
1	Length-width ratio, LWR [-]	0.50	
2	Room height [m]	2.60	
3	Window-to-floor ratio, WFR [-]	Calibrated	1961-78: <i>Beta</i> (5,25)
4	Window frame fraction [-]	0.25	
5	Shading factor (overhang, surroundings)	0.5	
6	Internal surface-to-floor ratio	4.5*	
Transmission & capacity			
7	Temp. adjustment factor (ground) [-]	0.70	
8	U-value (floors) [W/(m ² K)]	0.5	
9	U-value (basement-walls) [W/(m ² K)]	0.6	

10	U-value (walls/roof) [W/(m ² K)]	Calibrated	1961-72: $\text{Gamma}\left(5, \frac{0.6}{5}\right)$ 1973-78: $\text{Gamma}\left(5, \frac{0.3}{5}\right)$
11	U-value (windows) [W/(m ² K)]	1.6	
12	g-value (windows) [-]	0.60	
13	Thermal capacity of interior [kJ/(m ² K)]	Assumed from experimental calibration:	
		$45 \frac{\text{kJ}}{\text{m}^2\text{K}}$	
14	Thermal capacity of heavy mass	Calibrated	1961-78: $\text{Gamma}(10,20)$
15	Effective mass area [m ² /m ²]	Regression, $f(C_m)$	
16	Heat transfer coef. (mass-surf.) [W/(m ² K)]	9.10*	
17	Heat transfer coef. (surf.-air) [W/(m ² K)]	3.45*	
	Ventilation		
18	Infiltration airflow @ 50 Pa [l/s/m ²]	Calibrated	1961-78: $\text{Gamma}(5, \frac{3}{5})$
19	Design airflow (nat. ventilation) [l/s/m ²]	0.4	
	Occupancy		
20	Occupant density [m ² /pers.]	Calibrated	1961-78: $\text{Gamma}(10,6)$
21	Occupant heat load [W/pers.]	80	
22	Room heating set point [°C]	20	
23	Appliances heat load [W/m ²]	Regression, $f(A_f, nOcc)$	
	Domestic hot water		
24	DHW flow temperature [°C]	55.0	
25	Mains temperature [°C]	10.0	
26	Hot water consumption [m ³ /pers./year]	15	

The prior for the DHW model parameters were specified as Gamma distributions. The same prior was specified for the “workday” and the “weekend” profiles, respectively. The priors describe the probability distribution of DHW consumption in a given hour as a share of the daily consumption in percent.

Table 3 lists the shape parameter and the mean associated with each hour, which may be expressed as $\mu = \alpha\beta$. The values correspond to the contours of Figure 9.

Table 3 Prior specification for DHW profiles. Based on data from 107 British residential buildings [55].

hour, i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
α	2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
$\mu = \alpha\beta$	1.8	1.1	0.8	0.7	1.4	1.7	4.4	8.9	7.5	6.0	5.6	5.0	4.3	3.5	3.0	3.1	3.6	5.3	7.6	7.1	6.0	6.0	4.1	2.5

Appendix 2: State space model representation

This appendix presents a continuous-time state space representation of the modified ISO 13790 building model. For readability, the state space model is presented in a version relying only on a single input for solar, and thus deviates slightly from the version used in the case study, where four solar inputs were used (north/east/south/west). If representation of solar from all four cardinal directions is needed, the model may readily be adapted by replacing the second column (aggregated solar contribution) of the matrix B with four individual columns – one for each cardinal direction.

$$\mathbf{A}(k) = \begin{bmatrix} \frac{-H_{ve}(T_a) + H_{is} \cdot \left(\frac{H_{is}}{H_w + H_{is} + H_{ms}} - 1 \right)}{C_a} & \frac{H_{ms} \cdot H_{is}}{(H_w + H_{is} + H_{ms}) \cdot C_a} \\ \frac{H_{ms} \cdot H_{is}}{(H_w + H_{is} + H_{ms}) \cdot C_m} & \frac{H_{ms} \cdot \left(\frac{H_{ms}}{H_w + H_{is} + H_{ms}} - 1 \right) - H_{em}}{C_m} \end{bmatrix}$$

$$\mathbf{B}(k) = \begin{bmatrix} \left(H_{ve}(T_a) + \frac{H_{is} \cdot H_w}{H_w + H_{is} + H_{ms}} \right) \cdot \frac{1}{C_a} & \left(\frac{H_{is} \cdot \left(1 - \frac{A_m}{A_t} \cdot \frac{H_w}{h_{ms} \cdot A_t} \right)}{(H_w + H_{is} + H_{ms})} \right) \cdot \frac{A_w}{C_a} & \left(\frac{H_{is} \cdot \left(0.5 - \frac{0.5 \cdot A_m}{A_t} \cdot \frac{0.5 \cdot H_w}{h_{ms} \cdot A_t} \right)}{H_w + H_{is} + H_{ms}} + 0.5 \right) \cdot \frac{1}{C_a} & \frac{1}{C_a} \\ \left(\frac{H_{ms} \cdot H_w}{H_w + H_{is} + H_{ms}} + H_{em} \right) \cdot \frac{1}{C_m} & \left(\frac{H_{ms} \cdot \left(1 - \frac{A_m}{A_t} \cdot \frac{H_w}{h_{ms} \cdot A_t} \right) + \frac{A_m}{A_t}}{H_w + H_{is} + H_{ms}} \right) \cdot \frac{A_w}{C_m} & \left(\frac{H_{ms} \cdot \left(0.5 - \frac{0.5 \cdot A_m}{A_t} \cdot \frac{0.5 \cdot H_w}{h_{ms} \cdot A_t} \right)}{H_w + H_{is} + H_{ms}} + \frac{0.5 \cdot A_m}{A_t} \right) \cdot \frac{1}{C_m} & 0 \end{bmatrix}$$

$$C = [1 \quad 0]$$

$$D = [0 \quad 0 \quad 0 \quad 0]$$

The matrix describing how inputs act on the model, $\mathbf{B}(k)$, are organized such that the columns refer to the following four inputs: $u = [T_{\text{exterior}}, \Phi_{\text{solar}}, \Phi_{\text{int}}, \Phi_h]$. Here, Φ_{solar} is the solar heat gains per window area (in m^2), while Φ_{int} denote the internal loads, respectively. In the context of the optimization problem of Eq. (9a-9e), the matrices \mathbf{B} and \mathbf{E} of Eq. (9b) are obtained by separating out the first three columns of the $\mathbf{B}(k)$ matrix listed here and forming the disturbance matrix, \mathbf{E} . The last remaining column (the new \mathbf{B}) describes the effect of our control variable (Φ_h) on the model. The time-varying parameter, H_{ve} depends on the external temperature, T_a in the following way:

$$H_{ve}(T_a) = c_{p,\text{air}} \cdot \rho_{\text{air}} \cdot (q_{\text{infiltration}} + q_{\text{venting}}(T_a)) \cdot A_{\text{footprint}}$$

Where $q_{\text{infiltration}}$ is a specified infiltration air flow per square-meter, while the air change from venting is given by:

$$q_{\text{venting}}(T_a) = b(T_a) \cdot q_{\text{designflow}}$$

In this study, the design air changes ($q_{\text{designflow}}$) was fixed at $0.4 \frac{l}{s \cdot m^2}$ and the scaling factor for venting, b , was defined as:

$$b = \frac{e^{0.25T_a - 0.25}}{1 + e^{0.25T_a - 0.25}}$$

Appendix 3: Model validation

Figure 18 shows the obtained estimates of model parameter posterior distributions, which were approximated using non-informative priors to remove any influence of our own beliefs related to the likeliness of parameter-values.

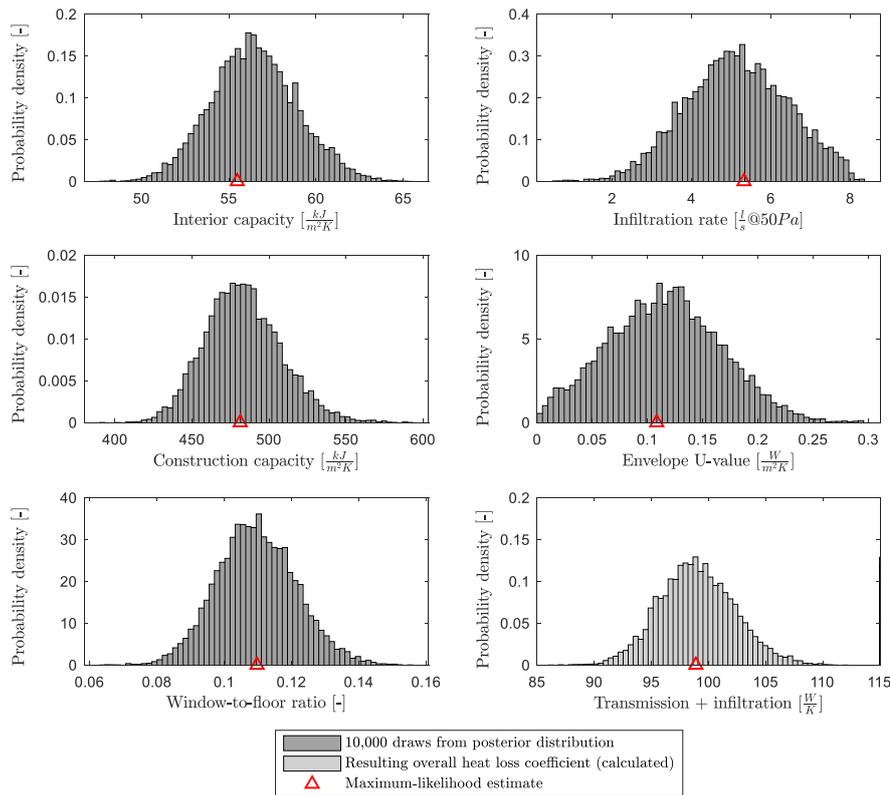


Figure 18 Estimates of model parameter posterior distributions for the modified model structure. The estimates were obtained from the calibration using measurement data from a terraced house. Red markers indicate the ML-estimate of model parameters of the modified model used for the model-comparison of Figure 2 (lower graph).

The estimates of the thermal capacity of the heavy building components is seen to be higher than the values attributed to any of the building classes of the 13790 standard. Whether this difference can be attributed to the increased excitation of the buildings thermal mass during the dynamic measurement experiments was not investigated further.

The broad posteriors of the infiltration rate and envelope U-value indicate that distinguishing between these two heat loss phenomena using the measurements of this dataset alone is difficult, thus agreeing with the conclusions drawn in [40]. A strong correlation was found between the two heat loss parameters (-0.99), thus indicating that whenever a high estimate the heat loss through transmission (Envelope U-value) appeared in the posterior, this added heat loss was compensated

for through a lower infiltration heat loss (Infiltration rate) and vice versa. This effect is clear if we add the heat losses of each of the two components together, thus obtaining a much more confined estimate of the posterior distribution for the overall heat loss, as depicted on the lower-right histogram of Figure 18.

Typically, such a strong correlation between two parameters would suggest that changes should be made - either to the set of calibrated parameters or to the model structure itself. Such a change could be to infer the combined heat loss coefficient instead of its components which, as indicated by Figure 18, is much easier to identify. While this would remove the issue with strongly correlated parameters, it would introduce a need for assumptions regarding the distribution of the total heat loss – this without even ensuring a better performance of the resulting model. Because of this, and because the objective of this work is not necessarily to obtain the “true” parameter estimates, we opted to keep the initially proposed structure and set of calibration parameters.

References

- [1] The Danish Energy Agency, “Energiscenarier frem mod 2020, 2035 og 2050 (in Danish),” The Danish Energy Agency, Copenhagen, 2014.
- [2] J. Torriti, M. G. Hassan and M. Leach, “Demand response experience in Europe: Policies, programmes and implementation,” *Energy*, vol. 35, no. 4, pp. 1575-1583, 2010.
- [3] Buildings Performance Institute Europe (BPIE), “Europe's Buildings under the Microscope - A country-by-country review of the energy performance of buildings,” BPIE, 2011.
- [4] Commission of European Union, “Energy consumption in households,” 03 2018. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households. [Accessed 17 09 2018].
- [5] F. Oldewurtel, A. Ulbig, A. Parisio, G. Andersson and M. Morari, “Reducing peak electricity demand in building climate control using real-time pricing and model predictive control,” in *49th IEEE Conference on Decision and Control (CDC)*, Atlanta, 2009.
- [6] J. Le Dréau and P. Heiselberg, “Energy flexibility of residential buildings using short term heat storage in the thermal mass,” *Energy 111*, pp. 991-1002, 2016.
- [7] R. E. Hedegaard, T. H. Pedersen and S. Petersen, “Multi-market demand response using economic model predictive control of space heating in residential buildings,” *Energy and Buildings 150*, pp. 253-261, 2017.
- [8] T. H. Pedersen, R. E. Hedegaard and S. Petersen, “Space heating demand response potential of retrofitted residential apartment blocks,” *Energy and Buildings 141*, pp. 158-166, 2017.
- [9] R. Halvgaard, N. K. Poulsen, H. Madsen and J. B. Jørgensen, “Economic Model Predictive Control for building climate control in a Smart Grid,” in *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, Washington, DC, 2012.
- [10] M. D. Knudsen and S. Petersen, “Model predictive control for demand response of domestic hot water preparation in ultra-low temperature district heating systems,” *Energy and Buildings 146*, pp. 55-64, 2017.

- [11] M. Avci, M. Erkoca, A. Rahmanib and S. Asfoura, "Model predictive HVAC load control in buildings using real-time electricity pricing," *Energy and Buildings* 60 (2013), pp. 199-209, DOI: 10.1016/j.enbuild.2013.01.008.
- [12] F. Oldewurtel, A. Parisio, C. N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann and M. Morari, "Use of model predictive control and weather forecasts for energy efficient building climate control," *Energy and Buildings*, vol. 45, pp. 15-27, 2012.
- [13] M. D. Knudsen and S. Petersen, "Demand Response Potential of Model Predictive Control of Space Heating based on Price and Carbon Dioxide Intensity Signals," *Energy and Buildings* 125 (2016), pp. 196-204, DOI: 10.1016/j.enbuild.2016.04.053.
- [14] P. J. C. Vogler-Finck, R. Wisniewski and P. Popovski, "Reducing the carbon footprint of house heating through model predictive control – A simulation study in Danish conditions," *Sustainable Cities and Society*, vol. 42, pp. 558-573, 2018.
- [15] N. O'Connell, P. Pinson, H. Madsen and M. O'Malley, "Benefits and challenges of electrical demand response: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 686-699, 2014.
- [16] C. F. Reinhart and C. C. Davila, "Urban building energy modeling – A review of a nascent field," *Building and Environment*, vol. 97, pp. 196-202, 2016.
- [17] M. Aydinalp-Koksal and V. I. Ugursal, "Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector," *Applied Energy*, vol. 85, no. 4, pp. 271-296, 2008.
- [18] R. Buffat, A. Froemelt, N. Heeren, M. Raubal and S. Hellweg, "Big data GIS analysis for novel approaches in building stock modelling," *Applied Energy*, vol. 208, pp. 277-290, 2017.
- [19] H. Lim and Z. (John), "Influences of energy data on Bayesian calibration of building energy model," *Applied Energy*, vol. 231, pp. 686-698, 2018.
- [20] P. Nageler, G. Zahrer, R. Heimrath, T. Mach, F. Mauthner, I. Leusbrock, H. Schranzhofer and C. Hochenauer, "Novel validated method for GIS based automated dynamic urban building energy simulations," *Energy*, vol. 139, pp. 142-154, 2017.
- [21] S. Nagpal, C. Mueller, A. Aijazi and C. F. Reinhart, "A methodology for auto-calibrating urban building energy models using surrogate modeling techniques," *Journal of Building Performance Simulation*, 2018.
- [22] C. A. Balaras, E. G. Dascalaki, K. G. Droutsa and S. Kontoyiannidis, "Empirical assessment of calculated and actual heating energy use in Hellenic residential buildings," *Applied Energy*, vol. 164, pp. 115-132, 2016.
- [23] A. Mastrucci, O. Baume, F. Stazi and U. Leopold, "Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam," *Energy and Buildings*, vol. 75, pp. 358-367, 2014.
- [24] D. F. Dominkovic, P. Gianniou, M. Münster, A. Heller and C. Rode, "Utilizing thermal building mass for storage in district heating systems: Combined building level simulations and system level optimization," *Energy*, vol. 153, pp. 949-966, 2018.
- [25] H. Cai, C. Ziras, S. You, R. Li, K. Honoré and H. W. Bindner, "Demand side management in urban district heating networks," *Applied Energy*, vol. 230, pp. 506-518, 2018.
- [26] W. Li, Y. Zhou, K. Cetin, J. Eom, Y. Wang, G. Chen and X. Zhang, "Modeling urban building energy use: A review of modeling approaches and procedures," *Energy*, vol. 141, pp. 2445-2457, 2017.

- [27] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819-1835, 2009.
- [28] M. Kavgić, A. Mavrogianni, D. Mumović, A. Summerfield, Z. Stevanović and M. Djurović-Petrović, "A review of bottom-up building stock models for energy consumption in the residential sector," *Building and Environment*, vol. 45, no. 7, pp. 1683-1697, 2010.
- [29] G. Reynders, J. Diriken and D. Saelens, "Quality of grey-box models and identified parameters as function of the accuracy of input and observations signals," *Energy and Buildings*, vol. 82, pp. 263-274, 2014.
- [30] R. De Coninck, F. Magnusson, J. Åkesson and L. Helsen, "Grey-Box Building Models for Model Order Reduction and Control," in *Proceedings of the 10th International Modelica Conference*, Linköping, 2014.
- [31] M. Kristensen, R. Choudhary, R. Høst Pedersen and S. Petersen, "Bayesian calibration of residential building clusters using a single geometric building representation," in *Building Simulation*, San Francisco, CA, 2017.
- [32] P. Giannou, C. Reinhart, D. Hsu, A. Heller and C. Rode, "Estimation of temperature setpoints and heat transfer coefficients among residential buildings in Denmark based on smart meter data," *Building and Environment*, vol. 139, pp. 125-133, 2018.
- [33] M. Kristensen, R. E. Hedegaard and S. Petersen, "Hierarchical calibration of archetypes for urban building energy modeling," *Energy and Buildings*, vol. 175, pp. 219-234, 2018.
- [34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087-1092, 1953.
- [35] European Committee for Standardization, "EN/ISO 13790:2008 - Energy performance of buildings - Calculation of energy use for space heating and cooling," European Committee for Standardization, Brussels, 2008.
- [36] P. H. Shaikh, N. B. M. Nor, P. Nallagownden, I. Elamvazuthi and T. Ibrahim, "A review on optimized control systems for building energy and comfort management of smart sustainable buildings," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 409-429, 2014.
- [37] SKAT; KL; Kombit, "Bygnings- og boligregistret (in Danish only)," [Online]. Available: <https://bbr.dk>. [Accessed 29 04 2018].
- [38] P. Zangheri, R. Armani, M. Pietrobon, L. Pagliano, M. F. Boneta and A. Müller, "Heating and cooling energy demand and loads for building types in different countries of the EU," ENTRANZE, 2014.
- [39] J. Vivian, A. Zarrella, G. Emmi and M. De Carli, "An evaluation of the suitability of lumped-capacitance models in calculating energy needs and thermal behaviour of buildings," *Energy and Buildings*, vol. 150, pp. 447-465, 2017.
- [40] Hedegaard, Rasmus Elbæk; Petersen, Steffen, "Evaluation of grey-box model parameter estimates intended for thermal characterization of buildings," in *11th Nordic Symposium on Building Physics*, Trondheim, 2017.
- [41] H. Harb, N. Boyanov, L. Hernandez, R. Streblov and D. Müller, "Development and validation of grey-box models for forecasting the thermal response of occupied buildings," *Energy and Buildings*, vol. 117, pp. 199-207, 2016.

- [42] R. Bruno, G. Pizzuti and N. Arcuri, "The prediction of thermal loads in buildings by means of the EN ISO 13790 dynamic model: a comparison with TRNSYS," *Energy Procedia*, vol. 101, pp. 192-199, 2016.
- [43] U.S. Department of Energy, "EnergyPlus version 8.8.0 Documentation: Engineering Reference," U.S. Department of Energy, Washington, D.C., 2017.
- [44] Danish Transport, Construction and Housing Authority, "Eref: Danish building regulations 2018 (BR18)," 01 01 2018. [Online]. Available: http://bygningsreglementet.dk/Tekniske-bestemmelser/11/BRV/Energiforbrug/Kap-1_6#f74d9134-db45-4deb-b6b2-dc5ae1fe9ffc. [Accessed 20 02 2018].
- [45] H. B. Rijal, P. Tuohy, M. A. Humphreys, J. F. Nicol, A. Samuel and J. Clarke, "Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings," *Energy and Buildings*, vol. 39, no. 7, pp. 823-836, 2007.
- [46] K. Gram-Hanssen, "SBI2005:12 - Husholdnings elforbrug - hvem bruger hvor meget, til hvad og hvorfor?," SBI forlag, Hørsholm, 2005.
- [47] P. Bacher, P. A. de Saint-Aubain, L. E. Christiansen and H. Madsen, "Non-parametric method for separating domestic hot water heatingspikes and space heating," *Energy and Buildings*, vol. 130, pp. 107-112, 2016.
- [48] R. Burzynski, M. Crane, R. Yao and V. M. Becerra, "Space heating and hot water demand analysis of dwellings connected to district heating scheme in UK," *Journal of Central South University*, vol. 19, no. 6, p. 1629-1638, 2012.
- [49] Statens byggeforskningsinstitut, "SBI 2009:10 - Varmt brugsvand," Statens byggeforskningsinstitut, Hørsholm, 2009.
- [50] M. Vihola, "Robust adaptive Metropolis algorithm with coerced acceptance rate," *Statistics and Computing*, vol. 22, no. 5, p. 997-1008, 2012.
- [51] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin, *Bayesian Data Analysis*, Boca Raton: CRC Press, 2014.
- [52] W. Cyx, N. Renders, M. V. Holm and S. Verbeke, "IEE TABULA - Typology Approach for Building Stock Energy," 2011.
- [53] Statens Byggeforskningsinstitut, "SBI 2015:25 - Tæthed i eksisterende bygninger [in Danish]," Statens Byggeforskningsinstitut, Aalborg Universitet, København, 2015.
- [54] Statistics Denmark, "Nyt fra Danmarks Statistik: På øerne er der god plads (in Danish)," 4 04 2016. [Online]. Available: <https://www.dst.dk/da/Statistik/nyt/NytHtml?cid=21543>. [Accessed 13 03 2018].
- [55] Department for Environment Food and Rural Affairs, "Measurement of Domestic Hot Water Consumption in Dwellings," 2008.
- [56] S. P. Brooks and A. Gelman, "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, vol. 7, pp. 434-455, 1997.
- [57] ASHRAE Standards Committee 2001-2002, "ASHRAE Guideline 14: Measurement of Energy and Demand Savings," ASHRAE, Atlanta, 2002.
- [58] M. Dahl, A. Brun, O. S. Kirsebom and G. B. Andresen, "Improving Short-Term Heat Load Forecasts with Calendar and Holiday Data," *Energies*, vol. 11, no. 7:1678, 2018.

- [59] T. H. Haider, O. H. See and W. Elmenreich, "A review of residential demand response of smart grid," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 166-178, 2016.
- [60] G. Reynders, "Quantifying the impact of building design on the potential of structural thermal storage for active demand response in residential buildings," PhD Thesis, KU Leuven, 2015.
- [61] O. Corradi, H. Ochsenfeld, H. Madsen and P. Pinson, "Controlling Electricity Consumption by Forecasting its Response to Varying Prices," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 421-429, 2013.