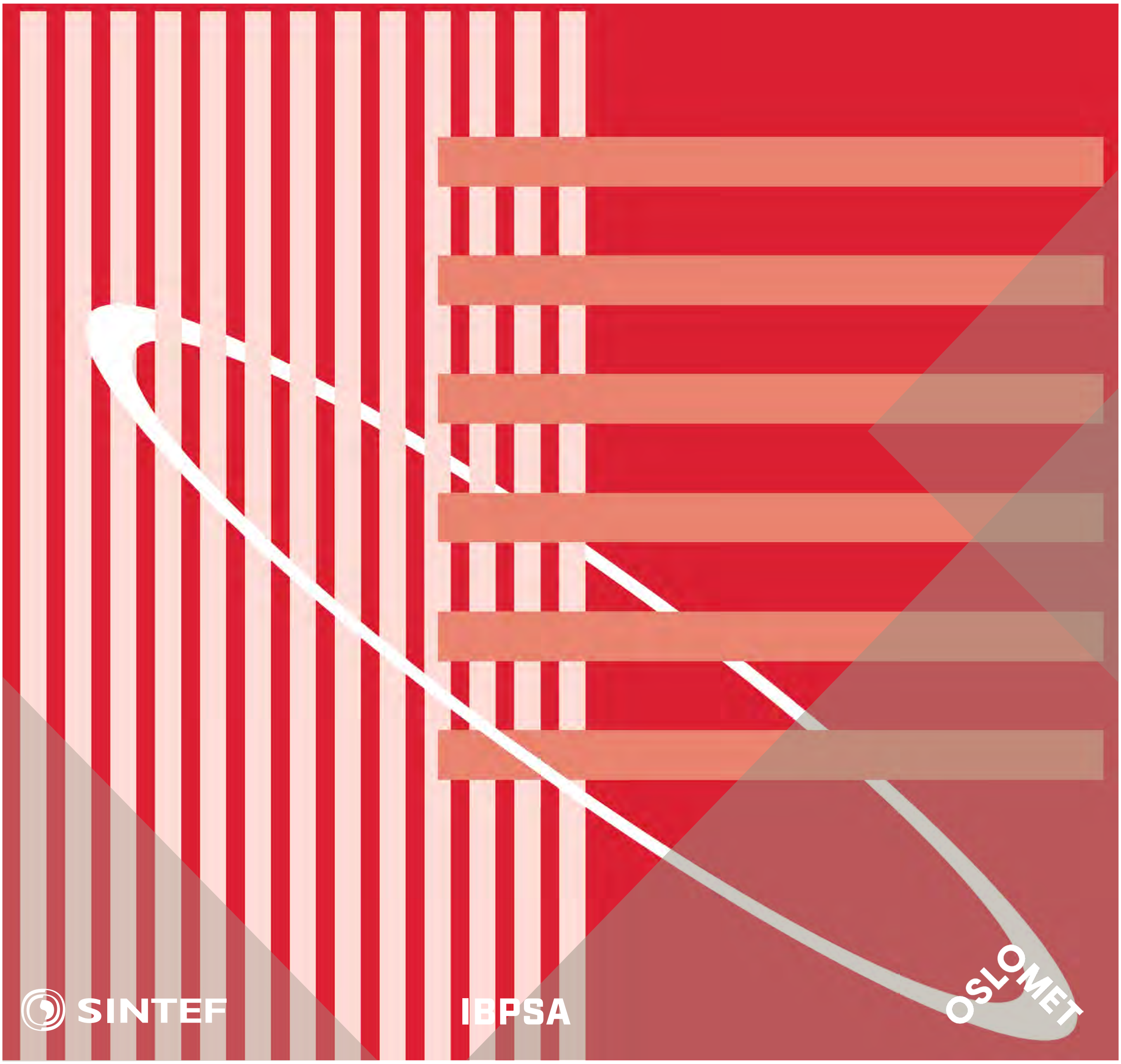
SINTEF
PROCEEDINGS
5

International Conference Organised by
IBPSA-Nordic, 13th–14th October 2020,
OsloMet

BuildSIM-Nordic 2020

Selected papers



SINTEF Proceedings

Editors:
Laurent Georges, Matthias Haase, Vojislav Novakovic and Peter G. Schild

BuildSIM-Nordic 2020

Selected papers

International Conference Organised by IBPSA-Nordic,
13th–14th October 2020, OsloMet

SINTEF Academic Press

SINTEF Proceedings no 5
Editors:
Laurent Georges, Matthias Haase, Vojislav Novakovic and Peter G. Schild
BuildSIM-Nordic 2020
Selected papers
International Conference Organised by IBPSA-Nordic,
13th–14th October 2020, OsloMet

Keywords:
Building acoustics, Building Information Modelling (BIM), Building physics, CFD and air flow, Commissioning and control, Daylighting and lighting, Developments in simulation, Education in building performance simulation, Energy storage, Heating, Ventilation and Air Conditioning (HVAC), Human behavior in simulation, Indoor Environmental Quality (IEQ), New software developments, Optimization, Simulation at urban scale, Simulation to support regulations, Simulation vs reality, Solar energy systems, Validation, calibration and uncertainty, Weather data & Climate adaptation, Fenestration (windows & shading), Zero Energy Buildings (ZEB), Emissions and Life Cycle Analysis

Cover illustration: IBPSA-logo

ISSN 2387-4295 (online)
ISBN 978-82-536-1679-7 (pdf)

SINTEF Academic Press
Address: Børrestuveien 3
PO Box 124 Blindern
N-0314 OSLO
Tel: +47 40 00 51 00

www.sintef.no/community
www.sintefbok.no

SINTEF Proceedings
SINTEF Proceedings is a serial publication for peer-reviewed conference proceedings on a variety of scientific topics.
The processes of peer-reviewing of papers published in SINTEF Proceedings are administered by the conference organizers and proceedings editors. Detailed procedures will vary according to custom and practice in each scientific community.

Influence of Data Pre-Processing Techniques and Data Quality for Low-Order Stochastic Grey-Box Models of Residential Buildings

Xingji Yu[1]*, Laurent Georges[1]
[1]Department of Energy and Process Engineering, Faculty of Engineering, NTNU - Norwegian University of Science and Technology, Kolbjørn Hejes vei 1a, 7034 Trondheim, Norway
*corresponding author: xingji.yu@ntnu.no

Abstract

Model Predictive Control (MPC) has proved to be a key technology to activate the energy flexibility of buildings. A reliable control-based model should be developed to implement an efficient optimal control. Grey-box models, as a combination of physical knowledge and experiment data, have been widely used in the literature. However, in the identification process of grey-box models, many factors affect the results. This paper uses data from virtual experiments in IDA-ICE to investigate the influence of the optimization methods, the filtering methods, the training dataset and the sampling time interval on stochastic grey-box models. It shows that global optimization increases the chance to avoid a local minimum. Pre-filtering methods have a small influence on the model quality. Larger data sampling time will cause the identified parameters to become non-physical. However, the simulation performance of the model is kept almost unchanged.

Introduction

The share of Renewable Energy Sources (RES) is increasing constantly in today's energy system. The volatile property of RES generation has brought notable challenges to the grid. Thus, flexible loads become a requirement to further increase the penetration of RES. Demand response (DR) is considered to be one of the key components to provide flexibility in smart grid [1]. DR can be described as the interaction and responsiveness of the end-use customer to a penalty signal (e.g. price signal, CO_2 intensity factor for electricity) [2]. In addition, due to the continuous increase of the electric consumption of households and the introduction of electric vehicles, DR can be used for peak-shaving in order to avoid congestion in the distribution grid [3]. Consequently, peak-shaving would enable to minimize or postpone the reinforcement of these grids.

About 25% of the final energy consumption is consumed by buildings and more than 65% of this energy is used for heating and cooling demand in European households, which makes HVAC systems a promising candidate for demand response [4]. In Nordic countries, the space-heating season is long and cold, the energy consumption is mainly related to space-heating. The thermal mass of buildings can be a significant heat storage [5,6]. When using the thermal mass to perform DR, the heating demand will be shifted optimally, while the thermal comfort constraints can still be respected [7]. The targets of DR in buildings are usually the reduction of peak load, lower CO_2 emissions, maximize the use of RES or minimize energy cost [8]. Model predictive control (MPC) is often considered as an important technique to perform demand response (DR) using the thermal mass of the building. The logic of MPC in a building is that the control agent (computers, built-in intelligent devices, etc.) takes the predictions of future disturbances (weather data, power generation from RES, etc.) and the system constraints into an optimization problem and generate an optimal control decision at each control time step. Thus, it is important that the dynamic model embedded in the MPC controller has decent prediction accuracy. A poor-quality model could lead to suboptimal performance, such as increased energy costs, violation of the thermal comfort or even be counterproductive for the electricity grid. In addition, the model identification should also be low-cost to make the investment costs of MPC sufficiently low.

Control models for MPC controller can be divided into three main categories, namely white-, black- and grey-box models. White-box models are based on physical laws, which require detailed knowledge of the system, the underlying physical process and parameters. In practice, it is too complicated and time-consuming to access all the information and to keep it updated during the building's operational lifetime. This type of model usually has higher accuracy but is mathematically more complex, which may cause challenges for the MPC optimizer. This fact makes this kind of model sometimes too complex for MPC [4]. Black-box models are pure data-driven methods considering only measured inputs and outputs from the system. The physical knowledge of the system is not needed. However, this method requires a larger amount of data for training and the precision of black-box models is significantly influenced by the data quality. Black-box models are known to have lower generalization (extrapolation) properties. Grey-box modelling is a combination of physical knowledge and statistical methods. Since the grey-box models have a model structure based on physical knowledge, grey-box models usually require less experimental data compared to black-box models and are hopefully less sensitive to data quality [9].

A common way to create Linear Time Invariant (LTI) grey-box models for buildings is to use lumped capacitance models (RC models). The thermal conditions of the building are expressed with an electrical circuit analogy [10]. This paper mainly focuses on five specific factors influencing the grey-box modelling of the building thermal dynamics. The first aspect (Q1) is data preprocessing. Few publications are addressing the importance of data preprocessing for building thermal dynamics. The topic is discussed in other disciplines, like [11] in process engineering, but not in building science. The second aspect (Q2) is the convexity of the optimization problem. Except for models with an extremely simple structure like first-order models, the optimization problem for identifying parameters of the grey-box models is not convex. Therefore, grey-box models are very sensitive to initial guess and the search method (i.e. the optimizer). For instance, Generic Algorithm (GA) combined with gradient-based optimization is used in the paper [12] to avoid the identification results ending up in a local minimum. The selection of the optimization algorithm to avoid the local minimum will be discussed in this paper. The third aspect (Q3) is how data quality (e.g. level of excitation signal and amount of data) influences the identification results. It is often said that the temperature of the ventilation extract air is a good image of the average building temperature and is reliable to identify a grey-box model, see e.g. [13]. Thus, the fourth aspect (Q4) is about the selection of the representative indoor temperature for system identification. The last aspect (Q5) considers the sensitivity of the grey-box parameters to the selection of the data sampling time (Ts). The theoretical analysis of Ljung showed that the continuous grey-box models are sensitive to the selection of the sampling time that should be taken lower than the shortest time of the system to be investigated [14]. This analysis needs to be repeated for building applications. All the research in this paper is performed using stochastic grey-box models in innovation form using the disturbance matrix K and the MATLAB identification toolbox.

Methodology

Dataset and virtual experiments

IDA ICE is a detailed dynamic simulation tool for studying thermal indoor climate as well as the energy consumption of buildings. A two-storey detached house with a heated floor area of 160 m² is used as virtual experiment for our case study. The three-dimensional geometry of the building from IDA ICE is shown in Fig. 1. The building is constructed in wood (i.e. lightweight construction) and complies with the requirement of the Norwegian passive house standard, NS 3700 [15]. The detailed description of the building construction can be found in [16]. The building is equipped with balanced mechanical ventilation with a heat recovery unit. The heat exchanger is here modelled using a constant effectiveness of 85% without bypass, like a plate heat exchanger.

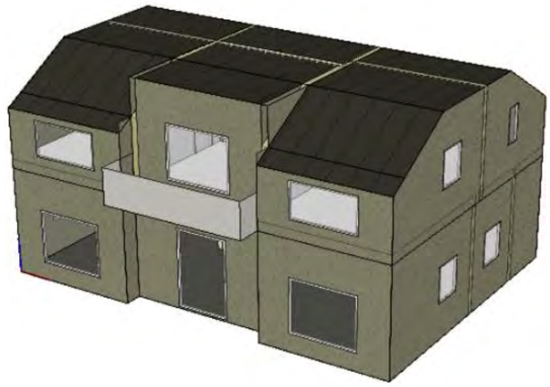


Figure 1: 3D geometry of the building model in IDA ICE (showing the southwest façade).

The building is simulated with a multi-zone model with open internal doors. IDA ICE has an embedded ventilation network model which accounts for the large bidirectional airflow through open doorways. This large convective heat transfer leads to relatively uniform air temperatures in the entire building. However, bathrooms are kept separated with closed doors. Following the cascade ventilation principle, ventilation air is supplied in living areas and bedrooms and mostly extracted in wet rooms (i.e. bathrooms and the laundry). The space-heating was performed using an electrical heater in this case study. Direct electricity is a most common way to heat small residential buildings in Norway [17] . The hourly profiles of internal heat gains for artificial lighting, electric appliances and occupancy is taken from a Norwegian standard [18].

Two types of excitation signals are used to activate the thermal mass of the building in order to collect data for system identification. The first signal is called Pseudo-Random Binary Signal (PRBS) with a minimum and maximum step of 10 and 80 min, respectively. The reason for choosing a PRBS signal is that it approximates white noise, which can activate the dynamic system in a large spectrum of frequencies [19,20]. The other excitation signal is an intermittent set-point, which means that the temperature set-point changes between daytime and night-time (i.e. night setback). In this case, an on-off control is implemented in IDA ICE to track the temperature set-point, like in real direct electric radiators. Both excitation signals are applied to an electric radiator placed in each zone, except for bathrooms as these rooms are relatively small and typically heated by floor heating (with significant thermal inertia). Five different periods with specific weather conditions are implemented in the virtual experiments, as described in the table below.

IDA ICE uses a time-varying time-step so that the data is not generated at constant time intervals. The data output from IDA ICE is therefore interpolated on a uniform time discretization of 2.5 min (thus well shorter than the 10 min time interval of the PRBS).

Table 1: Weather condition of four PRBS experiments.

Type	Outdoor Temperature	Sky	Date	Duration
Very Cold	-10 °C	Clear sky	12/13 /2019	One week
Cold	0 °C	Overcast	12/24 /2019	One week
Cold	0 °C	Clear sky	3/23/ 2019	One week
Mild	5 °C	Overcast	11/23 /2019	One week

Grey-box model structure and identification

The main purpose of this paper is not to investigate the grey-box model structure. This topic is already discussed in previous works [21–23]. Only first-order (1R1C) and second-order (3R2C) grey-box models are considered in this paper with a single temperature node inside the building (i.e. mono-zone model). Preliminary tests have shown that a third-order model would be over-fitted for this test case. Higher-order models can cause over-parameterization more easily, which has been shown in the papers [23,24]. The structure of the two grey-box models follows a RC-formalism. The lumped resistance and capacitance as well as the physical interpretation of these parameters can be found in Figures 2 and 3 below. The free parameters of these grey-box models are calibrated using the IDA ICE data. The ventilation exhaust air temperature or the volume-averaged temperature can be selected to represent the measured interior node Ti and their respective model performance will be compared.

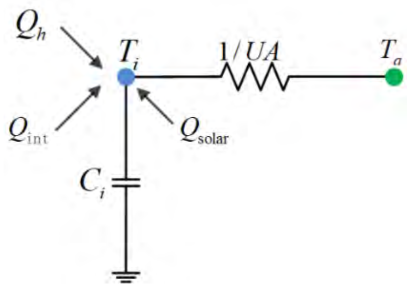


Figure 2: First-order model (1R1C)

T_i Temperature of interior heat capacity [°C].

T_a The outdoor (or ambient) temperature [°C].

C_i Heat capacity of the building [kWh/K].

R ($1/UA$) Overall heat resistance between the building and the ambient, including ventilation [K/kW].

Q_{int} Internal heat gain from artificial lighting, people and electric appliances [kW].

Q_{solar} Heat gain from solar radiation [kW].

Q_h Heat gain from the electric heater [kW].

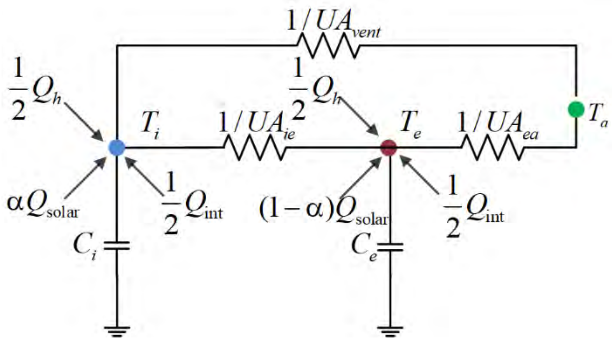


Figure 3: Second-order model (3R2C)

T_e Temperature of the building envelope [°C].

T_a The outdoor temperature [°C].

C_i Heat capacity of the building combining the thermal mass of the air, the furniture, internal walls and, potentially, the first centimetres of external walls [kWh/K].

C_e Heat capacity of the node Te [kWh/K].

R_{ie} ($1/UA_{ie}$) Heat resistance between the building envelope and the interior [K/kW].

R_{ea} ($1/UA_{ea}$) Heat resistance between the ambient and the building envelope [K/kW].

R_{vent} ($1/UA_{vent}$) Heat resistance between the ambient and the interior [K/kW].

α Fraction of solar gains to air node.

The internal gains and solar gains are computed exactly by IDA ICE. In this work, they are not identified and are introduced directly in the grey-box model. Consequently, in the 3R2C model, only the coefficient α to distribute the solar gains between the two temperature nodes needs to be identified regarding gains. In real application, gains are not known exactly. However, simplifying the problem enables to emphasize the specific research questions of the article.

The MATLAB identification toolbox is used for model identification. In grey-box models, the continuous time model is first discretized in order to identify the model parameters using discrete measurement data. The discretization assumes the input data to be piecewise constant during each time interval (i.e. zero-order hold). Regarding the optimization problem, the initialization value of the model parameters and their corresponding range (i.e. minimum and maximum values) should be defined. The optimizer will then iterate to find the parameters that minimize the Normalized Root Mean Square Error (NRMSE) of the one-step ahead prediction. Then, the toolbox covert the discrete time model back to continuous time:

$$\dot{x}(t) = \mathbf{A}\, x(t) + \mathbf{B}\, u(t) + \mathbf{K}\, e(t) \quad (1)$$

$$\dot{y}(t) = \mathbf{C}\, x(t) + e(t) \quad (2)$$

where x is the state vector and A, B and C are the system matrices. u is the input vector (T_a, Q_{solar}, Q_{int}, Q_h) and y is the output (indoor temperature, T_i). K is the disturbance matrix of the innovation form (Kalman gain). It is a transformed representation from the general process [25].

Influence of the optimizer

In MATLAB, the function *greyest* identifying the model parameters has four different gradient-based iterative search methods, used in sequence. However, preliminary tests using the 3R2C model show a quick converge to a local minimum close to the initial estimate of the parameters. A similar behavior is also reported in the paper [12]. The authors used GA combined with gradient-based optimization to overcome the non-convexity of the optimization problem. Consequently, a global optimization algorithm has been implemented in this paper. Instead of the GA method, the first stage optimization uses particle swarm optimization (PSO) while the second stage uses the default *greyest* function to further polish the results. Each optimization method is able to identify the parameters' value and their corresponding variance. For each case, the optimizer giving the lowest NMRSE for the one-step ahead prediction is selected and provides the selected model parameters.

Pre-filtering methods

In real-life applications, it is difficult to guarantee that the measurement data will be sampled at a higher frequency (Ts) than the highest frequency of the system (here 10 min, imposed by the PRBS). For instance, the Advanced Metering System (AMS) in Norway has a typical time interval of 15 min [26]. It is important to investigate the influence of data pre-processing by low-level digital measurement devices before they log the data at an appropriate time interval. Temperature sensors can register data at a very high frequency (here 2.5 min). This data can be pre-processed before being sampled and logged at a longer time interval (Ts > 2.5 min). A low-pass discrete filter can first be applied, such as a moving average (MA) or a finite impulse response (FIR). Without this low-pass filter (i.e. direct sampling), the aliasing error may be high. With MA, the aliasing error is smaller but still present while the FIR (applied with a sufficient order) would lead to negligible aliasing. By comparing the performance of (MA + sampling), (FIR + sampling) and the direct sampling on the parameter identification, it is possible to understand the influence of aliasing. The low-pass filter is applied to all the input and output variables of the model. If the filter introduces a time delay (like MA), this delay is the same for all variables and will thus not affect the model. The situation would be more complex if the low-pass filter is only applied to a subset of the input or output variables.

Results

Influence of the optimizer (Q2)

Five datasets using the four PRBS signals and the intermittent on-off heating during the full heating season (FHS) are used to investigate the influence of the optimizer. The two optimization methods do not show much difference for the 1R1C model. In most cases, the two optimization methods converge to the same parameter values. However, the identified parameters from *greyest* function are not identical for the 3R2C model. This implies that the optimization is already non-convex from the second-order model, this conclusion is also confirmed in Arendt et al. [12]. The best optimizer leading to lowest NRMSE for the second-order model can be found in Table 2 (with different time intervals, excitation signals and filters). The figure shows that global optimization is selected for all cases no matter the time interval or filtering method.

Table 2: Best optimizer for the four PRBS and FHS experiments.

Sampling time	Type	Direct sampling	Averaging filter	FIR filter
2.5min	PRBS1	Global	Global	Global
	PRBS2	Global	Global	Global
	PRBS3	Global	Global	Global
	PRBS4	Global	Global	Global
	FHS	Global	Global	Global
15min	PRBS1	Global	Global	Global
	PRBS2	Global	Global	Global
	PRBS3	Global	Global	Global
	PRBS4	Global	Global	Global
	FHS	Global	Global	Global
30min	PRBS1	Global	Global	Global
	PRBS2	Global	Global	Global
	PRBS3	Global	Global	Global
	PRBS4	Global	Global	Global
	FHS	Global	Global	Global
60min	PRBS1	Global	Global	Global
	PRBS2	Global	Global	Global
	PRBS3	Global	Global	Global
	PRBS4	Global	Global	Global
	FHS	Global	Global	Global

Since the datasets contain different excitation signals and weather conditions, it is a strong proof that global optimization can give more robust and higher-quality results when the optimization problem is not convex. In other words, the global optimization algorithm can increase the chance to avoid a local minimum in the grey-box identification process.

Influence of the selection of input (Q4)

While the one-step prediction is used to train the models, the simulation performance is more relevant for MPC applications. Therefore, the simulation NRMSE fitting is mainly used as the performance index in this subsection. Table 3 and Table 4 compare the cross-validation simulation performance using the volume-averaged air temperature and the extracted air as representative indoor temperature respectively. Only datasets trained with the original 2.5 min sampling time is used to avoid the influence of other factors (e.g. dataset, discretization error and pre-filtering method).

Table 3:Simulation NRMSE fitting using the volume-averaged air temperature (Ts = 2.5min)

Training dataset	Validation dataset and simulation NRMSE fitting				
	PRBS1	PRBS2	PRBS3	PRBS4	FHS
PRBS1	84.25%	74.96%	0.53%	72.34%	-17.72%
PRBS2	77.10%	74.16%	24.25%	60.58%	9.49%
PRBS3	39.36%	34.03%	64.20%	14.41%	33.24%
PRBS4	82.19%	69.36%	-17.69%	78.45%	-42.34%
FHS	45.95%	41.11%	69.06%	20.59%	39.17%

Table 4:Simulation NRMSE fitting using the extracted ventilation air temperature (Ts = 2.5 min)

Training dataset	Validation dataset and simulation NRMSE fitting				
	PRBS1	PRBS2	PRBS3	PRBS4	FHS
PRBS1	90.21%	70.83%	16.97%	79.05%	-94.10%
PRBS2	73.51%	81.86%	29.88%	71.77%	-74.10%
PRBS3	30.44%	43.28%	68.02%	25.09%	-15.82%
PRBS4	78.70%	73.55%	-10.68%	83.63%	-155.32%
FHS	78.11%	71.50%	52.43%	64.46%	25.33%

In general, simulation performance with the two different representative temperatures are following the same trend. The simulation NRMSE fitting is higher for the original training dataset and lower for the validation datasets. The model identified from the intermittent set-point and on-off control dataset during the FHS presents higher performance on the validation datasets: the validation fitting is acceptable at each period never completely collapsing. Models trained from the PRBS excitation signals of one week have a good simulation NRMSE fitting on their own training data but largely fail in some cross-validation datasets. Simulation results from extracted air temperature show a slightly higher simulation NRMSE fitting value for the original training dataset. However, models trained with extracted air temperature show worse simulation NRMSE fitting compared with volume-averaged temperature when they are trained and validated with the FHS dataset (values in red in Table 3. Thus, the volume-averaged air temperature is a more balanced choice of representative indoor temperature. The exhaust air temperature is not always the best option to train the model and this conclusion could be even more severe if all the internal doors inside the building were closed.

Influence of pre-filtering methods and data-quality (Q1, Q3 and Q5)

Figures 4 to 6 show three key identified parameters for the second-order model. For the value of the total heat transfer coefficient in Figure 4, the estimated value from a step-response of the heating power applied in IDA-ICE is about 85 W/K. Figure 4 shows that most of the results are close to the estimation from IDA-ICE. When the Ts is increased to 60 min, some values using the FIR filter or

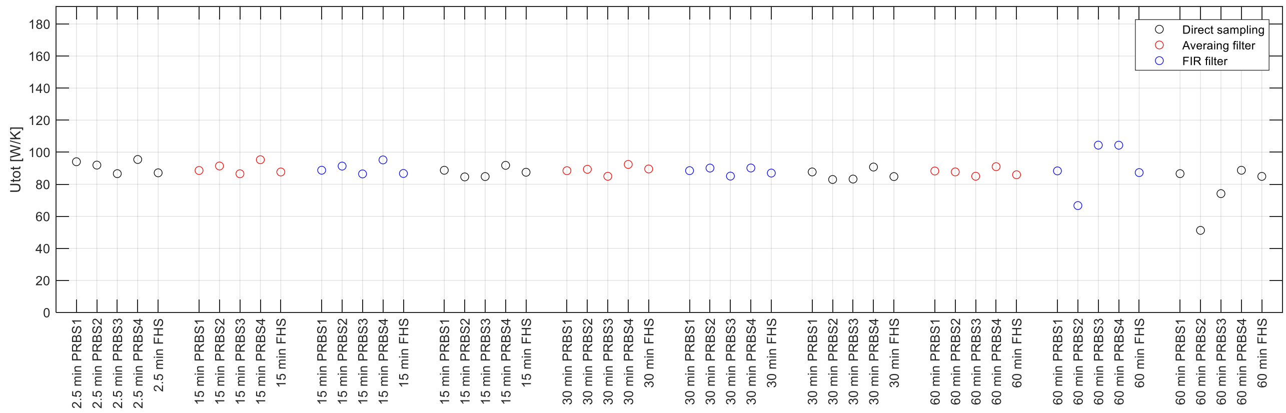


Figure 4: Identified Utot of the 3R2C model (variance is not given as Utot combines the 3R)

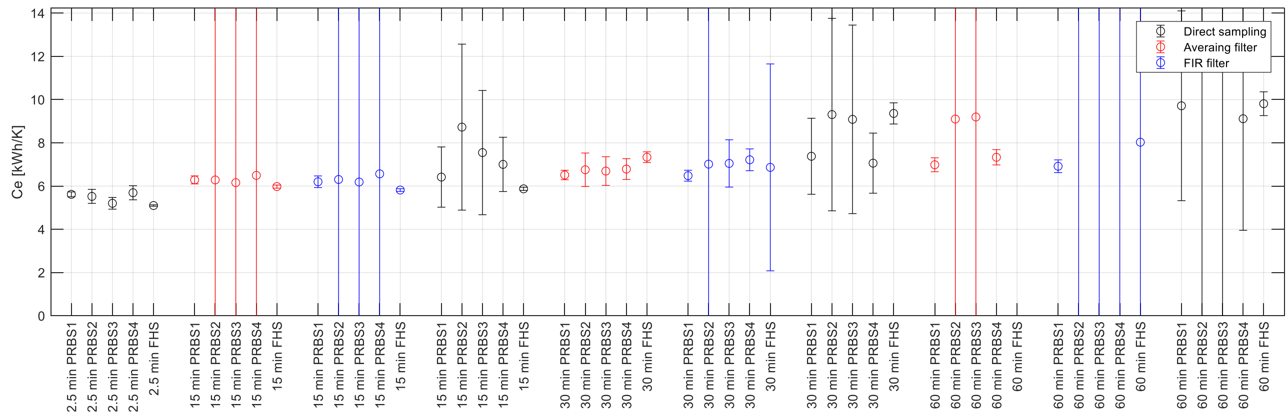


Figure 5: Identified Ce of the 3R2C model

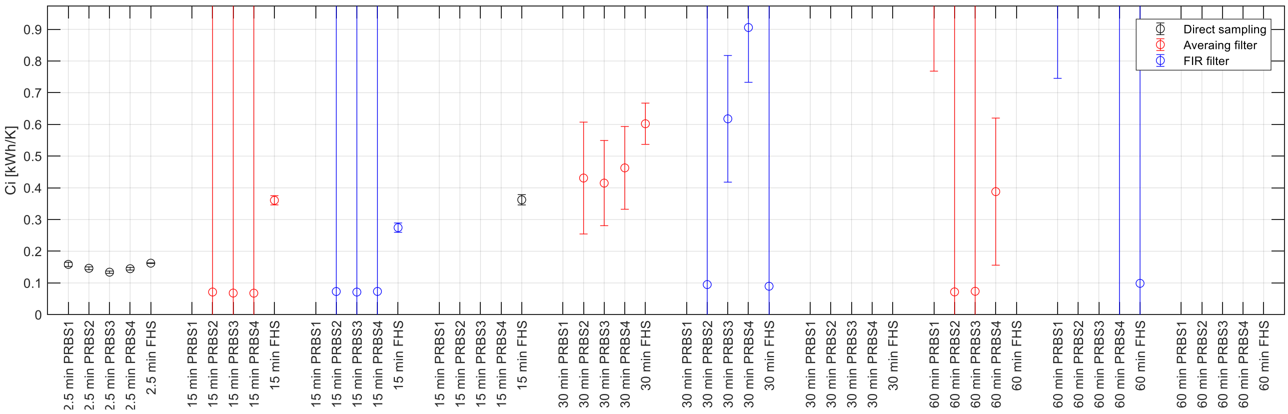


Figure 6: Identified Ci of the 3R2C model

direct sampling starts to depart from the estimated value. Figure 5 shows the value and variance of the heat capacitance of the external wall C_e. Regarding the value of C_e, direct sampling has the tendency to generate a larger capacitance value with increasing sampling time. Some values are not visible because completely outside the y-axis limits of the graph. The same problem is even more pronounced for the value of the heat capacitance C_i in Figure 6. The value of C_i diverges quickly when Ts is increased for every pre-filtering method. Although it shows that the low-pass filter, especially the moving-average, can improve the results of identified value for these key parameters. Regarding the variance of the parameters, it is very limited for the sampling time of 2.5 min. Like the parameter value, the parameter variance increases with the sampling time. However, this increase of the variance is less systematic and regular than for the parameter value.

Regarding the influence of filters, FIR does not show a significant advantage over the moving-average for the identification even though the FIR filter is theoretically better. On the contrary, FIR filter sometimes has worse results than the moving-average filter when Ts is large.

Another important conclusion can be found. The FHS dataset has more stable identified parameters (both values and variance) than the PRBS datasets. This shows that a dataset generated from a normal building operation over a long time period with comfortable indoor temperatures (and thus possible occupancy) can give equivalent or even better parameter identification than a short training period using a better excitation signal (here PRBS) but leading to uncomfortable indoor temperatures, probably preventing occupancy.

The simulation performance is shown in Figure 7 taking the FHS period to train the model. For the sake of the conciseness, the other cases using the other training

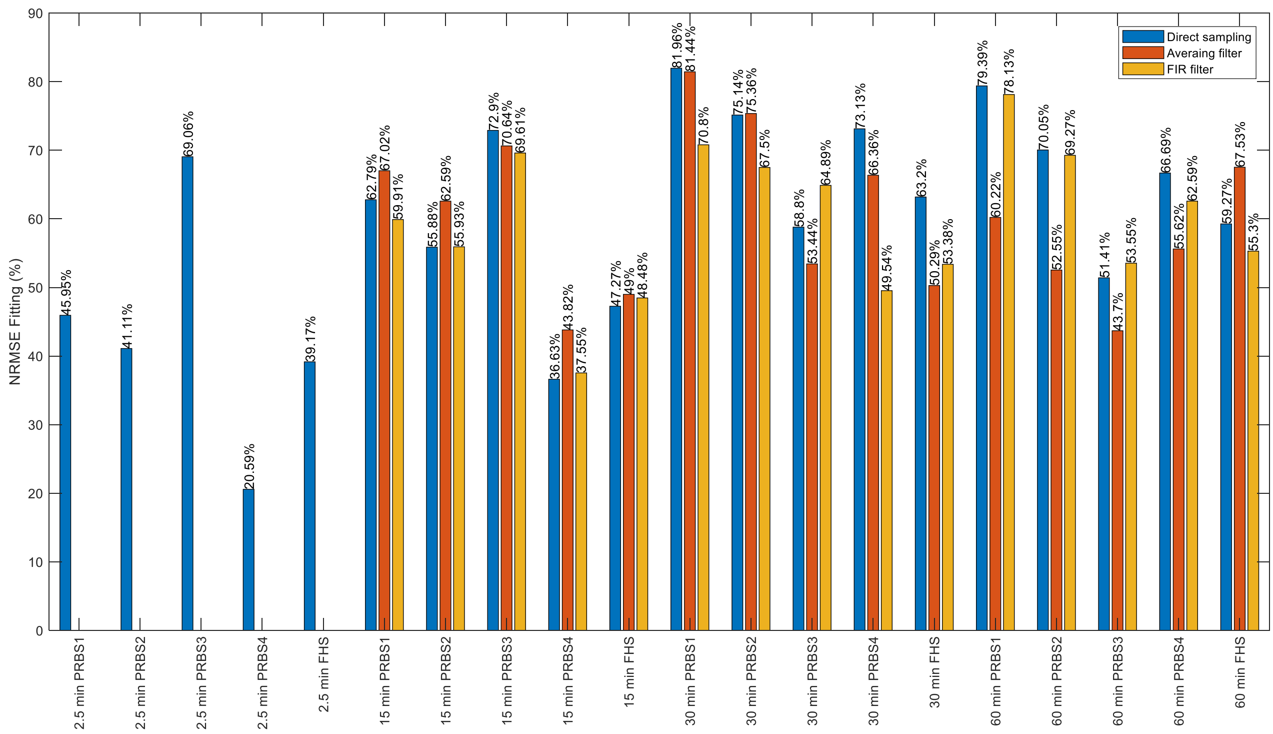


Figure 7: Simulation performance of the models trained using the FHS dataset.

periods are not reported but they give similar conclusions. Unlike, the parameter value and variance, it is clear that the increase of the sampling time (Ts) does not degrade the simulation performance. In some cases, even though the identified parameters have non-physical value or variance, this does not alter the simulation performance. The simulation performance is the main property of interest for the application of MPC. This demonstrates that training of a model for MPC application or characterization of the building thermal properties does not require the same quality of the input data. For instance, the pre-filtering methods (MA, FIR or direct sampling) do not affect much the simulation performance as well. It is difficult to rank the three pre-filtering methods as their relative performance changes between the validation cases.

Conclusions

The sampling time (Ts) of data should be limited to guarantee the physical meaning of the parameter value and variance. Larger Ts can result in non-physical parameter values and variance (Q5). If a small Ts is not applicable, the data should be low-pass filtered before being sampled even though this measure alone does not guarantee that the parameters will be physical for all Ts. This answers the first question in the introduction (Q1). More than the data pre-filtering, the selection of the right sampling time is the dominating factor to guarantee the physical meaning of the parameters. Nevertheless, sampling time and pre-filtering do not seem to affect the simulation performance of the identified models, which is a positive conclusion for MPC applications.

Even if a grey-box model has good simulation performance, having meaningful physical parameters in the model remains interesting. Firstly, it increases the physical understanding of the system, it enables to create benchmark values for other buildings of the same category. Secondly, if the parameters have not physical meaning, the model may have no additional value compared to a pure black-box model. However, to conclude this, the simulation performance of black-box models should be compared as well.

Regarding the selection of the optimizer (Q2), the results show that only the oversimple structure of the first-order model shows convexity property. Significant non-convexity already emerged from the second-order grey-box model. When applying the four different gradient-based iterative optimizers, the trained second-order grey-box model has lower NRMSE for the one-step ahead prediction compared to the model from global optimization. Therefore, it is better to use global optimization to increase the chance of avoiding a local minimum.

It is hard to say whether PRBS or FHS is a better option from the results that we observe. Since it also depends on the target period of the model (better fitting on a certain period or longer period of the FHS). However, it is clear that with a larger amount of data (longer observation period or more samples with smaller sampling time), the chance to identify a model with higher fitting and more physical parameters can be increased. This answers the third question (Q3) of the introduction. The data quality does influence the identified results of the grey-box model. Nevertheless, it is not always realistic to use the PRBS signal to excite the building's thermal mass with normal occupancy in the residential building. Data from normal operation (here intermittent on-off heating) over long periods seems more accessible. The results of this paper also show that an acceptable model can be obtained with normal building operations if large amount of data is accessible.

The selection of the correct input and output is also important for system identification (Q4). In the case study, the identified results from volume-averaged temperatures are better than those from the extracted air temperature. This proves that the correct selection of the representative indoor temperature of the building can increase the model quality and that choosing the extracted air temperature does not systematically give the best performance.

This work has answered some questions for the identification of stochastic grey-box models. However, the data in this paper is based on the results of virtual experiments without measurement noise. For future work, it will be worth investigating the influence of the measurement noise on the identification results. In addition, complementary pre-processing methods to increase the chance to identify parameters with a physical value is also an interesting topic.

Acknowledgement

The authors would like to thank the Norwegian Research Council and industry partners as this work was done in the framework of the Norwegian Research Centre on Zero Emission Neighbourhoods in Smart Cities (ZEN).

References

[1] A. Losi, P. Mancarella, A. Vicino, Integration of demand response into the electricity chain: challenges, opportunities, and Smart Grid solutions, John Wiley & Sons, 2015.

[2] P. Siano, Demand response and smart grids — A survey, Renew. Sustain. Energy Rev. 30 (2014) 461–478. https://doi.org/10.1016/j.rser.2013.10.022.

[3] J. Hu, H. Morais, T. Sousa, M. Lind, Electric vehicle fleet management in smart grids: A review of services, optimization and control aspects, Renew. Sustain. Energy Rev. 56 (2016) 1207–1226. https://doi.org/10.1016/j.rser.2015.12.014.

[4] G. Steindl, W. Kastner, V. Stangl, Comparison of Data-Driven Thermal Building Models for Model Predictive Control, J. Sustain. Dev. Energy, Water Environ. Syst. 7 (2019) 730–742. https://doi.org/10.13044/j.sdewes.d7.0286.

[5] J. Le Dréau, P. Heiselberg, Energy flexibility of residential buildings using short term heat storage in the thermal mass, Energy. 111 (2016) 991–1002. https://doi.org/10.1016/j.energy.2016.05.076.

[6] G. Reynders, Quantifying the impact of building design on the potential of structural strorage for active demand response in residential buildings, 2015. https://doi.org/10.13140/RG.2.1.3630.2805.

[7] M. Dahl Knudsen, S. Petersen, Demand response potential of model predictive control of space heating based on price and carbon dioxide intensity signals, Energy Build. 125 (2016) 196–204. https://doi.org/10.1016/j.enbuild.2016.04.053.

[8] B.P. Esther, K.S. Kumar, A survey on residential Demand Side Management architecture , approaches , optimization models and methods, Renew. Sustain. Energy Rev. 59 (2016) 342–351. https://doi.org/10.1016/j.rser.2015.12.282.

[9] T.P. Bohlin, Practical Grey-box Process Identification, Springer London, 2006. https://doi.org/10.1007/1-84628-403-1.

[10] J.A. Crabb, N. Murdoch, J.M. Penman, A simplified thermal response model, Build. Serv. Eng. Res. Technol. 8 (1987) 13–19.

[11] J.F. van Impe, P.A. Vanrolleghem, D.M. Iserentant, Advanced instrumentation, data interpretation, and control of biotechnological processes, Springer Science & Business Media, 2013.

[12] K. Arendt, M. Jradi, M. Wetter, C.T. Veje, ModestPy: An Open-Source Python Tool for Parameter Estimation in Functional Mock-up Units, in: Proc. Am. Model. Conf. 2018, Oct. 9-10, Somb. Conf. Center, Cambridge MA, USA, 2019: pp. 121–130. https://doi.org/10.3384/ecp18154121.

[13] A. Afram, A.S. Fung, F. Janabi-Sharifi, K. Raahemifar, Development and performance comparison of low-order black-box models for a residential HVAC system, J. Build. Eng. 15 (2018) 137–155. https://doi.org/10.1016/j.jobe.2017.11.021.

[14] L. Ljung, A.G. Wills, Issues in sampling and estimating continuous-time models with stochastic disturbances, in: IFAC Proc. Vol., 2008. https://doi.org/10.3182/20080706-5-KR-1001.0271.

[15] S. Norge, NS 3700: 2013 Criteria for passive houses and low energy buildings-Residential buildings, (2013).

[16] T. Johnsen, K. Taksdal, J. Clauß, X. Yu, L. Georges, Influence of thermal zoning and electric radiator control on the energy flexibility potential of Norwegian detached houses, E3S Web Conf. 111 (2019) 06030. https://doi.org/10.1051/e3sconf/201911106030.

[17] A.C. Bøeng, B. Halvorsen, B.M. Larsen, Kartlegging av oppvarmingsutstyr i husholdningene, Rapp. 2014/45. (2014). https://www.ssb.no/energi-og-industri/artikler-og-publikasjoner/kartlegging-av-oppvarmingsutstyr-i-husholdningene.

[18] S. Norge, SN/TS 3031: 2016 Energy performance of buildings, Calc. Energy Needs Energy Supply. (2016).

[19] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy Build. 43 (2011) 1511–1522. https://doi.org/10.1016/j.enbuild.2011.02.005.

[20] N.R. Kristensen, H. Madsen, S.B. Jørgensen, Parameter estimation in stochastic grey-box models, Automatica. 40 (2004) 225–237. https://doi.org/10.1016/j.automatica.2003.10.00.

[21] R.E. Hedegaard, T.H. Pedersen, M.D. Knudsen, S. Petersen, Towards practical model predictive control of residential space heating: Eliminating the need for weather measurements, Energy Build. 170 (2018) 206–216. https://doi.org/10.1016/j.enbuild.2018.04.014.

[22] R.E. Hedegaard, T. Pedersen, M.D. Knudsen, S. Petersen, Identifying a comfortable excitation signal for generating building models for model predictive control: a simulation study, in: CLIMA2016 12th REHVA World Congr. World Congr., Aalborg Universitet, 2016. https://doi.org/https://doi.org10.3384/ecp18154121.

[23] X. Yu, L. Georges, M.D. Knudsen, I. Sartori, L. Imsland, Investigation of the Model Structure for Low-Order Grey-Box Modelling of Residential Buildings, in: Proc. Build. Simul. 2019 16th Conf. IBPSA, International Building Performance Simulation Association (IBPSA), 2019. https://doi.org/10.26868/25222708.2019.211209.

[24] O.M. Brastein, D.W.U. Perera, C. Pfeifer, N.O. Skeie, Parameter estimation for grey-box models of building thermal behaviour, Energy Build. 169 (2018) 58–68. https://doi.org/10.1016/j.enbuild.2018.03.057.

[25] K.J. Åström, Introduction to stochastic control theory, Courier Corporation, 2012.

[26] Advanced Metering System (AMS) Status and plans for installation per Q2 2016, 2016. https://www.nve.no/energy-market-and-regulation/retail-market/smart-metering-ams/.