Proceedings of BS2013:
13th Conference of International Building Performance Simulation Association, Chambéry, France, August 26-28

ROBUSTNESS OF REDUCED-ORDER MODELS FOR PREDICTION AND SIMULATION OF THE THERMAL BEHAVIOR OF DWELLINGS

Glenn Reynders[1,2,3], Thomas Nuytten[1,2], Dirk Saelens[1.3]
[1] Energyville, Dennenstraat 7, BE-3600 Genk, Belgium
[2] VITO, Unit Energy Technology, Boeretang 200, BE-2400 Mol, Belgium
[3]Building Physics section, Department of Civil Engineering, Faculty of Engineering, KU Leuven, Kasteelpark Arenberg 40 – box 2447, BE-3001 Leuven, Belgium

ABSTRACT

The integration of buildings in a Smart Grid environment, enabling demand-side management and thermal storage, requires robust reduced-order building models that (i) allow simulation of the energy demand of buildings at a grid-level and (ii) contribute to the development of demand-side management control strategies.

System identification is carried out to identify suitable reduced-order models that are able to predict and simulate the thermal response of a residential building. Both grey-box models, based on physical knowledge, and statistical black-box models are considered and identified on data obtained from simulations with a detailed physical model, deployed in the Integrated District Energy Assessment Simulation (IDEAS) package in Modelica.

The robustness of identified black-box and grey-box models for day-ahead predictions and simulations of the thermal response of a dwelling is analysed. Whereas accurate day-ahead predictions are obtained for both grey-box and black-box models, the simulated indoor temperatures for the grey-box models tend to gradually deviate from the validation data. Thereby the influence of the data period used for the identification process is found to be of significant importance.

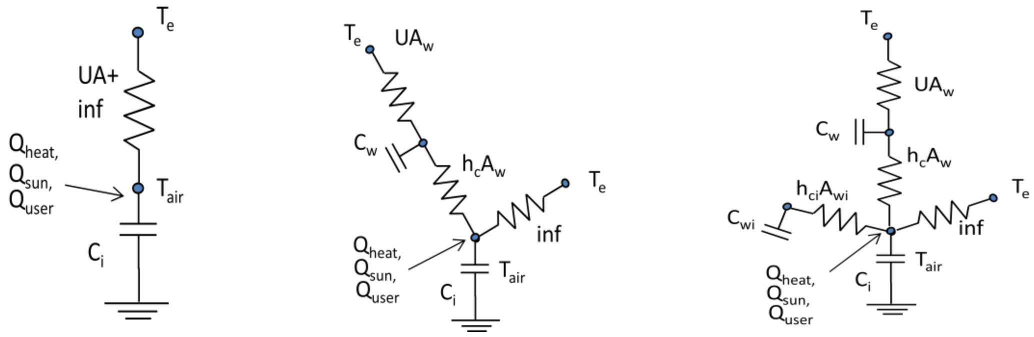
INTRODUCTION

In order to allow a more sustainable integration of renewable energy sources in the electricity network and avoid possible stability problems resulting from the mismatch between the demand and supply, development of intelligent networks or Smart Grids is suggested. Smart Grids integrate real-time communication between actuators on both demand and supply side to enable demand-side management and the use of storage technologies. In this context of Smart Grids, buildings can be of significant importance since the thermal mass of the building may be actively used as an active thermal storage capacity to enable demand-side management of the energy demand for heating (Reynders et al. 2013).

To activate the thermal mass of buildings active control of the indoor temperature, e.g. by means of model predictive control (MPC), is required. MPC has shown significant potential to improve the efficiency of heating systems and increase the penetration of renewable energy, taking into account time-of-use electricity rates and the availability of passive gains (Kintner-Meyer & Emery 1995). However a strong dependence is found of the potential savings to the coupling of the building with the outdoor environment and the efficiency of the control strategy, since the activation of the structural storage gives rise to increased conduction losses (Braun 2003). Therefore efficient application of MPC demands for reduced-order building models that are able to accurately predict the future heat demand of the building with a minimal computational effort (S Liu & G. Henze 2004).

In general three approaches can be followed in order to derive reduced-order models. A first group of reduced-order models is represented by physical *white-box* models that simulate the heating demand by simplified physical equations using solely physical knowledge about the system and material properties. Lumped capacity models are well known examples mostly represented by electric network analogies (Gouda et al. 2000). Where these white-box models make it possible to analyse the physical behaviour of buildings and are therefore interesting for research purposes, the accuracy of control strategies relying on these physical models have not been satisfactory since the real building parameters tend to deviate from the physical values used during the control design. Therefore the use of statistical input-ouput models (*black-box models*) that have self-learning capabilities is suggested (Chen et al. 2006; Cigler & Prívara 2010; Simeng Liu & G. P. Henze 2006). These black-box models do not require any prior knowledge about the system. Instead they rely purely on statistical data analysis. However, a substantial amount of data might be required to achieve the accuracy needed for model predictive control. Moreover, the resulting parameters do not necessary have a physical meaning and can therefore not be extrapolated to other buildings with the same physical properties.

To overcome this problem *grey-box* models are introduced. Grey-box models rely on physical knowledge about the system dynamics to define the model structure using stochastic differential equations. Statistical methods are then used to estimated the unknown parameters. These parameters may be directly linked to the physical properties of



(a) Model A (b) Model B (c) Model C

Figure 1 RC network representation of grey-box models

the building, given that the model structure correctly represents the physical behaviour of the system (Xu & Wang 2008; Bacher & Madsen 2011).

The presented work compares the robustness of identified black-box and grey-box models for day-ahead predictions and simulations of the thermal response of a dwelling. The day-ahead period is representative for the forecast period that is typically found in MPC. In addition, robust reduced-order models are needed to evaluate demand-side management measures on a district level.

The influence of the data set used for the identification process is analysed. Thereby it is investigated whether robust models can be derived based on simple measurements in an occupied dwelling. The added value of dedicated identification experiments, using Pseudorandom binary sequences (PRBS) for the heating system, is evaluated.

Finally the physical interpretability of the parameters of the grey-box models is evaluated. The impact of the simplifications used to develop the model structure is analysed.

METHODOLOGY

Detailed model

The data used for the identification of reduced-order building models is obtained from detailed simulations of a detached single family dwelling. The detailed simulations are carried out using the IDEAS tool developed at KU Leuven. The IDEAS tool is implemented in the object-oriented modelling language Modelica and expresses transient thermal processes in detail based on the control volume method (CVM) as described in (Baetens et al. 2012).

Table 1
Thermal properties implemented in detailed model

	A [m²]	U [W/(m²K)]	C [J/(m²K)]
External walls	200	0.21	2.743E+05
Windows N	16.2	1.10	
Windows E	14.7	1.10	
Windows S	20.2	1.10	
Windows W	11.8	1.10	
Floor on ground	132	0.12	4.567E+05
Roof	152	0.12	3.903E+05
Internal walls	120	1.82	1.843E+05
Internal floors	132	1.75	4.595E+05

The dwelling, an example of a typical Belgian detached single family house, has a floor area of 131 m² and a volume 741 m³. The building envelope has a total surface area of 365 m². The external walls are masonry walls with cavity insulation to assure both thermal resistance and the accessibility of the thermal mass. The EPS-insulation layer has a thickness of 0.15 m, resulting in a U-value of 0.21 W/m²K for the exterior walls. Thermal losses to the ground are reduced by an under-floor insulation layer. A well insulated, concrete structure is used for the flat roof. The thermal properties of the different building components are summarized in table 1.

The model is simulated for the heating dominated climate of Uccle (Belgium) as a single zone building, assuming a uniform air temperature for the whole building. Thereby 1-minute data are used for the boundary conditions. The output is generated with a sample time of 5 min.

An ideal heating system is implemented with a nominal power of 8 kW. The use of ideal heating results in an instantaneous response of the system. Consequently the dynamics of the heating system should not be included in the reduced-order models. Note that in real experiments the dynamics and efficiency of the heating system should be taken into account when using measurements of the energy use for heating as an input to the system.

Grey-box modelling

Grey-box models consist of a set of continuous stochastic differential equations formulated in a state space form that is derived from the physical laws which define the dynamics of the building. The unknown parameters in these equations are derived using estimation techniques. The grey-box approach is interesting from a research perspective as the parameters of these models may be directly interpreted as physical properties (Madsen 2008). The model structure is formulated in a state space form, given by equation 1.

$$d\boldsymbol{X}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{X}(t) + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{U}(t) + \boldsymbol{\sigma}(\boldsymbol{\theta})d\boldsymbol{\omega} \quad (1.)$$

In this equation $\boldsymbol{X}(t)$ is the state vector of the dynamic system. In the case of thermal models in this paper these states correspond to the temperatures of different building components.

$\boldsymbol{U}(t)$ is a vector containing the measured inputs of the system. These inputs can be controllable, such as the heat delivered by the heating system or the air flow rate of the ventilation system, or not controllable, such as the outdoor temperature, solar gains, internal gains... ω is a random function of time (Wiener process).

The measured output of the system $\boldsymbol{Y}(t)$ is given in equation 2 as a function of the states $\boldsymbol{X}(t)$ and the inputs $\boldsymbol{U}(t)$. ϵ is the measurement error.

$$\boldsymbol{Y}(t) = \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{X}(t) + \boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{U}(t) + \epsilon \qquad (2.)$$

The parameters $\boldsymbol{\theta}$ are estimated using the Continuous Time Stochastic Modelling (CTSM) toolbox implemented in the statistical software R (Kristensen & Madsen 2003). CTSM uses maximum likelihood estimation (MLE) to find the unknown parameters for a given model structure. The model structures are derived from resistance capacitance (RC) networks, analogue to electric circuits. Thereby the distributed thermal mass of the dwelling is lumped to a discrete number of capacitances, depending on the model order. In this work 1st , 2nd and 3rd-order models, referred to as model A, B and C (figure 1), have been investigated. In all cases the air temperature is used as the observation variable. As a simplification, the internal and solar gains, as well as the heating are directly injected to this air node.

Black-box models

Whereas grey-box models can be defined as internal models since they can give insight to the internal states and dynamics of the system, black-box models purely concern the input-output relation of a dynamic system. As such the parameters in these models have no direct relation to physical properties, but result from a pure statistical relation between the input and output of the system.

In this paper auto-regression models with exogenous input (ARX-models) are used to predict the indoor temperature based on input data for the outdoor temperature, heating input, internal gains and solar gains. The general formulation of a discrete ARX-model of order n is given by equation 3

$$\begin{aligned} y(k) + \beta_{n-1}y(k-1) + \cdots + \beta_1 y(k-n) \\ = \boldsymbol{\alpha_m}\mathbf{u}(k-n+m) \\ + \boldsymbol{\alpha_{m-1}}\mathbf{u}(k-n+m-1) \\ + \cdots + \boldsymbol{\alpha_1}\mathbf{u}(k-n+1) \\ + \boldsymbol{\alpha_0}\mathbf{u}(k-n) \end{aligned} \qquad (3.)$$

with y(k) the output at instance k, u(k) a vector containing the inputs of the system, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ the regression coefficients.

The identification process is carried out for 3 models (I, II and III) characterized by an increasing amount of input signals $\boldsymbol{u}(k)$ that is taken into account. Model I only takes the outdoor temperature and the heating input into account. For model II the influence of solar gains, both direct and diffuse, is included. Finally, Model III also integrates the internal gains. For all models the maximum order (n) is 192, corresponding a delay of 48 h, with a sample time of 15 min. However to reduce the amount of parameters, while maintaining the maximum lag of 48 h, a step of 1 h is used between 6 and 48 hours.

The regression coefficients are estimated using maximum likelihood estimation. Thereby a backward selection procedure is used to identify a suitable model of the lowest possible order. The backward selection procedure reduces the number of parameters using a stepwise elimination process that minimises the Akaike information criterion (AIC) given by equation 4

$$AIC = -2\log(L) + 2k \qquad (4.)$$

with L the value of the likelihood function for the estimated model and k the number of parameters. As such, the goodness of fit, expressed by the likelihood function, is penalised by the number of parameters in the model, avoiding to complex models. Each subsequent step removes the least contributable parameter in the model. The procedure ends when removing a parameter increases the AIC (Akaike 1974). As such, the model is obtained that has the highest goodness of fit with the smallest amount of parameters.

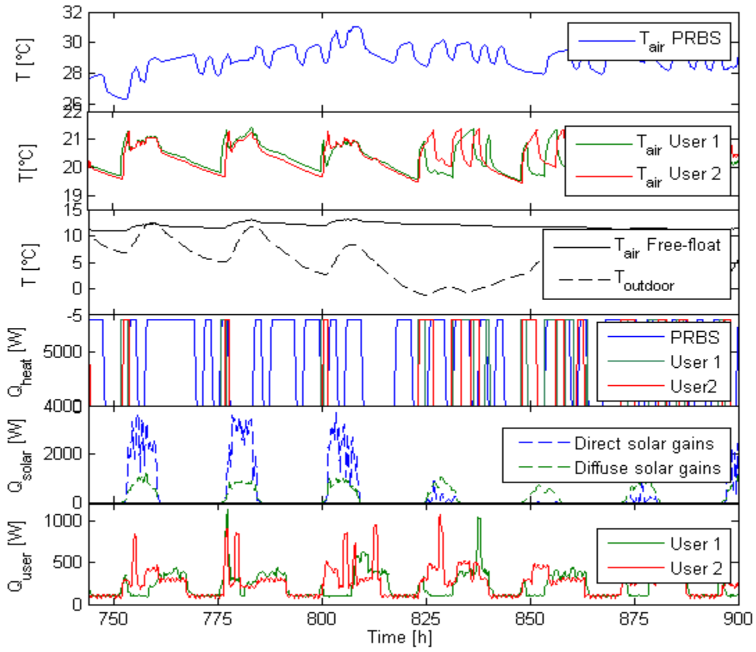


Figure 2 Sample of the input data obtained by detailed simulation

Table 2
Data periods selected for identification

Data set	Period
Dat1	14 Feb. – 21 Feb.
Dat2	14 Apr. – 21 Apr.
Dat3	14 Feb. – 28 Feb.
Dat3	14 Apr. – 28 Apr.
Dat5	1 Feb. – 28 Feb.
Dat6	1 Apr. – 28 Apr.

Table 3
Root of mean squared residuals obtained for a 1-step prediction, 1-day ahead prediction and simulation with the grey-box models identified for the different data sets. Empty cell correspond to cases that did not convergence.

Model type	Data period	**1 step prediction**				**1 day prediction**				**simulation**			
		Free	PRBS	User 1	User 2	Free	PRBS	User 1	User 2	Free	PRBS	User 1	User 2
A	Dat1	0.0071	0.044	0.050	0.053	0.27	1.46	0.76	0.65	0.26	1.63	0.87	0.79
	Dat2	0.0111	0.045	0.017	0.017	0.38	1.12	0.36	0.36	0.98	1.06	0.81	0.73
	Dat3	0.0095	0.046	0.049	0.048	0.37	1.52	0.68	0.55	1.52	2.24	0.79	0.73
	Dat4	0.0099	0.044	0.027	0.028	0.34	1.19	0.56	0.54	1.30	1.16	1.10	1.04
	Dat5	0.0101	0.046	0.050	0.049	0.39	1.39	0.63	0.54	2.19	1.76	0.91	0.83
	Dat6	0.0121	0.045	0.024	0.025	0.46	1.28	0.53	0.55	3.16	2.26	1.92	1.98
B	Dat1	0.0062	0.006	0.033	0.026	0.07	0.34	0.53	0.22	0.08	0.51	0.69	0.22
	Dat2	0.0027		0.011	0.010	0.13		0.19	0.15	0.18		0.50	0.30
	Dat3	0.0069	0.006			0.07	0.32			0.12	0.97		
	Dat4	0.0025	0.006	0.018	0.015	0.10	0.28	0.18	0.18	0.23	0.56	0.35	0.69
	Dat5	0.0024	0.006			0.11	0.28			0.34	1.05		
	Dat6	0.0028	0.006	0.013	0.014	0.11	0.28	0.16	0.15	0.49	0.92	0.74	0.65
C	Dat1	0.0068	0.006	0.028	0.026	0.48	0.23	0.16	0.16	0.48	0.52	0.32	0.17
	Dat2	0.0026	0.006		0.024	0.33	0.28		0.64	0.97	0.27		1.43
	Dat3	0.0022	0.006	0.027	0.025	0.30	0.50	0.14	0.13	1.35	1.21	0.42	0.20
	Dat4	0.0024	0.006	0.014	0.014	0.28	0.39	0.24	0.26	1.44	0.40	0.69	0.85
	Dat5	0.0023	0.006	0.028	0.025	0.34	0.28	0.16	0.14	2.18	1.04	1.04	0.46
	Dat6	0.0027	0.007			0.41	0.44			4.44	2.68		

Data

The data for the identification process is generated using the detailed building model. In total 4 data sets are generated, corresponding to 4 different virtual experiments: (1) Free-floating, (2) Dynamic heating using PRBS-signal, (3) In-use data for user 1 and (4) In-use data for user 2.

The first two data sets represent experiments during unoccupied periods. Thereby the free-floating experiment corresponds to measurements in an empty building without heating. Note that, although a free-floating experiment is a simple procedure, it is not expected to give good results for the identification of a dynamic building model, due to the absence of higher excitation frequencies. In contrast, the dynamic heating experiment uses a Pseudo-Random Binary Sequence (PRBS) for the on/off control of the heating system. The signal also excites higher frequencies and is generated to maintain the state of the heating system (on/off) for at least 90 min and has a total period of 6 weeks.

In addition to these dedicated experiments, two data sets are generated corresponding in-use building data. Thereby the set point for heating is 20.5°C during the occupied period with a setback to 16°C during un-occupied periods. A hysteresis control of ±0.5°C is used for the heating system. The occupied periods follow a deterministic pattern and differ between the two user profiles, as shown in figure 2. The profile of user 1 represents the behaviour of a family that is out for work throughout the day, whereas user 2 represents a case with occupancy throughout the entire day. The internal gains are calculated from a corresponding stochastic profile.

In order to evaluate the influence of the period in which the data is collected on the identified models, 6 data periods have been selected for the 4 data sets. Thereby winter and mid-season periods have been selected to get different contributions of the heating and the solar gains. The length of the periods varies from 1 week to 1 month as shown in table 2

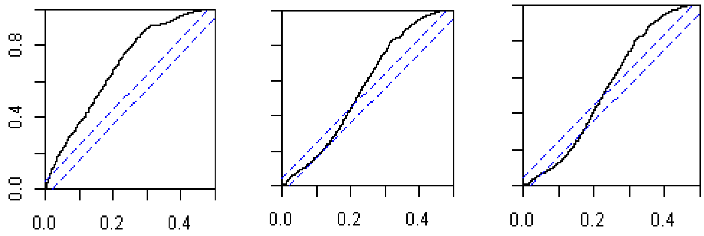
RESULTS

In order to evaluate the quality and robustness of the models, 3 validation steps are taken. Firstly, the statistic properties of the residuals are analysed. Thereby both the root of the mean squared residuals as well as the autocorrelation of the residuals are investigated. Secondly, the models are used for day ahead prediction and simulation. Thereby only the models that showed good performance in the residual analysis are used. Finally, the parameters of the grey-box models are compared against the physical properties of the building. In all cases, the influence of the data period and the type of virtual experiment is analysed.

Residual analysis

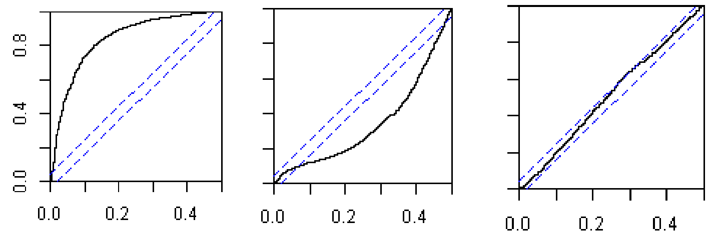
Tables 3 and 4 show the root of the mean squared error (RMSE) for the 1-step prediction, day-ahead prediction and simulation of the identified models for each data set. The errors for 1-step prediction are analysed since they correspond to the residuals obtained by the Prediction Error Method, that is used to estimate the parameters, and thus indicate the goodness of fit. In addition, RMSE values for the day-ahead prediction and the simulation are analysed, quantifying the uncertainty that can be expected in MPC application and simulation on network level using the identified models.

Table 3 shows decreasing residuals for the 1-step prediction as the order of the grey-box model



(a) Model I (b) Model II (c) Model III

Figure 3 Cumulative periodogram for ARX- models I,II and III estimated on the data set from User 1



(a) Model A (b) Model B (c) Model C

Figure 4 Cumulative periodogram for grey-box models A,B and C estimated on the data set from User 1

increases. This effect is explained by the increasing degrees of freedom that is available to optimize the cost function in the parameter estimation method. This effect is generally found for increasing model orders and should be interpreted carefully as consequently increasing the model order might affect the robustness due to overfitting. Overfitting occurs when a model is excessively complex and describes random error in the data instead of the underlying physics. In literature, a number of tests such as the Akaike information criteria (AIC) are found to decide on the optimal model order and avoid over-fitting.

The effect of overfitting is shown for the day-ahead prediction and simulation with the free-floating data set. In both cases the residuals increase when the model is extended from model B to C. A similar effect is shown for model C fitted on the in-use building data (User 1 and User 2). Here the smallest residuals in a one-step prediction are obtained for the data periods in mid-season (Dat 2, 4 and 6). During these periods the heating input, responsible for most of the high-frequent dynamics, is small. Consequently, overfitting occurs when the third order models is fitted on mid-season data, shown by large errors for day-ahead prediction and simulation, indicating that free-floating measurements do not provide adequate dynamics for system identification.

Note that for the PRBS-signal the inverse effect is shown, since for this virtual experiment both heating and high solar gains are available in mid-season.

The magnitude of the residuals for the ARX-models (table 4) is comparable to the residuals obtained for the third order grey-box models. Nevertheless, the black-box models appear to be more robust for simulations as the increasing trend for longer data periods is no longer found. In contrast, longer data periods improve the model robustness when the identification is done on the in-use data set, especially for the mid-season data.

In addition to the RMSE values, an important statistical test to evaluate the quality of a model is to analyse the autocorrelation in the residuals for one-step predictions. If all dynamics in the data are explained by the model, no autocorrelation should appear in the residuals as they are assumed to be white-noise.

To evaluate this assumption the scaled cumulated periodograms of the residuals are analysed (figure 3 and 4). For white noise, the variation of the residuals is uniformly distributed over all frequencies. As such, the theoretical cumulated periodogram for white noise is a straight line through the origin. A 95%-confidence interval (dotted lines) is obtained based on the Kolmogorov-Smirnov test (Madsen 2008). For the assumption of white-noise residuals to be valid, the cumulated periodogram needs to lie within this confidence interval.

The cumulated periodogram for the grey-box models (figure 4) show significant autocorrelation for models A and B, indicating that there are still dynamics in the data that are not explained using these models. For model A the cumulated periodogram indicates that especially low frequency dynamics are not well explained by the model. By separating the low-frequent behaviour of the walls and the high-frequent response of the air capacity (model B), the amount of autocorrelation is slightly reduced. However, a third capacity is required to obtain white-noise residuals as introduced by model C.

The cumulated periodograms for the ARX-models (figure 3) shows that the autocorrelation significantly reduces when the model is extended from model I to II, by taking into account the solar gains. The influence of further extending the model by taking into account the internal gains is limited.

Cross-validation and simulation

In addition to the residual analysis, the robustness of the models is evaluated by cross-validation and

Table 4
Root of the mean squared residuals for ARX-model III

	PRBS			User 1		
	1-step	day-ahead	simulation	1-step	day-ahead	simulation
Dat1	0.010	0.062	0.540	0.026	0.050	0.220
Dat2	0.010	0.122	1.423	0.028	0.087	0.851
Dat3	0.010	0.085	0.861	0.025	0.053	0.341
Dat4	0.010	0.069	0.621	0.026	0.056	0.369
Dat5	0.010	0.085	0.861	0.024	0.053	0.341
Dat6	0.010	0.073	0.728	0.024	0.040	0.226

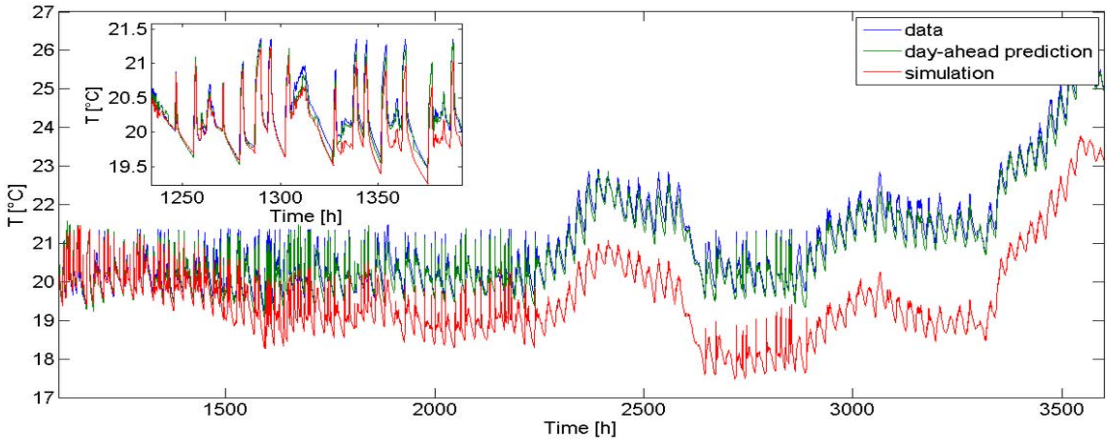


Figure 5 Cross-validation and simulation of grey-box model C for the period of 14th February - 30th the May

simulation.

Cross-validation is carried out by performing a day-ahead prediction of the indoor temperature for a data from 14th February until 30th May, assuming perfect prediction of the future inputs. As such, the model is tested in both cold and warm climatic conditions. The same data period is used for simulation.

The results of both day-ahead prediction and simulation for the grey-box model C, estimated on data period 3, are shown in figure 5. This model was chosen as it showed the smallest RMSE in table 3. Whereas the results indicate that the model can be used for day-ahead predictions with a strong accuracy, a root mean squared error of 0.14 °C, a gradual deviation from the correct solution is obtained when the model is used for simulation. This deviation indicates that the low-frequent dynamics are not correctly estimated. Consequently, the use of the grey-box models for estimation demands for further extending of the model order. However it has to be noted that for the data used to estimate the grey-box models, white-noise residuals are obtained. This indicates that the model explains all the dynamics that is available in the data. Further extending the model will therefore require information from additional variables as for example heat flux measurements or surface temperatures of the walls.

For the ARX-models a similar deviation is shown (Figure 6). However the residuals for both day-ahead prediction and simulation are smaller, indicating a higher robustness of the ARX-models. For simulation the RMSE-value of the ARX-model III is 0.2°C.

As such, it can be concluded that both grey-box and black-box models can be used for accurate day-ahead predictions of the dynamic behaviour of the dwelling. For the grey-box models a 3rd-order model is able to capture the dynamics needed for forecasting over a limited timeframe, i.e. up to a few days. For long-term predictions and simulation the grey-box models tend to drift away from the correct solution, since the low-frequent dynamics are not well explained. For the ARX-models the drift is less pronounced, given the data set used for identification contains both cold and warm periods.

Physical properties of parameters in grey-box

An important benefit of grey-box models is the physical interpretation which may be given to the model parameters. As the model structure is derived from prior knowledge about the physics of the system, the parameters may be directly linked to the physical properties of the system, in this case the thermal properties of the building. However, one can expect that the physical interpretability of the estimated parameters are strongly influenced by the assumptions and simplifications used for defining the model structure. Therefore this section compares the estimated parameters for the different model

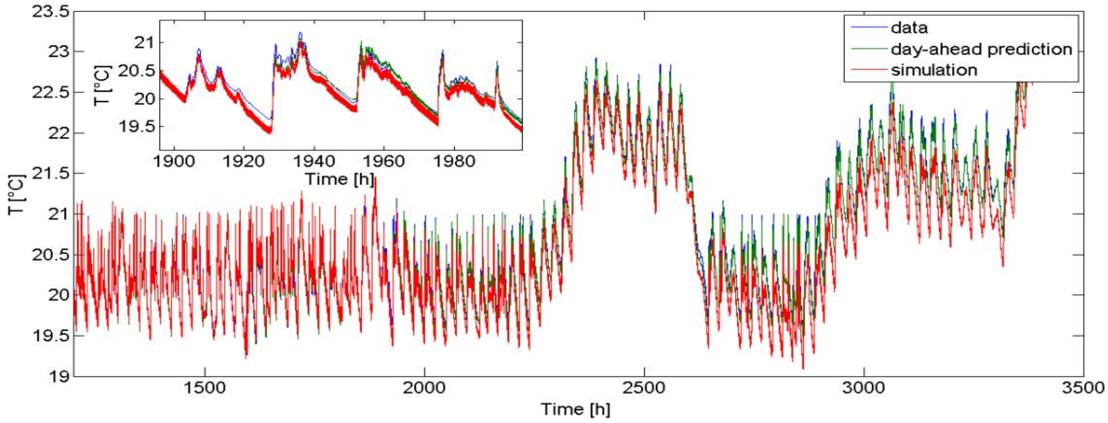
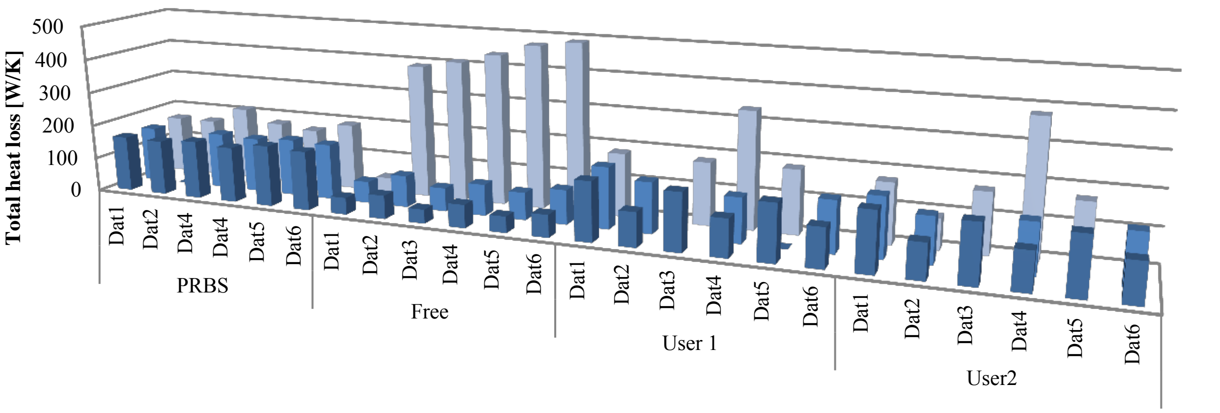


Figure 6 Cross-validation and simulation of ARX-model III for the period of 14th February - 30th the May



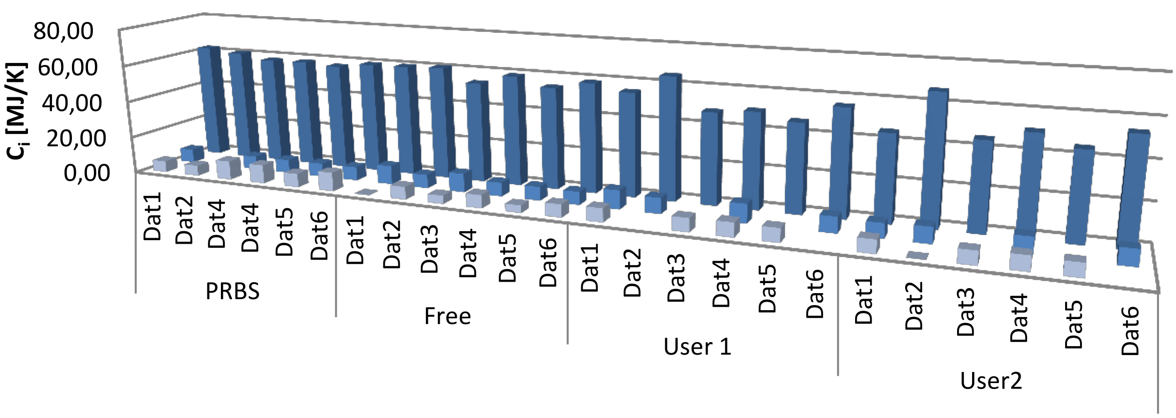
	PRBS						Free						User 1						User2					
	Dat1	Dat2	Dat4	Dat4	Dat5	Dat6	Dat1	Dat2	Dat3	Dat4	Dat5	Dat6	Dat1	Dat2	Dat3	Dat4	Dat5	Dat6	Dat1	Dat2	Dat3	Dat4	Dat5	Dat6
A	162	160	170	162	178	173	49	68	40	67	47	65	171	99	166	108	163	113	171	101	167	109	165	113
B	161	0	164	159	167	163	63	92	68	91	80	99	177	146		129	0	148	171	131		143		145
C	167	167	214	178	167	194	41	390	411	439	474	490	183		181	336	184		173	89	173	381	172	

Figure 7 Estimated total heat loss coefficient. Empty cells correspond to cases that did not converge

structures and the different data sets.

Figure 7 shows the estimated values of the total heat loss coefficient of the building for the different models and data periods. Firstly, it is shown that the use of free-float data results in an unreliable estimation of the total heat loss coefficient. This can be explained by the limited difference between the indoor and outdoor temperature for the unheated building. The same problem occurs for the estimation of Model C on the in-use data during the mid-season periods (Dat 2 and Dat 4). Compared to the theoretical value of 145 W/K, the third order model tends to overestimate the heat loss coefficients explaining the underestimation of the indoor temperature in the simulation test (Figure 5).

A second interesting parameter is the estimate of the indoor air capacity. Figure 8 shows that this capacity is strongly overestimated for the first order model. Since no distinction is made between the thermal mass of the structure and the thermal mass of the air, the estimated capacity represents the active thermal mass corresponding to the dominating time constant of the building. When the order of the model increases the fast dynamics of the indoor air can be separated from the slow dynamics of the building fabric. However, the estimated capacity is still higher than the theoretical value for the total thermal capacity, of 4.5 MJ/K for the indoor air, that is implemented in the detailed model. This overestimation is caused by the assumption that all



	PRBS						Free						User 1						User2					
	Dat1	Dat2	Dat4	Dat4	Dat5	Dat6	Dat1	Dat2	Dat3	Dat4	Dat5	Dat6	Dat1	Dat2	Dat3	Dat4	Dat5	Dat6	Dat1	Dat2	Dat3	Dat4	Dat5	Dat6
A	62,5	60,8	58,0	58,4	57,3	59,7	60,0	60,8	53,8	59,1	54,5	58,6	55,2	65,2	48,4	50,4	46,8	56,1	45,3	66,8	45,6	51,1	45,1	54,2
B	7,28		7,24	7,23	7,24	7,25	10,0	7,24	10,0	7,30	7,29	6,97	10,0	8,35		10,0		8,38	7,73	8,43		8,90		8,43
C	5,98	5,53	10,0	9,93	7,24	9,99	0,15	6,77	4,35	6,95	3,88	6,79	7,13		7,12	7,44	7,18		6,91	0,10	6,99	7,53	6,99	

Figure 8 Estimated values for the indoor air capacity. Empty cells correspond to cases that did not converge

gains are directly attributed to the air-node, whereas in the detailed model the solar gains are distributed over the indoor surfaces of the zone and the internal gains and heating are characterized by both a radiative and convective component. Moreover the figure indicates that the use of data periods in mid-season results less reliable results. This can again be explained by the absence of heating in the data for mid-season.

Based on these results it can be concluded that the physical interpretation of the parameters should be handled carefully and is strongly influenced by the assumptions and simplification in the grey-box models. When the purpose of the system identification is to quantify the thermal properties of buildings, rather than the prediction of the dynamic behaviour, further extension of the model is required. Note however that this might require the availability of additional measurements since the 3rd-order model already results in white-noise residuals, indicating that the dynamics that are available in the data are well explained by the model. Further research is required to analyse the hypotheses.

CONCLUSION

System identification is carried out on simulation results for a single zone dwelling in order to identify both grey-box and black-box models that can be used for accurate day-ahead prediction as well as simulations of the thermal response of the dwelling.

The 4 data sets are generated using a detailed building model developed in Modelica and representing both dedicated identification experiments as measurements in occupied buildings. For all scenarios 6 sample periods are analysed with a length of 1, 2 and 4 weeks in both winter and mid-season conditions.

The influence of the data period and the virtual experiment on the quality and robustness of the identified models is analysed. Thereby both grey-box and black-box models show reliable results for day-ahead prediction. The best predictions are obtained for the 3rd-order grey-box model and the ARX-model that takes into account the outdoor temperature, solar gains, internal gains as well as the heating input, with a root of the mean squared residuals equal to 0.14 °C and 0.05 °C, respectively.

Considering long-term predictions or simulations, the temperatures obtained with the grey-box models show an important drift from the validation data, indicating that the long-term dynamics are not well described.

The results show that the data set used for identification has an important influence on the robustness of the identified models. It is shown that measurements in free-floating conditions do not contain adequate dynamics to allow the identification of robust reduced-order models. Nevertheless, the added value of a dedicated experiment using PRBS control signals for the heating does not improve the model quality significantly compared to measurements of occupied buildings.

Further research is required to evaluate the process and models discussed in this paper on multi-zone buildings as well as on actual measured data.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*.

Bacher, P. & Madsen, H., 2011. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43(7), pp.1511–1522.

Baetens, R. et al., 2012. Assessing electrical bottlenecks at feeder level for residential net zero-energy buildings by integrated system simulation. *Applied Energy*, 96, pp.74–83.

Braun, J.E., 2003. Load Control Using Building Thermal Mass. *Journal of Solar Energy Engineering*, 125(3), pp.292–301.

Chen, K., Jiao, Y. & Lee, E.S., 2006. Fuzzy adaptive networks in thermal comfort. *Applied Mathematics Letters*, 19(5), pp.420–426.

Cigler, J. & Prívara, S., 2010. Subspace identification and model predictive control for buildings. *Control Automation Robotics & Vision (…*, (December), pp.7–10.

Gouda, M.M., Danaher, S. & Underwood, C.P., 2000. Low-order model for the simulation of a building and its heating system. *Building Services Engineering Research and Technology*, 21(3), pp.199–208.

Kintner-Meyer, M. & Emery, A.F., 1995. Optimal control of an HVAC system using cold storage and building thermal capacitance. *Energy and Buildings*, 23(1), pp.19–31.

Kristensen, N.R. & Madsen, 2003. *Continuous Time Stochastic Modelling, CTSM 2.3—Mathematics Guide*,

Liu, S & Henze, G., 2004. Impact of Modeling Accuracy on Predictive Optimal Control of Active and Passive Building Thermal Storage Inventory. *HVAC&R Research*, pp.151–164.

Liu, Simeng & Henze, G.P., 2006. Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory. *Energy and Buildings*, 38(2), pp.142–147.

Madsen, H., 2008. *Time Series Analysis*,

Reynders, G., Nuytten, T. & Saelens, D., 2013. Potential of structural thermal mass for demand-side management in dwellings. *Building and Environment*, 64, pp.187–199.

Xu, X. & Wang, S., 2008. A simplified dynamic model for existing buildings using CTF and thermal network models. *International Journal of Thermal Sciences*, 47(9), pp.1249–1262.