2018 Building Performance Modeling Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA
Chicago, IL
September 26-28, 2018

COMPARATIVE ANALYSIS OF WHITE-, GRAY- AND BLACK-BOX MODELS FOR THERMAL SIMULATION OF INDOOR ENVIRONMENT: TEACHING BUILDING CASE STUDY

K. Arendt, M. Jradi, H.R. Shaker, C. T. Veje
University of Southern Denmark, Odense, Denmark

ABSTRACT

This study presents a performance comparison between selected white-, gray- and black-box models for indoor temperature prediction in a university building located at the SDU Campus Odense. It was found that the black-box models outperform the gray- and white-box models in most cases, but the accuracy highly depends on the training data in terms of both period and modes of heat transfer covered by the data set. The average mean absolute error for the best performing black-box model was 0.4°C as compared to 1.0°C and 0.7°C for the gray-box and white-box models, respectively. In terms of accuracy, the gray-box models are a reasonable alternative for black-box only in case of short-term predictions, in which their error decreases to around 0.3-0.8°C , depending on the room.

INTRODUCTION

Indoor environment modeling and simulation plays an increasingly important role in advanced HVAC control systems and fault detection and diagnostics (FDD). Simulations can be used to predict future indoor conditions based on the weather forecast and anticipated indoor occupancy patterns. The predictions can be used to optimize the control strategy, e.g. in model predictive control (MPC) systems (Henze 2013). On the other hand, simulations on the historical data can help detect faulty systems by comparing expected results with actual measurements (Turner, Staino, and Basu 2017). In both applications an accurate model of the indoor environment is needed. However, the difficulties related to the development and calibration of building models are often cited as the major obstacle hindering a widespread use of model-based solutions in buildings (Henze 2013).

The modeling approaches can be divided into white-box, gray-box and black-box. The approaches differ in the amount of physical relationships included in the models, with the white-box and black-box being entirely physics-based and data-driven, respectively. The gray-box approach is a mixture of both. The whole-building white-box models are predominant in the design stage, but all three approaches are frequently employed in different building operation applications.

Whole-building white-box models provide detailed insights into the building operation and enable to explore the relationships between the performance of different systems. Due to these features they are used to quantify the effects of faulty systems (Zhang and Hong 2017) or assess thermal renovation strategies and measures (Jradi, Veje, and Jørgensen 2018). White-box models are also used in MPC systems. Although white-box models typically are more difficult to develop than gray- or black-box models, once developed they can be more easily used for other applications like retrofit analysis or FDD (Henze 2013). However, many argue that the calibration of whole-building models is difficult and requires iterations with human-in-the-loop. In consequence, practical application of white-box models in MPC is feasible only for simple buildings (Prívara et al. 2013).

Some of the shortcomings of white-box models are addressed by gray- and black-box approaches. They are usually used to develop more specialized models, e.g. zone models used for temperature prediction (Gunay, O'Brien, and Beausoleil-Morrison 2016), indoor occupancy prediction models (Sangogboye et al. 2017), HVAC subcomponent models (Afram and Janabi-Sharifi 2015b), or overall heating and cooling load prediction models (Ahmad and Chen 2018). The gray- and black-box models are arguably easier to calibrate and more scalable than white-box (Prívara et al. 2013). The scalability addresses one of the commonly cited limitations of MPC, i.e. the significant resources, in terms of both time and expert knowledge, required to provide a reliable control model (Henze 2013).

Some of the reasons the gray-box approach is often preferred over the black-box is the lack of the physical interpretation of the results in the latter (Žáčeková, Váňa, and Cigler 2014). The gray-box models can also provide smooth solutions and be based on twice differentiable equations which make the dynamic optimization more robust, e.g. by using collocation methods (Coninck and Helsen 2016), whereas dynamic optimization in black-box models is usually based on global optimization methods, like genetic algorithms (Reynolds et al. 2018). The gray-box models also have better generalization capabilities when the test data deviates considerably from the training data (Afram and Janabi-Sharifi 2017).

Afram and Janabi-Sharifi (2015a) compared various black-box and gray-box models for HVAC system modeling, including artificial neural networks (ANN), transfer functions (TF), autoregressive exogenous models (ARX), state-space models (SS) and several gray-box models based on the general understanding of physical processes present in HVAC subcomponents. The validation on real measured data showed that ANN outperformed all other models, while the gray-box models were the least accurate.

Since each approach has valid pros and cons, and models of different types performed differently in different cases reported in the literature, a model comparison and selection is advised when implementing a model-based system in a building.

This study presents a performance comparison of selected white-, gray- and black-box models used for thermal simulation of indoor environment in two test rooms of the OU44 building located at the SDU Campus Odense. The building is equipped with sensors measuring HVAC parameters, indoor environment parameters, and climatic conditions. The models selected for the study are: (1) a white-box model implemented in EnergyPlus, (2) gray-box models based on RC thermal networks implemented in Modelica and (3) black-box models including a nonlinear autoregressive exogenous model (NARX) and a neural network model (NN). The models are trained on two training data sets and each time tested on two validation windows, selected within one month of the available measurements.

The findings of this study can aid researchers and engineers in the decision-making in terms of choosing and implementing model-based systems in buildings.

The results of this study will be used in the future implementation of a multi-objective MPC system in the OU44 building (Arendt et al. 2016).

CASE STUDY BUILDING

General description

The OU44 building is a 4-story (3 floors + basement) university building with a floor area of 8500 m^2 (Fig. 1), located at the SDU Campus Odense, Denmark. The building construction finished in November 2015.

The building is used mainly for teaching. It contains 21 classrooms, 8 study zones and several smaller rooms on the top floor working as offices. The basement comprises of technical rooms, storage facility and installations. As of 2018, it is one of the most energy efficient public buildings in Denmark, with a primary energy consumption around 42 kWh/m^2 per year.

The building is equipped with a balanced ventilation system, comprising four air handling units (AHU) providing up to 15000 m^3/h of fresh air each. Each AHU is equipped with two fans (supply/exhaust), a heat recovery wheel and a heating coil. Due to the cold climate, there is no need for mechanical cooling in the building. The fans operate to maintain constant pressures in the supply and exhaust ducts. All major rooms in the building (classrooms, study zones, offices) are equipped with VAV terminal units. The fresh air supply is controlled based on indoor CO_2 levels, with a step-wise correlation between the room's CO_2 concentration and damper position. The ventilation air temperature setpoint is kept at a constant level of 21°C, so the ventilation provides heating, but only during occupancy periods. The building heating demand is covered mainly by a district heating loop. Heat is distributed to the rooms using a hydronic system, with each room equipped with radiators located under the windows. The room temperature setpoints can be set separately in the Building Management System (BMS) and are controlled by the technical staff.

All rooms except in basement are equipped with operable windows. The windows are equipped with shading curtains controlled based on outdoor and indoor illuminance, and on wind speed (safety roll up on wind speed above 15 m/s).

Figure 1: OU44 building

Test rooms

Two test rooms were selected for this study, a study zone located on the second (middle) floor and a classroom located on the third (top) floor. Both rooms are equivalent in terms of connected systems (ventilation, heating) and sensors, but differ in the occupancy patterns and occupant behavior. The measured data used in this study spans from March 21 00:00 to April 21 00:00, 2017 (31 days) and includes: outdoor temperature T_{out}, global horizontal solar radiation H_{glo}, indoor temperatures T_r, VAV damper positions d_{pos}, radiator valve positions v_{pos}, and occupancy counts n_{occ} (Fig. 2). The occupancy counts were obtained from stereo vision cameras located above the entrances to the rooms (Sangoboye and Kjærgaard 2016).

The study zone (125 m^2) is a room used solely by students to work on their assignments and group projects. The building and the study zone are open for students 24/7, so nighttime occupancy occasionally occurs. The study zone is furnished for around 30 students, but sometimes much higher numbers are present (notice up to 60 in Fig. 2a). It has been noticed by the staff, that students often open windows in the study zones, despite the ventila-

tion controlling indoor CO_2 concentration. This happens less often in the classrooms.
The classroom (139 m^2) is a typical teaching room with a capacity of around 80 people. It has more consistent occupancy patterns compared to the study zone (Fig. 2), with around 20-40 students attending each lecture.

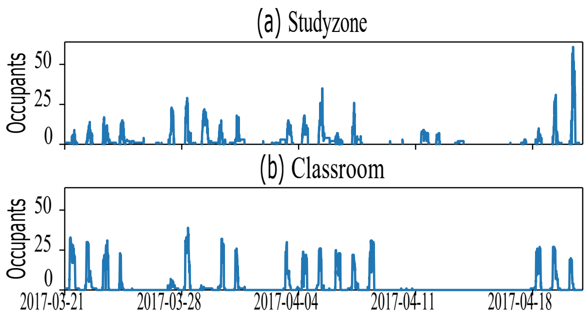

Figure 2: Number of occupants measured by stereo vision cameras

MODELING APPROACHES

The results of six models were selected to be presented in this work: an EnergyPlus model (EP), three RC thermal networks (R1C1, R2C2, R3C3), a nonlinear autoregressive exogenous model (NARX) and a feed-forward neural network (NN). The EP model represents the white-box approach, R1C1, R2C2 and R3C3 represent the gray-box approach, while NN and NARX are black-box models. The decision to include only low order RC models was motivated by (1) their low computational demand and (2) simpler parameter estimation due to a lower number of local minima and reduced risk of overfitting (Brastein et al. 2018). In the final application, the authors' intention is to combine multiple low-order gray-box models and black-box models (e.g. each model representing a separate room) for the use in a distributed optimization framework (Arendt et al. 2016).

White-box

The EnergyPlus model is a whole-building model, developed for a detailed energy performance analysis of the OU44 building (Jradi et al. 2017). The model contains 190 thermal zones, a district heating loop for heat supply and 4 ventilation units with heat recovery wheels and preheating loops.
The model, as the only one from all presented in this study, was not calibrated on the measured data from the considered period of March 21–April 21. The calibration was performed earlier based on the building energy consumption data from September–December 2016, including whole building heating consumption, ventilation units electricity consumption and lighting on the levels of building and floors, and using actual occupancy data and weather conditions. The calibrated model predicts the OU44 building energy consumption with an acceptable maximum monthly deviation of -8.4%, -6.9% and 3.9% for heating, ventilation electricity and lighting.
The model uses outdoor temperature T_{out}, horizontal global radiation H_{glo} and occupancy counts n_{occ} for inputs. The measured climate data was inserted into the climate file, while the occupancy data was used to generate the occupancy schedules in the model for both test rooms.

Gray-box

Three RC thermal networks are included in the study: R1C1, R2C2, R3C3. The models can be represented in a compact state-space form:

$$\dot{x}_m = A_m x_m + B_m u, \tag{1}$$

where $\dot{x}_m$ is the time derivative, x_m is the state vector, A_m is the state matrix, B_m is the input matrix, u is the input vector, and m is the model order. The state vectors are $x_1 = [T_r]$, $x_2 = \begin{bmatrix} T_r & T_i \end{bmatrix}^T$ and $x_3 = \begin{bmatrix} T_r & T_i & T_e \end{bmatrix}^T$ for R1C1, R2C2 and R3C3, respectively, where T_r is the room temperature [°C], T_i is the internal thermal mass temperature [°C] and T_e is the external thermal mass temperature [°C]. The state matrices are as follows:

$$A_1 = [-1/(R_{e1}C_r)], \tag{2}$$

$$A_2 = \begin{bmatrix} -1/(R_{e1}C_r) - 1/(R_iC_r) & 1/(R_iC_r) \\ 1/(R_iC_i) & -1/(R_iC_i) \end{bmatrix}, \tag{3}$$

$$A_3 = \begin{bmatrix} -(\frac{1}{R_i} + \frac{1}{R_{e1}})\frac{1}{C_r} & \frac{1}{R_iC_r} & \frac{1}{R_{e1}C_r} \\ \frac{1}{R_iC_i} & -\frac{1}{R_iC_i} & 0 \\ \frac{1}{R_{e1}C_e} & 0 & -(\frac{1}{R_{e1}} + \frac{1}{R_{e2}})\frac{1}{C_e} \end{bmatrix}, \tag{4}$$

where C_r is the room thermal capacitance [J/K], C_i is the internal thermal mass capacitance [J/K], C_e is the external thermal mass capacitance [J/K], R_i represents the internal wall resistance [K/W], and R_{e1} and R_{e2} represent the external wall resistance [K/W]. A single resistor R_{e1} is used to model the external wall in R1C1 and R2C2. Two resistors (R_{e1}, R_{e2}) and a single capacitor (C_e) are used to model the external wall in R3C3.
The disturbance vector is shared by all models:

$$u = \begin{bmatrix} T_{out} & H_{glo} & v_{pos} & q_{ve} & q_{occ} \end{bmatrix}^T, \tag{5}$$

where v_{pos} is the radiator valve position [%], q_{ve} is the heat gain due to ventilation [W] and q_{occ} is the occupancy heat gain [W]. The ventilation heat gain is a function of the damper position d_{pos} [%], ventilation air temperature T_{ve} [°C] and room temperature T_r, calculated as $q_{ve} = (d_{pos}V_{max})(T_{ve} - T_r)$, where V_{max} is the maximum airflow rate to the zone for a fully open damper [m^3/h]. The ventilation heat gain is a nonlinear term since both d_{pos} and T_{ve} can vary. In the considered period, the ventilation air temperature had a constant setpoint temperature of 21°C, but the damper position was time-varying. The occupancy heat gain is also a nonlinear term, calculated as $q_{occ} = n_{occ} g(T_i)$, where n_{occ} is the number of occupants and $g(T_i)$ is the sensible heat generation per occupant [W], given as a function of indoor temperature: 84 W for $T_i < 20$°C, 0 W for $T_i > 52$°C with a linear interpolation in-between 20 and 52°C. It should be noted that $g(T_i)$ is the authors' approximation of the relationship based on various sources. There is no measured data regarding the actual occupancy heat gains in the test building.

The input matrices B_m are as follows:

$$B_1 = \begin{bmatrix} 1/(R_{e1}C_r) & s/C_r & Q_{rad}/C_r & 1/C_r & 1/C_r \end{bmatrix}, \quad (6)$$

$$B_2 = \begin{bmatrix} 1/(R_{e1}C_r) & s/C_r & Q_{rad}/C_r & 1/C_r & 1/C_r \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (7)$$

$$B_3 = \begin{bmatrix} 0 & s/C_r & Q_{rad}/C_r & 1/C_r & 1/C_r \\ 0 & 0 & 0 & 0 & 0 \\ 1/(R_{e2}C_e) & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (8)$$

for R1C1, R2C2 and R3C3, respectively, where s represents a combined effect of shading, reflectance, window area, glazing spectral parameters and indoor surface absorptivity, while Q_{rad} stands for the maximum radiator power [W] under standard operation conditions.

The following parameters were estimated based on the training data: s, C_r, R_{e1}, R_{e2}, Q_{rad}, R_i, C_i, C_e (R1C1: 4 parameters, R2C2: 6 parameters, R3C3: 8 parameters). The maximum ventilation rate V_{max} was assumed based on the building documentation.

The models were implemented in Modelica. The parameters were estimated using an open-source tool *ModestPy* for system identification in Functional Mock-Up Interface compliant models (Arendt 2018). A double-stage estimation technique was used. In the first stage a genetic algorithm (GA) is used to find a promising set of parameters. The implemented GA is based on standard operations and strategies like elitism, tournament selection, uniform crossover, and adaptive mutation. In the second stage, a pattern search algorithm is adopted to find a local minimum based on the final estimates from GA.

Black-box

Two black-box models are considered: a nonlinear autoregressive exogenous model (NARX) and a feed-forward neural network (NN). Both types of models are often used in building energy-related predictions (Macas et al. 2016).

The NARX model is composed of two main parts: model regressors and a nonlinearity estimator. The nonlinearity estimator consists of linear and nonlinear functions that work on the model regressors to form the model output. The model selected for this study is based the binary tree nonlinear function. The orders and delays of the model were tuned manually to minimize the training error.

The feed-forward neural network structure was selected based on preliminary tests. The criterion was to minimize the training error. After testing networks with different number of layers, neurons and activation functions, the authors selected a fully connected feed-forward network with two hidden layers with 150 and 30 neurons in the 1st and 2nd layer, respectively. The rectifier, $f(x) = \max(0, x)$, was chosen as the activation function in all neurons except the output neuron which has a linear activation function $f(x) = x$. The network was trained using the stochastic gradient descent algorithm.

As opposed to the NARX model, the chosen NN model does not take into account the dynamic effects (the past temperatures do not affect the result), as it only maps inputs to outputs based on the training data.

The NARX model was implemented in Matlab, while the NN model was implemented in Python using the Keras (Chollet et al. 2015) and Tensorflow libraries (Abadi et al. 2015).

EXPERIMENT

The models are compared in terms of indoor temperature accuracy, which is relevant in MPC applications taking thermal comfort into account. The tests were carried out based on the available data from the period of March 21 00:00 to April 21 00:00. In case of the gray- and black-box models the data is split into training and validation periods. The white-box model was calibrated earlier based on data from September–December 2016. Two training scenarios for gray- and black-box models are considered: Scenario 1 – 4 training days, Scenario 2 – 20 training days. The training in Scenario 1 starts at midnight on March 21 and ends at midnight on March 25, 2017. The training in Scenario 2 also starts on March 21, but ends on April 11. In both training scenarios, two validation windows are tested: (1) long-term validation including the rest of the data up to April 11, and (2) 3-day validation window following the training period.

The white-box temperature results do not differ between the scenarios, whereas the accuracy metric does, due to different time windows used to calculate the metric. The metric used in this study is the mean absolute error (MAE), calculated as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |T_{m,i} - T_{s,i}|}{n}, \quad (9)$$

where $T_{m,i}$ is the measured temperature at time step i, $T_{s,i}$

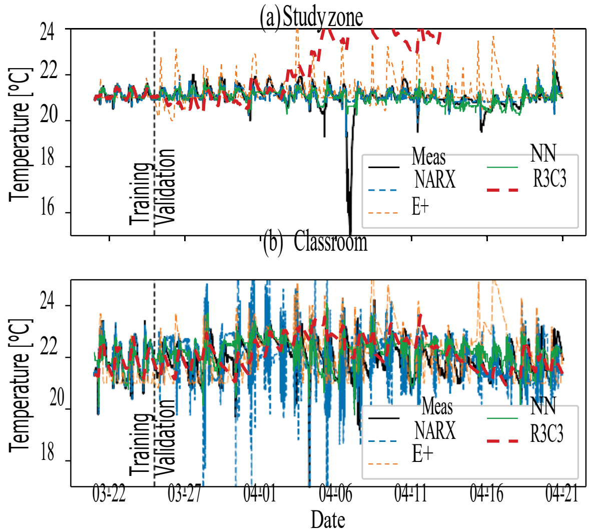

Figure 3: Temperature results in Scenario 1 (20 training days) – R1C1 and R2C2 excluded for clarity

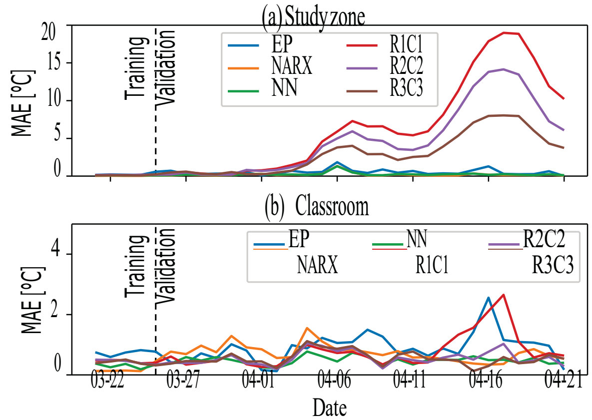

Figure 4: Daily mean absolute error (MAE) in Scenario 1 (4 training days)

Table 1: Training and validation mean absolute errors (MAE) in Scenario 1 – best performers underlined

Zone	Model	MAE [°C]		
			Validation	
		Training	27 days	3 days
Study zone	EP	0.265	0.526	0.529
	NARX	<u>0.050</u>	<u>0.212</u>	<u>0.115</u>
	NN	0.095	0.247	0.129
	R1C1	0.159	6.560	0.310
	R2C2	0.163	4.791	0.305
	R3C3	0.116	3.002	0.438
Classroom	EP	0.736	0.900	0.525
	NARX	<u>0.198</u>	0.726	0.637
	NN	0.307	<u>0.479</u>	0.439
	R1C1	0.451	0.776	0.452
	R2C2	0.438	0.553	<u>0.382</u>
	R3C3	0.406	0.546	0.383

is the simulated temperature at time step i, and n is the number of time steps.

The selected periods overlap with the Easter break (April 10–17), during which no classes were conducted (Fig. 2b). However, the study zone was occupied for few days in that period (Fig. 2a).

The results for 3-day validation periods are more relevant for dynamic applications like MPC, while the longer validation periods (11-27 days) are more relevant for FDD, benchmarking, and overall building performance evaluation (Verhelst et al. 2017). In all cases the time propagation of errors is presented and both short- and long-term accuracy is discussed.

RESULTS AND DISCUSSION

In Scenario 1 (4 training days), the lowest MAE in the 3-days validation was achieved by the NARX model in the study zone case (0.115°C) and by the R2C2 model in the classroom case (0.382°C), as shown in Table 1. In the 27-days validation the NARX model was also best for the study zone (0.212°C), while the NN model achieved the lowest error for the classroom (0.479°C).

In the case of the study zone, the gray-box models are reasonably accurate in the first few days of the validation, but diverge significantly at the beginning of the Easter break (Figs. 3-4). The explanation for this behavior is likely an open window in the room. Although there is no available data on window opening, it can be inferred from the available measurements that a window in the study zone was left open for at least one night. The study zone temperature dropped to 15°C at night on April 6/7 (Fig. 3a), much below the heating setpoint of 21°C. The outdoor temperature at that time was 11.2°C. On April 7 the temperature rose back to around 21°C, but due to the accumulated error the models were unable to close the temperature gap. It is noteworthy that none of the models could fit (in training) or predict (in validation) the temperature pattern for that night.

Since the EnergyPlus model was trained on the data from outside the considered period, its results do not depend on the number of training days in any scenario. The model accuracy depends solely on the time window used to calculate MAE. In Scenario 1 the EnergyPlus model achieved MAE of around 0.526-0.529 for the study zone in both 3-days and 27-days validation periods. A similar error was obtained in the 3-days validation in the class-

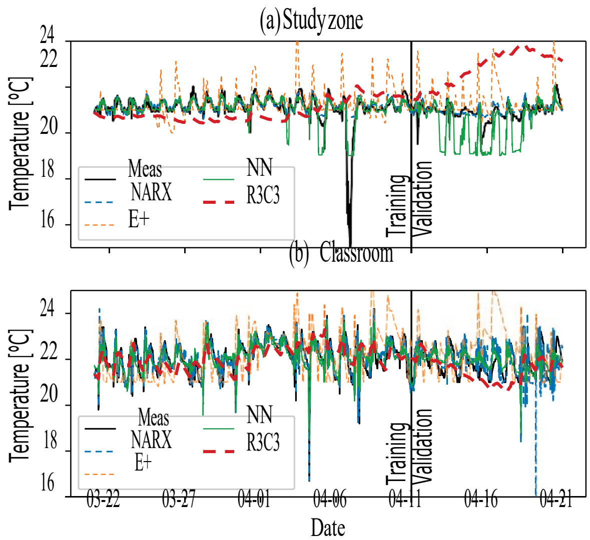

Figure 5: Temperature results in Scenario 2 (20 training days) – R1C1 and R2C2 excluded for clarity

room (0.525), but error increases in the 27-days validation. This result is partially due to the low variability in the indoor temperature (Fig. 3). The measured temperature standard deviation is 0.60°C in the study zone and 0.68°C in the classroom.

The EnergyPlus model does not experience the error accumulation as observed in the gray-box models (Fig. 4), because it is a whole-building model with a hard-coded control logic. The accuracy of this approach is, however, dependent on the accuracy of the modeled control logic. Some of the control rules from the OU44 building are not present in the model, e.g. the emergency roll up of the shading curtains in the presence of heavy winds. In addition, some of the setpoints in the building might be changed over time. To ensure the model accuracy, all of these changes must be reflected in the model. However, since the white-box model is typically calibrated manually, such changes are costly as compared to gray- and black-box models.

In Scenario 2 (20 validation days) the best performance was achieved again by the black-box models. The NARX model significantly outperformed other models in the study zone case, with MAE around 0.196 and 0.204 for the 11-days and 3-days validations, respectively (Table 2). The EnergyPlus model was the second best in both validation periods MAE around 0.476 and 0.447, respectively. The NN model performance was severely affected by the inclusion of the night with the open window in the trainig data. The model repeated the remembered pattern in the validation (Fig. 5a), exposing the main weakness of

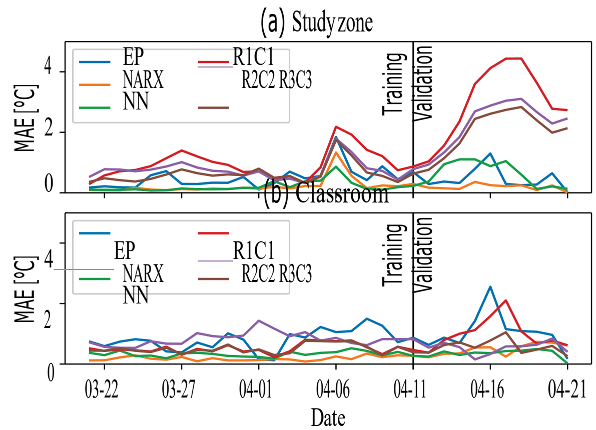

Figure 6: Daily mean absolute error (MAE) in Scenario 2 (20 training days)

feed-forward neural networks, i.e. the lack of the notion of dynamics. Even though the NARX model was trained on the same period, it did not repeat the anomaly in the validation.

Table 2: Training and validation mean absolute errors (MAE) in Scenario 2 – best performers underlined

Zone	Model	MAE [°C]		
			Validation	
		Training	11 days	3 days
Study zone	EP	0.497	0.476	0.447
	NARX	0.227	0.196	0.204
	NN	0.209	0.614	0.514
	R1C1	0.944	2.870	1.157
	R2C2	0.756	1.939	1.015
	R3C3	0.627	1.046	0.846
Classroom	EP	0.794	1.046	0.793
	NARX	0.185	0.472	0.280
	NN	0.329	0.347	0.313
	R1C1	0.516	0.964	0.546
	R2C2	0.830	0.561	0.690
	R3C3	0.503	0.564	0.467

The gray-box models performed comparatively on the classroom validation data to the white-box model, but diverged significantly in the study zone (Fig. 6). The divergence of the gray-box temperature trajectory highlights the need for a careful selection of the training period for these models. Extending the training period does not necessarily improve their performance. Typically, when an RC model cannot explain the indoor temperature variation, and the indoor temperature variations are moderate, the parameter estimation algorithm overestimates the capacitance of the system. As a result, the temperature trajectory predicted by the model in the training period is flattened (Fig. 5a). The R3C3 model is able to explain the temperature variations in the classroom training period, maintaining also a much better accuracy in the validation period (Fig. 5b). This suggests that the mode of heat transfer changed in the study zone between March 21 and April 11, and stands in agreement with findings from Scenario 1.

In overall, the black-box models outperformed the gray- and white-box models in all except one of the considered validation periods (NARX, Scenario 1). Since the black-box models are purely statistical in nature, they generally require longer training periods than gray-box. The training period of just 4 days in Scenario 1 negatively affected the accuracy of NN and NARX. The opposite effect can be observed in the considered gray-box models – longer training periods led to worse accuracy. These results are, however, not generalizable to higher-order gray-box models, which can model a larger range of dynamic effects in the building. Since each building is different, the the results from this study may not be generalizable to other buildings. However, other studies presented in the literature review also reported a high accuracy of black-box models in various building applications as compared to gray-box. The black-box models can implicitly take into account complex dynamics, including events. The range of dynamic effects covered depends only on the training data.

Finally, in real applications the gray-box models may be preferred over the black-box models for short-term predictions due to the smoother solutions and differentiability.

CONCLUSION

This paper compared the performance between selected white-, gray- and black-box models, namely: the EnergyPlus model, the RC thermal networks (R1C1, R2C2, R3C3), the nonlinear autoregressive exogenous model (NARX) and the neural network (NN). All models except the EnergyPlus model were trained on two periods: (1) 4 days starting from March 21, (2) 20 days starting from March 21. The EnergyPlus model was manually calibrated by an expert using high-level measured data (electric and heating energy consumption per building, per floor and per system) from few months preceding the month analyzed in this study. In each training scenario, the models were validated on two periods: (1) on the rest of the available data (i.e. until April 11), (2) on three days following the end of the training period. Each model was tested on two test rooms of the OU44 building, giving in total 8 validation cases to compare the performance (2 rooms × 2 training windows × 2 validation windows).

The main findings of this study are as follows:

- The black-box models outperformed the gray- and white-box models in 7 out of 8 considered validation cases. On average, their mean absolute errors were 0.355°C and 0.385°C for NARX and NN, respectively, as compared to 1.023°C and 0.655°C for R3C3 and EnergyPlus, respectively.
- Low-order RC models are not reliable for long-term prediction due to the error accumulation over time.
- The R3C3 model performed better than R1C1 and R2C2 in 6 out of 8 validation cases.
- In 5 out of 8 validation cases, the best performing model was not the one with the lowest error in the training period.
- Longer training periods do not necessarily improve the performance of gray-box models. If the RC model is unable to explain temperature variations in the training period, the parameter estimation algorithm overestimates the thermal mass.
- The white-box model performance was worse than that of the gray- and black-box, but the model was calibrated only based on the whole-building level data from few months prior to the period considered in this study. Contrary to the gray- and black-box models, the white-box model can be used to monitor the overall building energy performance on various levels.

ACKNOWLEDGMENT

This work was supported by the Innovation Fund Denmark for the project COORDICY (4106-00003B).

This work emerged from the IBPSA Project 1, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA). Project 1 will develop and demonstrate a BIM/GIS and Modelica Framework for building and community energy system design and operation.

REFERENCES

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, and Geoffrey Irving. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org.

Afram, Abdul, and Farrokh Janabi-Sharifi. 2015a. "Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods." *Energy and Buildings* 94:121 – 149.

Afram, Abdul, and Farrokh Janabi-Sharifi. 2015b. "Gray-box modeling and validation of residential HVAC system for control system design." *Applied Energy* 137:134 – 150.

Afram, Abdul, and Farrokh Janabi-Sharifi. 2017. "Supervisory model predictive controller (MPC) for residential HVAC systems: Implementation and experimentation on archetype sustainable house in Toronto." *Energy and Buildings* 154:268 – 282.

Ahmad, Tanveer, and Huanxin Chen. 2018. "Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches." *Energy and Buildings* 166:460 – 476.

Arendt, Krzysztof. 2018. ModestPy: FMI-compliant Model Estimation in Python. https://github. com/sdu-cfei/modest-py.

Arendt, Krzysztof, Ana Ionesi, Muhyiddine Jradi, Ashok Singh, Mikkel B. Kjægaard, Christian T. Veje, and Bo N. Jørgensen. 2016. "A Building Model Framework for a Genetic Algorithm Multi-objective Model Predictive Control." *CLIMA 2016 – proceedings of the 12th REHVA World Congress*, vol. 8.

Brastein, O.M., D.W.U. Perera, C. Pfeifer, and N.-O. Skeie. 2018. "Parameter estimation for grey-box models of building thermal behaviour." *Energy and Buildings* 169:58 – 68.

Chollet, François, et al. 2015. Keras. https://github. com/keras-team/keras.

Coninck, Roel De, and Lieve Helsen. 2016. "Practical implementation and evaluation of model predictive control for an office building in Brussels." *Energy and Buildings* 111:290 – 298.

Gunay, H. Burak, William O'Brien, and Ian Beausoleil-Morrison. 2016. "Control-oriented inverse modeling of the thermal characteristics in an office." *Science and Technology for the Built Environment* 22 (5): 586–605.

Henze, Gregor P. 2013. "Model predictive control for buildings: a quantum leap?" *Journal of Building Performance Simulation* 6 (3): 157–158.

Jradi, M., C.T. Veje, and B.N. Jørgensen. 2018. "A dynamic energy performance-driven approach for assessment of buildings energy Renovation—Danish case studies." *Energy and Buildings* 158:62 – 76.

Jradi, Muhyiddine, Fisayo Caleb Sangogboye, Claudio Giovanni Mattera, Mikkel Baun Kjærgaard, Christian Veje, and Bo Nørregaard Jørgensen. 2017. "A World Class Energy Efficient University Building by Danish 2020 Standards." *Energy Procedia* 132:21 – 26. 11th Nordic Symposium on Building Physics, NSB2017, 11-14 June 2017, Trondheim, Norway.

Macas, Martin, Fabio Moretti, Alessandro Fonti, Andrea Giantomassi, Gabriele Comodi, Mauro Annunziato, Stefano Pizzuti, and Alfredo Capra. 2016. "The role of data sample size and dimensionality in neural network based forecasting of building heating related variables." *Energy and Buildings* 111:299 – 310.

Prívara, Samuel, Jiří Cigler, Zdeněk Váňa, Frauke Oldewurtel, Carina Sagerschnig, and Eva Žáčeková. 2013. "Building modeling as a crucial part for building predictive control." *Energy and Buildings* 56:8 – 22.

Reynolds, Jonathan, Yacine Rezgui, Alan Kwan, and Solène Piriou. 2018. "A zone-level, building energy optimisation combining an artificial neural network, a genetic algorithm, and model predictive control." *Energy* 151:729 – 739.

Sangoboye, Fisayo Caleb, and Mikkel Baun Kjærgaard. 2016. "PLCount: A Probabilistic Fusion Algorithm for Accurately Estimating Occupancy from 3D Camera Counts." *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, BuildSys '16. New York, NY, USA: ACM, 147–156.

Sangogboye, Fisayo Caleb, Krzysztof Arendt, Ashok Singh, Christian T. Veje, Mikkel Baun Kjærgaard, and Bo Nørregaard Jørgensen. 2017. "Performance comparison of occupancy count estimation and prediction with common versus dedicated sensors for building model predictive control." *Building Simulation* 10 (6): 829–843 (Dec).

Turner, W.J.N., A. Staino, and B. Basu. 2017. "Residential HVAC fault detection using a system identification approach." *Energy and Buildings* 151:1 – 17.

Verhelst, J., G. Van Ham, D. Saelens, and L. Helsen. 2017. "Model selection for continuous commissioning of HVAC-systems in office buildings: A review." *Renewable and Sustainable Energy Reviews* 76:673 – 686.

Zhang, Rongpeng, and Tianzhen Hong. 2017. "Modeling of HVAC operational faults in building performance simulation." *Applied Energy* 202:178 – 188.

Žáčeková, Eva, Zdeněk Váňa, and Jiří Cigler. 2014. "Towards the real-life implementation of MPC for an office building: Identification issues." *Applied Energy* 135:53 – 62.