Simplifying the Auto Regressive and Moving Average (ARMA) Model Representing the Dynamic Thermal Behaviour of iHouse Based on Theoretical Knowledge

Shamsul Faisal Mohd Hussein[1(✉)], Mohd Anuar Abu Bakar[1], Yoshiki Makino[3], Hoaison Nguyen[2], Shahrum Shah Abdullah[1], Yuto Lim[3], and Yasuo Tan[3]

[1] Biologically Inspired System and Technology (Bio-iST) iKohza, Department of Electronic Systems Engineering (ESE), Malaysia – Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak), 54100 Kuala Lumpur, Malaysia
foxfaisal@yahoo.co.uk, mohdanuar.abubakar@gmail.com, shahrum@fke.utm.my

[2] Faculty of Information Technology, Department of Networks and Computer Communications, VNU University of Engineering and Technology, Building E3, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam
hsnguyen@jaist.ac.jp

[3] School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Nomi, Ishikawa Prefecture 923-1211, Japan
{m-yoshi,ylim,ytan}@jaist.ac.jp

Abstract. Modelling and simulation is an alternative way of testing the dynamic behaviour of a real system – in some situation, testing the real system are expensive, time consuming, not comfortable, and dangerous. Mathematical model describing the dynamic behaviour of a system can be represented by using white, black, or grey box model. This study focuses on developing a simplified Auto Regressive Moving Average (ARMA) model (a type of linear black model) to represent the dynamic thermal behaviour of iHouse – simplification is done based on the theoretical knowledge of the building. The performance of the simplified ARMA model developed in this study is compared with the performance of the models developed in previous studies, which are: (1) House Thermal Simulator; (2) and ARMA model. Result shows that the simplified ARMA model developed in this study consists of simpler set of mathematical equations, but can still simulate the dynamic thermal behaviour of iHouse with the accuracy that is almost on par with the models developed in previous studies.

Keywords: Modelling and simulation · Black box modelling · Building temperature simulation · Building temperature prediction

M.S. Mohamed Ali et al. (Eds.): AsiaSim 2017, Part I, CCIS 751, pp. 697–711, 2017.
DOI: 10.1007/978-981-10-6463-0_59

1 Introduction

Mathematical model is used to represent the dynamic behaviour of a system for simulation purpose. Mathematical model can be categorised as white, black and grey box model.

White box model is also known as theoretical model [1] and is developed based on the fundamental knowledge in science and engineering [2]. The value of unknown parameters in the white box model representing the dynamic behaviour of a system can be obtained through measurement in the actual system, referring to datasheet/manual of the system etc. [2]. Complex models can be simplified by making assumptions [2]. The advantage of white box model is it gives insight to the user on how the system behaves according to the fundamental law of science and engineering [1]. Since the model is constructed based on fundamental law of science and engineering, the model can be simulated over a wide range of operating point [1]. However, the white box model also has some weaknesses – in some situations, the value of unknown parameters contained in the system that is going to be modelled is difficult or even impossible to be measured or obtained [1]. In addition, assumptions made to simplify complex model may cause the model to be inaccurate [2].

Black box model is also known as empirical model [1] or data driven model [3], is developed by tuning the parameters in a set of linear or non-linear equations to map the relationship between the input(s) and output(s) of a system. Example of linear black box models are Auto Regressive (AR) model, Moving Average (MA) model, Auto Regressive Moving Average (ARMA) model etc. Example of non-linear black box model is Artificial Neural Network (ANN). Black box model is a purely data driven model – it maps the relationship between input(s) and output(s) of a system without describing the physical theory behind it. One of the advantages of the black box model is that it is suitable to be implemented when only the recorded input(s) and output(s) produced by the system that is going to be model are available, but the theoretical knowledge describing the system is unknown [1, 3]. Shamsul et al. [4] developed a black box model to simulate the air temperature of one of the rooms in iHouse, a smart house testbed (shown in Fig. 1) belongs to Japan Advanced Institute of Science and Technology (JAIST) with minimal physical knowledge of iHouse. Mustafaraj et al. in [5] predicted the room temperature and relative humidity of visa building in London using the following models: (1) autoregressive model with external inputs (ARX); (2) and neural network- based nonlinear auto regressive model with external inputs (NNARX) – results shows that both ARX and NNARX performed reasonably good, but the NNARX outperformed ARM. Mustafaraj et al. in [6] investigated the performance of the following models to predict the thermal behaviour of an open-plan office (room) in Portman House, located in central London: (1) a neural network-based non-linear autoregressive model with external inputs (NNARX); (2) a non-linear autoregressive moving average model with external inputs (NNARMAX); (3) and a non-linear output error model (NNOE) to predict the thermal behaviour of an open-plan office (room) in Portman House, central London – results showed that all models performed reasonably good, but the NNARX and NNARMAX models outperformed the NNOE model. Hazyuk et al. in [7] uses physical knowledge to decide the structure of the model, then uses black box modelling to represent the dynamic behaviour of the indoor temperature of a typical detached house in France, which is used as one of the reference building by *Centre Scientifique et Technique du Bâtiment* (CSTB)/Building Scientific and Technical Centre for performance evaluation. Even though black box model can be implemented with minimal theoretical knowledge of the system, knowing more theoretical knowledge will be more advantageous during the model development [3]. Like other models, the black box model also has disadvantage – it cannot be simulated within the operating condition that is outside the range of the data that is used to train/regress the black box model [1, 3] and it doesn't give physical insight regarding the theoretical knowledge of the system.

Fig. 1. The photo of iHouse, the smart house testbed belongs to Japan Advanced Institute of Science and Technology (JAIST).

Grey box model is also known as semi-empirical model [1] or hybrid model [3], is the combination of both white box model and black box model – the set of mathematical equations describing the dynamic behaviour of the model is constructed based on the knowledge of science and engineering while the unknown parameters in these equations are estimated based on the input(s)-output(s) relationship data produced by the system. The mathematical equations constructed based on fundamental knowledge of science and engineering gives physical insight to the system like the white box model, but the value of the unknown parameters in the equations that are difficult or impossible to be obtained can be estimated based on the input(s)-output(s) relationship data produced by the system that is going to be modelled [1]. Unlike black box model, the grey box model can be simulated (with caution) within the operating range that is outside the range of the data that is used to train/regress the grey box model [1]. Nguyen et al. [8] built a SIMULINK® toolbox named House Thermal Simulator to simulate the dynamic indoor temperature of iHouse, JAIST (shown in Fig. 1) based on a set of large quantity of equations that describe how the controlled inputs and disturbances affect the indoor air temperature of iHouse – the unknown parameters in these equations were estimated based on the real recorded data in iHouse using Simulink Design Optimization toolbox in MATLAB®.

A mathematical model is just a set of mathematical equations that represents the dynamic behaviour of a system [1]. Depending of the application of the model, care must be taken before and during the development process to maintain the balance between accuracy and complexity of the model [1]. Too accurate models will increase its complexity and consume lots of resources (in terms of time, budget, etc.) during the model development while too simple model will decrease its accuracy [1].

This study focuses on simplifying the ARMA model developed in [4] to represent the dynamic indoor air temperature behaviour of iHouse belongs to JAIST (shown in Fig. 1) based on theoretical knowledge – simpler model leads to simpler implementation and faster simulation. There are 2 previous studies that modelled the dynamic thermal behaviour of iHouse: (1) the first study developed a grey box model called House Thermal Simulator [8], which was developed using a set of many detailed mathematical equations describing the dynamic thermal behaviour of iHouse; (2) and the second study developed a black box model which was based on a purely data driven auto regressive and moving average (ARMA) model with minimal theoretical knowledge regarding the heat transfer properties of iHouse [4]. This study simplify the model developed in [4] based on some of the theoretical knowledge learnt in [8] and other sources.

2 Methodology

Since this study is related to the previous work in [4], the scope of this study is also similar with the previous work. First, only one out of fifteen rooms available in iHouse is modelled in this study. The modelled room is Bedroom A, which is the same room modelled in previous work in [4] and is shown in the iHouse floor plan shown in Fig. 2.

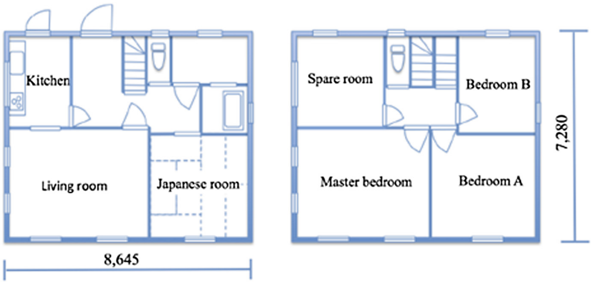

Fig. 2. The floorplan of iHouse.

Second, only the weather-related inputs (disturbances) without control-related input (s) are considered for this study. Even though thermal comfort devices such as air conditioner and motor-operated windows are installed in iHouse, the available historical data that are recorded when these thermal comfort devices are operated by the time this study is done are still not sufficient to develop model with control-related inputs using black box and grey box modelling. Therefore, it is still unable to develop a grey box or black box model that is capable to simulate control-related input by the time this study is done.

Third, all the walls of the room that is going to be modelled is assumed to be consisted of plain wall without doors or windows – these door and windows are assumed to be part of the wall to maintain the simplicity of the model and reduce time taken to develop it. Including the equations describing how the weather-related inputs affect the room's air temperature through the door and windows will increase the total number or length of the mathematical equations in the model and will increase the development duration.

2.1 Data Collection

Due to time constrain, available historical data that were used in previous study in [4] are used again in this study. These selected data were recorded with the following conditions: (1) the air conditioner was switched off; (2) and the motor-operated windows were closed.

Two groups of historical data that were recorded with same conditions (as mentioned previously) and were recorded on the date that are as close with each other as possible are identified and assigned as training and testing data set. The reason why the training and testing data set must be recorded on the date that are as close as possible is to reduce the variation of weather related inputs, especially in four-season countries.

The first group of historical data were recorded from the 1st of August 2012 until the 3rd of August 2012 (assigned as train data set) while the second group of historical data were recorded from the 10th of August 2012 until the 19th of August 2012 (assigned as test data set).

Different types of sensors in iHouse record different data at different time intervals. To simplify the model development process, the interval for all recorded data is standardized at every 90 s – this means that there will be 960 recorded input-output relationship data for every day if the data is recorded at the interval of every 90 s. Training data set (recorded from the 1st of August 2012 until the 3rd of August 2012) has 2880 recorded inputs-outputs relationships while testing data-set (recorded from the 10th of August 2012 until the 19th of August 2012) has 9600 recorded inputs-outputs relationships.

2.2 Data Selection

In previous work in [4], the ARMA model representing the dynamic indoor thermal behaviour of iHouse was created based on minimal (almost zero) theoretical knowledge regarding the physical thermal characteristic of iHouse. By the time this study is done, more theoretical knowledge regarding the thermal characteristic of iHouse (and other buildings) has been learnt.

Like the ARMA model developed in previous study in [4], the simplified ARMA model developed in this study is a multi-input single-output (MISO) system. However, the number of inputs has been reduced after having more information regarding the theoretical knowledge of iHouse (and other buildings). The output assigned for this model is the future temperature of bedroom A, (T_{BedA}). Unlike in the previous work in [4], the inputs assigned for this model has been revised based on the physical insight regarding how the weather related inputs affect the indoor air temperature of iHouse, which are: (1) the past temperature of bedroom A itself (T_{BedA}); (2) the differences between the past temperatures of the spaces surrounding bedroom A and the temperature of bedroom A itself – based on Fig. 2, bedroom A is surrounded by bedroom B and staircase from the North ($\Delta T_{BedB-BedA}$ and $\Delta T_{Stair-BedA}$), outdoor from the East and South ($\Delta T_{Out-BedA}$), master bedroom from the West ($\Delta T_{MBed-BedA}$), roof attic from the top ($\Delta T_{Attic-BedA}$) and Japanese-style room from the bottom ($\Delta T_{JRoom-BedA}$); (3) the past heat emitted from residence(s) and electronic device(s) in bedroom A ($Q_{ResDevBedA}$); (4) the past heat radiation from outside air and ground onto the outer walls of bedroom A ($Q_{OutAirRad}$); (5) the past solar radiations – only 2 types of solar radiations data are considered in this study, which are direct solar radiation on eastern and southern outer wall surface of bedroom A ($\varphi_{DirEastWall}$ and $\varphi_{DirSouthWall}$) and diffuse solar radiation on both eastern and southern outer wall surface of bedroom A ($\varphi_{DiffEastWall}$ and $\varphi_{DiffSouthWall}$); (6) and the past heat gain due to convection between outside air and both eastern and southern outer wall surface of bedroom A ($Q_{ConvEastWall}$ and $\varphi_{ConvSouthWall}$). This model with the determined inputs and output is illustrated in Fig. 3.

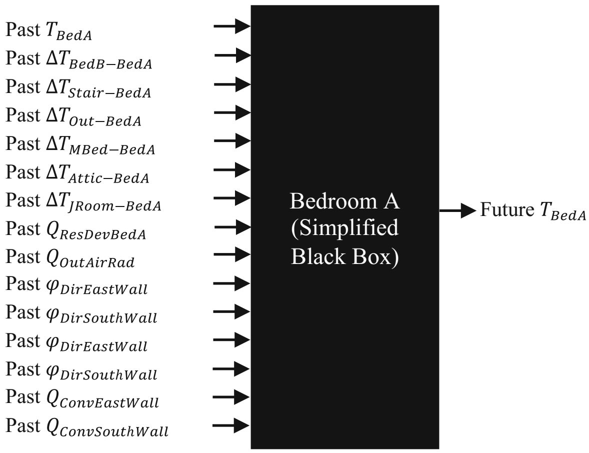

Fig. 3. The illustration of model for the thermal behaviour of bedroom A with listed possible inputs.

2.3 Model Construction

Let's say that the ARMA model has k past input(s), which is from $i = 0$ until $i = k - 1$. After the inputs that may have the potential to influence the temperature in bedroom A are determined and listed in Sect. 2.3 and depicted in Fig. 3, the simplified ARMA model equation with k past input(s) describing the air temperature in bedroom A is written, which is shown below:

$$
\begin{aligned}
T_{BedA}[n+k] = & \sum\nolimits_{i=0}^{k-1} A_i T_{BedA}[n+i] + \sum\nolimits_{i=0}^{k-1} B_i \Delta T_{BedB-BedA}[n+i] \\
& + \sum\nolimits_{i=0}^{k-1} C_i \Delta T_{Stair-BedA}[n+i] + \sum\nolimits_{i=0}^{k-1} D_i \Delta T_{Out-BedA}[n+i] \\
& + \sum\nolimits_{i=0}^{k-1} E_i \Delta T_{MBed-BedA}[n+i] + \sum\nolimits_{i=0}^{k-1} F_i \Delta T_{Attic-BedA}[n+i] \\
& + \sum\nolimits_{i=0}^{k-1} G_i \Delta T_{JRoom-BedA}[n+i] + \sum\nolimits_{i=0}^{k-1} H_i Q_{ResDevBedA}[n+i] \\
& + \sum\nolimits_{i=0}^{k-1} I_i Q_{OutAirRad}[n+i] + \sum\nolimits_{i=0}^{k-1} J_i \varphi_{DirEastWall}[n+i] \\
& + \sum\nolimits_{i=0}^{k-1} K_i \varphi_{DirSouthWall}[n+i] + \sum\nolimits_{i=0}^{k-1} L_i \varphi_{DiffEastWall}[n+i] \\
& + \sum\nolimits_{i=0}^{k-1} M_i \varphi_{DiffSouthWall}[n+i] + \sum\nolimits_{i=0}^{k-1} N_i Q_{ConvEastWall}[n+i] \\
& + \sum\nolimits_{i=0}^{k-1} O_i Q_{ConvSouthWall}[n+i].
\end{aligned}
\quad (1)
$$

Equation (1) can also be written in an expanded form as shown below:

$$
\begin{aligned}
T_{BedA}[n+k] = & A_{k-1} T_{BedA}[n+k-1] + A_{k-2} T_{BedA}[n+k-2] + \cdots \\
& + A_1 T_{BedA}[n+1] + A_0 T_{BedA}[n] + B_{k-1} \Delta T_{BedB-BedA}[n+k-1] \\
& + B_{k-2} \Delta T_{BedB-BedA}[n+k-2] + \cdots + B_1 \Delta T_{BedB-BedA}[n+1] \\
& + B_0 \Delta T_{BedB-BedA}[n] + C_{k-1} \Delta T_{Stair-BedA}[n+k-1] + C_{k-2} \Delta T_{Stair-BedA}[n+k-2] + \cdots \\
& + C_1 \Delta T_{Stair-BedA}[n+1] + C_0 \Delta T_{Stair-BedA}[n] \\
& + D_{k-1} \Delta T_{Out-BedA}[n+k-1] + D_i \Delta T_{Out-BedA}[n+k-2] + \cdots \\
& + D_1 \Delta T_{Out-BedA}[n+1] + D_0 \Delta T_{Out-BedA}[n] + E_{k-1} \Delta T_{MBed-BedA}[n+k-1] \\
& + E_{k-2} \Delta T_{MBed-BedA}[n+k-2] + \cdots + E_1 \Delta T_{MBed-BedA}[n+1] \\
& + E_0 \Delta T_{MBed-BedA}[n] + F_{k-1} \Delta T_{Attic-BedA}[n+k-1] + F_{k-2} \Delta T_{Attic-BedA}[n+k-2] + \cdots \\
& + F_1 \Delta T_{Attic-BedA}[n+1] + F_0 \Delta T_{Attic-BedA}[n] \\
& + G_{k-1} \Delta T_{JRoom-BedA}[n+k-1] + G_{k-2} \Delta T_{JRoom-BedA}[n+k-2] + \cdots \\
& + G_1 \Delta T_{JRoom-BedA}[n+1] + G_0 \Delta T_{JRoom-BedA}[n] + H_{k-1} Q_{ResDevBedA}[n+k-1] \\
& + H_{k-2} Q_{ResDevBedA}[n+k-2] + \cdots + H_1 Q_{ResDevBedA}[n+1] \\
& + H_0 Q_{ResDevBedA}[n] + I_{k-1} Q_{OutAirRad}[n+k-1] + I_{k-2} Q_{OutAirRad}[n+k-2] + \cdots \\
& + I_1 Q_{OutAirRad}[n+1] + I_0 Q_{OutAirRad}[n] + J_{k-1} \varphi_{DirEastWall}[n+k-1] \\
& + J_{k-2} \varphi_{DirEastWall}[n+k-2] + \cdots + J_1 \varphi_{DirEastWall}[n+1] \\
& + J_0 \varphi_{DirEastWall}[n] + K_{k-1} \varphi_{DirSouthWall}[n+k-1] + K_{k-2} \varphi_{DirSouthWall}[n+k-2] + \cdots \\
& + K_1 \varphi_{DirSouthWall}[n+1] + K_0 \varphi_{DirSouthWall}[n] \\
& + L_{k-1} \varphi_{DiffEastWall}[n+k-1] + L_{k-2} \varphi_{DiffEastWall}[n+k-2] + \cdots \\
& + L_1 \varphi_{DiffEastWall}[n+1] + L_0 \varphi_{DiffEastWall}[n] + M_{k-1} \varphi_{DiffSouthWall}[n+k-1] \\
& + M_{k-2} \varphi_{DiffSouthWall}[n+k-2] + \cdots + M_1 \varphi_{DiffSouthWall}[n+1] \\
& + M_0 \varphi_{DiffSouthWall}[n] + N_{k-1} Q_{ConvEastWall}[n+k-1] \\
& + N_{k-2} Q_{ConvEastWall}[n+k-2] + \cdots + N_1 Q_{ConvEastWall}[n+1] \\
& + N_0 Q_{ConvEastWall}[n] + O_{k-1} Q_{ConvSouthWall}[n+k-1] \\
& + O_{k-2} Q_{ConvSouthWall}[n+k-2] + \cdots + O_1 Q_{ConvSouthWall}[n+1] \\
& + O_0 Q_{ConvSouthWall}[n].
\end{aligned}
\quad (2)
$$

2.4 Model Regression

'Regression' is defined as the process of estimating the unknown parameter(s) in ARMA model. In Artificial Neural Network, this process is known as 'training'.

Let's say that the training data-set is recorded from $t = 0$ until $t = p$. Therefore, the number of available sampled data recorded from $t = 0$ until $t = p$ are $p + 1$. The value of $p + 1$ in this study is equal to the number of data recorded from the 1st of August 2012 until the 3rd of August 2012, which is 2880 (as mentioned earlier in Subsect. 2.2 – Data Collection). When the data from k previous steps are used to estimate the temperature of bedroom A (from $t = 0$ until $t = k - 1$), the input-output pairs are equal to $p + 1 - k$. This can be illustrated by the matrices in Eq. (3) below:

$$\mathbb{Y} = \mathbb{X}\mathbb{M}. \tag{3}$$

where:

$$\mathbb{Y} = \begin{bmatrix} T_{BedA}[n+k] \\ T_{BedA}[n+k+1] \\ \vdots \\ T_{BedA}[n+p-1] \\ T_{BedA}[n+p] \end{bmatrix},$$

$$\mathbb{X} = \begin{bmatrix} \mathbb{X}_{(1,:)} \\ \mathbb{X}_{(2,:)} \\ \vdots \\ \mathbb{X}_{(p-k,:)} \\ \mathbb{X}_{(p+1-k,:)} \end{bmatrix},$$

$$\mathbb{X}_{(1,:)} = \begin{bmatrix} T_{BedA}[n+k-1] \\ T_{BedA}[n+k-2] \\ \vdots \\ T_{BedA}[n+1] \\ T_{BedA}[n] \\ \Delta T_{BedB-BedA}[n+k-1] \\ \Delta T_{BedB-BedA}[n+k-2] \\ \vdots \\ \Delta T_{BedB-BedA}[n+1] \\ \Delta T_{BedB-BedA}[n] \\ \vdots \\ Q_{ConvSouthWall}[n+k-1] \\ Q_{ConvSouthWall}[n+k-2] \\ \vdots \\ Q_{ConvSouthWall}[n+1] \\ Q_{ConvSouthWall}[n] \end{bmatrix}^T,$$

$$\mathbb{X}_{(2,:)} = \begin{bmatrix} T_{BedA}[n+k] \\ T_{BedA}[n+k-1] \\ \vdots \\ T_{BedA}[n+2] \\ T_{BedA}[n+1] \\ \Delta T_{BedB-BedA}[n+k] \\ \Delta T_{BedB-BedA}[n+k-1] \\ \vdots \\ \Delta T_{BedB-BedA}[n+2] \\ \Delta T_{BedB-BedA}[n+1] \\ \vdots \\ Q_{ConvSouthWall}[n+k] \\ Q_{ConvSouthWall}[n+k-1] \\ \vdots \\ Q_{ConvSouthWall}[n+2] \\ Q_{ConvSouthWall}[n+1] \end{bmatrix}^T,$$

$$\mathbb{X}_{(p-k,:)} = \begin{bmatrix} T_{BedA}[n+p-2] \\ T_{BedA}[n+p-3] \\ \vdots \\ T_{BedA}[n+p-k] \\ T_{BedA}[n+p-1-k] \\ \Delta T_{BedB-BedA}[n+p-2] \\ \Delta T_{BedB-BedA}[n+p-3] \\ \vdots \\ \Delta T_{BedB-BedA}[n+p-k] \\ \Delta T_{BedB-BedA}[n+p-1-k] \\ \vdots \\ Q_{ConvSouthWall}[n+p-2] \\ Q_{ConvSouthWall}[n+p-3] \\ \vdots \\ Q_{ConvSouthWall}[n+p-k] \\ Q_{ConvSouthWall}[n+p-1-k] \end{bmatrix}^T,$$

$$\mathbb{X}_{(p+1-k,:)} = \begin{bmatrix} T_{BedA}[n+p-1] \\ T_{BedA}[n+p-2] \\ \vdots \\ T_{BedA}[n+p+1-k] \\ T_{BedA}[n+p-k] \\ \Delta T_{BedB-BedA}[n+p-1] \\ \Delta T_{BedB-BedA}[n+p-2] \\ \vdots \\ \Delta T_{BedB-BedA}[n+p+1-k] \\ \Delta T_{BedB-BedA}[n+p-k] \\ \vdots \\ Q_{ConvSouthWall}[n+p-1] \\ Q_{ConvSouthWall}[n+p-2] \\ \vdots \\ Q_{ConvSouthWall}[n+p+1-k] \\ Q_{ConvSouthWall}[n+p-k] \end{bmatrix}^T,$$

$$\mathbb{M} = \begin{bmatrix} A_{k-1} \\ A_{k-2} \\ \vdots \\ A_1 \\ A_0 \\ B_{k-1} \\ B_{k-2} \\ \vdots \\ B_1 \\ B_0 \\ \vdots \\ O_{k-1} \\ O_{k-2} \\ \vdots \\ O_1 \\ O_0 \end{bmatrix}.$$

In this study, the least square method is used to regress the simplified ARMA model. The purpose of regression is to estimate the values of the unknown constants in matrix $\mathbb{M}$ for the ARMA model in Eq. (3). The least square equation used for regressing the model is represented by Eq. (4) is written below:

$$\mathbb{M} = \left(\mathbb{X}^T\mathbb{X}\right)^{-1}\mathbb{X}^T\mathbb{Y}. \tag{4}$$

After the values of the unknown constants in matrix $\mathbb{M}$ are estimated using least square method, the model is then simulated again using the training data set to check its ability to fit the data-set. Then, the model can be tested and optimized with new data set that has never been 'seen' by the model, which will be discussed in the next sub-section.

2.5 Model Testing

In this process, the regressed simplified ARMA model is simulated with a new data-set that has never been 'seen' by the model. This data-set is called testing data-set. The purpose of this process is to ensure that the regressed model can produce accurate result when the model is simulated using different data-set (other than the training data-set).

Let's say that the testing data-set is recorded from $t = 0$ until $t = q$. Therefore, the number of available sampled data recorded from $t = 0$ until $t = q$ are $q + 1$. The value of $q + 1$ in this study is equal to the number of data recorded from the 10^{th} of August 2012 until the 19^{th} of August 2012, which is 9600 (as mentioned earlier in Subsect. 2.2 – Data Collection).

2.6 ARMA Model Parameter Estimation

The only parameter that needs to be tuned to optimize the ARMA model is the number of past input(s), which is the value of k. Bigger value of k leads to more quantity of constants available in matrix $\mathbb{M}$ (and vice versa). A MATLAB® script is written to try the possible values of k one by one (instead of assigning the value of k randomly and manually using trial and error method). Due to time constraint however, the value of k in this research is tried one by one only from $k = 1$ until $k = 200$. The percentage of fitness, $\%Fit$ is calculated for each tested value of k and its formula is shown below:

$$\%Fit = 1 - \frac{norm\left(T_{BedA} - \hat{T}_{BedA}\right)}{norm\left[T_{BedA} - mean\left(T_{BedA}\right)\right]}. \tag{5}$$

3 Results

The performance of the optimized simplified ARMA model in this study is compared with the performance of the previous works, which are: (1) the optimised ARMA model [4]; (2) the optimized House Thermal Simulator [8]. The parameters in both ARMA model [4] and House Thermal Simulator [8] are also estimated and optimized by using their own parameter estimation and optimization methods, but using the same training and testing data-sets as those used in this study. The simulation results for the simplified ARMA model in this study, the ARMA model in previous study [4], and House Thermal Simulator [8] are plotted and displayed in Fig. 4 for comparison. Meanwhile, the value of percentage of fitness, $\%Fit$ for the simplified ARMA model in this study, the ARMA model in previous study [4], and House Thermal Simulator [8] are summarized in Table 1 for comparison.

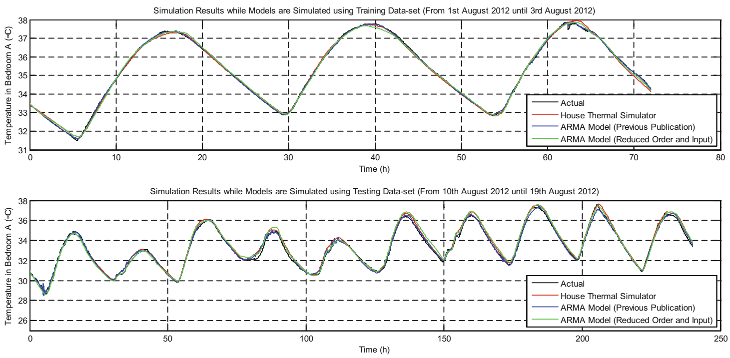

Fig. 4. The simulation result comparison for all three optimized models when simulated with training data set and testing data set.

Table 1. The %*Fit* values for all optimized models during simulation are presented in.

Model	%*Fit* (Train data set)	%*Fit* (Test data set)
House Thermal Simulator [8]	94.824	88.0188
ARMA Model [4]	97.6667	91.6101
Simplified ARMA Model	95.071	87.4575

4 Discussion

From the result displayed on Fig. 4 and Table 1, the accuracy of the simplified ARMA model obtained in this study in terms of percentage of best fit, %*Fit* is slightly less compared to the previous works in [8] and [4]. However, the differences are not big.

Even though the accuracy of the simplified ARMA model done in this study is slightly less compared to the previous works [4, 8], it's development and implementation is simpler. The ARMA model developed in previous study [4] has 19 types of inputs – when the model's number of order (which is also the number of past inputs) is tuned from $k = 1$ until $k = 200$, the model is accurate the most when $k = 26$. Meanwhile the simplified ARMA model developed in this study has 15 types of inputs – when the model's number of order is tuned in this study (from $k = 1$ until $k = 200$), the model is accurate the most when $k = 1$. Lesser types of inputs increase model's simplicity and reduces the time taken during model's regression process while model with lower order number (or lesser number of past inputs) reduces computation time during simulation. Meanwhile, the model developed in this study is also simpler to be implemented compared to House Thermal Simulator even though the accuracy of the model developed in this study is slightly lower than the accuracy of House Thermal Simulator because the simplified ARMA model developed in this study has lesser and simpler mathematical equations.

House Thermal Simulator [8] is a grey box model – it was developed based on fundamental knowledge of science and engineering while the unknown parameters in the model are estimated based on recorded inputs-output data relationship recorded in iHouse, which provide insight regarding how each input cause increment/decrement to the air temperature in iHouse. The set of equations used in House Thermal Simulator [8] are in large quantities to describe the relationship between the inputs and output vividly base on theoretical knowledge. This makes the House Thermal Simulator suitable to be used by the advanced researchers who would like to get a very detail insight on how the inputs affect the output. However, utilizing large quantity of equation is time consuming and requires more powerful computers for simulation. Meanwhile, the ordinary ARMA model that was developed [4] is a purely data driven model and was developed with minimal theoretical knowledge – ARMA model is just a single equation with parameters tuned to map the value of inputs versus output of a system. This make a black box model suitable to be used by beginner researchers because it can be constructed with minimal physical knowledge describing the system. Model with single equation (or a set of equation in small quantity) can be simulated in less powerful computers. Knowing more theoretical knowledge of a system while developing black box models will be more advantageous [3]. Compared to the ARMA model developed previously in [4], the number of inputs and model's order in this study is managed to be reduced based on the theoretical knowledge. In previous work, the only inputs that provide the insight regarding how they affect the air temperature increment/decrement in iHouse are: (1) the past temperature of bedroom A (T_{BedA}); (2) the past temperature difference between the surrounding spaces and bedroom A $(\Delta T_{BedB-BedA}, \Delta T_{Stair-BedA}, \Delta T_{Out-BedA}, \Delta T_{MBed-BedA}, \Delta T_{Attic-BedA}, \text{ and } \Delta T_{JRoom-BedA})$; (3) the heat generated by residence(s) and electrical device(s) in bedroom A $(Q_{ResDevBedA})$; (4) and heat radiation from outside air and ground onto the outer walls of bedroom A $(Q_{OutAirRad})$. Meanwhile the rest of the inputs utilised in previous study [4] did not give insight on how they affect the air temperature gain/loss in iHouse – these inputs are: (1) the relative humidity of bedroom A (RH_{BedA}); (2) the solar altitude (Z_{Solar}); (3) the solar azimuth (θ_{Solar}); (4) and the wind speed blowing from all 4 directions $(V_{NorthWind}, V_{EastWind}, V_{SouthWind}, \text{ and } V_{WestWind})$. Instead of being used directly as one of the inputs in the previous study, the relative humidity of bedroom A (RH_{BedA}) should be used to calculate the heat capacity of the air in bedroom A. However, neither the relative humidity of bedroom A (RH_{BedA}) nor the heat capacity of the air in bedroom A are used as the inputs of the simplified ARMA model in this study. Then, instead of being used directly as some of the inputs in the previous study, the value of solar position (Z_{Solar}) and solar azimuth (θ_{Solar}) are used to calculate the amount of direct and diffuse solar radiation that hit both the eastern and southern outer wall surfaces of bedroom A $(\varphi_{DirEastWall}, \varphi_{DirSouthWall}, \varphi_{DiffEastWall}, \varphi_{DiffSouthWall})$, which are used as the inputs of the simplified ARMA model in this study – these values are calculated based on solar altitude, solar azimuth wall slope angle, and wall azimuth. Finally, the wind speed (v_{Wind}) and direction (θ_{Wind}) in the previous research were resolved using trigonometric method into wind velocities that are coming from 4 directions, which are the wind velocity blowing from all 4 directions $(v_{NorthWind}, v_{EastWind}, v_{SouthWind}, \text{ and } v_{WestWind})$. However, the values of v_{Wind} and θ_{Wind} should be used to calculate heat gain due to convection between outside air and both eastern and southern outer wall surface of bedroom A $(Q_{ConvEastWall}$ and $Q_{ConvSouthWall})$, which are used as the inputs of the simplified ARMA model in this study – these values are calculated based on outdoor air temperature (T_{Out}), the wind speed (v_{Wind}), and wind direction (θ_{Wind}).

5 Suggestion for Future Work

The simplified ARMA model developed in this study is the improvement of the basic ARMA model developed in the previous study in [4] and will be used in the future as a platform to test any proposed control system and strategy to maintain the thermal comfort in iHouse. Like the ARMA model developed in the previous study [4], the simplified ARMA model developed in this study only considers weather-related inputs during the model development due to the unavailability of historical data that was recorded when thermal comfort devices were operated. The next step is to include the control-related input(s) produced by thermal comfort device(s) so that the developed model can be used to simulate any proposed control system and strategy to maintain the thermal comfort in iHouse. New data will be recorded in iHouse while the thermal comfort devices (air conditioner and motor operated window) are operated. The study in this area is in progress and will be published once it is completed.

The equation of the model can be improved to increase the model's accuracy. One of the suggestion is to include the mathematical equations to represent the heat transfer through the door and windows. Due to time constrain, all the walls of bedroom A in this study is assumed to be plain walls (without window and door) to simplify the model development. Investigation should be done on the model's performance when the model is expanded to have mathematical equations describing the heat transfer through door and windows.

In addition, additional mathematical equations describing the heat transfer through the envelope of bedroom A can be investigated and added.

6 Conclusion

The main goal of this study is to simplify the ARMA model describing the thermal behaviour of iHouse developed in [4] based on fundamental knowledge in science and engineering. Through this study, it is shown that the simplified ARMA model with lesser inputs and model's order can still perform almost on par compared to the ARMA model in [4] and House Thermal Simulator in [8]. This is supported by the results presented in Sect. 3 which show that the accuracy of the simplified ARMA model developed in this study is slightly less compared to the other models of the previous studies, but the accuracy differences are not big. In addition, the model developed in this study is simpler to be implemented and faster to be simulated due to the lower number of inputs and model's order. The main contribution from this finding is in the simplification of ARMA model based on theoretical knowledge that can be implemented easier and simulated faster.

Acknowledgement. We would like to express our gratitude to the Japan Student Services Organization (JASSO) for their financial support and the Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) for their well-managed attachment program.

References

1. Seborg, D.E., Mellichamp, D.A., Edgar, T.F., Doyle, F.J.: Process Dynamics and Control. Wiley, Hoboken (2010)
2. Nise, N.S.: Control Systems Engineering, 6th edn. Wiley, Hoboken (1995)
3. Atam, E., Helsen, L.: Control-Oriented thermal modeling of multizone buildings: methods and issues: intelligent control of a building system. IEEE Control Syst. **36**, 86–111 (2016). doi:10.1109/MCS.2016.2535913
4. Mohd Hussein, S.F., Nguyen, H., Abdullah, S.S., et al.: Black box modelling the thermal behaviour of iHouse using auto regressive and moving average (ARMA) model. J. Teknol. **78**, 51–58 (2016). doi:10.11113/jt.v78.9272
5. Mustafaraj, G., Lowry, G., Chen, J.: Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. Energy Build. **43**, 1452–1460 (2011). doi:10.1016/j.enbuild.2011.02.007
6. Mustafaraj, G., Chen, J., Lowry, G.: Thermal behaviour prediction utilizing artificial neural networks for an open office. Appl. Math. Model. **34**, 3216–3230 (2010). doi:10.1016/j.apm.2010.02.014
7. Hazyuk, I., Ghiaus, C., Penhouet, D.: Optimal temperature control of intermittently heated buildings using model predictive control: part I - building modeling. Build. Environ. **51**, 379–387 (2012). doi:10.1016/j.buildenv.2011.11.009
8. Nguyen, H., Makino, Y., Lim, A.O., Tan, Y., Shinoda, Y.: Building high-accuracy thermal simulation for evaluation of thermal comfort in real houses. In: Biswas, J., Kobayashi, H., Wong, L., Abdulrazak, B., Mokhtari, M. (eds.) Inclusive Society: Health and Wellbeing in the Community, and Care at Home, pp. 159–166. Springer, Heidelberg (2013)