Jurnal
Teknologi

Full Paper

BLACK BOX MODELLING THE THERMAL BEHAVIOUR OF IHOUSE USING AUTO REGRESSIVE AND MOVING AVERAGE (ARMA) MODEL

Article history
Received
8 November 2015
Received in revised form
22 March 2016
Accepted
22 March 2016

Shamsul Faisal Mohd Hussein[a], Hoaison Nguyen[b], Shahrum Shah Abdullah[a*], Yuto Lim[c], Yasuo Tan[c]

[a]Biologically Inspired System and Technology (Bio-iST) iKohza, Malaysia – Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak), 54100 Kuala Lumpur, Malaysia
[b]Department of Networks and Computer Communications, Faculty of Information Technology, VNU University of Engineering and Technology, Building E3, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam
[c]School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Nomi, Ishikawa Prefecture 923-1211, Japan

*Corresponding author
shahrum@fke.utm.my

Graphical abstract

Abstract

Modelling and simulation of the dynamic thermal behaviour of a building is important to test any proposed thermal comfort control system and strategy in the building. A simulation model can be obtained by using either the white box, grey box or black box modelling method. This research focuses on the usage of auto regressive and moving average (ARMA) model, a type of black box model that represents the dynamic thermal behaviour of iHouse testbed and uses real recorded data from the testbed and limited knowledge regarding the physical characteristics of the testbed. The performance of the ARMA model developed in this research is compared with the performance of House Thermal Simulator, a previously developed model, based on grey box modelling. Results obtained shows that ARMA model works better than House Thermal Simulator in some aspects.

Keywords: Modelling and simulation; black box modelling; building temperature simulation; building temperature prediction

1.0 INTRODUCTION

Modelling and simulation of the dynamic thermal behaviour of a building is very important to aid in the design of thermal comfort control system & strategies in buildings. Testing the proposed control systems and strategies on actual plant is expensive and in some situation might be dangerous. By having a dynamic thermal behaviour model of a building, researchers can simulate their proposed control systems and strategies cost effectively.

Rick Kramer et al. [1] conducted a study on types of building's modelling approaches together with their strengths and weaknesses. Like other plants, the method of obtaining the thermal behaviour model of a building can be categorized into 3 types, which are: (1) white box modelling (also known as theoretical modelling [2]); (2) black box modelling (also known as empirical modelling [2]); (3) and grey box modelling (also known as semi-empirical modelling [2]). White box modelling of a system is done based on the physical knowledge (also known as fundamental laws of science and engineering [3]) to describe the system while black box modelling of a system is done based on the input(s)-output(s) data-set to describe the system. Grey box modelling lies somewhere between white box and black box modelling – e.g. the model is built based on physical knowledge to describe the system but the unknown parameters in the model are obtained by estimation process from input(s)-output(s) data-set.

Each modelling approach has its own strengths and weaknesses. White box model gives access to the physical insight of the system and can be simulated over a wider range of operating conditions, but the development of this model is time consuming and expensive [2]. In addition, sometimes the value of some parameters in the white box models for complex plants might not be available during model development [2]. Black box model is easier to be constructed, but does not usually extrapolate well outside the range of the training data-set (the data-set that is used to estimate the model) [2]. Most of the time, the training data-set available does not consist of the whole operating point of the represented plant/process [2]. Caution must be taken if it is required to simulate the model with operating conditions that are outside the range of the training data-set [2]. However, it is reported in [2] that black box model is popular in the industry. Meanwhile, grey box model provides physical insight like the white box model, can be simulated over a wider range of operating points than the black box model and takes faster time to be developed than the white box model [2].

White box (and also grey box) modelling can be done either directly or by converting the physical model into an electrical circuit analogy equivalent first. Nguyen et al. [4] developed House Thermal Simulator, a MATLAB® and Simulink® based program to simulate the thermal behaviour of iHouse, a smart house testbed that belongs to the Japan Advanced Institute of Science and Technology (JAIST), which is shown in Figure 1 – the model was constructed using detail physical knowledge, but the values of the unknown parameters were estimated based on the input-output relationship data recorded from experiments. Hazyuk et al. [5] obtained a low order model for a typical French detached house by constructing an electrical circuit analogy equivalent to the house, then simulated the house's detail model in Simbad toolbox for MATLAB® and Simulink® to get input-output relationship data and finally used real recorded data to estimate the parameters for the low order model.

Black box modelling can be done by using linear or non-linear models. Mustafaraj et al. modelled the thermal behaviour of an actual office on the 7th floor, Visa Building, London using both linear and non-linear models in [6]. Mustafaraj et al. also conducted a study on the performance of neural network auto-regressive model with exogenous inputs (NNARX), neural network auto regressive moving average models with external input (NNARMAX), and neural network output error model (NNOE) in [7]. While modelling the temperature of a greenhouse, Frausto et al. studied the performance of 2 linear models [8], which were ARX and ARMAX. In addition, Frausto et al. also modelled the indoor temperature of a greenhouse using a non-linear model, which was NNARX in [9]. Patil et al. also conducted a study on the performance of ARX, ARMAX and NNARX while modelling the indoor temperature of a greenhouse in [10]. Rabl [11] used data from an office building to review various black box modelling methods (thermal networks, modal analysis, differential equations, auto regressive and moving average (ARMA) model, Fourier series, and computer calibrated model) – this study concluded that ARMA model and differential equation are recommended to be used as the basic starting point due to their simplicity and parameters can be added systematically to improve accuracy.

A dynamic model is just a set of mathematical equation(s) that approximately represents the dynamic behaviour of the actual system (also known as plant or process), so the model cannot behave 100% exactly like the actual system at all operating points [2]. Depending on the application or purpose of the model, consideration should be taken during the development stage to maintain the balance between the model's accuracy and simplicity – a very accurate model might end up being too complicated and wasting too much resources (cost and time) while a very simple model might end up not being able to serve its original development purpose [2].

This research focuses on modelling the thermal behaviour of iHouse, the same plant modelled in [4]. Even though there was already a work previously done to model the thermal behaviour of the house using grey box modelling, not enough detail information of the developed model is known by the time this research is done. It is possible to obtain all the necessary information about the physical characteristic of the plant that is going to be modelled and also the previous work regarding the grey box modelling of the plant, but time limitation is one of the main factors in this research. The motivation of developing the model is to simulate any proposed control system and strategy to maintain the thermal comfort in iHouse in the future. Due to limited resources (time, information related to the characteristic of the house and knowledge related to previous grey box modelling of the house), it is decided to remodel the thermal behaviour of the house using black box modelling method. As a starting point, an ARMA model is used for black box modelling in this research due to its simplicity and ability to add parameters systematically to improve accuracy [11]. The result of the ARMA model obtained from this research is compared with the result of House Thermal Simulator developed in the previous work in [4].

Figure 1 The photo of iHouse, the smart house facility that belongs to the Japan Advanced Institute of Science and Technology (JAIST)

2.0 METHODOLOGY

2.1 Scope of Research

Firstly, only 1 out of 15 available rooms in iHouse is remodelled in this research. The floorplan of iHouse is depicted in Figure 2 and the room that is selected to be remodelled is bedroom A.

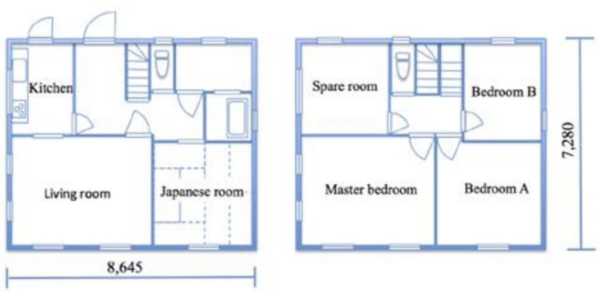


Figure 2 The floor plan of iHouse

Secondly, only weather related inputs are considered for the model in this research. The inputs from actuators (also known as final elements) used as heating and cooling devices that are used to regulate the temperature inside iHouse are not considered in this research. The air conditioners and windows are some of the heating and cooling devices available in iHouse. The experimental data recorded in bedroom A for this research is done based on the following conditions: (1) the air conditioner is switched off; (2) and the windows are closed.

Next, the number of past input(s) for each type of input used by the ARMA model in this research is standardized so that all types of input for this model will have the same amount of past input(s). For ARMA model, the number of past input(s) is also known as the number of term(s).

Finally, the simulation result of the model produced in this research is compared with the simulation result of the model developed by Nguyen et al. in [4] – both models will be estimated using the same training data-set and then tested using the same testing data-set.

2.2 Data Collection

Instead of performing new experiments to record new sets of data for this research, historical data recorded in the past that fulfil the condition requirement defined in Sub-section 2.1 – Scope of Research is used in order to save time. The historical data is selected based on the following conditions: (1) the air conditioner is switched off; (2) and the windows are closed. The historical data with similar experimental conditions (the air conditioner is switched off and the windows are closed) and as close as possible with each other (in terms of the dates of the conducted experiments) are chosen. The reason why it is decided to choose two groups of recorded data-sets with the recording dates that are as close as possible with each other is because the weather related inputs will vary significantly if the recording dates between these two groups are separated for a very long time, especially in four-season countries. Two groups of available historical data fulfilling these criteria and requirements are identified – the first group of historical data was recorded from the 1st of August 2012 until the 3rd of August 2012 while the second group of historical data was recorded from the 10th of August 2012 until the 19th of August 2012 – these groups of data are separated between each other by only six days.

These groups of data are assigned as training data-set and testing data-set. The training data set is used for model regression while the testing data-set is used to test the performance of the regressed model with new data-set that has never been seen by the model. The historical data recorded from the 1st of August 2012 until the 3rd of August 2012 is assigned as training data-set while the historical data recorded from the 10th of August 2012 until the 19th of August 2012 is assigned as testing data-set.

Different types of data in iHouse were recorded using various types of sensors at different sampling time intervals, as explained in detail in [4]. However, in this research the sampling time of all types of data are standardized at every 90 seconds in order to simplify the model regression algorithm and also to shorten the time taken for model simulation. In this case, there will be 960 recorded inputs-outputs relationships per day if recorded data is sampled at every 90 seconds. Therefore, the training data-set (recorded from the 1st of August 2012 until the 3rd of August 2012) will consist of 2880 recorded inputs-outputs relationships while the testing data-set (recorded from the 10th of August 2012 until the 19th of August 2012) will consists of 9600 recorded inputs-outputs relationships.

2.3 Data Selection

The plant modelled in this research is a multiple-input and single-output (MISO) system. The output assigned for this model is the future temperature of bedroom A

(T_{BedA}). Meanwhile, the inputs that may have the potential to influence the temperature in bedroom A are determined and listed, 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 bedroom A itself – based on Figure 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 relative humidity of bedroom A (RH_{BedA}); (4) the past heat emitted from residence(s) and electronic device(s) in bedroom A ($Q_{ResDevBedA}$); (5) the past radiation heat from outside air ($Q_{AirRadBedA}$); (6) the past solar radiation – there are 3 types of solar radiation data available from the work done in [4], which are direct solar radiation (φ_{Dir}), diffuse solar radiation (φ_{Diff}) and global solar radiation (φ_{Glo}); (7) the past solar position – there are 2 types of solar position data available, which are solar altitude (Z_{Solar}) and solar azimuth (θ_{Solar}); (8) and the past wind velocity – the original available data related to wind are wind speed and wind direction, but this vector quantity is resolved using trigonometric method into wind velocities that are coming from 4 directions, which are the wind velocity blowing from the North (V_{NWind}), the wind velocity blowing from the East (V_{EWind}), the wind velocity blowing from the South (V_{SWind}) and the wind velocity blowing from the West (V_{WWind}). This model with the determined inputs and output is illustrated in Figure 2.

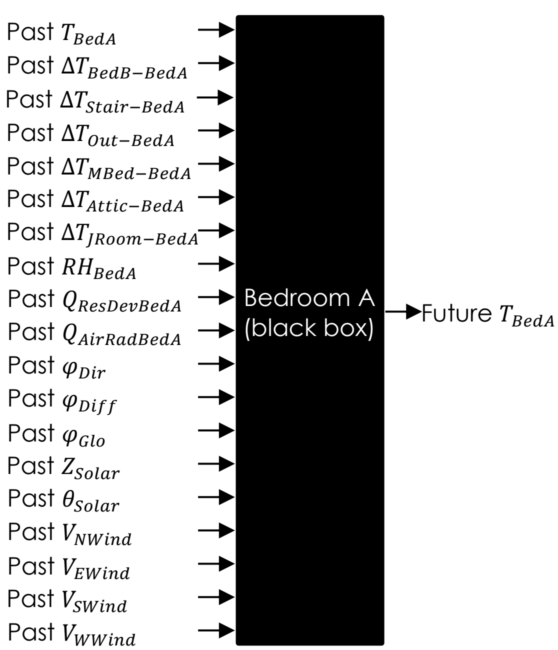


Figure 2 The illustration of model for the thermal behaviour of bedroom A with listed possible inputs

2.4 Model Construction

Let's say that the ARMA model uses k past input(s) (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, the ARMA model equation with k past input(s) describing Figure 2 is written, which is shown below:

$$\begin{aligned} T_{BedA}[n+k] = {} & \textstyle\sum_{i=0}^{k-1} A_i T_{BedA}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} B_i \Delta T_{BedB-BedA}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} C_i \Delta T_{Stair-BedA}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} D_i \Delta T_{Out-BedA}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} E_i \Delta T_{MBed-BedA}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} F_i \Delta T_{Attic-BedA}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} G_i \Delta T_{JRoom-BedA}[n+i] + \sum_{i=0}^{k-1} H_i RH_{BedA}[n+ \\ & i] + \textstyle\sum_{i=0}^{k-1} I_i Q_{ResDevBedA}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} J_i Q_{AirRadBedA}[n+i] + \sum_{i=0}^{k-1} K_i \varphi_{Dir}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} L_i \varphi_{Diff}[n+i] + \sum_{i=0}^{k-1} M_i \varphi_{Glo}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} N_i Z_{Solar}[n+i] + \sum_{i=0}^{k-1} O_i \theta_{Solar}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} P_i V_{NWind}[n+i] + \sum_{i=0}^{k-1} Q_i V_{EWind}[n+i] + \\ & \textstyle\sum_{i=0}^{k-1} R_i V_{SWind}[n+i] + \sum_{i=0}^{k-1} S_i V_{WWind}[n+i] \end{aligned} \tag{1}$$

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

$$\begin{aligned} T_{BedA}[n+k] = {} & A_{k-1} T_{BedA}[n+k-1] \\ & + A_{k-2} T_{BedA}[n+k-2] + \cdots \\ & + A_1 T_{BedA}[n+1] + A_0 T_{BedA}[n] \\ & + B_{k-1} \Delta T_{BedB-BedA}[n+k-1] \\ & + B_{k-2} \Delta T_{BedB-BedA}[n+k-2] \\ & + \cdots + B_1 \Delta T_{BedB-BedA}[n+1] \\ & + B_0 \Delta T_{BedB-BedA}[n] + \cdots \\ & + S_{k-1} V_{WWind}[n+k-1] \\ & + S_{k-2} V_{WWind}[n+k-2] + \cdots \\ & + S_1 V_{WWind}[n+1] \\ & + S_0 V_{WWind}[n] \end{aligned} \tag{2}$$

2.5 Model Regression

Regression is the process of estimating the unknown parameters in ARMA model (also known as 'training' if it is the process of estimating the unknown parameters in artificial neural networks).

If the training data-set is recorded from $t = 0$ until $t = p$, this means that the number of available sampled data are $p + 1$ (recorded from $t = 0$ until $t = p$). In this research, $p + 1$ 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 Sub-section 2.2 – Data Collection). When the data from k previous steps (also known as sampling intervals) 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 Equation (3) below:

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

where:

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

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

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

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

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

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

and $$\mathbb{M} = \begin{bmatrix} A_{k-1} \\ A_{k-2} \\ \vdots \\ A_1 \\ A_0 \\ B_{k-1} \\ B_{k-2} \\ \vdots \\ B_1 \\ B_0 \\ \vdots \\ S_{k-1} \\ S_{k-2} \\ \vdots \\ S_1 \\ S_0 \end{bmatrix}.$$

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

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

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. After that, the model can be tested and optimized with new data-set that has never been seen by the model – these processes will be discussed in the next sub-sections.

2.6 Model Testing

In this process, the regressed 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 is able to 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 was recorded from $t = 0$ until $t = q$. This means that the number of available sampled data are $q + 1$ (recorded from $t = 0$ until $t = q$). In this research, $q + 1$ is equal to the number of data recorded from the 10th of August 2012 until the 19th of August 2012, which is 9600 (as mentioned earlier in Sub-section 2.2 – Data Collection).

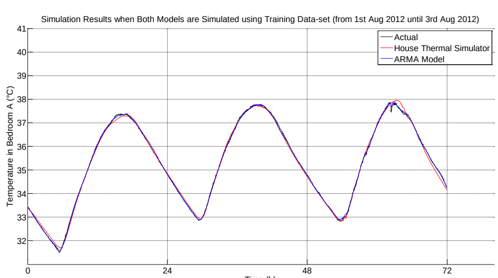
2.7 ARMA Model's Parameter Optimization

For ARMA model, the only parameter that needs to be tuned in order to optimize the model is the number of past input(s), which is the value of k. The bigger the value of k, the more the quantity of constants available in matrix $\mathbb{M}$ (and vice versa). Instead of assigning the value of k randomly and manually using trial and error method, a MATLAB® script is written to try the possible values of k one by one – due to time constraint, the value of k in this research is tried one by one only from $k = 1$ until $k = 100$. The percentage of fitness, $\%Fit$ is calculated for each tested value of k. The formula to calculate $\%Fit$ are shown in the equation below:

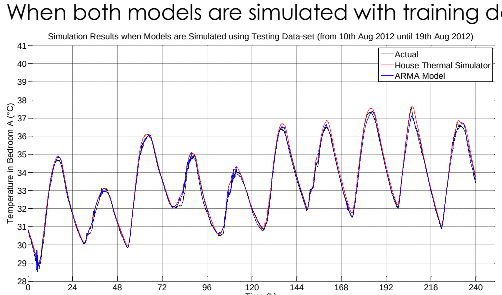
$$\%Fit = \left(1 - \frac{norm(\hat{T}_{BedA} - T_{BedA})}{norm[T_{BedA} - mean(T_{BedA})]}\right) \tag{5}$$

3.0 RESULTS

The performance of the optimized ARMA model is compared with the performance of the optimized House Thermal Simulator [4]. House Thermal Simulator [4] is also estimated and optimized by using its own parameter estimation and optimization method, and using the same training and testing data-sets. The simulation results for both the optimized ARMA model and House Thermal Simulator [4] during simulation are plotted and displayed in Figure 3 for comparison.



(a) When both models are simulated with training data-set



(b) When both models are simulated with testing data-set

Figure 3 The simulation result comparison for both optimized models

The $\%Fit$ values for both optimized models during simulation are presented in Table 1.

Table 1 The percentage of fitting, %Fit comparison for both optimized models during simulation

a) When both models are simulated with training data-set

Data	Percentage of Fitting, $\%Fit$ (%)
House Thermal Simulator [4]	94.8240
ARMA Model	97.6667

b) When both models are simulated with testing data-set

Data	Percentage of Fitting, $\%Fit$ (%)
House Thermal Simulator [4]	88.0188
ARMA Model	91.6101

4.0 DISCUSSION

The discussion focus on the accuracy of the ARMA model. Overall, based on the obtained results, it can be seen that the optimized ARMA model proposed in this research can fit the actual data better than the optimized House Thermal Simulator developed in [4] most of the time during both regression and testing process. This is supported by the result presented in Section 3.0, which shows that the optimized ARMA model produces higher percentage of fitting, $\%Fit$ than House Thermal Simulator [4]. In addition to the result shown in Section 3.0, obvious spikes can also be observed from the output generated from the optimized ARMA model, which is highlighted with red circles in Figure 4.

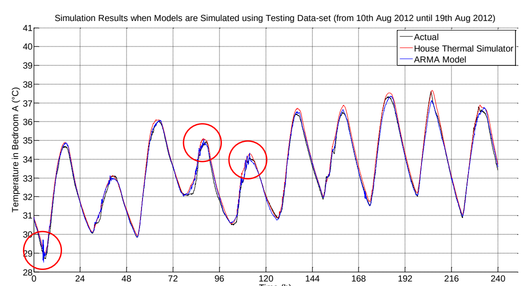


Figure 4 The spikes produced by the optimized ARMA model (highlighted with red circles) when simulated with testing data-set

One of the possible reasons why spikes appear at certain simulation periods when the optimized ARMA model is simulated with testing data-set is because of the insufficient size of training data-set. House Thermal Simulator [4] is developed based on grey box modelling approach where the mathematical model is constructed based on physical knowledge of the system but the unknown parameters in the model are estimated based on the input-output relationship data that was recorded from the actual system – if the constructed model consists sufficient mathematical equations describing the physical characteristic of a system, the model can still perform robustly even if the unknown constants in the equations are estimated using smaller but adequate size of training data-set. Meanwhile, ARMA model is a black box model which is purely estimated based on the input-output relationship data that is recorded from the actual system. The larger the training data-set, the more accurate and robust the model describing the system will be.

5.0 SUGGESTIONS FOR FUTURE WORK

This research has thrown up many questions in need for further investigations. It is recommended that further research be undertaken in the following areas: (1) model accuracy improvement; (2) and additional inputs from thermal comfort device(s).

5.1 Model Accuracy Improvement

Many things can be done in order to improve the accuracy of the estimated model. First, it is suggested that the inputs that may have the potential to influence the temperature in bedroom A should be revised – newer input(s) that might have the potential to influence the temperature should be included while existing input(s) that might not have the potential to influence the temperature should be removed.

Second, it is also suggested to use larger data-set (more data) for the regression process. A good data-set should be large enough and cover various weather conditions. If there is a requirement to regress a model using a data-set recorded within certain seasons but simulate the regressed model with data-set recorded on other seasons, it is suggested to utilize the recursive least square or other on-line tuning methods. On-line tuning allows the model to keep on updating its parameters whenever the accuracy of the model decreases because the model is simulated within the operating condition that is outside the range of the training data-set. The study in this area is in progress and will be published once it is completed.

Third, if the model is still not accurate even though the potential of linear model such as ARMA model (and its derivatives/variants) etc. has been explored to the limit, it is suggested to switch modelling the system using non-linear models, such as artificial neural networks (ANN) etc. There is a possibility that linear models do not perform accurately due to the presence of non-linear relationship between inputs and output. The study in this area is also in progress and will be published once it is completed.

Fourth, besides using self-written program, it is also suggested to identify the system using readily-available programs in the market, such as MATLAB® and Simulink® System Identification Toolbox™ etc. These programs give users opportunity to explore the system identification process using various models and parameter identification methods.

Fifth, it is also suggested to investigate and explore the potential of black box models with different number of past input(s) (the value of k) for each type of input. Different values of k for different types of inputs increase the model's flexibility. However due to time constrain during this research, all types of inputs were assigned with similar values of k in order to minimize the time taken during the regression process. The study in this area is also in progress and will be published once it is completed.

Last but not least, to use the experience and knowledge obtained from this study to develop models based on grey box modelling approach. One of the objectives of this study is to help provide insights to House Thermal Simulator previously developed in [4]. Besides helping to understand the previous work, the experience obtained from this study can also be used to improve House Thermal Simulator [4] or to develop new models based on grey or white box modelling approach.

5.2 Additional Input(s) from Thermal Comfort Devices.

In the future, the ARMA model developed in this study will be used as a platform to test any proposed control system and strategy to maintain the thermal comfort in iHouse. As the initial step, this study only considers weather related inputs during the model construction. What is now needed is to expand the model by also considering the input(s) from thermal comfort device(s). Instead of including only a single thermal comfort device, it is also possible to include multiple thermal comfort devices, such as air conditioner, fan, window, curtain, heater etc. The study in this area is also in progress and will be published once it is completed.

6.0 CONCLUSION

The main goal of this study is to develop a model describing the thermal behaviour of iHouse based on black box modelling using ARMA model. Through this study, it is shown that without enough information available prior to modelling, a data-driven black box model using the right input-output combination can still perform at least equal to a physical grey box model. This is supported by the results presented in Section 3.0 which show that the optimized ARMA model perform better than House Thermal Simulator in some aspects. The main contribution from this finding is in the construction of a mathematical model to simulate the system in a short time with limited physical knowledge of the system.

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. This work has also been partly supported by Vietnam National University, Hanoi (VNU), under Project No. QG.16.30.

References

[1] R. Kramer, J. van Schijndel, H. Schellen. 2012. Simplified thermal and hygric building models: A literature review. Front. Archit. Res. 1 . 318–325.

[2] D.E. Seborg, D.A. Mellichamp, T.F. Edgar, F.J. Doyle. 2010. Process Dynamics and Control. John Wiley & Sons Incorporated.

[3] N.S. Nise. 2010. Control Systems Engineering. 6th Ed. John Wiley & Sons Incorporated.

[4] H. Nguyen, Y. Makino, A.A. Lim, Y. Tan, Y. Shinoda. 2013. Building HighAccuracy Thermal Simulation for Evaluation of Thermal Comfort in Real Houses. Incl. Soc. Heal. Wellbeing CommunityCare Home SE 20.159–166.

[5] Hazyuk, C. Ghiaus, D. Penhouet. 2012. Optimal Temperature Control of Intermittently Heated Buildings Using Model Predictive Control: Part I – Building modeling. Build Environ. 51: 379–387.

[6] G. Mustafaraj, G. Lowry, J. Chen. 2011. Prediction of Room Temperature and Relative Humidity by Autoregressive Linear and Nonlinear Neural Network Models for an Open Office. Energy Build. 43: 1452–1460.

[7] G. Mustafaraj, J. Chen, G. Lowry. 2010. Thermal Behaviour Prediction Utilizing Artificial Neural Networks for an Open Office. Appl. Math. Model. 34: 3216–3230.

[8] H.U. Frausto, J.G. Pieters, J.M. Deltour. 2003. Modelling Greenhouse Temperature by means of Auto Regressive Models. *Biosyst. Eng*. 84: 147–157.

[9] H. Uchida Frausto, J.G. Pieters. 2004. Modelling Greenhouse Temperature Using System Identification by Means of Neural Networks. *Neurocomputing*. 56: 423–428.

[10] S.L. Patil, H.J. Tantau, V.M. Salokhe. 2008. Modelling of Tropical Greenhouse Temperature by Auto Regressive and Neural Network Models. *Biosyst. Eng*. 99: 423–431.

[11] Rabl. 1988. Parameter Estimation in Buildings: Methods for Dynamic Analysis of Measured Energy Use. *J. Sol. Energy Eng. Trans. ASME*. 110: 52–66.

.