Black box modelling and simulating the dynamic indoor air temperature of a laboratory using (ARMA) model

SFM Hussein[1], N.B. Sharifmuddin[2], A. O. Al rabeei[3], A Faruq[4],
M.S Noorazizi[5], S. A. Zaki[6], SS Abdullah[7]

[1,3,4,7]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, Malaysia
[2]Takasago Thermal/Environmental Systems (TTES) iKohza, Department of Mechanical Precision Engineering (MPE), Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia
[4]Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang, Malang, Indonesia
[5]Engineering Department, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia
[6]Wind Engineering for (Urban, Artificial, Man-made) Environment Laboratory, Department of Mechanical Precision Engineering (MPE), Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia

Article Info

ABSTRACT

Mathematical model representing the dynamic indoor air temperature of a building is important for reducing the time and cost required to test any proposed thermal comfort control algorithm and strategy for that building through computer simulation. There are many types of mathematical model, and each type has its strength(s) and weakness(es). An autoregressive–moving-average (ARMA) model, a type of black box model is used to represent the dynamic indoor air temperature behaviour of industrial instrumentation laboratory at Malaysia-Japan international institute of technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur based on the recorded data from the laboratory and minimal physical characteristics knowledge of the laboratory. The ARMA model's output developed in this research is compared with the actual data recorded from the laboratory for performance measurement. The obtained result shows that the ARMA model is sufficient for modelling and simulating the dynamic indoor air temperature behaviour of the laboratory.

Corresponding Author:

SS Abdullah
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
Email: shahrum@utm.my

1. INTRODUCTION

Modelling and simulation of a building's hygrothermal (temperature and humidity) behaviour is important to help to design and optimise any proposed building thermal comfort control algorithm and strategy. It is possible to design and optimise the proposed control algorithm and strategy on the real site, but

doing so is time-consuming and costly. To solve this problem, mathematical models describing the dynamic hygrothermal behaviour of the building are developed so that any proposed control algorithm and strategy can be designed and optimised quickly and cost-effectively. In addition, the developed mathematical model can also be used for developing and optimising predictive controller such as model predictive controller (MPC).

Like other models, the dynamic indoor air temperature behaviour of a building can also be modelled by using three types of model: (1) the white box model, also referred to as theoretical model; (2) the black box model, also referred to as empirical model; and (3) the grey box model, also referred to as semi-empirical model. The white box model of a system is developed based on the system's physical knowledge [1], also referred to as the fundamental knowledge of science and engineering [2]. The black box model is developed by tuning the constant(s) in linear or non-linear 'off-the-shelf' equation(s) based on the input(s)-output(s) data of the plant that is going to be modelled until the output(s) of the black box model almost matches (as accurately as possible) the output(s) of the plant. The grey box model is the combination of some parts of the white box model and some parts of the black box model – the equation(s) in the grey box model is(are) based on the theoretical knowledge of the plant while the unknown parameter(s) in the grey box model is(are) tuned based on the real input(s)-output(s) data generated by the plant.

Each type of model has its own strength(s) and weakness(es). The white box model can be simulated in a broader range of operating condition, but it is time-consuming and costly to develop this model if the system to be modelled has complex physical characteristic(s) [1]. In some situation, it is difficult to develop the white box model because the constant(s) in the model is(are) not available, obtainable, or measurable during the model development [1]. The black box model is simpler to build, but the model does not extrapolate well beyond the data used to estimate the model, known as the training data set [1]. The available training data set usually does not cover the whole operating condition of the plant that is going to be modelled, so caution must be taken if there is any requirement to simulate the developed model beyond the range of the training data set [1]. However, it is reported that the black box model is popular in the industry [1]. Meanwhile, the grey box model provides the physical insight similar to the white box model because the grey box model is also built based on the fundamental knowledge of science and engineering, but the unknown constant(s) in the grey box model is(are) estimated by using the input(s)-output(s) data of the modelled plant similar to the black box model. Therefore, the grey box model can be developed faster and easier than the white box model and can be simulated across a broader range of operating condition than the black box model [1].

Various researchers had developed various types of mathematical models describing the building's dynamic indoor air temperature behaviour. Nguyen et al. [3] created a SIMULINK®-based simulation toolbox called House Thermal Simulator to simulate the dynamic indoor air temperature of iHouse, a smart house testbed belongs to Japan Advanced Institute of Science and Technology (JAIST) by using a large number of equations based on the grey box model describing how the controllable inputs and uncontrollable disturbances affect the iHouse's dynamic indoor air temperature – the unknown parameters in these equations were estimated using the Simulink Design Optimization toolbox in MATLAB® based on the actual recorded data from iHouse. Then, Hussein et al. developed a black box model to simulate the dynamic indoor air temperature of one of the rooms in the same iHouse in [4] with minimal thermal-related physical knowledge of the smart house testbed. After that, Hussein et al. simplified the model developed in [4] to become a grey box model in [5] based on more thermal-related physical knowledge of the smart house testbed. Nguyen et al. used the House Thermal Simulator developed in [3] to propose a simple model predictive controller (MPC) in [6] to control the heating, ventilation, and air conditioning (HVAC) systems in residential houses through simulation and optimised the controller for both energy efficiency and thermal comfort. Next, Ooi et al. developed their mathematical models in [7] for developing and optimising a conventional MPC via computer simulation to maintain the air temperature of one of the bedrooms in iHouse – this work is then improved by Ooi et al. in [8] by using the adaptive MPC that utilises two types of online model estimation system, the Kalman filter (KF) state estimator and the linear time-varying Kalman filter (LTVKF) estimator for increasing the accuracy of the controller's internal plant model.

Lim et al. in [9] used a simple white box model representing the thermal behaviour of one of the rooms in iHouse to examine and discuss the issues of implementing the energy-efficient thermal comfort control (EETCC) system via cyber-physical system (CPS) technique for monitoring, maintaining, and controlling the desired thermal comfort level using three actuators, which are: (1) air conditioner; (2) window; and (3) curtain. Radecki et al. in [10] proposed the implementation of a multimode unscented Kalman filter (UKF) as a generalisable online grey-box model based on the basic physical knowledge of the building combined with measured building data to estimate the dynamics of a multi-zone building and to find out the unknown time-varying thermal loads of that building – the general and scalable method of developing this control-oriented thermal model is deployable in a wide-scale for cost-effective predictive controls.

Mustafaraj et al. in [11] used the following models to predict the room temperature and relative humidity of a visa building in London: (1) autoregressive model with external inputs (ARX); and (2) neural network-based non-linear autoregressive model with external inputs (NNARX) – their results showed that both ARX and NNARX performed well, but the NNARX performed better than ARX.

Mustafaraj et al. in [12] studied the quality of the following models to predict an open-plan office's thermal behaviour in Portman House 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) – their results indicated that all models performed equally well, but the NNARX and NNARMAX models outperformed the NNOE model. Afram et al. in [13] developed the grey box models of the residential HVAC system using the data recorded at the instrumented Toronto and Region Conservation Authority Archetype Sustainable House (TRCA-ASH) located at Kortright Centre for Conservation in Vaughan, Ontario, Canada as a tool to investigate the effects of advanced controllers and energy conservation strategies via simulation before implementing them on the real system – this model is comprised of several subsystems such as energy recovery ventilator (ERV), air handling unit (AHU), buffer tank (BT), radiant floor heating (RFH) system, and zone and ground source heat pump (GSHP) that can be estimated individually and then combined to get the complete system model. The same work in [13] was repeated by Afram et al. in [14, 15] for the same purpose, but this time each of the same subsystems at TRCA-ASH was remodelled using different types of black box model, which are: (1) artificial neural network (ANN); (2) transfer function (TF) model; (3) process model, (4) state-space (SS) model; and (5) autoregressive exogenous(ARX) – result shows that all the black box models developed in [14, 15] outperformed the grey box model developed in [13]. After that, Afram et al. extended his work in [13] to be a multi-zone residential HVAC system in [16] to use it to design the simulation and experimental framework for MPC based supervisory controller to shift the heating and cooling load of a house that were controlled using on-off controllers outside peak hours for significant heating and cooling costs savings.

Sturzenegger et al. in [17] proposed and tested the derivation of MPC-compatible reduced-order model of a medium-size representative office in Allschwil, Switzerland which is equipped with the following thermal comfort actuators: (1) thermally activated building system (TABS); air handling unit (AHU); and (3) centrally controlled blinds. Sturzenegger et al. also modelled the same building in [17] using a physics-based bilinear model for MPC in [18] to control the thermal comfort actuators and compared its performance with the existing industry-standard rule-based control (RBC). A. Safa et al. in [19] developed a house model in TRNSYS, one of the commercial building energy simulation programs based on the design and parameters of the TRCA-ASH. Then, Alibabaei et al. in [20] developed a MATLAB®-TRNSYS co-simulator, an integration between MATLAB® and TRNSYS to control or manage the TRCA-ASH model developed in [19] by using the advanced controller designed in MATLAB® using the following predictive strategy planning models: (1) Load Shifting (LSH); (2) Smart Dual Fuel Switching System (SDFSS); and (3) the integration between LSH and SDFSS (LSHSDFSS). De Coninck et al. in [21] identified the model of Kalkkaai building, the headquarter of 3E, an independent global consultancy services and software products company situated in Brussels, Belgium by using used the Grey-Box Buildings toolbox, an open-source Python toolbox based on a Modelica library. Then, De Coninck et al. used the model identified in [21] to implement an MPC in [22] for thermal comfort control and cost saving, and compared its performance with the default rule-based control (RBC) at the Kalkkaai building. Hazyuk et al. in [23] used physical knowledge to determine the black box model's structure, then used the model to represent the dynamic behaviour of the indoor temperature of a typical detached house in France, which was used by Centre Scientifique et Technique du Bâtiment (CSTB)/Building Scientific and Technical Centre as one of the reference building for performance assessment.

This research focuses on constructing the model describing the dynamic indoor air temperature behaviour of the Industrial Instrumentation Laboratory at Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur. Similar work has been done by the main author in [4, 5], but this time it is repeated in a different location so that it will be easier for the main author to expand his work in the future. To save time, cost, and other limited resources by the time this research is done, only the autoregressive–moving-average (ARMA) model, a type of simple linear black box model is used in this research to develop the model with minimal physical knowledge.

2. RESEARCH METHOD
2.1. Scope of research
Only the weather-related inputs are used for the model developed in this research. The Industrial Instrumentation Laboratory is equipped with air conditioners and ventilation fans, but the data available by the time this research is done is recorded when these thermal comfort devices are not operated – newer data will be recorded while these thermal comfort devices are operated for the model's future upgrade.

2.2. The research location

The Industrial Instrumentation Laboratory at Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur is selected as the plant to be modelled in this research. The laboratory is located on the 7th floor (based on the US-style) or the 6th floor (based on the UK-style) of the MJIIT building. It is also located at the front-side of the building (which is also the western side of the building) on the south wing.

The typical floor plan of the Industrial Instrumentation Laboratory is shown in Figure 1. Based on this figure, it is shown that the Industrial Instrumentation Laboratory is surrounded by the following spaces: (1) the staircase in the north-west; (2) the lift lobby in the north; (3) the corridor in the east; (4) the outdoor space in both the south and the west; (5) a classroom at exactly one level below the laboratory (on the 6th floor based on the US-style or the 5th floor based on the UK-style and is not shown in Figure 1, and (6) another classroom at exactly one level above the laboratory on the 8th floor based on the US-style or the 7th floor based on the UK-style and is also not shown in Figure 1.
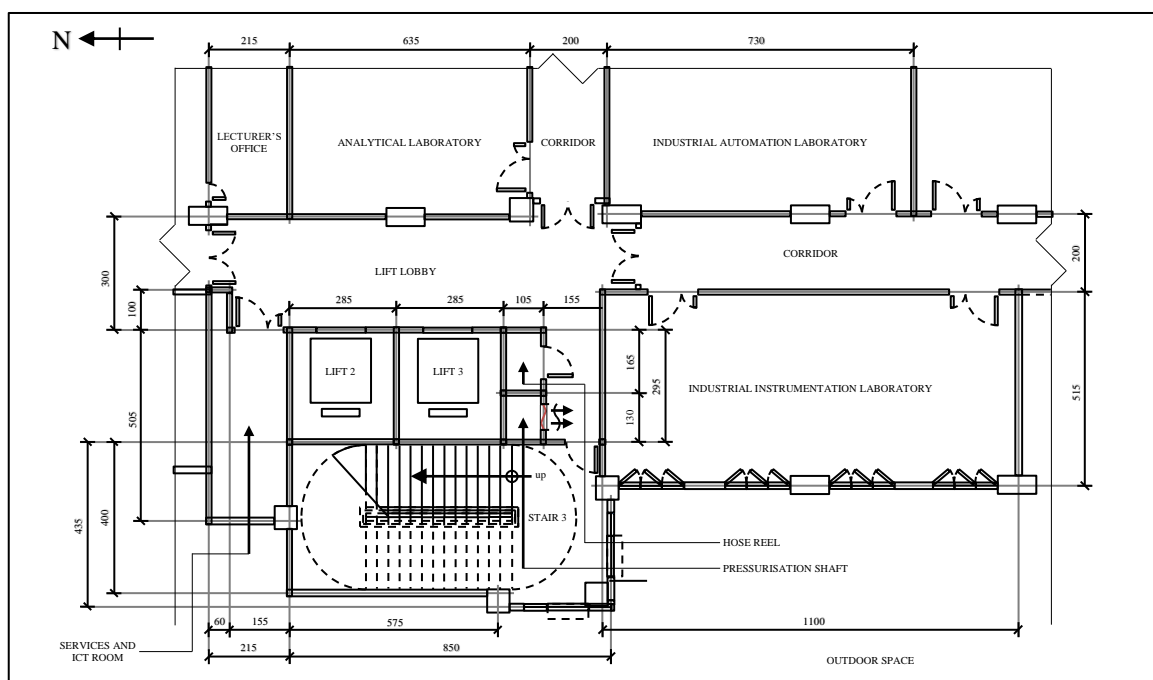

Figure 1. The typical floor plan of the industrial instrumentation laboratory at the 7th floor (based on the US-style) or the 6th floor (based on the UK-style) of the MJIIT building

2.3. The data recording devices

Various types of data are recorded at the Industrial Instrumentation Laboratory and its surrounding spaces by using various types of recording devices. The recorded data is divided into two categories, which are the indoor data and outdoor data. The indoor data is recorded using home-made low-cost data loggers using off-the-shelf sensors connected to single-board microcontrollers (such as Arduino) or single-board computers (such as Raspberry Pi). There are two types of data recorded by these indoor data loggers by the time this research is done, which are the indoor air temperature and the indoor relative humidity. These data loggers are installed in the Industrial Instrumentation Laboratory and the surrounding indoor spaces as mentioned in Subsection 2.2 (The Research Location), which are: (1) the laboratory itself; (2) the staircase in the north-west; (3) the lift lobby in the north; (3) the corridor in the east; (4) the classroom at exactly one level below the laboratory; and (5) the other classroom at exactly one level above the laboratory. Meanwhile, the outdoor data is obtained from the readily available weather station belongs to the Wind Engineering for (Urban, Artificial, Man-made) Environment Laboratory, one of the laboratories in MJIIT. The weather station is located at the rooftop of the MJIIT building. There are various types of data recorded by the weather station, including: (1) the outdoor air temperature; (2) the outdoor relative humidity; (3) the wind speed and direction; and (4) the global solar radiation. However, only the indoor and outdoor data related to air temperature and solar radiation is used in this research. The types of solar radiation that are used as the

inputs for this research are direct solar radiation and diffuse solar radiation, and their values are obtained by splitting the value of actual recorded global solar radiation into the values of direct solar radiation and diffuse solar radiation through calculation.

2.4. The data recording period

All the data required for this research is recorded for 11 days, from the 5th of February 2019 until the 15th of February 2019 and this data is recorded at every one-minute interval. The recorded data is then divided into two parts, which are: (1) the first group (recorded for the first five days from the 5th of February 2019 until the 9th of February 2019); and (2) the second group (recorded for the remaining six days from the 10th of February 2019 until the 15th of February 2019). It is mentioned earlier in Section 1 (Introduction) that the black box model cannot extrapolate very well when it is simulated beyond the range of the training data set [1]. Based on the recorded data, it is observed that the output data for the second group has a wider maximum-minimum range and overlaps the maximum-minimum range of the output data for the first group. Therefore, it is decided in this research that the second group of the data is assigned as the training data set while the first group of the data is assigned as the testing data set. It is also decided in this research that the duration of the training data set is longer than the testing data set – this is the reason why the number of days for the training data set is one day more than the testing data set (six days versus five days).

2.5. The inputs-output data selection

The types of input that are considered able to affect the indoor air temperature behaviour of the Industrial Instrumentation Laboratory are listed, which are: (1) the past and present indoor air temperature of the laboratory itself, $T_{IndInsLab}$; (2) the past and present air temperature difference between the staircase in the north-west and the laboratory itself, $\Delta T_{Stair-IndInsLab}$; (3) the past and present air temperature difference between the lift lobby in the north and the laboratory itself, $\Delta T_{LiftLobby-IndInsLab}$; (4) the past and present air temperature difference between the corridor in the east and the laboratory itself, $\Delta T_{Corridor-IndInsLab}$; (5) the past and present air temperature difference between the outdoor space in both the south and the west and the laboratory itself, $\Delta T_{Outdoor-IndInsLab}$; (6) the past and present air temperature difference between the classroom at exactly one level below the laboratory and the laboratory itself, $\Delta T_{LowerClass-IndInsLab}$; (7) the past and present air temperature difference between the classroom at exactly one level above the laboratory and the laboratory itself, $\Delta T_{UpperClass-IndInsLab}$; (8) the past and present direct solar radiation that lands on the southern outer-wall of the laboratory, $I_{D\_OuterSouthWall}$; (9) the past and present direct solar radiation that lands on the western outer-wall of the laboratory, $I_{D\_OuterWestWall}$; and (10) the past and present diffuse solar radiation that lands on both the southern and western outer-walls of the laboratory, $I_{DS\_OuterWalls}$. Meanwhile, only one output is considered for the model in this research, which is the future indoor air temperature of the laboratory itself, $T_{IndInsLab}$. The listed inputs and output lead to the construction of the multiple-input and single-output (MISO) model, which is depicted in Figure 2.
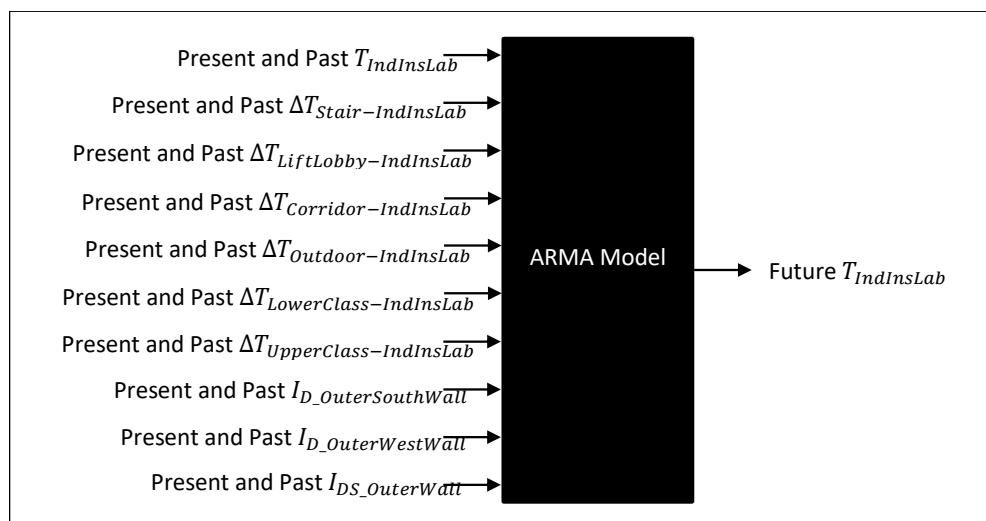

Figure 2. The ARMA model describing the air temperature of the industrial instrumentation laboratory

2.6. The black box model construction

The general ARMA model was described by Peter Whittle in his thesis in 1951 [24] and was popularised by the book written by George Box and Gwilym Jenkins in 1970 [25]. The general equation of the ARMA model is shown in (1):

$$Y[n] = c + \varepsilon_k + \sum_{i=1}^{p} \varphi_i Y[n-i] + \sum_{i=1}^{q} \theta_i X[n-i], \tag{1}$$

where
$Y[n]$ is the output variable,
$X[n]$ is the input variable,
p is the number of past output(s),
q is the number of past input(s),
φ is the parameter for the output variable,
θ is the parameter for the input variable,
c is a constant,
ε_k is the white noise random variable.

From (1) it can be seen that: (1) the model is for a single-input and single-output (SISO) system with one type of input and one type of output; (2) the numbers of past input(s) and past output(s) are different, which are p and q; and (3) there are two more additional constant and variable, c and ε_k. Some modifications are done with (1) to fit the requirement of this research: (1) the number of types of input is increased from one type of input to l types of input since the system that is going to be modelled in this research is a multiple-input and single-output (MISO) system with l types of input and one type of output; (2) the different number of the past input(s) and past output(s) of the model, p and q are standardised as k to simplify the model's regression algorithm; and (3) the constant c and the white noise random variable ε_k are assumed to be zero and omitted, also to maintain the simplicity of the model's regression algorithm. Hence, the modified (2):

$$Y[n] = \sum_{i=1}^{k} \varphi_i Y[n-i] + \sum_{j=1}^{l}\left(\sum_{i=1}^{k} \theta_{j_i} X_j[n-i]\right), \tag{2}$$

where
$Y[n]$ is the output variable,
$X[n]$ is the input variable,
l is the number of types of inputs,
k is the number of past input(s) and past output(s),
φ is the parameter for the output variable,
θ is the parameter for the input variable.

Based on (2) and the inputs-output data summarised in Figure 2, the ARMA model with k past input(s) and past output(s) representing the dynamic indoor air temperature behaviour of the Industrial Instrumentation Laboratory is written as (3):

$$\begin{aligned}
T_{IndInsLab}[n] &= \sum_{i=1}^{k} A_i T_{IndInsLab}[n-i] + \sum_{i=1}^{k} B_i \Delta T_{Stair-IndInsLab}[n-i] \\
&+ \sum_{i=1}^{k} C_i \Delta T_{LiftLobby-IndInsLab}[n-i] + \sum_{i=1}^{k} D_i \Delta T_{Corridor-IndInsLab}[n-i] \\
&+ \sum_{i=1}^{k} E_i \Delta T_{Outdoor-IndInsLab}[n-i] + \sum_{i=1}^{k} F_i \Delta T_{LowerClass-IndInsLab}[n-i] \\
&+ \sum_{i=1}^{k} G_i \Delta T_{UpperClass-IndInsLab}[n-i] + \sum_{i=1}^{k} H_i I_{D_{SouthWall}}[n-i] \\
&+ \sum_{i=1}^{k} I_i I_{D\_WestWall}[n-i] + \sum_{i=1}^{k} J_i I_{DS\_Walls}[n-i]
\end{aligned} \tag{3}$$

As shown in (3) can be written in expanded form as(4):

$$\begin{aligned}
T_{IndInsLab}[n] &= A_1 T_{IndInsLab}[n-1] + A_2 T_{IndInsLab}[n-2] + \cdots + A_{k-1} T_{IndInsLab}[n-k+1] \\
&+ A_k T_{IndInsLab}[n-k] + B_1 \Delta T_{Stair-IndInsLab}[n-1] + B_2 \Delta T_{Stair-IndInsLab}[n-2] \\
&+ \cdots + B_{k-1} \Delta T_{Stair-IndInsLab}[n-k+1] + B_k \Delta T_{Stair-IndInsLab}[n-k] + \cdots \\
&+ J_1 I_{DS_{Walls}}[n-1] + J_2 I_{DS_{Walls}}[n-2] + \cdots + J_{k-1} I_{DS_{Walls}}[n-k+1] \\
&+ J_k I_{DS\_Walls}[n-k]
\end{aligned} \tag{4}$$

2.7. The black box model regression

'Regression' for the statistical model (or 'training' for the artificial neural networks) is the process of estimating the unknown parameter(s) in the model (or neural networks) based on the training data set. First

of all, the training data set for this research has to be arranged in the inputs-output matrix. If the training data set is recorded from $t = 1$ until $t = r$, this means that the number of available sampled data is r. In this research, the value of r is equal to the number of times the data is sampled every minute from the 10th of February 2019 until the 15th February 2019, which is 8640. When the data from k previous step(s) [can also be called as k sampling interval(s)] is used to regress the dynamic indoor air temperature of the Industrial Instrumentation Laboratory from $t = 1$ until $t = k$, the number of input-output pairs are equal to $r - k$. This can be shown in the matrices in (5):

$$\mathbb{Y} = \mathbb{X}\mathbb{M}, \tag{5}$$

where

$$\mathbb{Y} = \begin{bmatrix} T_{IndInsLab}[1 + k] \\ T_{IndInsLab}[2 + k] \\ \vdots \\ T_{IndInsLab}[r - 1] \\ T_{IndInsLab}[r] \end{bmatrix}, \mathbb{X} = \begin{bmatrix} \mathbb{X}_{(1,:)} \\ \mathbb{X}_{(2,:)} \\ \vdots \\ \mathbb{X}_{(r-k-1,:)} \\ \mathbb{X}_{(r-k,:)} \end{bmatrix},$$

$$\mathbb{X}_{(1,:)} = \begin{bmatrix} T_{IndInsLab}[k] \\ T_{IndInsLab}[k-1] \\ \vdots \\ T_{IndInsLab}[2] \\ T_{IndInsLab}[1] \\ \Delta T_{Stair-IndInsLab}[k] \\ \Delta T_{Stair-IndInsLab}[k-1] \\ \vdots \\ \Delta T_{Stair-IndInsLab}[2] \\ \Delta T_{Stair-IndInsLab}[1] \\ \vdots \\ I_{DS\_Walls}[k] \\ I_{DS\_Walls}[k-1] \\ \vdots \\ I_{DS\_Walls}[2] \\ I_{DS\_Walls}[1] \end{bmatrix}^T, \mathbb{X}_{(2,:)} = \begin{bmatrix} T_{IndInsLab}[k+1] \\ T_{IndInsLab}[k] \\ \vdots \\ T_{IndInsLab}[3] \\ T_{IndInsLab}[2] \\ \Delta T_{Stair-IndInsLab}[k+1] \\ \Delta T_{Stair-IndInsLab}[k] \\ \vdots \\ \Delta T_{Stair-IndInsLab}[3] \\ \Delta T_{Stair-IndInsLab}[2] \\ \vdots \\ I_{DS\_Walls}[k+1] \\ I_{DS\_Walls}[k] \\ \vdots \\ I_{DS\_Walls}[3] \\ I_{DS\_Walls}[2] \end{bmatrix}^T,$$

$$\mathbb{X}_{(r-k-1,:)} = \begin{bmatrix} T_{IndInsLab}[r-2] \\ T_{IndInsLab}[r-3] \\ \vdots \\ T_{IndInsLab}[r-k] \\ T_{IndInsLab}[r-k-1] \\ \Delta T_{Stair-IndInsLab}[r-2] \\ \Delta T_{Stair-IndInsLab}[r-3] \\ \vdots \\ \Delta T_{Stair-IndInsLab}[r-k] \\ \Delta T_{Stair-IndInsLab}[r-k-1] \\ \vdots \\ I_{DS\_Walls}[r-2] \\ I_{DS\_Walls}[r-3] \\ \vdots \\ I_{DS\_Walls}[r-k] \\ I_{DS\_Walls}[r-k-1] \end{bmatrix}^T, \mathbb{X}_{(r-k,:)} = \begin{bmatrix} T_{IndInsLab}[r-1] \\ T_{IndInsLab}[r-2] \\ \vdots \\ T_{IndInsLab}[r-k+1] \\ T_{IndInsLab}[r-k] \\ \Delta T_{Stair-IndInsLab}[r-1] \\ \Delta T_{Stair-IndInsLab}[r-2] \\ \vdots \\ \Delta T_{Stair-IndInsLab}[r-k+1] \\ \Delta T_{Stair-IndInsLab}[r-k] \\ \vdots \\ I_{DS\_Walls}[r-1] \\ I_{DS\_Walls}[r-2] \\ \vdots \\ I_{DS\_Walls}[r-k+1] \\ I_{DS\_Walls}[r-k] \end{bmatrix}^T,$$

and $\mathbb{M} = [A_k \quad A_{k-1} \quad \cdots \quad A_2 \quad A_1 \quad B_k \quad B_{k-1} \quad \cdots \quad B_2 \quad B_1 \quad \cdots \quad J_k \quad J_{k-1} \quad \cdots \quad J_2 \quad J_1]^T$.

From (1) until (5) we can see that the ARMA model is a type of linear equation. The unknown parameters for the ARMA model developed in this research are located in the matrix $\mathbb{M}$ in (5). The values of these unknown parameters are estimated by using the least-squares estimation approach (also called 'linear

regression'), which is widely used for linear model parameters estimation [1]. This approach generates the values of the parameters by minimising the sum of squared errors between the output of the actual data and the output produced by the linear models [1]. The least-squares estimation approach formula used to estimate the contents of matrix $\mathbb{M}$ is written as (6):

$$\mathbb{M} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y} \tag{6}$$

After the values of the required parameters are estimated, the regressed ARMA model is then simulated using the same training data set to check its ability to fit the data set. Then, the model is tested and optimised using testing data set, which is a new data set that has never been 'seen' by the estimated ARMA model (has never been used for estimation purpose) and will be described in the Subsection 2.8 (The Black Box Model Testing).

2.8. The black box model testing

In this process, the estimated ARMA model is simulated with a new set of data that is not used during the training process for parameter estimation, which is the testing data set. The purpose of this process is to ensure that the estimated model can still perform as accurate as possible when the model is simulated with the input data that is different than the training data set. If the testing data set is recorded from $t = 1$ until $t = s$, this means that the number of available sampled data is s. In this research, the value of s is equal to the number of times the data is sampled every minute from the 5th of February 2019 until the 9th of February 2019, which is 7200. When the data from k previous step(s) [(can also be called as k sampling interval(s)] is used to test the dynamic indoor air temperature of the Industrial Instrumentation Laboratory from $t = 1$ until $t = k$, the number of input-output pairs are equal to $s - k$.

2.9. The black box model optimisation

The only parameter that can be adjusted in the ARMA model for optimisation is the number of past input(s) and past output(s), which is the value of k in (3), (4), and (5). The bigger the value of k, the more the quantity of constant(s) available in the matrix $\mathbb{M}$ in (3) (and vice versa). Instead of assigning the value of k randomly and manually using trial and error method to search for the best value of k based on the data recorded in this research, a MATLAB® script is written to try the possible values of k one by one within the predefined minimum k value to the predefined maximum k value. Due to time constraint, the value of k in this research is tried one by one only from $k = 1$ until $k = 50$ – the calculation of $\%Fit$ begins at $k = 1$, then the k value is added by 1 and the calculation of $\%Fit$ is done again repetitively until $k = 50$. Even though the percentage of fitness, $\%Fit$ is calculated for each tested value of k, only the identified k value that gives the best (highest) value of $\%Fit$ is retained in this process. The formula to calculate $\%Fit$ are shown in the (7):

$$\%Fit = \left[1 - \frac{norm(\hat{T}_{IndInsLab} - T_{IndInsLab})}{norm[T_{IndInsLab} - mean(T_{IndInsLab})]}\right], \tag{7}$$

where
$\hat{T}_{IndInsLab}$ is the output calculated using the optimised ARMA model,
$T_{IndInsLab}$ is the actual recorded output data.

3. RESULTS AND ANALYSIS

The best-obtained result when the value of k is tested one by one from $k = 1$ until $k = 50$ is when $k = 41$. The output of the optimised ARMA model when $k = 41$ is plotted and displayed in Figure 3 together with the actual output data for comparison purpose. Meanwhile, the percentage of fitness, $\%Fit$ value for the output of the optimised ARMA model when $k = 41$ during simulation is presented in Table 1.

Based on the obtained result, the optimised ARMA model proposed in this research can fit the actual data sufficiently. This is supported by the result presented in Figure 3 and Table 1. However, the level of accuracy obtained in this research can still be improved. For future work, it is suggested to try modelling the same data with different types of black box model, both linear and non-linear models to investigate their performance fitting the same data. In addition, it is also suggested to try the grey box or even the white box modelling to gain the physical insight into the dynamic indoor air temperature behaviour of the Industrial Instrumentation Laboratory. Furthermore, it is also suggested to upgrade the developed models with controllable inputs from the thermal comfort devices such as the air conditioner, ventilation fan, and motor-operated window etc. so that the obtained model can be used for testing any thermal comfort control

algorithm or strategy for the thermal comfort devices in the future. In addition, this upgraded models with controllable inputs can also be used to develop and optimise predictive controllers such as model predictive controller (MPC).
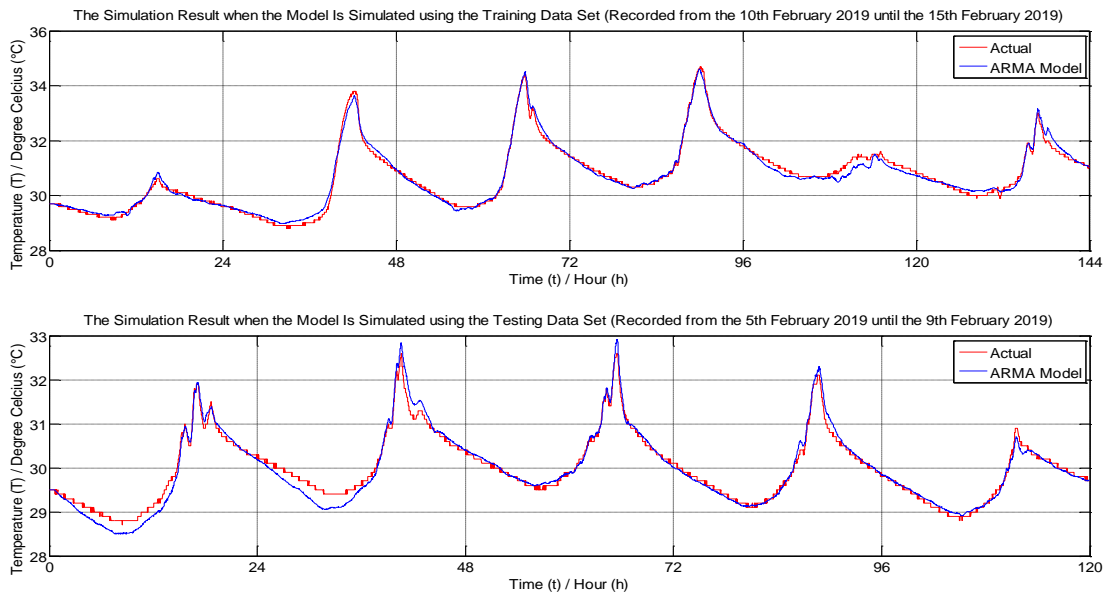

Figure 3. The simulation result for the optimised ARMA model when $k = 41$ during the simulation using the training data set (top) and testing data set (bottom)

Table 1. The percentage of fitting, $\%Fit$ for the optimised ARMA model when $k = 41$ during the simulation using the training data set and testing data set

Data	Percentage of Fitting, $\%Fit$ (%)
Training Data Set	87.45
Testing Data Set	80.36

4.　CONCLUSION

The objective of this research is to construct a mathematical model describing the dynamic indoor air temperature behaviour of Industrial Instrumentation Laboratory by using autoregressive–moving-average (ARMA) model, a type of linear black box model. From the result of this research, it is shown that a data-driven black box model developed with suitable inputs-output combination and minimal physical knowledge regarding the plant prior to modelling is still capable of producing output that is almost similar with the actual output from the plant. The contribution of this research is the development of a mathematical model for system simulation in a limited time frame with limited physical knowledge of the system.

REFERENCES
[1]　D. E. Seborg, D. A. Mellichamp, T. F. Edgar, and F. J. Doyle, *Process Dynamics and Control*. John Wiley & Sons Incorporated, 2010.
[2]　N. S. Nise, *Control Systems Engineering*, 6th ed. John Wiley & Sons Incorporated, 1995.

[3] H. Nguyen, Y. Makino, A. O. Lim, Y. Tan, and Y. Shinoda, "Building High-Accuracy Thermal Simulation for Evaluation of Thermal Comfort in Real Houses," in *Inclusive Society: Health and Wellbeing in the Community, and Care at Home SE - 20*, vol. 7910, J. Biswas, H. Kobayashi, L. Wong, B. Abdulrazak, and M. Mokhtari, Eds. Springer Berlin Heidelberg, pp. 159-166, 2013.

[4] S. F. M. Hussein, H. Nguyen, S. S. Abdullah, Y. Lim, and Y. Tan, "Black box modelling the thermal behaviour of ihouse using auto regressive and moving average (ARMA) model," *J. Teknol.*, vol. 78, no. 6-13, 2016.

[5] S. F. M. Hussein *et al.*, "Simplifying the auto regressive and moving average (ARMA) model representing the dynamic thermal behaviour of ihouse based on theoretical knowledge," *Asian Simulation Conference (AsiaSim 2017)*, pp. 697-711, 2017.

[6] S. H. Nguyễn and Y. Tan, "Simulation-based Short-term Model Predictive Control for HVAC Systems of Residential Houses," *VNU J. Sci. Comput. Sci. Commun. Eng.*, vol. 35, no. 1, pp. 10-22, 2019.

[7] E. Sian, M. Yoshiki, F. Yuan, L. Yuto, and T. Yasuo, "Study of predictive thermal comfort control for cyber-physical home systems," in *2018 13th System of Systems Engineering Conference, SoSE 2018*, pp. 444-451, 2018.

[8] S. E. Ooi, Y. Fang, Y. Lim, and Y. Tan, "Study of adaptive model predictive control for cyber-physical home systems," in *Lecture Notes in Electrical Engineering*, vol. 481, pp. 165-174, 2019.

[9] L. Yuto, O. S. En, M. Yoshiki, T. T. Kin, R. Alfred, and T. Yasuo, "Implementation of energy efficient thermal comfort control for cyber-physical home systems," *Adv. Sci. Lett.*, vol. 23, no. 11, pp. 11530-11534, 2017.

[10] P. Radecki and B. Hencey, "Online Model Estimation for Predictive Thermal Control of Buildings," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 4, pp. 1414-1422, 2017.

[11] G. Mustafaraj, G. Lowry, and J. Chen, "Prediction of Room Temperature and Relative Humidity by Autoregressive Linear and Nonlinear Neural Network Models for an Open Office," *Energy Build.*, vol. 43, no. 6, pp. 1452-1460, 2011.

[12] G. Mustafaraj, J. Chen, and G. Lowry, "Thermal Behaviour Prediction Utilizing Artificial Neural Networks for an Open Office," *Appl. Math. Model.*, vol. 34, no. 11, pp. 3216-3230, 2010.

[13] A. Afram and F. Janabi-Sharifi, "Gray-box modeling and validation of residential HVAC system for control system design," *Appl. Energy*, vol. 137, pp. 134-150, 2015.

[14] A. Afram and F. Janabi-Sharifi, "Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods," *Energy Build.*, vol. 94, pp. 121-149, 2015.

[15] A. Afram, A. S. Fung, F. Janabi-Sharifi, and K. Raahemifar, "Development and performance comparison of low-order black-box models for a residential HVAC system," *J. Build. Eng.*, vol. 15, pp. 137-155, 2018.

[16] A. Afram and F. Janabi-Sharifi, "Supervisory model predictive controller (MPC) for residential HVAC systems: Implementation and experimentation on archetype sustainable house in Toronto," *Energy Build.*, vol. 154, pp. 268-282, 2017.

[17] D. Sturzenegger, D. Gyalistras, M. Morari, and R. S. Smith, "Semi-automated modular modeling of buildings for model predictive control," *BuildSys 2012 - Proc. 4th ACM Work. Embed. Syst. Energy Effic. Build.*, pp. 99-106, 2012.

[18] D. Sturzenegger, D. Gyalistras, M. Morari, and R. S. Smith, "Model Predictive Climate Control of a Swiss Office Building: Implementation, Results, and Cost-Benefit Analysis," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 1, pp. 1-12, 2016.

[19] A. A. Safa, A. S. Fung, and R. Kumar, "Performance of two-stage variable capacity air source heat pump: Field performance results and TRNSYS simulation," *Energy Build.*, vol. 94, pp. 80-90, 2015.

[20] N. Alibabaei, A. S. Fung, and K. Raahemifar, "Development of Matlab-TRNSYS co-simulator for applying predictive strategy planning models on residential house HVAC system," *Energy Build.*, vol. 128, pp. 81-98, 2016.

[21] R. De Coninck, F. Magnusson, J. Åkesson, and L. Helsen, "Toolbox for development and validation of grey-box building models for forecasting and control," *J. Build. Perform. Simul.*, vol. 9, no. 3, pp. 288-303, 2016.

[22] R. De Coninck and L. Helsen, "Practical implementation and evaluation of model predictive control for an office building in Brussels," *Energy Build.*, vol. 111, pp. 290-298, 2016.

[23] I. Hazyuk, C. Ghiaus, and D. Penhouet, "Optimal temperature control of intermittently heated buildings using Model Predictive Control: Part II – Control algorithm," *Build. Environ.*, vol. 51, pp. 388-394, 2012.

[24] P. Whitle, *Hypothesis Testing in Time Series Analysis*. Almqvist & Wiksells, 1951.

[25] G. E. P. Box and G. M. Jenkins, *Time series analysis: forecasting and control*. Holden-Day, 1970.