Black Box Modelling and Simulating the Dynamic Indoor Relative Humidity of a Laboratory Using Autoregressive–moving-average (ARMA) Model

S F M Hussein¹, M K Mohd Fitri Alif², A O Al Rabeei¹, A Faruq¹,²,³, S M Zulkapli¹, M S Noorazizi², S A Zaki² and S S Abdullah¹,⁶

¹Biologically Inspired System and Technology (Bio-iST) iKohza, Department of Electronic Systems Engineering (ESE), Malaysia-Japan International Institute of Technology (MJIIT), UniversitiTeknologi Malaysia (UTM) Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak), 54100 Kuala Lumpur, MALAYSIA.
²Department of Control and Mechatronics Engineering (CMED), School of Electrical Engineering, Faculty of Engineering, UniversitiTeknologi Malaysia (UTM) Johor Bahru, 81310 Johor DarulTakzim, MALAYSIA.
³Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang. Jl. Raya Tlogomas 246, 65144 Malang, INDONESIA.
⁴Engineering Department, Razak Faculty of Technology and Informatics, UniversitiTeknologi Malaysia (UTM) Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak), 54100 Kuala Lumpur, MALAYSIA.
⁵Wind Engineering for (Urban, Artificial, Man-made) Environment Laboratory, Department of Mechanical Precision Engineering (MPE), Malaysia-Japan International Institute of Technology (MJIIT), UniversitiTeknologi Malaysia (UTM) Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak), 54100 Kuala Lumpur, MALAYSIA.

E-mail: shahrum@utm.my

Abstract. Modelling and simulation of the dynamic indoor relative humidity behaviour of a building is essential to test any proposed thermal comfort controller and strategy in the building. Like other plants, the dynamic indoor relative humidity behaviour of a building can be developed based on the white box model, black box model, and grey box model. This research focuses on the usage of autoregressive–moving-average (ARMA) model, a type of black box model to represent the dynamic indoor relative humidity behaviour of Industrial Instrumentation Laboratory at Malaysia-Japan International Institute of Technology (MJIIT), UniversitiTeknologi Malaysia (UTM) Kuala Lumpur and uses the real recorded data from the laboratory and minimal knowledge regarding the physical characteristics of the humidity behaviour in the laboratory. The performance of the ARMA model developed in this research is compared with the real recorded data from the laboratory. Result obtained shows that the ARMA model is enough for modelling and simulating the dynamic indoor relative humidity behaviour of the laboratory

⁶ To whom any correspondence should be addressed.
1. Introduction
The indoor hygrothermal (moisture and heat) condition of a building is one of the essential elements to ensure the comfort, health, and productivity of the occupants of the building. The hygrothermal condition can be maintained within a comfortable range by using thermal comfort devices such as the air conditioners, fans, windows, etc. Various types of control algorithms and strategies can be proposed to optimise the performance of thermal comfort devices. These proposed control algorithms and strategies can be tested on both the real plant (the building itself) or via the computer simulation based on the mathematical model describing the hygrothermal behaviour of the building. Testing the proposed controllers and strategies on the real building is costly and time-consuming, so it is better to do the testing (at least the initial stage) via the computer simulation. Therefore, it is important to develop the mathematical model describing the hygrothermal behaviour of a building for the testing purpose via computer simulation.

Previously, models describing the air temperature behaviour of one of the rooms in iHouse, a smart house testbed belongs to Japan Advanced Institute of Science and Technology (JAIST) in Ishikawa, Japan have been developed by the main author of this research by using a linear black box model in[1] and a linear grey box model in[2] – results show that these trained and optimised models can simulate the new data (that is not used during the training process) with high level of accuracies. Since a number of models have been developed by the main author of this research to describe the dynamic indoor air temperature behaviour of a room, now it is the time for the main author to develop the model describing the dynamic indoor relative humidity behaviour of a room so that the author will have a complete hygrothermal model of a building for simulating any proposed thermal comfort control algorithm and strategy in the future.

Just like the mathematical models describing the physical behaviour of other plants, the humidity behaviour of a building can also be modelled by using three types of models, which are: (1) the white box model, also called the theoretical model; (2) the black box model, also called the empirical model; and (3) the grey box model, also called the semi-empirical model. The physical knowledge regarding the behaviour of the plant is required to develop the mathematical equation(s) for the white box model[3] – this physical knowledge is also referred to as the fundamental knowledge of science and engineering[4]. Meanwhile, minimal physical knowledge regarding the plant’s behaviour is required while developing the black box model (but more plant’s knowledge will give more advantage) because the black box model does not have any equation that describes the physical characteristic of the plant, but only consists of a set of linear or non-linear ‘off-the-shelf’ equation(s) with the parameter(s) that is(are) tuned using the input(s)-output(s) data of the plant until the output(s) of the black box model becomes almost similar (as accurately as possible) with the output(s) of the plant. The grey box model is the combination of the white box and black box model – the model is developed using the theoretical knowledge of the plant, but the unknown parameter(s) in the model is(are) tuned based on the real input(s)-output(s) data from the plant.

Every type of model has its own advantage(s) and disadvantage(s). The white box model can be simulated over a broader range of operating condition, but its development is time-consuming and costly for the plant with complicated physical characteristic(s)[3]. In addition, there are some situations where it is difficult to obtain or measure the unknown parameter value(s) available in the white box model[3]. The construction of the black box model is simpler because it requires minimal physical knowledge of the plant, but the model does not extrapolate well beyond the range of the training data set that is used to estimate the model[3]. Usually, the available training data set does not cover the plant’s whole operating condition, so caution must be taken if there is any requirement to simulate the developed model beyond the range of the training data set[3]. However, the black box model is popular in the industry [3]. Meanwhile, the mathematical equation(s) in the grey box model provide(s) physical meaning like the white box model because the grey box model is also developed based on fundamental knowledge of science and engineering, but the unknown parameters(s) in the grey box model is(are) estimated based on the input(s)-output(s) data recorded from the plant exactly like the black box model. This means that the grey box model can be built faster and easier than the
white box model but can be simulated over a broader range of operating condition than the black box model[3].

Various researchers had developed various types of mathematical models describing the building’s dynamic indoor relative humidity behaviour for various purposes. Wessberg et al.[5] built a grey box model to simulate the indoor relative humidity (and air temperature) in massive historic buildings (which are three different 13th-century churches) to identify their suitable heating power distribution and heat-up time that heat up the indoor air temperature while minimising the relative humidity fluctuation to prevent damage to the materials of those historical buildings. Kramer et al.[6] presented a grey box model in the simple state-space form to simulate the hygrothermal behaviour of monumental building (three castles and one cathedral) – this simple grey box model was proposed to solve the following three problems: (1) the white box model is more tedious to be developed; (2) it is difficult to obtain the parameters for the white box model representing a complex system; and (3) the white box model requires higher computation resource during the simulation. Tao Lu and Martti Viljanen [7] developed neural networks (NN) to predict the indoor relative humidity (and air temperature) of their test house with complex physical characteristic and difficult for white box simulation – two non-linear black box models were used, the non-linear autoregressive with external input (NNARX) model and genetic algorithm, and both models simulated the indoor air temperature and relative humidity of the test house with high level of accuracies. Simon Harasty et al.[8] modelled the indoor climate (air temperature and relative humidity) of a room at Schloss Fasanerie castle, Germany using an artificial neural networks (ANN) and used the model for two types of predictive controllers, the model predictive controller (MPC) and neuro controller – simulation results showed that both the MPC and neuro controller maintained the indoor climates with fewer fluctuations compared with the bang-bang controller, but the neuro controller generalised better and required lower computational power compared to MPC.

This research focuses on constructing the model describing the dynamic indoor relative humidity behaviour of Industrial Instrumentation Laboratory at Malaysia-Japan International Institute of Technology (MJIIT), UniversitiTeknologi Malaysia (UTM) Kuala Lumpur. Since this is the first time the relative humidity behaviour of this laboratory is going to be modelled and not enough physical knowledge regarding the relative humidity behaviour of the laboratory is known at the moment, only the autoregressive–moving-average (ARMA) model, one of the simple linear black box models is used in this research.

2. Methodology

2.1. Scope of research
First, the Industrial Instrumentation Laboratory at Malaysia-Japan International Institute of Technology (MJIIT), UniversitiTeknologi Malaysia (UTM) Kuala Lumpur is selected as the plant to be modelled in this research and will be described in more detail in Subsection 2.2 (The research location).

Second, 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 at the moment 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.

Third, the number of past input(s) for each type of input used by the ARMA model in this research is standardized to simplify the parameters estimation algorithm and shorten the model’s development time in this research.

2.2. The research location
The Industrial Instrumentation Laboratory at Malaysia-Japan International Institute of Technology (MJIIT), UniversitiTeknologi Malaysia (UTM) Kuala Lumpur is selected as the plant to be modelled in this research. The laboratory is located at 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 laboratory is shown in Figure 1. Based on this figure, it is shown that the 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; and (4) the outdoor space in both the south and the west.

![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.](image)

There are wall-mounted ventilation fans installed on the northern and southern wall of the laboratory (one fan on each wall). The shutters of these fans are opened automatically when the fans are spinning and closed automatically when the fans are not spinning. When the shutters are closed, there are small but visible openings between the shutters. There are also two sets of doors as the entry points from the corridor in the east of the laboratory through the eastern wall. The bottoms of these doors have air gaps with the height of around one centimetre (measured between the doors and the floor). Only the western wall of the laboratory has windows, and the area of the windows consists of 61.34% of the western wall. Some of the windows have bent frames, which also cause small but visible openings between the windows and frames. These small openings at the shutter of the wall-mounted fans, beneath the doors, and between the windows and their bent frames are identified as infiltration points which may affect the dynamic indoor relative humidity of the laboratory and are considered as some of the inputs for the ARMA model developed in this research – these inputs will be described in Subsection 2.5 (The inputs-output data selection).

### 2.3. The data recording devices

The recorded data around the Industrial Instrumentation Laboratory is divided into two categories: (1) the indoor data; and (2) the outdoor data. The indoor data is recorded using home-made low-cost data loggers consists of the off-the-shelf sensors connected to single-board microcontrollers (such as Arduino) or single-board computers (such as Raspberry Pi). These data loggers are installed in the laboratory itself and the surrounding indoor spaces as mentioned in Subsection 2.2 (The research location) – these spaces are: (1) the laboratory itself; (2) the lift lobby in the north; and (3) the corridor.
in the east. 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 – this weather station is located at the rooftop of the MJIIT building. Various types of data are recorded by both the indoor and outdoor equipment, but only the relative humidity data is used in this research.

2.4. The data recording period
The required data for this research is recorded for 15 days, from the 5th of February 2019 until the 15th of February 2019 – 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 from the 5th of February 2019 until the 9th of February 2019); and (2) the second group (recorded 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 outside the training data set [3]. Based on the recorded data, it is observed that the output data from the second group has a wider maximum-minimum range and overlaps the maximum-minimum range of the output data from the first group. Therefore, it is decided that the second group is assigned as the training data set while the first group is assigned as the testing data set in this research. It is also decided that the duration of the training data set is longer than the testing data set for this research, which is six days versus five days.

2.5. The inputs-output data selection
The types of input that are considered able to affect the indoor relative humidity of the Industrial Instrumentation Laboratory are listed, which are: (1) the past and present indoor relative humidity of the laboratory itself, RHIndInsLab; (2) the past and present relative humidity gain/loss due to the air infiltration through the small openings at the wall-mounted ventilation fans installed on the northern and southern wall of the laboratory, ΔRHVentFans; (3) the past and present relative humidity gain/loss due to the air infiltration through the small openings under the doors available on the eastern wall of the laboratory, ΔRHUnderDoors; and (4) the past and present relative humidity gain/loss due to the air infiltration through the small openings at the bent frames of the windows available at the western wall of the laboratory, ΔRHWindowFrames. Meanwhile, only one output is considered for the model in this research, which is the future indoor relative humidity of the laboratory itself, RHIndInsLab. The listed inputs and output lead to the construction of the multiple-input and single-output (MISO) model, which is represented by Equation (1) in Subsection 2.6 (The black box model construction).

2.6. The black box model construction
The general ARMA model was described by Peter Whittle in his thesis in 1951 [9] and was popularised by the book written by George Box and Gwilym Jenkins in 1970 [10]. The general equation of the ARMA model is written as Equation (1) below:

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

where:
- \( Y[n] \) is the output variable,
- \( X[n] \) is the input variable,
- \( p \) is the maximum number of past output(s),
- \( q \) is the maximum number of past input(s),
- \( \phi \) is the parameter for the output variable,
- \( \theta \) is the parameter for the input variable,
- \( c \) is a constant,
- \( \varepsilon_k \) is the white noise random variable.
From Equation (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 \( \varepsilon_k \). Some modifications are done with Equation (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 numbers of the maximum 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 \( \varepsilon_k \) are assumed to be zero and omitted also to maintain the simplicity of the model’s regression algorithm. Hence, the modified Equation (1) for fulfilling these requirements is rewritten as Equation (2) below:

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

(2)

where:

- \( Y[n] \) is the output variable,
- \( X[n] \) is the input variable,
- \( l \) is the number of types of input,
- \( k \) is the maximum number of past input(s) and past output(s),
- \( \varphi \) is the parameter for the output variable,
- \( \theta \) is the parameter for the input variable.

Based on Equation (2) and the inputs-output data listed in Subsection 2.5 (The inputs-output data selection), the ARMA model with \( k \) past input(s) and \( k \) past output(s) representing the indoor dynamic relative humidity behaviour of the Industrial Instrumentation Laboratory is written as Equation (3) below:

\[
RH_{IndInstLab}[n] = \sum_{i=1}^{k} A_i RH_{IndInstLab}[n - i] + \sum_{i=1}^{k} B_i \Delta RH_{VentFans}[n - i] + \sum_{i=1}^{k} C_i \Delta RH_{UnderDoors}[n - i] + \sum_{i=1}^{k} D_i \Delta RH_{WindowFrames}[n - i]
\]  

(3)

2.7. The black box model estimation and testing

The parameters that are needed to be estimated in the developed ARMA model describing the dynamic indoor relative humidity of the Industrial Instrumentation Laboratory in Equation (3) above are available in the matrices \( A_i, B_i, C_i, \text{and} D_i \). Since the ARMA model is a type of linear model, the values of these parameters are estimated using the least-squares estimation approach (also known as ‘linear regression’) because this approach is widely used to estimate the parameters of the linear models[3]. The least-squares estimation approach estimates the values of the parameters in the matrices \( A_i, B_i, C_i, \text{and} D_i \) by minimising the sum of squares of the errors between the actual data and the output produced by the ARMA model in Equation (3). After the parameters are estimated, the ARMA model is simulated using the same training data set to check its ability to fit the data set. Then, the model is tested and optimised using the testing data set, which is a new data that has never been ‘seen’ by the estimated ARMA model during the regression process and will be elaborated in Subsection 2.8 (The black box model optimization).

2.8. The black box model optimization

The least-squares estimation approach mentioned in the previous Subsection 2.7 (the black box model estimation and testing) produces the same numbers and values of the parameters in the matrices \( A_i, B_i, C_i, \text{and} D_i \) in Equation (3) as long as the training data set and the number of past input(s) and past output(s) are the same. If the training data set is maintained the same during the training process, the only adjustment that can be done for the model’s accuracy adjustment is by changing the number of past input(s) and past output(s), which is the value of \( k \). Different \( k \) values leads to different
numbers and values of the parameters in the matrices $A_i$, $B_i$, $G_i$, and $D_i$ in Equation (3). Instead of assigning the value of $k$ randomly and manually by using the 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 $k = 200$. The percentage of fitness, $\%Fit$ is calculated for each tested value of $k$. The formula to calculate $\%Fit$ is shown in Equation (4) below:

$$\%Fit = \left[1 - \frac{\text{norm}(\overline{RH}_{\text{indinstLab}} - RH_{\text{indinstLab}})}{\text{norm}(RH_{\text{indinstLab}} - \text{mean}(RH_{\text{indinstLab}}))}\right],$$

where:

- $\overline{RH}_{\text{indinstLab}}$ is the output calculated using the optimised ARMA model,
- $RH_{\text{indinstLab}}$ is the actual recorded output data.

### 3. Main Results

The best-obtained result when the value of $k$ is tested one by one from $k = 1$ until $k = 200$ is when $k = 85$. The simulation result for the optimized ARMA model when $k = 85$ is plotted and displayed in Figure 2. Meanwhile, the percentage of fitness, $\%Fit$ value for the optimised ARMA model when $k = 85$ during simulation is presented in Table 1.

![Figure 2](image_url)

**Figure 2.** The simulation result for the optimised ARMA model when $k = 85$ 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 $k = 85$ during the simulation using the training data set and testing data set.

| Data                | Percentage of Fitting, $\%Fit$ (%) |
|---------------------|-----------------------------------|
| Training Data Set   | 79.32                             |
| Testing Data Set    | 70.28                             |
simulated relative humidity value of the Industrial Instrumentation Laboratory at one minute ahead based on the input recorded from 1 hour and 25 minutes before. It is suspected that the relationship between the inputs and the output of the dynamic indoor relative humidity behaviour of the actual laboratory are non-linear while the ARMA model used to represent the dynamic indoor relative humidity behaviour of the laboratory is a linear equation. 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 (in terms of accuracy) and practicality [in terms of the number of past input(s)]. In addition, it is also suggested to try the grey box or even the white box modelling to gain physical insight into the dynamic indoor relative humidity behaviour in the laboratory. Furthermore, it is also suggested to upgrade the models with controllable inputs from the thermal comfort devices such as the air conditioners, ventilation fans, and motor-operated windows so that the obtained model can be used for simulating the performance of the controller and strategy for the thermal comfort devices in the future.

4. Conclusion
The main goal of this study is to develop a model describing the thermal behaviour of Industrial Instrumentation Laboratory, Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur based on the 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 suitable input(s)-output combination can still produce the output that is almost the same as the output from the actual plant. The main contribution from this finding is in the construction of a mathematical model to simulate the system in a short time with limited system's physical knowledge.

Acknowledgments
The authors would like to express their appreciation for the support of the: (1) Wind Engineering for (Urban, Artificial, Man-made) Environment Laboratory for supplying the outdoor data recorded using their weather station installed on the rooftop of Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur; and (2) Biologically Inspired System and Technology (Bio-iST) iKohza for providing the fund to purchase the data recording equipment to record the indoor data and giving access to record the indoor data in the Industrial Instrumentation Laboratory in Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia (UTM) Kuala Lumpur.

References
[1] Mohd Hussein S F, Nguyen H, Abdullah S S, Lim Y and Tan Y 2016 Black Box Modelling the Thermal Behaviour of iHouse Using Auto Regressive and Moving Average (ARMA) Model J. Teknol.78 51–8
[2] Hussein S F M, Bakar M A A, Makino Y, Nguyen H, Abdullah S S, Lim Y and Tan Y 2017 Simplifying the Auto Regressive and Moving Average (ARMA) Model Representing the Dynamic Thermal Behaviour of iHouse Based on Theoretical Knowledge vol 752 pp 697–711
[3] Seborg D E, Mellichamp D A, Edgar T F and Doyle F J 2010 Process Dynamics and Control (John Wiley & Sons Incorporated)
[4] Nise N S 1995 Control Systems Engineering (John Wiley & Sons Incorporated)
[5] Wessberg M, Vyhlidal T and Broström T 2019 A model-based method to control temperature and humidity in intermittently heated massive historic buildings Build. Environ.159 106026
[6] Kramer R, van Schijndel J and Schellen H 2013 Inverse modeling of simplified hygrothermal building models to predict and characterize indoor climates Build. Environ.68 87–99
[7] Lu T and Viljanen M 2009 Prediction of indoor temperature and relative humidity using neural network models: Model comparison Neural Comput. Appl.18 345–57
[8] Harasty S, Böttcher A D and Lambeck S 2019 Using Artificial Neural Networks for Indoor Climate Control in the Field of Preventive Conservation *E3S Web Conf.* 111 04054
[9] Whitle P 1951 Hypothesis Testing in Time Series Analysis (Almqvist & Wiksells)
[10] Box G E P and Jenkins G M 1970 *Time series analysis: forecasting and control* (Holden-Day)