

Original Research Paper

# Developing and Testing Dynamic Models for HVAC Systems Using System Identification Approach

Nabil Nassif, Jasmine Buford and Taher M. Abu-Lebdeh

Department of Civil and Architectural Engineering, North Carolina A&T State University, Greensboro, USA

## Article history

Received: 27-02-2017

Revised: 02-03-2017

Accepted: 14-03-2017

Corresponding Author:

Nabil Nassif

Department of Civil and  
Architectural Engineering,  
North Carolina A&T State  
University, Greensboro, USA  
Email: nnassif@ncat.edu

**Abstract:** This paper proposes integrated modeling and optimization methods for a chilled water HVAC system using a system identification approach. Two multiple input-single output models are developed to find the supply air temperature and fan power of the investigated chilled water air handling unit AHU. To test the proposed models, actual data are collected from existing HVAC systems. Different fan and cooling coil model structures with various time delays and orders are tested to find the optimal model structure in term of normalized root-mean-square deviation or Coefficient Of Variances (COV). The paper also proposes an optimization procedure integrated into system identification model to automate the process of finding optimal fan and cooling coil model structures yielding the best accurate predictions. The testing results show that the proposed methods can provide accurate predictions that can used for several applications such as control optimization, energy assessment and fault detection and diagnosis.

**Keywords:** HVAC Systems, Self-Tuning Models, System Identification, Air Handling Unit, Cooling Coil, Optimization

## Introduction

Heating, ventilation and air conditioning systems are widely used in buildings to provide occupants with conditioned air and acceptable indoor air quality. The chilled water system is one of the most commonly used HVAC systems in the both commercial and industrial buildings. These systems are currently used to provide thermal comfort for a wide array of building types, sizes and in different climates. The design of these systems constitutes a large impact on the energy usage and operating cost of buildings they serve. Buildings stand for a substantial part of the total energy consumption in the United States and with an increase focus on cost reductions and energy savings, it is necessary to use intelligent and energy-saving models (ASHRAE, 2015; EIA). The development of building energy savings methods and models becomes apparently more necessary for a sustainable future. Those models can be integrated into the building automation system to perform many intelligent functions such as building energy assessment, control strategies optimization, fault detection and diagnosis (Kusiak and Xu, 2012; Nassif,

2014; Wang and Jin, 2000; Buford and Nassif, 2016; Nassif, 2008; Hani, 2009). Today a majority of commercial buildings are equipped with BAS that have the ability to collect large amounts of data; however, these buildings still do not operate optimally due to the lack of embedded computational means. Thus, there is a need to develop new modeling technique to improve whole system efficiency. Modeling using system identification is a technique used in many studies (Afram and Janabi-Sharifi, 2015). However, this paper proposes a new integrated approach combining system identification modeling with model structure optimization. Two multiple input-single output models are developed to find the supply air temperature and fan power of a typical chilled water air handling unit AHU a typical VAV system. The models use time-series data that are usually available from any typical building automation system BAS. The system identification approaches (SD) with different modeling techniques and model structures are investigated. The optimization procedure is integrated into the system identification models to find the optimal model parameters yielding the least prediction errors. A

genetic algorithm is used to solve the optimization problem. The GA, as a heuristic approach, is suit for solving complex problems with large search spaces (Goldberg, 1989; Deb, 2001; Mossolly *et al.*, 2009).

## Methodology

The research proposed in this study is conducted using the following methodology (1) data collection and preprocessing, (2) model development, (3) model structure selection, (4) model parameter estimation, (5) model optimization and validation. The data used in this study is collected from an existing building's chilled water system. After a sufficient amount of data has been collected, the data is divided into two training and testing sets. Utilizing the System Identification (SI) process and the collected data, various model structures along with different time delays and orders, are investigated to determine the best structure yielding satisfactory accuracy in terms of Mean Square errors or Deviation (MSD), Root Mean Square errors or Deviation (RMSD) and normalized root-mean-square deviation or Coefficient Of Variances (COV). The selected model structures are tested for optimality by using an exhaustive parameter combination. After each model is tested, the results are then evaluated to determine the most optimal model structure and architecture. The models produced from the parametric study are then evaluated and validated by performing an optimization on the

testing data set. An objective function and decision variables are created along with a set of constraints. Figure 1 illustrates the process to formulate various System Identification (SI) models that can be implemented in the building automation system.

## Data Collection and Preprocessing

An existing building located in Greensboro, NC is selected for this study. The building is 88,000 ft<sup>2</sup>, three-story and multi-use classroom conditioned by typical VAV systems. The HVAC system consists of six air handling units, two for each floor. There are two chillers and two boilers to provide chilled and hot water to those units, respectively. Figure 2 shows the air handling unit including supply and return fans, discharge, recirculation and outside air dampers and cooling and heating coils. The building is equipped with a BAS that collects performance data from those units. In this study, two data-based models are investigated (1) cooling coil model and (2) fan model. As shown in Fig. 3, the cooling coil model inputs are chilled water cooling coil valve position, chilled water supply temperature, mixed air temperature, supply or discharge air flow rate and humidity ratio. The relative ratio is neglected as this is typically unavailable from most existing HVAC systems. The cooling coil model output is the supply air temperature. The fan model inputs are supply air flow and fan speed and model output is fan power.

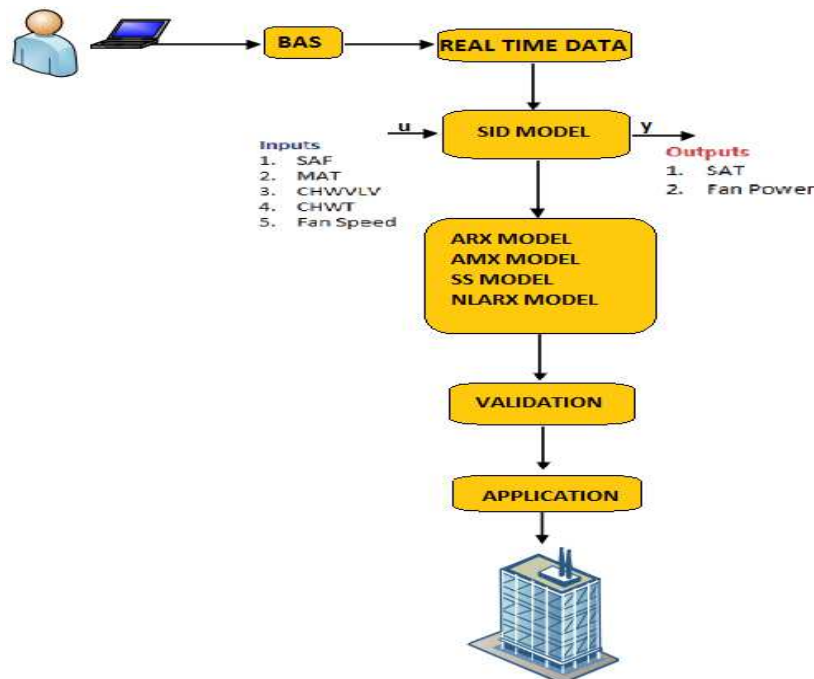


Fig. 1. A schematic diagram of the process used to formulate various System Identification (SI) models and analysis the data collected from BAS

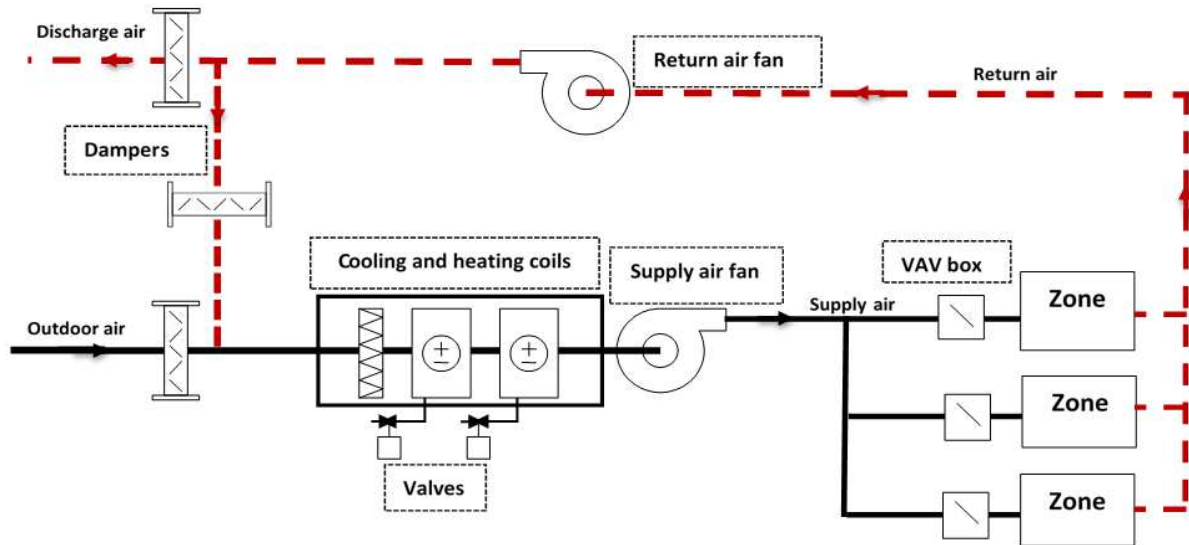


Fig. 2. A schematic of the investigated air handling unit AHU

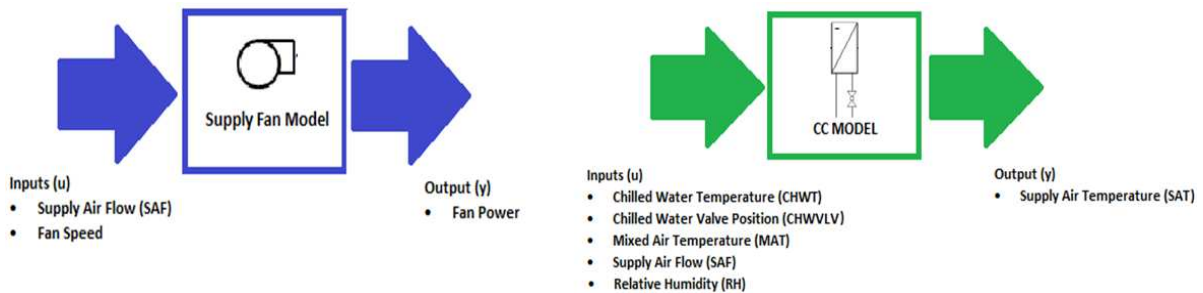


Fig. 3. The inputs and output for the cooling coil and fan models

Through BAS, all necessary input and output model data are trended at 5-minute intervals from October 2014 to February 2015 (a span of 5 months). In total 28,767 sample points were collected from the system, representing around 100 h of system run time. The data recorded from the system covers a temperature range as high as 68°F and as low as 50°F for the supply air temperature. The supply air temperature reached as high as 68°F when the building was no longer occupied due to the time of day. This temperature ranged from 53°F to 60°F during schedule occupancy. It is important to note that the BAS continuously records measurements, even if the system is turned off. These points are removed, because they are repetitive and won't improve the learning capabilities of the identification models. Once all the null data points are removed, the total sample data size is reduced to close to 26,789 points.

### Model Development

Four models are investigated in this study (1) Autoregressive Exogenous model (ARX), (2) Autoregressive Moving Average Exogenous model

(ARMAX), (3) State Space model (SS) and (4) Nonlinear Autoregressive Exogenous model (NLARX). All these models are presented in discrete time. The discrete time is used for modeling, because it is desired to describe the experimental measurements at fixed time intervals.

#### Autoregressive Exogenous (ARX)

The autoregressive Exogenous (ARX) model is one of the most common polynomial model structures. The ARX model specifies that the output of the system depends linearly on its own previous values. Autoregressive models are models are remarkably flexible at handling wide ranges of different time series patterns. The ARX model structure is define as:

$$y(t) = \frac{\sum_n^{j=1} B_j(z) u_j(t-nk) + e(t)}{A(z)} \quad (1)$$

where,  $A$  and  $B$  are polynomials and  $u(t)$ ,  $y(t)$  and  $e(t)$  are the input, output and system disturbance, respectively. The equation features the variable  $nk$  which adjusts the sample delay period.

### Autoregressive Moving Average Exogenous (ARMAX)

The autoregressive moving average model with exogenous inputs (ARMAX) extends the ARX structure by providing more flexibility for modeling the noise disturbance by introducing a C parameter. The ARMAX is a forecasting model in which both auto regression analysis and moving average methods are applied to the observed time series data.

$$y(t) = \frac{\sum_n^{j=1} B_j(z) u_j(t - nk) + C(z) e(t)}{A(z)} \quad (2)$$

where,  $A$ ,  $B$  and  $C$  are polynomials and  $u(t)$ ,  $y(t)$  and  $e(t)$  are the input, output and system disturbance, respectively. The  $n_k$  is the number of input samples that occur before the input affects the output.

### State Space (SS)

State Space (SS) models are common representations of dynamic systems. These models describe the same type of linear difference relationship between the inputs and outputs of a system as in an ARX model, but they are rearranged so that only one delay is used in the expressions. The order of a state space model relates to the number of delayed inputs and outputs used in the linear difference equation. The discrete-time SS model is given by the following equations:

$$x(t+1) = Ax(t) + Bu(t) + Ke(t) \quad (3)$$

$$y(t) = Cx(t) + Du(t) \quad (4)$$

where,  $A$ ,  $B$ ,  $C$ ,  $D$  and  $K$  are the state space matrices,  $u(t)$  is the model input,  $e(t)$  is the disturbance and  $x(t)$  is the vector of state variables. The matrix  $K$  determines the noise properties.

### Nonlinear Autoregressive Exogenous (NLARX)

The nonlinear autoregressive with exogenous inputs (NLARX) is an extension of the linear autoregressive exogenous model. The NLARX model can be mathematically represented as follows:

$$y_p(t) = f(y(t-1), y(t-2), y(t-3), \dots, u(t), u(t-1), u(t-2), \dots) \quad (5)$$

where,  $y(t-1)$ ,  $y(t-2)$ , ...,  $u(t)$ ,  $u(t-1)$ , ...,  $u(t-2)$  are delayed input ( $u$ ) and output ( $y$ ) variables, called regressors and  $f$  is a nonlinear function.

### Optimization

Genetic algorithm GA is utilized to find the optimal model structure yielding the minimum error between the simulated and actual model output data for the testing data set. The objective function of GA is to estimate the Mean Square errors or Deviation (MSD), Root Mean Square errors or Deviation (RMSE) and normalized root-mean-square deviation or Coefficient Of Variances (COV), for all model types. The time delay and model order are the optimal variables. The optimization constraints are set to cover the upper and lower limits of design variables, such as the maximum and minimum model order and time delay. Figure 4 shows a flow chart of the genetic algorithm process used to optimize the investigated system.

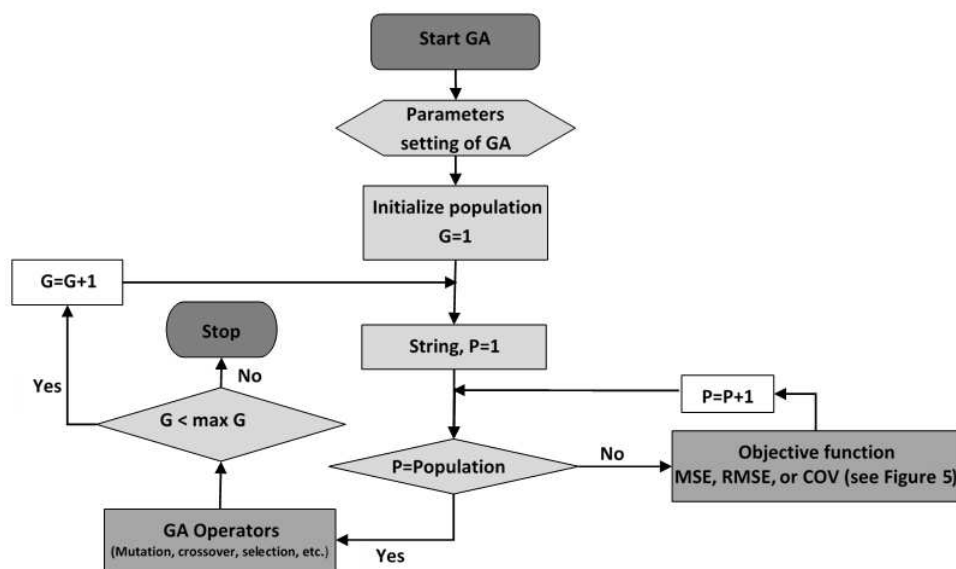


Fig. 4. A flow chart of the genetic algorithm procedure used for the model parameters optimization

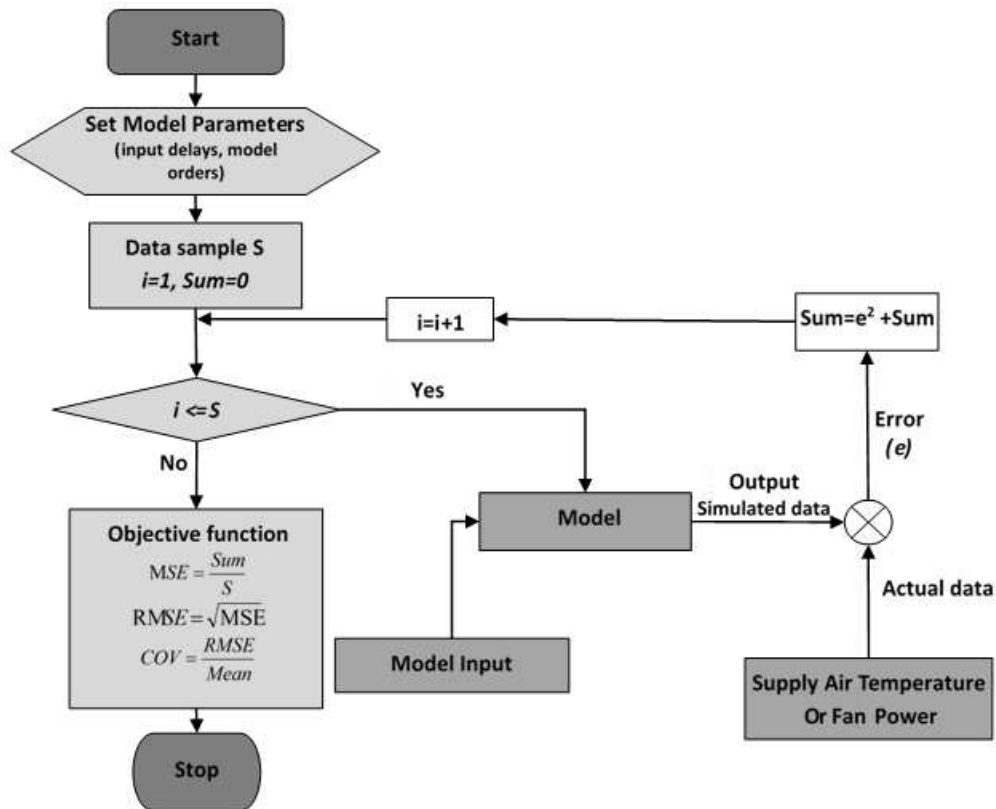


Fig. 5. Flow chart of objective function calculations (MSD, RMSD, or COV) for a sample S of data

As shown in Fig. 4, the genetic algorithm GA modifies a population of individual solutions. It selects individuals at random, at each time step, from the current population to be parents and then it uses to produce the children for the next generation. Over successive generations, the population evolves toward the optimal solution. The genetic algorithm starts with a random generation of the initial solution or initial population and ends with the optimal solutions (optimal variables). The model orders and time delays as problem variables represent an individual solution in the population. The objective function of the first generation (MSD, RMSD or COV) is calculated as shown in Fig. 5. The second generation is obtained using operations on individuals such as selection, crossover and mutation, in which individuals with higher performance have a greater chance to survive. The performance (fitness) of each new individual is again assessed. The process is then repeated until the maximum number of generations is reached.

## Results

A parametric study is performed with various model time delay and orders. The parametric study provides better understand the relationship between the model

parameters. It is also help to validate the optimization method. The data collected from BAS are divided into two sets: Training data set and testing data set. The training data set covers the period from November 1st to December 31th and the testing data set covers the period from January 1st to February 1st all at interval of five minutes. The model types described before (ARX, ARMAX, SS, NLARX) are evaluated with different model orders and time delays. The purpose of this evaluation is to get the best model structure. The input delays vary from 0 to 5 and the orders varies from 1 to 5. The resulted COVs for cooling coil model are illustrated in Fig. 6 and 7 for training and testing data sets, respectively. A total of 60 different model structures are evaluated for ARX and ARMAX (30 structure each). Only 10 different model structures are evaluated for NLARX and 26 structures for SS due to the amount of regressors available to vary. As shown in Fig. 6, the cooling coil model produces more accurate results in terms of the COV for ARX and ARMAX with elevated model order for the training set. However, this does not hold true for testing data set. There are optimal model order and time delay values that provide least COV. As shown in Fig. 7, the least COVs for testing data set and for ARX, ARMAX, SS, NLARX are 2.04, 2.02, 2.03 and 2.033%, respectively.

This parametric study shows that the cooling coil ARMAX model with a time delay 2 and model order of 4 holds the best results in term of COV value of 2.02% for the testing data set.

Figure 8 shows the actual data and simulated data obtained from the best cooling coil structure model (ARMAX model with a time delay 2 and model order of 4). The data for training and testing periods are depicted.

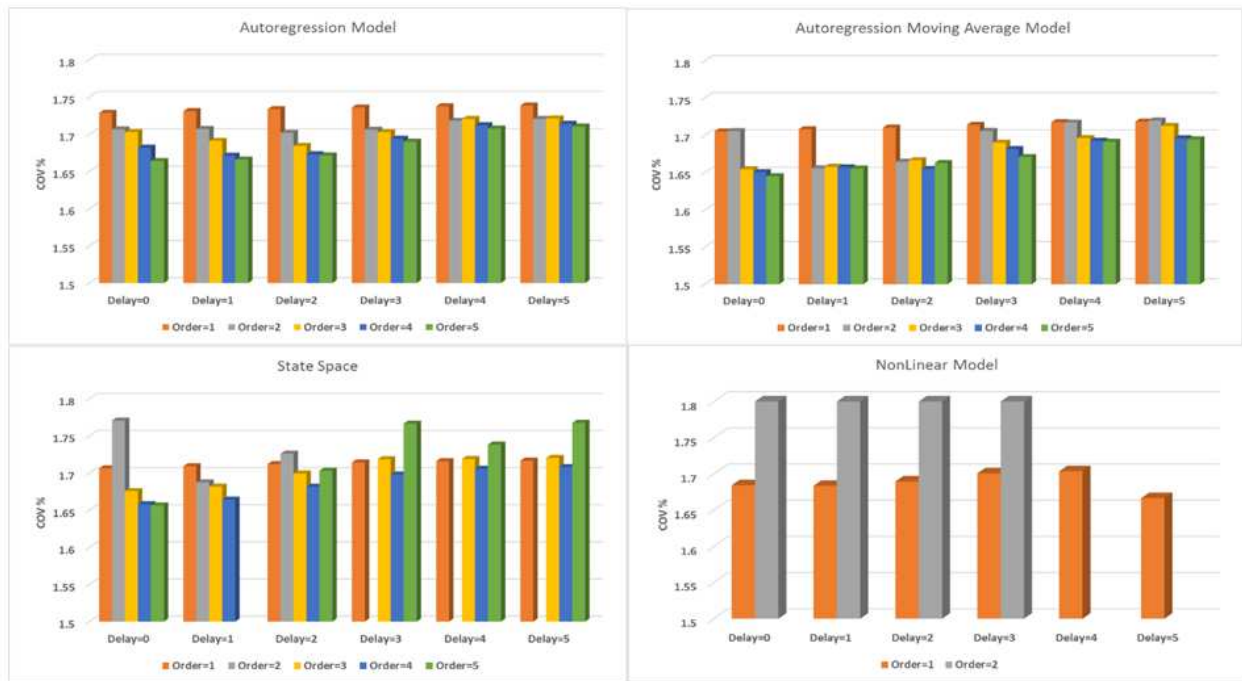


Fig. 6. Resulted COVs for training data set with different cooling coil model time delays and orders

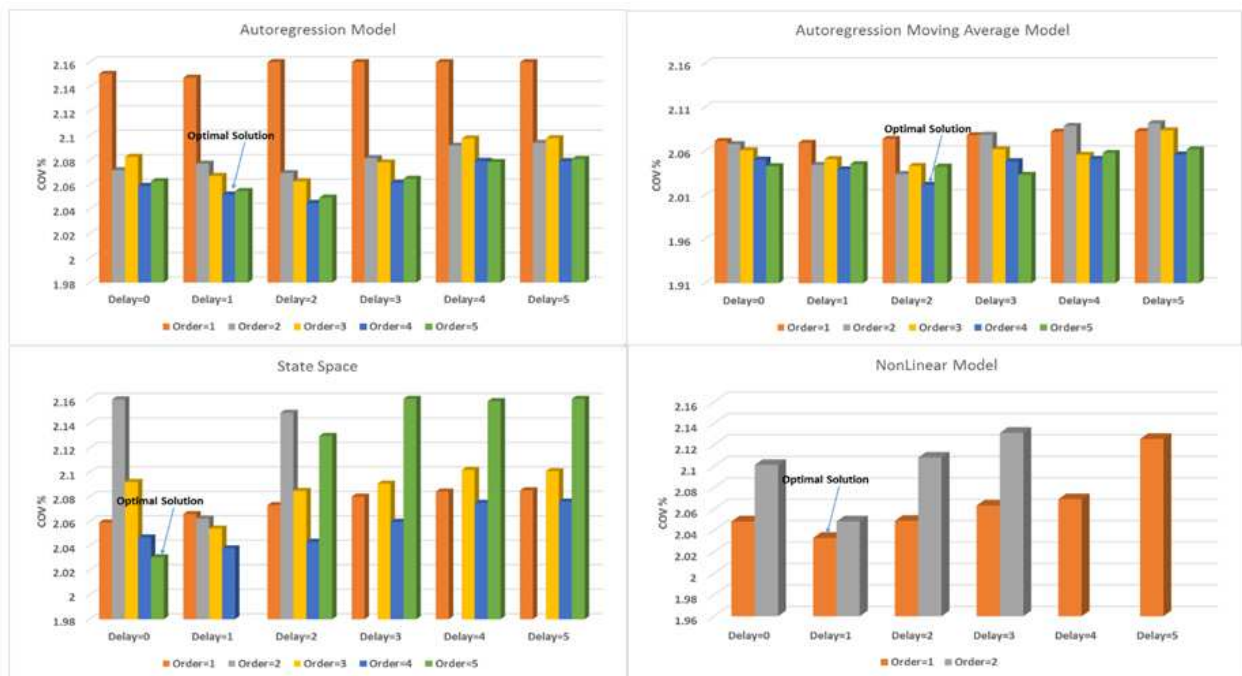


Fig. 7. Resulted COVs for testing data set with different cooling coil model time delays and orders



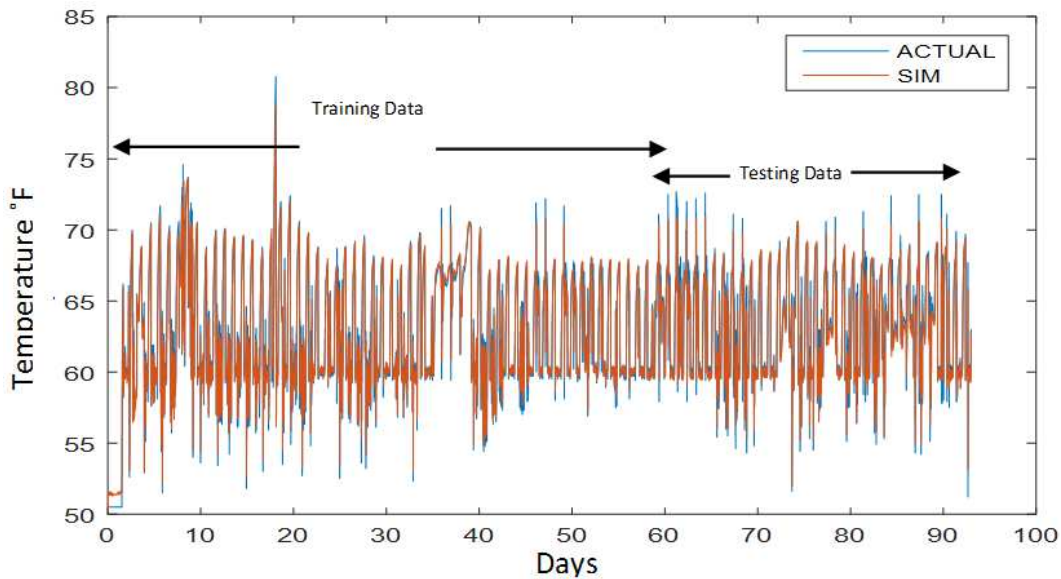


Fig. 8. The actual data and best structure cooling coil model outputs (ARMAX model with a time delay 2 and model order of 4)

Table 1. Resulted optimal COVs for testing data set for cooling coil and fan models

Type of model	Cooling Coil Model			Fan Model		
	MIN COV	Time delay	Order	MIN COV	Time delay	Order
ARX	2.04	2	4	3.34	1	3
ARMAX	2.02	2	4	3.26	1	3
SS	2.030	0	5	3.27	0	2
NLARX	2.033	1	1	3.28	1	1

The proposed optimization process shown in Fig. 4 and 5 are used to find the optimal solution (best model structure). The GA runs to find the optimal variables (the combination of model order and time delay) that produce the minimum RMSD, MSD, or COV for testing data set. In this study, the COV is utilized. The GA parameters are set to a maximum of 100 generations and a population size of 50. The optimal results are then compared with the results obtained with the parametric study. The optimization algorithm GA runs for each model types (ARX, ARMAX, NLARX and SS). The optimization results (minimum coefficient of variance COV) for cooling coil are summarized in Table 1. The optimal time delay and model order values obtained by GA are similar with those obtained by the parametric study. The optimization algorithm find that the cooling coil ARMAX with a time delay of 2 and an order of 4 provides the least COV of 2.02%. Similar scenario is applied for the fan model. The COV for the fan model is calculated by comparing the model output with the actual data (fan power). The fan model output and inputs are illustrated in Fig. 1. The optimal results COVs for the fan model are also shown in Table 1. The optimal fan model with time delay of 1 and model order of 3 provides accurate prediction in term of COV of 3.26.

## Conclusion

This paper discussed integrated modeling and optimization methods for a chilled water air handling unit using system identification approaches. Two multiple input-single output models were developed to find the supply air temperature and fan power of a typical chilled water air handling unit AHU using the time-series data that are usually available from a building automation system BAS. The system identification approaches (SD) with different modeling techniques and model structures were examined to find the best model types and structures yielding the best accurate predictions. An optimization procedure using GA was integrated into the model for better accurate predictions. Data from existing systems were collected to test and evaluate the proposed methods using the statistic index of Coefficient Of Variances (COV). Four model types (ARX, ARMAX, SS and NLARX) were examined and their structures were optimized by determining the optimal model time delays and orders. The test results showed that the autoregressive moving average exogenous cooling coil model with order of 4 and a time delay of 2 and autoregressive moving average exogenous fan model with time delay of 1 and model order of 3

provide most accurate predictions. The same results were obtained by the optimization process GA. The resulted COVs for cooling coil model are 2.04, 2.02, 2.03 and 2.033% and for fan model are 3.34, 3.26, 3.27 and 3.28% for ARX, ARMAX, SS, NLARX, respectively. The results validate that the developed identification models can provide an accurate prediction of the cooling coil and fan performance. This proposed integrated modeling and optimization method can be used for several applications required accurate system performance predictions such as energy assessment, energy saving estimation and fault detection and diagnosis.

## Acknowledgement

This work was performed under the following financial assistant award 70NANB14D1319 from the U.S. Department of Commerce, National Institute of Standards and Technology.

## Author's Contributions

Ms. Buford performed laboratory work and conducted data analysis of the research. Dr. Nassif and Dr. Abu-Lebdeh provided the research topic and guided the research development, modeling plan and data analysis. All authors contributed to the writing of the manuscript

## Ethics

No part of this article may be reproduced without written permission from the publisher or authors.

## References

EIA, U.S. Energy Information Administration (EIA). [www.eia.gov](http://www.eia.gov)  
ASHRAE, 2015. ASHRAE Handbook-Applications. American Society of Heating Refrigeration and Air Conditioning Engineers, Atlanta.

Afram, A. and F. Janabi-Sharifi, 2015. Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods. *Energy Build.*, 94: 121-49. DOI: 10.1016/j.enbuild.2015.02.045  
Deb, K., 2001. *Multi-Objective Optimization Using Evolutionary Algorithms*. 1st Edn., John Wiley and Sons, New York, ISBN-10: 047187339X, pp: 497.  
Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. 13th Edn., Addison-Wesley, Reading, ISBN-10: 0201157675, pp: 412.  
Hani, H., 2009. Employing computational intelligence to generate more intelligent and energy efficient living spaces. *Int. J. Automat. Comput.*, 05: 1-9. DOI: 10.1007/s11633-008-0001-7  
Kusiak, A. and G. Xu, 2012. Modeling and optimization of HVAC systems using a dynamic neural network. *Energy*, 42: 241-250. DOI: 10.1016/j.energy.2012.03.063  
Mossolly, M., K. Ghali and N. Ghaddar, 2009. Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm. *Energy*, 34: 58-66. DOI: 10.1016/j.energy.2008.10.001  
Nassif, N., 2014. Modeling and optimization of HVAC systems using artificial neural network and genetic algorithm. *Int. J. Build. Simulat.*, 7: 237-245.  
Nassif, N., 2008. Self-tuning dynamic models of HVAC system components. *Energy Build.*, 40: 1709-1720. DOI: 10.1016/j.enbuild.2008.02.026  
Buford, J. and N. Nassif, 2016. The dynamic modeling of chilled water HVAC systems using system identification methods. *Proceedings of the ASHRAE Annual Conference*, Jun. 25-29, St Louis.  
Wang, S. and X. Jin, 2000. Model-based optimal control of VAV air-conditioning system using genetic algorithm. *Build. Environ.*, 35: 471-487. DOI: 10.1016/S0360-1323(99)00032-3