Identifying Models of HVAC Systems Using ARIMAX

Fan Feng
Student Member ASHRAE

Zheng O'Neill, Ph.D., PE
Member ASHRAE

In recent years, Model Predictive Control (MPC) has received a lot of attention for controls of HVAC systems because of significant improvement of the energy efficiency while respecting occupant comfort(Oldewurtel et al., 2012). MPC relies on dynamic models of building and HVAC systems to predict the system states for a specified time horizon and then applies optimization algorithms to minimize the cost functions. Clearly, an appropriate dynamic model with reasonable prediction accuracy and computational speed is crucial for a successful implementation of MPC. Modeling HVAC systems of buildings have a long history through simulation software such as EnergyPlus and Trnsys. The models generated by this building simulation software, however, are not amenable to control design because the models are high-dimensional from their inclusion of complex physical effects.

Auto-Regressive Integrated Moving Average (ARIMAX) is a promising alternative model for building and HVAC systems because it generates equations that can be used for control while encapsulating the highly time-varying heating load from exogenous sources. Simply put, an ARIMAX model can be viewed as a multiple regression model with one or more autoregressive (AR) terms and/or one or more moving average(MA) terms, as Eqn. (1).

$$Y_t = \sum_{i=1}^{p}\phi_i Y_{t-i} + \sum_{j=1}^{q}\theta_j \varepsilon_{t-j} + \sum_{k=1}^{m}\beta_k X_k + \varepsilon_t \qquad (1)$$

Where X_k is the value of the exogenous variables at time *t*, Y_{t-i} is the immediately preceding value of the dependent variable at time *t-i*, ε_{t-j} is the estimation error produced by the model at time *t-j*.

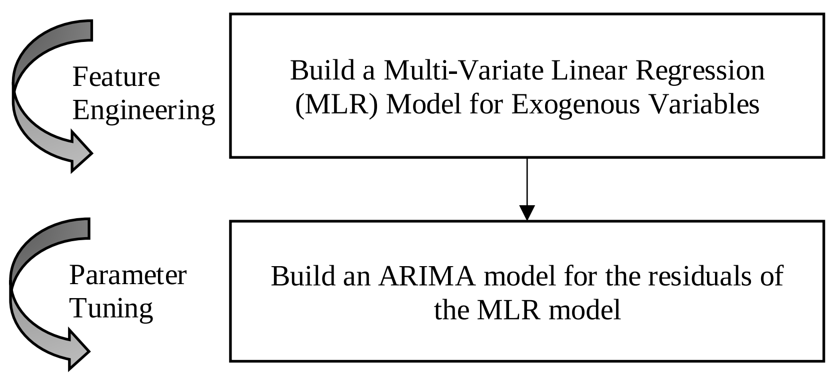

Figure 1. ARIMAX model-building algorithm

Building an ARIMAX model calls for combining the predictive value of both the trailing time series values themselves (Y_t) and the trailing model errors (ε_t) with the predictive value of exogenous variables. Usually, the ARIMAX approach to the time-series model building has two phases, and this two-phase method will be implemented in this paper. Admittedly, there are some other more complex and reasonable approaches (Andrews, Dean, Swain, & Cole, 2013), but whether the benefits from using these approach could justify the difficulties/costs of implementing them have not been tested yet. The flowchart shown in

Figure 1 presents the algorithms used to build a valid ARIMAX model. In the first step, a Multi-Variate Linear Regression(MLR) will be constructed to leverage the predictive value of exogenous variables. Moreover, Feature Engineering (FG) (Zhang, Wen, & Chen, 2017), which is an iterative process to generate and select a set of most useful independent variables, could be used to fully explore the predictive power of exogenous variables. Then, after a high-quality MLR model has been built, we will build an ARIMA model for the residuals of the MLR model. Herein, we will conduct an automated parameter tuning procedure to estimate the parameters for ARIMA models, because one emerging trend is that data mining algorithms should have as few parameters as possible, ideally none.

Testbed and dynamics excitation

The testbed used in this study is the Building examples in Modelica Buildings Library by Lawrence Berkeley National Laboratory, which is a variable air volume flow systems with terminal reheat. This system serves a total of 5 zones, among which four are perimeter zones. In this paper, we focused on one perimeter zone, and the investigation of other zones will be carried out later.

For the model, we considered the control inputs of VAV supply air flow rate and zone supply temperature with ambient temperature as measured disturbances, and the outputs of the model are zone air temperatures. Figure 2a and 2b show the excitation signals for these inputs respectively, and the ambient temperature profile is presented in Figure 2c. Herein, our research is focusing on one thermal zone of this building, and later more zones will be investigated. Figure 2d shows the profile of the output—zone air temperature.

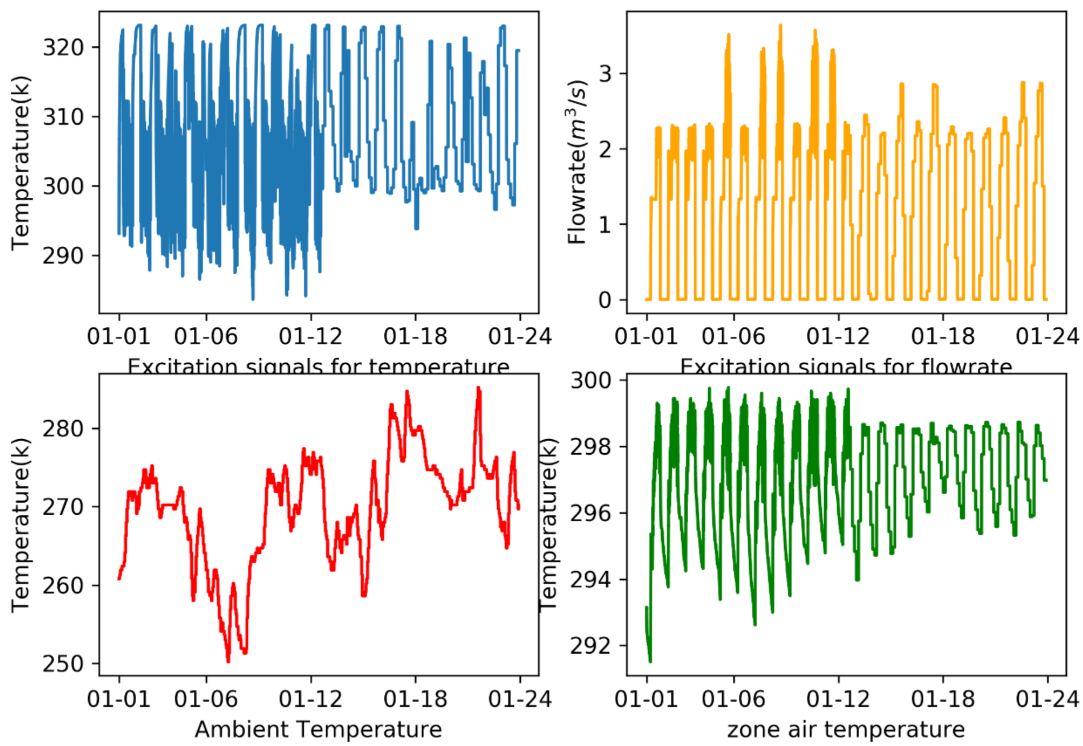

Figure 2 Dynamics excitation

Results and analysis

For system identification and validation, we performed functional tests during the first 3 weeks of January using our virtual testbed. The data from the first half was used for model identification, and the data from the last half was used as a test dataset.

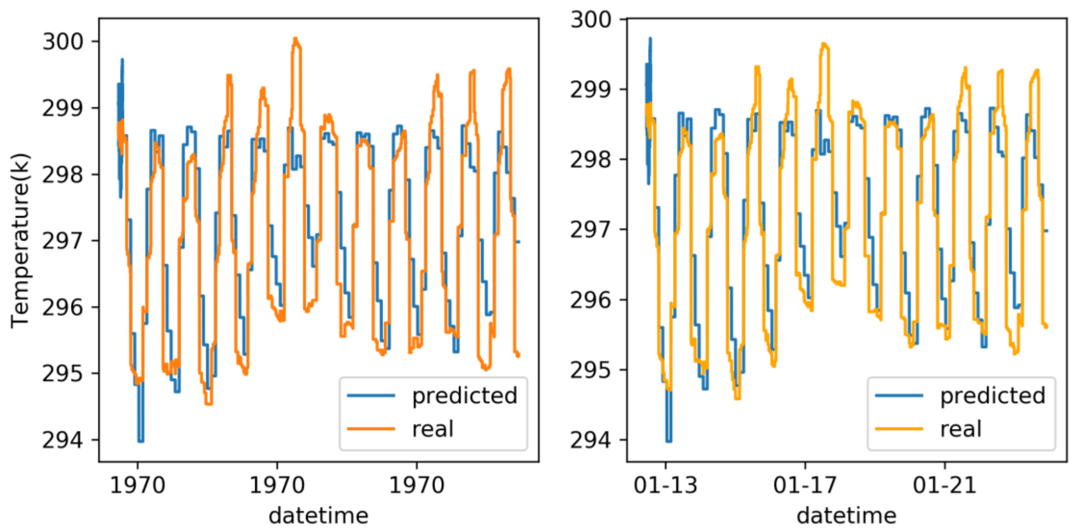

Figure 3 Validation results

Figure 3(left) shows the comparison between real values and predicted values just by the MLR model in testing dataset. The time horizon used in this study is 3 hours. This single MLR model has an R^2-score of 0.65, which is not really good. Then, we modeled the residuals further by using the ARIMA model, and the results are shown in Figure 3(right). It is clear that after integrating with the ARIMA model, the accuracy of this model is significantly improved, which has an R^2-score of 0.72.

However, further work should be conducted to improve the accuracy of this model and test the applicability of this model for another building case. In addition, an automated feature engineering procedure should be developed to build a totally parameter-free model building method.

References

[1] Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., … Morari, M. (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, *45*, 15–27.

[2] Andrews, B. H., Dean, M. D., Swain, R., & Cole, C. (2013). Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. *Society of Actuaries*, 1–54.

[3] Zhang, L., Wen, J., & Chen, Y. (2017). Systematic Feature Selection Process Applied in Short-Term Data-Driven Building Energy Forecasting Models: A Case Study of a Campus Building. In *ASME 2017 Dynamic Systems and Control Conference* (p. V003T27A002--V003T27A002).