Real-time Dynamic House Thermal Model Identification for Predicting HVAC Energy Consumption

Yicheng Wen
Algorithm Engineer
GE Appliances
Louisville, KY, 40225
Email: yicheng.wen@ge.com

William Burke
Advanced Systems Engineer
GE Appliances
Louisville, KY, 40225
Email: wjburke@ge.com

***Abstract*—This paper presents a real-time algorithm to predict the energy consumption of the heating, ventilation, and air conditioning (HVAC) system at home. The autoregressive model with exogenous inputs (ARX model) is used to identify the house thermal model. The ARX model, with the thermostat controller, is simulated to obtain the future state of the HVAC system with the knowledge of the weather forecast data obtained from a weather server. The utility bill for the HVAC system can be estimated if a real-time price model is provided, thereafter. The proposed method is validated by experimentation in a particular home using GE Nucleus energy management system for data aggregation and algorithm implementation. The experimental results show that the energy prediction error is around 15% in both heating and cooling mode of the HVAC system.**

***Keywords*-dynamic house thermal model; HVAC; bill prediction;**

I. Introduction

The heating, ventilation, and air conditioning (HVAC) system can cost over 50% of the total energy consumption at residential houses. It is important to make people aware of their home energy consumption in advance so that actions can be taken to save the utility bill if necessary.

There are two kinds of data-driven energy modeling approaches. One is called the direct method. It assumes a direct mathematical relationship between a set of inputs (e.g., outdoor temperature) and an output (e.g., electrical consumption). Regression methods [1] are used to determine the parameters. In [2], many such methods are discussed. However, several drawbacks in this method are: *a*) the temporal dependencies are not modeled; *b*) the prediction error is larger beyond the range of the training inputs; *c*) the control parameters such as the setpoint setting and the deadband setting are not considered.

Another method is based on the identification of a house thermal model. Afterwards, the model is simulated with the known control logic of the HVAC system for energy prediction. There are many popluar house thermal models presented in the liturature. In [3], a white-box approach is followed. A heat transfer model, which considers a building as a thermal network analogous to an electric circuit network, is used to build a state space model. The coefficients of the model, such as house structure, thermal resistance, and thermal capacitance are obtained from the construction handbook. Simplification of such a heat transfer model can be found in [4], where the authors stated that larger model size (more than 2nd order) does not lead to significant improvement of their simulation results, in [5], and in [6]. In [7], autoregressive models with external input (ARX) and autoregressive moving average models with external input (ARMAX) are investigated with different structures and it is also concluded that the 2nd order ARX model is preferred. Artificial neural network (ANN) is an alternative model [8]. The performance of the ARX model and a neural network auto regressive with exogenous input (NNARX) model is compared in [9]. The conclusion is that NNARX gives more accurate indoor temperature prediction than the ARX model since ANN can capture some non-linearity in the process. However, in this paper, the ARX model is used rather than a neural network model due to the following consideration.

- The identification algorithm for ARX model, the recursive least square (RLS) [10], is non-iterative, computationally inexpensive, and relatively easy to implement;
- ARX model can achieve decent performance with very simple model structure (2nd order model seems to perform well in most cases);
- ARX model provides a straightforward physical interpretation. For example, it is easy to see the contribution of each input to the indoor temperature.

The indoor temperature can depend on many factors, such as the outdoor temperature, the effective power of HVAC systems, relative humidity, solar radiation, wind speed, house condition, indoor human activity, etc. All of them are not easy to measure or to quantify for an ordinary household. Indeed, more measurement may result in a better model, but may be difficult to implement or to commercialize in practice.

The contribution of this paper is that we only consider easily accessible measurements, which has major impact on the model, including the indoor temperature (from a thermostat), the state of the HVAC system (from a sub-meter), and the outdoor temperature (from an online web resource) with a simple 2nd order ARX model for the identification of the thermal model. We show that this approach can be implemented in an ordinary household with off-the-shelf products (GE Nucleus energy management system) and yet provides the energy prediction within an acceptable error bound.

The paper is organized as follows. In Section II, the detailed mathematical formulation and the procedure of the method are described including data collection, data processing, model identification, and energy prediction. Section III presents the experimental set-up and the results of the proposed method in both heating and cooling mode of the HVAC system. The paper is concluded with recommendation for future work in Section IV.

II. Thermal Model Identification and Bill Prediction

In this section, we present in detail the method of identifying the ARX model and predicting the energy consumption of HVAC system. For the ease of presentation, we discuss how to identify a heating model in winter. However, the exactly same procedure can be followed to work with a cooling model in summer.

A. Input-output Data and Data Pre-processing

Let k be the discrete time instant with some sampling period T_s. At each time instant k, we need to obtain the following variables.

- Indoor temperature y_k in Celsius;
- Outdoor temperature $\tilde{v}_k$ in Celsius;
- Instantaneous HVAC power consumption $\tilde{u}_k$ in watts.

It is not hard to derive the state of the HVAC system based on its instantaneous power $\tilde{u}_k$. When the heater or the cooler is on, $\tilde{u}_k$ is usually on the order of hundreds or thousands of watts while on the order of ten watts when off. It is thus easy to find a threshold C (e.g. 100 watts) and we define a state variable $u_k \in \{0, 1\}$ for the HVAC system by

$$u_k = \begin{cases} 0 & \text{if } \tilde{u}_k < C \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

Denote eu the averaged energy usage as the heater or the cooler is off and ev the averaged energy usage when it is on. Both eu and ev can be estimated at time k by averaging over the history.

$$eu_k = \sum_{i=1}^{k} (1 - u_i)\tilde{u}_i T_s \tag{2}$$

$$ev_k = \sum_{i=1}^{k} u_i \tilde{u}_i T_s \tag{3}$$

where T_s is the sampling time.

Let us denote the difference between the outdoor and indoor temperature by

$$v_k = \tilde{v}_k - y_k \tag{4}$$

B. ARX Model Identification

The thermal model is modeled by a discrete-time auto-regressive with an exogenous inputs (ARX) model with v and u as inputs and y as the output. In general, such an ARX model is

$$y_k = \sum_{i=1}^{n_a} a_i y_{k-i} + \sum_{i=1}^{n_{b_1}} b_{1i} v_{k-i} + \sum_{i=1}^{n_{b_2}} b_{2i} u_{k-i} + \varepsilon_k \tag{5}$$

where ε is the noise term with zero mean. For simplicity, we will present with $n_a = 2$, $n_{b_1} = 1$, and $n_{b_2} = 1$ and the noise term ε is omitted because it does not affect the mean estimation. In Section III, we shall compare the performance when we change the model structure. This reduces the model structure to

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + b_1 v_{k-1} + b_2 u_{k-1} \tag{6}$$

The ARX model can be identified by the recursive least square (RLS) algorithm, which can be found [10].

C. Weather Forecast Data Fitting

The weather data can be obtained from the online weather service for the particular region where the house is located, such as *Weather Underground (www.wunderground.com)*. The current temperature can be pulled directly from the server, while the forecast data usually needs to be interpolated or fitted depending on the available data. If hourly future data is available, it might be enough to perform a linear interpolation to estimate the daily profile. Sometimes only daily maximum and daily minimum are available. In this case, a piece-wise sinusoid curve is fitted to describe the daily temperature variation.

Given the daily maximum (T_{max}) and minimum temperature (T_{min}), a sinusoid curve needs to be fitted. We first estimate the time for T_{max} and T_{min} to be achieved at that day, denoted by t_{min} and t_{max} (UTC time in second), respectively, if they are not known. Let $t_0(m)$ be the UTC time at midnight (0 am) at day m. It is logical to use the average time over the previous L days as an estimate. To that end, let W be the total seconds of a day, namely $W = 86400$ sec. The estimate of $t_{min}(m+1)$ is given by

$$\begin{aligned} \hat{t}_{min}(m+1) &= t_0(m+1) \\ &\quad + \frac{1}{L} \sum_{i=0}^{L-1} \left(t_{min}(m-i) - t_0(m+1) + (i+1)M \right) \\ &= \frac{(1+L)M}{2} + \frac{1}{L} \sum_{i=0}^{L-1} t_{min}(m-i) \end{aligned} \tag{7}$$

and $\hat{t}_{max}(m+1)$ can be found similarly.

$$f(t) = \begin{cases} \frac{T_{max}(m)+T_{min}(m)}{2} - \frac{T_{max}(m)-T_{min}(m)}{2}\cos\left(\frac{\pi(t-t_{min}(m))}{t_{max}(m)-t_{min}(m)}\right) & t \in [t_{min}(m), t_{max}(m)) \\ \frac{T_{max}(m)+T_{min}(m+1)}{2} + \frac{T_{max}(m)-T_{min}(m+1)}{2}\cos\left(\frac{\pi(t-t_{max}(m))}{t_{min}(m+1)-t_{max}(m)}\right) & t \in [t_{max}(m), t_{min}(m+1)) \end{cases} \tag{8}$$

At day m and $m+1$, we have $t_{min}(m) < t_{max}(m) < t_{min}(m+1)$ (use estimates in Equation 7 if they are not available), then the temperature fitting profile $y(t)$ is given in Equation 8.

D. Thermostat Controller

Most common residential thermostats are bang-bang controllers with a deadband 2δ. The controller in the heating mode is

$$u_k = \begin{cases} 1 & \text{if } u_{k-1} = 0, y_{k-1} < r_{k-1} - \delta \\ 1 & \text{if } u_{k-1} = 1, y_{k-1} < r_{k-1} + \delta \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where r_k is set point at time k. Most programmable thermostats have pre-programmed setpoints. Similarly, the controller in the cooling mode is

$$u_k = \begin{cases} 1 & \text{if } u_{k-1} = 0, y_{k-1} > r_{k-1} + \delta \\ 1 & \text{if } u_{k-1} = 1, y_{k-1} > r_{k-1} - \delta \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

E. Energy/Bill Prediction

Let τ be the UTC time for the slow-time scale of the prediction. At time τ, the identified dynamic model is

$$y_k = a_1^\tau y_{k-1} + a_2^\tau y_{k-2} + b_1^\tau v_{k-1} + b_2^\tau u_{k-1} \tag{11}$$

Assume that the prediction window is of length N. Let y_0 and y_{-1} be the initial condition, the temperature at time τ and $\tau - T_s$, respectively. The forecast temperature is given by use of Equation 8

$$v_k = f(\tau + kT_s), \quad k = 1, \ldots, N \tag{12}$$

Then the future indoor temperature y_k and the state of the HVAC system u_k can be simulated following Equation 11 and 9 (or Equation 10).Then the energy consumption is given as

$$\sum_{k=1}^{N} ev \cdot u_k + eu(1 - u_k) \tag{13}$$

Recall that u_k only takes value 0 or 1, so only the first term contributes if $u_k = 1$ and only the second term does, otherwise.

If a time-of-use price model is available, we can easily compute the estimated bill. Assume that $p(k)$ gives the unit price at time k, then the predicted bill can be computed by

$$\sum_{k=1}^{N} p(k) \left(ev \cdot u_k + eu(1 - u_k)\right) \tag{14}$$

Figure 1. Test House

III. Experimental Results

In this section, the experimental set-up of the proposed method is described. Experimental results are presented to validate the proposed method.

A. Experimental Set-up

The experiment was conducted in an occupied single-family home constructed circa 1940, as shown in Figure 1. It is an all brick, one and a half story house with a finished basement. The above grade living area is approximately 1800 sq-ft. The house was occupied by two adults and two children and is located in Louisville Kentucky, inside the 40205 zip code. Louisville has a humid sub-tropical climate, with hot and humid summers, cold winters, and temperate springs and falls. The house has several efficiency upgrades, including double-pane windows.

The test house contained a GE Nucleus energy management system. The Nucleus is a small form factor computer capable of communicating via the Zigbee (IEEE802.15.4) wireless standard to Smart Energy Profile (SEP) devices. Such SEP devices include smart electricity meters, load

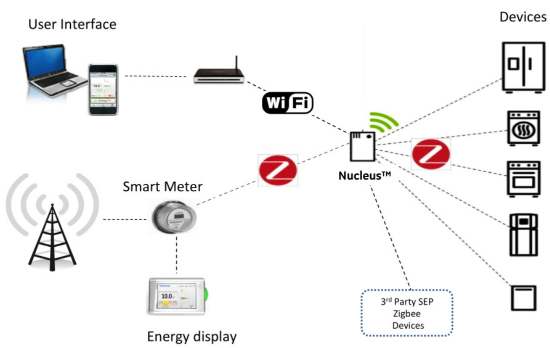


Figure 2. Nucleus Connection Diagram

panel based energy sensors, plug sensors, thermostats, environment sensors, and a range of GE smart appliances. The Nucleus can also communicate over IP via Wifi or Ethernet. Figure 2 illustrates a connection diagram for the Nucleus. In addition to advanced communications technologies, the Nucleus also contains data storage, control, and visualization capabilities.

A special version of the Nucleus software was created for this test in order to enable advanced data acquisition. The following measurement were performed on the house:

- Whole house power measurements with 15 second sample rate.
- Air Conditioning compressor power measurements with 15 second sample rate.
- Heating Ventilation and Air Conditioning (HVAC) blower power measurements with 15 second sample rate.
- Indoor temperature measurements via a wireless thermostat.
- Outdoor temperature measurements via a Zigbee wireless sensor.

We collected 30 days of data in the test house in November, 2011, with Figure 3 showing a portion of the collected data. The solid curve and the dashdot curve are the indoor temperature (y_k) measured by the thermostat and the outdoor temperature ($\tilde{v}_k$) measure by a temperature sensor located at the porch and well shielded from the sunshine, respectively. The dashed one is the scaled energy consumption for HVAC ($\tilde{u}_k$). Figure 4 gives a zoomed-in version of a portion of Figure 3. It is observed that when $\tilde{u}_k$ is high, the indoor temperature y_k arises, which indicates that the heater is turned on.

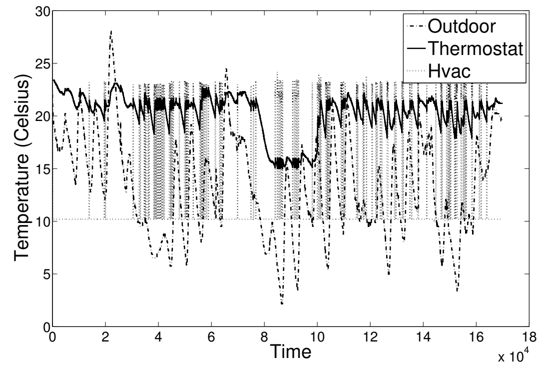


Figure 3. Collected Data

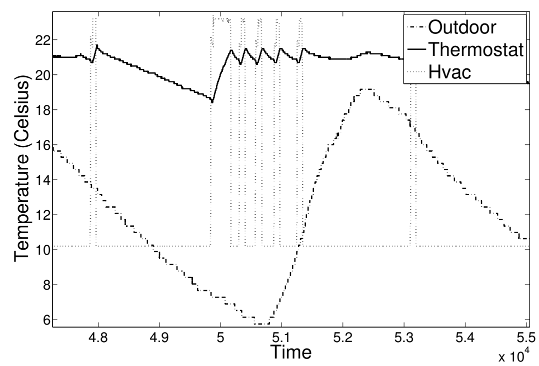


Figure 4. Zoomed-in of Figure 3

B. Algorithm Evaluation

We compare the results by executing the algorithm with the fitted weather data and with the actual measurement. Figure 5 gives the actual measurement (solid) and the fitted data (dashed) of the outdoor temperature using Equation 8.

We ran the RLS algorithm with the forgetting factor $\lambda = 0.99995$ to identify the ARX model of the house. We executed the proposed method to perform the energy prediction with the prediction window of 2 weeks (namely, $N = 14$). The prediction started on day 6 and ran every half a day. Table I shows the prediction results for energy prediction and for future indoor temperature for different orders of parameters. It is observed that the prediction error is around 15% with the fitted weather data and 8% with the actual measurement. The averaged temperature error is around $0.5°C$ for all the cases. It is noted that the order n_a of the ARX model has little impact on the result. This coincides with the observation in the literature [4] [7] that a second order model is enough for the thermal model.

To visualize the simulation result, we identified the ARX model using half of the data (15 days) and ran the simulation from day 1 with the initial condition, the identified model,

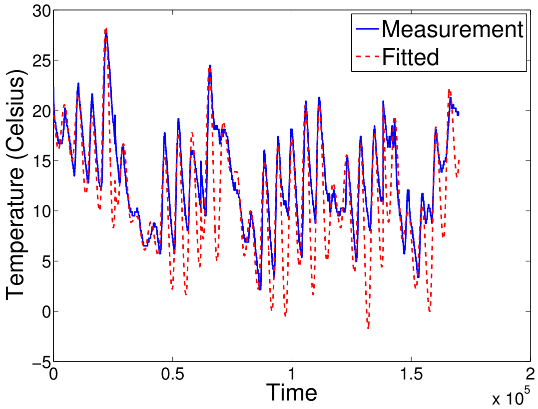


Figure 5. Fitted Outdoor Temperature vs. Actual Measurement

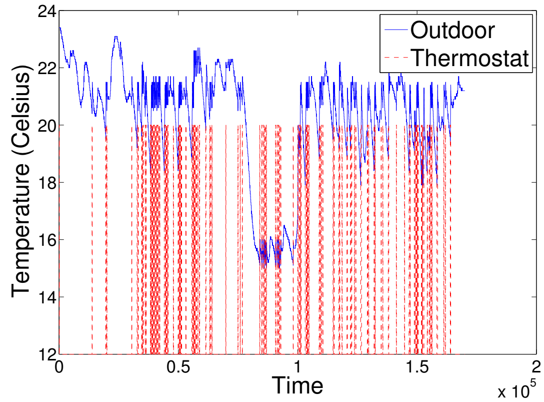
Table I
PREDICTION ERROR FOR HEATING MODEL

Order (n_a)		Fitted Weather Data	Measurement
2	Energy Prediction Error	14.61%	7.89%
	Energy Prediction Std.	9.48%	4.87%
	Temperature Error ($^\circ$C)	0.4867	0.4745
	Temperature Std. ($^\circ$C)	0.5402	0.5401
3	Energy Prediction Error	14.61%	7.96%
	Energy Prediction Std.	9.39%	4.85%
	Temperature Error ($^\circ$C)	0.4838	0.4712
	Temperature Std. ($^\circ$C)	0.5391	0.5381
4	Energy Prediction Error	14.60%	7.93%
	Energy Prediction Std.	9.35%	4.85%
	Temperature Error ($^\circ$C)	0.4816	0.4685
	Temperature Std. ($^\circ$C)	0.5373	0.5375
5	Energy Prediction Error	14.58%	7.91%
	Energy Prediction Std.	9.35%	4.86%
	Temperature Error ($^\circ$C)	0.4804	0.4689
	Temperature Std. ($^\circ$C)	0.5360	0.5367

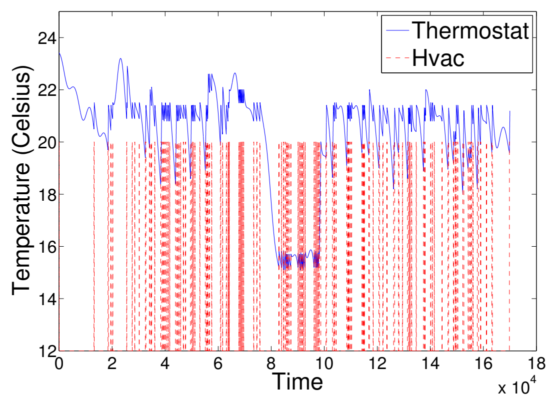
and the outdoor temperature data. Figure 6 gives the comparison between the prediction of the indoor temperature and the state of the HVAC in both cases of using fitted weather data and actual measurement. Figure 6(a) shows the actual measurement. Figure 6(b) and Figure 6(c) show the predicted values in both cases. The predicted values are fairly close to the actual values in either case.

To study the effect of setpoints on the energy consumption, we offset the setpoints by some certain degrees in the prediction window (second half of the data). The ARX model was identified using the first half of data, as above. Figure 7 depicts how the energy prediction for the second half changes depends on the different setpoints. Surprisingly, it shows almost a linear relation. It is noted that the energy consumption cuts into half if one lowers the setpoints by 4 degrees. This chart allows the users to make trade-off between their comfort and the cost.

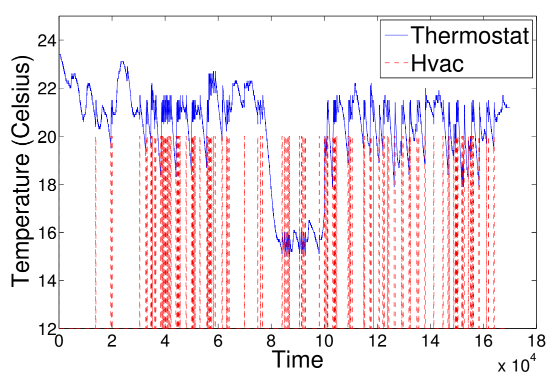
Another set of data was collected in June, 2012, to validate the proposed method for a cooling model. Figure 8 depicts the raw data of the indoor temperature (solid), outdoor temperature (dashdot), and the scaled instantaneous power



(a) Measured Data



(b) Using Fitted Weather Data



(c) Using Actual Outdoor Measurement

Figure 6. Simulation of HVAC System for a Month

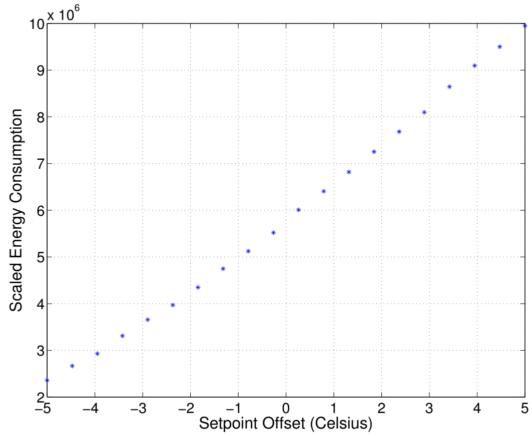


Figure 7. Scaled Energy Consumption vs. Setpoint Offset

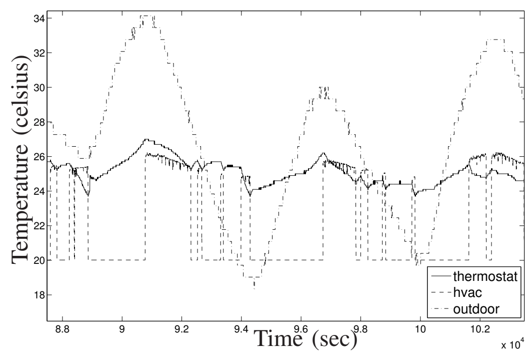


Figure 8. Portion of Raw Data Collected in Summer

of the air conditioner (dashed). It is observed that when the power of the AC is high, the indoor temperature decreases as expected since the AC is cooling down the house. A similar simulation is executed to create Table II in this case. Again the prediction error is found to be around 15% with the fitted weather data when a second order model is used.

IV. Conclusion and Future Work

This paper presents a method to predict the energy consumption of the HVAC system at home. The ARX model is used for identiying a house thermal model. With the knowledge of the thermostat controller and the weather forecast data, the ARX model is simulated to predict the future state of the HVAC system and therefore the energy consumption. This method can be easily implemented in an ordinary house with GE Nucleus technology for home energy management. A real-time experiment was conducted in a particular house and the prediction error of the proposed method is around 15% for a two-week prediction window. The relationship between energy consumption and the setpoints can be revealed and provides users with useful information to achieve trade-off between setpoints and cost.

One of the future directions is to incorporate users' comfort and cost effect into the formulation and to design a high-level controller above the thermostat controller to automatically adjust the setpoint of the house.

Table II
Prediction Error for Cooling Model

Order $n_a = 2$	Fitted Weather Data	Actual Measurement
Energy Prediction Error	15.94%	3.00%
Energy Prediction Std.	5.31%	1.28%
Temperature Error (°C)	0.4955	0.4615
Temperature Std. (°C)	0.3617	0.3608

References

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd ed., 2009.

[2] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy and Buildings*, 2012.

[3] M. R. Kulkarni and F. Hong, "Energy optimal control of a residential space-conditioning system based on sensible heat transfer modeling," *Building and Environment*, vol. 39, pp. 31–38, 2004.

[4] O. Mejri, E. P. D. Barrio, and N. Ghrab-Morcos, "Energy performance assessment of occupiedbuildings using model identification techniques," *Energy and Buildings*, vol. 43, pp. 285–299, 2011.

[5] K. K. Andersen, H. Madsen, and L. H. Hansen, "Modelling the heat dynamics of a building using stochastic differential equations," *Energy and Buildings*, vol. 31, pp. 13–24, 2000.

[6] R. Balan, J. Cooper, K.-M. Chao, S. Stan, and R. Donca, "Parameter identification and model based predictive control of temperature inside a house," *Energy and Buildings*, vol. 43, pp. 748–758, 2011.

[7] G. Ríos-Moreno, M. Trejo-Perea, R. Castañeda-Miranda, V. Hernández-Guzmán, and G. Herrera-Ruiz, "Modelling temperature in intelligent buildings by means of autoregressive models," *Automation in Construction*, 2007.

[8] Ö. A. Dombayc, "The prediction of heating energy consumption in a model house by using artificial neural networks in denizli–turkey," *Advances in Engineering Software*, vol. 41, pp. 141–147, 2010.

[9] A. Mechaqrane and M. Zouak, "A comparison of linear and neural network arx models applied to a prediction of the indoor temperature of a building," *Neural Computing and Applications*, vol. 13, pp. 32–37, 2004.

[10] L. Ljung, *System Identification*. Prentice Hall, 2nd ed., 1999.