Notes on a dynamic model of a building's thermal dynamics and its use in system identification

Prabir Barooah and Austin Coffman
University of Florida
Gainesville, Florida USA

Abstract

We describe briefly the so-called resistor-capacitor network model structure that is frequently used to model thermal dynamics of buildings. The discussion is tailored to using these models for system identification.

I. INTRODUCTION

A thermal dynamic model of a building can be used to predict how the indoor temperature of a building evolves in response to various inputs. Since temperature at every point inside a building can be distinct and also time-varying, a full-blown model will be a PDE. But for control purposes, we seek a low order ODE model. A review of various model structures can be found in [R.Kramer et al., 2012].

II. A LOW-ORDER PHYSICS-BASED MODEL: RC NETWORK

The basic idea for constructing a simple thermal dynamic model is that a building has (1) thermal inertia so that once heated or cooled it doesn't suddenly change temperature and (2) it offers resistance to heat exchange between its interior and the outside. The first is modeled using one or more capacitances, and the latter by one or more resistance. An electrical analogy is at play here. Temperature is like electrical potential: a higher difference in temperature between two points in the network causes more heat flow. For obvious reasons, such a model is called a resistance capacitance (RC) network model. This is actually a class of models instead of being one model.

A. 1R-1C

Within RC network models, the simplest is the 1R-1C model:

$$C\dot{T}_z(t) = \frac{T_{oa}(t) - T_z(t)}{R} + A^{\mathrm{eff}}\eta^{\mathrm{sol}}(t) + q_{\mathrm{x}}(t) + q_{\mathrm{hvac}}(t) \tag{1}$$

where C is the thermal capacitance of the building and R is the thermal resistance of the building, respectively, and $A^{\mathrm{eff}}$is the effective area of the building for incident solar radiation, q_{x} is heat generated by sources internal to the building (occupants, lights, computers etc.), and q_{hvac} is the heat injected into the building by its HVAC system. In the HVAC system provides cooling, like it does in Florida most times in summer, q_{hvac} is negative.

Figure 1 shows a schematic illustration of the model.

You can see the electrical analogy from the ODE: the first term on the right represents heat outflow (out of the building) when $T_z > T_{oa}$, and heat inflow otherwise, and this heat flow is smaller if the thermal resistance R is higher. If the capacitance C is large, the same rate of heat flow will lead to a smaller change in temperature, which is consistent with our intuitive notion of capacitance (whether thermal or electrical).

Moreover, this first term ($\frac{T_{oa}(t)-T_z(t)}{R}$) says that the rate of heat flow between the outside and the inside is proportional to the temperature difference between outdoors and indoors; this is sometimes justified from Newton's law of cooling. You should remember this is a gross simplification. The rate of heat exchange between the outside of the building and its inside through the walls and windows is a very complex phenomenon involving all three forms of heat exchange: conduction, convection and radiation. It is not possible for us to get into a thorough discussion. If you are curious, an excellent exposition can be found in [Allen, 1980]. If you have the appetite for a far more technical discussion; see [American Society of Heating, Refrigerating and Air Conditioning Engineers, 2005].

This document was prepared for EML 6352 students.

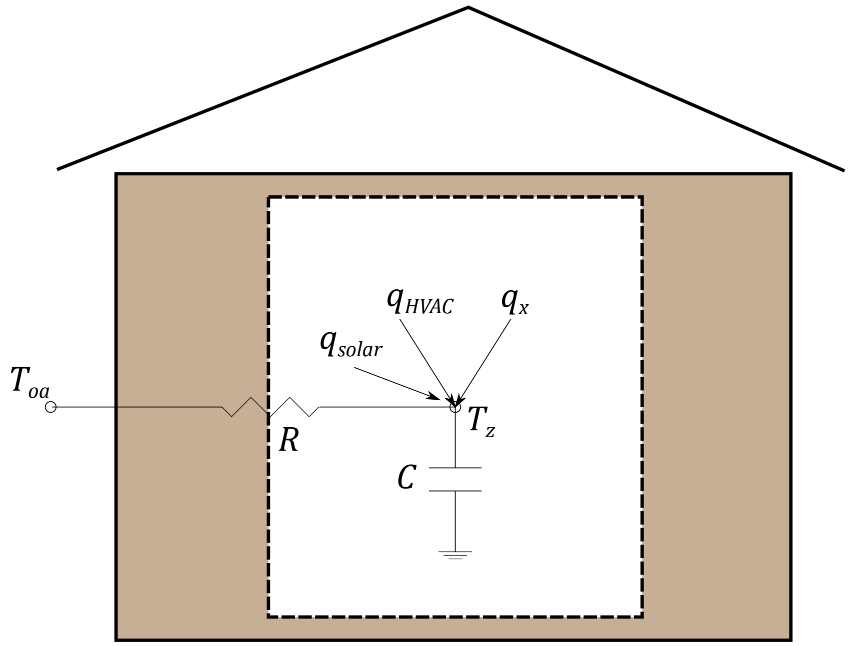

Fig. 1: A 1R-1C model of a building's thermal dynamics

B. 2R-2C

There are several criticisms of the 1R-1C model. One is that it does not capture the distinct thermal inertia of the air inside the building (whose temperature we are are primarily interested in modeling) and that of the wall and other structural elements of the building that indirectly affect the air temperature. In fact, walls can retain quite a bit of heat that continue to affect the air temperature after other sources of energy, such as the solar irradiance or the HVAC system, stop being active.

A model that presumably overcomes this defect is one that uses two capacitors, one of lower value for the air and one of higher value for the walls. Consequently it requires two resistances.

One choice of the 2R2C model is the following 2nd order coupled ODE, which is schematically illustrated in Figure 2:

$$C_z\dot{T}_z = \frac{T_w(t) - T_z(t)}{R_z} + A^{\text{eff}}\eta^{\text{sol}}(t) + q_{\text{x}}(t) + q_{\text{hvac}}(t) \tag{2}$$

$$C_w\dot{T}_w = \frac{T_{oa} - T_w}{R_w} + \frac{T_z - T_w}{R_z} \tag{3}$$

where C_z, C_w, R_z, R_w are the thermal capacitances and resistances of the zone and wall, respectively.

Notice that T_w, the so-called "wall temperature", is a fictitious state; it cannot be measured by any sensor.

I said "a choice" above since there is at least one other choice, which has to do with how the solar irradiance affects the indoor air temperature. It is tacitly assumed in the model above that the heat gain from Sun's rays directly affect the air temperature. This might be justified if there is a large window.

However, even if there is a large window, one can argue that the the air molecules in the path of the Sun's rays cannot absorb the heat from those rays; they just don't have the surface area needed to be in the path of the rays. According to this argument, the rays mostly passes through the window (except for the part reflected back) and hits the floor or walls(s). And, the air does not absorb the heat. Rather, the floor/wall(s) absorb the heat in the Sun's rays, and then release them to the air through a combination of conduction, convection and radiation. If you agree with this argument, the heat gain from solar irradiance (the $A^{\text{eff}}\eta^{\text{sol}}(t)$ term) should act as an input to the "wall node" in the RC network model. The model then becomes:

$$C_z\dot{T}_z = \frac{T_w(t) - T_z(t)}{R_z} + q_{\text{x}}(t) + q_{\text{hvac}}(t) \tag{4}$$

$$C_w\dot{T}_w = \frac{T_{oa} - T_w}{R_w} + \frac{T_z - T_w}{R_z} + A^{\text{eff}}\eta^{\text{sol}}(t) \tag{5}$$

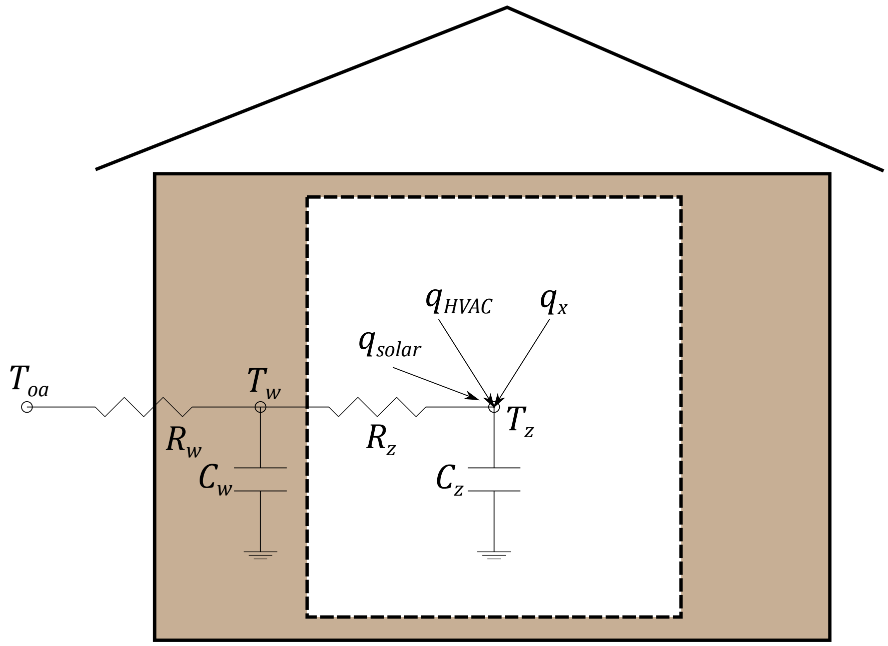

Fig. 2: 2R-2C model of a building's thermal dynamics

A schematic for this model is shown in Figure 3.

Of course, if you are not sure of either argument, you can use a third form of the 2R-2C model in which part of the solar irradiance affects the room temperature directly and another part indirectly. The model becomes:

$$C_z \dot{T}_z = \frac{T_w(t) - T_z(t)}{R_z} + A_a^{\text{eff}} \eta^{\text{sol}}(t) + q_{\text{x}}(t) + q_{\text{hvac}}(t) \tag{6}$$

$$C_w \dot{T}_w = \frac{T_{oa} - T_w}{R_w} + \frac{T_z - T_w}{R_z} + A_w^{\text{eff}} \eta^{\text{sol}}(t) \tag{7}$$

where now you have two effective area parameters, A_a^{eff} and A_w^{eff}, for "air" and "wall", instead of one.

C. *nR-mC?*

Yes, why not? You can use many more R's and C's, presumably to make the model "more accurate". Even after several decades of research, this question has not been resolved: it is still not clear whether 1R1C is enough or do we need 2R2C, or even nR-mC models, or for what type of buildings. Earlier papers used to go for many R-many C type models, but more recent papers have bucked the trend. The paper [Gouda et al., 2002] argues for a smaller number of R's and C's, though still higher than 2R-2C for a multi-zone building. A recent paper argues that - at least for residential buildings - 1R1C is enough [Fux et al., 2014].

III. Inputs and outputs for system identification - version 1

The indoor zone temperature T_z is affected by three *known* inputs: (1) the heat gain from the HVAC system, q_{hvac} (Watt or kW), (2) the outside air temperature T_{oa} (°C), (3) the solar irradiance η^{sol}(kW/m^2), and an *unknown* input: (a) the exogenous internal heat gain q_{x} (kW). Recall that the exogenous internal heat gain is due to occupants, at least partially: it is due to heat generated by the occupants and lights and equipment used by them.

So, the known/measurable input is $u(t) := [q_{\text{hvac}}(t), T_{oa}(t), \eta^{\text{sol}}(t)]^T \in \mathbb{R}^3$ and the unknown input (disturbance) is $w(t) = [q_{\text{x}}(t)] \in \mathbb{R}$. The measurable output is the indoor temperature T_z, so $y(t) = [T_z(t)] \in \mathbb{R}$.

We can express the ODE corresponding to a 2R2C model in the following state-space form for future convenience. We are using the model with "solar heat gain added to air node" version, but you can use the other model structures

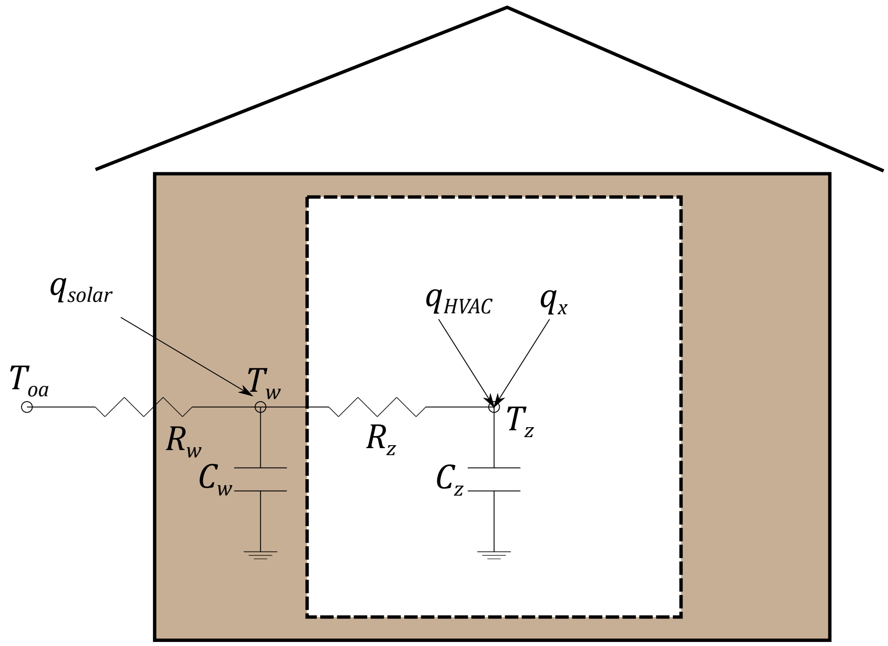

Fig. 3: 2R-2C model of a building's thermal dynamics

too.

$$
\begin{aligned}
\begin{bmatrix} \dot{T}_z \\ \dot{T}_w \end{bmatrix} &= \begin{bmatrix} \frac{-1}{C_z}(\frac{1}{R} + \frac{1}{R}) & \frac{1}{RC_z} \\ \frac{1}{RC_w} & \frac{-1}{RC_w} \end{bmatrix} \begin{bmatrix} T_z \\ T_w \end{bmatrix} + \begin{bmatrix} 0 & \frac{A_e}{C_w} & \frac{1}{C_z} \\ \frac{1}{RC_z} & 0 & 0 \end{bmatrix} \begin{bmatrix} T_a \\ \eta^{\text{sol}} \\ q_{\text{hvac}} \end{bmatrix} + \begin{bmatrix} \frac{1}{C_z} \\ 0 \end{bmatrix} q_{\text{x}} \\
T_z &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} T_z \\ T_w \end{bmatrix}
\end{aligned} \tag{8}
$$

A. *Measuring q_{hvac} and a comment on control*

The model structures defined above are not useful for control since the control input q_{hvac} cannot be directly commanded. For residential buildings you can only turn an air conditioner unit on or off[1]. Once on, the amount of heat injected by the unit depends on the indoor and outdoor temperatures and humidities. For commercial buildings with variable air volume (VAV) HVAC systems, you can control the rate of air flow and the temperature plus humidity of the air flow into the building, which together with the indoor zone temperature, determine q_{hvac}.

However, all we need are measurements of q_{hvac} for the purpose of system identification, i.e., estimating the parameters in the model. The next section describes how this quantity is computed from measurements of other variables that can in fact be controlled. Reading that section will provide you with some understanding of how HVAC systems of modern commercial buildings work.

Another reason the next section is included is that in the discussions of thermal dynamics so far, we have ignored a very important aspect: humidity. Both the model structure, the state dimension and the inputs and outputs change if humidity is incorporated. The model becomes nonlinear as well. Identification of such a model, along with its unknown disturbances, has not been attempted yet, but is a worthy goal.

IV. HVAC SYSTEMS OF COMMERCIAL BUILDINGS, AND MEASURING q_{hvac}

We are going to limit ourselves to the so-called variable-air-volume HVAC systems in which the rate of air flow can be varied. Older constant-air-volume systems are gradually being phased out.

For the sake of simplicity, we are also going to limit to single-zone buildings. A zone is a part of a building whose temperature is typically controlled independently of other zones, and has a air distribution system on its own.

[1]Unless your home has a fancy variable speed units, but they are rare and are truly variable speed.

Therefor, in this discussion the entire interior of the building is assumed to be a single space whose climate is controlled by a heating, ventilation, and air conditioning (HVAC) system. The building and its HVAC system configuration is shown in Figure 4. The HVAC system cools and possibly dehumidifies the air before supplying it to the building's interior. This cooling and dehumidification is performed by the cooling coil, which uses chilled water that is produced at a chilled water plant located outside of the building. Sometimes the air has to be cooled to a very low temperature in order to dehumidify it (so that water vapor in it condenses out), so a heating coil is used to heat the air up before being supplied to the zone.

The subscripts RA, MA, CA, DA, EA stand for return air, mixed air, conditioned air, discharge air, exit air, and correspond to locations in the HVAC system that are shown in the figure. Sometimes we will use the superscripts ma and da, which will stand for "moist air" and "dry air", the latter meaning without the water vapor. I am using H_20 for water vapor (both sub- and super-script). Beware: da for dry air is not to be confused with DA for "discharge air".

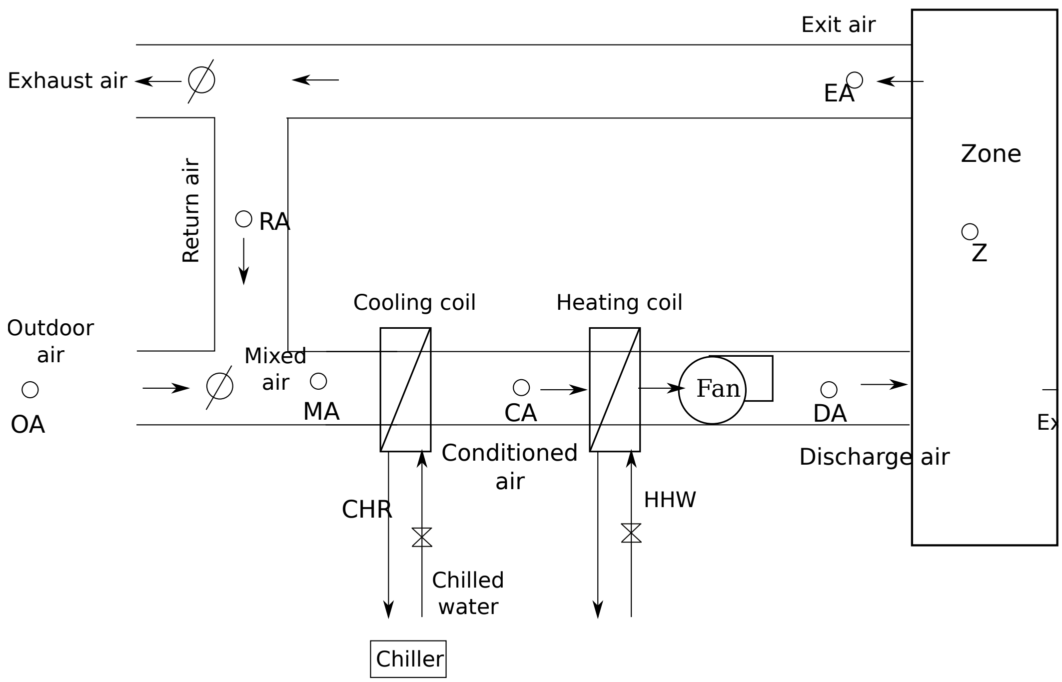

Fig. 4: A typical layout of a VAV HVAC system for a single zone. Airflow rate is controlled by controlling the fan speed. The humidity of the supply air is controlled by controlling the conditioned air temperature (after the cooling coil), which is controlled by varying the chilled water flow rate. (CHR: chilled water return, CHS: chilled water supply, HHW: heating hot water). That air may be too cold for the zone in certain times, in which case it is reheated by using the heating coil. So the temperature of the discharge air is affected by both the cooling and heating coil operation.

A. *Computing* q_{hvac}

First, a note on humidity. The *humidity ratio* of air, denoted by W, is the ratio of mass of water vapor to mass of dry air. It is related to the relative humidity, which is denoted by RH, but they are not the same. The value of the humidity ratio is typically in the range of 4-12 gm of water vapor per Kg of dry air. The reason we use humidity ratio and not relative humidity is that humidity ratio only depends on the amount of water vapor while the relative humidity is also affected by temperature.

Second, a definition. The enthalpy of one unit mass of (moist) air at temperature T and humidity ratio W, measured in kW/kg, is

$$h(T, W) = M^{da} C_{pa} T + M^{H_2O}(h_g + C_{pw} T) \tag{9}$$

where M^{da}, M^{H_2O} are the masses of the dry air and water vapor within the unit mass of moist air, so that $M^{da} + M^{H_2O} = 1$, and C_{pa}, C_{pw} are the specific heat of air and water at constant pressure ($kW/kg/Kelvin$) and h_g is the specific latent heat of evaporation of water (kJ/kg). The quantity h on the left hand side of (9) is called the *specific enthalpy* of moist air.

Recall that we use M to denote the mass (kg) of a certain fixed volume of air (dry or moist), while m is used to denote (mass) flow rate (kg/s).

The rate of heat injected into the zone/building by the HVAC system is equal to rate of total change in enthalpy of the air as it changes from supply air to leaving air (leaving air = air leaving the zone, same as return air in Figure 4). Here we make a reasonable assumption, that the temperature and humidity of the air in the zone is the same as those of the air leaving the zone, which is essentially the same as assuming the air inside the zone is well-mixed. Then, we get

$$q_{\text{hvac}}(t) = m_a(h(T_{\text{DA}}, W_{\text{DA}}) - h(T_z, W_z)). \tag{10}$$

As long as we have measurements of the temperature and humidity of the discharge air, the temperature and humidity of the zone air, and of the flow rate of discharge air, q_{hvac} can be computed from (10).

Sometimes a simplification/approximation can be made, which is particularly useful if humidity measurements are not available. Suppose the change in the humidity ratio of the air as it passes through the zone is negligible, which can happen if the zone generates or absorbs very little water vapor. In that case, $W_{\text{DA}} = W_z$[2]. Then (10) simplifies to

$$q_{hvac} := m_{\text{DA}} C_{pa}(T_{\text{DA}} - T_z), \tag{11}$$

which only requires flow rate and temperature measurements to compute.

V. Non linear RC model

A. Control-oriented model, without humidity

See (11): both m_{DA} and T_{DA} can be used as control commands; they can can be controlled using the fan and the cooling/heating coils. But q_{hvac} cannot be used as a control command because of its dependence on the state T_z. So a *control oriented model* should have m_{DA} and T_{DA} as inputs, not q_{hvac}.

We can still use one of the 2R2C models, but replace q_{hvac} by its expression that bring the controllable inputs explicitly into the model, such as:

$$\begin{aligned} C_z\dot{T}_z &= \frac{T_w - T_z}{R_z} + A^{\text{eff}}\eta^{\text{sol}} + q_{\text{x}} + m_{\text{DA}}C_{pa}(T_{\text{DA}} - T_z) \\ C_w\dot{T}_w &= \frac{T_{oa} - T_w}{R_w} + \frac{T_z - T_w}{R_z} \end{aligned} \tag{12}$$

which is a non-linear[3] ODE due to $m_{\text{DA}}C_{pa}(T_{\text{DA}} - T_z)$.

1) Inputs and outputs for system identification - version 2: For the control oriented model described above, the inputs to the system are: (1) the flow rate of air supplied to the zone, m_{DA}(kg/s), (2) the (dry bulb)[4] temperature of the air supplied to the zone, T_{DA}(°C), (3) the outside air temperature T_{oa} (°C), (4) the solar irradiance η^{sol}(kW/m^2), and an *unknown* input: (a) sensible heat gain q_{x} (kW).

So, the known input vector is $u(t) := [m_{\text{DA}}(t), T_{\text{DA}}(t), T_{oa}(t), \eta^{\text{sol}}(t)]^T \in \mathbb{R}^4$. The unknown disturbance, the states and the measured outputs stay the same as in "Version 1".

The identification problem for the control-oriented model (12) is to estimate the parameters R's, C's, and A^{eff}'s, and of the signal $q_{\text{x}}(t)$ from measurements of $u(t)$'s and $y(t)$'s.

B. Control-oriented model with humidity

If the effect of humidity is retained in the model, more non-linear terms have to be introduced. In particular, the expression for q_{hvac} from (10) have to be used in the right hand side of the ODE $C\dot{T}_z = \dots$ instead of the approximation (11). In addition, at least one more state has to be introduced: the zone humidity itself. If you are interested in the gory details, you can find them in [Goyal and Barooah, 2012]. For the moment let us agree to ignore humidity.

[2]This is another reason we do not use relative humidity in the model. Even if there is very little change in water vapor, the relative humidity will change since it depends on temperature as well. So we cannot make simplifying assumptions with the relative humidity.

[3]It is also called a bilinear model since the nonlinearity involves product of a state (T_z) and an input (m_{DA}).

[4]See https://en.wikipedia.org/wiki/Dry-bulb_temperature

VI. Other model structures

The RC-network models described above are *gray-box models* since the model structure is fixed and the model is completely specified by a few parameters: the R's, C's, and the A^{eff}'s, that may be unknown. Of course, the disturbance signal is not parameterized, it is an arbitrary signal.

Such models have been popular since only a small number of parameters are needed to specify the model, and more importantly, these parameters have physical meaning. For instance, the resistance and capacitance of a building's structure can be computed from the tables of material properties and knowledge of the building's geometry. Although a building's properties are going to vary from such estimates because of myriad factors such as air leakage, they provide a useful starting point. For instance, suppose an architect computes the R's and C's of a building's structure from material properties and geometry information. If you then estimate the R's and C's from measurements of inputs and outputs using a system identification algorithm, you can compare the two. If your estimates are within an order of magnitude of the architect's calculations, both of you can be confident that your calculations have passed a sanity check.

However, one weakness of these models is that the problem of identifying the parameters from measurement, the parametric system identification problem, is non-convex. That means that the numerical optimization algorithms used in system identification many find a local minimum as the solution, which can vary depending on the initial guess.

A possible way to overcome this difficulty is to avoid the RC network models altogether, and assume a generic "black-box" model structure and pose the identification problem in such a way that a convex optimization problem results. The downside is that the model parameters now do not have any physical meaning; no sanity check can be performed.

References

[Allen, 1980] Allen, E. (1980). *How buildings work: the natural order of architecture*. Oxford University Press.

[American Society of Heating, Refrigerating and Air Conditioning Engineers, 2005] American Society of Heating, Refrigerating and Air Conditioning Engineers (2005). The ASHRAE handbook fundamentals (SI Edition).

[Fux et al., 2014] Fux, S. F., Ashouri, A., Benz, M. J., and Guzzella, L. (2014). EKF based self-adaptive thermal model for a passive house. *Energy and Buildings*, 68, Part C:811 – 817.

[Gouda et al., 2002] Gouda, M., Danaher, S., and Underwood, C. (2002). Building thermal model reduction using nonlinear constrained optimization. *Building and Environment*, 37:1255–1265.

[Goyal and Barooah, 2012] Goyal, S. and Barooah, P. (2012). A method for model-reduction of nonlinear building thermal dynamics of multi-zone buildings. *Energy and Buildings*, 47:332–340.

[R.Kramer et al., 2012] R.Kramer, Schijndel, J., and H.Schellen (2012). Simplified thermal and hygric building models: a literature review. *Frontiers of Archetectural Research*, 1:318325.