CENTER FOR ENVIRONMENTAL BUILDING AND DESIGN

White box
Black box
Gray box

Modelling of Building Energy Use

Final Report
Daikin Open Innovation Lab, Silicon Valley

April 4, 2021 rev.1

White box, Black box, GRAY-BOX modelling of building energy use: An overview

William Braham, Ph.D., FAIA Director, CEBD

White Box Modelling

Yun Kyu Yi, PhD Assistant Professor, UIUC

Low order white box modelling technique

Pengyuan Shen, PhD Assistant Professor, Harbin Institute of Technology

Blackbox Modelling

Ravi Srinivasan, PhD Associate Professor, University of Florida

Black Box and HVAC Controls: Measuring the Right Factors

Nancy Ma PhD student, CEBD

Application of Black & White Energy Models @ UPenn
Deriving actionable information from available data

Alex Waegel,PhD Research Associate, CEBD

Creating an RC thermal model

Max Hakkarainen PhD student, CEBD

White box, Black box, GRAY-BOX modelling of building energy use: An overview

William W. Braham, Center for Environmental Building and Design

Contents

Introduction

Architects use multiple kinds of models to inform multiple stages their work. These include physical models, full-size and scaled, analytical, mathematical models, and computational models, commonly called simulation (Weisberg 2013).

In our terminology, white box models are deterministic, physics-based models solved with numerical techniques. They are widely used in the design and analysis of buildings. Black box models are stochastic models analyzed with statistical and machine-learning techniques and are most commonly used for the analysis of limited data streams for example from thermostats or meters. Gray box models combine a deterministic model with factors to account for the stochasticity of data and are solved with a variety of techniques.

This paper will briefly review the use of white and black models, which are explored in more detail in the sections by Ravi Srinivasan, Pengyuan Shen, and Nancy Ma. The body of the paper will review gray-box methods as they have been applied to buildings

White Box Models

White box models are deterministic, physics-based models solved with numerical techniques. They are widely used in the design and analysis of buildings. These are commonly divided into categories by level of complexity: single zone, multiple zone, and computational fluid dynamic models (CFD). The building physics for building energy modelling has been well established since the early 20th century, so the limitation has been computational power and efficiency. A variety of techniques were employed in manual calculations, but with the advent of ready computational power and the increased urgency of the energy crisis in the 1970s, the heat balance method became the dominant approach. For a more detailed history of these methods, see (Oh and Haberl 2016) and (Malkawi and Augenbroe 2004).

The dominant whole-building, multi-zone model at present is EnergyPlus, which has been developed and supported by the Department of Energy (U.S. DOE 2020), who provide the computational engine used by many different 3rd party interfaces. EnergyPlus is used by many different kinds of users for many different kinds of purposes from architectural and engineering design to code compliance, policy analysis, research, and measurement and verification. Its strength is its comprehensive nature, allowing the evaluation of small variations on almost any component or condition. The challenge is its complexity, which becomes especially evident when it is used to model the performance of an existing building, in which the process of calibrating the model can be time-consuming and uncertain. Simplified or low-order white-box models make clearer what causes their behavior and are much easier to calibrate (Shen, Braham, and Yi 2018) (ISO 52016-1 2017).

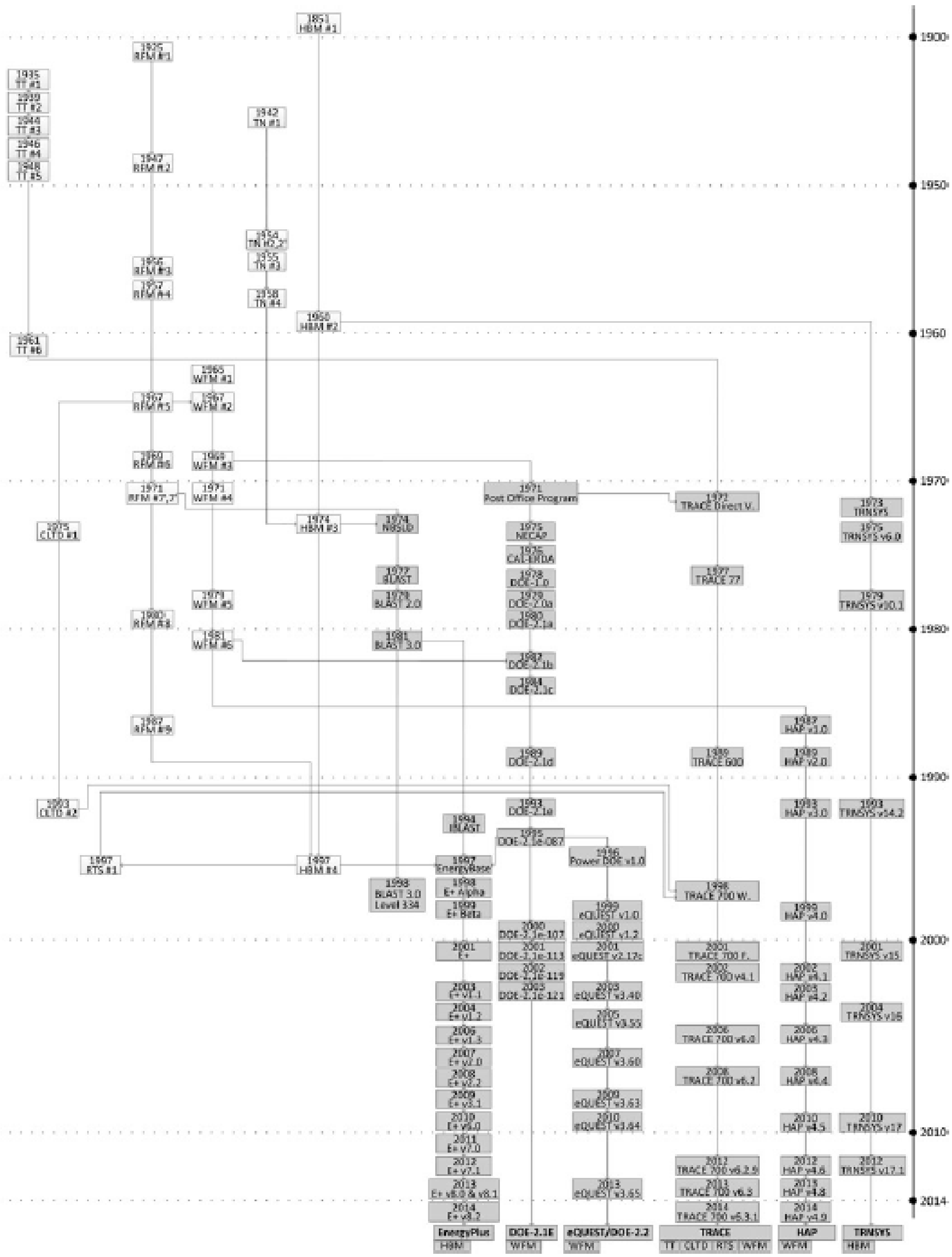


Figure 1. Genealogy chart for whole building, white-box simulation programs (Sukoon 2016)

Black Box Models

Black box models are stochastic models analyzed with statistical and machine-learning techniques and for building energy analysis are most commonly used for the analysis of limited data streams, for example from thermostats or utility meters. Machine learning techniques can also be used to optimize or calibrate white box models.

Gray Box Models

Gray box models combine a deterministic model with factors to account for the stochasticity of data and are solved with a variety of techniques. As Kissock et al explained, "In the building energy community, models derived from measured energy use are called 'inverse' models. The term 'inverse' differentiates them from 'forward' models in which building energy use is predicted from engineering principles (Kissock, Haberl, and Claridge 2003, 2002)." Inverse methods are also called "estimation and system identification" and are used to identify models that provide a good fit to data and whose parameters also correlate to some physical aspect of the building (Rabl 1988).

Correlations between outdoor temperature and energy use have been in use since the early twentieth century, initially used to time the delivery of fuels to homes, and correlation parameter was indication of the temperature driven heat loss of the building (1906). Inverse or gray-box methods were intensely explored in the 1970s, inspired by the energy crisis and the desire to accurately determine energy consumption and savings. They can be broadly divided into steady-state and dynamic models.

Steady State Models

Steady-state models are simpler and are generally employed with the average daily, monthly, or yearly data that is more commonly available.

Princeton CES: Twin Rivers Program

Among the earliest studies of gray box modelling emerged from the Center for Environmental Studies at Princeton University, which had been engaged in 1972 by the newly formed Energy Research and Development Administration "to document, to model, and to learn how to modify the amount of energy used in homes (Socolow 1976)." Their focus was on winter heating, which was the largest residential use of energy. The monitored a group of 48 identical rowhouses in Twin Rivers, New Jersey and in the process developed or refined many of the auditing techniques still used in energy analysis today. Simple modelling was readily adopted, because as Socolow explained, "Winter gas consumption is strongly predicted by a linear relation involving one single independent variable: average outside temperature." They began their analysis with the simplest model linearly relating gas consumption to the outside temperature and a reference temperature, which is now commonly called the balance point and represents the temperature at which the furnace needs to turn on.

$$Gas\ Consumption = B(Reference\ Temp. - Outside\ Temp.)$$

They obtained good fit of gas consumption with temperature over periods of months, but for individual houses they discovered many variations attributable to differences in orientation (sun and wind), to differences in construction affecting heat loss and air infiltration, and differences in occupancy and operation. The team devoted considerable effort to identifying the additional independent variables that could be measured and could explain the variations. They focused on solar gains, internal electric heat gains, and wind velocity applied in a sequence of increasingly complex models and more detailed data gathering.

> Certainly, the models in Chapter IV are an improvement on degree-day models with fixed reference temperature. But it must be possible to advance the state of the art further, while sacrificing only a little of the simplicity and economy of the current models. It remains our conviction that the way to further progress is not by the back door of elaborate, costly, and highly deterministic computer models drawn from the world of office buildings with fixed usage patterns, that track the weather hour by hour through the year. Rather, it may well lie in the direction of identifying those few parameters that capture the gross features of the energy balance of a house (its "signature") and then finding simple field approaches to measure their numerical values (Socolow 1976).

The use of meaningful signatures remains a popular approach and is widely used in government programs and businesses focused on energy reduction. Energy consumption normalized per unit area (kBTU/sf and kWh/m^2) is widely used as benchmarking and improvement metric, while two of what we might call “challenge” measures—air infiltration rate under pressure and cool down time—provide valuable indicators without detailed modelling.

PRISM: Variable-Base Degree-Day Models

Through the early 1980s, the Princeton group continued to develop simple methods that could be applied with commonly available data. They refined the use of a “degree-day” method that used records of daily energy consumption and average daily temperatures, with which the numbers of degree-days of difference from a reference temperature could be calculated. They designated the method PRinceton Scorekeeping Method (PRISM) and it was widely used as a benchmarking tool and to evaluate the effectiveness of energy conservation measures in heating dominated buildings. It was subsequently applied to commercial buildings, with some limited success. It is called a variable-base method because it relies on a best-fit estimate of the reference temperature, τ:

$$\tau = \frac{Q_{int}}{Lossiness}$$

In effect, it divides energy use into two categories, heating and everything else, fitting a line to the heating portion of the usage and assuming that everything else is constant through the year. As Fels observed, “for climates in which the energy used for cooling rather than heating dominates, and for houses with a large solar component in their design, more research is needed (Fels 1986).” Moreover, as Kissock et al argued, the “linear two-parameter regression models fail to capture the non-linear relationship between heating and cooling energy use and ambient temperature caused by system effects, such as VAV control, or latent loads (Kissock, Reddy, and Claridge 1998).”

Change Point Models: Inverse Modelling Toolkit (IMT)

Better results were obtained by using models with more parameters, fitting lines to portions of the data and determining the “change point” between the behaviors iteratively. These models were sufficiently useful for more complex buildings that ASHRAE commissioned an “Inverse Modelling Toolkit” from Kissock, Haberl, and Claridge, which was published with software in 2002 (Kissock, Haberl, and Claridge 2003). The three-, four-, and five-parameter models were able to capture the energy consumption behavior of buildings with more complex systems with variable or non-linear components. The method was developed for daily temperature and energy measurements.

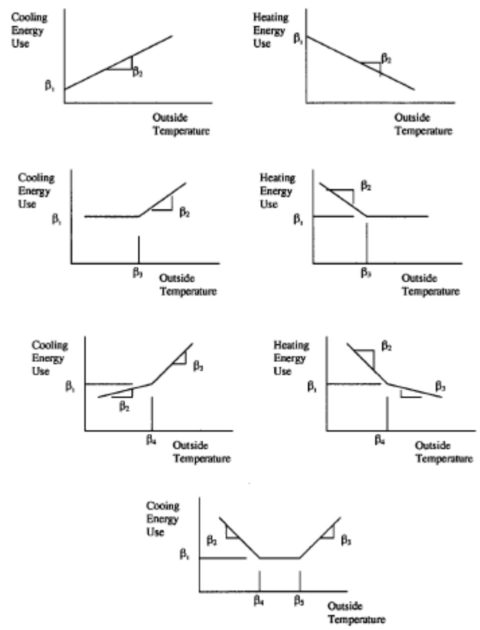


Figure 2. Change Point Models. Top row, 2 parameter heating or cooling, second row, 3 parameter heating and cooling, third row, 4 parameter, heating and cooling, and bottom row, 5 parameter, heating and cooling (Kissock et al, 2002)

FirstView

FirstView has similar ambitions to the Inverse Modelling Toolkit (IMT), which is to provide performance assessment and analysis of buildings using readily available data. Beginning in the early 1990s, Howard Reichmuth began working with monthly average temperatures and monthly energy use, which is regularly reported in utility billing, so required no special metering. Unlike the regression techniques in IMT, it uses a variety of assumptions about characteristic heat loads and gains to build a physics-based model that is fit to the data. It is effectively an elaboration of the degree-day “bin methods” that correlated monthly temperature distributions to monthly energy, and also draws on the load breakdowns revealed with hourly simulation models. Reichmuth began producing “Howdy charts” of discrete energy gains and losses showing the patterns that were eventually designated as energy signatures (White and Reichmuth 1996).

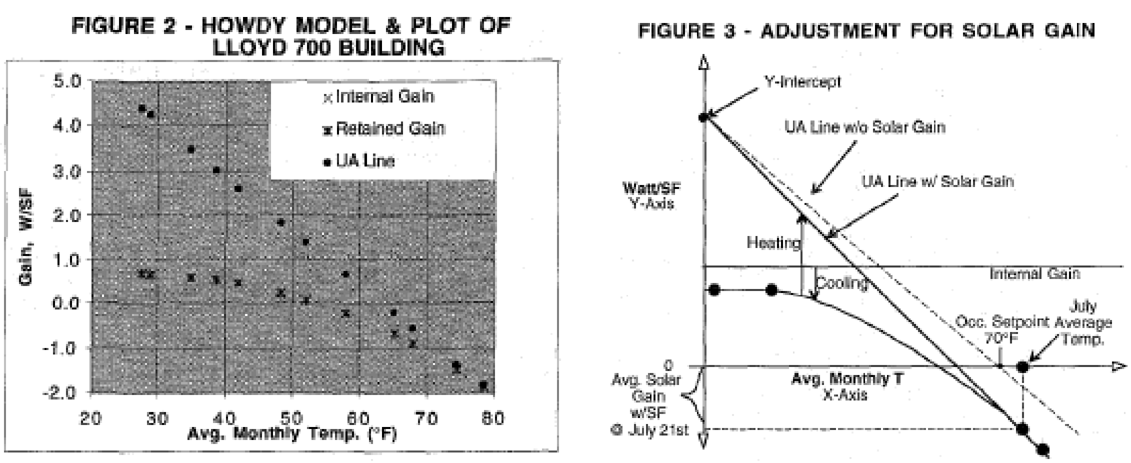


Figure 3. Howdy models of monthly energy data (Reichmuth 1996)

The method is similar to the calibration of low-order white-box models, adjusting a limited set of building parameters to fit the model to the data. The impressive aspect of the method is that good fits have been obtained with monthly data for a large set of commercial buildings, mostly office buildings (Robison and Reichmuth 2001). The use of analog building parameters also facilitates recommendations for the improvement of energy performance.

Beginning in the 2000s, the method was adopted by the New Building Institute (NBI), who were able to validate it against a larger pool of building and support its further refinement (Reichmuth and Turner 2010). NBI is actively using the tool and have been promoting it to EPA for use with the EnergyStar tool.

Table 1. Equivalent Analog Building Parameters

Parameter, symbol	Units	Notes
Internal Gain, Qin	W/ft2	Solved
External Energy, Qext	W/ft2	Fixed ratio of internal gain
Aggregate Normalized UA, UAn	BTU/deg F-hr-ft2	Solved
Heating Efficiency, Eh	No units	Assumed to be 0.75
Cooling Efficiency, COP	No units	Solved
Service Water Heating, SWH	Gal/day/ft2	Solved
Heat Intercept, Ht	T deg F	Solved
Cool intercept, Ct	T deg F	Solved

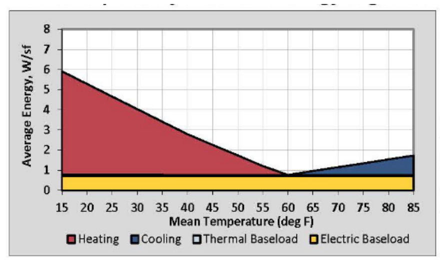


Figure 4. FirstView, equivalent analog building parameters and analysis (Reichmuth 2010)

Dynamic Models

Inverse models become dynamic with the additional of some measure of internal heat storage. As Rabl observed, “there are many situations where dynamic models are preferable or required: warmup and cooldown; peak loads; rapid monitoring; diagnostics; and optimal control (1988).” Dynamic models can also incorporate more heat transfer pathways and methods of solution (linear and non-linear), facilitating the identification of suitable models, but increasing the complexity and computational intensity.

The experiments with different dynamic models have revealed two kinds of issues—building parameters that vary through time and correlation between variables that are not entirely independent. Principle among the varying parameters are the air exchange rate and the admittance of solar radiation. A parameter that can be problematic in drier climates is the difference between the temperatures of air, sky, and ground, which can have quite different spatial and temporal patterns.

This review will focus on the different models that have been tested, with some discussion of the methods of solution. There is also a difference in the literature between gray-box models that are tested against the results from complex white box models, on the argument that they have no noise, so are better suited to evaluate the models themself. That is in contrast to models tested against measured data from real buildings, which have to account for the noisiness of the real world, including environmental and building parameter variation, as well as the noise inherent in measurement and data collection.

Thermal Time Constant

Principle among the dynamic properties of a building is the thermal time constant, which is a measure of the rate at which a building or thermal system warms up or cools down. Even with all the focus on insulation values in construction, early researchers like Andreas Bugge considered it one of the fundamental characteristics of a building (Bugge 1924). He recognized that it interacted in ways that complemented insulation values and would be of importance for understanding the effect of variable heat sources, such as the sun. That instinct was confirmed by the solar researchers of the 1940s and 1970s, who devoted a great deal of time to the incorporation of thermal mass in interior construction and learning to determine its effects (Barber 2016, Balcomb 1982).

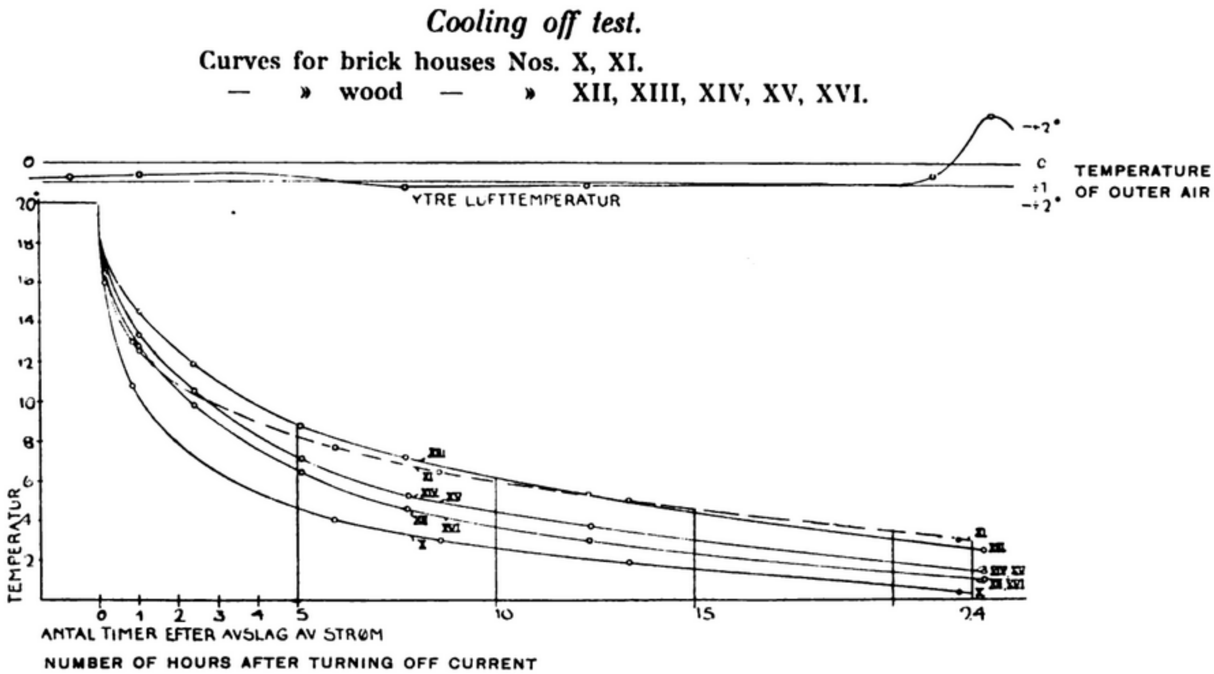


Fig. 46.

Figure 5. Cool down tests for Trondheim test houses (Bugge 1924)

In real construction, there are multiple time constants within a building and thermal properties are distributed, but for lumped-parameter, RC models there is a thermal time constant that typically forms one of the characteristic parameters of the model and its solutions.

The simplest example comes from Newtons law of cooling, which applies to the cooling of a thermal mass by some linear from of heat exchange that is proportional to the temperature difference. That linearity is generally true for forced convection, but can vary with buoyancy driven convection and only applies for small temperature differences in radiant exchange. The basic heat balance expression is:

$$\boldsymbol{\rho C_p V \frac{dT}{dt} = hA(T(t) - T_a)}$$

Where:

ρ	= Density, kg/m^3
C_p	= Specific Heat, J/kg ºC
V	= Volume, m^3
$\rho C_p V$	= Thermal mass, J/ºC
h	= Convection heat transfer coefficient (W/m2·ºC)
A	= Heat transfer surface area (m^2)
T(t)	= Temperature of the solid (ºC)
T_a	= Temperature of the air surrounding the surface (ºC)

The solution to the differential equation has the following solution expressed as a function of the time constant:

$$T(t) = T_a + (T_0 - T_a)e^{-\frac{t}{\tau}}$$

Where the time constant, $\boldsymbol{\tau}$, is defined as the ratio between the thermal mass and the temperature dependent rate of heat loss, which has the units of time. The time constant is the time it takes the mass to cool off by 1/e in temperature.

$$\tau = \frac{\rho C_p V}{h A_s}$$

Taking the simple example of a masonry brick with an initial temperature of 50 C in a environment with a temperature of 20 C, the charts in Figure 6 show the cool down curves for time constants of 3 and 9 hours (very different shapes and sizes of brick!).

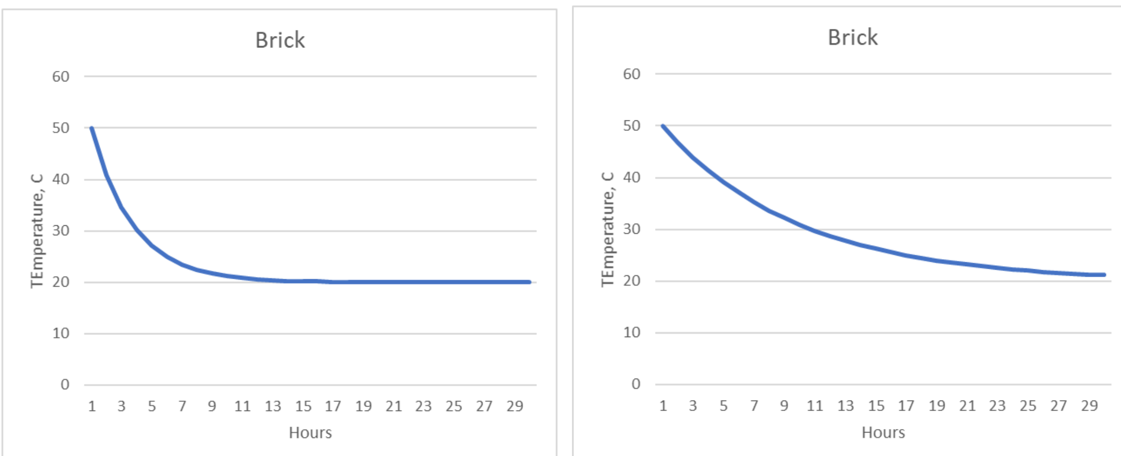


Figure 6. Cool down curves for bricks with different thermal time constants, 3 hr on the left and 9 hour on the right

Equivalent Thermal Parameters

The first of the dynamic gray-box or inverse modelling techniques developed in the Twin Rivers study was reported in Robert Sonderegger's dissertation (1977). He tested a variety of simple models, whose distinguishing feather was the inclusion of thermal storage elements. In the model that was most successful with the Twin Rivers townhouses, he also included a constant temperature "clamp" to account for the effect of the basement. Figure 2. The model is expressed in terms of measurable temperatures and energy inputs, and solved for interior temperature over time.

Using algebraic manipulation, he was able to reduce the heat balance to a single difference equation that could be used for linear regression.

$$\dot{V} = \left(\frac{H \cdot \dot{V}O + A \cdot \dot{S}}{HS + H + HS}\right) + \frac{H}{C^*}(VO - V) + \frac{HC}{C^*}(VC - V) + \frac{A}{C^*}S + \frac{E+P+L}{C^*}$$

Where V = temperature, H = lossiness, A = window area, S = solar flux, C = thermal mass, E+P+L are energy inputs

He obtained good fit with the model, and by precisely determining the heat gains (E+P+L) using electric heaters, he was able to solve for the "equivalent thermal parameters" of the building.

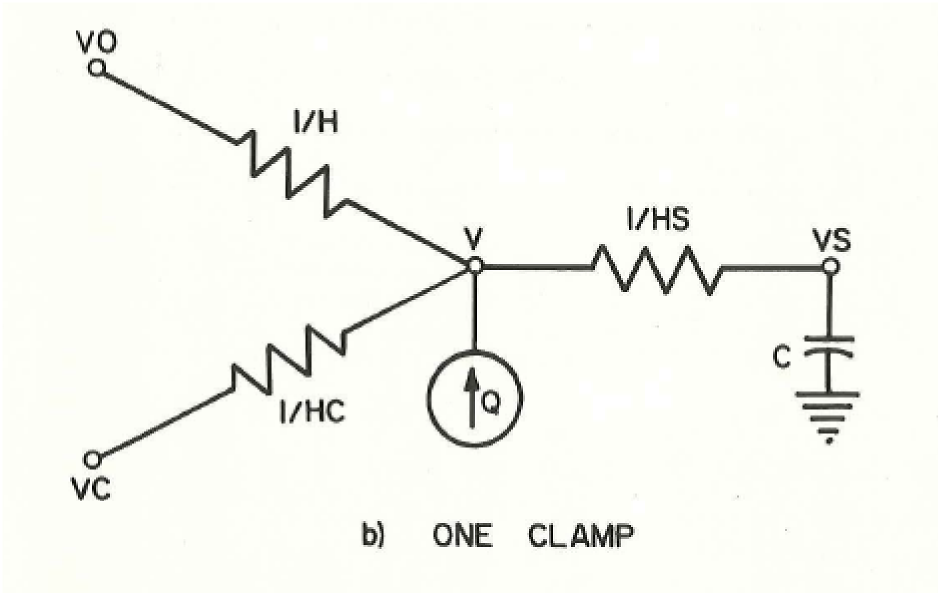


Figure 7. Equivalent RC model of Twin Rivers Townhouse, Sonderegger, 1977

The model and its solution also provided a measure for the thermal time constant of the whole system.

$$\boldsymbol{\tau = C(\frac{1}{HS} + \frac{1}{H + HC})}$$

For the Twin River Townhouse they were measuring, the time constant was 6.8 hours. The time constant also emphasizes the point that the thermal behavior of buildings is not the function of a single property, but of their relation in the building as a thermal system. As Bugge's chart illustrated, you can achieve the same cool down time with massive construction lightly insulation and highly insulated, but lightweight construction.

Enerplex

After the Twin Rivers program, the CES group was involved in the design, construction, and monitoring of two commercial buildings called Energyplex. A number of inverse techniques were used on the data collected from these buildings, from multiple solution techniques for differential heat-balance equations, to the use of an auto-regressive moving average (ARMA) method that yielded time constants and admittances, but without estimation of physical parameters (Subbarao 1985). Two useful features revealed by dynamic methods are the characteristic time constant of the building and the effective thermal mass (Rabl 1988, Norford et al. 1986).

b) DYNAMIC METHODS

	forward	inverse	comments
thermal network [Sonderegger 1977]	*	*	In forward direction no limit on complexity of network. For inverse problem network must be simple, with equival. thermal parameters.
response factor series [Stephenson and Mitalas 1967]	*		Tabulated results for building components [ASHRAE 1985] useful for calculation of peak loads.
Fourier analysis [Shurcliff 1984]	*	*	Calculates response to sinusoidal (constant plus diurnal) input. Can be combined with calculation in time domain.
ARMA model [Subbarao 1985]		*	Coefficients lack direct physical interpretation, but that can be provided with time constants and admittances.
BEVA [Subbarao 1985]	*	*	Combination ARMA + Fourier methods. Loads calculated in time domain.
modal analysis [Bacot et al. 1984]	*	*	Diagonalization of the differential equations for the building. For inverse problem building is appro- ximated by small number of modes.
differential equation [Eq.2.10 of this paper]		*	Approximates building by linear diff. eq. Order and coefficients adjusted by data. Can be integrated analytically. Much flexibility for fitting, prediction and control.
computer simulation (e.g. DOE 2.1, BLAST)	*	*	Very detailed. Potentially the most accurate method. Also models HVAC equipment. Requires much expertise and labor for coding the input.
hybrid methods	*	*	Computer simulation, plus diff.eq. or ARMA. To be developed.

Figure 8. Summary of dynamic methods in the later 1980s (Rabl)

Continuous Time Stochastic Models (CTSM)

Beginning in the late 1980s, the mathematician Heinrich Madsen began working with a variety of colleagues to apply more sophisticated statistical techniques to the heat dynamics of buildings (Madsen 1985). The focus was on accounting for the different forms of noise in time series data, so their projects were based on measured data from a variety of test buildings. In their first study (1983-1995), shown in Figure 7, they achieve very good fit of a 2R2C model. However, similar to the electric heaters introduced in Twin Rivers, they used an electric heater configured to produce heat in a white-noise pattern, which was explicitly independent of environmental factors (Madsen 1995).

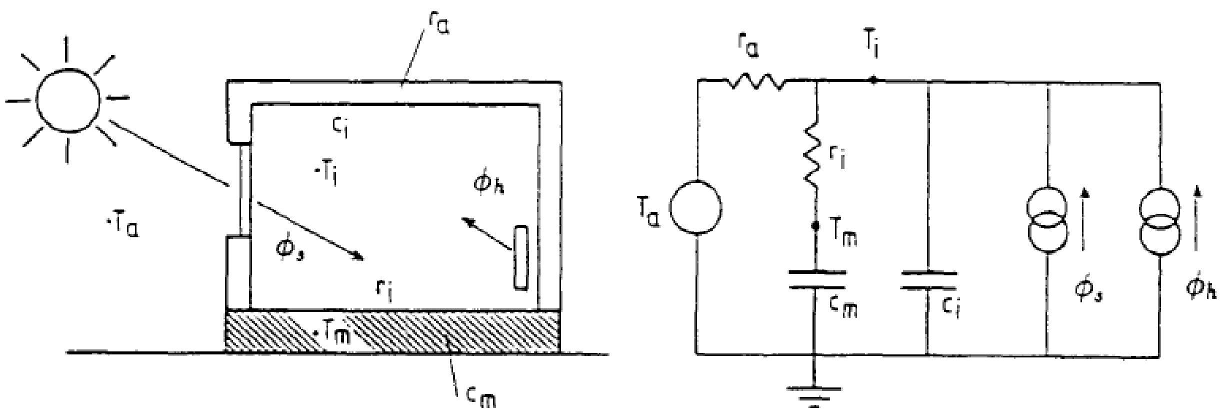
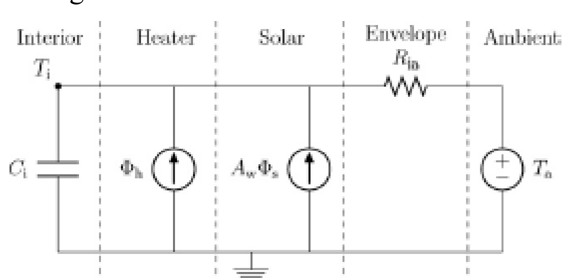


Figure 9. A 2R2C model of building (Madsen 1995)

In a subsequent project, they tested a hierarchy of models of increasing complexity on data from their test building, using a likelihood ratio test to identify the best fit. The heat input was again not driven by a thermostat, but used a “pseudo-random binary sequence (PRBS)” to excite different frequencies of thermal response in the building and to make the variable independent of other environmental factors. They tested a total of 17 models that increased in complexity from a 1R1C model to 6R5C model that even includes the capacitance of the sensor and a resistance for its connection to the indoor air temperature (Bacher and Madsen 2011). The one that best fit the data was a 4R4C model that included a mass and resistance for the sensor and the heater, as well as mass for the air and for the building envelope. See Figure 8.



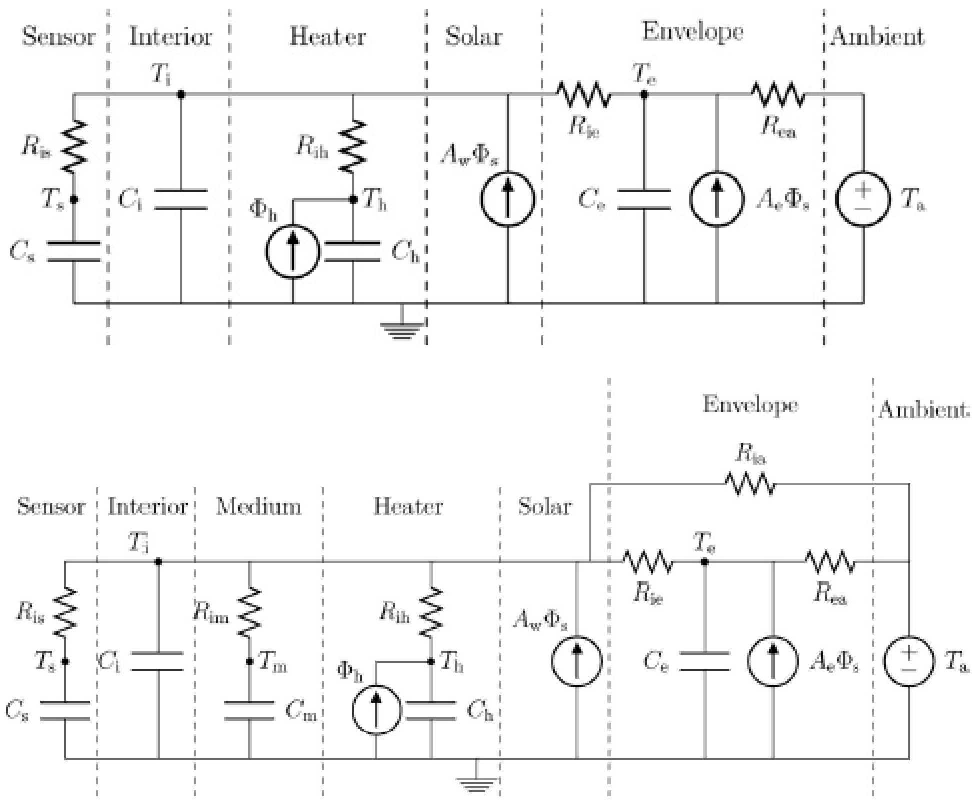


Figure 10. The simplest (top), most complex (bottom), and best fit (middle) of the RC models used in the CTSM study (Bacher 2011)

Of course, there are multiple time constants in more complex models. As Bacher and Madsen described in the matrix formulation of their models "the estimates of the time constants, $\boldsymbol{\tau_i}$ are calculated by the eigenvalues, $\boldsymbol{\lambda_i}$of the system matrix A, i.e., $\boldsymbol{\tau_i = 1 - \frac{1}{\lambda_i}}$ (Bacher and Madsen 2011)." These all have the same form of a ratio of thermal transmittance and thermal capacitance, and there is a time constant for each capacitor. The cool down time constant for the house would generally be that for the building mass, but this gets complicated in the models that separate the interior mass from the mass that is part of the exterior envelope. See Figure 11.

Table 3
The estimated parameters. The heat capacities, C_x, are in [kWh/°C]. The thermal resistances, R_{xx}, are in [°C/kW]. The areas, A_x, are in [m^2]. The time constants, τ_x, are in (h). Note that the physical interpretation for many of the parameters is different for each model.

	Model				
	Ti	*TiTh*	*TiTeTh*	*TiTeThTs*	*TiTeThTsWithAe*
C_i	2.07	1.36	1.07	0.143	0.0928
C_e	–	–	2.92	3.24	3.32
C_h	–	0.309	0.00139	0.321	0.889
C_s	–	–	–	0.619	0.0549
R_{ia}	5.29	5.31	–	–	–
R_{ie}	–	–	0.863	0.909	0.897
R_{ea}	–	–	4.54	4.47	4.38
R_{ih}	–	0.639	93.4	0.383	0.146
R_{is}	–	–	–	0.115	1.89
A_w	7.89	6.22	5.64	6.03	5.75
A_e	–	–	–	–	3.87
τ_1	10.9	0.16	0.129	0.0102	0.0102
τ_2	–	8.9	0.668	0.105	0.105
τ_3	–	–	18.4	0.786	0.788
τ_4	–	–	–	19.6	19.3

Figure 11. Thermal parameters, including time constant, for the 17 models tested by Bacher et al

Control Model

In a novel conference paper, McKinley and Alleyne combined an RC model of a building with a heat and moisture balance model of a simple HVAC system (McKinley and Alleyne 2008). They used data from a white-box, EnergyPlus model of a small commercial building, so there was no noise present in the data. They used a 4R2C thermal model and a standard optimization technique to achieve good fit. One interesting observation was about the choice of error function and the confounding of parameters. In their case they started by using the rms error of interior air, which confounded the heat gain through the envelope and through the windows, but when the switched to using the zone humidity ratio, they achieved much better results. It seems a very useful insight!

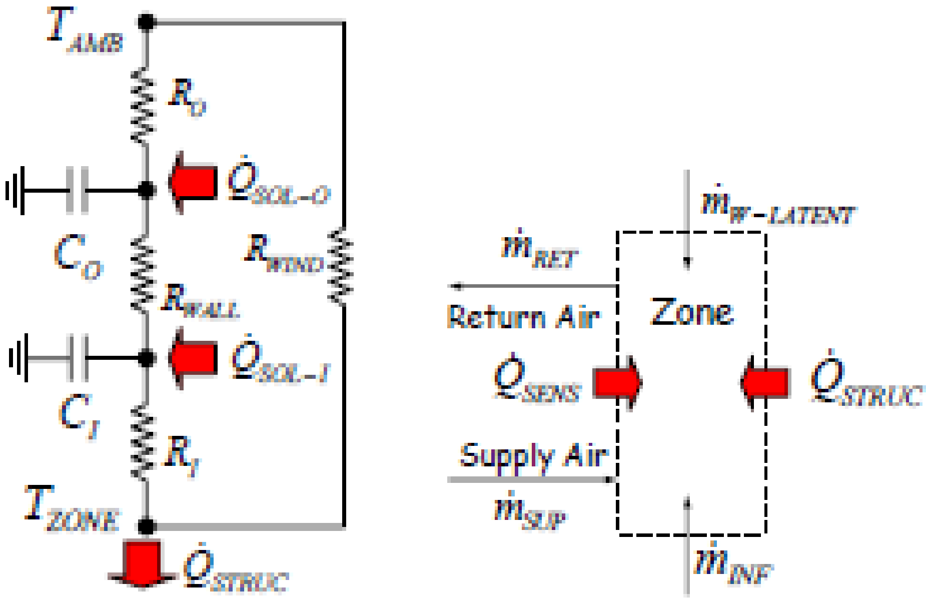
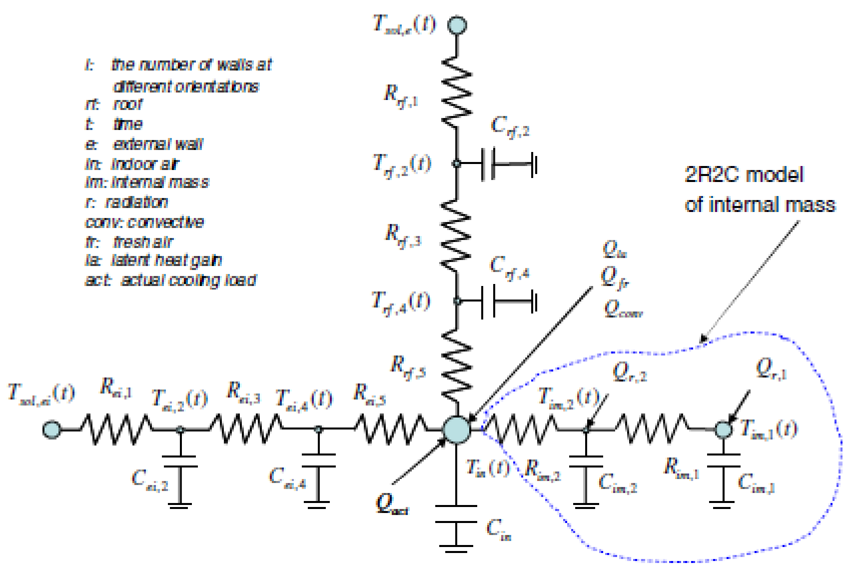


Figure 12. Thermal network model and zone control volume model (McKinley & Alleyne)

Genetic Algorithm

Wang and Shu developed a hybrid method for parameter estimate, using physical specifications to calculate the parameters of the 3R2C models of the envelope, but used inverse modelling of a 2R2C model for the internal mass, employing a genetic algorithm to optimize the fit to the data. They used the measured heat input to the building as the objective function, and achieve reasonable (though not perfect) results with the model.



Bath Group

Ramallo-González et al successfully tested lumped parameter models for heat transfer through building envelopes, getting good results with 3R2C models (Ramallo-González, Eames, and Coley 2013). In a subsequent study they tested a variety of whole building models against data for a diverse selection of residential buildings, settling on a simple, 2R1C model, which gave the best fit across the population. They used data for 1,000 simulations of 16 house types and 6 different types of actual buildings. Their goal was "to see if the restricted data gathered from advanced smart metres or similar devices might be used to form the basis of a dynamic thermal model of a building (Ramallo-González et al. 2018)."

One peculiarity of their study was that after testing for fit to interior temperature data, their focus was on estimating heat transfer coefficients (HTC)of the buildings, which they compared to those calculated from the EnergyPlus simulations. The also used the fit models to test their sensitivity to missing heat gain data,

They did achieve good results with the real buildings, that had some critical sensors added to increase precision, specifically: "Internal temperatures (in three locations per house), external temperature, electricity use, gas use (aggregated with DHW) and CO_2 concentration were obtained at a resolution of 5 minutes. Given the CO_2 concentration, it was possible to produce an estimation of the air renewal."

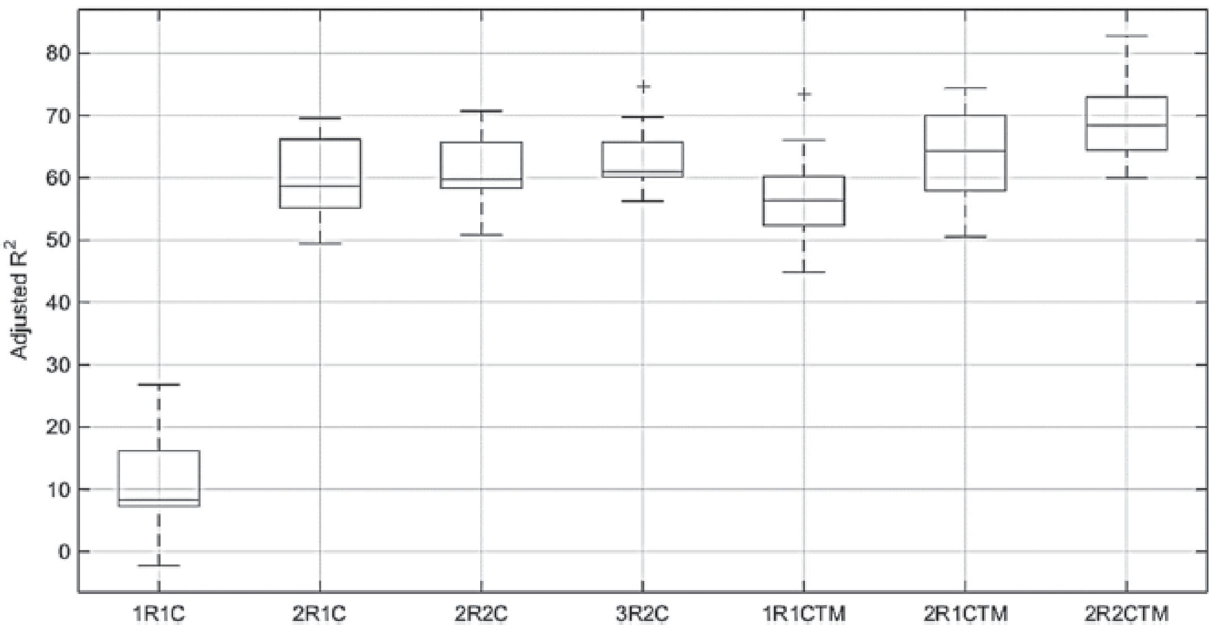


Figure 13. Performance of Bath models of increasing complexity to reproduce time series of interior temperature. 1R model not included (Ramallo-Gonzalez)

The following diagrams show the various lump-parameter building models that were tested.

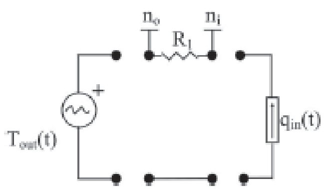


Figure A6. 1R model. This model has a single resistor that represents the thermal resistance of the building envelope. No inertial effects are considered with this model.

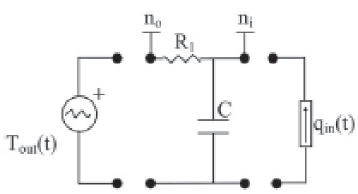


Figure A7. 1R1C model. First-order model where the thermal resistance of the envelope is represented by R_1 and all the thermal mass of the building is represented by the capacitor C.

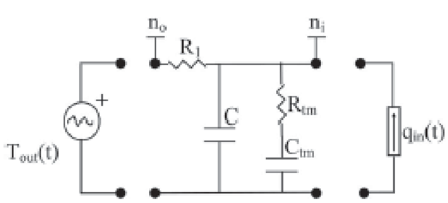


Figure A8. 1R1CTM model. This is the same as the 1R1C model, but the thermal inertia of the building is separated between the thermal inertia of the walls with the element C, and the thermal inertia of the thermal mass – which has its own time constant. This is represented with the resistor R_{TM} and C_{TM}.

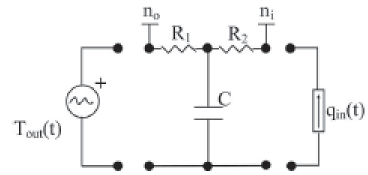


Figure A9. 2R1C model. This model is also first order like the 1R1C. However, the second resistor R_2 allows it to have different temperatures between the construction and the internal air, i.e. it has an extra degree of freedom. This allow to control the temperature swing due to gains $q_{in}(t)$.

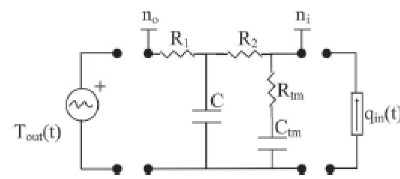


Figure A10. 2R1CTM model. This model is equivalent to the 2R1C model but it has the thermal mass included.

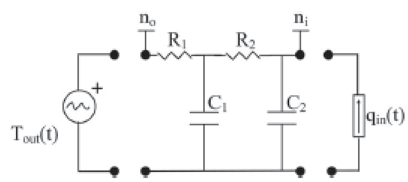


Figure A11. 2R2C model. This is a second-order model with two time constants. R_1 C_1 provides the long-time constant of the building and relates with the thermal inertia of the constructions. The second time constant given by R_2 and C_2 is smaller and is used to represent quick response parts of the building such as the air within the spaces.

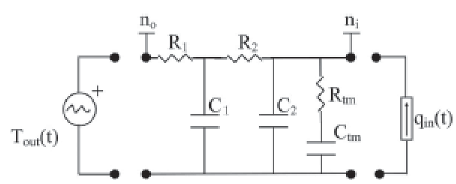


Figure A12. 2R2CTM model. As in previous cases, this is the 2R2C model with the thermal mass being represented.

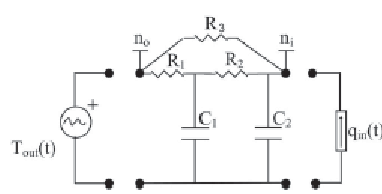


Figure A13. 3R2C model. This model adds the resistor R_3 to account for heat flows that bypass the thermal envelope of the building. This is considered to be the heat flow going through windows and infiltration and ventilation.

Figure 14. Eight RC models evaluated in Bath project. (Ramallo-Gonzalez)

Summary

Relatively simple RC, gray-box models have been successfully applied to data form a variety of buildings for a variety of purposes. Their formulation and solution are driven by the building type, the particular data available, and the research question. The successful models range from 2R1C to 4R3C.

There are many approaches used to the optimization, but for real data from real buildings the two most immediately promising methods seem to be CTSM and the genetic algorithm.

References

1906. "Highest economy in furnace heating: Proper temperatures, ventilation and coal consumption for different outside temperatures." *The Metal Worker, Plumber and Steam Fitter* 66:47-49.

Bacher, Peder, and Henrik Madsen. 2011. "Identifying suitable models for the heat dynamics of buildings." *Energy and Buildings* 43:1511-1522.

Balcomb, J. Douglas. 1982. *Passive Solar Design Handbook*. Vol. 3. Washington: US Department of Energy.

Barber, Daniel A. 2016. *A House in the Sun: Modern Architecture and Solar Energy in the Cold War*. New York, NY: Oxford University Press.

Bugge, Andreas Fredrik. 1924. Test houses erected by Norges Tekniske Hoiskole (Norwegian Technical University), Trondhjem; result of tests with wall-constructions and materials for building warm and cheap dwelling-houses. Trondhjem: F. Bruns bokhandel.

Fels, Margaret F. 1986. "PRISM: An introduction." *Energy and Buildings* 9 (1):5-18. doi: https://doi.org/10.1016/0378-7788(86)90003-4.

ISO 52016-1. 2017. Energy performance of buildings — Energy needs for heating and cooling, internal temperatures and sensible and latent heat loads — Part 1: Calculation procedures.

Kissock, J. K., T. A. Reddy, and D. E. Claridge. 1998. "Ambient-Temperature Regression Analysis for Estimating Retrofit Savings in Commercial Buildings." *Journal of Solar Energy Engineering* 120 (3):168-176. doi: 10.1115/1.2888066.

Kissock, John Kelly, Jeff S. Haberl, and David E. Claridge. 2002. Development of a Toolkit for Calculating Linear, Change-point Linear and Multiple-Linear Inverse Building Energy Analysis Models. In *ASHRAE Research Project 1050-RP*: American Society of Heating Refrigeration and Air Conditioning Engineers.

Kissock, John Kelly, Jeff S. Haberl, and David E. Claridge. 2003. "Inverse modeling toolkit: Numerical algorithms." *ASHRAE transactions* 109:425-434.

Madsen, H. 1985. "Statistically determined dynamical models for climate processes." PhD, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark (Thesis No. 45).

Madsen, H.; Hoist, J. 1995. "Estimation of continuous-time models for the heat dynamics of a building." *Energy and Buildings* 22:67-79.

Malkawi, Ali , and Godfried Augenbroe. 2004. *Advanced Building Simulation*: Routledge.

McKinley, Thomas, and A. G. Alleyne. 2008. "Identification of building model parameters and loads using on-site data logs." SIMBuild: Third National Conference of IBPSA-USA, Berkeley, CA, 08/30.

Norford, Les, Ari Rabl, Joe Spadaro, and Robert Socolow. 1986. "Energy Use in the Enerplex Office Buildings: A Progress Report." ACEEE Summer Study on Energy Efficiency in Buildings.

Oh, Sukjoon, and Jeff S. Haberl. 2016. "Origins of analysis methods used to design high-performance commercial buildings: Whole-building energy simulation." *Science & Technology for the Built Environment* 22 (1):118-137. doi: 10.1080/23744731.2015.1063958.

Rabl, A. 1988. "Parameter Estimation in Buildings: Methods for Dynamic Analysis of Measured Energy Use." *Journal of Solar Energy Engineering* 110 (1):52-66. doi: 10.1115/1.3268237.

Ramallo-González, Alfonso P., Matthew Brown, Elizabeth Gabe-Thomas, Tom Lovett, and David A. Coley. 2018. "The reliability of inverse modelling for the wide scale characterization of the thermal properties of buildings." *Journal of Building Performance Simulation* 11 (1):65-83. doi: 10.1080/19401493.2016.1273390.

Ramallo-González, Alfonso P., Matthew E. Eames, and David A. Coley. 2013. "Lumped parameter models for building thermal modelling: An analytic approach to simplifying complex multi-layered constructions." *Energy and Buildings* 60:174-184. doi: https://doi.org/10.1016/j.enbuild.2013.01.014.

Reichmuth, Howard, and Cathy Turner. 2010. "A Tool for Efficient First Views of Commercial Building Energy Performance." 2010 ACEEE Summer Study on Energy Efficiency in Buildings.
Robison, David, and Howard Reichmuth. 2001. "Billing Simulation Tool for Commercial Buildings." *Strategic Planning for Energy and the Environment* 21 (2).
Shen, Pengyuan, William Braham, and Yunkyu Yi. 2018. "Development of a lightweight building simulation tool using simplified zone thermal coupling for fast parametric study." *Applied Energy* 223:188-214. doi: https://doi.org/10.1016/j.apenergy.2018.04.039.
Socolow, Robert; Sonderegger, Robert. 1976. The Twin Rivers Program on Energy Conservation in Housing: Four Year Summar Report.: Center for Environmental Studies, Princeton University.
Sonderegger, Robert. 1977. "Dynamic Models of House Heating Based on Equivalent Thermal Parameters." PhD, Princeton University.
Subbarao, K. 1985. Building Parameters and Their Estimation from Performance Monitoring. SERl/TP-253-2661.
U.S. DOE. 2020. "EnergyPlus™ Version 9.4.0 Documentation: Engineering Reference." U.S. Department of Energy. https://www.energyplus.net/sites/default/files/docs/site_v8.3.0/EngineeringReference/05-Climate/index.html.
Weisberg, Michael. 2013. *Simulation and similarity: Using models to understand the world*. New York: Oxford University Press.
White, J. A., and R. Reichmuth. 1996. "Simplified method for predicting building energy consumption using average monthly temperatures." IECEC 96. Proceedings of the 31st Intersociety Energy Conversion Engineering Conference, 11-16 Aug. 1996.

White Box Modeling

Yun Kyu Yi, University of Illinois Urbana-Champaign

History of Building Simulation

The main development of building simulation tools starts in the 1970s and was further developed in the ’80s and ’90s. During the '90s, most functions were completed and validated. Since the late ’90s and the beginning of the 2000s, tools were developed to share code, files, and integrate them into the design tools. In the early 10s, several cloud-based tools were introduced to the market, and more tools were integrated into CAD tools.

Reference

Augenbroe, Godfried (2011). The role of simulation in performance based design. In: J. Hensen and R. Lamberts (eds), Building Performance Simulation for Design and Operation. Spon Press.

Ali Malkawi and Godfried Augenbroe (editors), Advanced Building Simulation. SPON Press, Taylor and Francis group, 2004. ISBN 0-415-32122-0

J.L.M. Hensen, Towards more effective use of building performance simulation in design, Van Leeuwen, J.P. and H.J.P. Timmermans (eds.) Developments in Design & Decision Support Systems in Architecture and Urban Planning, Eindhoven: Eindhoven University of Technology, ISBN 90-6814-155-4, p. 291-306

Currently, the issue of using building simulation tools is not that there are too few tools, but instead too many. This brings the problem of selecting the right tools to use. Currently, available tools can be found in the Building Energy Software Tools webpage (https://www.buildingenergysoftwaretools.com/). Crawley (2006) discusses the capabilities of different building energy simulation tools and its shows limits and functions used in different tools.

Reference

Crawley, Drury B., Hand, Jon W., Kummert, Michael, and Griffith, Brent T. Contrasting the capabilities of building energy performance simulation programs. United States: N. p., 2008. Web. doi:10.1016/j.buildenv.2006.10.027.

Yi, Yun Kyu, “Building Performance and Computational Simulation,” in The Design and Construction of High Performance Homes Building Envelopes, Renewable Energy and Integrated Practice, ed. Franca Trubiano, (Abingdon, UK: Routledge, 2012): pp. 163-177

White Box Models for Whole Building Energy Use

White box models for whole-building energy use can be sub-divided into “milky white box” and “glass box.” The difference between the two can be identified by its accessibility to its core function. “Milky white boxes” are typically commercial tools that encapsulate its core engine. Because of this, it has difficulty accessing functions, where “glass box,” is an open-source where users can get access to the core engine which allows the user to test different configurations or new models. The benefit of the milky white box is that it is stable, and the outputs are more trustable. However, its benefits work against the users whose interests are in testing different models or algorithms to replace existing core functions.

Since White box models are based on physics, their model was developed based on energy flow. This means the 1st step of the calculation process is using the heat balance model to find the cooling and heating load. This load is passed to the system and primary energy side to calculate whole-building energy use. This process was developed as modules in the tool that the system manager communicates between modules to calculate energy use.

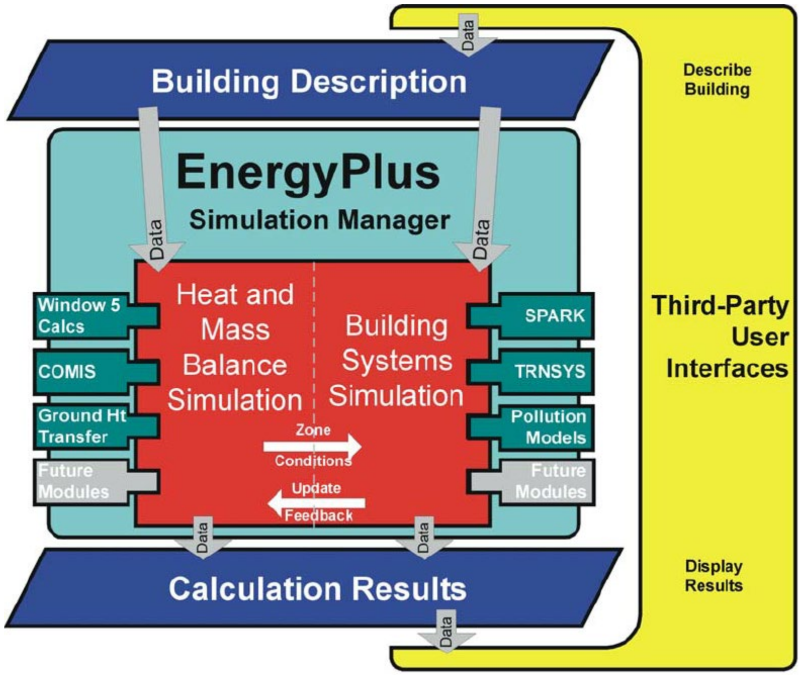


Figure 1. EnergyPlus Internal elements (Getting Started with EnergyPlus, 2021)

The building energy tools are dynamic and deterministic. This means that boundary conditions were set up at the initial stage and modules communicate each other accordingly based on time steps. The white model which is a physics-based model is sophisticated and able to understand the performance of building in-depth. However, its strongest benefit also works against its benefit. It is common to find simulation result shows a significant discrepancy with actual performance (Sokolowski and Catherine, 2011). DOE identified this discrepancy into two categories (DOE 2019). Variability and Uncertainty were identified as sources of differences between simulated and actual performance.

Currently, several research projects are undergoing to improve the discrepancy between actual results and simulation results. The most significant effect on discrepancy is occupant behavior and a significant amount of literature can be found in recent years. Gaetani, Hoes, and Hansen (2016) summarized all papers related to occupant behavior modeling in Building energy modeling.

For this reason, other modeling methods like a grey box or black box can be utilized to overcome the limitations of the white-box model. However, grey box or black box were depending on the data set to overcome the limitation of white-box modeling. If the data set is missing, it is difficult to develop a black or grey box model and has to depend on the white-box model. This especially true when the building is in the design stage and no data is available. To overcome the limitation of missing data and discrepancy of white-box model result with actual usages, Judkoff (1988) suggested three methods for validation of the white box model (NREL, 2006) (Figure 2).

Since other sections focused on the operation and management stage of building energy usage, this section focuses on the design stage and how the white-box model can be used. Among the three techniques in figure 2, in the design stage "comparative" method is most suitable to use. Since in the design stage, it is more important to understand the relative comparison of different design strategies than absolute truth results. The following chapter discusses some of the examples of different white-box models that are used to overcome the limitation.

Table 1. Validation Techniques

Technique	Advantages	Disadvantages
Empirical Test of model and solution process	• Approximate truth standard within experimental accuracy • Any level of complexity	• Experimental uncertainties: - Instrument calibration, spatial/ temporal discretization - Imperfect knowledge/ specification of experimental object (building) being simulated • High quality detailed measurements are expensive and time consuming • Only a limited number of test conditions are practical
Analytical Test of solution process	• No input uncertainty • Exact mathematical truth standard for the given model • Inexpensive	• No test of model validity • Limited to highly constrained cases for which analytical solutions can be derived
Comparative Relative test of model and solution process	• No input uncertainty • Any level of complexity • Many diagnostic comparisons possible • Inexpensive and quick	• No absolute truth standard (only statistically based acceptance ranges are possible)

Figure 2. Validation Techniques

Reference
Getting Started with EnergyPlus, accessible March 2021, https://www.energyplus.net/sites/default/files/docs/site_v8.3.0/GettingStarted/GettingStarted/index.html
Sokolowski, John A., and Catherine M. Banks. 2009. Principles of modeling and simulation: a multidisciplinary approach. Hoboken, N.J.: John Wiley.
DOE, accessible March. 2021, https://www.energy.gov/sites/prod/files/2019/05/f62/bto-peer%E2%80%932019-lbnl%E2%80%93empirical-validation.pdf
Isabella Gaetani, Pieter-Jan Hoes, Jan L.M. Hensen, "Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy," Energy and Buildings, Volume 121, 2016, 188-204, https://doi.org/10.1016/j.enbuild.2016.03.038.
Judkoff, R. (1988). Validation of Building Energy Analysis Simulation Programs at the Solar Energy Research Institute. Energy and Buildings, Vol. 10, No. 3, p. 235. Lausanne, Switzerland: Elsevier Sequoia.
Model Validation and Testing: The Methodological Foundation of ASHRAE Standard 140, accessable March 2021., https://www.nrel.gov/buildings/assets/pdfs/40360.pdf

Integration with other models

Here few examples of integration with other models are discussed, first section discusses how the building energy model can be integrated with the CFD (computational fluid dynamics) modeling to improve boundary conditions for energy simulation tools. Second is the integration between the building energy model with the daylight model to improve the accuracy of the indoor daylight level. Lastly, discuss integrating the building energy model with two different models to speed up computation time that limits integrating two white-box models.

Building energy model with CFD

One of the major limits of building energy modeling is calculating airflow in buildings. Specifically, convection thermal transfer is a difficult part of energy modeling. For that reason, several papers discuss indoor coupling between energy simulations and CFD. Here are some of the major publications in this area.

Reference

Clarke, JA, Dempster, WM, Negrão, CRO (1995) The Implementation of a Computational Fluid Dynamics Algorithm Within the ESP-r System. Glasgow: Energy system division, Univertsity of Strathclyde.

Beausoleil-Morrison, I (2002) The adaptive conflation of computational fluid dynamics with whole-building thermal simulation. Energy and Building 34: 857–871.

Wang, L, Chen, Q (2007) Validation of a coupled multizone-CFD program for building airflow and contaminant transport simulation. HVAC&R Research 13(2): 267–281.

Chen, Q, Lee, K, Mazumdar, S, et al. (2010) Ventilation performance prediction for buildings: Model assessment. Building and Environment 45: 295–303.

Djunaedy, E, Hensen, JLM, Loomans, MGLC (2004c) On integration of CFD in building design. In: Proceedings of the PhD symposium modelling and simulation for environmental engineering (ed Hensen, JLM, Lain, M), 16 April, 53–60.

Kong, Q, Feng, J, Yang, C, et al. (2017) Numerical simulation of a radiant floor cooling office bades on CFD-BES coupling and FEM. In: Proceedings of the 8th international conference on applied energy, Beijing, China, 8–11 October, 3577–3583.

Yi, YK, Feng, N (2013) Dynamic integration between building energy simulation (BES) and computational fluid dynamics (CFD) simulation for building exterior surface. Building Simulation 6: 297–308.

Rodríguez-Vázquez, Martin, Iván Hernández-Pérez, Jesus Xamán, Yvonne Chávez, Miguel Gijón-Rivera, and Juan M Belman-Flores. “Coupling Building Energy Simulation and Computational Fluid Dynamics: An Overview.” Journal of Building Physics 44, no. 2 (September 2020): 137–80. https://doi.org/10.1177/1744259120901840.

Coupling with CFD and the building energy model extend to outdoor conditions, specifically site-specific conditions are one of the major discrepancies between actual vs. simulation. The following papers discuss coupling CFD and building simulation for outdoor conditions.

Reference

Liu, J, Heidarinejad, M, Gracik, S, et al. (2015) The impact of exterior surface convective heat transfer coefficients on the building energy consumption in urban neighborhoods with different plan area densities. Energy and Buildings 86: 449–463.

Toparlar, T, Blockn, B, Maiheu, B, et al. (2018) Impact of urban microclimate on summertime building cooling demand: a parametric analysis for Antwerp, Belgium. Applied Energy 228: 852–872.

Building energy model with Daylight model

Electricity takes a significant portion of building energy usage and it is crucial to understand how daylight performs indoor space to estimate the reduction of energy usage by utilizing daylight. The current method built-in energy model is a simplified method, and several approaches were developed to compensate for the limitation of the current energy model. Here are some of the methods developed to overcome this limitation.

Reference

Denis Bourgeois, Christoph Reinhart, Iain Macdonald, Adding advanced behavioural models in whole building energy simulation: A study on the total energy impact of manual and automated lighting control, Energy and Buildings, 38, 7, 2006, doi.org/10.1016/j.enbuild.2006.03.002.

Janak, Milan. (1997). Coupling building energy and lighting simulation. 5th International IBPSA Conference.
Yun Kyu Yi, Dynamic coupling between a Kriging-based daylight model and building energy model, Energy and Buildings, 128, 2016, 798-808, doi.org/10.1016/j.enbuild.2016.05.081

Building energy model with two different models

Even though integrating two white-box models improves the prediction of building energy use, the most significant limitation of integration is related computational time to simulate both white-box models. To overcome this limit, the black-box model can support reducing computational time with reasonable prediction. This method allows the use of the black-box model in the initial design stages where data is limited to the use black-box model.

Reference

Changyu Qiu, Yun Kyu Yi, Meng Wang, Hongxing Yang, Coupling an artificial neuron network daylighting model and building energy simulation for vacuum photovoltaic glazing, Applied Energy, 263, 2020, 114624, https://doi.org/10.1016/j.apenergy.2020.114624.
Yi, Y. K., Tariq, A., Park, J., and Barakat, D., "Multi-Objective Optimization (MOO) of a Skylight Roof System for Structure Integrity, Daylight, and Material Cost," Journal of Building Engineering. 34. Feb. 2021. https://doi.org/10.1016/j.jobe.2020.102056.
Wang, B., Yi, Y. K, "Developing an Adapted UTCI (Universal Thermal Climate Index) for the Elderly Population in China's Severe Cold Climate Region," Sustainable Cities and Society. Accepted. https://doi.org/10.1016/j.scs.2021.102813.

Recent tool development

The building energy modeling tools are continuously developed and their capacity is updated, which makes it difficult to discuss the tools' capacities since there will be a new update soon. For that reason, the section discusses each tool's engine or base where it starts and summarizes its pros and cons of the current released version.

The most popular commercial tools used in the industry can be divide by what type of engine it uses. Some commercial tools are run based on an open-source engine like EnergyPlus (https://energyplus.net/). Others like TRACE700 (https://www.trane.com/commercial/north-america/us/en/products-systems/design-and-analysis-tools/trace-700.html), developed by Trane were built with an engine based on ASHRAE (American Society of Heating, Refrigerating, and Air-Conditioning).

The greatest benefit of the open-source engine is data sharing and module insertion. The user can revise and add new components to the engine. The benefit of a closed-source engine is that it is verified and more reliable. However, with the rapid development of new technologies and new methods, open-source engines are more frequently mentioned in the market because of their ability to adapt to state-of-the-art technology.

Some of the major open-source tools can be grouped to what engine it uses. The most dominant engine in the US market is EnergyPlus, another engine can be DOE-2 (https://doe2.com/). One of the popular EnergyPlus based tools is DesignBuilder (DB, https://designbuilder.co.uk/), which is a standalone tool that includes parts of Radiance (https://www.radiance-online.org/) and a simplified CFD model. DB has a relatively easy process that can be used to build complex geometry and an easy to create complex zoning. For DOE-2, eQuest (https://doe2.com/equest/index.html) is the most popular tool in the market. Since it was open-source for more than a decade, several practitioners still use the tool.

Some of the building energy tools were integrated into CAD tools like Rhino (Grasshopper)(https://www.rhino3d.com/) or Dynamo (https://www.autodesk.com/products/dynamo-studio/overview). These CAD tools use a graphic program interface that allows users to easily control geometry. Ladybug (https://www.ladybug.tools/), ClimateStudio (previously called Diva-for-Rhino, https://www.solemma.com/climatestudio), and OpenStudio SketchUp Plug-in (https://www.openstudio.net/) are three major building energy tools that embed into NURBS (Non-Uniform Rational Basis Spline) CAD tools. The most beneficial things about these tools are that it makes it easy to test energy performances in the early design stage to find a better design solution. However, it is comparatively complicated for beginners to use and has the limitation of populating complex HVAC (Heating, ventilation, and air conditioning) systems and zoning.

Another trend in building energy tools is web-based energy simulation tools. Sefaira (https://www.sketchup.com/products/sefaira), and cove.tool (https://www.cove.tools/), are the most well know tools as cloud-based simulation tools. The benefit of cloud-based simulation is fast and easy to use. However, tools have limited parameters that difficult to use for sophisticate or complex geometry.

One of the tools frequently used by HVAC professions is TRACE700. TRACE700 has a closed-source engine, which is built by TRANE, a company expert in the HVAC system. However, TRACE700 will replace by TRACE3Dplus (https://www.trane.com/commercial/north-america/us/en/products-systems/design-and-analysis-tools/trane-design-tools/trace-3d-plus.html) which is based on EnergyPlus. IES (https://www.iesve.com/software/building-energy-modeling), which is another tool that uses its engine. It has various engines and is integrated into one platform.

Other well-known building energy simulation tools that are not frequently used in the US include esp-r, DeST, and WUFI Plus. The following references discuss in more detail on comparison between different simulation tools.

Reference:

Al Ka'bi, A.H. Comparison of energy simulation applications used in green building. Ann. Telecommun. 75, 271–290 (2020). https://doi.org/10.1007/s12243-020-00771-6

Jacobs, P, and Henderson, H. Fri . "State-of-the-Art Review Whole Building, Building Envelope, and HVAC Component and System Simulation and Design Tools". United States.

Roth, Amir, and Reyna, Janet. Sun . "Innovations in Building Energy Modeling: Research and Development Opportunities for Emerging Technologies". United States. https://doi.org/10.2172/1710155. https://www.osti.gov/servlets/purl/1710155.

Lee, Sang Hoon & Hong, Tianzhen & Piette, M. & Taylor-Lange, Sarah. (2015). Energy Retrofit Analysis Toolkits for Commercial Buildings: A Review. Energy. 10.1016/j.energy.2015.06.112.

Summary

The white box model is sophisticated and validates. It is the main reason why it is widely used in the field. However, its complexity and significant dependence on physical properties require careful attention. The study reviewed current methods and related papers that integrate different white-box models to improve its prediction. As discussed, it is important to understand that the white-box model is not always applicable to any problems related to building energy use and it is requiring careful investigation to find the right model to use.

Low order white box modeling technique

Pengyuan Shen, Harbin Institute of Technology (Shenzhen)

1. Building thermal modeling

1.1. White (or forward), grey, and black box model for building simulation

A large number of models of the static and dynamic approaches have been used in the presentation of the thermal behavior of buildings. It was proposed to classify the sets of these models into three categories, the white, the black and the grey boxes models. Depending on the static and dynamic approaches, some of the models have been very successful in describing the thermal behavior of large residential buildings. Others have been used to estimate the thermal-energy demands or in the prediction of heat consumption and reducing energy consumption.

References:

Khan, M.E. and Farmeena, K. (2012) A Comparative Study of White Box, Black Box and Grey Box Testing Techniques. International Journal of Advanced Computer Science and Applications, 3, 12-15.

Amara, F., et al. (2015). Comparison and Simulation of Building Thermal Models for Effective Energy Management. Smart Grid and Renewable Energy: 95-112.

1.2. What to model in building energy simulation? – occupancy behavior and building thermal behavior

According to occupant control level, appliances driven loads can be categorized into two classes, responsive loads and unresponsive loads. Responsive loads includes plug loads, lighting loads, laundry and drying machines, dishwashers, cooking ranges, heating thermostats (loads), cooling thermostats (loads). Unresponsive loads includes refrigerator loads, freezer loads and stand-by loads. In regards to electric heaters, with a responsive thermostat which controls the output of the heater effectively and maintains a more consistent room temperature, the electric heaters models without a thermostat require closer monitoring by the customer, and therefore are associated with unresponsive loads.

Another important analysis procedure of building energy simulation is the description of building thermal behavior by an energy balance model shown in Equation (1). Heating or cooling load can be predicted by this model, which is used for system and equipment selection.

$$Qi + Qc + Qs + Qv + Qe = \Delta S$$

Qi – internal heat gain
Qc – conduction heat gain or loss
Qs – solar heat gain
Qv – ventilation heat gain or loss
Qe – evaporative heat loss
ΔS– change in heat stored in the building

References:

Amara, F., et al. (2015). Comparison and Simulation of Building Thermal Models for Effective

Energy Management. Smart Grid and Renewable Energy: 95-112.

Singh, R. and Vyakaranam, B. (2012) Evaluation of Representative Smart Grid Investment Grant Project Technologies: Distributed Generation. PNNL, Richland. http://www.esc.gov.yk.ca/

Energy Solution Centre (2011) Easy$ Tip Sheets—Energy Advice Saving Yukoners Money. Energy Solution Centre Report, Whitehorse, 1-4. www.esc.gov.yk.ca

1.3. Systems involved in building energy use simulation

To maintain a comfortable environment, there have to be some systems to meet human requirements. Systems involved in building energy use include HVAC and domestic hot water system, lighting and plug-in system, and some other ultimate and special usage system.

2. White box model

White box testing is a test case design method that uses the control structure of the procedural design to derive test cases, which require a significant amount of expertise. To build the control structure, physical significance must be known to develop the theoretical basis.

Reference:

Khan, M.E. and Farmeena, K. (2012) A Comparative Study of White Box, Black Box and Grey Box Testing Techniques. International Journal of Advanced Computer Science and Applications, 3, 12-15.

Amara, F., et al. (2015). Comparison and Simulation of Building Thermal Models for Effective Energy Management. Smart Grid and Renewable Energy: 95-112.

2.1. Static and dynamic thermal response of buildings

In static conditions, the conduction heat transfer follows the Fourier Law, which states that the negative gradient of temperature and the time rate of heat transfer is proportional to the area at right angles of that gradient through which the heat flows.

Due to thermal inertia of envelope, there exists delay and attenuation the heat flow through a real wall compared with a "zero-mass" wall of the same U-value. The greater thermal mass is, the more daily temperature swings dampen.

2.2.1. Different equation types of white models

2.2.2. Static conditions

For static conditions, linear equations for conduction, convective heat transfer and solar energy received and a non-linear equation (Stefan-Boltzmann equation) for radiative heat transfer were considered.

$$q = L(T - T_a) - A_s I + \varepsilon$$

$$\dot{q}_{21} = h\, A \cdot (T_2 - T_1)$$

Linear equations:

q : Heat transfer rate (W);
L: Coefficient of static losses (W/°C);
T −Ta : Difference between indoor and outdoor temperature (˚C);
As : Equivalent surface (m^2);
I : Solar energy received by a vertical wall (W/m^2),
ε : Depends by the state of variables measured at the beginning and end of the period observational (W)
T2 −T1: Difference between the boundary and ambient temperature (˚C);
h: Convective heat transfer coefficient (W/m^2·°C)
A: heat transfer area of the surface (m^2).

$$\dot{q}_{21} = \varepsilon \cdot \sigma \cdot A \cdot \left(T_2^4 - T_1^4\right)$$

Non-linear equation:

q : Emitted heat transfer rate (W);
ε : Surface emissivity;
σ : Stefan-Boltzmann constant (5.669×10−8m^2K^4)
A: Radiation surface (m^2).

2.2.3. Dynamic conditions

For dynamic conditions, an ordinary differential equation can be used to analyze temperature variation with time as shown in the following equation.

$$C \cdot \frac{\mathrm{d}T}{\mathrm{d}t} = U \cdot (T - T_a)$$

C: Thermal capacity (J/K),
U: Overall heat transmission coefficient (W/m^2K),
t: time.
To acquire the temporal-spatial temperature distribution, partial linear differential equations are utilized.

$$\frac{\partial^2 T(x,\tau)}{\partial x^2} = \frac{1}{\alpha}\frac{\partial T(x,\tau)}{\partial \tau}$$

u (x,t) : Temperature at position x and time (t),
α: Thermal diffusivity (mm^2/s) - measures the rate of transfer of heat of a material from the hot end to the cold end.

Reference:

Christian, N., Dirk, J., Burhenne, S. and Florita, A. (2011) Modellbasierte Methoden für die Fehlererkennung und Optimierung im Gebäudebetrieb. Fraunhofer ISE, Technical Report 0327410A-C, 1-276.

3. Black box model

Black box testing treats the software as a "Black Box" – without any knowledge of internal working and it only examines the fundamental aspects of the system. In the black-box, the parameters are generally adjusted automatically in training procedure. Therefore, the relationship with physical fundamental principles is Implicit in black box models.

Reference:

Khan, M.E. and Farmeena, K. (2012) A Comparative Study of White Box, Black Box and Grey Box Testing Techniques. International Journal of Advanced Computer Science and Applications, 3, 12-15.

Amara, F., et al. (2015). Comparison and Simulation of Building Thermal Models for Effective Energy Management. Smart Grid and Renewable Energy: 95-112.

3.1. Pros and cons of white and black box models

3.1.1. White box:

Pros:

- Clear model internal structure
- Extrapolation enabled (under various scenarios)
- Physical meaning
- Can be used for optimization

Cons:

- Need great amount of expertise
- Hard to calibrate
- Sometimes need great amount of computation

3.1.2. Black box:

Pros:

- Calibrated while modeling
- Less computation
- Can be used for fault detection

Cons:

- Cannot be used for optimization
- Internal structure unknown
- Unable for extrapolation (extreme conditions or scenarios)

4. Why low-order white box model?

There are certain drawbacks in the white box model, which needs to be improved.

4.1. Dynamic conduction heat transfer simulation in white box model

In cooling load and energy calculation, building simulation and energy analysis, conduction heat transfer is usually modeled as a one-dimensional, transient process with constant material properties. The simplified heat diffusion equation in Cartesian coordinates is shown in the first following equation. Since the first equation is a partial differential equation, the system is usually solved numerically, often by means of conduction transfer function methods.

$$\frac{\partial^2 T(x,\tau)}{\partial x^2} = \frac{1}{\alpha}\frac{\partial T(x,\tau)}{\partial \tau}$$

$$q = -k\frac{\partial T(x,\tau)}{\partial x}$$

The method results in a simple linear equation that expresses the current heat flux in terms of the current temperature and temperature and heat flux histories.

$$q_{o,\theta} = -\sum_{n=0}^{N_y} Y_n T_{is,\theta-n\delta} + \sum_{n=0}^{N_x} X_n T_{os,\theta-n\delta} + \sum_{n=1}^{N_\phi} \phi_n q_{o,\theta-n\delta}$$

$$q_{i,\theta} = -\sum_{n=0}^{N_z} Z_n T_{is,\theta-n\delta} + \sum_{n=0}^{N_y} Y_n T_{os,\theta-n\delta} + \sum_{n=1}^{N_\phi} \phi_n q_{i,\theta-n\delta}$$

where q0 and qi are heat flux at exterior and interior surface, respectively. Xn, Yn and Zn are surface-to-surface exterior, cross and interior CTF coefficient, respectively. Tis and Tos are interior and exterior surface temperature, respectively. Nx, Ny and Nz are number of exterior, cross and interior CTF terms, respectively. φn is flux coefficient. Nφ is the number of flux history terms. The subscript θ represents the current time, and δ is time step. The zero subscript represents a current value.

References:

Incropera, F.P. and DeWitt, D.P. Introduction to heat transfer, 3rd ed. Wiley, New York, NY., 1996.

Chen Youming, et. al. (2006). Investigation of the Accuracy of Calculation Methods for Conduction Transfer Functions of Building Construction. ICEBO2006, Shenzhen, China.

4.2. Two main methods to solve the equation

- Numerical methods (TRNSYS): e.g., Direct root-finding algorithms, State space method (EnergyPlus)
- Frequency domain methods (BLAST): e.g., Laplace transform…
- Frequency domain regression methods

References:

Chen Youming, et. al. (2006). Investigation of the Accuracy of Calculation Methods for Conduction Transfer Functions of Building Construction. ICEBO2006, Shenzhen, China
Wang, S.,&Chen, Y.. (2003). Transient heat flow calculation for multilayer constructions using a frequency-domain regression method. Building and Environment, 38(1), 45-61.
Harish, V.s.K.V. & Kumar, Arun. (2016). A review on modeling and simulation of building energy systems. Renewable and Sustainable Energy Reviews. 56. 1272-1292.

4.2.1. State space method

The basic state space system is defined by the following linear matrix equations:

$$\frac{d[x]}{dt} = [A][x] + [B][u]$$

$$[y] = [C][x] + [D][u]$$

where x is a vector of state variables, u is a vector of inputs, y is the output vector, t is time, and A, B, C, and D are coefficient matrices.
Through the use of matrix algebra, the vector of state variables (x) can be eliminated from the system of equations, and the output vector (y) can be related directly to the input vector (u):

$$\frac{d\begin{bmatrix} T_1 \\ \vdots \\ T_n \end{bmatrix}}{dt} = [A]\begin{bmatrix} T_1 \\ \vdots \\ T_n \end{bmatrix} + [B]\begin{bmatrix} T_i \\ T_o \end{bmatrix}$$

$$\begin{bmatrix} q_i'' \\ q_o'' \end{bmatrix} = [C]\begin{bmatrix} T_1 \\ \vdots \\ T_n \end{bmatrix} + [D]\begin{bmatrix} T_i \\ T_o \end{bmatrix}$$

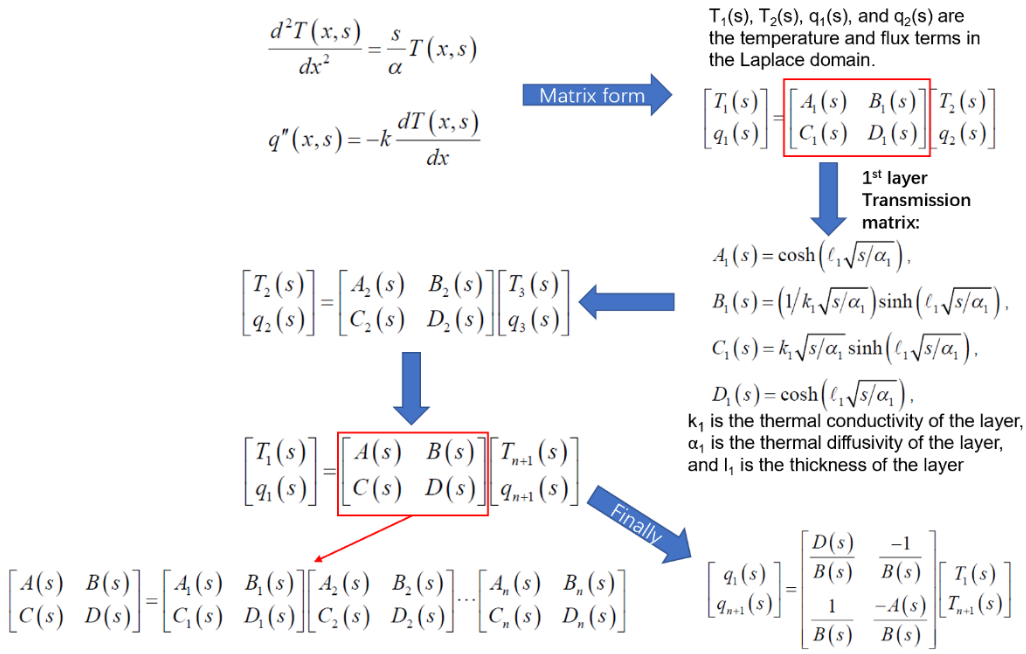
where T1, T2, ..., Tn-1, Tn are the finite difference nodal temperatures, n is the number of nodes, Ti and To are the interior and exterior environmental temperatures, and and are the heat fluxes.

References:

US Department Of Energy. (2010). EnergyPlus Engineering Reference: The Reference to EnergyPlus Calculations.

4.2.2. Frequency domain methods (FDR):

In FDR method, the frequency characteristics of the total transmission matrix are calculated within the frequencyrange concerned firstly. Then, a set of linear equations is solved to yield a simple polynomial function. Finally, the response factors are obtained simply by applying inverse Laplace transforms or Z-transforms on the polynomial s-transfer function.



References:

Wang, S.,&Chen, Y.. (2003). Transient heat flow calculation for multilayer constructions using a frequency-domain regression method. Building and Environment, 38(1), 45-61.

4.2.3. Frequency domain vs. State space

Through the use of relatively simple matrix algebra, the state space variables (nodal temperatures) can be eliminated to arrive at a matrix equation that gives the outputs (heat fluxes) as a function of the inputs (environmental temperatures) only. This eliminates the need to solve for roots in the Laplace domain.

The resulting matrix form has more physical meaning than complex functions required by the Laplace transform method.

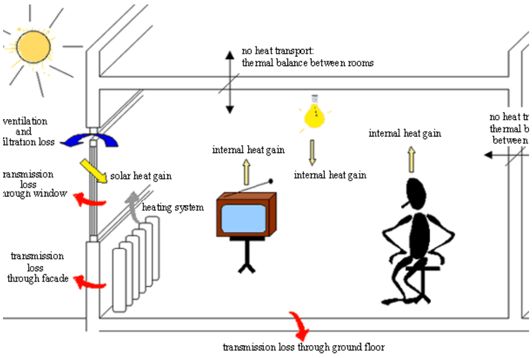
For an adequate number of nodes the state space method computed a heat flux at the surface of a simple one layer slab within 1% of the analytical solution.

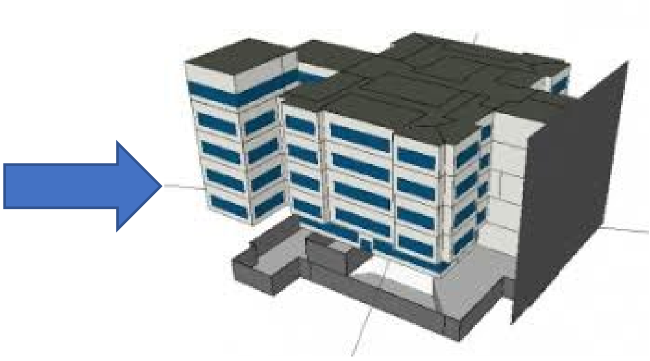
References:

Chen Youming, et. al. (2006). Investigation of the Accuracy of Calculation Methods for Conduction Transfer Functions of Building Construction. ICEBO2006, Shenzhen, China.

Wang, S.,&Chen, Y.. (2003). Transient heat flow calculation for multilayer constructions using a frequency-domain regression method. Building and Environment, 38(1), 45-61.

4.2.4 Real multi-zone building can be complicated (than the calculations of wall components)



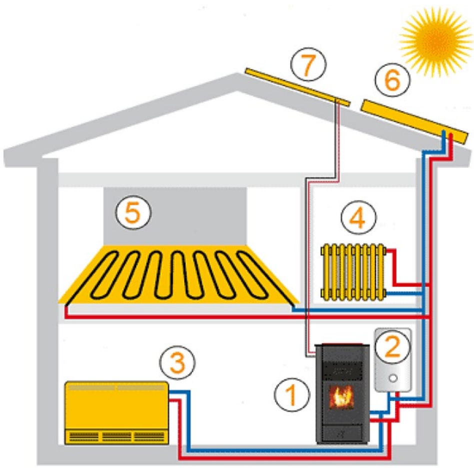


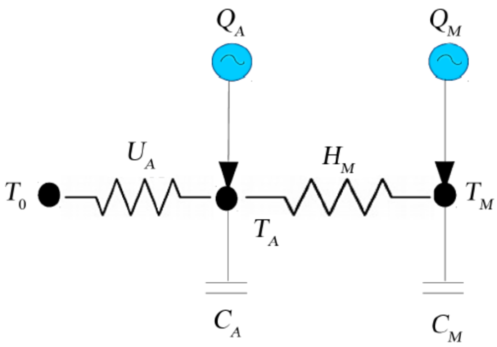
When the model gets more complicated, the solving of the dynamic heat transfer function consumes more computation. For the conventional numerical method, the state space method and frequency domain method (the solvers that most white box simulation tools nowadays use), the solution of the dynamic heat transfer sometimes brings about inaccuracies and usually costs great amount of computation.

The reasons of people turn to low-order/reduced-order white box model include faster in computation (also means faster calibration and optimization), less inputs compared with pure white box modeling, physical significance kept.

4.3. Low-order white box modeling: The electrical analogue – lumped capacitance

Here we introduced the concept of thermal capacity of building denoted by Cr, in the electric circuit analog model (RC) that is used to describe heat flow and heat transfer phenomena. Where, Cr equal to the air mass (m) in the room times the specific heat capacity of air (cp) which change with time as shown in the following Equations (1) and (2). This method makes a simplified building model and solutions can be easily found.

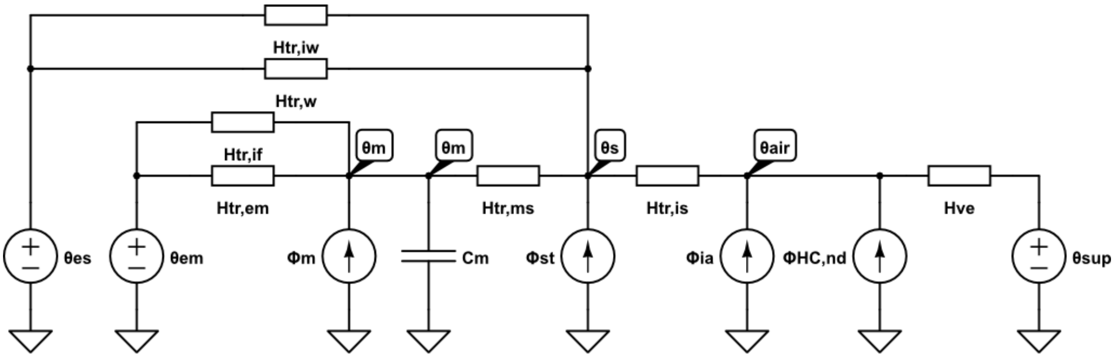




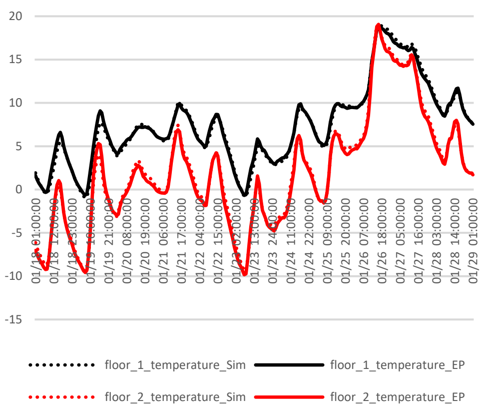
References:

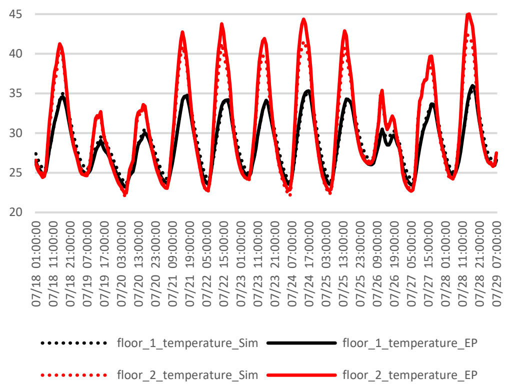
Amara, F., et al. (2015). Comparison and Simulation of Building Thermal Models for Effective Energy Management. Smart Grid and Renewable Energy: 95-112.

4.4. The prediction of indoor air temperature



The low-order white box modeling is also able to catch some of the dynamic nature of building thermal performance due to the impacts of both thermal resistance and capacitance. For example, using the above 5R1C electrical analogue model taking into account of the thermal transmittance between adjacent zones, it is possible to have a comparatively high confidence in predicting zone internal air temperature when thermal mass is not too high. For example, here is an example of comparing indoor air temperature simulation of a simple residential building in Philadelphia predicted by the 5R1C model to its benchmark simulation results produced by the white box simulation engine - EnergyPlus:





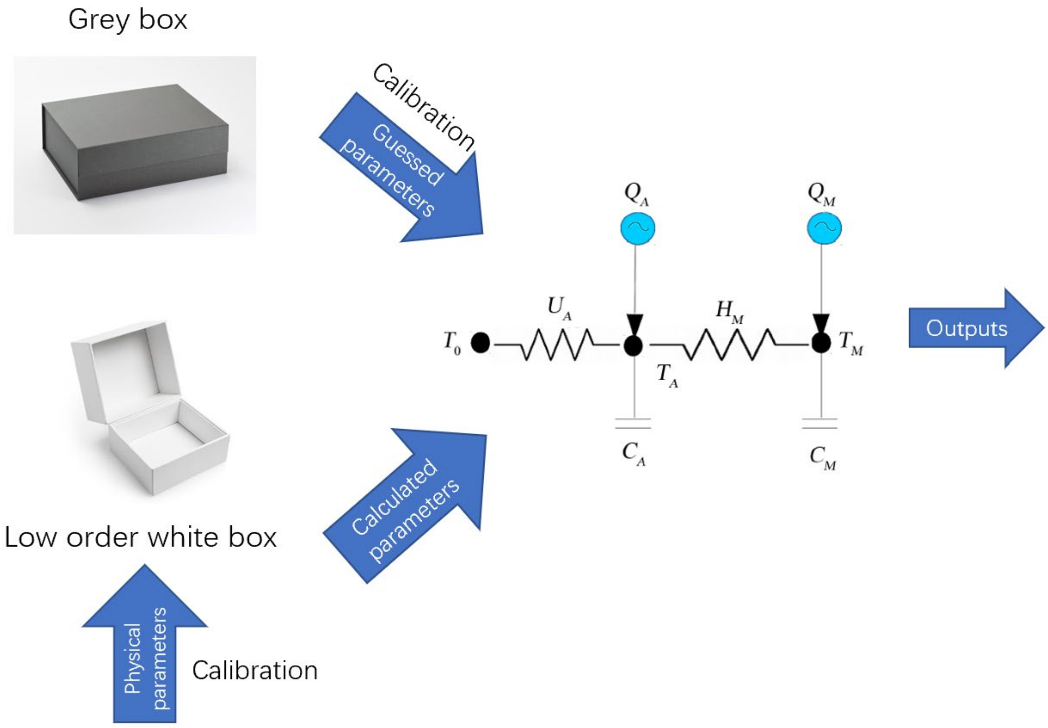
A random winter week A random summer week

However, it should be noted that deviations may occur when heavy thermal mass is used in such low-capacitance-order white model.

<u>References:</u>

Shen, P., Braham, W., Yi, Y., 2018. Development of a lightweight building simulation tool using simplified zone thermal coupling for fast parametric study. Applied Energy 223, 188-214.

4.5. The difference between low order white model and grey box model



Low order white box modeling involves physical parameters that do have a real thermophysical meaning for buildings, and the calibration procedure deals with the tuning of those parameters, while the calibration of grey box models usually deals with guessing and turning of the imaginary (or "proxy") parameters that represents the overall performance of the building envelopes or heat gain (loss).

4.6. Pros and cons of low order white box:

Pros:

- Clear model internal structure
- Extrapolation enabled (under various scenarios)
- Physical meaning
- Can be used for optimization

Cons:

- Lose some predictive accuracy compared with pure white box model

Blackbox Modelling

Ravi Srinivasan, University of Florida

Agent based Modelling (Occupant behaviour Modelling)

Machine Learning and Building Energy prediction:

Ensemble models:

Air quality and Indoor emission:

Agent based Modelling (Occupant behaviour Modelling)

Occupant behaviour is hard to track inside a thermal zone whereas occupant number (count) is relatively easy to track. Uncertainties in energy estimation has been tracked to occupant behaviour. Among others, one approach to track occupant behaviour is by using Agent-based modelling (ABM). At the UrbSys Lab, University of Florida, we have done extensive studies of using ABM for improving energy estimations. For this purpose, we used LBNL's Building Control Virtual Test Bed (BCVTB) that linked both EnergyPlus™ and an ABM, PMFServ (from University of Pennsylvania). Some significant work in this area are listed below.

1) Jia, M., Srinivasan, R.S., Ries, R., Weyer, N., Bharathy, G. (2019). A systematic development and validation approach to a novel agent-based modelling of occupant behaviours in commercial buildings. Energy and Buildings, 199: 352-367; https://doi.org/10.1016/j.enbuild.2019.07.009
2) Jia, M., Srinivasan, R.S., Ries, R., Bharathy, G., Weyer, N. Investigating the Impact of Actual and Modeled Occupant Behavior Information Input to Building Performance Simulation. Buildings 2021, 11(1), 32; https://doi.org/10.3390/buildings1101003
3) Jia, M. and Srinivasan, R.S. (2020). Building Performance Evaluation using Coupled Simulation of EnergyPlus™ and an Occupant Behavior Model. Sustainability 2020, 12(10), 4086. https://doi.org/10.3390/su12104086
4) Jia M, Srinivasan R.S., Raheem A.A. (2017). From Occupancy to Occupant Behavior: An Analytical Survey of Data Acquisition Technologies, Modeling Methodologies, and Simulation Coupling Mechanisms for Building Energy Efficiency. Renewable and Sustainable Energy Reviews, 68(1): 525-540; https://doi.org/10.1016/j.rser.2016.10.011
5) Mengda, J., Srinivasan, R.S., Bharathy, G., Silverman, B.S., Weyer, N. An Agent-based Model Approach for Simulating Interactions between Occupants and Building Systems. Building Simulation 2017; https://doi.org/10.26868/25222708.2017.673 (Conference)
6) Berger, Christiane, and Ardeshir Mahdavi. "Review of current trends in agent-based modeling of building occupants for energy and indoor-environmental performance analysis." Building and Environment 173 (2020): 106726. https://doi.org/10.1016/j.buildenv.2020.106726
7) Dziedzic, Jakub Wladyslaw, Da Yan, Hongsan Sun, and Vojislav Novakovic. "Building occupant transient agent-based model–Movement module." Applied Energy 261 (2020): 114417. https://doi.org/10.1016/j.apenergy.2019.114417
8) Micolier, Alice, Franck Taillandier, Patrick Taillandier, and Frédéric Bos. "Li-BIM, an agent-based approach to simulate occupant-building interaction from the Building-Information Modelling." Engineering Applications of Artificial Intelligence 82 (2019): 44-59. https://doi.org/10.1016/j.engappai.2019.03.008
9) Vellei, Marika, Simon Martinez, and Jérôme Le Dréau. "Agent-based stochastic model of thermostat adjustments: A demand response application." *Energy and Buildings* 238 (2021): 110846. https://doi.org/10.1016/j.enbuild.2021.110846
10) Chong, Adrian, Godfried Augenbroe, and Da Yan. "Occupancy data at different spatial resolutions: Building energy performance and model calibration." Applied Energy 286 (2021): 116492. https://doi.org/10.1016/j.apenergy.2021.116492

Machine Learning and Building Energy prediction:

ML approaches to building energy prediction has gained more traction owing to processing power and data availability. Such predictions can be extended beyond one building, i.e., to a university-campus or even entire cities. There are several ML approaches that may be used, however, there is no one type of

ML that fits a type of building. The selection of ML is based on dependent and independent variables; data availability including weather; frequency; and other uncertainties in data. Below, you can find some examples of individual to campus buildings' energy use prediction. Our work related to campus buildings (refer to the first article below) uses time-series data and we found that ARIMA was best suited for such data.

1) Fathi, S., Srinivasan, R.S., Kibert, C.J., Steiner, R.L., and Demirezen, E. AI-based Campus Energy Use Prediction for Assessing the Effects of Climate Change. Sustainability 2020, 12, 3223; http://dx.doi.org/10.3390/su12083223
2) Wang, Z., Wang, Y., Srinivasan R.S. (2018). Random Forest-based Hourly Building Energy Prediction. Energy and Buildings, 171 (15): 11-25; https://doi.org/10.1016/j.enbuild.2018.04.008
3) Wang, Z., Srinivasan, R.S., Shi, J. (2016). Artificial Intelligence Models for Improved Prediction of Residential Heating. ASCE Journal of Energy Engineering, 142(4) https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29EY.1943-7897.0000342
4) Amasyali, Kadir, and Nora El-Gohary. "Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings." Renewable and Sustainable Energy Reviews 142 (2021): 110714. https://doi.org/10.1016/j.rser.2021.110714
5) Lei, Lei, Wei Chen, Bing Wu, Chao Chen, and Wei Liu. "A building energy consumption prediction model based on rough set theory and deep learning algorithms." Energy and Buildings (2021): 110886. https://doi.org/10.1016/j.enbuild.2021.110886
6) Fan, Cheng, Yongjun Sun, Yang Zhao, Mengjie Song, and Jiayuan Wang. "Deep learning-based feature engineering methods for improved building energy prediction." *Applied energy* 240 (2019): 35-45. https://doi.org/10.1016/j.apenergy.2019.02.052
7) Guo, Yabin, Jiangyu Wang, Huanxin Chen, Guannan Li, Jiangyan Liu, Chengliang Xu, Ronggeng Huang, and Yao Huang. "Machine learning-based thermal response time ahead energy demand prediction for building heating systems." Applied energy 221 (2018): 16-27. https://doi.org/10.1016/j.apenergy.2018.03.125
8) Singaravel, Sundaravelpandian, Johan Suykens, and Philipp Geyer. "Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction." Advanced Engineering Informatics 38 (2018): 81-90. https://doi.org/10.1016/j.aei.2018.06.004
9) Zekić-Sušac, Marijana, Saša Mitrović, and Adela Has. "Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities." *International journal of information management* 58 (2021): 102074. https://doi.org/10.1016/j.ijinfomgt.2020.102074
10) Ahmad, Tanveer, Huanxin Chen, Ronggeng Huang, Guo Yabin, Jiangyu Wang, Jan Shair, Hafiz Muhammad Azeem Akram, Syed Agha Hassnain Mohsan, and Muhammad Kazim. "Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment." Energy 158 (2018): 17-32. https://doi.org/10.1016/j.energy.2018.05.169

Ensemble models:

Instead of using a single ML for prediction, one approach is to use ensemble models (homogenous and heterogenous). Below, you can see some excellent examples of ensemble modeling of building energy prediction.

1) Wang Z, Srinivasan R.S. (2017). A Review of Artificial Intelligence based Building Energy Use Prediction: Contrasting the Capabilities of Single and Ensemble Prediction Models. Renewable and Sustainable Energy Reviews. 75:796-808; https://doi.org/10.1016/j.rser.2016.10.079
2) Wang, Z., Srinivasan, R.S. (2017). A Review on Applications of Artificial Intelligence based Building Energy Use Prediction with a Focus on Single vs Ensemble Prediction Models – Contrasting their Capabilities. Renewable & Sustainable Energy Reviews, 75: 796-808; https://doi.org/10.1016/j.rser.2016.10.079
3) Wang, S., Zheng, P., Srinivasan, R.S. (2017). A Novel Ensemble Learning Approach to Support Building Energy Use Prediction. Energy and Buildings, 159(15): 109-122; https://doi.org/10.1016/j.enbuild.2017.10.085
4) Wang, Lan, Eric WM Lee, and Richard KK Yuen. "Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach." Applied Energy 228 (2018): 1740-1753. https://doi.org/10.1016/j.apenergy.2018.07.085
5) Dong, Zhenxiang, Jiangyan Liu, Bin Liu, Kuining Li, and Xin Li. "Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption patterns classification." *Energy and Buildings* (2021): 110929. https://doi.org/10.1016/j.enbuild.2021.110929
6) Huang, Yao, Yue Yuan, Huanxin Chen, Jiangyu Wang, Yabin Guo, and Tanveer Ahmad. "A novel energy demand prediction strategy for residential buildings based on ensemble learning." Energy Procedia 158 (2019): 3411-3416. https://doi.org/10.1016/j.apenergy.2020.115025
7) Al-Rakhami, Mabrook, Abdu Gumaei, Ahmed Alsanad, Atif Alamri, and Mohammad Mehedi Hassan. "An ensemble learning approach for accurate energy load prediction in residential buildings." IEEE Access 7 (2019): 48328-48338. DOI: 10.1109/ACCESS.2019.2909470
8) S. Kumar T.M., C. P. Kurian and S. G. Varghese, "Ensemble Learning Model-Based Test Workbench for the Optimization of Building Energy Performance and Occupant Comfort," in IEEE Access, vol. 8, pp. 96075-96087, 2020, doi: 10.1109/ACCESS.2020.2996546.
9) Zhang, Guiqing, Chenlu Tian, Chengdong Li, Jun Jason Zhang, and Wangda Zuo. "Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature." Energy 201 (2020): 117531. https://doi.org/10.1016/j.energy.2020.117531
10) Araya, Daniel B., Katarina Grolinger, Hany F. ElYamany, Miriam AM Capretz, and Girma Bitsuamlak. "An ensemble learning framework for anomaly detection in building energy consumption." *Energy and Buildings* 144 (2017): 191-206. https://doi.org/10.1016/j.enbuild.2017.02.058

Most recently, we published an article listing issues in actual implementation of ML models for building energy efficiency. Wang, Z., Liu, J., Yuan, H., Zhang, R., Srinivasan, R.S. Practical Issues in Implementing Machine Learning Models for Building Energy Efficiency: Moving Beyond Obstacles. Renewable and Sustainable Energy Reviews. Volume 143, June 2021, 110929; ttps://doi.org/10.1016/j.rser.2021.110929

Air quality and Indoor emission:

Below you can see examples of low cost, affordable indoor air quality monitoring approaches. Currently, we are using mixed methods research (using both qualitative and quantitate data) to interpret/ model/ correlate to identify factors influencing indoor air pollution in buildings. This work is still not published.

1) Zhang, H., Srinivasan, R.S., Ganesan, V. Low Cost, Multi-Pollutant Sensing System using Raspberry Pi for Realtime Indoor Air and Environmental Condition Monitoring. Sustainability 2021, 13(1), 370; https://doi.org/10.3390/su13010370
2) Zhang, H. and Srinivasan, R.S. A Systematic Review of Air Quality Sensors, Guidelines, and Measurement Studies for Indoor Air Quality Management. Sustainability 2020, 12(21), 9045; https://doi.org/10.3390/su12219045
3) Wu, Peihao, Zhaosong Fang, Hui Luo, Zhimin Zheng, Kaiyue Zhu, Yanping Yang, and Xiaoqing Zhou. "Comparative analysis of indoor air quality in green office buildings of varying star levels based on the grey method." Building and Environment (2021): 107690. https://doi.org/10.1016/j.ifacol.2019.12.430
4) Ganesh, Hari S., Kyeongjun Seo, Hagen E. Fritz, Thomas F. Edgar, Atila Novoselac, and Michael Baldea. "Indoor air quality and energy management in buildings using combined moving horizon estimation and model predictive control." Journal of Building Engineering 33 (2021): 101552. https://doi.org/10.1016/j.jobe.2020.101552
5) L. Zhao, W. Wu and S. Li, "Design and Implementation of an IoT-Based Indoor Air Quality Detector With Multiple Communication Interfaces," in IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9621-9632, Dec. 2019, doi: 10.1109/JIOT.2019.2930191..
6) Salman, N., Andrew H. Kemp, A. Khan, and C. J. Noakes. "Real time wireless sensor network (WSN) based indoor air quality monitoring system." IFAC-PapersOnLine 52, no. 24 (2019): 324-327. https://doi.org/10.1016/j.ifacol.2019.12.430
7) Shan, Ning. "Research on Indoor Environment Design Based on VR Technology and Wireless Sensor Network." Microprocessors and Microsystems 83 (2021): 103999. https://doi.org/10.1016/j.micpro.2021.103999
8) J. Kim, C. Chu and S. Shin, "ISSAQ: An Integrated Sensing Systems for Real-Time Indoor Air Quality Monitoring," in IEEE Sensors Journal, vol. 14, no. 12, pp. 4230-4244, Dec. 2014, doi: 10.1109/JSEN.2014.2359832.
9) P. Spachos and D. Hatzinakos, "Real-Time Indoor Carbon Dioxide Monitoring Through Cognitive Wireless Sensor Networks," in IEEE Sensors Journal, vol. 16, no. 2, pp. 506-514, Jan.15, 2016, doi: 10.1109/JSEN.2015.2479647.
10) Caron, Alexandre, Nathalie Redon, Patrice Coddeville, and Benjamin Hanoune. "Identification of indoor air quality events using a K-means clustering analysis of gas sensors data." Sensors and Actuators B: Chemical 297 (2019): 126709. https://doi.org/10.1016/j.snb.2019.126709

Black Box and HVAC Controls: Measuring the Right Factors

Nancy Ma, Center for Environmental Building & Design

Why model IAQ and thermal comfort?

1. Humans spend up to 90% of their time in indoors. We live, work, and learn in buildings
2. Socio-economic benefits of improved indoor environmental quality (IEQ)
3. Buildings don't use energy, people do
4. Indoor air quality (IAQ) and health problems statistics

What factors have been measuring?

1. Controllable variables
2. Control components, parameters, mode, and algorithm
3. The discrepancy of building performance and human health/thermal comfort
4. Physics-based thermoregulation models + CFD

What variables are worth measuring?

Analytical models of thermal comfort: steady state and adaptive comfort models

1. Thermal comfort: steady state models
2. Thermal comfort: adaptive comfort models
3. Application and limitations

Analytical models of IAQ

1. Models of determining CO_2 concentrations
2. Models of determining airborne contaminants concentrations

Thermal comfort and health defined data-driven system

Research articles on machine learning application for indoor climate control

Ma, Nan, Dorit Aviv, Hongshan Guo, and William W. Braham. "Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality." *Renewable and Sustainable Energy Reviews* 135 (2021): 110436.

Why model IAQ and thermal comfort?

1. Humans spend up to 90% of their time in indoors. We live, work, and learn in buildings

This study analyzed data from a web-based survey administered to 52,980 occupants in 351 office buildings over 10 years by the Center for the Built Environment. The most important parameters were satisfaction with amount of space, noise level, and visual privacy. Satisfaction with amount of space was ranked to be most important for workspace satisfaction, regardless of age group, gender, type of office (single or shared offices, or cubicles), distance of workspace from a window (within 4.6 m or further), or satisfaction level with workspace (satisfied or dissatisfied). Satisfaction with amount of space was not related to the gross amount of space available per person.

Frontczak, Monika, Stefano Schiavon, John Goins, Edward Arens, Hui Zhang, and Pawel Wargocki. "Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design." *Indoor air* 22, no. 2 (2012): 119-131.

In this review article, the authors summarize recent advances in source characterization, exposure assessment, health effects associated with indoor exposures, and intervention research related to indoor environments. They concluded that more research is needed on the interactions of multiple exposures, and the risks to certain populations (such as children, the elderly, or socioeconomically disadvantaged populations). Identification of research priorities should include input from building designers, operators, and the public health community. Research on interventions should examine a range of outcomes and potential tradeoffs and confounders, and does not necessarily need to await the identification of specific causal agents.

Mitchell, Clifford S., Junfeng Zhang, Torben Sigsgaard, Matti Jantunen, Paul J. Lioy, Robert Samson, and Meryl H. Karol. "Current state of the science: health effects and indoor environmental quality." *Environmental health perspectives* 115, no. 6 (2007): 958-964.

2. Socio-economic benefits of improved indoor environmental quality (IEQ)

This is a great book which covers origins and foundations of the built environment as a public health focus and its joint history with urban planning, transportation and land use, infrastructure and natural disasters, assessment tools, indoor air quality, water quality, food security, health disparities, mental health, social capital, and environmental justice.

Lopez, Russell P. *The built environment and public health*. Vol. 16. John Wiley & Sons, 2012.

The authors outlined the following priority research topics in below article: building-influenced communicable respiratory infections, building-related asthma/allergic diseases, and nonspecific building-related symptoms; indoor environmental science; and methods for increasing implementation of healthful building practices. Available data suggest that improving building environments may result in health benefits for more than 15 million of the 89 million US indoor workers, with estimated economic benefits of $5 to $75 billion annually.

Mendell, Mark J., William J. Fisk, Kathleen Kreiss, Hal Levin, Darryl Alexander, William S. Cain, John R. Girman et al. "Improving the health of workers in indoor environments: priority research needs for a national occupational research agenda." *American journal of public health* 92, no. 9 (2002): 1430-1440.

This study estimates some of the benefits and costs of implementing scenarios that improve indoor environmental quality (IEQ) in the stock of U.S. office buildings. The scenarios include increasing ventilation rates when they are below 10 or 15 l/s per person, adding outdoor air economizers and controls when absent, eliminating winter indoor temperatures >23°C, and reducing dampness and mold problems.

Fisk, William J., Doug Black, and Gregory Brunner. "Benefits and costs of improved IEQ in US offices." *Indoor Air* 21, no. 5 (2011): 357-367.

3. Buildings don't use energy, people do

This article argues that building users play a critical but poorly understood and often overlooked role in the built environment. To fully address the task ahead, it argues that architects need to develop their professional expertise to improve buildings and seek ways of integrating user involvement in building performance.

Janda, Kathryn B. "Buildings don't use energy: people do." *Architectural science review* 54, no. 1 (2011): 15-22.

This paper presents ten questions, highlighting some of the most important issues regarding concepts, applications, and methodologies in occupant behavior research. It is crucial to understand occupant behavior in a comprehensive way, integrating qualitative approaches and data- and model-driven quantitative approaches, and employing appropriate tools to guide the design and operation of low-energy residential and commercial buildings that integrate technological and human dimensions.

Hong, Tianzhen, Da Yan, Simona D'Oca, and Chien-fei Chen. "Ten questions concerning occupant behavior in buildings: The big picture." *Building and Environment* 114 (2017): 518-530.

The authors reviewed papers published over the last five years (from 2014 to 2019) and presented information about questionnaires, interviews, brainstorming, post-occupancy evaluation, personal diaries, elicitation studies, ethnographic studies, and cultural probe. Increasing use of qualitative methods is expected to support the spread of human-centric policies and design/control of buildings, with a consequent overall optimization of energy performance of buildings as well as the comfort of occupants.

Bavaresco, Mateus V., Simona D'Oca, Enedir Ghisi, and Roberto Lamberts. "Methods used in social sciences that suit energy research: A literature review on qualitative methods to assess the human dimension of energy use in buildings." *Energy and Buildings* 209 (2020): 109702.

The authors attempt to rethink occupant behavior and its role in building energy performance by means of review. The review focuses on four critical research topics: a) the current understanding of occupant behavior, with particular focus on window opening behavior, lighting control behavior, and space heating/cooling behavior; b) methods and techniques for collecting data on behavior and building energy performance; c) quantitative modeling of occupant behavior and building energy performance; and d) evaluation of energy saving potentials of occupant behavior based on existing literature. They concluded that the energy-saving potential of occupant behavior to be in the range of 10%–25% for residential buildings and 5%–30% for commercial buildings, based on findings of existing research.

Zhang, Yan, Xuemei Bai, Franklin P. Mills, and John CV Pezzey. "Rethinking the role of occupant behavior in building energy performance: A review." *Energy and Buildings* 172 (2018): 279-294.

4. Indoor air quality (IAQ) and health problems statistics

Below study summarizes the historical development and understanding on environmental exposures/risks and indoor air. Indoor air was believed to be a major environmental factor for more than a hundred years, from the start of the hygienic revolution, around 1850, until outdoor environmental issues entered the scene, and became dominant around 1960. Main environmental issues today are outdoor air quality, energy use, and sustainable buildings, but not indoor air quality (IAQ).

Sundell, Jan. "On the history of indoor air quality and health." *Indoor air* 14, no. s 7 (2004): 51-58.

This overview has reviewed the literature about the effects of extended exposure to low humidity on perceived IAQ, sensory irritation symptoms in eyes and airways, work performance, sleep quality, virus survival, and voice disruption.

Wolkoff, Peder. "Indoor air humidity, air quality, and health–An overview." *International journal of hygiene and environmental health* 221, no. 3 (2018): 376-390.

What factors have been measuring?

1. Controllable variables

This study presents a critical review of current modeling techniques used in HVAC systems regarding their applicability and ease of acceptance in practice and summarizes the strengths, weaknesses, applications and performance of these modeling techniques. Additionally, the performance and outcome of some of the developed models used in real world HVAC systems have been discussed.

Afroz, Zakia, G. M. Shafiullah, Tania Urmee, and Gary Higgins. "Modeling techniques used in building HVAC control systems: A review." *Renewable and sustainable energy reviews*83 (2018): 64-84.

2. Control components, parameters, mode, and algorithm

Advanced control strategies provide a more efficient way of minimizing energy demand of buildings and maintaining indoor environmental quality in accordance with global principle of sustainability, which has also proven reliable for diverse applications such as Heating, Ventilation and Air Conditioning (HVAC) control and thermal comfort control etc. The objective of this paper is to review the control strategies in buildings, particularly focusing on low energy buildings (LEB), in recent 10 years. Present work consists of why to use control strategies in buildings, categories of control strategies, research literature for building performance affected by diverse control strategies from the perspective of theoretical modelling, physical experimental study and numerical simulation investigation. Following that, more than 20 parameters affecting control performance have been analyzed and evaluated.

Wang, Yang, Jens Kuckelkorn, and Yu Liu. "A state of art review on methodologies for control strategies in low energy buildings in the period from 2006 to 2016."*Energy and Buildings*147 (2017): 27-40.

3. The discrepancy of building performance and human health/thermal comfort

The authors reviewed research fields of thermal comfort and building control, and their relationship using a data-driven approach. They found that building control focuses predominantly on energy savings rather than incorporating results from thermal comfort, especially when it comes to occupant satisfaction.

Park, June Young, and Zoltan Nagy. "Comprehensive analysis of the relationship between thermal comfort and building control research-A data-driven literature review." *Renewable and Sustainable Energy Reviews* 82 (2018): 2664-2679.

4. Physics-based thermoregulation models + CFD

This paper first reviews several thermal comfort models that address local thermal sensations and attempts to distinguish these models by their advantages, limitations and suitable ranges of applications. Then, two typical thermal comfort models, the simple ISO 14505 standard method and the comprehensive UC Berkeley thermal comfort model (UCB model), were coupled to computational fluid dynamic (CFD) numerical simulation with different process to evaluate thermal environment of a small office. The results indicated that compared with the UCB model, the ISO 14505 index could be applied with caution as a convenient method to evaluate thermal comfort in non-uniform, overall thermally neutral environments.

Cheng, Yuanda, Jianlei Niu, and Naiping Gao. "Thermal comfort models: A review and numerical investigation." *Building and environment* 47 (2012): 13-22.

The Berkeley Comfort Model is based on the Stolwijk model of human thermal regulation but includes several significant improvements. This new model proposed by the authors allows an unlimited body segments (compared to six in the Stolwijk model). Each segment is modeled as four body layers (core, muscle, fat, and skin tissues) and a clothing layer. Physiological mechanisms such as vasodilation, vasoconstriction, sweating, and metabolic heat production are explicitly considered. Convection, conduction (such as to a car seat or other surface in contact with any part of the body) and radiation between the body and the environment are treated independently. The model is capable of predicting human physiological response to transient, non-uniform thermal environments.

Huizenga, Charlie, Zhang Hui, and Edward Arens. "A model of human physiology and comfort for assessing complex thermal environments." *Building and Environment* 36, no. 6 (2001): 691-699.

The authors investigated the pollutant exposure reduction and thermal comfort that can be achieved with personalised ventilation (PV) design when a PV system is combined with two types of background air conditioning systems. For the investigation of inhaled air quality, pollutants emitted from building materials are the targeted pollutants; and for the investigation of thermal comfort, local discomfort associated with nonuniform thermal environment is focused upon. These investigations were performed by combining CFD simulation of the 3D air flow and a multi-nodal human body thermo-regulation model. The results reveal some new characteristics of the three typical air distribution designs, i.e. mixed ventilation, displacement ventilation and PV, and provide insight into the possible optimization of system combinations.

Gao, N. P., H. Zhang, and J. L. Niu. "Investigating indoor air quality and thermal comfort using a numerical thermal manikin." *Indoor and built environment* 16, no. 1 (2007): 7-17.

What variables are worth measuring?

This paper reviews the existing systems and proposes an innovation in HVAC systems management: a system that tracks the occupants' preferences, learns from them, and manages HVAC automatically. We show that ambient intelligent systems can be used to control a building's Energy Management Systems (EMS), effectively reducing energy consumption while maintaining acceptable comfort levels. Our results indicate that employing a k-means machine learning technique enables the automatic configuration of an HVAC system to reduce energy consumption while keeping the majority of occupants within acceptable comfort levels.

Carreira, Paulo, António Aguiar Costa, Vitor Mansur, and Artur Arsénio. "Can HVAC really learn from users? A simulation-based study on the effectiveness of voting for comfort and energy use optimization." *Sustainable cities and society* 41 (2018): 275-285.

The objective of this paper is to highlight evidence and variables from empirical and deterministic models, which are combined in analytical models that current machine learning techniques often overlook. Eighteen critical variables are extracted from forty-five works closely related to the field (as listed in the table).

Ma, Nan, Dorit Aviv, Hongshan Guo, and William W. Braham. "Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality." *Renewable and Sustainable Energy Reviews* 135 (2021): 110436.

Summary of input variables that are worthy measuring.

Subgroups[a]	**Variables of IAQ-related thermal comfort and health**[b]	**Topics**[c]
Environmental survey	Outdoor temperature (T_{out})	TC
	Wind velocity (v_a)	TC+H
	Outdoor relative humidity (RH_{out})	H
	Outdoor contaminants concentration (C_{out})	H
Design	Room dimensions[d] (Dim)	H
	Ceiling height (H)	H
	Total surface area (A)	TC+H
	Penetration factor through envelope/door (P)	H
Material selection	Radiant temperature (T_{MR})	TC
	Temperature of surface[e] (T_i)	TC
Operation	Indoor relative humidity (RH_{in})	TC+H
	Volume flow rate (Natural, Mechanical, Infiltration) (Q)	TC+H
	Indoor temperature (T_a)	TC+H
	Air density[f] (ρ)	H
	Contaminants generation/deposition/removal concentrations/rates (G)	H
	Number of occupants (N)	H
	Exposure time (t)	TC+H
	Air exchange rate (E_X)	H

[a] Total eighteen input variables are arranged based on the different phases of buildings;

[b] The listed variables are given its abbreviation in parentheses to keep consist in Nomenclature, figures and tables;

[c] TC and H represent that this variable stem from topics of thermal comfort and health respectively; TC+H means thermal comfort and health fields both echo and cover this variable;
[d] Analytical models uses volume of a space more often, while it is determined from size of the space and ceiling height;
[e] Temperature of surface implies for surface temperatures of each material in accordance to air temperature;
[f] Air density is hardly measurable, but is correlated with air pressure, temperature, humidity and dew point.

Analytical models of thermal comfort: steady state and adaptive comfort models

1. Thermal comfort: steady state models

These two articles are the classic readings where Fanger proposed his thermal comfort models.

Fanger, Poul O. "Thermal comfort. Analysis and applications in environmental engineering." *Thermal comfort. Analysis and applications in environmental engineering.* (1970).

Fanger, Poul O. "Calculation of thermal comfort-introduction of a basic comfort equation." *ASHRAE Transacions* 73 (1967).

2. Thermal comfort: adaptive comfort models

These three articles are the typical adaptive comfort models. The authors proposed different correction coefficient to modify Fanger's model for different building types and ventilation modes.

Yao, Runming, Baizhan Li, and Jing Liu. "A theoretical adaptive model of thermal comfort–Adaptive Predicted Mean Vote (aPMV)." *Building and environment* 44, no. 10 (2009): 2089-2096.

Humphreys, Michael A., and J. Fergus Nicol. "The validity of ISO-PMV for predicting comfort votes in every-day thermal environments." *Energy and buildings* 34, no. 6 (2002): 667-684.

Atmaca, Ibrahim, Omer Kaynakli, and Abdulvahap Yigit. "Effects of radiant temperature on thermal comfort." *Building and environment* 42, no. 9 (2007): 3210-3220.

3. Application and limitations

This paper looks critically at the foundation and underlying assumptions of the adaptive model approach and its findings.

Halawa, Edward, and J. Van Hoof. "The adaptive approach to thermal comfort: A critical overview." *Energy and Buildings* 51 (2012): 101-110.

Analytical models of IAQ

A comprehensive summary of standards and guidelines as developed by various worldwide organizations.

Abdul-Wahab, Sabah Ahmed, Stephen Chin Fah En, Ali Elkamel, Lena Ahmadi, and Kaan Yetilmezsoy. "A review of standards and guidelines set by international bodies for the parameters of indoor air quality." *Atmospheric Pollution Research* 6, no. 5 (2015): 751-767.

The primary IAQ standards and guidelines stipulated by WHO and the United States' authentic agencies.

American Society of Heating, Refrigerating and Air Conditioning Engineer (ASHRAE)

ANSI/ASHRAE Standard 621-2016 Ventilation for acceptable Indoor air Quality 2016

Occupational Safety and Health Administration (OSHA)

OSHA, OS. "OSHA Technical Manual-Section III: Chapter 2: Indoor Air Quality." (1999).

US Environmental Protection Agency (EPA)

Mudarri, David H. "Building codes and indoor air quality." *US EPA* (2010).

Koontz, M. D., G. M. Zarus, M. J. Stunder, and N. L. Nagda. "Air toxics risk assessment." (1991).

World Health Organization (WHO)

World Health Organization. "WHO guidelines for indoor air quality: selected pollutants." (2010).

World Health Organization. *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization, 2006.

1. Models of determining CO_2 concentrations

The results indicate that, compared to the existing fixed ventilation rate strategy at which the ventilation rate is always 5% of the total supply air flow, a cooling coil energy savings of 0.03% and 1.86% can be achieved using an occupancy detection control strategy under the new ASHRAE 62.1 and old ASHRAE 62 respectively, while preserving thermal comfort and indoor air quality.

Ng, Malcolm Owen, Ming Qu, Pengxuan Zheng, Zhiyuan Li, and Yin Hang. "CO_2-based demand controlled ventilation under new ASHRAE Standard 62.1-2010: a case study for a gymnasium of an elementary school at West Lafayette, Indiana." *Energy and Buildings* 43, no. 11 (2011): 3216-3225.

Experiments were conducted in a school office by measuring indoor CO_2 concentrations and pressure differences between the return air vent and space. Excellent agreement was obtained. At least 0.998 R^2 values were obtained for fitting measured CO_2 concentrations when conducting MLE for estimating space air change rate, and the corresponding residual plots showed no pattern and trend. The estimated numbers of occupants were same as the actual ones. Furthermore, the predicted space air change rates showed great consistencies with those from CO_2 equilibrium analysis. The model is simple, handy and effective for practical use. Moreover, the model is also capable for dealing with time-varying space air change rates.

Lu, Tao, Anssi Knuutila, Martti Viljanen, and Xiaoshu Lu. "A novel methodology for estimating space air change rates and occupant CO_2 generation rates from measurements in mechanically-ventilated buildings." *Building and Environment* 45, no. 5 (2010): 1161-1172.

2. Models of determining airborne contaminants concentrations

The review paper publication where you can find online elaborates and specifies all the equations listed on the slides. In this summary I just pick three great articles that are worthy reading in full text:

Nazaroff, William W. "Indoor particle dynamics." *Indoor air* 14, no. Supplement 7 (2004): 175-183.

Walker, Iain S., and Max H. Sherman. "Effect of ventilation strategies on residential ozone levels." *Building and environment* 59 (2013): 456-465.

Ye, Wei, Doyun Won, and Xu Zhang. "A practical method and its applications to prioritize volatile organic compounds emitted from building materials based on ventilation rate requirements and ozone-initiated reactions." *Indoor and Built Environment* 26, no. 2 (2017): 166-184.

Thermal comfort and health defined data-driven system

There are two review papers which are also comprehensive to summarize ANN structure, input features, outcome variables, and how machine learning techniques help forecast thermal comfort and IAQ.

Enescu, Diana. "A review of thermal comfort models and indicators for indoor environments." Renewable and Sustainable Energy Reviews 79 (2017): 1353-1379.

Table 7 *(continued)*

Article	Models	Number of hidden neurons	Total number of input neurons	Input data												
				Environment floor level	Orientation of the buildings	Opening condition of inner surface (envelope)	Outdoor Relative Humidity	Opening condition of outer surface (envelope)	Indoor Air Quality	Wind Velocity (and direction)	Solar Radiation	Chilled water valve opening level	Outdoor air damper opening level	Chilled water temperature	Outdoor Air Temperature	Prediction error (measurements vs. model outputs)
		Cavity temperature	Radiator electrical power	Activity in the building	Indoor Relative Humidity	Indoor Air Temperature	Hot water temperature	Sky cloudiness	Auxiliary Heating Power	Supply air relative humidity	Supply air temperature from the air handling unit	Supply air flow rate from the air handling unit	Heating control signal (valve position)	Time of the day	Hour	
[210]	RBF				1	1										
[211]	RBF, LM				2	2										
[212]	RBF				1	1										
[215]	RBF					7–9										
[216]	ARX					3			2							
[218]	ARX					1		5								
[219]	ARX				1	1	1			1	1	1				
[220]	ARX/ ARMAX/ OE				1/ 2/ 2	2/ 2/ 2					2/ 2/ 2	2/ 2/ 2		2/ 2/ 2		
[221]	ARX/ ARMAX		2*			2*								1		
[223]	ARX, LM				2	3										
[224]	ARX/ ARMAX					4		1								
[203]	MLP (n time intervals)															
[205]	MLP, LM					6										
[206]	MLP, LM			12	12											
[207]	MLP,LM			3												
Article	Models	Input data														
		Time of the year	Time of the month	Rain	HVAC reference temperature	Surface of investigated environment	Ratio between surfaces	Mean transparent surface	Thermal transmittance	Periodic thermal transmittance	Thickness	Phase shiftand attenuation factors	Surface mass	Heating system status	Room carbon dioxide concentration	
[210]	RBF															
[211]	RBF, LM															
[212]	RBF															
[215]	RBF															
[216]	ARX															
[218]	ARX															
[219]	ARX														1	
[220]	ARX/ ARMAX/ OE															
[221]	ARX/ ARMAX															
[223]	ARX, LM															
[224]	ARX/ ARMAX															
[203]	MLP (n time															

(continued on next page)

Wei, Wenjuan, Olivier Ramalho, Laeticia Malingre, Sutharsini Sivanantham, John C. Little, and Corinne Mandin. "Machine learning and statistical models for predicting indoor air quality." Indoor Air 29, no. 5 (2019): 704-726.

Table 2 Summary of IAQ prediction studies using artificial neural networks

Reference	Data	Data transformation	Indoor type	Pre-analysis	Network	Inputs	Outputs	No. of hidden layers
83	BASE study: 100 office buildings from 10 geographic/climatic regions		Office	None	Feed-forward back-propagation network	Indoor TVOC, formaldehyde, CO_2, $PM_{2.5}$, airborne fungi and bacteria, temperature, relative humidity, light, and noise	Building-related symptom index	1
63	Daily average values in S subway station, Seoul, Korea. March 1, 2007-July 13, 2008	No	Subway station	PLS	Recurrent neural network	Indoor $PM_{2.5}$ and PM_{10} of the day before, current indoor NO_x	Current daily indoor $PM_{2.5}$	1
84	BASE study: 100 office buildings from 10 geographic/climatic regions	No	Office	None	Feed-forward back-propagation network	Indoor temperature, relative humidity, air velocity, CO_2, TVOC, formaldehyde, $PM_{2.5}$, airborne fungi and bacteria	Indoor air quality index	2
85	8 apartments from four bedrooms and six living rooms in an apartment building located in Kuopio, Finland. May-October 2011	No	Apartment	None	Multilayer perceptron neural network	Indoor temperature and relative humidity	Indoor CO_2	NA
86	150 workdays in Beijing campus, China. December 23, 2013-May 9, 2014	No	Office	None	Forward network	Outdoor air quality index, indoor air quality index, outdoor temperature, humidity, atmospheric pressure, and wind speed	Air purification time to reach the goal (indoor $PM_{2.5} \leq 35$ μg/m^3 or reaches steady state)	1
46	500 Latin-Hypercube samples of the results predicted by the CONTAM model	No	Dwelling	Sensitivity analysis	Cascade forward network	Indoor temperature, generation rate of internal $PM_{2.5}$, kitchen window opening area	Indoor $PM_{2.5}$	2
52	Three buildings measured in 2010-2011	No	Office and shop	None	Feed-forward time-delay neural network	Time of day, barometer level pressure, sea level pressure, outdoor temperature, relative humidity, wind speed, wind direction, Pasquill atmospheric stability class, global solar radiation, outdoor NO_2 and $PM_{2.5}$	Indoor NO_2 and $PM_{2.5}$	1
47	The measurements were taken in each site for three consecutive days during school hours	No	School	Stepwise regression	Feed-forward back-propagation network	Outdoor and $PM_{2.5\text{-}10}$, indoor CO_2 and relative humidity	$PM_{2.5}$	1

Research articles on machine learning application for indoor climate control

This research article explains extensively on sensor network deployment, data collection, and learning model development. For a period of five months, the resulting learning-based temperature preference control (LTPC) was applied to a cooling system of an office space under real-world conditions. The experimental results indicate that occupant preferences in the individual rooms differ from each other in both time horizon and temperature levels. The results report energy savings of between 4% and 25% as compared to static temperature setpoints at the low values of preferred temperature ranges.

> Peng, Yuzhen, Zoltán Nagy, and Arno Schlüter. "Temperature-preference learning with neural networks for occupant-centric building indoor climate controls." Building and Environment 154 (2019): 296-308.

Our review study only focuses on environmental parameters, however much research collected physiological data and used them to predict thermal comfort/IAQ. For example:

This paper proposes a personal TSI prediction method termed as the enhanced Predicted Thermal State (ePTS) method by sensing physiological parameters namely, hand skin temperature and pulse rate, along with the ambient air temperature. The ePTS method achieves the highest accuracy at over 97%, outperforming the PTS model (82%), and other physiology based methods (82%–94%).

> Chaudhuri, Tanaya, Yeng Chai Soh, Hua Li, and Lihua Xie. "Machine learning driven personal comfort prediction by wearable sensing of pulse rate and skin temperature." Building and Environment 170 (2020): 106615.

Using combined skin temperatures from different body segments can improve the model to over 90% accuracy. Results show that three skin locations contained enough information for classification and more would cause the curse of dimensionality.

> Dai, Changzhi, Hui Zhang, Edward Arens, and Zhiwei Lian. "Machine learning approaches to predict thermal demands using skin temperatures: Steady-state conditions." Building and Environment 114 (2017): 1-10.

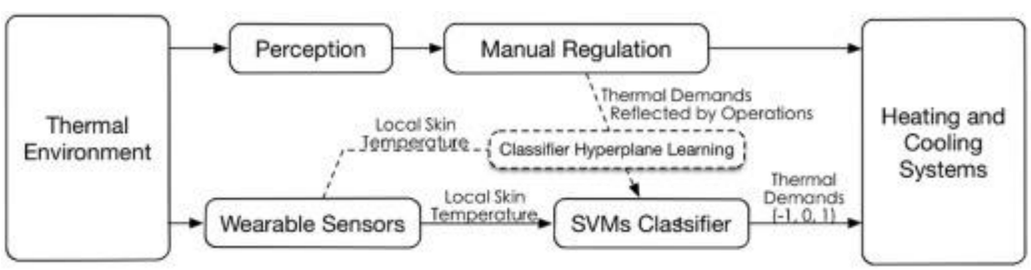


Fig. The controlling concept of SVM classifier based on skin temperature.

Machine learning-based electroencephalogram (EEG) pattern recognition methods as feedback mechanisms were investigated. Results showed that EEG theta band (4–8 Hz) correlated with subjective perceptions, and EEG alpha band (8–13 Hz) correlated with task performance. These EEG indices could be utilized as more objective metrics in addition to questionnaire and task-based metrics. For the machine learning-based EEG pattern recognition methods, the linear discriminant analysis (LDA) and support vector machine (SVM) classifiers can classify mental states under different indoor air quality conditions with high accuracy.

Shan, Xin, En-Hua Yang, Jin Zhou, and Victor WC Chang. "Neural-signal electroencephalogram (EEG) methods to improve human-building interaction under different indoor air quality." Energy and Buildings 197 (2019): 188-195.

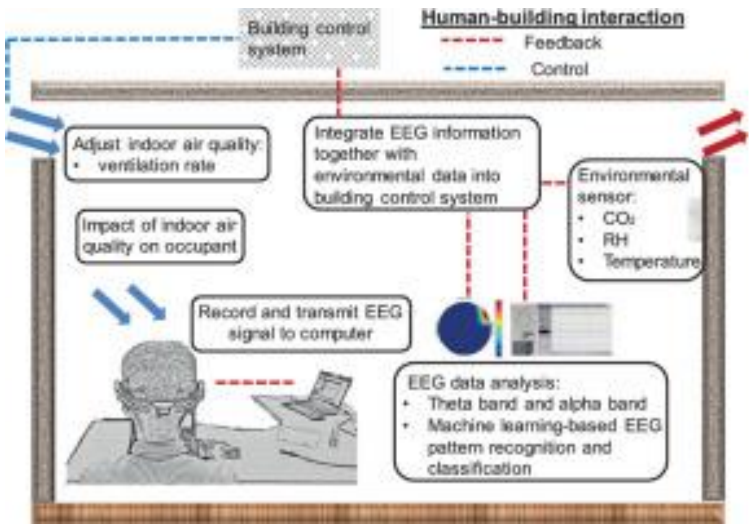


Fig. Machine learning-based EEG pattern recognition methods as real-time feedback mechanisms have good potential to improve the human-building interaction

Application of Black & White Energy Models @ UPenn
Deriving actionable information from available data

Alex Waegel, PhD, Center for Environmental Building and Design

Carbon Neutrality and Energy Models

Since President Amy Gutman first committed the University of Pennsylvania to the *American College and University Presidents Climate Commitment* (ACUPCC) in 2007, annual carbon footprints have shown that the vast majority of our greenhouse gas emissions are the result of energy consumption in the built environment. Early plans to achieve carbon neutrality focused on reducing energy consumption in the built environment through projects such as recommissioning systems, renovating buildings to improve the envelope, or efficiency gains through equipment replacement. However, there was initially little to no submetering of steam or chilled water consumption at the building level which limited the ability to accurately gauge the true potential for reduction in the built environment through those measures.

The following sections describe how building level energy consumption data has been acquired and utilized through three phases corresponding to iterations of the University's five-year plan towards carbon neutrality. From an energy modeling perspective, the focus of each phase shifted from data acquisition to data-driven black box models to low-order white box models in order to best utilize the data available to us to identify the potential for reductions in the built environment. As the construction and development of both white and black box models has been covered in other modules, the discussion will center on how these tools were applied to a real-world scenario to achieve actionable results.

Climate Action Plan I

The first Climate Action Plan (CAP, 2009) was released in 2009 was intended to provide a framework for how the University would achieve carbon neutrality by 2042. A centerpiece of the plan was a chart showing the historical carbon footprint for the University through to the current year, FY09. Beyond FY09 historical data is replaced by a wedge diagram that projects the FY09 carbon footprint through to FY42. A baseline was calculated by assuming that historical growth would continue. The goal of neutrality by 2042 was envisioned to be achievable primarily through a series of programs that would rightly be focused on the built environment.

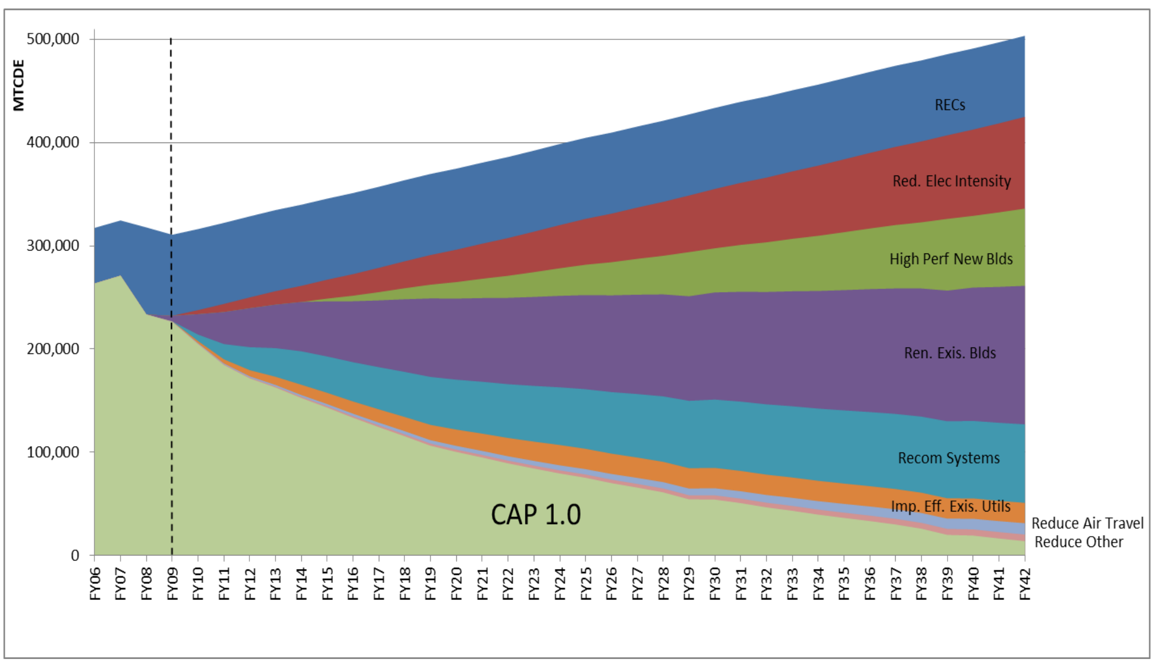


Figure 1- Wedge Diagram for CAP1

A significant hinderance to this effort was the lack of actual data on the energy performance of most individual buildings on campus. While some limited submetering existed, for the most part the only way to capture the steam and chilled water usage was at the campus level. Without this data at the building level, it was not possible to link the goals for reductions laid out in the Climate Action Plan to specific renovation and recommissioning projects in University properties. Aware of this deficit, the University began to install meters for steam and chilled water in the majority of the buildings that are connected to those campus level loops. It was a process that took several years and was not completed until 2014.

The timing of the initial meter data was such that it coincided with the preparation for the second iteration of the 5-year Climate Action Plan. Unlike before it was now possible to consider the energy consumption of individual buildings. Unfortunately, though the meter installation was largely complete by this juncture, there was a minimal historical record for most of those sensors. Further, the sensors were subject to a calibration period during which the returned data was unreliable. Despite this, an annual energy profile was created for all metered buildings by combining actual meter data, where it was available, with estimates generated by a very simple low-order white box model to fill in any gaps.

This allowed for each building's energy consumption to be compared against benchmarks for its building type to gauge its overall performance and estimate potential reductions. This analysis showed that the first Carbon Action Plan had likely overestimated the potential for achieving energy reductions in the built environment by not considering the overlapping impacts of different effects and the limits of reductions that could be achieved through renovation and recommissioning. It also showed the limits of the data that was being acquired in its raw form. While in this iteration the targets are refined, in that buildings that seem to be higher energy consumers can be investigated, the data itself is messy and does not provide insight into specific actions that could be taken to address the poor performance. It was clear that a deeper understanding of energy consumption at the building level would be required to provide actionable information.

BPAT+ Normative Model

A very simple white box model that was based on same ISO framework as SimPyBuild. This model only considered a single zone, envelope areas, materials, orientations, system types, plug-loads, light-loads, and schedules. The model was used to estimate energy consumption for unmetered buildings, but outputs could not be evaluated against metered consumption to verify accuracy so uses were limited.

Climate Action Plan 2

Drawing on the new meter data for steam and chilled water use at the building level, Climate Action Plan 2.0 (CAP2) was developed for FY14-FY19 was able to reevaluate the generic proposals made in CAP1 to determine what potential effect they may have on the actual building stock at the University. Further, one of the initiatives from CAP1 had led to a series of renovation and lighting projects under a program called the Century Bond which allowed better evaluation of the costs and energy reductions associated with these types of renovations. This data confirmed what had been previously suspected, that there were fewer overall reductions possible in the built environment than had been assumed in CAP1, but also that pursuing the deepest of those reductions would carry a steep price tag.

By basing CAP2 on the proposals of CAP1, the new data on building energy use, and the costs and impacts of the projects being explored under the Century Bond it is clear how significant the over-estimates of CAP1 were regarding the potential for carbon reductions in the built environment. While this made it apparent that alternative means of achieving neutrality were necessary it also highlighted that only a fraction of the potential reductions in the built environment were being addressed by the projects proposed under the Century Bond. A significant task remaining if those projected reductions were to be realized would be to thoroughly evaluate the building stock and to identify the actual projects that could be undertaken in the worst performers to reduce energy consumption on a scale comparable to that seen in the Century Bond.

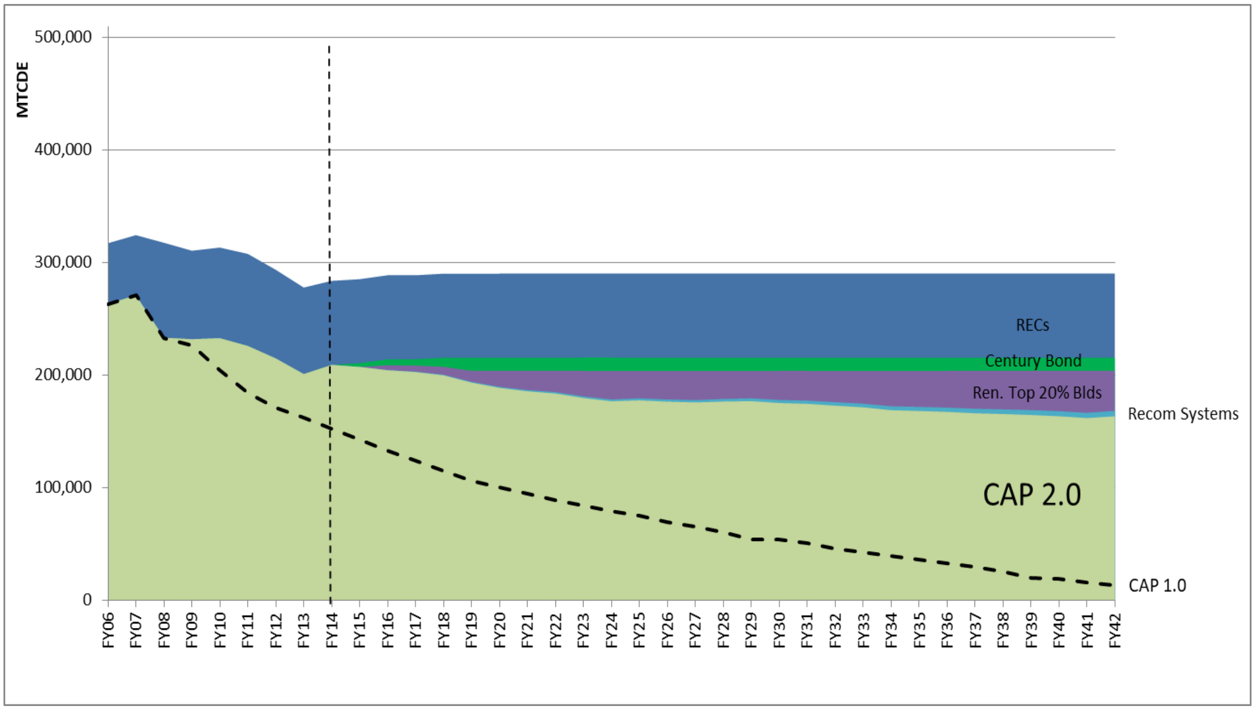


Figure 2- Wedge diagram for CAP2

The newly acquired meter data provided the most promising avenue to explore this question but several issues remained with the data itself, primarily that long spans of data from individual meters was frequently missing or had been returned from a meter prior to calibration. In order to clean the data, values for the faulty or missing data would need to be estimated and inserted. To this end a data-driven black box model was created that models the relationship each building has between its energy electric, steam, and chilled water consumption and external variables including weather conditions and the date.

In addition to cleaning the raw meter data and allowing for better evaluation of the scale of energy consumption between buildings, the black box model would also be utilized for two further analyses that would further the goal of identifying potential reductions in the built environment: 1) feature importance analysis to indicate the external variables to which energy consumption in a building is sensitive and 2) using the model to identify buildings whose energy performance has declined compared to what would be predicted based on prior observations correlating energy use to external variables.

Black Box Model

This model uses a random forest regressor to predict the energy consumption of a individual building for electricity, steam, and chilled water. (Braham et al, 2016) Other machine learning models were considered and evaluated, but this configuration consistently yielded more accurate results. (Amasyali, 2018; Bordeau, 2019, Wang, 2017) The attributes used to train this model and for predictions are month, hour, temperature, relative humidity, wind, precipitation, cloud cover, solar irradiance, and snowfall. The energy meter data is received in 15-minute increments, but these are aggregated to hourly values as that is the finest granularity available for the weather data that was purchased.

The black box model aides the analysis of University buildings in two ways. Firstly, it takes the meter data, which is often messy or incomplete, and interpolates a clean data set for each year using the predictions of the model. This increases confidence in the data that is being returned by the meters and allows for the aggregation of the data into larger blocks of time so that buildings absolute consumption can be determined on a monthly or annual basis. These aggregated figures allow for the traditional methods of evaluating the energy performance of a building: normalization and benchmarking using the EUI (annual energy / area). The EUIs and benchmarks allow for a direct comparison of similar types of buildings accounting for their scale and also gives an indication of what buildings of each type could be capable of achieving in terms of efficiency.

Secondly, once trained the model provides a feature importance analysis. This result shows the relative level of impact that each of the external variable features had on the predicted values from the model. Thus, one might be able to say that a particular building's steam consumption is highly sensitive to just temperature while another's might be sensitive to both wind and temperature. While this analysis does not spell out the specific issue with the building, it does provide insight into what might be occurring within the building to yield those results and suggest starting points for further investigation.

Black Box Model Data Cleaning Pipeline:

1. Obtain 15-min meter data for individual buildings from Facilities and Real-Estate Services
2. Obtain 1-hr weather data from a commercial vendor for the corresponding timeframe
3. Aggregate 15-min meter data to 1-hr increments to match weather data
4. Flag any missing values in the meter data
5. Flag any outlier values in the meter data using double median absolute deviation outlier detection
6. Create a data frame that joins the trusted meter data (unflagged) with the purchased weather data on the date and the hour
7. Train a random forest regressor model on the joined data frame for each energy type and building which yields a trained model and a feature importance analysis
8. Use the trained random forest model to predict all values flagged as missing or as outliers and return a copy of the original data frame with the flagged values replaced by the predicted values
9. Save a copy of the cleaned data and feature importance analysis to disk

Finally, the trained model can be used for auto-benchmarking. This is the process of training a model on the known weather conditions and energy consumption for a given time frame and then using the trained model to predict the energy consumption given the weather conditions in other time frames. This allows for a comparison of the energy consumption that was expected based on the weather conditions against the actual energy consumption that was metered.

Two possible uses of this analysis are the measurement of improvement due to renovations, recommissioning, or operational changes within a building and the identification of system failures or degradation. The utility of the auto-benchmark is the same for both cases, it allows for the comparison of energy data across different time frames by considering the relationship between energy and weather. Without considering weather and other external influences, it can be difficult to determine whether a building's energy consumption patterns have changed or if the deviation detected is more the result of varying weather conditions.

Black Box Model Auto-Benchmark Pipeline:

1. Select baseline year and years for auto-benchmark analysis
2. Load cleaned energy data and weather data for the baseline year from disk into data frames
3. Join the two data frames on date and hour
4. Train a random forest regressor model on the joined data frame for each energy type
5. For each year to be included in the auto-benchmark analysis, load the corresponding weather data into a data frame
6. Predict values for electric, steam, and chilled water consumption using the hourly weather data for the auto-benchmark years as instances to be evaluated by the trained model
7. Populate a data frame with the predicted values for the auto-benchmark years
8. Compare values predicted by model trained on baseline year against actual consumption in auto-benchmark years

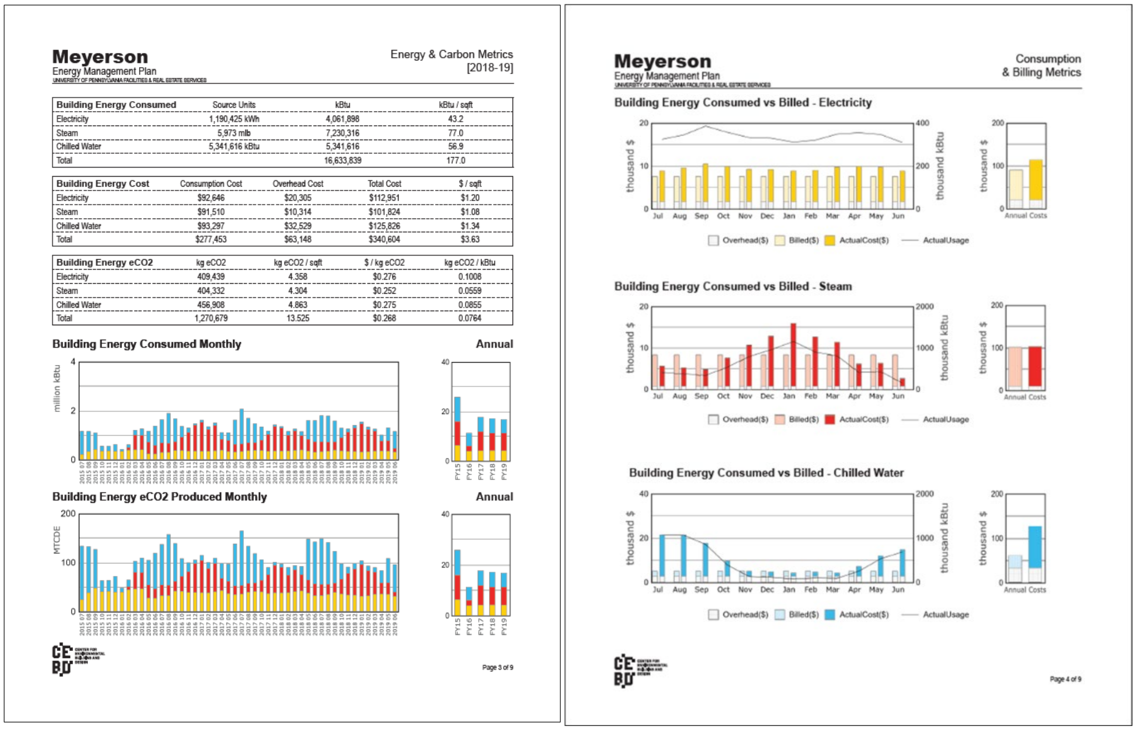
For further analysis and generation of reports, the weather data, cleaned energy data, and auto-benchmark projections are all stored to disk in a relational database along with organizational and construction data for each building.

Black Box Model Reporting

During Climate Action Plan 2 an effort was undertaken to create an annual report for each metered building which could be used to provide a broad update to building managers and administrators on the energy performance of the buildings in their portfolios. The Annual Energy Reports combine the meter data and the outputs of the black box model with the organizational data and basic construction information to provide a snapshot of the buildings energy performance for that year compared to its own history as well as compared to similar buildings. The metrics and analyses contained within the AER are almost completely ignorant of the actual inner workings of any building and so the information contained within serves better as a warning or indicator that the building should be more deeply investigated rather than providing advice on how to address specific identified issues. (Chalal, 2016)

Identification and Metrics

The first page of the annual energy report focuses on two sections. The first contains basic identification information for each building such as the building ID, official name, address, height, area, department, and names of contacts. The second section contains totals on the monthly and annual energy consumption of the building along with the resulting emissions and costs from that consumption without accounting for the type of building, the weather conditions, or even the size of the building. This provides basic statistics about the energy performance of the building but no points of comparison by which good or poor performance might be judged. The second page of the report is devoted entirely to billing and costs data for each of the three data types.



Meyerson
Energy Management Plan
UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Energy & Carbon Metrics
[2018-19]

Building Energy Consumed	Source Units	kBtu	kBtu / sqft
Electricity	1,190,425 kWh	4,061,898	43.2
Steam	5,973 mlb	7,230,316	77.0
Chilled Water	5,341,616 kBtu	5,341,616	56.9
Total		16,633,839	177.0

Building Energy Cost	Consumption Cost	Overhead Cost	Total Cost	$ / sqft
Electricity	$92,646	$20,305	$112,951	$1.20
Steam	$91,510	$10,314	$101,824	$1.08
Chilled Water	$93,297	$32,529	$125,826	$1.34
Total	$277,453	$63,148	$340,604	$3.63

Building Energy eCO2	kg eCO2	kg eCO2 / sqft	$ / kg eCO2	kg eCO2 / kBtu
Electricity	409,439	4.358	$0.276	0.1008
Steam	404,332	4.304	$0.252	0.0559
Chilled Water	456,908	4.863	$0.275	0.0855
Total	1,270,679	13.525	$0.268	0.0764

Figure 3- Energy, Carbon, Cost Metrics

Regional and Campus Benchmarks

The third page of the Annual Energy Report shows the total energy consumption of each building compared against benchmarks. The benchmarks use energy data normalized by the area to compare the EUI of the building against the EUI of other buildings which are used for similar purposes. Normalization by the area of the buildings accounts for differences in scale. Restricting comparisons of the benchmarks to buildings of similar types ensures that the purpose of the building is considered, that is we should not expect a laboratory building to consume energy in the same way as an office building due to differences in its schedule of use and equipment installed.

There are two separate pools of buildings that any University building might be benchmarked against representing buildings used for a similar purpose in our region and buildings used for a similar purpose on campus. Both the definition of the types and the energy data for the buildings in the region were obtained from the Building Performance Database (BPD) which is a large repository of energy data from buildings across the country maintained by Berkley Labs (https://buildings.lbl.gov/cbs/bpd). Each building on campus was assigned a type corresponding to one of the building types defined by the BPD. Then its EUI was compared to the distribution of EUIs for buildings in the Philadelphia region of the same type drawn from the BPD. Specific points of comparison were generated for each type using the distributions of EUIs returned by the BPD. These were the 25th percentile, the median, and the 75th percentile values, which were used to bound normal behavior.

Having typed the campus buildings according to the categories defined by the BDP, the next benchmark compares a building's EUI against that of similarly typed buildings from the campus portfolio. While these samples are too small for statistical analysis, they a much more representative sample for comparison, accounting for things like the urban neighborhood, energy sources, and even administration and management. The 25th percentile and 75th percentile benchmarks from the regional analysis are superimposed across all for additional context.

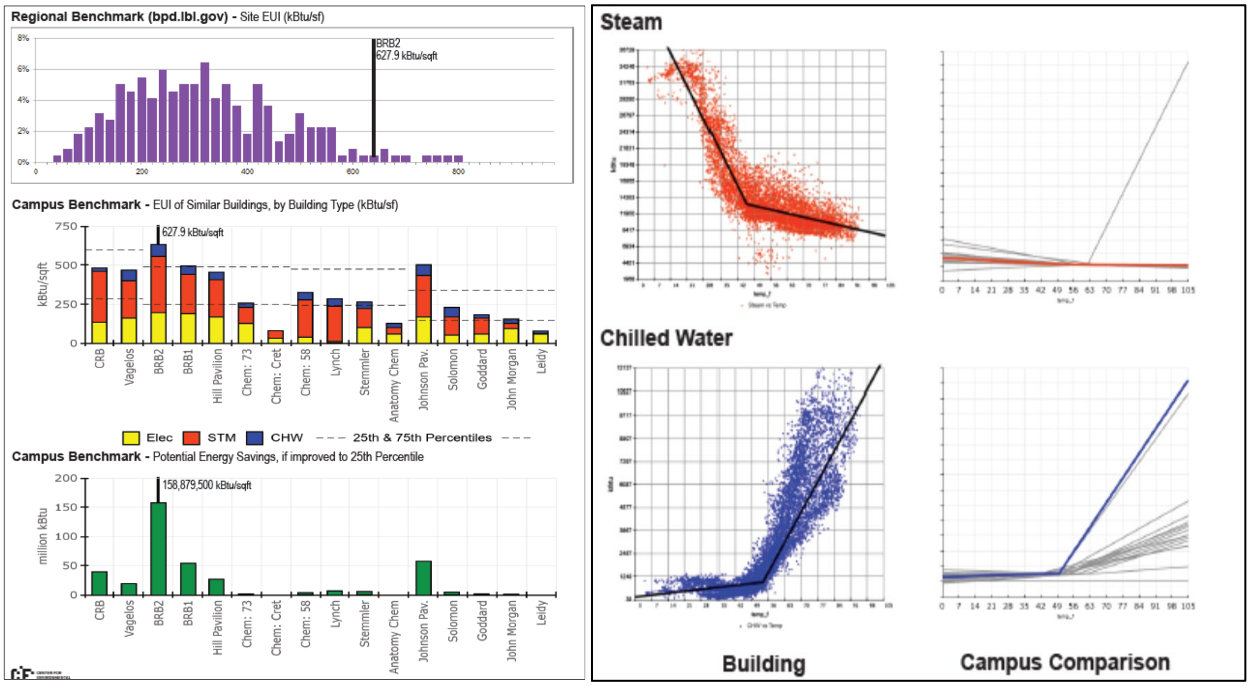


Figure 4- Building Benchmarks & Energy Signature

A final analysis can be derived from the combination of these two benchmarks which estimates the overall potential for energy reductions exists within each building of a similar type. This is done by using the 25th percentile benchmark as an "achievable goal", a target that should be theoretically reachable as at least one quarter of similar buildings have done so. For each building the difference between the EUI and the 25th percentile is determined. This value is un-normalized by multiplying it by the building area. It now represents the amount of energy that is estimated could be saved if the building were improved to the 25th percentile benchmark. Any building already performing at a better level than the 25th percentile is assigned a 0 by default. This now allows groups of buildings to be compared based on the total energy that can be saved.

Energy Signatures

The energy signatures displayed on the fourth page of the Annual Energy Reports are really a combination of several charts. The energy signature of a building considers the relationship that an energy source has with some other external variable, such as those that are used in the training of the black box model. While any attribute could be analyzed in this fashion, the most common variable to consider is temperature and the relationship that it has with the energy used for heating and cooling. The goal of an energy signature is to determine three pieces of information: does a building have different modes of operation based on a change point in the attribute, what is the correlation between consumption and the attribute in each mode of operation, and how do these calculated values compare to those found in other similar buildings.

The first element of the energy signatures developed is a scatterplot of the energy consumption vs. temperature. This step may visually reveal the magnitude of correlation a building has with temperature and identify the temperatures at which the building's mode begins to utilize the heating and cooling systems, however to mathematically confirm this, the second element of the energy signature is a change point analysis. A change point analysis iteratively considers a range of potential temperatures to serve as the change point. The data is then divided by that temperature and single variable linear regression used to determine the slope and intercept of a best fit line for each segment. The combined root mean squared error is then calculated for each change point evaluated to find the on that yields the minimal error between the observed data and the calculated best fit lines. (Kissock, Haberl, and Claridge 2002) Visually, the calculated lines are overlayed on the scatter plot for that building as well as compared against the change point diagrams for other buildings of the same type.

Auto-Benchmark

While all the previous analyses depended on the cleaned output of the black box model, none leverage the full capabilities of a data-driven model. The auto-benchmark does so by moving beyond simply cleaning data to analyzing how the building is behaving. Unlike the other benchmarks that used other buildings as the point of comparison for a building's performance, the auto-benchmark instead relies on the historical behaviors of the building itself and rather than comparing different physical structures, it compares different periods of time.

The auto-benchmark is derived by training the black box model on energy and weather data from the span of time that will serve as the baseline period. Once trained, it will make predictions based on the relationships observed between the attributes (the external calendar and weather variables) and energy consumption of each type. This prediction represents what the expected behavior of the building would be for those conditions assuming that nothing has changed. (Li, 2014)

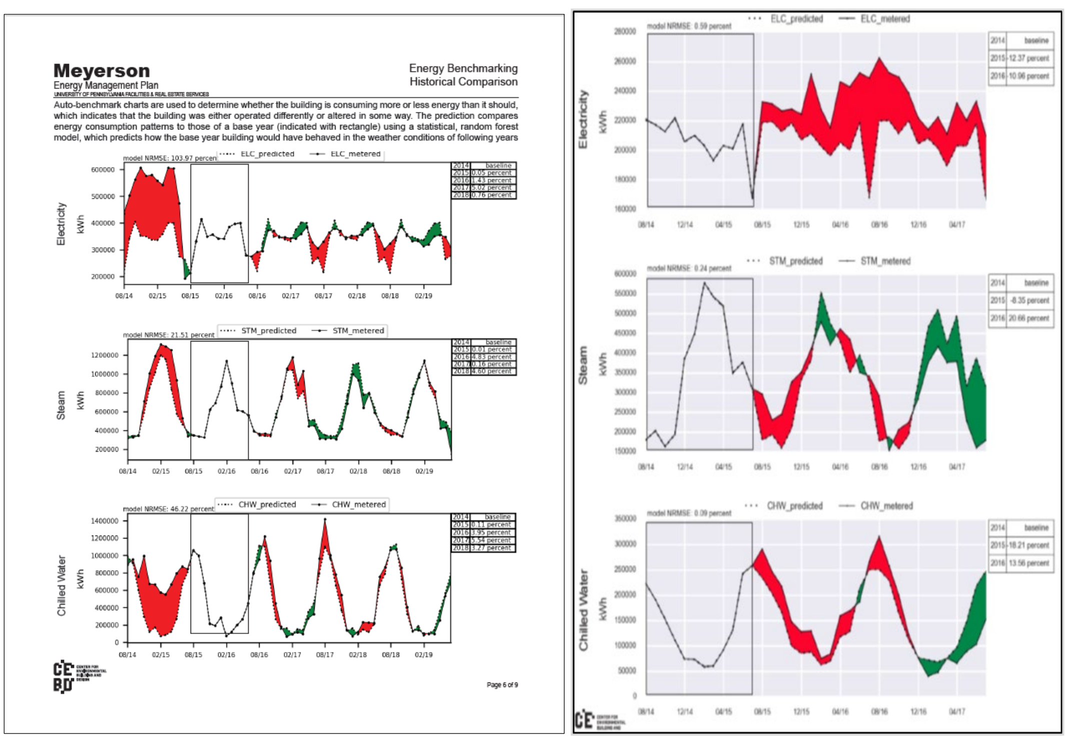


Figure 5- Building Auto-benchmarks

The model can then be applied to predict the energy consumption for different periods based on the calendar and weather conditions in those times. The predictions for these other time periods can be compared against the behavior that is being observed so see if there is a deviation between the two. Times when the prediction exceeds the observed consumption indicate that the building is consuming less than expected by the model and times when the observed consumption exceeds the prediction indicate that the building is consuming more than was expected.

The results of the auto-benchmark can be utilized in several different ways. One possibility is to identify buildings in need of systems recommissioning where the performance is slowly declining over time. This would allow responsive recommissioning rather than following a set schedule or only addressing the issue once it has become noticeably bad. By recommissioning as needed two scenarios are avoided, allowing a building to go unaddressed for a long period of time as well as recommissioning a building on a schedule when maybe it did not require it yet.

A second utilization of the auto-benchmark is measuring the impact of known events or changes in a buildings structure, systems, or operations. A frequent question is, how much energy did this renovation save? Given that weather conditions and other external variables can strongly influence energy consumption, one cannot just compare the energy expenditures from two separate time periods. However, the auto-benchmark accounts for these variables and provides a stronger point of reference so that a building manager can compare the expected energy consumption against the actual. In this way the impact of building projects or events within the facility can be accurately measured. (Georgescu, 2014)

A final utilization of the auto-benchmark is in the realm of fault detection. While the energy data we receive is static and received in blocks of time, the black box model could be used to predict and compared the energy consumption of a building in real time. In this was the model can be used to quickly identify buildings that experience sudden changes in their energy consumption due to equipment failure. On the UPenn campus steam loop, leaks and failures are not uncommon. Small problems frequently go unnoticed for days or even weeks until the issue is visually observed by chance or a review of monthly energy consumption reveals an increase, which can lead to large amounts of wasted steam. By identifying and addressing these issues quickly, the University could both improve operations and reduce overall energy consumption.

WiFi as Occupancy Proxy

While the black box model has proven effective and reliable in the prediction of energy consumption, it achieves best results when considering time frames on the scale of a day or higher, while the single hour predictions tend to be less reliable. Although the weather and calendar attribute used are strong predictors of energy consumption, especially for heating and cooling, there are many other factors that are not considered which have a significant impact on energy consumption on an hourly basis, particularly electric. One could consider adding any of these to improve the accuracy of the black box model.

One attribute in particular that has been considered is that of occupancy. Occupancy is roughly approximated through the calendar and hour attributes, but especially in an academic environment where each semester and the times between can occupy a space very differently. A direct measure of occupancy would correlate strongly to the usage of lights and plug loads within a building and could also impact heating and cooling demand, depending on the density of occupation. Occupancy, however, it much more difficult to directly measure than the weather attributes or calendar, a single set of which may be applied to the entire campus. Instead, occupancy would need to be measured for each building using a system of occupancy sensors, which would add a layer of technical, financial, and operation complexity. (Kwok, 2011)

An alternative to the direct measurement of occupancy would be to track a proxy attribute for occupancy that might not provide an exact count for a building, but which could be used to gauge its relative level of occupation over time. One such piece of data is already being monitored and can easily be retained and collected, and that is the wireless connectivity within a building. Most individuals carry with them phones, computers, or other devices which will detect and automatically connect to familiar wireless networks. Some individuals may have two devices, some may have none that have access to the network, but with a sufficient population, the relative changes in WiFi connectivity should strongly correlate to the level of occupation. Due to this potential, the use of WiFi data as a proxy for occupancy levels is one of several options being explored for improving the performance of the black box model.

Using a Normative Building Model

Up until this point the meter data has primarily been used a sieve to find those buildings which are high energy consumers when compared to other similar buildings or when compared to their historical performance. At best, the analyses coming from the energy signatures and feature importance may provide some clues as to the primary influences on a building's energy consumption, but they cannot gauge the impact of potential remedies because they do not consider anything regarding physical structure and systems of the actual building. In order to analyze what interventions may be effective in improving the energy performance of a building a white box model is required.

In 2018 a normative building model was developed by Pengyuan Shen, then a doctoral candidate at the Center for Environmental Building + Design (Shen, Braham, and Yi 2018). Unlike the more complex white box models, such as Energy Plus, the normative building model is a low-order white box model, meaning that, rather than relying on direct physics based calculations of heat flow between equipment, zones, and the exterior, simplified heat flow equations are used. The overall effect is a model which requires fewer inputs, is quicker to construct, easier to calibrate, and simple to use. (Neto, 2008)

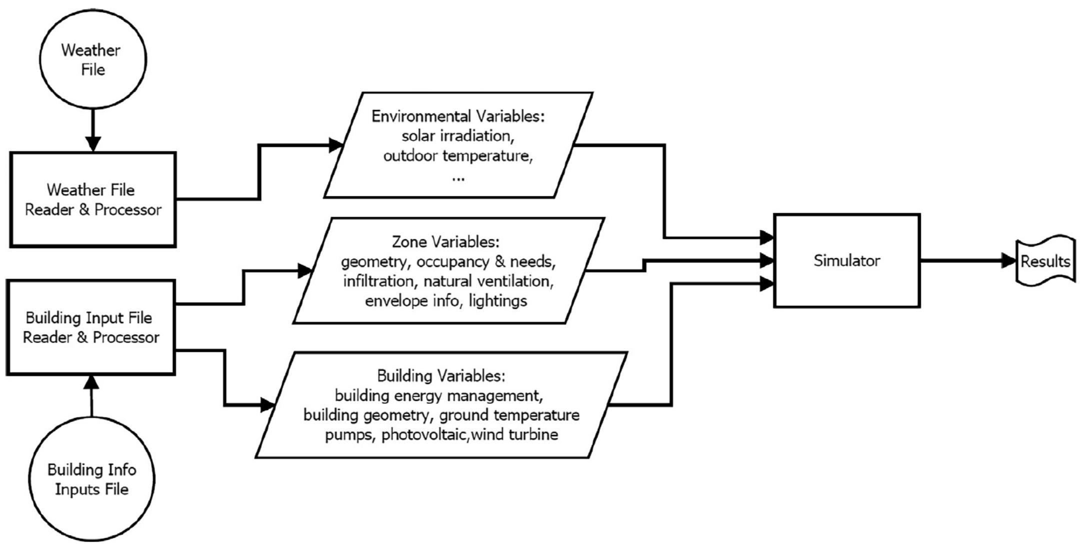


Figure 6- SimPyBuild Flowchart

In 2018 a pilot program to create normative models of six of the worst performing buildings on campus so that a suite of energy conservation measures could be evaluated for the energy reductions that could be expected and the net present value of those projects. All six buildings came from a list of the thirty buildings on campus which showed the greatest potential for energy reductions using the values calculated from the benchmark analysis. Rather than specifically choosing the six worst buildings from this list, a representative sample was selected to include buildings of different types and scales. The pilot program successfully showed the utility of the normative model in evaluating a wide range of energy conservation measures based on varying bottom lines, such as maximizing financial saving for investment, maximizing carbon reductions to achieve a goal, or combining financial and energy impacts to select those that provide the greatest improvement for the fewest dollars.

However, the process of constructing and calibrating the model proved to be slower and more difficult than anticipated. As a white-box model a large part of the construction of the model revolves around recreating the physical structure of the building. In the normative model the physical building is represented as a series of boxes representing zones which are defined by the area and orientation of their six sides. Connections between zones are identified by maintaining a list of the surfaces in each zone that are shared with another zone and the amount of area that they share. Any surface area which is not adjacent to another zone is assumed to be an exterior surface.

In the pilot program, the surface area and adjacency area calculations were being done manually and so, by necessity, the zonal breakdown of buildings was simple and relatively formulaic. Each floor would be assigned a core zone and a perimeter zone, each of which would be described as simple rectangular cubes of equal height. The zones of each floor would be stacked and the adjacencies of the core and perimeter zones with each other and with the floors above and below would be calculated. As a result, these early zonal models did not reflect the true zones that existed within these buildings and the method by which the adjacencies were calculated meant that any changes to the zonal structure would require a lengthy recalculation of all affected values across many zones. This decreased the overall accuracy of the models, lengthened the time necessary for calibration, and extended the overall time it took to finish a model from the desired day or two to weeks.

The pilot program showed that the normative model was capable of being used to provide analysis for the potential of specific interventions to reduce energy consumption in a building and so it was decided to expand the buildings that were being model to the full list of the 30 buildings on campus that were identified as having the greatest potential for reduction, an additional 24 models. To complete this work in a timely fashion, a new workflow had to be generated that could address the speed and ease of creating the geometric models representing the buildings' structures. To this end, the following procedure was developed to more quickly and accurately generate and evaluate the models.

As the primary weak point in the generation of the models was in determining their zonal structure and connections manually, the most significant improvement was the use of the Rhino 3d modelling and design software. Frequently used in architecture and engineering, Rhino is well suited to the creation of 3D geometric models of buildings and their zones. These models can be quickly constructed from scans of the floorplans which ensures accuracy of scale and in the calculations of the interconnectedness of zones. Further, as all the calculations regarding area and adjacency are done by the program, significantly more complex zonal structures can be created beyond the simple core and perimeter zones of the pilot program. The results of these calculations can be exported and saved to .csv files, which are easily read by other software.

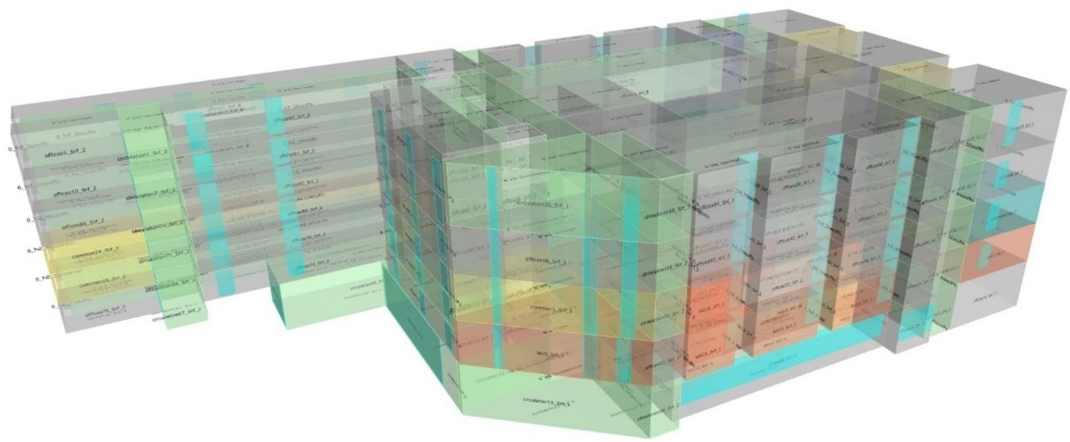


Figure 7- Example of Zonal Rhino 3D Geometric Model

The next step was to takes those geometric outputs from Rhino and to combine them with the remaining information needed for the creation of the model. This information can be broken down into two categories, single point information and schedules. Single point information includes building level data, such as the system types or the thermal properties of the envelope, as well as information that could vary from zone to zone and thus may have distinct values for each. Schedules represent those variables that change over time and are typically in the form of factors which are linked to single point peak values, such as peak plug load or occupancy. The schedules would then track the relative plug load or occupancy level compared to that peak for a given hour of the day or month of the year. Like the single point values, schedules can apply to the entire building or can be customized for each zone type.

All the single point information for the whole building as well as for the zone types is entered into an Excel template. The four .csv files and this excel file are then read by a Java program that then writes out a single text file containing all the information needed in the Cen format. The Cen file is then read by the normative model to estimate the hourly and monthly energy consumption of the building for a year.

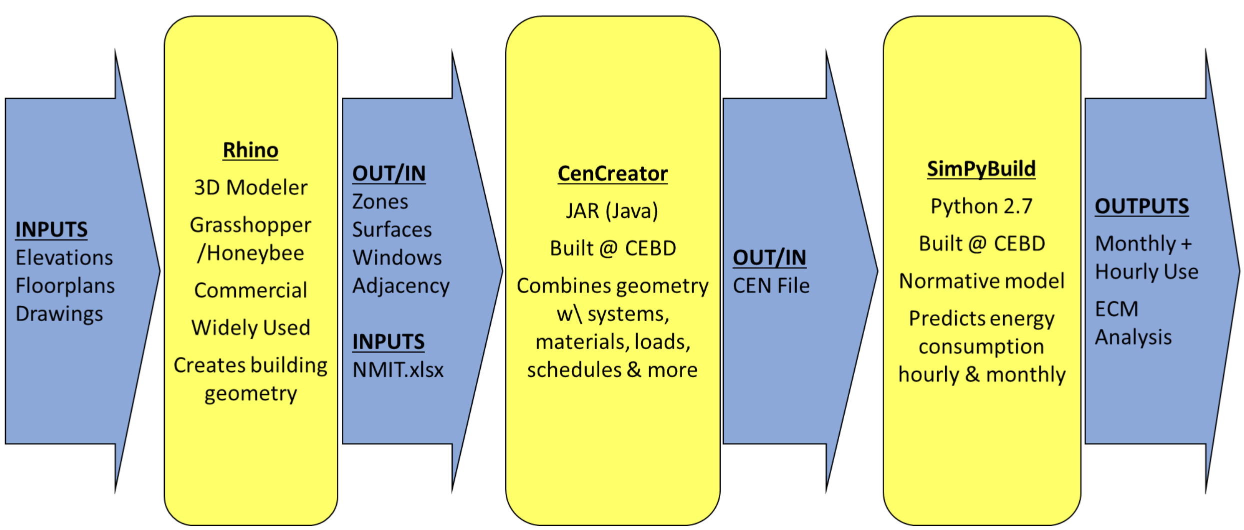


Figure 8- Flowchart of Data Flow for Normative Model

At this juncture, the model is ready for calibration. As the model was run using a weather file from 2018, the outputs of the model are compared against the measured energy consumption for that year. (Gutiérrez, 2021) This is accomplished by adjusting those values which were initially estimated or placeholders in the original construction of the model or by revisiting the values which may have changed over time. Assuming the building was properly zoned in the creation of the Rhino model, all the changes for calibration can be made within the normative model template. The calibration process is described in greater detail in the following section.

Once the model has been calibrated, it is possible to evaluate a variety of different energy conservation measures (ECMs) by altering the Cen file to reflect the changes that would be enacted. (Li, 2015) This process is like the method of calibration in that nearly all the changes are made within the normative model input template rather than to the geometric outputs from Rhino. Each ECM is linked to a specific value or set of values that inputs should be set to in order to model that scenario. An example would be an ECM evaluating the impact of improving the insulation in a wall. R-values of R-13, R-17, R-23, and R-30 can replace the original R-value for the baseline building (assuming they are an improvement) and the model rerun on the newly generated Cen file. The energy expended over a year according to the new model can be compared to the results of the old model to determine how each level of insulation improvement would decrease energy consumption.

The final piece in this analysis is to determine the cost and net present value of each of the ECMs that was modeled. For this project, we were able to consult with one of the primary consulting engineers who provides price estimates on renovation and construction work on campus. Using the historical averages and estimates provided by this consultant, cost per unit rates were developed for each ECM option. These cost per unit rates were then applied to the appropriate values from the building model to estimate the total initial and ongoing costs of undertaking that ECM in that building. Secondly, the value of the energy that is saved by the ECM is also calculated using the current per unit price of those energy sources. Due to volatility in the energy markets over the long term, future prices on energy were not projected. From these figures the net present value of each project is determined.

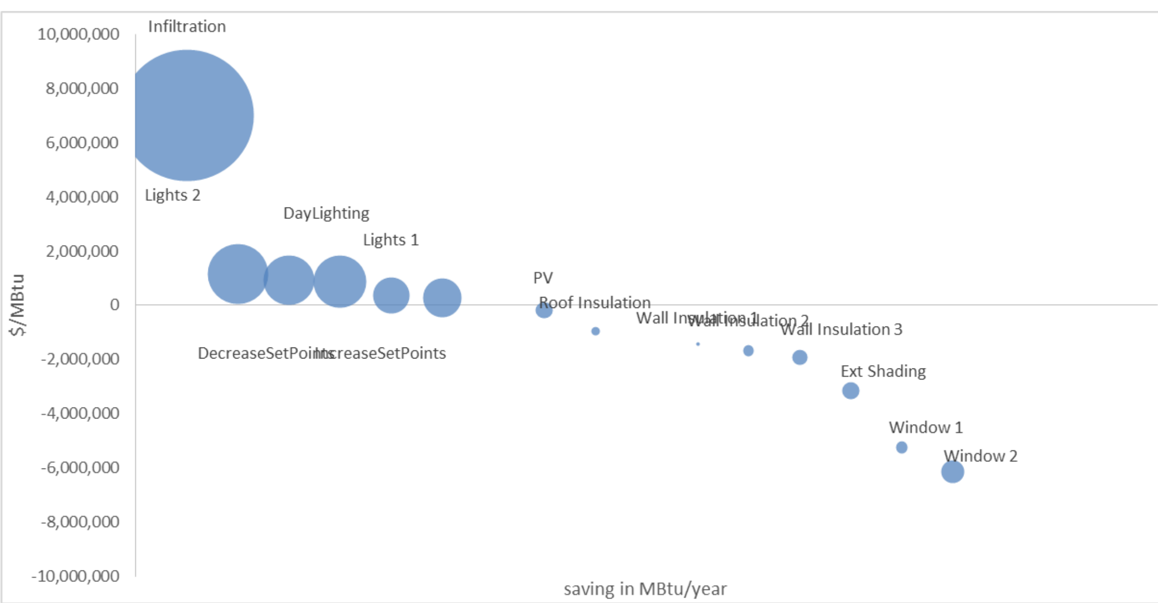


Figure 9- Cost Abatement Curve for $/MBtu per ECM

This information is represented as a type of cost abatement curve, wherein the results can be displayed according to the desired bottom line. Results could be ordered by absolute carbon reductions, NPV, initial cost, or more complex measures such as cost per ton of carbon eliminated. This allows a building manager or administrator to create a package of interventions for a building to meet any given goal.

Normative Model Pipeline:

1. Collect scanned / digitized floorplans for each floor
2. Load each floorplan into Rhino and assign height to create a 3D model of each floor
3. Identify zones on each floor based on areas that are contiguous and utilized for the same purpose
    a. Zones should be labeled by their use and an index number, i.e. offices_23
4. Add windows to the Rhino model to match the elevations seen in the drawings
5. Stack the 3D zonal models of each floor to create a zonal model of the entire building
6. Export the geometric model from Rhino using plugins Grasshopper and Honeybee
    a. List of zone names and area's (offices_1, 570; offices_2, 890; circulation_0, 340.…)
    b. List of surfaces, labeled by zone and orientation, and areas (offices_1_floor, 570; offices_1_ceiling, 570; offices_1_south, 680; offices_1_east, 475; offices_1_north….)
    c. List of windows, labeled by surface and type, and areas (offices_1_south_glz_0, 90….)
    d. List of adjacent surfaces and area shared (offices_1_ceiling, circulation_2_floor, 570; offices_1_south, offices_2_north, 320; ….)
7. Enter non-geometric properties into normative model input template
    a. Identifying information
    b. Systems information
    c. Set point, occupancy, lighting, and plug-loads, peak loads and schedules
    d. Infiltration and ventilation rates
    e. Thermal properties of opaque surface materials and glazing types
8. Use the CenCreator to combine four output files from Rhino with the input template
9. Run the model by providing the Cen file to SimPyBuild, a Python implementation of the normative model
10. Compare the results of the initial run against the measured consumption for the time period covered by the weather file provided to SimPyBuild, in this case 2018
11. Calibrate the model by adjusting values in the normative model input template and recreating Cen
    a. This can be done manually for simple adjustments or algorithmically to find a optimal values across many inputs.
12. Using the calibrated model, record the energy consumption as the baseline scenario
13. For each ECM to be evaluated, make the corresponding change to the normative model input template and create a new Cen file with those changes

14. Run the model using each new Cen file and compare the energy consumption against the baseline to determine the magnitude of the energy reduction from that intervention
15. Perform NPV analysis of each option, considering initial and continuing costs along with savings from energy reductions
16. Compare the energy / carbon reductions, initial costs, and NPV to identify a selection of ECMs to most closely match the desired outcome

Calibrating a white box model

The calibration of the white box normative model can either be done manually or algorithmically. (Chaudhary, 2016) In both processes a series of adjustments to the model inputs are considered and the successive outputs of the model after these changes are compared to actual consumption to find those setting which minimize the error observed. While this process could be done at random or according to some iteration through the options, typically only a few setting changes will make sense for calibration. (Martinez, 2020) This section will describe the best practices for calibration that were discovered in the creation of the 30 building models.

In most cases, it will be optimal to try to calibrate electrical consumption on its own before attempting to calibrate the models steam and chilled water results. This is because changes to electrical consumption can significantly impact the demand for heating and cooling due to waste heat from electric use itself. If electric consumption is low or high consistently throughout the year, then the peak plug-loads and lighting-loads should be reevaluated. If the usage is high or low at certain times of the day or months of the year, then the usage and occupancy schedules can be adjusted to better reflect the patterns observed in the consumption.

Once electric has been calibrated and the occupancy and usage schedules are determined, the amount of heat generated by the population and these loads becomes fixed and the steam and chilled water consumption can be calibrated. Typical inputs to adjust to calibrate these outputs would be to confirm the set points, consider variations in infiltration and ventilation rates, consider the possibility that the thermal properties of the envelope have degraded from the as-built condition, or to adjust the thermal mass of the building which will impact its responsiveness to change.

The following list details the inputs which should or should not be adjusted during the calibration process.

<u>Do Change / Audit</u>

1. Lighting Loads
    1. Intensity
    2. Monthly / Daily Schedules
    3. Plug Loads
    4. Intensity
    5. Monthly / Daily Schedules
    6. ACH / Infiltration
    7. Monthly Rates

        8. Occupancy
        9. Rates
        10. Monthly / Daily Schedules
        11. Schedules and Loads by Zone
    2. Internal Thermal Mass

Don't Change

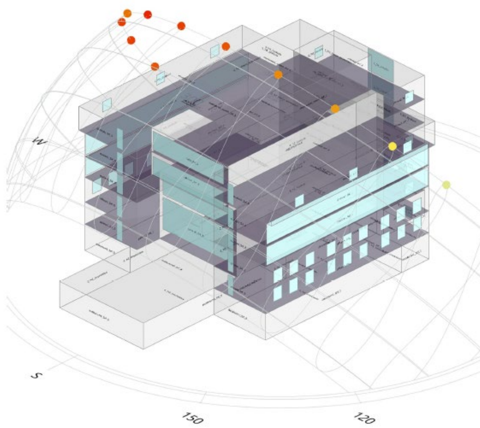
1. Don't change anything that is well known
    1. Building dimensions
    2. Building system type
    3. Material Thermal Properties
2. Building Zones
    1. Rezoning model is time intensive and usually not necessary if well zoned to start
3. Don't Force Anything
    1. No variable should be altered from default too much without confirmation

Building Energy Management Plans

The Building Energy Management Plans were designed to integrate all the energy related data available in databases at the University with the outputs of the black box data driven model and the low order white box normative model. As they rely on outputs from the normative model, unlike the Annual Energy Reports, these were only created for the buildings which had normative models, which were the 30 worst performing buildings identified by the meter data and the black box model. These reports contain all of the information in the Annual Energy Reports but add significant details on the buildings use by area, administration, construction details such as wall sections and systems installed, a history of the renovations and remediations, and finally the cost abatement curves generated by the normative models.

A final addition was a review of all this information which discussed the building's energy performance in context of its benchmarks, energy signature, construction and the results of the normative model to suggest potential issues with the building along with an evaluation of a suite of corrective measures that could be applied. These documents are intended to provide long term planning guidance to administrators and building managers who make capital decisions in the facilities and operations sphere. They could be used as a result of a poor review in an annual energy report to identify means of improving performance, or they could be consulted whenever renovations or construction are planned for a facility in order to align energy saving measures with preexisting disruptions to building operations.

An example of an Energy Management Plan is provided to show the additional analyses and information included along with how all of the information included is used at the end to provide an assessment of the structure along with an evaluation of the likely issues causing energy performance problems.



Meyerson

Energy Management Plan

UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

CENTER FOR ENVIRONMENTAL BUILDING AND DESIGN

Meyerson

Energy Management Plan

UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Table of Contents

This Energy Management Plan collects, summarizes, and analyzes available data about campus buildings to provide strategic information for reducing energy consumption, costs, and carbon emissions. In synthesizing benchmarks, consumption data and patterns, and forensic analysis this document serves as a template for potential remediation efforts of various scales, uniquely catered to specific buildings. This plan is designed to supplement the annual energy report compiled and provided by the Center for Environmental Building + Design to provide a wide variety of readers and stakeholders with more readily actionable information for near and long term decision making at the building level.

Meyerson

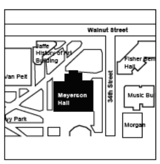
Energy Management Plan

UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Building Data & Overview

ID	340
Address	210 South 34th Street
Department	School of Design
BPD Type*	Education: University
Area Manager	Agnes Donaldson
Building Manager	Karl M. Wellman, Jr.
Year Built	1967
Area (GSF)	98,027
Number of Floors	7
Height (ft)	49.3
Window-Wall Ratio	16.6%

**Building Performance Database*



Architectural Description

A blend of Philadelphia School and early modern heroic forms, Meyerson Hall recalls the dynamic era of the Perkins faculty in a concrete framed building crowned by industrial roof skylights recalling Le Corbusiers house for Ozenfant, with Mediterranean sunscreens on the side. The building was designed by Martin, Stewart, Nobel and Class.

Program Distribution

Current Use & Administration

Meyerson Hall houses the University of Pennsylvania School of Design, comprising of Architecture, Landscape Architecture, Historic Preservation, and City and Regional Planning. The basement contains the newly opened PennDesign Cafe, eight lecture halls, custodia staff support spaces, and the materials library. The first floor contains departmental offices, Second Floor contains offices for Architecture as well as studio space. Third Floor contains two computer labs, a plotter room, student lounge and additional studio space. The fourth floor contains studios and a fabrication lab. Currently the studio spaces are being used 24/7 and fabrication lab open from 8:00am to 7:00pm.

Mechanical Systems Overview

Meyerson is served by a number of air handling units supplying VAV boxes throughout the building. The perimeter is heated with a fin-tube radiation system. The building obtains heat from the campus steam loop and cooling frm the campus chilled water loop.

Page 1 of 9

Meyerson

Energy Management Plan

UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Construction & Systems Overview

Primary Envelope Description

The building's envelope is of concrete masonry unit with brick cavity wall on the exterior and plastered on the inside.

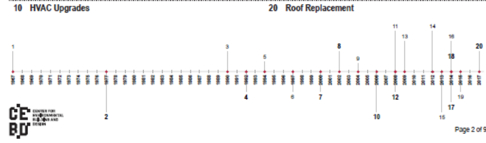
Secondary Envelope & Glazing Description

The secondary envelope consists of 72-in thick reinforced concrete and glazing consists of single pane glass with Bronze plated aluminum frame.

Component	U-Value (Btu/hr-sqft-°F)
Wall, As-Built	0.227
Wall, Code - IECC 2018	0.104
Window	1.1
Window, IECC 2018	0.38

Modification & Recomissioning Information

1	Built	11	Classroom Renovation
2	**Chilled Water Centralization**	12	**Fire Alarm System Upgrades**
3	Infrastructure Renovation	13	Sprinkler System
4	**Lighting Renovations**	14	School of Design Café
5	Infrastructure Renovations	15	Studio Renovation
6	Classroom Improvements	16	Lecture Hall Renovation
7	**Networking Phase IV**	17	**Facade Inspection**
8	**Facade Repairs**	18	**HVAC Upgrades**
9	Infrastructure Renovation	19	Infrastructure Renovations
10	**HVAC Upgrades**	20	**Roof Replacement**



Page 2 of 9

Meyerson
Energy Management Plan
UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

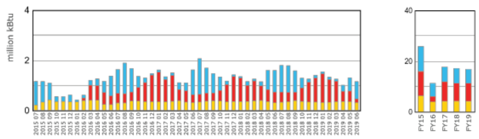
Energy & Carbon Metrics
[2018-19]

Building Energy Consumed	Source Units	kBtu	kBtu / sqft
Electricity	1,190,425 kWh	4,061,898	43.2
Steam	5,973 mlb	7,230,316	77.0
Chilled Water	5,341,616 kBtu	5,341,616	56.9
Total		16,633,839	177.0

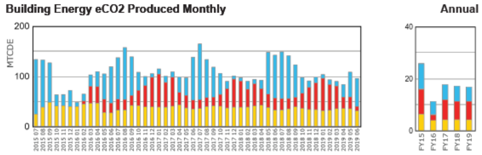
Building Energy Cost	Consumption Cost	Overhead Cost	Total Cost	$ / sqft
Electricity	$92,646	$20,305	$112,951	$1.20
Steam	$91,510	$10,314	$101,824	$1.08
Chilled Water	$93,297	$32,529	$125,826	$1.34
Total	$277,453	$63,148	$340,604	$3.63

Building Energy eCO2	kg eCO2	kg eCO2 / sqft	$ / kg eCO2	kg eCO2 / kBtu
Electricity	409,439	4.358	$0.276	0.1008
Steam	404,332	4.304	$0.252	0.0559
Chilled Water	456,908	4.863	$0.275	0.0855
Total	1,270,679	13.525	$0.268	0.0764

Building Energy Consumed Monthly **Annual**



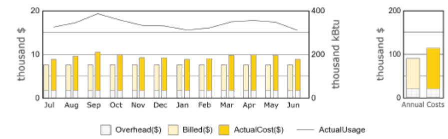
Building Energy eCO2 Produced Monthly **Annual**



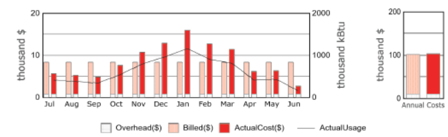
Meyerson
Energy Management Plan
UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Consumption
& Billing Metrics

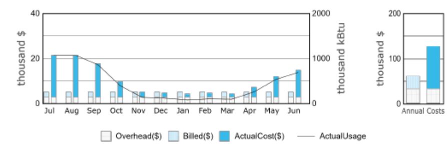
Building Energy Consumed vs Billed - Electricity



Building Energy Consumed vs Billed - Steam



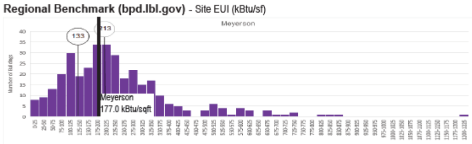
Building Energy Consumed vs Billed - Chilled Water



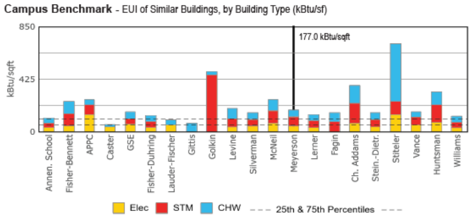
Meyerson
Energy Management Plan
UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Energy Benchmarking
Campus & Region

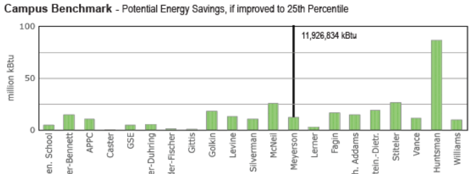
Regional Benchmark (bpd.lbl.gov) - Site EUI (kBtu/sf)



Campus Benchmark - EUI of Similar Buildings, by Building Type (kBtu/sf)



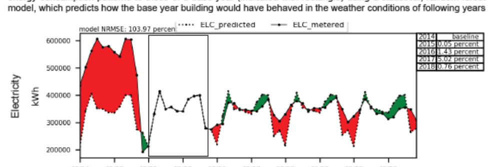
Campus Benchmark - Potential Energy Savings, if improved to 25th Percentile

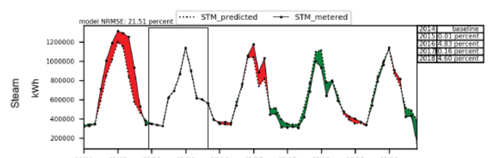


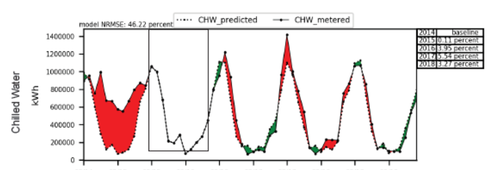
Meyerson
Energy Management Plan
UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Energy Benchmarking
Historical Comparison

Auto-benchmark charts are used to determine whether the building is consuming more or less energy than it should, which indicates that the building was either operated differently or altered in some way. The prediction compares energy consumption patterns to those of a base year (indicated with rectangle) using a statistical, random forest model, which predicts how the base year building would have behaved in the weather conditions of following years







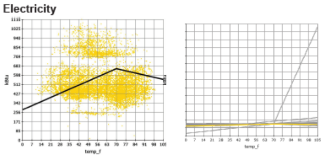
Meyerson

Energy Management Plan

Building Energy Signature

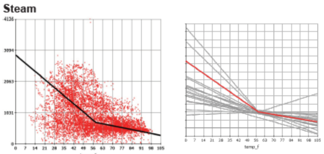
Building energy signature charts are used to reveal the characteristic patterns of energy consumption in response to weather conditions. In the left hand charts, the hourly energy consumption is plotted against the outdoor temperature and an iterative technique is used to fit "change point" lines to the data, one for heating and one for cooling. The slopes of the lines indicate the intensity of the heating or cooling—the steeper the line the greater the response—and the intersection of the two lines is called the "change or balance point," and indicates the point of transition from heating to cooling. In the right hand charts, the buildings signature is plotted with those of similar buildings. In most cases, building electric use shows less correlation with temperature.

Electricity



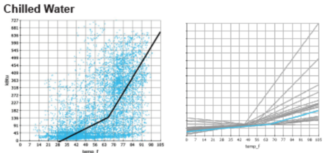
Electric usage at Meyerson is not particularly influenced by temperature, and shows two distinct modes, likely reflecting occupancy schedules

Steam



Steam usage at Meyerson varies with temperature, and exhibits a characteristic change point to cooling in the 50s F.

Chilled Water



Chilled water usage at Meyerson correlates strongly with temperature, with a general change point occuring in the 60s F.

Building | Campus Comparison

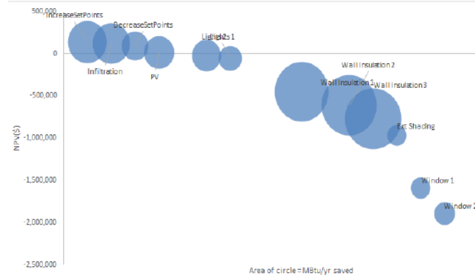
Meyerson

Energy Management Plan

UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Analysis & Recommended Improvements

A normative simulation of building energy consumption (ISO 13790) was used to evaluate the cost and effectiveness of various energy reduction measures. These are evaluated on a life-cycle basis using Penn utility pricing and industry standards for the cost of improvements. These are plotted in bubble chart ranked from left to right according to the cost of each improvement (NPV in $) with the size of the bubble describing the total energy saved. Positive costs indicate the savings.



Potential Energy Conservation Measures (NPV over 20 years at 3% discount rate)

Recommendation	Description	NPV ($/mBtu)	Initial Cost	Savings (mBtu/yr)	ROI	CO2/yr Reduced (MTCDE)
Increase Set Points	Increase cool SP 1 degree all seasons but winter	117	$0	1,163	inf	84,503
Infiltration	Reduce infiltration rate to 0.5 ACH	115	$51,152	1,058	2.37	54,352
Decrease Set Points	Decrease heat SP 1 degree all seasons but summer	168	$0	554	inf	33,392
PV	Set PV Surface Area to 1/2 Roof Area	19	$110,514	702	0.12	70,776
Lights 2	Set W/m2 to 5 for lighting	-33	$135,963	666	-0.16	62,940
Lights 1	Set W/m2 to 11 for lighting	-137	$118,976	409	-0.47	38,510
DayLighting	Set DayLightFactor to 0.5 for appropriate zones	362	$119,986	-753	-2.27	-69,601
Wall Insulation 1	Set External Wall uvalue to 0.4367	-199	$889,213	2,288	-0.51	132,139
Wall Insulation 2	Set External Wall uvalue to 0.334	-259	$1,067,055	2,371	-0.57	136,126
Wall Insulation 3	Set External Wall uvalue to 0.2468	-319	$1,244,898	2,432	-0.62	138,735

Meyerson

Energy Management Plan

UNIVERSITY OF PENNSYLVANIA FACILITIES & REAL ESTATE SERVICES

Summary & Recommendations

Performance Overview

Meyerson is a classically mixed use buildling, which requires air conditioning in the core all year roung, while the perimeter rooms are largely climate driven. The benchmark data indicates that performance is just below the mean for this type of building, which experienced a major HVAC upgrade in 2015, so although it can be improved, only a few conservation measures make economic sense.

Energy Usage Benchmarks

Compared to similar educational buildings nationally, Meyerson's 177 kBTU/sf annually is a bit below the mean, which is 203 kBTU/sf. It is in the lower third of its class on the campus. Nevertheless if it were improved to the top 25th percentile it would save about 11 million kBTU/yr.
Compared to its performance in 2014-15 and adjusting for weather the building used less steam and chilled water, and much less electricity, which reflects the HVAC improvements.

Detailed Performance Observations

Electric usage at Meyerson is not particularly influenced by temperature, and shows two distinct modes, likely reflecting occupancy schedules

Steam usage at Meyerson varies with temperature, and exhibits a characteristic change point to cooling in the 50s F.

Chilled water usage at Meyerson correlates strongly with temperature, with a general change point occuring in the 60s F.

The walls are uninsulated. The current U-Value is 0.227, which is considerably worse than the current code standard of 0.104 for mass walls. The windows are single pane class which are also less efficient than current standards

Recommendations

Because of the HVAC upgrade, most of the improvements show low ROI, though considerable energy could be saved by insulating the walls.

References

ACUPCC. American Colleges and Universities President's Climate Commitment. *Second Nature*. https://secondnature.org/signatory-handbook/the-commitments/

Amasyali, K. and El-Gohary, N. 2018. 'A review of data-driven building energy consumption prediction studies'. *Renewable and Sustainable Energy Reviews.* 81(1). 1192-1205.

Bourdeau, M., Zhai, Z., Nefzaoui, E., Guo, X. and Chatellier, P. 2019. 'Modeling and forecasting building energy consumption: A review of data-driven techniques'. *Sustainable Cities and Society*. 48.

Braham, W., Yi, Y., Waegel, A., et al. 2016. Building Energy Reporting and Performance Analysis. https://www.design.upenn.edu/sites/default/files/files/FRES_16_Report_Final_PT1.pdf

CAP. University of Pennsylvania Climate Action Plan. 2009. https://www.sustainability.upenn.edu/sites/default/files/downloads/PENN-2009-Climate_Action_Plan.pdf

CAP2. University of Pennsylvania Climate Actin Plan 2.0. 2014. https://www.sustainability.upenn.edu/sites/default/files/downloads/New%20Penn%20Climate%20Action%20Plan%202.0_1.pdf

Chalal, M.L., Benachir, M., White, M., and Shrahily, R. 2016. 'Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: a review'. *Renewable and Sustainable Energy Reviews*. 64. 761-776.

Chaudhary, G., New, J., Sanyal, J., Im, P., O'Neill, Z. and Garg, V. 2016. 'Evaluation of "Autotune" calibration against manual calibration of building energy models'. *Applied Energy*. 182. 115-134.

Georgescu, M., Eccles, E., Manjunath, V., Swindle, E., and Mezic, I. 2014. 'Machine learning methods for site-level building energy forecasting and data rectification'. *Building Simulation Optimization*. IBPSA, London.

Gutiérrez, V., Ramos, G. and Fernández, C. 'Impact of Actual Weather Datasets for Calibrating White-Box Building Energy Models Base on Monitored Data'. *Energies*. 2021; 14(4). 1187.

Kissock, John Kelly, Jeff S. Haberl, and David E. Claridge. 2002. Development of a Toolkit for Calculating Linear, Change-point Linear and Multiple-Linear Inverse Building Energy Analysis Models. In *ASHRAE Research Project 1050-RP*: American Society of Heating Refrigeration and Air Conditioning Engineers.

Kwok, S. and Lee, E. 2011. 'A study of the importance of occupancy to building cooling load in prediction by intelligent approach'. *Energy Conversion and Management* 52. 2555-2564.

Ladybug Tools. Roudsari, M. https://www.design.upenn.edu/cebd/work/ladybug-tools

Li, N., Yang, Z., Becerik-Gerber, B., Tang, C., and Chen, N. 2015. 'Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?'. *Applied Energy*, 159, 196-205.

Li, Z., Han, Y., and Xu, P. 2014. 'Methods for benchmarking building energy consumption against its past or intended performance: an overview'. *Applied Energy*. 124. 325-334.

Martínez, S., Eguía, P. Granada, E., Moazami, A. and Hamdy, M. 2020. 'A performance comparison of multi-objective optimization-based approaches for calibrating white-box building energy models'. *Energy and Buildings.* 216.

Mathew, P.A., Dunn, L.N., Sohn, M.D., Mercado, A., Custudio, C., and Walter, T. 2015. 'Big-data for building energy performance: lessons from assembling a very large national database of building energy use'. *Applied Energy*. 140. 85-93.

Neto, A.H. and Fiorelli, F.A.S. 2008. 'Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption'. *Energy and Buildings*. 40. 2169-2176.

Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. and Pendyala, R. 2017. 'Machine learning approaches for estimating commercial building energy consumption'. *Applied Energy*. 208. 889-904.

Shen, P., Braham, W., Yi, Y. and Eaton, E. 2019. 'Rapid multi-objective optimization with multi-year future weather condition and decision-making support for building retrofit'. Energy. 172. 892-912.

Shen, Pengyuan, William Braham, and Yunkyu Yi. 2018. "Development of a lightweight building simulation tool using simplified zone thermal coupling for fast parametric study." *Applied Energy* 223:188-214. doi: https://doi.org/10.1016/j.apenergy.2018.04.039.

Wang, Z. and Srinivasan, R.S. 2017. 'A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models'. *Renewable and Sustainable Energy Reviews*. 75. 796-808.

Zhou, K. and Yang, S. 2016. 'Understanding household energy consumption behavior: the contribution of energy big data analytics'. *Renewable and Sustainable Energy Reviews.* 56. 810-819.

Creating an RC thermal model

Max Hakkarainen, Center for Environmental Building and Design

Introduction

Model Generation:

Linearization:

Stochastic State Space Description:

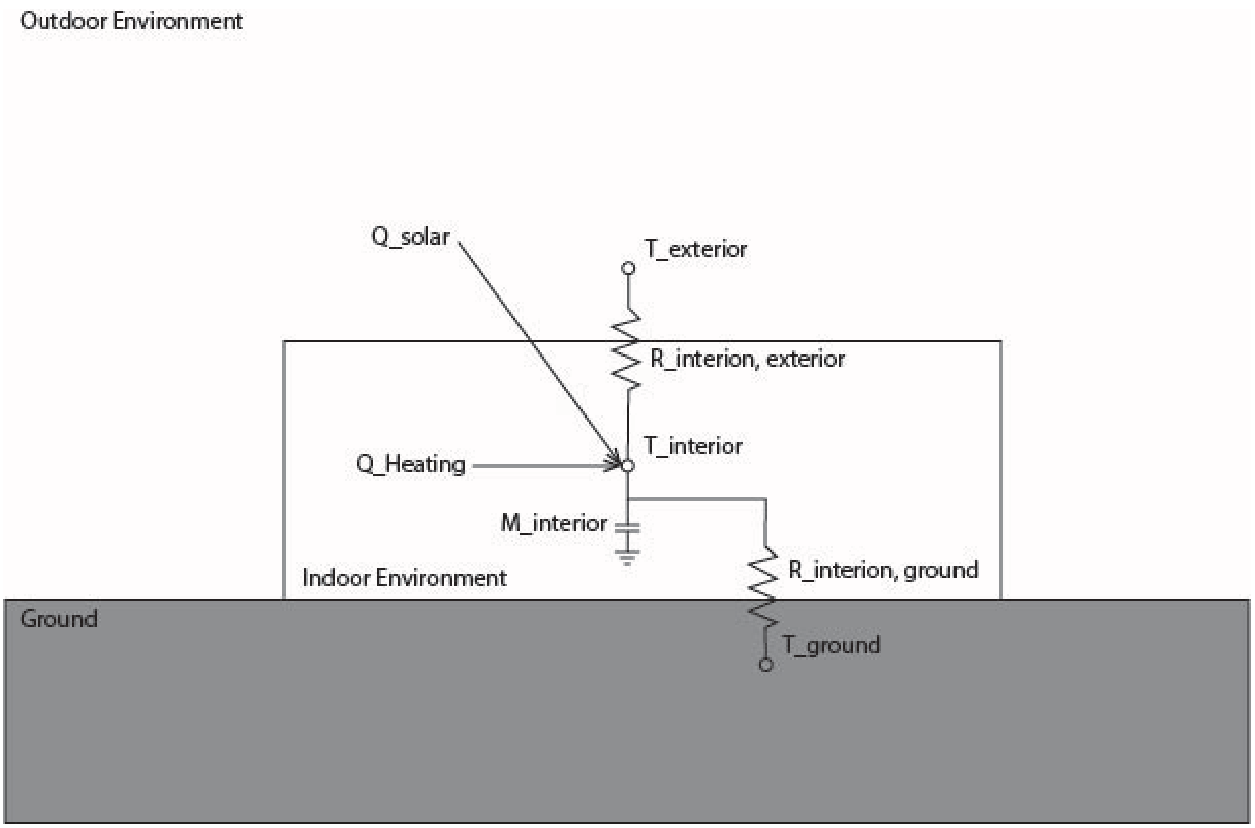
Minimum Error Optimization:

Introduction

This document will explain how to generate low order RC thermal models for describing buildings. Occupants, air infiltration, and ventilation will not be covered, here, as those are more complex from a thermal and mathematical perspective. This can be used to describe single or multizone buildings.

Model Generation:

The first step is to come up with a system description. This will depend both on the system being described and the measurements that are being taken within the system. The simplest system to describe would be a single zone building:



Here, the interior temperature (T_interior) is the dependent variable. The active heating or cooling (Q_Heating), solar heat gain (Q_solar), the exterior temperature (T_exterior), and the ground temperature (T_ground) are all taken to be independent variables driving the system. A couple of comments, here. It is common to model the ground temperature as a constant value corresponding to the steady state ground temperature about 10 m below the ground. Additionally, the solar heat gain is often modeled as an area (A_s) multiplied by the solar radiation flux (φ_s). This area is sometimes treated as the area of the windows on the building and sometimes as an effective area variable corresponding to the absorptivity of the opaque building surfaces. The thermal resistances correspond to effective envelope (R_interior, exterior) and floor (R_interior, ground) thermal resistances. Finally, each dependent variable, in this case, just the interior temperature, has an associated thermal mass (M_interior). This system can be used to generate a system equation (directly from a heat balance on this system) of the form:

$$M_{interior} * \frac{dT_{interior}}{dt} = \frac{A_{Interior,Exterior}}{R_{Interior,Exterior}} * (T_{Exterior} - T_{Interior}) + \frac{A_{Interiro,Ground}}{R_{Interior,Ground}} * (T_{Ground} - T_{Interior}) + A_S * \varphi_s + Q_{Heating} = UA_1 * (T_{Exterior} - T_{Interior}) + UA_2 * (T_{Ground} - T_{Interior}) + A_S * \varphi_s + Q_{Heating}$$

This can be algebraically rearranged and rewritten into a matrix equation:

$$\underline{\underline{M}} * \underline{\dot{T_i}} = -\underline{\underline{U_1}} * \underline{T_i} + \underline{\underline{U_2}} * \underline{T_o} + \underline{\underline{A}} * \underline{Q}$$

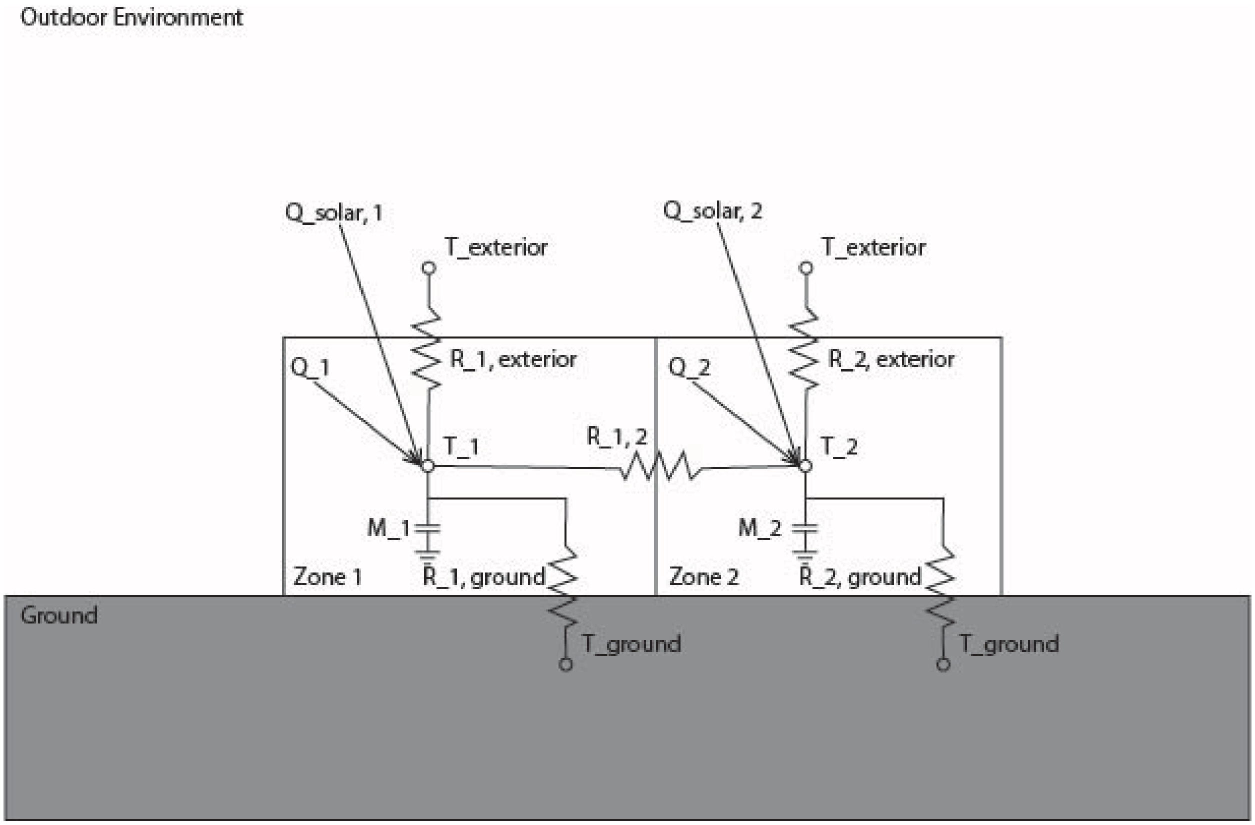
$$\underline{\underline{M}} = [M_{interior}] , \underline{T_i} = [T_{interior}], \underline{\underline{U_1}} = [(UA_1 + UA_2)], \underline{\underline{U_2}} = [UA_1 \quad UA_2],$$

$$\underline{T_o} = \begin{bmatrix} T_{Exterior} \\ T_{Ground} \end{bmatrix}, \underline{\underline{A}} = [1 \quad A_s] , \underline{Q} = \begin{bmatrix} Q_{Heating} \\ \varphi_s \end{bmatrix}$$

Here, $\underline{\underline{M}}$ is the mass matrix, $\underline{T_i}$ is the state vector, $\underline{T_o}$ is the environmental temperature vector, $\underline{\underline{U_1}}$ is the state coefficient matrix, $\underline{\underline{U_2}}$ is the environmental temperature coefficient matrix, $\underline{\underline{A}}$ is the heating coefficient matrix, and $\underline{Q}$ is the heating matrix. This model is parameterized by the set of thermal parameters, $\underline{\theta}$:

$$\underline{\theta} = \begin{bmatrix} UA_1 \\ UA_2 \\ M_{interior} \\ A_s \end{bmatrix}$$

In the multizone case, the system can be described by:



Here, there are two dependent variables, T_1 and T_2, corresponding to the two zones. This will yield two system equations of the form:

$$M_1 * \frac{dT_1}{dt} = \frac{A_{1,Exterior}}{R_{1,Exterior}} * (T_{Exterior} - T_1) + \frac{A_{1,Ground}}{R_{1,Ground}} * (T_{Ground} - T_1) + \frac{A_{1,2}}{R_{1,2}} * (T_2 - T_1) + A_{S,1} * \varphi_s + Q_{Heating,1} = UA_{1,E} * (T_{Exterior} - T_1) + UA_{1,G} * (T_{Ground} - T_1) + UA_{1,2} * (T_2 - T_1) + A_{S,1} * \varphi_s + Q_{Heating,1}$$

$$M_2 * \frac{dT_2}{dt} = \frac{A_{1,Exterior}}{R_{1,Exterior}} * (T_{Exterior} - T_2) + \frac{A_{1,Ground}}{R_{1,Ground}} * (T_{Ground} - T_2) + \frac{A_{1,2}}{R_{1,2}} * (T_1 - T_2) + A_{S,2} * \varphi_s + Q_{Heating,2} = UA_{2,E} * (T_{Exterior} - T_2) + UA_{2,G} * (T_{Ground} - T_2) + UA_{1,2} * (T_2 - T_2) + A_{S,2} * \varphi_s + Q_{Heating,2}$$

This can be algebraically rearranged and rewritten into a matrix equation:

$$\underline{\underline{M}} * \underline{\dot{T}_i} = -\underline{\underline{U_1}} * \underline{T_i} + \underline{\underline{U_2}} * \underline{T_o} + \underline{\underline{A}} * \underline{Q}$$

$$\underline{\underline{M}} = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}, \underline{T_i} = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix},$$

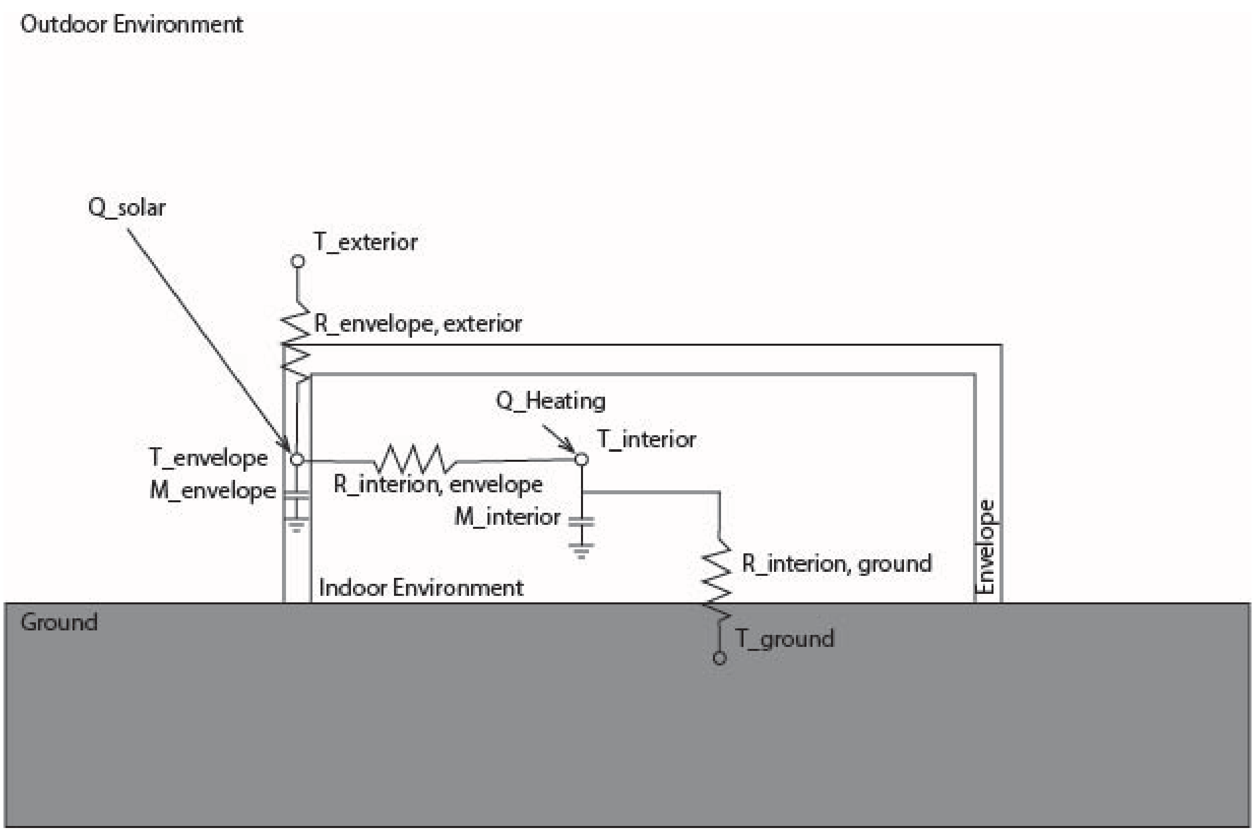
$$\underline{\underline{U_1}} = \begin{bmatrix} (UA_{1,E} + UA_{1,G} + UA_{1,2}) & -UA_{1,2} \\ -UA_{1,2} & (UA_{2,E} + UA_{2,G} + UA_{1,2}) \end{bmatrix},$$

$$\underline{\underline{U_2}} = \begin{bmatrix} UA_{1,E} & UA_{1,G} \\ UA_{1,E} & UA_{2,G} \end{bmatrix}, \underline{T_o} = \begin{bmatrix} T_{Exterior} \\ T_{Ground} \end{bmatrix}, \underline{\underline{A}} = \begin{bmatrix} 1 & 0 & A_{S,1} \\ 0 & 1 & A_{S,2} \end{bmatrix}, \underline{Q} = \begin{bmatrix} Q_{Heating,1} \\ Q_{Heating,2} \\ \varphi_S \end{bmatrix}$$

This model is parameterized by the set of thermal parameters, $\underline{\theta}$:

$$\underline{\theta} = \begin{bmatrix} UA_{1,E} \\ UA_{1,G} \\ UA_{1,2} \\ UA_{2,E} \\ UA_{2,G} \\ M_1 \\ M_2 \\ A_{S,1} \\ A_{S,2} \end{bmatrix}$$

This can be extended into indefinite additional zones. Additionally, as was done in the Mongolia case, the envelope can be treated as a zone, yielding the following system description:



Here, there are two dependent variables, T_interior and T_envelope, corresponding to the two zones. This will yield two system equations of the form:

$$M_{interior} * \frac{dT_{interior}}{dt} = \frac{A_{interior,envelope}}{R_{interior,envelope}} * \left(T_{envelope} - T_{interior}\right) + \frac{A_{interior,Ground}}{R_{interiro,Ground}} * (T_{Ground} - T_{interior}) + Q_{Heating} = UA_1 * \left(T_{envelope} - T_{interior}\right) + UA_2 * (T_{Ground} - T_{interior}) + Q_{Heating}$$

$$M_{envelope} * \frac{dT_{envelope}}{dt} = \frac{A_{interior,envelope}}{R_{interior,envelope}} * \left(T_{interior} - T_{envelope}\right) + \frac{A_{envelope,Exterior}}{R_{envelope,Exterior}} * \left(T_{Exterior} - T_{envelope}\right) + A_S * \varphi_s = UA_1 * \left(T_{interior} - T_{envelope}\right) + UA_3 * \left(T_{Exterior} - T_{envelope}\right) + A_S * \varphi_s$$

This can be algebraically rearranged and rewritten into a matrix equation:

$$\underline{\underline{M}} * \underline{\dot{T}_i} = -\underline{\underline{U_1}} * \underline{T_i} + \underline{\underline{U_2}} * \underline{T_o} + \underline{\underline{A}} * \underline{Q}$$

$$\underline{\underline{M}} = \begin{bmatrix} M_{interior} & 0 \\ 0 & M_{envelope} \end{bmatrix}, \underline{T_i} = \begin{bmatrix} T_{interior} \\ T_{interior} \end{bmatrix}, \underline{\underline{U_1}} = \begin{bmatrix} (UA_1 + UA_2) & -UA_1 \\ -UA_1 & (UA_1 + UA_3) \end{bmatrix},$$

$$\underline{\underline{U_2}} = \begin{bmatrix} 0 & UA_2 \\ UA_3 & 0 \end{bmatrix}, \underline{T_o} = \begin{bmatrix} T_{Exterior} \\ T_{Ground} \end{bmatrix}, \underline{\underline{A}} = \begin{bmatrix} 1 & 0 \\ 0 & A_S \end{bmatrix}, \underline{Q} = \begin{bmatrix} Q_{Heating} \\ \varphi_s \end{bmatrix}$$

This model is parameterized by the set of thermal parameters, $\underline{\theta}$:

$$\underline{\theta} = \begin{bmatrix} UA_1 \\ UA_2 \\ UA_3 \\ M_{interior} \\ M_{interior} \\ A_S \end{bmatrix}$$

No matter the construction of the model, it will produce an equation of the form:

$$\underline{M^T} * \underline{\dot{T}_i} = -\underline{\underline{U_1}} * \underline{T_i} + \underline{\underline{U_2}} * \underline{T_o} + \underline{\underline{A}} * \underline{Q}$$

Parameterized by $\underline{\theta}$.

Linearization:

The above equation is a differential equation, which needs to be linearized into a difference equation to be numerically computed. Here a midpoint Euler derivative approximation is used:

$$\underline{\dot{T}_i}\left(t+\frac{\Delta t}{2}\right)=\begin{bmatrix}\dot{T_{i,1}}\left(t+\frac{\Delta t}{2}\right)\\ \dot{T_{i,2}}\left(t+\frac{\Delta t}{2}\right)\\ \vdots\\ \dot{T_{i,n}}\left(t+\frac{\Delta t}{2}\right)\end{bmatrix}=\begin{bmatrix}\frac{T_{i,1}(t+\Delta t)-T_{i,1}(t)}{\Delta t}\\ \frac{T_{i,2}(t+\Delta t)-T_{i,2}(t)}{\Delta t}\\ \vdots\\ \frac{T_{i,n}(t+\Delta t)-T_{i,n}(t)}{\Delta t}\end{bmatrix}=\frac{1}{\Delta t}*\begin{bmatrix}T_{i,1}(t+\Delta t)\\ T_{i,2}(t+\Delta t)\\ \vdots\\ T_{i,n}(t+\Delta t)\end{bmatrix}-\frac{1}{\Delta t}*\begin{bmatrix}T_{i,1}(t)\\ T_{i,2}(t)\\ \vdots\\ T_{i,n}(t)\end{bmatrix}$$
$$=\frac{1}{\Delta t}*\underline{T_i}(t+\Delta t)-\frac{1}{\Delta t}*\underline{T_i}(t)$$

$$\underline{T_i}\left(t+\frac{\Delta t}{2}\right)=\begin{bmatrix}T_{i,1}\left(t+\frac{\Delta t}{2}\right)\\ T_{i,2}\left(t+\frac{\Delta t}{2}\right)\\ \vdots\\ T_{i,n}\left(t+\frac{\Delta t}{2}\right)\end{bmatrix}=\begin{bmatrix}\frac{T_{i,1}(t+\Delta t)+T_{i,1}(t)}{2}\\ \frac{T_{i,2}(t+\Delta t)+T_{i,2}(t)}{2}\\ \vdots\\ \frac{T_{i,n}(t+\Delta t)+T_{i,n}(t)}{2}\end{bmatrix}=\frac{1}{2}*\begin{bmatrix}T_{i,1}(t+\Delta t)\\ T_{i,2}(t+\Delta t)\\ \vdots\\ T_{i,n}(t+\Delta t)\end{bmatrix}+\frac{1}{2}*\begin{bmatrix}T_{i,1}(t)\\ T_{i,2}(t)\\ \vdots\\ T_{i,n}(t)\end{bmatrix}$$
$$=\frac{1}{2}*\underline{T_i}(t+\Delta t)+\frac{1}{2}*\underline{T_i}(t)$$

$$\underline{T_o}\left(t+\frac{\Delta t}{2}\right)=\begin{bmatrix}T_{o,1}\left(t+\frac{\Delta t}{2}\right)\\ T_{o,2}\left(t+\frac{\Delta t}{2}\right)\\ \vdots\\ T_{o,k}\left(t+\frac{\Delta t}{2}\right)\end{bmatrix}=\begin{bmatrix}\frac{T_{o,1}(t+\Delta t)+T_{o,1}(t)}{2}\\ \frac{T_{o,2}(t+\Delta t)+T_{o,2}(t)}{2}\\ \vdots\\ \frac{T_{o,k}(t+\Delta t)+T_{o,k}(t)}{2}\end{bmatrix}=\begin{bmatrix}\overline{T_{o,1}}(t)\\ \overline{T_{o,2}}(t)\\ \vdots\\ \overline{T_{o,k}}(t)\end{bmatrix}=\underline{\overline{T_o}}(t)$$

$$\underline{Q}\left(t+\frac{\Delta t}{2}\right)=\begin{bmatrix}Q_1\left(t+\frac{\Delta t}{2}\right)\\ Q_2\left(t+\frac{\Delta t}{2}\right)\\ \vdots\\ Q_l\left(t+\frac{\Delta t}{2}\right)\end{bmatrix}=\begin{bmatrix}\frac{Q_1(t+\Delta t)+Q_1(t)}{2}\\ \frac{Q_2(t+\Delta t)+Q_2(t)}{2}\\ \vdots\\ \frac{Q_l(t+\Delta t)+Q_l(t)}{2}\end{bmatrix}=\begin{bmatrix}\overline{Q_1}(t)\\ \overline{Q_2}(t)\\ \vdots\\ \overline{Q_l}(t)\end{bmatrix}=\underline{\bar{Q}}(t)$$

Through substitution the differential equation becomes:

$$\underline{\underline{M}}*\left(\frac{1}{\Delta t}*\underline{T_i}(t+\Delta t)-\frac{1}{\Delta t}*\underline{T_i}(t)\right)$$
$$=-\underline{\underline{U_1}}*\left(\frac{1}{2}*\underline{T_i}(t+\Delta t)+\frac{1}{2}*\underline{T_i}(t)\right)+\underline{\underline{U_2}}*\underline{\overline{T_o}}(t)+\underline{\underline{A}}*\underline{\bar{Q}}(t)$$

Through some algebraic manipulation, this becomes:

$$\underline{T_i}(t+\Delta t)=\underline{\underline{A}}*\underline{T_i}(t)+\underline{\underline{B}}*\underline{\overline{T_o}}(t)+\underline{\underline{C}}*\underline{\bar{Q}}(t)$$

$$\underline{\underline{A}} = \left(\frac{1}{\Delta t} * \underline{\underline{M}} + \frac{1}{2} * \underline{\underline{U_1}}\right)^{-1} * \left(\frac{1}{\Delta t} * \underline{\underline{M}} - \frac{1}{2} * \underline{\underline{U_1}}\right)$$
$$\underline{\underline{B}} = \left(\frac{1}{\Delta t} * \underline{\underline{M}} + \frac{1}{2} * \underline{\underline{U_1}}\right)^{-1} * \underline{\underline{U_2}}$$
$$\underline{\underline{C}} = \left(\frac{1}{\Delta t} * \underline{\underline{M}} + \frac{1}{2} * \underline{\underline{U_1}}\right)^{-1} * \underline{\underline{A}}$$

This is a system difference equation. Here, $\underline{\underline{A}}$ is the autoregressive matrix, $\underline{\underline{B}}$ is the environmental temperature correlation matrix, and $\underline{\underline{C}}$ is the heating correlation matrix. Other linearization schemes can be used. Here is a summary of a few options and their benefits:

1. Forward Euler Approximation:
   a. $\dot{T}_i(t) = \frac{T_i(t+\Delta t)-T_i(t)}{\Delta t}$
   b. One of the simplest numerical integration schemes, but it requires very high time resolution data to be accurate.
   c. Good for systems the state variable changes much more slowly than the observed data timestep.
2. Reverse Euler Approximation:
   a. $\dot{T}_i(t+\Delta t) = \frac{T_i(t+\Delta t)-T_i(t)}{\Delta t}$
   b. Quite similar to Forward Euler but runs through the data from back to front. Can have advantages if there are large changes in the state variable during the first several timesteps.
3. Midpoint Euler:
   a. $\dot{T}_i\left(t+\frac{\Delta t}{2}\right) = \frac{T_i(t+\Delta t)-T_i(t)}{\Delta t}$
   b. This is actually the arithmetic mean of the forward and reverse Euler approximations
   c. Does not require as high time resolution of data as the Forward or Reverse Euler approximations, and can handle noisy data reasonably well
4. Runge-Kutta
   a. This is a family of functions that extend the Euler Approximations (the Forward Euler is Runge-Kutta 1).
   b. Runge-Kutta 4 commonly referred to the classic is give by:
   c. $T_i(t+\Delta t) = T_i(t) + \frac{\Delta t}{6} * (k_1 + 2 * k_2 + 2 * k_3 + k_4)$
   d. $k_1 = \frac{dT_i}{dt}\big(t, T_i(t)\big)$

e. $k_2 = \frac{dT_i}{dt}\left(t + \frac{\Delta t}{2}, T_i(t) + \frac{k_1}{2} * \Delta t\right)$

f. $k_3 = \frac{dT_i}{dt}\left(t + \frac{\Delta t}{2}, T_i(t) + \frac{k_2}{2} * \Delta t\right)$

g. $k_4 = \frac{dT_i}{dt}(t + \Delta t, T_i(t) + k_3 * \Delta t)$

h. Can work with very time resolution limited data

Each of these classes of solutions is suitable for solving the initial value problem. That is a differential equation with a set of initial conditions. In general, the Forward Euler approximation should be sufficient for many applications. However, the Midpoint Euler approximation is not that much more complicated to implement and can offer higher accuracy, so we tend to favor it.

Stochastic State Space Description:

To convert this to a stochastic state space description, a stochastic noise term is added to the system difference equation:

$$\underline{T_i}(t + \Delta t) = \underline{\underline{A}} * \underline{T_i}(t) + \underline{\underline{B}} * \underline{\overline{T_o}}(t) + \underline{\underline{C}} * \underline{\overline{Q}}(t) + \underline{\varepsilon_1}(t)$$

Further, the observation equation is written to handle any unobserved dependent variables:

$$\underline{Y}(t) = \underline{\underline{D}} * \underline{T_i}(t) + \underline{\varepsilon_2}(t)$$

Here, $\underline{\underline{D}}$ is an appropriately sized matrix with entries 1 or 0 that reduces the state vector $\underline{T_i}(t)$ to the observed state vector $Y_i(t)$. In the case where all of the elements of the state vector are observed (as was the case for the Mongolia project), $\underline{\underline{D}}$ is the identity matrix. This gives:

$$\underline{Y}(t) = \underline{T_i}(t) + \underline{\varepsilon_2}(t)$$

$\underline{\varepsilon_2}(t)$ is a term to account for measurement noise. This can correspond to the accuracy of the instruments used. In the simplest case, with very accurate instruments, this term can be neglected.

Minimum Error Optimization:

The error used, here is the root mean square error (RMSE), which is defined by:

$$RMSE = \frac{1}{m}\sum_{t=1}^{m}\left[\left\langle \underline{Y}(t), \underline{Y}(t)\right\rangle\right]$$

Here, $\left\langle \underline{Y}(t), \underline{Y}(t)\right\rangle$, represents the dot product of $\underline{Y}(t)$ with itself or the vector norm of $\underline{Y}(t)$. The goal of the optimization is to produce a set of parameters, $\underline{\theta}$, which minimize the above error, i.e.:

$$\underline{\theta} = argmin_{\underline{\theta}}(RMSE)$$

In order to do this an initial guess for $\underline{\theta}$, $\underline{\theta_0}$, is made, and the guess is updated using a gradient descent update rule:

$$\underline{\theta^{n+1}} = \underline{\theta^n} - \eta * \nabla_{\underline{\theta}}(RMSE)$$

η, here is the learning rate, generally a small value of around 0.001 works well. The optimization is run a few hundred thousand times starting from various initial guesses to find a true global minimum.