

OFFICE ZONE HVAC ENERGY CALIBRATION FROM EASILY AVAILABLE SENSOR DATA

Tomoshi Otsuki¹, and Mitsunobu Yoshida¹
¹ Toshiba Corporation

ABSTRACT

Evaluation of HVAC performance in an office room by a virtual scenario is important when performing energy-saving analysis or peak power analysis. When evaluating HVAC power by calibration, the framework in which a simulator is made and simulation parameters are optimized, it is indispensable to minimize the engineering cost while achieving high simulation speed and high accuracy. The aim of this paper is to propose a low-cost HVAC calibration framework by using a small number of easily available sensor data. First, the HVAC power model at an office room is approximated with a linear heat balance equation. When multiple temperature time series in the room can be used as input, the heat balance equation becomes redundant. Thus, we propose a regression model to solve the problem. According to the temperature-based HVAC calibration results at a test room, the proposed method achieves results that are comparable to those of conventional simulated annealing methods but more stable.

INTRODUCTION

Building energy performance evaluation is important when constructing or renovating buildings, since evaluation results give the prediction of the energy cost and sometimes affect the property value. In particular, HVAC power consumption plays a key role in evaluation, since nearly half of the energy is used for HVAC and HVAC power varies depending on the internal and external environment. When evaluating HVAC power by calibration, it is necessary to make a simulator and optimize simulation parameters. For proper and quick calibration, it is indispensable to minimize the engineering cost while achieving high simulation speed and high accuracy.

Since a power bill is generally determined by the peak power, an effort is often made to reduce the peak-time demand or shift the power peak to reduce the power bill through the change of the facility's power curve. For example, use of HVAC is controlled by changing set temperature or controlling on/off, as well as by changing operation of the building or the scheduling of shift work.

The aim of this paper is to propose a low-cost HVAC

calibration framework combining a simplified simulation model with a smaller number of easily available sensor data. Here, an easily available sensor means mainly a thermometer. A thermometer with a data logger is commercially available at an acceptable price¹. Though calibration typically uses the power measurement data, we pursue calibration without power measurement.

In the problem setting, we approximate the HVAC power model at an office room with many thermometers corresponding to individual internal units. We consider a heat balance equation, which is the linear model of the room environment, activity, or HVAC performance parameters. Then, as the parameter optimization technique, in addition to the conventional simulated annealing, we propose regression-based methods that are effective when parameters are redundant in the multiple thermometer test setting. Finally, we show the power curve estimation results for the room when we suppose not energy consumption but temperature time series in training periods are measured. When temperature time series corresponding to the internal units of VRFs in the room can be used as input, the heat balance equation becomes redundant. Thus, we can use a regression model.

For the estimation of energy savings, ASHRAE Guideline 14 (ASHRAE 2002) provides three methods: (a) retrofit isolation, (b) whole building metering, and (c) calibrated simulation. For a detailed survey, see, for example, (Shengwei Wang and Xiao 2012). However, (a) and (b) need actual energy difference evaluation before and after retrofit. Thus, they do not match our purpose of comparing a virtual scenario.

On the other hand, (c) calibrated simulation, which combines the energy simulation with parameter optimization methods, may estimate the parameter at relatively lower cost with higher precision. EnergyPlus (EnergyPlus 2015), software widely used for simulation, is used in this work. EnergyPlus, whose formulation of the solution scheme starts with a heat balance for the room air, solves the equation with sequential simulation. Though construction of an elaborate simulator, as

¹We can obtain thermometer with data logger about 40,000 yen (~370 USD) in Japan

shown in the ASHRAE 1051-RP-based approach (Sun and Reddy 2006), can provide high-precision results, it requires tuning of multiple parameters, leading to high cost. Moreover, since calibrated simulation is based on the difference between estimated and actual power, it is necessary to measure power. However, usually it is very expensive to measure the power of individual rooms or zones, for example by an electricity meter that may cost a few thousand US dollars to install.

Even if absolute power values cannot be obtained, peak analysis can be performed if relative values can be estimated. We propose a way of estimating the power curve shape under the assumption that the power is proportional to the heat generated by HVAC.

For calibration methods, there are various optimization approaches, including simulated annealing (Robertson and Polly 2013) and genetic algorithm whose speed-up algorithm NSGA-II (Non-dominated Sorting Genetic Algorithm II) is well known (Brounlee and Wright 2015). Moreover, Openstudio (Ralph T. Muehleisen and Turner 2014) and Autotune (Joshua New and Shrestha 2012) are well-known frameworks combined with EnergyPlus.

Bayesian calibration (Yeonsook Heo and Augenbroe 2012), which estimates not only the optimum parameter but also the distribution of the parameter, and systematic evidence-based approaches (Paul Raftery and O'Donnell 2011) are proposed. In addition, methods without physical simulators are sometimes applied for one-month power estimation, including regression methods (Jimmy Besterberg and Olofsson 2014), which are, however, difficult to apply directly to one-hour power estimation necessary for one-day use case power evaluation.

ENERGY EVALUATION FRAMEWORK

Figure 1 shows the energy evaluation framework consisting of the calibration part and the virtual energy use case evaluation part. The calibration part uses measurement information including room, outside, and intake air temperature and the HVAC use case to estimate the HVAC performance parameters θ^* . Then, the energy evaluation part evaluates the energy by the virtual HVAC use case and outside air temperature pattern by HVAC simulation based on estimated parameter value θ^* .

The energy evaluation framework is used as follows:

1. Define the zone and identify the most influential input parameters $\theta = \{\theta_1, \theta_2, \dots\}$.
2. Place the thermometer and measure data (room and outside air temperature, HVAC use case, etc.) to estimate θ s.
3. Perform calibration to get θ^* s, the estimation of θ s, from measured data.

4. Evaluate virtual one-day HVAC use case (On/Off time, set temperature, etc.) at a specific external environment (outside air temperature, etc.), by simulation using θ^* s.

In the calibration part, we tune θ s by using HVAC simulation. Here, we evaluate the discrepancy between simulated and measured data by room air temperature data, since we do not use the power measurement data. Our simulator provides not only room air temperature but also heat generation from HVAC. The curve of HVAC heat generation and the power curve have the same shape, under the assumption that the power consumption is proportional to the heat generation amount. Thus, our framework can be used for energy evaluation or peak analysis for the virtual HVAC use case.

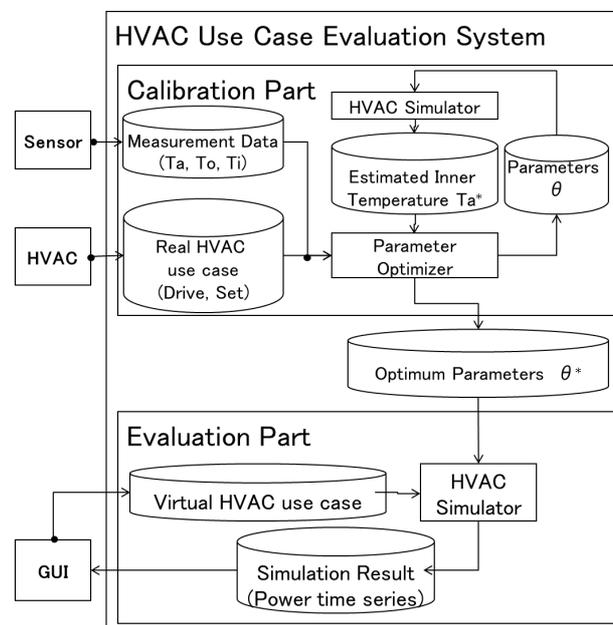


Figure 1: HVAC use case evaluation framework.

CASE-STUDY OFFICE ROOM

We conducted experiments in an office room in Kawasaki, Japan. The office room is rectangular about $500 m^2$, on the 3rd floor of a 5-story building. The east and west side of the room are windows facing outside, and the other sides of the room are walls and doors connecting to corridors. About 50 persons work in the room. In the room, there are 10 indoor units of VRF that are connected to the outdoor unit. VRFs are operated by power, and most of the power supplied is consumed in the outdoor unit. We can measure outdoor unit power. However, it is generally difficult to measure individual power meters in existing buildings at low cost. Thus, we use the power measurement data not

for calibration, but only for evaluation of our calibration results.

Available input data

To perform calibration, we need to obtain one-minute interval input data. We describe the available input and how to make one-minute data as follows:

- $T_a[t]$ (°C): Room air temperature at time t at the center of the room.
- $T_o[t]$ (°C): Outdoor air temperature at time t . The source is 1-hour-interval observatory data at Yokohama Local Meteorological Observatory ² and it is interpolated to make 1-min interval data.
- $T_i[t]$ (°C): Intake air temperature of the internal unit i at time t .
- $Drive_i[t]$: On (= 1) or Off (= 0) of the internal unit i at time t .
- $SET_i[t]$ (°C): Set temperature of the internal unit i at time t .
- $Insolation[t]$ (W): Insolation amount at time t . The source is 1-hour-interval observatory data at Tokyo District Meteorological Observatory ³ and it is interpolated to make 1-min interval data.
- $Person[t]$: The population in the room at time t . The source is the record of the work schedule, which includes individual worker's start and end times. However, it does not perfectly reflect the correct population in the room.
- $OA[t]$ (W): The OA ⁴ power consumption in the room at time t .
- $RealPower[h]$ (W): The power consumption for outdoor units of VRF in the room at each hour h . We use one-hour interval data to evaluate calibration results.

Among them, we can obtain T_a, T_o, T_i at relatively low cost by placing thermometers. Note that, we can measure intake temperature T_i by the thermometer hanging from the each internal unit. Moreover, it is considered to be relatively easy to estimate $Drive_i$ by the behavior of the intake air temperature time series T_i .

²Approximately 15 km away from the room. Though we measured outdoor air temperature of the building, some of the temperature data in the experimental period was missing. Thus, we used the meteorological data. We confirm that the average error between meteorological data and measured temperature is within 1 (°C) on average in the not-missing period.

³Approximately 15 km away from the room.

⁴OA means office automation appliances such as PCs, displays, and lightings. We assume that heat from OA can be measured by OA power consumption.

On the other hand, whether we can easily obtain the insolation, OA power, or the population, depends on the buildings' sensor environment. However, the following experimental results show that we do not necessarily require these data in our evaluation.

MODEL AND SIMULATOR

Linear heat balance equation

For HVAC modeling and the temperature simulation of the test room, we begin with the following heat balance equation:

$$\frac{(T_a[t + \Delta_t] - T_a[t])}{\Delta_t} \cdot C_p = \sum_{k=1}^K Q_k, \quad (1)$$

where K is the number of modeled heat sources (parameters) and Δ_t is the time slot length.

The left-hand side of Equation (1) represents the heat increase amount calculated from the temperature difference ($= T_a[t + \Delta_t] - T_a[t]$). The right-hand side represents the sum of heat generation from the following heat sources:

- Q_{base} : Background heat in 8:00~17:00 corresponding to the times when there is usually at least one person in the room
- Q_{person} : Heat from the human body
- Q_{insol} : Heat by insolation
- Q_{oa} : Heat generated by load (mainly from OA appliances)
- Q_{hvac} : Heat from indoor units of VRF
- Q_{out} : Heat transfer amount from outdoor

Using measured available input data we can represent each heat resource as follows:

$$Q_{base} = \theta_{base} \cdot Base[t], \quad (2)$$

$$Q_{person} = \theta_{person} \cdot Person[t], \quad (3)$$

$$Q_{insol} = \theta_{insolation} \cdot Insolation[t], \quad (4)$$

$$Q_{oa} = \theta_{oa} \cdot OA[t], \quad (5)$$

$$Q_{hvac} = \sum_{i \in I} \theta_{hvac} \cdot \{(SET_i[t] - T_i[t]) \cdot Drive_i[t]\}, \quad (6)$$

$$Q_{out} = \theta_{out} \cdot (T_o[t] - T_a[t]), \quad (7)$$

where $Base[t]$ is the value taking one if time t is within 8:00~17:00 and zero otherwise.

Here, $\theta = \{\theta_{base}, \theta_{person}, \theta_{insolation}, \theta_{oa}, \theta_{hvac}, \theta_{out}\}$ is the proportional coefficient corresponding to the heat resource Q_i , most of which is generally difficult to determine only by measuring. This model is based on and similar to the basic linear models in EnergyPlus (EnergyPlus 2015).

We assume background heat (Q_{base}), heat from human body (Q_{person}), heat from insolation (Q_{inso}), and heat from OA power consumption (Q_{oa}) are simply proportional to their corresponding input data as in Equation (2), (3), (4), and (5). Then in Equation (6), Q_{hvac} is the sum of heat from internal unit i that is proportional to the difference between set temperature ($=SET_i[t]$) and intake air temperature ($=T_i[t]$) when $Drive_i[t] = 1$. That is, θ_{hvac} corresponds to the compressor efficiency. Finally, heat transfer amount from outdoor (Q_{out}) shown in Equation (7) is proportional to the difference between outside air temperature ($=T_o[t]$) and room air temperature $T_a[t]$. That is, θ_{out} corresponds to the infiltration.

In the following, we assume the symbol $\Delta_t = 1$ for simplicity, and convert Equation (1) into the equation below:

$$(T_a[t+1] - T_a[t]) \cdot C_p = \sum_{k=1}^K Q_k. \quad (8)$$

In this case, $t = \{1, 2, \dots, T\}$ is a series of time slots with unit time length of Δ_t . For example, in considering the 10-day-length 1-min-interval simulation in the following discussion, we assume $T = 24 * 60 * 10$, $\Delta_t = 60$, and then assume that $t = 1$ corresponds to the 1st day's 0:00 ~ 0:01, $t = 2$ to the 1st day's 0:01 ~ 0:02, ..., and the final $t = 24 * 60 * 10$ corresponds to the 10th day's 23:59 ~ 24:00. In the following, we call this time slot simply "time".

As a result of the discussion above, we obtain the linear heat balance equation

$$(T_a[t+1] - T_a[t]) \cdot C_p = \sum_{k=1}^K Q_k = \sum_{k=1}^K a_{kt} \cdot \theta_k, \quad (9)$$

where K is the number of heat sources corresponding to $Q = \{Q_{base}, Q_{hvac}, Q_{inso}, Q_{person}, Q_{oa}, Q_{out}\}$, and a_{kt} s are known parameters calculated only by measurable input. Among them, Q_{base} , Q_{hvac} , and Q_{out} are calculated only by T_a, T_i, T_o, SET and $Drive$, whereas residual Q_{inso}, Q_{person} , and Q_{oa} are calculated from other measurement data.

Since the above model enables us to represent the heat balance equation as linear in parameters θ , we demonstrate the calibration method using the linearity of the equation.

Simulator

When we regard Equation (9) as a differential equation, we can construct the HVAC simulator. Actually, if $T_a[1]$, the initial value of T_a , as well as $\theta_1 \sim \theta_K, T_o, SET_i$, and $Drive_i$, is given, we can evolve T_a^* , the estimation of T_a , as the following algorithm:

Algorithm 1 simplified simulator of HVAC

Require: $\theta, T_o, T_a[1], SET_i, Drive_i$

Ensure: $T_a^*[t], Power[t]$

```

1:  $T_a^*[1] \leftarrow T_a[1]$ 
2: for  $t = 1$  to  $T$  do
3:   for  $i = 1$  to  $I$  do
4:      $T_i^*[t] \leftarrow T_a^*[t] + \alpha_i$ ;
5:   end for
6:    $Q \leftarrow 0.0$ 
7:   for  $k = 1$  to  $K$  do
8:      $a_{kt} \leftarrow \mathbf{get\_a}(T_o, T_i^*, T_a^*, SET_i, Drive_i, etc)$ 
9:      $Q \leftarrow Q + a_{kt} \cdot \theta_k$ 
10:  end for
11:   $T_a^*[t] \leftarrow T_a^*[t-1] + Q/C_p$ 
12:   $EstimatedPower[t] \leftarrow COP \cdot Q_{hvac}$ 
13: end for

```

Note that, the module **get_a** calculates the a_{kt} values, by Equation (2)–(7) with Equation (9). When we calculate the a_{kt} values, we use available input values including T_i, T_o, T_i, SET_i , and $Drive_i$.⁵

Additionally, we assume that the value of T_i , the temperature of the internal unit i , can be calculated by $T_a + \alpha_i$, where α_i is the constant value representing the difference between T_a and T_i . This is based on the fact that the difference between T_a and the T_i when the $Drive_i = 1$ is almost constant in our preliminary test.

Thus, in algorithm 1, first, for each time t , we obtain T_i^* , the estimation of T_i , for $i \in I$ by $T_i^*[t] \leftarrow T_a^*[t] + \alpha_i$. Then, for each time t , we calculate a_{kt} , and sum up $a_{kt} \cdot \theta_k$ to obtain total increase heat value Q . Finally, we update estimated room air temperature by updating $T_a^*[t]$, the estimation of $T_a[t]$, by $T_a^*[t-1] + Q/C_p$.

By repeating these procedures at each t , we can obtain $T_a^*[t]$ (and $T_i^*[t]$) for all t from the available input $T_o, T_n, SET_i, Drive_i$, and $T_a[1]$. Simultaneously, if we know the value of COP, the cost of the performance parameters that represents the proportionality coefficient between output heat and required power, we can obtain the estimated power time series $EstimatedPower[t]$ from Q_{hvac} .

CALIBRATION METHOD

We introduce two calibration methods. One is conventional simulation annealing. The other is a regression-based method in which we can make the most of the redundant available data.

Simulation-based calibration method

The simulated annealing (SA) based calibration method is a conventional parameter search method as shown in

⁵For simplicity, we exclude Q_{inso} , Q_{person} , and Q_{oa} from the model. When we include Q_{inso}, Q_{person} , and Q_{oa} , we add $Insolation[t], Person[t], OA[t]$ to the argument of **get_a**.

(Robertson and Polly 2013), which is a metaheuristic for approximating the global optimum solution of a given error function. Here, we use CVRMSE as the error function (definition shown later). We generate 500 results from the above simulator with different θ parameters within the preset parameter range. Then, we repeat the whole process 10 times from the different initial solution with a different pseudo-random number.

Regression-based calibration method

In the heat balance equation, Equation (9) holds at all times $t = 1 \dots T$:

$$(T_a[t+1] - T_a[t]) \cdot C_p = \sum_{k=1}^K a_{kt} \cdot \theta_k \quad (t = 1 \dots T). \quad (10)$$

Here, a_{kt} in the right-hand side in Equation (10) can be obtained from available data by Equation (2)-(7).

On the other hand, values in the left-hand side in Equation (10) are also calculated from available data, and so we set this value as b_t :

$$b_t = (T_a[t+1] - T_a[t]) \cdot C_p. \quad (11)$$

Thus, unknown parameters are only $\theta_1 \sim \theta_K$. In this case, in order to choose the parameter $\theta_1 \sim \theta_K$ to hold these T equations as well as possible, we need to minimize the squared error of the right-hand side and left-hand side of Equation (10).

Thus, we set the evaluation criteria as follows:

$$\theta^* = \arg \min_{\theta} \sum_{t=1}^T \left\{ \left(\sum_{k=1}^K a_{kt} \theta_k \right) - b_t \right\}^2, \quad (12)$$

where $\theta^* = \{\theta_1^*, \theta_2^*, \dots, \theta_K^*\}$ are the optimal value of the θ .

This optimization problem, in fact, is equivalent to obtaining the regression parameter θ_k^* in the multiple regression problem without constant term:

$$b = \sum_{k=1}^K a_k \theta_k, \quad (13)$$

where $b = \{b_1, b_2, \dots, b_T\}$ is the criterion (or dependent) variable, and $a_k = \{a_{k1}, a_{k2}, \dots, a_{kT}\}$ are the $k(k = 1, 2, \dots, K)$ th predictor variables.

As a result of the above transformation, the calibration problem of heat balance equation (9) is reduced to the multiple linear regression (without constant term) problem.

Improved regression model

For this regression model, we can improve the results by changing the calibration objective from heat to the room air temperature.

Equation (10) holds for each time t , so that we can obtain equations again by summing up multiple equations for $t' = 1$ to t . By summing up from time 1 to t , we obtain the equation whose left-hand side means $T_a[t+1] - T_a[1]$, since the left-hand side of Equation (10) is the difference between $T_a[t+1]$ and $T_a[1]$.

When we neglect $T_a[1]$, the left-hand side corresponds to the value of room air temperature T_a at $t+1$, which introduces the regression of T_a .

Actually, summing up

$$(T_a[t+1] - T_a[t]) \cdot C_p = \sum_{k=1}^K a_{kt} \cdot \theta_k \quad (t = 1 \dots T) \quad (14)$$

from time $t' = 1$ to t , we obtain T equations:

$$\begin{aligned} & \sum_{t'=1}^t (T_a[t'+1] - T_a[t']) \cdot C_p \\ &= (T_a[t+1] - T_a[1]) \cdot C_p \end{aligned} \quad (15)$$

$$= \sum_{t'=1}^t \sum_{k=1}^K a_{kt'} \cdot \theta_k \quad (t = 1 \dots T). \quad (16)$$

When we consider the minimization of the difference between Equation (15) and (16), we can apply similar logic as discussed above. When we assume

$$f_{kt} = \sum_{t'=1}^t a_{kt'}, \quad (17)$$

and

$$g_t = (T_a[t+1] - T_a[1]) \cdot C_p, \quad (18)$$

then all f_{kt}, g_t can be calculated from available input data and θ^* can be obtained by solving the following optimization problem:

$$\theta^* = \arg \min_{\theta} \sum_{t=1}^T \left\{ \left(\sum_{k=1}^K f_{kt} \theta_k \right) - g_t \right\}^2. \quad (19)$$

Moreover, this optimization problem is similarly equivalent to obtaining the regression parameter θ_k^* in the multiple regression problem without constant term:

$$g = \sum_{k=1}^K f_k \theta_k, \quad (20)$$

where $g = \{g_1, g_2, \dots, g_T\}$ is criterion (or dependent) variable, and $f_k = \{f_{k1}, f_{k2}, \dots, f_{kT}\}$ are the $k(k = 1, 2, \dots, K)$ th predictor variables.

By a preliminary test, the result of minimizing the squared error of the room air temperature by Equation (19) is better than that by Equation (12). The reason is that Equation (19), which is obtained by summing up the equations, has a fewer measurement error than that of the original equation (12). Thus, we apply regression-based Equation (19) as follows.

Some enhancement techniques for calibrating VRF

Finally, we show two enhancement techniques for calibrating VRF in our test case.

First of all, in the discussion above, the heat from internal unit i is represented by the following equation

$$Q_{hvac} = \theta_{hvac} \cdot (\text{SET}_i[t] - T_i[t]) \cdot \text{Drive}_i[t]. \quad (21)$$

On the other hand, according to the specification of the VRF, it provides much more heat right after switch-on than usual. Thus, we separate the operating mode into two modes, namely, normal mode and boost mode, and separate the heat Q_{hvac} into two sources Q_{normal} and Q_{boost} :

- Q_{boost} : heat from the internal unit (within 30 minutes after switch is ON)
- Q_{normal} : heat from the internal unit (otherwise)

Then, we calculate the following parameter Boost_i from available Drive_i

- $\text{Boost}_i[t]$: The variable that takes 1 when internal unit i is within 30 minutes after switch is ON, and takes 0 otherwise.

and assume θ_{boost} is the corresponding boost-time heat amount from the internal unit within 30 minutes after switch is ON. In preliminary tests, the result improves when we add the Q_{boost} term.

Secondly, in the discussion above, the heat from internal unit i is proportional to the difference $\text{SET}_i[t] - T_i[t]$. However, when we drop the $\text{SET}_i[t] - T_i[t]$ term, the result improves in the preliminary test. Thus, we drop the $\text{SET}_i[t] - T_i[t]$ term. This is because the set temperature is often different from the value when calculating required heat in actual control.

Then, the term of the heat from unit i in our test is summarized as follows:

$$\begin{aligned} Q'_{hvac} &= Q_{normal} + Q_{boost} \\ &= \sum_{i=1}^I (\theta_{normal} \cdot \text{Drive}_i[t] + \theta_{boost} \cdot \text{Boost}_i[t]). \end{aligned} \quad (22)$$

COMPUTATIONAL EXPERIMENTS

Experimental condition

We evaluate the obtained parameters θ^* by the conventional simulating annealing method (SA) and the proposed regression-based method (REG). In each method, using θ^* , we obtain Q_{hvac} by HVAC simulation of the virtual one-day use case in Algorithm 1, and measure the performance of the Q_{hvac} estimation that is assumed to be proportional to the energy use case.

Conventionally, we calculate the mean squared error between actual hourly (or monthly) power and estimated power by CVMSE of actual values as shown in ASHRAE Guideline 14 (ASHRAE 2002).

However, by our problem setting, we can obtain not the power value itself but the heat value from VRF, which we assume is proportional to the power value. Thus, making use of the proportional relationship between heat value Q'_{hvac} and power, we compare the standardized value of the actual power and estimated Q'_{hvac} values.

Thus, we evaluate each method's root mean square error (CVRMSE) performance by standardization and calculating the error between actual power and Q_{hvac} estimation time series. Daily CVRMSE is calculated as follows:

$$\text{CVRMSE} = \frac{\sqrt{\sum_{h=1}^{24} \{\text{Real}[h] - \text{Est}[h]\}^2 / 24}}{\sum_{h=1}^{24} \text{Real}[h] / 24}, \quad (23)$$

where

$$\text{Real}[h] = \frac{\text{RealPower}[h]}{\sum_{h=1}^{24} \text{RealPower}[h]}, \quad (24)$$

$$\text{Est}[h] = \frac{\sum_{t=60 \cdot (h-1)+1}^{60 \cdot h} Q'_{hvac}[t]}{\sum_{t=1}^{24 \cdot 60} Q'_{hvac}[t]}. \quad (25)$$

Then, we evaluate each method by LOOCV (leave-one-out cross-validation). In our experiment, the total period is 10 days from weekday Aug. 24 to Sep. 24 in 2015. Thus, we set 9 days as the training period and another day as the test period. In each method, we obtain θ^* by the training period data, and then evaluate CVRMSE by the residual test day's simulation by θ^* . Totally, we evaluate 10 test day results by staggering the test day, so that we show the average CVRMSE of 10 days' results.

Moreover, we change the heat resource set of Q to obtain the best model for minimizing the CVRMSE. The experiments were performed on a computer with Intel Core 2 Duo CPU E8400 3.00 GHz Processor and 2GB RAM. SA is implemented in C and takes about 1 minute for getting one test data, whereas REG is implemented in R 3.1.0 and takes about 10 seconds.

CVRMSE evaluation results

Table 1 shows the CVRMSE of each day and the average values for SA and REG. Here, we show the result for the best model (of both SA and REG) based on the sensitivity analysis in the next subsection. The best model is the one including Q_{base} , Q_{boost} , and Q_{hvac} (that is, excluding Q_{oa} , Q_{person} and Q_{base}). The average CVRMSE for HVAC power is 43.4% in both cases and they meet the ASHRAE standards for hourly calibration data's CVRMSE: within 30%, when we suppose HVAC power usually accounts for at most 50% of whole building power (even if there exists at most 16% RMSE in residual power estimation, we can meet the ASHRAE standards, since $(43.4 + 16.0) / 2.0 \leq 30.0$). Figure 2 shows the best model example of $\text{Est}[h]$ s by SA and REG against actual power data $\text{Real}[h]$ in Aug. 24.

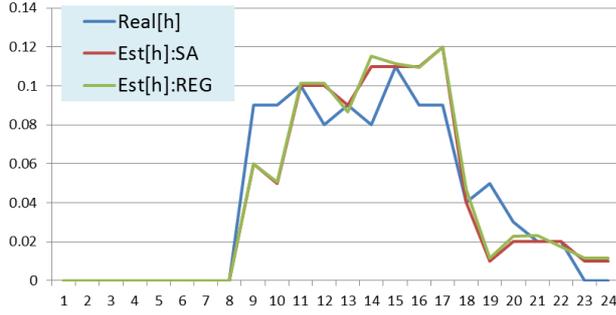


Figure 2: Example of Est[h]s by SA and REG against actual power data Real[h] in Aug. 24. In this case two estimation results are almost same.

Table 1: Each day's CVRMSE results and average by REG and SA (in %).

date	REG	SA
Aug. 24	43.6	42.7
Aug. 25	42.4	41.8
Aug. 26	44.2	37.2
Aug. 27	24.0	24.2
Aug. 28	53.2	55.4
Aug. 31	58.4	59.1
Sep. 1	42.2	42.9
Sep. 2	40.3	41.7
Sep. 3	46.6	49.5
Sep. 4	39.2	39.7
average	43.4	43.4

Sensitivity analysis results

Table 2 shows the sensitivity analysis results. Each row shows the CVRMSE values and the difference against the best model. The "optimal case" row shows the best model including only Q_{base} , Q_{boost} , and Q_{hvac} . Then, other rows show the absolute CVRMSE and relative CVRMSE against the best model by dropping or adding each q heat source difference from the best model. For example, "- Q_{base} " rows represent the model by dropping Q_{base} from the best model and "+ Q_{oa} " rows represent the model adding the Q_{oa} to the best model. Moreover, " $Q'_{hvac} \rightarrow Q_{hvac}$ " row means the result in which we use Q_{hvac} instead of Q'_{hvac} and set SET_i at 20.0 (20.0 is the best result case).

Whereas REG and SA indicates the similar CVRMSE values in the best case, the result of REG is more stable than that of SA in that the result values in not-best cases are nearer the best values in REG. This is because REG neglects the parameters whose contributions are low, whereas SA makes use of all parameters equally.

Table 2: Sensitivity analysis results of CVRMSE (in %)

	REG	SA
optimal case	43.4	43.4
+ Q_{inso}	43.4 (+0.0)	46.3 (+2.9)
+ Q_{oa}	43.4 (+0.0)	43.4 (+0.0)
+ Q_{person}	43.4 (+0.0)	44.3 (+0.9)
- Q_{base}	43.4 (+0.0)	45.4 (+2.0)
- Q_{boost}	45.8 (+2.4)	45.7 (+2.3)
$Q'_{hvac} \rightarrow Q_{hvac}$	46.4 (+3.0)	48.2 (+4.8)

Moreover, the results show the significance of incorporating the heat term, whereas the number of people, OA power consumption or insolation is unnecessary in terms of CVRMSE.

Comparison of SA and REG

From our test cases, SA has some demerits as follows:

- The parameter range must be set beforehand
- Calculation cost in SA is relatively large, since it requires many simulation iterations
- SA cannot use all the measured T_a , T_i values.

Inversely, the benefit of REG is as follows:

- REG is relatively stable for parameter setting. When the model includes unnecessary parameters, REG can neglect the parameters whose contributions are low by setting the weights small.
- REG can use redundant available parameters
- The optimization process of REG is faster since it is just a simple regression method

Currently, though we use a simple multiple regression model for a few parameters, we may improve the result by increasing the number of sensors. When we increase the parameters, we run the risk of overfitting to the training data. However, variable selection techniques such as LASSO or ridge regression may solve the overfitting problem.

For both methods, correct prior information such as area of the room, heat capacity or HVAC specification in detail may improve the results, since we can easily implement the information as prior distribution. Moreover, to deal with heat transfer strictly, splitting the single-zone room model to a multi-zone model may work well.

CONCLUSION

We propose a low-cost HVAC calibration framework by using a small number of easily available sensor input data to evaluate energy-performance of an office room.

First we propose the HVAC simulation model using a linear heat balance equation combined with multiple temperature time series in the room. Then, we propose a regression-based calibration method making use of the redundancy of input data to solve the calibration problem. According to the temperature-based HVAC calibration results for a test room, the proposed method achieves the results comparable to those of conventional simulated annealing methods. Moreover, sensitivity analysis shows that the proposed methods provide stable results even when the input includes unnecessary data. Thus, our methods constitute a good candidate for calibration with easily available multiple sensor information. In the future, the result of the proposed model may work if a more sophisticated model is used.

REFERENCES

- ASHRAE. 2002. *ASHRAE Guideline 14 - measurement of energy and demand savings*. American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Atlanta, GA.
- Brounlee, Alexander E.I., and Jonathan A. Wright. 2015. "Constrained, mixed-integer and multi-objective optimisation of building designs by NSGA-II with fitness approximation." *Applied Soft Computing*.
- EnergyPlus. 2015. *Engineering Reference*. "https://energyplus.net/sites/all/modules/custom/nrel_custom/pdfs/pdfs_v8.4.0/EngineeringReference.pdf".
- Jimmy Besterberg, Staffan Andersson, and Thomas Olofsson. 2014. "Robustness of a regression approach, aimed for calibration of whole building energy simulation tools." *Energy and Buildings*.
- Joshua New, Jibonananda Sanyal, Mahabir Bhandari, and Som Shrestha. 2012. "Autotune E+ building energy models." In *Proceedings of the 5th National SimBuild of IBPSA-USA, International Building Performance Simulation Association*.
- Paul Raftery, Marcus Keane, and James O'Donnell. 2011. "Calibrating whole building energy models: An evidence-based methodology."
- Ralph T. Muehleisen, Brian Craig, Daniel Macumber Elaine Hale, and Jason Turner. 2014. "Integration of the CEN/ISO monthly building energy model into OpenStudio." *ACEEE Summer Study on Energy Efficiency in Buildings*.
- Robertson, Joseph, and Ben Polly. 2013. "Evaluation of automated model calibration techniques for residential building energy simulation." Technical Report, National Renewable Energy Laboratory (NREL).
- Shengwei Wang, Chengchu Yan, and Fu Xiao. 2012. "Quantitative energy performance assessment methods for existing buildings."
- Sun, Jian, and T. Agami Reddy. 2006. "Calibration of building energy simulation programs using the analytic optimization approach (RP-1051)." *HVAC and R Research*.
- Yeonsook Heo, Ruchi Choudhary, and Godfried Augenbroe. 2012. "Calibration of building energy models for retrofit analysis under uncertainty." *Energy and Buildings* 47.

NOMENCLATURE

T	The number of times
I	The number of internal units of VRF in the room
K	The number of heat sources for evaluation including base, HVAC, insulation, person, etc
Δ_t	Unit time length
C_p	Heat capacitance of the room
t	Time
$T_a[t]$	Room air temperature of the room at time t
$T_o[t]$	Outside air temperature of the room at time t
$T_i[t]$	Intake air temperature for the internal unit i at time t
$\text{Drive}_i[t]$	On/Off for the internal unit i at time t (1: ON, 0: OFF)
$\text{SET}_i[t]$	Set temperature for the internal unit i at time t
a_{kt}	Known coefficient parameters for heat source k at time t
θ_k	Unknown parameters for heat source k
Q_k	Heat generation term from heat source k
T_a^*	Estimated value for room air temperature
T_i^*	Estimated value for intake air temperature of internal unit i
θ_k^*	Estimated value for parameter θ_k
$\text{Base}[t]$	The time series indicating one if time t is within 8:00~17:00 and zero otherwise
$\text{Insolation}[t]$	Insolation amount of the room at time t
$\text{Person}[t]$	The population of the room at time t
$\text{OA}[t]$	The OA power of the room at time t
$\text{RealPower}[t]$	Total power of the VRF outdoor units at time t