ADVANCES IN CALIBRATION OF BUILDING ENERGY MODELS TO TIME SERIES DATA

David Goldwasser[1], Brian Ball[1], Amanda Farthing[1], Stephen Frank[1], and Piljae Im[2]
[1]National Renewable Energy Laboratory, Golden, CO
[2]Oakridge National Laboratory, Oak Ridge, TN

ABSTRACT

Advanced building analytics and control applications, such as fault detection and diagnostics and model predictive control, benefit from building energy models calibrated to time series data. This paper recommends best practices for aligning EnergyPlus™ model time series output with measured building performance. These include the use of subcalibration procedures to fine-tune envelope, heat transfer, and mechanical system parameters; initialization of zone temperatures to match observed conditions at the beginning of each calibration period; and calculation of calibration metrics across multiple run periods using a new OpenStudio reporting measure.

INTRODUCTION

Building energy models and simulations provide a virtual laboratory for exploring and refining energy-saving design, equipment, and operational strategies before using them in buildings. However, discrepancies between modeled and actual building performance are common (Calì et al. 2016; Turner and Frankel 2008; Herrando et al. 2016). These discrepancies affect projections of energy savings, the performance of predictive control algorithms, and confidence in the usefulness of building energy simulations.

Therefore, calibration to measured data is necessary to ensure accurate representation of a building's physical attributes, equipment, schedules, and energy use. Calibration involves iterative improvements to bring model outputs (e.g., electricity consumption, zone temperatures) in line with measured data from an actual building. Coakley et al. (2014) provide a detailed review of existing calibration tools and techniques for building energy models.

For use cases such as automated fault detection and diagnosis (AFDD) and model predictive control (MPC), which depend on the timing and pattern of energy consumption, it is important to calibrate energy consumption to time series data (as opposed to monthly or annual totals). Use of hourly data for time series calibration is common in the literature (Coakley et al.

2012; Haberl and Bou-Saada 1998; Soebarto 1997; Bou-Saada and Haberl 1995; Harmer and Henze 2015). However, if accurate load profile shapes are needed at subhourly resolutions, higher-resolution (e.g., 15-minute) metrics should be used. In addition to calibrating to whole-building electricity consumption, calibration of building subcomponents (e.g., equipment electricity use and rooftop air handling unit [RTU] air temperatures) can expose otherwise confounded errors and improve model accuracy (Ji and Xu 2015). Although necessary, calibration at this level of resolution can be time consuming and error prone, particularly if metrics are manually calculated across multiple run periods and if a systematic calibration approach is not employed.

This paper describes lessons learned and recommends best practices for calibrating building energy models to time series data for nonconsecutive calibration periods when building operations are unknown during noncalibration periods. Calibration of the Oak Ridge National Laboratory (ORNL) flexible research platform (FRP) #2, an experimental facility designed to resemble a small office building, serves as a case study. The authors performed calibration for this facility using 15-minute time series data and a newly developed, publicly available OpenStudio reporting measure that calculates calibration metrics across multiple run periods (NRELa).

ASHRAE Guideline 14 standards for normalized mean bias error (NMBE) and coefficient of variation of the root mean square error (CVRMSE) (ASHRAE 2002) were used as calibration targets. For time series data, the guideline considers a model well-calibrated if NMBE does not exceed 10% and CVRMSE does not exceed 30%. Although these targets are intended for hourly data for whole-building energy consumption, in this work the authors calculated the metrics using 15-minute whole-building and end-use electricity consumption data, resulting in more stringent calibration requirements.

This paper outlines the calibration methods used for the case study, describes key challenges, and proposes mitigation strategies for these challenges. Topics covered include data preparation; initialization of zone temperatures; identification of anomalies to exclude

from calibration periods; handling missing or incorrect values in measured data; and calibration across multiple nonconsecutive run periods.

FACILITY DESCRIPTION

The ORNL FRP #2 is a 3,200 ft^2 structure designed to emulate a 1980s-era office building (Figure 1). The building is unoccupied and used only for experiments. Occupancy is simulated by process control of lighting, humidifiers for human-based latent load, and heaters for miscellaneous electrical loads (Im and Bhandari 2016).

Figure 1 Flexible research platform #2 at Oak Ridge National Laboratory (Image courtesy of Piljae Im/ORNL)

Im and Bhandari (2016) provide a detailed description of the facility's design, envelope, equipment, instrumentation, and research capabilities. The heating, ventilation, and air conditioning (HVAC) system used within the FRP #2 during the test periods was an RTU connected to a multi-zone variable air volume (VAV) system. The RTU is a Trane® YCD150 12.5-ton unit with an energy efficiency rating (EER) of 9.6. The connected VAV system serves a total of 10 zones (8 perimeter and 2 core). Each VAV box includes electric resistance reheat. To better control the outside air introduced to the building during experiments, the RTU's outdoor air intake is permanently blocked. In addition, the RTU's natural gas heating system was disabled, such that the RTU provided cooling only and the VAV boxes provided all required heating.

CALIBRATION METHODOLOGY

To facilitate model calibration, the athors developed the model using a built-up workflow of OpenStudio measures (scripts), programmatically applying a series of controlled changes to the initial model (NRELb). This methodology allowed systematic, independent adjustment of model characteristics during calibration. The automated workflow started with a baseline model developed according to the operating conditions of the

FRP #2 during the calibration run periods. Next, a series of measures, each of which focused on different calibration variables, was applied. The OpenStudio model was then translated to EnergyPlus and run using actual meteorological year (AMY) weather data collected onsite. Lastly, reporting measures automatically generated the calibration metrics. This modular workflow, depicted in Figure 2, made it easy to modify assumptions or underlying data, add additional calibration periods, and quickly rerun the process. All tools and measures referenced in this paper are open source and freely available for download (NRELe).

Calibration variables

The primary variables for the calibration parameter space are:

- Primary fan attributes
- Air loop set point strategy
- Infiltration schedule
- Heating and cooling set point adjustments
- Supply air duct leakage
- Internal mass (various arguments for different areas of the building)
- Solar heat gain coefficient for windows
- Window U-factor
- Air mixing between zones
- Roof and wall insulation values.

These variables represent building characteristics that are unknown or uncertain and are therefore appropriate for adjustment during calibration.

Measured calibration data

FRP #2 includes extensive permanent instrumentation that collects building performance data at a 1-minute resolution. Data available include electricity and natural gas consumption for the entire building and individual equipment; outside, supply, mixed, return, and zone temperature and humidity; supply and return air volumetric and mass flow rates; and HVAC system commands. Of these, the authors selected several data streams likely to be available from a typical building automation system and/or from electrical submeters:

- Whole building electricity consumption
- End-use electricity consumption for heating, cooling, fans, lighting, and plug loads
- Supply and return air temperatures at the RTU
- Zone air temperatures.

Because heating is provided exclusively by the VAV boxes, it was possible to meter cooling and heating electricity separately. End-use electricity consumption, supply and return air temperatures, and zone air
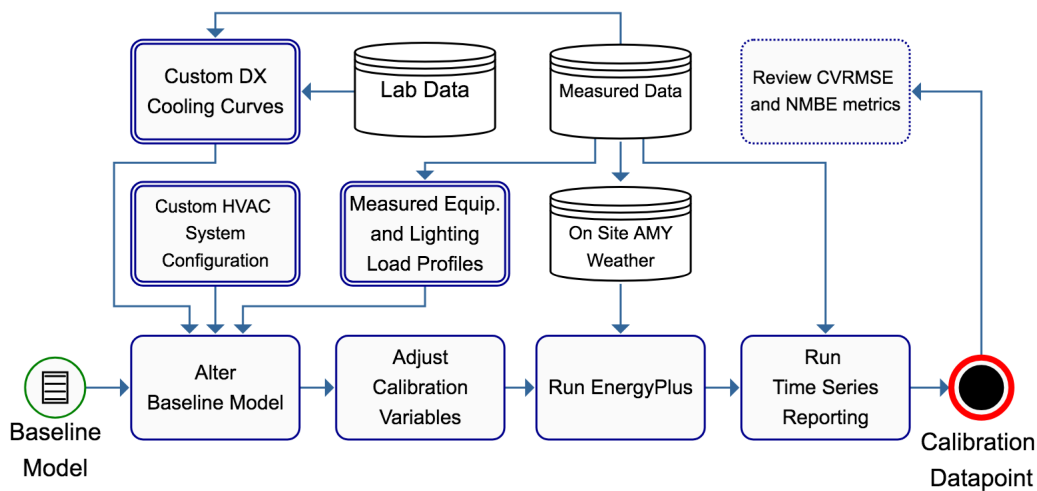
Figure 2 Workflow for calibrating building energy models to time series data

temperatures were used for quantitative calibration. However, measured zone temperatures were not available for the stairs or plenum zones.

One-minute interval calibration data were available for 11 periods of 2 days each, spread over 9 months. The first calibration period was July 25–26, 2015, and the last was March 5–6, 2016. These periods represent days when calibration data were available, and the operating condition of the building matched the configuration described in "FACILITY DESCRIPTION" above. To enable free oscillation calibration as outlined by Ruiz et al. (2016), one additional period was run from August 11–16, 2016, with the FRP #2 HVAC system disabled.

Prior to calibration, the authors verified the quality of the relevant data, time-averaged to a 15-minute resolution, converted to units compatible with EnergyPlus, and aligned to Eastern Standard Time (no daylight saving). These quality-controlled data then served as ground truth for the calibration process.

Weather data

Weather is a major determinant of building loads and an important source of uncertainty in building energy simulations. Erba et al. (2017) show the significant effect of weather data choice on simulated building performance and energy use. Therefore, for calibration to time series data, local weather data are essential. The National Weather Service provides historical hourly weather data for a large number of locations (NOAA 2017). In this case study, the authors used hourly data collected from a weather station located on the roof of FRP #2. Locally measured data include ambient dry bulb temperature, ambient humidity, direct and diffuse solar irradiance, wind speed, and wind direction. These data were converted to the EnergyPlus Weather (EPW) file format (NRELc) for use with the energy model.

Baseline building energy model

The baseline FRP #2 EnergyPlus model was developed by ORNL's Building Technologies Research and Integration Center as a part of the FRP research project (Buckberry and Bhandari 2012). The energy model reflects the geometry and basic characteristics of the physical FRP #2 building. However, several adjustments were needed prior to calibration to align this model with FRP #2 operations during the case study test periods. First, any model inputs known with certainty were assigned to their measured values and held constant, which removes them from the parameter space for the calibration. For example, the authors developed OpenStudio measures to inject measured electric equipment, lighting, and other equipment energy (simulated occupants) using schedules derived from the measured data. This approach minimized lighting, plug loads, and occupancy as sources of error, which allowed the calibration to focus on the envelope and HVAC system. Second, HVAC system control schedules were aligned with the experimental configuration. Third, infiltration schedules were altered so there was not a reduced daytime rate because there was no positive internal pressure due to outdoor air during HVAC operation. Finally, the generic DX cooling performance curves used in the baseline model were replaced with performance curves generated from laboratory test data (NRELf) with the EER and capacity adjusted to match that of the unit installed on the FRP #2.

Model evaluation with OpenStudio time series reporting measure

The calibration acceptance criteria established by ASHRAE Guideline 14 are NMBE <10% and CVRMSE <30% for hourly, whole-building energy

consumption (ASHRAE 2002). To compute these metrics, the authors ran each calibration data point for 12 run periods in EnergyPlus, then used an OpenStudio reporting measure to generate the respective metrics at the 15-minute time interval (Figure 3). The 11 conditioned run periods as a group were assigned a single CVRMSE and NMBE value for each objective function. The 12th free oscillation unconditioned run period was assigned a CVRMSE and NMBE for average building air temperature. The free oscillation run period provides an excellent way to calibrate envelope and thermal mass in the model (Ruiz et al. 2016). In the free oscillation period, the zone temperatures were initialized to the correct starting values, as described in "Recommendations and lessons learned."
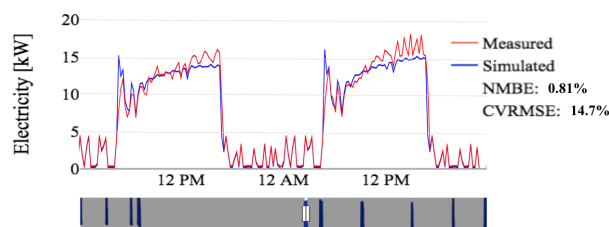

Figure 3 Multiple run periods and associated metrics calculated by OpenStudio reporting measure; simulated (blue) and measured (red) load shown for July 25–26, while metrics are for all days

Iterative model adjustments

The authors used an OpenStudio Parametric Analysis Tool (PAT) project to manage both manual and algorithmic calibration analyses (NRELd). After manual runs to ensure modeled results were within realistic bounds, the variables were confined to logical ranges. Next, a Morris screening method (Campolongo et al. 2007) was executed to inform the sensitivity of the variables and to indicate which variables should be included in the parameter space. Engineering judgment and knowledge of the building informed the parameter space by limiting variables to reasonable ranges considering onsite conditions. Next, the analysis was updated to sample all variables simultaneously to characterize interactions between variables.

The authors selected a "best point" model from the output of the multivariable sampling analysis: a data point that was a good fit (low CVRMSE and NMBE) for both the whole building and individual subcomponents. The updated model was the starting point for an optimized calibration with an objective function that incorporated the NMBE and CVRMSE metrics.

Through visual inspection of data points (trial models) generated from this optimization, the authors discovered

a number of atypical days during which the measured data differed substantially from the modeled data (Figure 4). These anomalies persisted across many model perturbations, which suggested that they could not be adequately explained by adjusting the calibration variables. However, modeled and measured data were well aligned during other times. Because the anomalies were limited to well-defined time periods, showed no clear pattern, and persisted across all trial models, they are hypothesized to represent unknown, transient changes in the operating conditions for the building, such as temporarily altered HVAC controls or substantial infiltration from an open door.

The goal of calibration is to properly capture the normal operating condition of the building. If the anomalies present a clear pattern, then they suggest an error in the underlying modeling assumptions that should be corrected. However, in the absence of a clear pattern, it would be possible but extremely time consuming to manually tune schedules for the model parameters until the model matched each observed anomaly. Instead, the authors developed the best fit for the non-anomalous periods by removing 7 of the 22 conditioned days from the objective function and rerunning the optimization using the remaining 15 days. When reporting final calibration results, values were provided both for these "typical" days and the full set of days. Although a lower value could have been achieved for all days by optimizing the full set, the optimization performed against the typical days better represents true building parameters and better predicts typical future behavior.
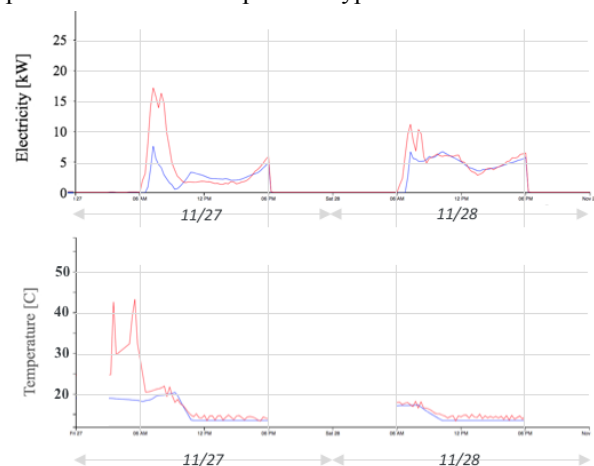

Figure 4 Simulated data (blue) vs. measured data (red) for November 27–28 for heating electricity (top) and supply air temperature (bottom); November 27 was one of the atypical days removed from the optimization. Data are not available for supply air temperature when system is off.

Table 1 CVRMSE and NMBE metrics for modeled vs. measured data for whole building and components for "All Days" and "Typical Days"

Component	Type	All Day		Typical Days	
		CVRMSE (%)	NMBE (%)	CVRMSE (%)	NMBE (%)
Whole Building	Electricity	26.20%	5.70%	14.70%	0.81%
Cooling (RTU)	Electricity	44.50%	4.20%	40.60%	0.03%
Heating (zone terminals)	Electricity	120.90%	24.40%	69.20%	5.20%
Fan	Electricity	15.50%	4.00%	14.60%	4.00%
Lights	Electricity	3.30%	0.23%	3.80%	0.28%
Electric Equipment	Electricity	3.90%	0.20%	4.50%	0.24%
RTU Entering Air	Temperature	9.90%	7.90%	6.90%	7.00%
RTU Exiting Air	Temperature	21.20%	5.70%	20.10%	6.80%
No HVAC Avg. Bldg. Air	Temperature	0.68%	0.04%	0.68%	0.04%

DISCUSSION AND RESULTS

FRP #2 calibration results

Following the optimization, the authors selected a final model with balanced results for all the time series evaluated. Had total electricity consumption been the only calibration variable, it would be possible to obtain a lower CVRMSE for typical days. However, when submetered and environmental data are also considered, it becomes apparent that this whole-building metric target would be a result of incorrect underlying assumptions. When focusing on maintaining a good fit across subcomponents as well, the calibration achieved a total electricity CVRMSE of 14.7% (shown in Figure 5). Fit metric values for all days and typical days for - whole-building and building subcomponents are summarized in Table 1. Figure 6 and Figure 7 show examples of calibrated electricity and temperature profiles, respectively.
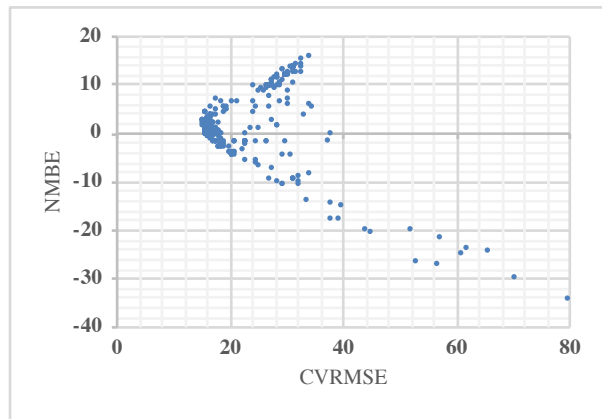
Most subcomponents had CVRMSE values less than 30%. Exceptions included cooling (40.6%) and heating (69.2%). The large heating error has two likely causes. First, when heating is a small fraction of cooling (e.g., during summer months), a small error in supply air temperature or envelope conditions can create a large error in heating. Another possible cause is that the small time step places more stress on the accuracy of the control schedules for the building. There could be a clock drift between the model and the real building that causes misalignment in the heating profiles even when the total heat produced aligns well. As a result of this misalignment, CVRMSE may be high even when NMBE is low.
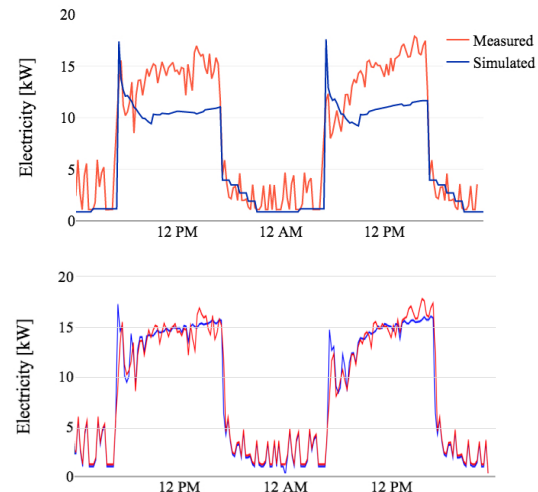

Figure 6 Whole-building electricity consumption during the August 7–8 run period; simulated data (blue) vs. measured data (red), pre- (top) and post- (bottom) calibration
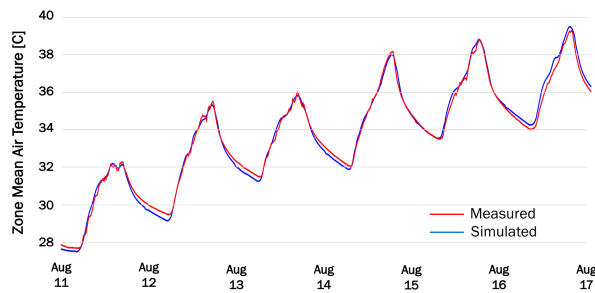

Figure 5 Whole-building electric consumption metrics for multiple simulation runs

Figure 7 Building volume weighted average temperature for 6-day "free oscillation" simulation with calibrated model; simulated data (blue) vs. measured data (red)

LESSONS LEARNED

This section discusses lessons learned, while the "Recommendations" section below lists specific recommendations for ensuring a high-quality calibration.

1. Data points for measured consumption and environmental conditions do not always map cleanly to EnergyPlus outputs.
2. Synchronization of time steps can be misinterpreted or inadvertently offset.
3. Calibrating during a year containing a leap day requires paying extra attention to EnergyPlus reporting.
4. Missing data for a sensor or gaps in coverage for specific building elements can introduce uncertainty.
5. It is important to understand the building's controls and how they impact other elements.
6. Look closely at time series profiles and not just the metrics they produce.
7. It is critical to properly model the building operation prior to the start of the time period being used for the analysis.
8. Time spent gaining a deeper understanding of the building is worth the effort to minimize both the required calibration effort and uncertainty.
9. Let the use case for the calibration inform the metrics used and the building elements that receive the most attention.
10. Exercise the building throughout its designed operational range.

RECOMMENDATIONS

Based on lessons learned, the authors recommend the following as best practices:

Sensor mapping: Carefully validate that there are not gaps or overlap in submetered sensors and that unit conversion between power and energy as well as si/ip units are correct for subhourly time steps.

Time step synchronization: It is easy to inadvertently offset time in the measured or simulated data—for example, misinterpreting the meaning of the time stamp in the measured data. Does data for time labeled 6 a.m. represent the 15 minutes ending at 6 a.m., the 15 minutes starting at 6 a.m., or the 15 minutes centered at 6 a.m.? In the case of EnergyPlus output, the time stamp is at the end of the time period.

Confirm whether the data are reported in the local time zone or Coordinated Universal Time and if daylight saving time has been used for any of the data.

Exclude the day that daylight saving time starts and ends from the calibration data set, as the time stamps in the measured data may be ambiguous.

Leap year: When calibrating time series data for a leap year after February 28, make sure simulation data are from the correct day. EnergyPlus may report February 29 as March 1.

Missing data or sensors: Address missing or invalid measured data by skipping that time step when calculating the calibration metrics. The multirun period time series OpenStudio calibration reporting measure will exclude any time steps that lack values for either simulated or measured data from metric calculations.

The authors did not have access to stair, plenum, or building surface temperatures but used careful analysis of the responses of surrounding zones to infer the expected conditions for the elements without sensors.

Building controls: Carefully manage HVAC system availability, setpoint managers, outdoor air, and other related elements such as thermostats and infiltration.

Time series profiles: The authors found multiple anomalies in the measured data that skewed calibration results. Although real buildings also experience anomalies, if the point of the calibration is to produce a model that is accurate for normal building operation, then anomalies should be excluded from the objective function during calibration—with two caveats. First, be careful that the anomalies observed are truly random and do not represent a systematic flaw in your underlying modeling assumptions. Second, for

transparency, calculate and report final calibration metrics both with and without the anomalies included.

Run period startup: When calibrating specific days, it is important to know or estimate what was happening with the building in the days leading up to the calibration period, specifically related to zone conditions. Building air temperatures may look good at the first time step of your calibration, but if the thermal mass is lagging far behind, it will be hard to get a good calibration for heat gain and loss.

Familiarity with the building: If you are not familiar with the building you are calibrating, learn about any special characteristics that might impact the calibration. For example, investigation revealed that there were four 200-gallon water tanks near windows in one zone that are part of a different HVAC system used in other experimental configurations. When possible, it is ideal to walk through the building or interview someone with first-hand knowledge of the building.

Calibration use case: If calibration will be done on data that have not yet been collected, think about what sensors you would like to have. In this case study, the authors identified the air temperature of the unconditioned stair and the first and second floor plenums as additional data points that would be helpful. If you are performing a calibration with a specific purpose in mind for the resulting model, more data collection effort might be spent for building components most important for your use case. A calibration setup might look different if you are considering envelope versus HVAC energy efficiency measures.

Operational range: If feasible, run the HVAC system in different configurations and environmental conditions to isolate the behavior of different building components and exercise the full range of the building and HVAC system. If a free oscillation test is done, ideally, it should be at a time with a high temperature difference between the initial inside conditions and the seasonal outside conditions. Of course, care needs to be taken so that the building conditions do not drift to a point that might cause problems or damage equipment by, for example, producing condensation.

CONCLUSION

Calibration of building energy models to time series data is significantly more difficult than calibration to monthly utility billing data, but careful attention to detail and the use of proven strategies can improve time series calibration outcomes. This paper has discussed the identification and exclusion of anomalous time periods when tuning calibration parameters, techniques for mitigating errors due to thermal transients at run period start, and a new OpenStudio measure for calculating calibration metrics across multiple run periods. The paper also provides lessons learned and recommends best practices, including proper alignment of measured and model time series and strategies for effective calibration parameter tuning. It is challenging but possible to produce a model that accurately predicts building performance at hourly or subhourly intervals using the techniques described in this paper.

ACKNOWLEDGMENTS

DISCLAIMER

REFERENCES

ASHRAE (American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc.). 2002. "Guideline 14-2002."

Bou-Saada, T.E., and J.S. Haberl. 1995. "An Improved Procedure for Developing Calibrated Hourly Simulation Models." *IBPSA-USA*.

Buckberry, H.L., and M. Bhandari. 2012. "ORNL MAXLAB Flexible Research Platforms Vision for a Novel Research Facility to Study Future Retrofit Options." In *ACEEE*, 24–35.

Calì, D., T. Osterhage, R. Streblow, and D. Müller. 2016. "Energy Performance Gap in Refurbished German Dwellings: Lesson Learned from a Field Test." *Energy Build.* 127(Supplement C): 1146–58.

Campolongo, F., J. Cariboni, and A. Saltelli. 2007. "An Effective Screening Design for Sensitivity Analysis of Large Models." *Environmental Modelling & Software* 22 (10): 1509–18.

Coakley, D., P. Raftery, and P. Molloy. 2012. "Calibration of Whole Building Energy Simulation Models: Detailed Case Study of a Naturally Ventilated Building Using Hourly Measured Data." In *First Building Simulation and Optimization Conference {BSO-2012}*, 57–64. Loughborough, UK.

Coakley, D., P. Raftery, and M. Keane. 2014. "A Review of Methods to Match Building Energy Simulation Models to Measured Data." *Renewable Sustainable Energy Rev.* 37: 123–41.

Erba, S., F. Causone, and R. Armani. 2017. "The Effect of Weather Datasets on Building Energy Simulation Outputs." *Energy Procedia* 134: 545–54. doi:10.1016/j.egypro.2017.09.561.

Haberl, J.S., and T. E. Bou-Saada. 1998. "Procedures for Calibrating Hourly Simulation Models to Measured Building Energy and Environmental Data." *J. Sol. Energy Eng.* 120(3): 193-204. doi:10.1115/1.2888069.

Harmer, L.C., and G.P. Henze. 2015. "Using Calibrated Energy Models for Building Commissioning and Load Prediction." *Energy Build.* 92. Elsevier B.V.: 204–15.

Herrando, M., D. Cambra, M. Navarro, L. de la Cruz, G. Millán, and I. Zabalza. 2016. "Energy Performance Certification of Faculty Buildings in Spain: The Gap between Estimated and Real Energy Consumption." *Energy Convers. Manage.* 125 (Supplement C): 141–53.

Im, P., M. Bhandari, and J. New. 2016. "Multi-Year Plan for Validation of EnergyPlus Multi-Zone HVAC System Modeling Using ORNL's Flexible Research Platform." Oak Ridge, TN.

Ji, Y., and P. Xu. 2015. "A Bottom-Up and Procedural Calibration Method for Building Energy Simulation Models Based on Hourly Electricity Submetering Data." *Energy* 93: 2337–50. https://doi.org/10.1016/j.energy.2015.10.109.

NOAA (National Oceanic and Atmospheric Administration). 2017. "Climate Data Online: Dataset Discovery." *Climate Data Online: Global Monthly Summaries.* https://www.ncdc.noaa.gov/cdo-web/datasets.

NRELa (National Renewable Energy Laboratory). "OpenStudio-Calibration-Measures." *GitHub.* https://github.com/NREL/OpenStudio Calibration-Measures.

NRELb. n.d. "About Measures." *OpenStudio User Docs.* https://nrel.github.io/OpenStudio-user-documentation/getting_started/about_measures/.

NRELc. n.d. "Weather Data for Simulation." https://energyplus.net/weather/simulation.

NRELd. n.d. "Parametric Analysis Tool (PAT) Interface Guide." *OpenStudio User Docs.* http://nrel.github.io/OpenStudio-user-documentation/reference/parametric_studies/.

NRELe (National Renewable Energy Laboratory). "OpenStudio-fault-models." *GitHub.* https://github.com/NREL/OpenStudio-fault-models/tree/master/pat_projects/

NRELf (National Renewable Energy Laboratory). *Technology Performance Exchange (TPEX).* Retrieved from NREL HVAC Lab: https://www.tpex.org/content/generic-rtu-5ton-172seer-13eer-1scroll2stage-wdehum-coil-interpolateddata

Ruiz, G.R., C.F. Bandera, T. Gómez-Acebo Temes, and A. Sánchez-Ostiz Gutierrez. 2016. "Genetic Algorithm for Building Envelope Calibration." *Appl. Energy* 168 (Supplement C): 691–705.

Soebarto, V.I. 1997. "Calibration of Hourly Energy Simulations Using Hourly Monitored Data and Monthly Utility Records for Two Case Study Buildings." *Proceedings of Building Simulation.* 97: 411-419.

Turner, C., and M. Frankel. 2008. "Energy Performance of LEED for New Construction Buildings." U.S. Green Building Council. New Buildings Institute, White Salmon, WA.