IMPROVING MODEL CALIBRATION METHODS: A CASE STUDY APPLICATION OF INCORPORATING IEQ WITH ENERGY

Nishesh Jain[1,2], Esfand Burman[1], Dejan Mumovic[1], Michael Davies[1]
[1] UCL Institute of Environmental Design and Engineering, University College London, UK
[2] DesignBuilder Software Limited, UK
Email: n.jain@ucl.ac.uk

ABSTRACT

Calibrated building performance simulation models are useful in operational performance assessments. The current measurement and verification (M&V) protocols provide the statistical criteria to check model accuracy. However, they do not describe the criteria for uncertainty or the accuracy of dependent parameters such as zone environmental outputs. Mathematically, meeting just the validation criteria in a highly parameterized model and an under-determined search space can lead to unrealistic solutions. This paper explores ways to improve the quality of calibrated models and suggests a multi-level checking framework. This is implemented on a case study building. The paper describes the current industry standard of energy use validation as the lowest level of calibration with higher levels requiring further validation of disaggregated energy use, whilst meeting the indoor environment quality (IEQ) parameter validation criteria.

INTRODUCTION

Calibration in the context of building simulation, is the process of fine-tuning the input parameters of a model to create a digital equivalent of a real building which can be used in operational performance assessments such as to evaluate energy conservation measures (ECMs), analyse the performance gap, and diagnose and optimise building services. Standard methods, defined in measurement and verification (M&V) protocols such as ASHRAE Guideline 14 (AG14) (ASHRAE, 2014), FEMP (Webster, et al., 2015) and IPMVP (EVO, 2016), suggest the use of calibrated simulation. The protocols define tolerances and requirements to create a calibrated model. However, they do not provide a detailed framework on how to develop and cross-validate these models. This paper assesses the use of calibration in M&V, evaluating the limitations and improvements needed. Then a multi-level calibration framework is proposed and applied on a case study to assess its performance gap, using zone air temperature for indoor environment quality (IEQ) cross-validation check.

BACKGROUND

Model calibration and its uses

Calibrated simulation models created for post-occupancy building performance diagnostics can provide valuable insights into performance issues with high level of confidence (Jain, et al., 2018). A model is said to be calibrated when the difference between the simulated results and the actual measurements is less than a predefined threshold, known as the validation criteria. These criteria are defined in the M&V protocols.

These protocols are primarily created for measuring and ascertaining performance using best practice techniques in implementation of ECMs. They provide guidance on data required for monitoring and measurement, measurement boundary, measurement period and ways to calculate impact and operational verification. Depending on the type of ECM and its relationship with other building performance input or output parameters, the protocols, mentioned earlier, provide four options for calculating ECM impact. While, two options look at isolating the analysis to the building systems affected by the ECM, the other options are applied to the whole building. The fourth option uses calibrated simulation method using energy modelling tools, calibrated to hourly or monthly energy use. A step-by-step evidence-based model fine-tuning method should be used by collecting detailed operational information during site surveys and by measurements (Reddy & Maor, 2006).

While these protocols are structured to calculate savings due to ECMs, the calibrated models can be used for other purposes, such as energy use monitoring or assessing the opportunities for performance improvement. Calibration can give insights into the operational inefficiencies and pinpoint underlying causes for the performance gap (Burman, 2016). Subsequently, a calibrated model could be used for quantifying impacts of the causes of the performance gap by reintroducing design assumptions. Figure 1 shows a typical calibration workflow.

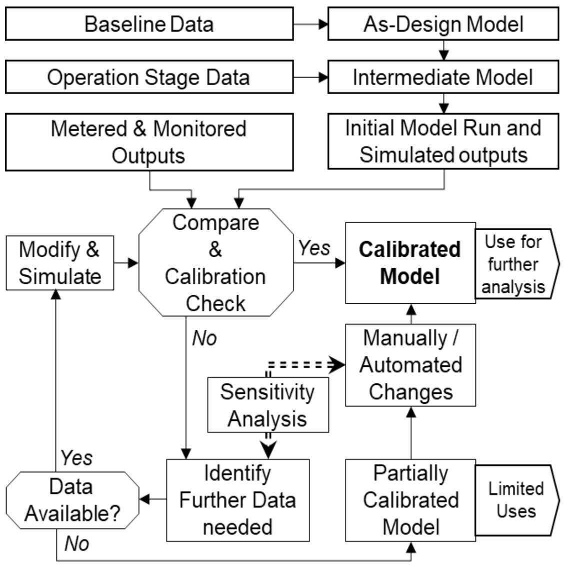


Figure 1: Step by step calibration workflow.

Calibrated model validation

Validation approaches in M&V protocols mainly focus on quantitative requirement for goodness of fit of the simulation model results to the actual data. Two statistical indices are used as the criteria for calibration, coefficient of variation of the root mean square error (Cv(RMSE)) and normalised mean bias error (NMBE). The tolerances for these as per AG14 are: 30% for C_v(RMSE), ±10% for NMBE for hourly data; and 15% C_v(RMSE), ±5% NMBE for monthly data. These values are also corroborated in the other protocols. According to the protocols, calibration should be done for one year with the minimum granularity of a month. But higher granularity (hourly or daily level) is preferred, especially in the cases when the calibrated model is used for a system and a sub-system level analysis (EVO, 2016).

Using the statistical criteria for calibration within the intended used of these M&V protocols, i.e. ECM evaluation, is expected to provide suitable results. However, relying merely on these metrics has certain limitations (Garrett & New, 2016), (Ruiz & Bandera, 2017). Using hourly data calibration is time consuming and monthly data might be too coarse. Also, as the criteria are deterministic statistical indices, they fail to capture that multiple solutions may exist that meet the criteria but may not necessarily reflect the real performance. For example, the gap in heating demand can be closed by increasing either the indoor set point temperatures or the mechanical ventilation supply or both. Therefore, based on limited data, even when AG14 criteria are met, it is not possible to deterministically identify the exact deviations in both these areas. Moreover, some of the solutions can be mathematically correct but physically impossible. In case the operation stage information is limited, then it is not possible to procedurally estimate, with certainty, the exact value of inputs to progress with calibration. At that stage modeller needs to rely on their own judgment. Defining weekly and daily level checks and cross-validation using secondary data streams such as disaggregated end uses, loads checks and zone set points (EVO, 2016) can help fix some of the unrealistic solutions issues. Further, incorporation of uncertainty based probabilistic approach can be useful in determining confidence levels in the validated model (BSI, 2008), (Jain, et al., 2018).

ADVANCED CALIBRATION CHECKS

To mitigate the limitations of using statistical checks, more data streams and analysis methods can be used. Annex C of AG14 discusses graphical data techniques, such as 24-h profile plots, box-and-whisker-plots and 3-D plots. These could be used to validate the calibrated model. Also, IPMVP recommends checking of building loads and energy use patterns, comparing measured and simulated data in form of bar charts, monthly percent difference time-series graphs and scatter plots. While they are all good techniques, in the absence of specific implementation guidelines, their practical use is limited. It is left to the modeller's judgment and expertise. This section reviews three advanced calibration checks.

1. Calibration for disaggregated data

Minimum data needed for any calibration is facility level energy use, for all fuels. However, calibrating for more data streams through disaggregated energy use can further improve the calibration accuracy and confidence in the calibrated model (Reddy, 2006). This information is increasingly available and can be taken during audits and short term and long term end use metering of energy data (Penna, et al., 2015).

Disaggregation compartmentalises the energy use, thereby reducing the chances of cross-compensation. Disaggregation can be done for different end-uses or spatially. Separating the energy uses can help in isolating interdependent aspects and undertaking a more granular analysis, where a dominant highly influential parameter for one end-use does not end up masking the influential parameters for other energy end-uses. As a minimum such disaggregation should include separating weather dependent loads such as heating and cooling from other occupant driven and non-weather dependent ones such as small power, and lighting (Soebarto, 1997).

Spatial disaggregation separates areas with distinct operating conditions and can provide insights into usage patterns and operations at a refined level. Disaggregation of hourly data in this manner can help in trends analysis, thereby giving an opportunity to create typical profiles. Spatial disaggregation should isolate different floors, zone types and tenancy. (Coakley, et al., 2014). Zones/

zone clusters with >500m^2 area and >25kW load should be considered to have separate meters (CEC, 2018).

It is better to have full year high resolution data for better calibration. However, if that is not possible due to practical limitations, then, short term intensive monitoring should be carried out where granular (hourly or finer) data is collected for typical weeks in different seasons to generate the typical profiles (Penna, et al., 2015). The criteria used for assessing calibration for these short-term periods can be based on the same statistical parameters but with finer acceptability ranges.

2. Cross-validation of other dependent results

Quality of calibrated models can be improved by cross-validating the simulation results with other dependent parameters. As suggested in IPMVP (EVO, 2016), verification of systems loads and zone level set points (e.g. temperature and humidity) can be checked. Besides this, IEQ parameters such as CO_2 concertation and peak and off-peak load profiles could also be used.

Zone temperatures and other IEQ parameters are easily available and can be a reliable data source. Monitoring of IEQ data streams can also provide evidence for detailed building operational profiles. Temperature data can help in finding the set-points used and CO_2 and $PM_{2.5}$ concentrations can help to check occupancy, ventilation and infiltration rates (Kapalo, 2013), (Parsons, 2014).

Incorporating cross-validation check needs requirements to be defined such as selection of the parameters, their measurement frequency and duration. Minimum check in this regards should be done for zone air temperatures as IEQ assessments are often largely based on room temperature distributions (Royapoor & Roskilly, 2015), (Roberti, et al., 2015). These checks should be done for representative rooms for all zone types, orientations, operational conditions, covering at least 10% of the floor area. Guidance regarding these are given in (BSI, 2019) and (BSI, 2012). The checks need to be done hourly, and if not for the whole year, they should cover at least two consecutive weeks for summer, winter and any other seasons. Checking of C_V(RMSE) and NMBE could be used, however the tolerance in AG14 might not suitable. For example, indoor temperature varies in a small range and AG14 acceptable error can result in large deviation from comfort bands. There are no standards or formal guidelines that define acceptance criteria for zone air temperatures calibration. Statistical indices that address absolute errors such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) was used by Roberti, et al. (2015). These are also used in climate research (Chai & Draxler, 2014). In most studies acceptable absolute error was in the range of ±1-2°C for most of the temperature data points (Booten & Tabares-Velasco, 2012; Royapoor & Roskilly, 2015; Ruiz, et al., 2016). MAE of 1°C and RMSE of 1.5°C can be used as targets for temperature calibration check.

3. Probabilistic calibration results

Due to the uncertainty in the inputs and when there is a lack of monitored data, numerous possible combinations can create a AG14 validated model. A probabilistic approach can be used in these cases (Jain, et al., 2018). BS EN 15603 (BSI, 2008) suggests to include the input uncertainty in energy simulations and provide energy use results probabilistically. Similar to that, in calibration, the observed deviations can be used for uncertain input parameters. The results can be presented with confidence bands around the data points. The aim of calibration is then to ensure that the measured value is within the upper and lower ranges of uncertainty. If required, the model can be fine-tuned to reduce the output uncertainty band. This methodology can be applied to the partially calibrated model in Figure 1 when the uncertain inputs are not significant for the outputs being analysed.

MULTI-LEVEL FRAMEWORK FOR VALIDATING A CALIBRATED MODEL

To make calibrated models versatile, a new framework is proposed based on the following principles:

1. Statistical tests provide the basis for accuracy.
2. Disaggregated checks reduce cross compensation.
3. Cross-validation of dependent variables ensures the correctness of inputs and the calculations used.
4. Probabilistic validation ascertains that uncertain, yet unknown, parameters can explain the residual gaps.

A three level calibration framework is described below and within each of the levels, four sub-levels define the accuracy requirement for four temporal resolutions.

Level 1: This base level calibration is done when only the building level data is available. Individual fuels should be calibrated separately for building energy use for one year, which is similar to the current protocols.

Level 2: When disaggregated metering (end-use and spatial energy use) is available, calibration for weather dependent loads (heating/cooling) should be separate from other loads. Also, floors and zones with sub meters (i.e. >500m^2 or >25kW), should be calibrated separately.

Level 3: This is the highest level, where dependent variable data such as zone IEQ parameters and loads are available. In this level, cross-validation should be done for at least zone space temperatures covering all zones with different activity or HVAC system, accounting for 10% of floor area. If not the whole year, this hourly temperature monitoring should cover two consecutive weeks for summer, winter and other transitional seasons.

Statistical compliance: Table 1 lists the statistical limits for different temporal resolutions for energy calibration. The weekly and daily values were interpolated between the monthly and hourly values in AG14. Limits for hourly temperature checks have also been defined as per the recommendations in the literature reviewed earlier.

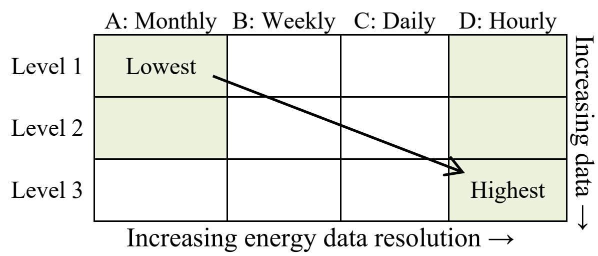
Table 1: Validation criteria

Energy	Monthly	Weekly	Daily	Hourly
Cv(RMSE) %	15	15	18	30
NMBE %	±5	±5	±6	±10

Temperature	MAE	RMSE
Hourly	1°C	1.5°C

Depending on the model's use and data available, a particular level and criteria can be used Table 2 shows the 'calibration level-calibration resolution' matrix. The lowest and highest levels are 1A and 3D respectively.

Table 2: Level-resolution matrix



**Green highlight: Levels used in the case study in this paper*

However, in cases where, due to lack data, the statistical validation criteria cannot be met, probabilistic validation method can be used. When using the probabilistic method, the uncertainty of parameters that explain the deviation have to be clearly defined. This needs to be supplemented by the modeller's explanation on how this partially calibrated model is fit for purpose and the uncertainties do not cause a conflict with the modelling results during the intended use of the calibrated model.

CASE STUDY APPLICATION

The case study is a ~4500 m^2 university office building located in Central London. Built in 1900's, the building had a major refurbishment in 2010. Having a basement, ground and six upper floors, the main zones in the building are open plan and cellular offices, meeting rooms, computer clusters and a library.

Building characteristics

Building Design: The concrete and brick wall structure, during the refurbishment, was made energy efficient. The external walls are fitted with composite insulation board and a secondary double glazing is installed on the inside of existing single glazed windows.

Heating and cooling: Heating and cooling is primarily supplied by a variable refrigerant flow (VRF) system with heat recovery. There are roof-mounted outdoor units, and indoor spaces have evaporators of differing capacities with a central control panels for adjusting and resetting the set-points. Hot water is provided through electric heaters. Also, heating in the circulation area is provided by two condensing boilers supplied through radiators. However, it uses very little energy and as high-resolution data is not available for gas use, this has not been assessed in this study.

Ventilation: The building is mainly naturally ventilated through operable windows. However, some areas in the basement and ground floor are served with pre-conditioned air through an air-handling unit (AHU).

Small power and lighting: Typical office equipment load has a diversified pattern of use as its occupancy is linked to university term times. For office lighting, recessed luminaries with T5 lamps are controlled by absence detection sensors. LEDs and CFLs are used in circulation areas, which are controlled by occupancy sensors.

Renewables: A 3.42 kW$_e$ rooftop photovoltaic system is installed and separately monitored to determine the exact amount of electricity generation.

Baseline simulation model: An initial energy model (Figure 2) was developed in DesignBuilder Software, a graphical user interface for EnergyPlus, based on design information, specifications and building audits.

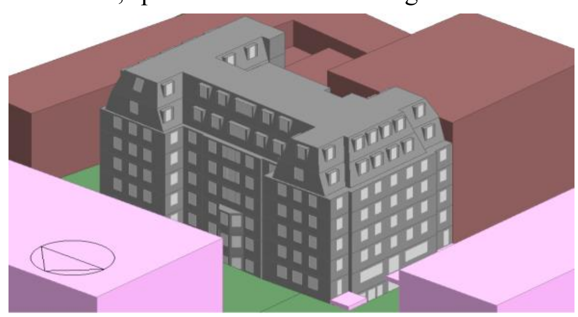


Figure 2: Case study building simulation mode.

Metering and monitoring: Monitoring for the building was carried out from Aug 2016 to July 2017. Most of the energy used is electricity. Mains half hourly electricity use data is available from building meters and the utility supplier. Also, hourly sub-metered data is available with spatial and end-use disaggregation. Space conditioning, heating + cooling (H+C), electricity use was recorded separately for the whole building. However due to the use of a VRF system, which can heat and cool different parts of the building simultaneously, their individual use was inseparable. Sub-meters at each floor recorded lighting + small power (L+P) electricity use, including small amount of domestic hot water energy use. L+P were not possible to be further disaggregated, except on a couple of floors where they were metered separately. Server and Lifts were metered separately. Zone temperature was recorded at 5-min intervals in the offices, meeting rooms and study spaces covering ~15% of the regularly occupied area.

Building Energy Performance

Figure 3 shows the monthly energy use disaggregated for H+C, L+P, server and lifts. H+C and L+P are the dominant energy end uses in the building.

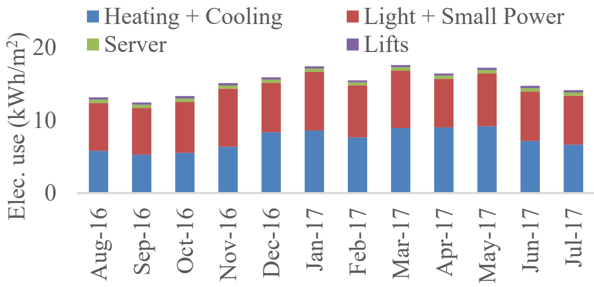


Figure 3: Monthly electricity use for various end-uses

The patterns of actual use are more evident in more granular data analysis. Figure 4 shows heat maps for building level L+P and H+C electricity use for the summer (August) and the winter (January) months.

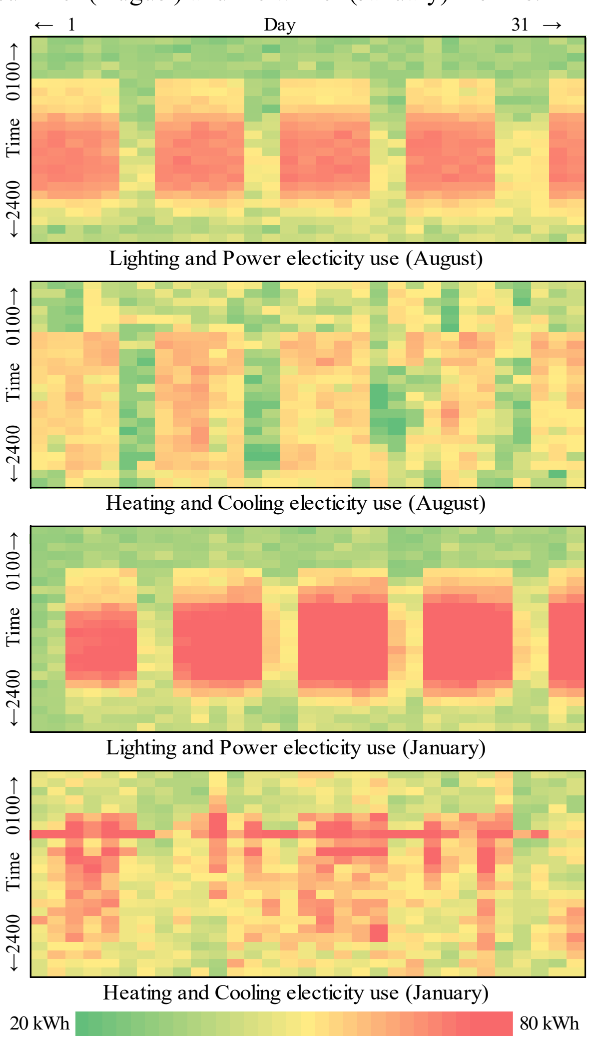


Figure 4: Heat map for building level electricity use

Typical weekday and weekend patterns can be seen from this data. These patterns can give insights into building operation. However, as the focus of this paper is calibration and validation framework, for the sake of brevity they have not been explicitly discussed.

Performance deviation and model calibration

As per site information and the interactions with facility managers, issues related to systems and operations were estimated. Following the framework in Figure 1, the baseline model was tuned for issues such as system faults and modifications to typical operations The various levels of calibration done are shown in the next sections.

Level 1A & 1D: Calibration of building level meters

In this model fine-tuning was done as per the observed modifications in parameters listed in Table 3. Figure 5 shows the results (C_V(RMSE) = 8.36; NMBE = -2.63).

Table 3: Fine-tuning for Level 1A calibration

Parameter	**Changes made and their reasons**
Weather data	Nearby station data from DesignBuilder Climate Analytics (DBS, 2019)
Temperature set points	Occupied: 23°C; Unoccupied: 21°C; as per the averages from BMS logs
Occupancy schedule	Measured occupancy varied monthly. It was generally higher than that assumed for out of hours use and weekends.
VRF System operations	Systems were operated throughout the day and night, even during unoccupied times (see figure 4)
Base load	~50% L+P load during unoccupied time.
AHU in basement	AHU wasn't working, leading to more opening of windows in the basement

Figure 5: Level 1A calibration results

This model cross-compensated the daily and hourly variations. Level 1D model for hourly building results, could have been created by tuning inputs at a higher granularity. However, it was challenging as H+C, L+P were not metered separately. In this case, as sub-metered data was available, Level 2 calibration was done directly.

Level 2A & 2D: Calibration at sub-metered level

For different end-uses, i.e. weather driven H+C and occupancy driven L+P, Level 2A and 2D models were created. Figure 6 shows the actual vs simulated scatter plot for building level energy use and Table 4 lists the detailed calibration results. The end-use sub meters were calibrated at the building level with further floor level calibration of L+P. Table 5 describes the further tuning done in the Level 1A model to get to Level 2.

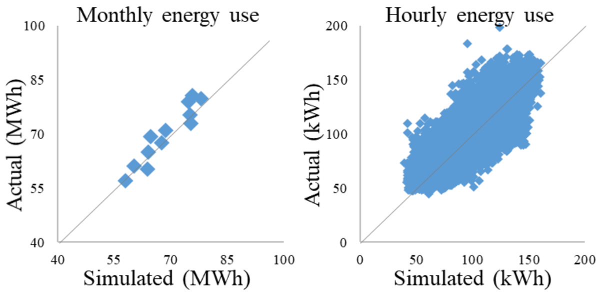


Figure 6: Scatter plot for total building total energy use for monthly (2A) and hourly (2D) calibration

Table 4: Level 2 calibration (C_V(RMSE) % / NMBE %)

End-use wise	**Monthly 2A**	**Hourly 2D**
H+C	8.68/3.47	28.54/3.47
L+P	8.84/-4.04	18.17/-4.04
Server	5.06/4.75	13.05/4.75
Lifts	12.15/4.23	24.87/4.23
Building Total:	**9.12/-2.08**	**23.22/-2.08**

Floor wise L+P	**Monthly 2A**	**Hourly 2D**
Basement	7.37/-3.82	24.31/-3.82
Ground	18.02/-8.09	32.11/-8.09
First	9.36/-3.83	24.84/-3.83
Second	7.34/-4.05	21.45/-4.05
Third	8.53/-4.19	29.45/-4.19
Fourth	6.95/-1.22	27.30/-1.22
Fifth	18.95/-3.55	39.43/-3.55
Sixth	10.94/-4.99	38.39/-4.99
Building L+P:	**8.84/-4.04**	**18.17/-4.04**
*Green: Table 1 criteria met; Red: Criteria not met.		

Table 5: Further fine-tuning for Level 2 calibration

Parameter	**Changes made and their reasons**
Occupancy schedule	Using floor level L+P use data, bespoke schedules for wkday, wkend & holiday were made for each floors.
Temperature set points	Set points were varied monthly; 20/24°C when occupied and 19/22°C when unoccupied as per the BMS logs and monthly averages from typical zones
Small Pwr & lighting load and schedule	Adjustments for peak load and diversity to address the monthly variation and term dates. Baseline loads changed to lighting: 60% and power: 35%

Specifically, for Level 2D models, L+P use trends were helpful in determining typical profiles. Seen in Figure 4, there is a marked difference in L+P use during occupied and unoccupied times. Floor L+P use data helped to identify typical occupancy for various zone types. The floor used by administrative staff had a more regular occupancy compared to the floors used by research students, which had a diversified usage. In Table 4, the floor level calibrations for L+P on the ground (which has the library), fifth and sixth floors were outside the C_V(RMSE) limit, suggesting cross-compensation for L+P use across the floors. Therefore, typical-day trends can be used but they can mask some of the behaviour in irregularly used spaces. These can only be analysed by using more granular trends, instead of typical averages.

Level 3: Dependent parameters check

The Level 2D model was cross validated by checking the air temperature in different space types. These checks showed variations in temperatures in the monitored and the simulated data, especially, in holidays and weekends. Using granular set-point temperature schedules for the monitored spaces, varying seasonally and by day-of-the-week, the temperature variations were reduced to meet the criteria in Table 1. This also made the H+C meter calibration more accurate. Table 6 gives the temperature calibration statistics over the year for some of the spaces.

Table 6: Level 3 Temperature calibration check

Zone	**MAE (°C)**	**RMSE (°C)**
Open Office 3F	1.29	1.27
Open Office 4F	0.76	1.45
Cellular Office 4F	1.09	1.61
*Green: Table 1 criteria met; Red: Criteria not met.		

The simulated temperatures followed the monitored data well except some weekend and overnight variations. They could be firther fine tuned with finer control over setpoints, however, the deviations are not huge over longer periods. The variation across the sample spaces is captured in Figure 7 which shows the spread of the error.

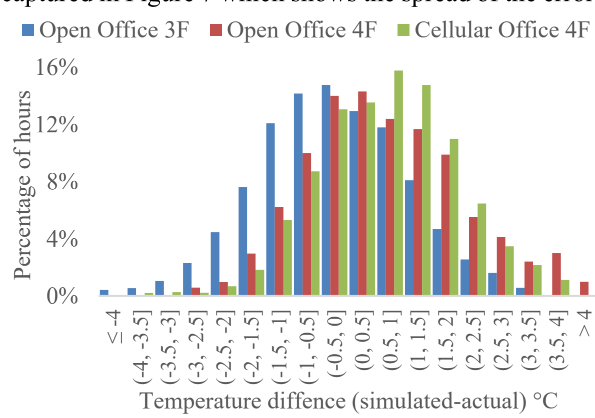


Figure 7: Histogram of hourly temperature difference

This temperature based fine tuning is useful when there is a need for high zone level accuracy e.g. in zone load and systems assessment or for comfort checking.

Uncertainty checks

Uncertainty-based validation was done to establish the robustness of the model. Remaining uncertainty range (upper and lower) was defined for key inputs to create the best and worst scenarios. Table 7 lists the inputs and Figure 8 shows the spread the best and worst scenarios along with the actual data. The figure suggests that the uncertainty in the electricity use is not very wide, 20% variation (+5% and -15% for upper and lower limits).

Detailed monitoring evidence can be obtained to reduce the deviations. Also, the probability across the range values will more likely have a normal distribution rather than a uniform one i.e. the upper and lower extremes of the range are less likely to occur than the central values A probabilistic uncertainty assessment could be done to determine the more likely scenarios instead of extreme ones but it is not within the scope of this paper.

Table 7: Uncertain areas and their ranges

Input	Range	Reason
Zone Temperature set point	±1.5°C in the Level 3 model	Average error in Level 3 temperature checks (heating vs cooling dead band of 2°C)
Occupancy hrs & related L+P use duration	12 hrs to 16 hrs	Additional low level occupancy during morning and evening hours based on site spot observations
U-Value	±10%	Poor documentation of changes during refurbishment
Equipment load	±20%	Standard variation observed in occupancy

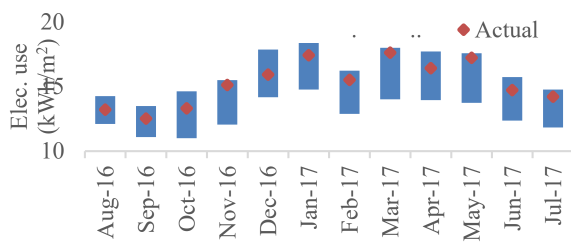


Figure 8: Calibration model uncertainty result check.

DISCUSSION

Learnings from case-study data: The case study issues investigated were centred around operation issues and maintenance issues and systems inefficiencies. Level 2D calibration, using end-use data, helped to separate the space conditioning loads from other loads and to identify use patterns and operational behaviours. Also, Level 3 temperature checks verified that the system, running overnight, had high set points even during unoccupied times. However, there were some limitations. As the heating and cooling were provided by a VRF system, it was not possible to isolate the space heating from cooling energy. This is important in London's climate were summers are mild and cooling load can be driven by internal gains, especially in an office building. This can often lead to instances of simultaneous heating and cooling provided in different zones. Also, building envelope construction was not documented well during the refurbishment, therefore the issues and uncertainties regarding material properties could not be assessed in this case. However, it can be seen that using the three-level framework proposed in this paper, the residual uncertainty seen in the overall results was not very high.

Using granular disaggregated data: Level of end-use disaggregation in the case study was limited to H+C and L+P. This level of end-use disaggregation, on its own, is not fine enough to understand all the underlying behaviour of the building. However, with spatial disaggregation available for each floor, the overall sub-metered data was very insightful. Some of the diversity arising from space type and nature of occupants was clarified when a more granular analysis was performed for each floor. Similarly, night running of the HVAC systems was also ascertained through analysing the disaggregated data. These findings at the granular level might be unique to that year, and more data is required to ensure that the observed deviations are a regular occurrence than a passing event.

Importance of temporal granularity: Data granularity can affect the accuracy of the model. If the model is calibrated in a lower resolution (monthly), it may mask some of the underlying issues which can only be uncovered at finer timescales. However, calibration at finer timescales need large amount of well sub-metered data, and a significant effort from the modeller to get useful results. Therefore, a correct balance is needed between temporal resolution, data availability, accuracy, effort and the intended use of the calibrated model.

Probabilistic validation approach: Uncertainty-based calibration can provide a way to deal with issues of data availability and granularity. It provides an alternate validation method of a partially calibrated model. Also, as shown in this case study, this can also be used to reinforce confidence in a calibrated model, ensuring that even with input uncertainty the measured data is within a reasonable output range. Additionally, a probabilistic approach can be used to resolve the limitation of using just the statistical criteria as they fail to address the fact that multiple solutions may exist that do not necessarily reflect the real performance. However, this requires a lot more data and computational time to model all scenarios compared to statistical validation.

CONCLUSIONS

Limitations of validation practices can be improved by detailing and enhancing the scope of current standards and guidelines. Current M&V practices need to account for enhanced sub-metering and incorporation of IEQ cross validation checks for model calibration. This can be structured in the form of pre-requisites for various levels of calibration for each temporal resolution. The new framework proposed in the paper keeps the essence of existing protocols as the base level of calibration. Depending on the intended use of the model, higher calibration levels can be achieved by adding more checks tailored to disaggregated end-uses whilst meeting the minimum validation requirement for IEQ parameters (e.g. data streams, monitoring duration and frequency and percentages of spaces to be covered). A probabilistic approach can be used as an alternative validation process to overcome some of the issues with data availability and granularity and improve the calibrated model's usability.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support by UCL Overseas Research Scholarships (UCL-ORS), DesignBuilder Software Ltd., Innovate UK KTP project (Partnership number:11615) and The 'Total Performance' of Low Carbon Buildings in China and the UK ('TOP') project funded by EPSRC (EP/N009703/1).

REFERENCES

ASHRAE, 2014. *Measurement of energy, demand and water savings - Guidelines 14,* Atlanta: ASHRAE.

Booten, C. & Tabares-Velasco, P. C., 2012. *Using EnergyPlus to Simulate the Dynamic Response of a Residential Building to Advanced Cooling Strategies.* Boulder, Colorado, 2nd International Conference on Building Energy and Environment.

BSI, 2008. *BS EN 15603 Energy performance of buildings.,* London: BSI.

BSI, 2012. *BS EN 12599:2012 Ventilation for buildings.* London: BSI.

BSI, 2019. *BS EN 16798-1:2019 Energy performance of buildings: Ventilation for buildings.,* London: BSI.

Burman, E., 2016. *Assessing the Operational Performance of Educational Buildings against Design Expectations – A Case Study Approach. Doctoral Thesis,* London, UK: University College London.

CEC, 2018. *2019 Building Energy Efficiency Standards for Residential and Nonresidential Buildings,* Sacramento, CA: California Energy Commission.

Chai, T. & Draxler, R. R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development,* Volume 7, pp. 1247-1250.

Coakley, D., Raftery, P. & Keane, M., 2014. A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews,* Volume 37, pp. 123-141.

DBS, 2019. *Climate Analytics.* [Online] Available at: https://analytics.designbuilder.co.uk/ [Accessed 12 11 2019].

EVO, 2016. *Core Concepts: International Performance Measurement and Verification Protocol,* Washington, DC: Efficiency Valuation Organization.

Garrett, A. & New, J., 2016. Suitability of ASHRAE guideline 14 metrics for calibration. *ASHRAE Transactions,* 122(1), p. 469+.

Jain, N. et al., 2018. *Comparative Analysis of Protocols Used in Measurement and Verification of Energy Performance.* Cambridge, UK, Proceedings of 4th Building Simulation and Optimization Conference.

Kapalo, P., 2013. *Analysis of ventilation rate and concentration of carbon dioxide in the office.* s.l.:Technical University of Kosice, Slovakia.

Parsons, P., 2014. Determining Infiltration Rates and Predicting Building Occupancy Using CO2 Concentration Curves. *Journal of Energy,* Volume 2014, p. 6 pages.

Penna, P. et al., 2015. Multi-stage calibration of the simulation model of a school building through short-term monitoring. *Journal of Information Technology in Construction,* Volume 20, pp. 132-145.

Reddy, T. A., 2006. Literature review on calibration of building energy simulation programs: uses, problems, procedure, uncertainty, and tools. *ASHRAE Transactions,* Volume 112 (Part1), pp. 226-240.

Reddy, T. A. & Maor, I., 2006. *Procedures for Reconciling Computer-Calculated Results With Measured Energy Data ASHRAE Research Project 1051- RP,* Philadelphia: Drexel University.

Roberti, F., Oberegger, U. F. & Gasparella, A., 2015. Calibrating historic building energy models to hourly indoor air and surface temperatures: Methodology and case study. *Energy and Buildings,* 108(1), pp. 236-243.

Royapoor, M. & Roskilly, T., 2015. Building model calibration using energy and environmental data. *Energy and Buildings,* 94(1), pp. 109-120.

Ruiz, G. R. & Bandera, C. F., 2017. Validation of Calibrated Energy Models: Common Errors. *Energies,* 10(10), p. 1587.

Ruiz, G. R., Bandera, C. F., Temes, T. G.-A. & Gutierrez, A. S.-O., 2016. Genetic algorithm for building envelope calibration. *Applied Energy,* Volume 168, pp. 691-705.

Soebarto, V., 1997. *Calibration of hourly energy simulations using hourly monitored data and monthly utility records for two case study buildings.* Prague, Proceedings of the 5th International Building Performance Simulation Association Conference.

Webster, L. et al., 2015. *Measurement and Verification for Performance-Based Contracts Version 4.0,* Washington, DC: U.S. Department of Energy.