2022

Building Energy Model Calibration with Functional Inputs and Outputs for Performance Monitoring

Thomas Cerbelaud

Bruno Duplessis

Pascal Stabat

Riad Ziour

Building Energy Model Calibration With Functional Inputs And Outputs For Performance Monitoring

Thomas CERBELAUD[1*, 2], Bruno DUPLESSIS[2], Pascal STABAT[2], Riad ZIOUR[1]

[1] Openergy R&D,
4 Rue Dolorès Ibarruri, 93100 Montreuil, France
thomas.cerbelaud@egis.fr, riad.ziour@egis.fr

[2] Mines Paris, PSL Research University, CES - Centre for Energy Efficiency of Systems,
60 bd Saint Michel, 75006 Paris, France
bruno.duplessis@minesparis.psl.eu, pascal.stabat@minesparis.psl.eu

* Corresponding Author

ABSTRACT

Building continuous performance monitoring is becoming a cornerstone in ensuring energy efficiency and sobriety of existing, retrofitted and newly built buildings. Although it may help convince investors in energy efficiency projects or bridge the gap between expected and actual performance, continuous monitoring - sometimes referred to as "Advanced M&V" or continuous commissioning – is still the exception rather than the rule. Recent efforts to continuously characterize building performance usually rely on building-level analyses: previous works include leveraging a Building Energy Model (BEM), monthly calibrated on building heating, ventilation and lighting consumption using real weather data and fine grain occupancy data, for daily monitoring. While BEM calibration against sub-daily frequency data has been increasingly studied in recent years, it is, to our knowledge, seldom used for building continuous monitoring. It is, however, particularly tailored for this task, to the extent it extracts embedded physics within the BEM into actionable insights for fault detection and diagnosis.

Fine grain calibration of BEM faces a number of challenges in the recent literature, among which are (i) accounting for time varying dependent functional inputs - e.g. electric equipment and lighting energy consumption altogether with building occupancy - but for sensor data in the calibration algorithm, and (ii) treating functional outputs as functional stochastic variables when comparing simulation outputs with real data. Our contribution is to enhance building-level performance monitoring by introducing a stochastic model inversion scheme, also referred to as stochastic calibration, to support robust preventive fault detection and diagnosis. Our approach extends the current state-of-the-art on Bayesian calibration of BEM by accounting for dependent functional inputs and outputs in both selecting the most influential parameters and calibrating the model, and deals with uncertainties in functional inputs such as daily profiles of lighting and electric equipment energy consumption. This methodology is illustrated against a medium-size real secondary school building, located in Rennes, France, and equipped with an Advanced Meter Infrastructure (AMI) with hundreds of sensors. A comparison between a classic calibration process and the described methodology is presented and the benefits of accounting for the functional nature of the inputs and outputs in both the Design of Experiment (DoE) and the calibration process are illustrated against this case study.

1. INTRODUCTION

The advent of Advanced Metering Infrastructure (AMI) and cloud computing is paving the way towards more affordable advanced building performance monitoring, notably reducing the cost of data acquisition (Granderson et al., 2016). Besides, Touzani et al. (2019) highlighted the necessity of accounting for non-routine event (NRE) in comparing metered consumption to a baseline, e.g. within an IPMVP-based process. It is only a short step from there to Automated Fault Detection and Diagnosis (AFDD) methods, where one is interested in monitoring and preventing performance drifts in an automated and continuous manner. FDD-related research has been quite active in the past years: on one hand, it mostly focuses on heavily instrumented commercial buildings and data-driven (Frank et al., 2019) approaches to precisely predict specific faults such as duct fouling - Frank et al. (2016) also tried using a hybrid or data-driven approach with a minimal set of metered data to diagnose specific faults; on the other hand, Jradi et al. (2018) used a

model-based approach, where an EnergyPlus (EnergyPlus, 2019) model is dynamically calibrated by passing real time metered data to the simulation, and compares simulated and metered data to help detecting and diagnosing drifts in performance. However, while BEM calibration against sub-daily frequency data has been increasingly studied in recent years (Chong et al., 2017), it is, to our knowledge, seldom used for building continuous monitoring without heavy data metering to account for dynamic variables such as occupancy. It is, however, particularly tailored for this task, to the extent it extracts embedded physics within the BEM into actionable insights for fault detection and diagnosis.

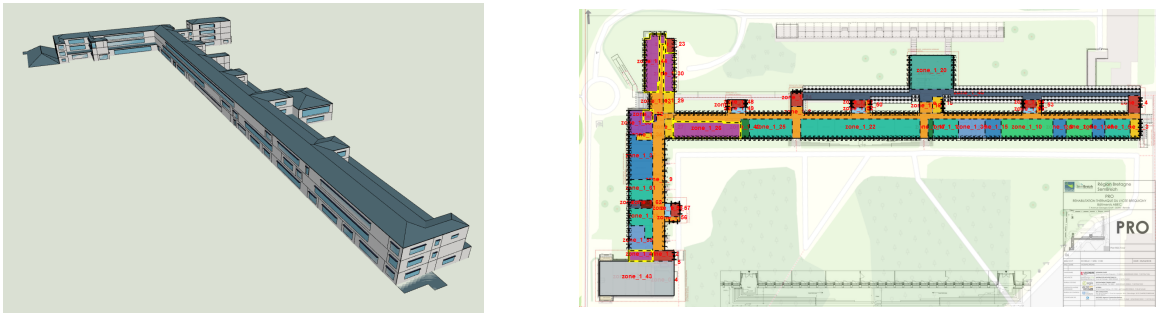
In this paper, we propose to extend the current state-of-the-art calibration process to account for dynamic variables, so that building-level performance monitoring algorithms may be developed for building equipped with a small number of sensors not necessarily accounting for dynamic variables such as occupancy. The following sections are organized as follows. Section 2 describes the case study used to illustrates our methodology. Section 3 digs into the technical details of how the calibration process is extended to account for functional and dynamic inputs and outputs, and defines the baseline to be compared with. Eventually, section 4 compares how the exposed methodology improves the baseline.

2. CASE STUDY OVERVIEW

The case study is a medium-size real secondary school building, *Lycée Brequigny*[1], located in Rennes, France, and equipped with an Advanced Meter Infrastructure (AMI) with hundreds of sensors. For the sake of clarity and simplicity, we only describe the building main features, and tested our methodology on the heating energy consumption only. The BEM model used in this work is built using EnergyPlus (EnergyPlus, 2019).

2.1 Building's description and thermal zoning

The building is a recently retrofitted educational building, comprised mostly of classrooms and offices, and is one of three buildings that together form the *Lycée Brequigny*. The building total floor area is 9426 square meters. 3D view and thermal zoning for the ground floor are displayed in Figures 1a and 1b.



(a) 3D view of *Lycée Brequigny*

(b) Ground floor thermal zoning

Figure 1: Thermal zoning of the case study

The building was externally insulated in 2020, and is heated by hot water radiators connected to four district heating substations. The inlet temperature of the secondary loop is regulated on the outdoor dry bulb temperature.

The study is particularly focused on the parameterization of occupancy, ventilation, electric equipment, lights and indoor temperature, which impact the heating consumption. As the retrofit came with an Energy Performance Contract (EPC), a BEM was built during the design stage of the retrofit. Therefore, the BEM built in our study took over all the parameters from the previous model, and updated them according to the building as built[2]. Part of these parameters will be calibrated further down the road, except for the - measured - indoor temperature that is taken as a boundary condition used during calibration.

Heating energy is delivered to the building by a heat exchanger connected to a heating district and that feeds hot water to the radiators throughout the building. Classrooms are equipped with motion detectors and temperature sensors that control the hot water radiator terminal regulation and the mechanical ventilation flow rate based on pre-configured

[1] It was renovated in 2020 and was awarded several national and international awards.
[2] as opposed to the building as designed.

values. All other rooms are regulated using thermostatic valves and pre-configured ventilation schedules except for some teachers rooms that are ventilated according to measured CO_2 levels.

2.2 Advanced Meter Infrastructure (AMI) overview

Thousands of data points are collected through the Building Management System (BMS). The site is operational since 2019, and the sensor data from November 2020 to March 2022 are used in this study. For the purpose of the demonstration, the ten first days of December 2020 are extracted from the data set. Because of COVID-19, the site regulation and occupancy were particularly disturbed, in that the variability of the occupancy and ventilation flow rate was high - in particular the overall occupancy was exceptionally low -, and the classrooms were over-ventilated (natural ventilation) during this time period to prevent the spread of the virus. Only a small subset of the AMI data was used, and is described in Table 1.

Table 1: Data used from the AMI to illustrate the methodology

Data	# sensors	Use in the proposed methodology
Indoor temperatures	71	The 71 time series are averaged out and used as observed input data in the calibration process.
Heating energy consumption	3	The consumption of three substations out of four are aggregated altogether. The fourth substation, feeding the conference hall, is voluntarily left out.

Besides sensor data, dry bulb temperature and horizontal irradiance are also used as inputs to the model calibration procedure. Eventually, a categorical temporal variable is used to indicate whether or not, for each day, the heating system is turned on.

The high variability in building uses makes it even more critical to account for these changes during monitoring in order to properly detect and diagnose potential fault during building operation.

3. METHODOLOGY

While the methodology relies on a well-known tried and tested succession of steps[3] (Kennedy and O'Hagan, 2001; Heo et al., 2012; Chong et al., 2017; Monari, 2016), our contribution dwells in bringing to BEMs calibration new developments in stochastic model inversion so (i) functional variables are accounted for (Nanty, 2015; Nanty et al., 2016) and (ii) the process is compatible with building performance monitoring. In addition, we emphasise on the choices made for use within a building performance monitoring workflow. This section describes the methodology laid out to achieve BEM calibration with functional inputs and outputs. The methodology breaks down as follows: (i) first a Design of Experiments (DoE) is generated; (ii) then a surrogate model is trained on the simulation in order to perform a Global Sensitivity Analysis (GSA) to select a relevant subset of features; (iii) and eventually a new surrogate is trained on the reduced subset and (iv) Bayesian calibration is performed using the emulator. This section ends up with describing the benchmark procedure of our methodology.

Table 2 displays the parameters that were chosen to conduct this work. $\mathcal{N}_{[a,b]}(\mu,\sigma)$ is the truncated normal distribution on the interval $[a,b]$ with mean μ and standard-deviation σ, and $\mathcal{U}_{[a,b]}$ is the uniform distribution on the interval $[a,b]$.

3.1 Modeling input and output functional variables

To account for functional input variables, the methodology leverages Functional Data Analysis (FDA) algorithms to decompose multivariate time series data into a reduced basis of functions. When applied to a single day profile, the decomposition allows to find the functional basis $(\varphi_i)_{i=1..k}$ and mean function e such that

[3] In the literature, while the calibration processes may slightly differ from one work to another, they basically all follow these three steps: (i) reduce the number of parameters to calibrate, (ii) optionally define a surrogate model (to train before or at calibration time) and (iii) explore posterior distribution of unknown parameters

[4] Only classes are regulated with a setback and comfort setpoint

$$f_j \approx e + \sum_{i=1}^{k} a_{i,j}\varphi_i,$$

That minimizes

$$\sum_{j=1}^{n} \int_I \left(f_j(t) - e(t) - \sum_{i=1}^{k} a_{i,j}\varphi_i(t) \right)^2 dt$$

In this work the functional decomposition is extended to multiple and dependent functions, using the Simultaneous functional Principal Component Analysis (SPCA) algorithm (Ramsay et al., 2005), consequently representing all dependent functions in Table 2 with a small number of scalars $\alpha = (a_1 \dots a_k)$, that will be further used in the BEM calibration process.

Besides, for the methodology to be compatible with continuous monitoring, all functional metered data - inputs and outputs, e.g. dry bulb temperatures or hourly heating energy consumption - were segmented into hourly time series of one day.

As a consequence, a datum in the training and testing sets is no longer a simulation run period long, as it would have been following Higdon et al. (2008a), or several months long, as suggested in Monari (2016), but a day long, bringing the number of samples in the training set up to 365 times the number of simulations. Therefore, the number of training

Table 2: Model parameters description and distributions.

Symbol	Parameter description	Distribution
Scalar parameters		
υ_{occ}	Multiplicative factor of the occupancy nominal value	$\mathcal{N}_{[0.3,1.6]}(1.0, 0.5)$
υ_{vent}	Multiplicative factor of the ventilation nominal value	$\mathcal{N}_{[0.4,0.6]}(0.5, 0.2)$
$\upsilon_{natvent}$	Multiplicative factor of the natural ventilation nominal value	$\mathcal{U}_{[0.5,1.5]}$
υ_{li}	Multiplicative factor of the lights nominal value	$\mathcal{N}_{[0.6,1.4]}(1.0, 0.2)$
υ_{eq}	Multiplicative factor of the equipments nominal value	$\mathcal{N}_{[0.6,1.4]}(1.0, 0.2)$
b_{vent}	Multiplicative factor of the ventilation baseload	$\mathcal{N}_{[0.6,1.4]}(1.0, 0.2)$
b_{li}	Multiplicative factor of the lights baseload	$\mathcal{N}_{[0.6,1.4]}(1.0, 0.2)$
b_{eq}	Multiplicative factor of the equipments baseload	$\mathcal{N}_{[0.6,1.4]}(1.0, 0.2)$
λ	Multiuplcative factor for all opaque surface conductivities	$\mathcal{N}_{[0.8,1.2]}(1.0, 0.1)$
U_w	Multiplicative factor for all windows U-value	$\mathcal{N}_{[0.8,1.2]}(1.0, 0.1)$
S_w	Multiplicative factor for all windows solar heat gain	$\mathcal{N}_{[0.8,1.2]}(1.0, 0.1)$
η	Efficiency of the heat exchanger of the heating district	$\mathcal{N}_{[0.7,1.0]}(0.9, 0.1)$
q_{inf}	Infiltration flow rate under 4Pa	$\mathcal{U}_{[0.6,1.4]}$
t_c	Indoor comfort temperature offset from nominal value	$\mathcal{N}_{[-2.0,3.0]}(0.0, 0.5)$
t_s	Indoor setback temperature offset from nominal value for classrooms	$\mathcal{U}_{[-3.0,0.0]}$
s_{blinds}	Threshold of direct solar radiation for closing / opening the blinds	$\mathcal{U}_{[100.0,800.0]}$
s_{li}	Threshold of illuminance for turning on/off the lights	$\mathcal{U}_{[200.0,700.0]}$
h_s	Time offset from system nominal start time in hours	$\mathcal{U}_{[-2.0,1.0]}$
h_e	Time offset from system nominal shut down time in hours	$\mathcal{U}_{[-1.0,2.0]}$
Functional parameters		
$f_{offices}$	Normalized day profile used for offices	GMM on the coefficient of the functional decomposition of randomly generated profiles
f_{class}	Normalized day profile used for classrooms	
f_{temp}	Day profile of temperature setpoint in classes[4]	

points in the simulation data set grows larger than 10,000 data points for several hundreds of simulations, making the training of Gaussian Processes (GP) (Kennedy and O'Hagan, 2001; Rasmussen and Williams, 2005; Higdon et al., 2008a; Chong et al., 2017) unfeasible. A Sparse Variational Gaussian Process (Titsias, 2009; Hensman et al., 2013) was used instead, as it scales to potentially millions of training points, and is easily expressed as a probability distribution over the model.

Figure 2 illustrates the functional decomposition by showing the effects of positively and negatively varying the coordinates a_i on the decomposed mean function, here the building energy use day profile, for the three first eigen functions.

3.2 Accounting for functional variables in the Design of Experiment (DoE)

Let $X = (x_1, \ldots x_n)$ and $Z = (z_1(t) \ldots z_m(t))$ be respectively the unknown scalar and functional variables (in our case time series). Usually, an optimized Latin Hypercube Design (LHD) is created on a bounded domain of X using some distance metric (see Jin et al. (2005); Morris and Mitchell (1995), and is used to perform GSA (Saltelli et al., 2008; Iooss and Lemaître, 2015), which assumes the random variables are independent. Time series, besides being high dimensional, are auto-correlated, i.e. each time step depends from the previous ones, and may be dependent from each other, e.g. building lighting and equipment daily electricity consumption are likely to be high when the building is occupied and low when it is not. As such, we cannot build a LHD on the functional variables.

We relied on the work of Nanty et al. (2016) for building our full DoE. Using the SPCA algorithm, an optimized *maximin* DoE on the coefficients $a_{i,j}$ was built on a non hypercubic domain as in Auffray et al. (2012). Then the DoEs on Z and X were combined adapting an algorithm by Muehlenstaedt et al. (2016)

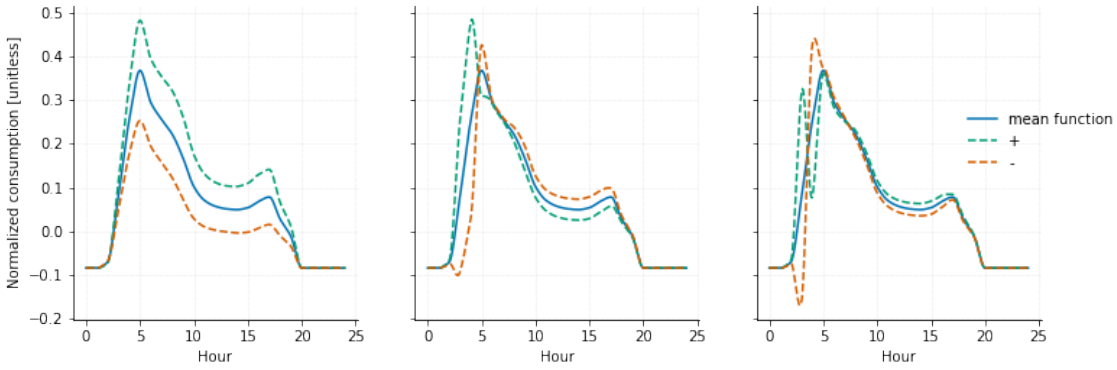


Figure 2: Illustration of the functional decomposition of the normalized heating energy use hourly day profiles. The three figures show the effect of positively (+) and negatively (-) varying the coordinates a_i on the mean decomposed function for the three first eigen functions φ_i. The first eigen function handles the overall amplitude of the day profile, and the second and third ones focus on the peak amplitude and phase variations.

3.3 Global Sensitivity Analysis (GSA)

In the standard framework originally exposed in Kennedy and O'Hagan (2001), a Global Sensitivity Analysis (GSA) (Saltelli et al., 2008) can be used to discard non influential parameters and calibrate only a handful of relevant parameters, thus helping to sanitize an otherwise under-determined problem. However, the Sobol' indices[5] (Saltelli et al., 2008; Janon et al., 2014) are, in their original formulation, only valid between independent variables and scalar outputs. Following the work of Nanty (2015), all coordinates a_i originating from the same functional decomposition were grouped, so only the influence of the group was quantified. Furthermore, the indices computation was also modified to account for the multivariate / functional outputs case as in Gamboa et al. (2014).

3.4 Bayesian BEM calibration

The input data was comprised of:

- Metered data D: these comprise daily profiles of weather - dry bulb temperature and global horizontal irradiance

[5] For model calibration, the total-order indices are used to screen out non-influential variables.

- and indoor temperature, projected onto two functional bases learnt on the simulation outputs from the DoE.

- Unknown data θ: vector of parameters to calibrate. It encompasses both scalar - i.e. we assume the value of each parameter is constant throughout the calibration period - and functional - each day has a different vector of values - variables. Assuming there are n_s scalar variables that the functional decomposition of the functional variables has n_f eigen functions, then there are $n = n_s + n_f \times L$ parameters to calibrate, L being the length, in days, of the calibration period.

The inputs were bound together by the following probabilistic model (Kennedy and O'Hagan, 2001)[6]:

$$y = \eta(D, \theta) + \delta(D) + \varepsilon \tag{1}$$

The probabilistic model from Higdon et al. (2008a) was modified to account for the modularity as in Bayarri et al. (2007, 2009) - since the surrogate model is trained before model calibration - while accounting for the emulator response and the discrepancy model respective functional decomposition. The PPL Numpyro (Bingham et al., 2018) was used in this work to factorize the likelihood on the observed days and to use enumeration of discrete latent variables[7].

3.5 Benchmark

The methodology developed in this work was benchmarked against a standard calibration setting, where neither functional nor dynamic parameters were accounted for (see Heo et al. (2012); Chong et al. (2017) for examples), that is referred to as Static Prior (SP), as opposed to the Dynamic and Functional Prior (DFP) setting. Although the model and its modifications in the SP setting - described in section 3.4 - is not quite the standard described in Higdon et al. (2008a), we felt the comparison would not have been fair otherwise, or at least it would have been blurred by the differences in the calibration procedure, besides the difference in the prior distribution.

In addition to the CVRMSE, NMBE and Q^2 indicators, that inform on the goodness of fit of the calibrated model, because our final goal is to estimate reasonable variation ranges for the observed data, and to detect and diagnose abnormal behaviors, indicators that capture the predictive capability of the model were computed, namely the empirical frequency coverage (EFC) (Kyzyurova et al., 2018) and the reliability-based metric ($\Pr(|y_i - \hat{y}_i| < \varepsilon)$) (Ling et al., 2014).

4. RESULTS

4.1 Comparison of posterior predictive distributions

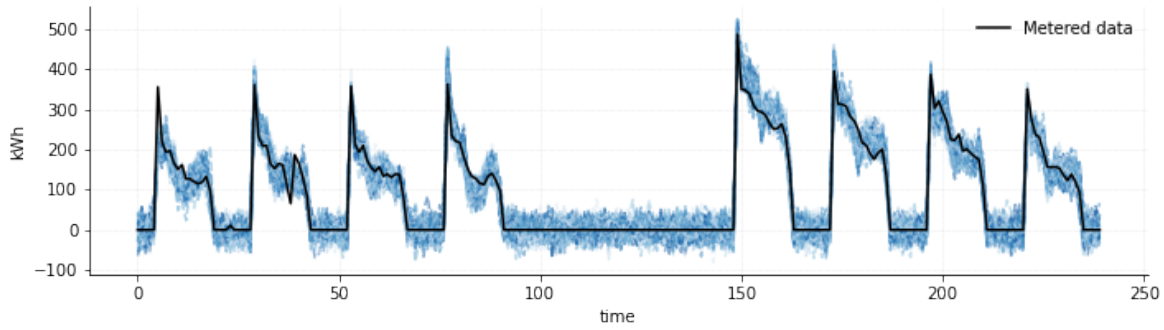
The posterior predictive distributions inform on how well the observed data is explained by the EnergyPlus model - through the EnergyPlus emulator η - and what sorts of discrepancies still remain between the observed data and the EnergyPlus model - through the discrepancy model δ. While a visual analysis is well suited for exploring discrepancies, s.a. phase shifts in the energy use day profile - see Figure 3 for the posterior predictive distribution of η -, the indicators displayed in Table 3 provide a more rigorous comparison as to which model explained the data best.

Table 3: Indicators computed on the posterior predictive distributions of the emulator for both models, SP and DFP.

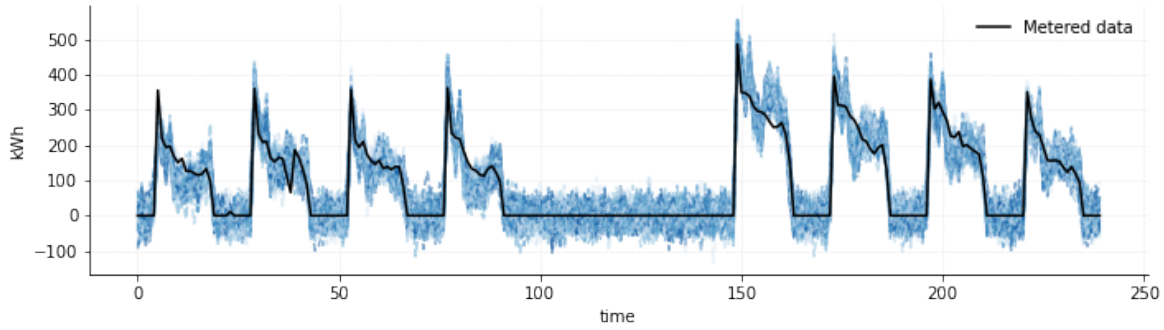
	Q^2	CVRMSE	NMBE	$\text{EFC}_{90\%}$	$\Pr(\lvert y_i - \hat{y}_i \rvert < \varepsilon)$
SP	0.93	0.32	0.02	**0.90**	0.77
DFP	**0.97**	**0.22**	**0.01**	0.91	**0.91**

From Table 3, it seems that the DFP setting is better suited to explain the observed data in almost every respect, and is hence likely to be of better use in an AFDD context. Indeed, indicators such as Q^2, CVRMSE and NBME appear to show better predictive power for the DFP setting. Furthermore, they abide by ASHRAE guildeline 14 about calibrated models, contrary to the SP setting. Besides, the reliability-based metric seems to go the same route and provide a better prediction interval at fixed ε. However, it should be noted that the SP setting provides slightly better uncertainty coverage as the $\text{EFC}_{90\%}$ equals its target value while the DFP setting seems to slightly overestimate the uncertainty.

[6] As in Higdon et al. (2008b), the discrepancy is built as a linear combination of 5 radial basis functions that smoothly adjust the emulator response.
[7] Which is needed when considering a GMM prior on the functional variables.



(a) $\eta + \varepsilon$, DFP



(b) $\eta + \varepsilon$, SP

Figure 3: Posterior predictive distributions of the emulator for both SP (3b) and DFP (3a) models. For the reliability-based metric, ε is arbitrarily taken equal to 50 kWh.

4.2 Comparison of posterior distributions

9 model parameters were selected using Sobol's total-order indices, and were calibrated against the building real heating energy use. Table 4 displays the posterior distributions - mean and standard deviation - for the static parameters (that are common to both settings). Mean estimates, along with calibrated high density intervals (HDI), seems robust since they are quite similar between both models.

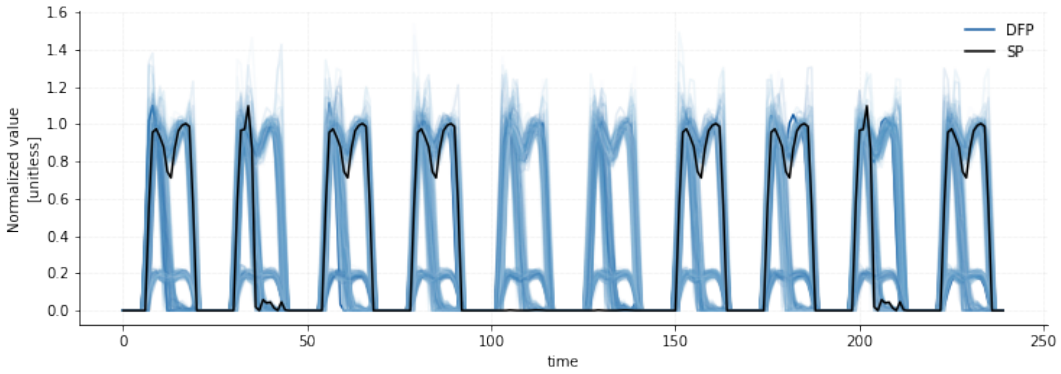
Table 4: Posterior distributions for static variables, for both models SP and DFP

	DFP				SP			
	mean	sd	$q_{3\%}$	$q_{97\%}$	mean	sd	$q_{3\%}$	$q_{97\%}$
h_e	-0.050	0.010	-0.069	-0.031	0.255	0.344	-0.117	0.825
h_s	0.066	0.056	-0.011	0.168	0.154	0.139	-0.038	0.419
η	0.724	0.013	0.700	0.745	0.717	0.013	0.700	0.739
$\upsilon_{natvent}$	1.465	0.036	1.399	1.500	1.460	0.042	1.379	1.500
υ_{occ}	0.678	0.188	0.358	0.982	0.446	0.102	0.350	0.619
t_c	1.195	0.104	1.018	1.403	1.151	0.125	0.949	1.404
t_s	-0.054	0.055	-0.156	-0.000	-0.084	0.077	-0.228	-0.000
υ_{vent}	0.620	0.049	0.529	0.698	0.516	0.097	0.352	0.670

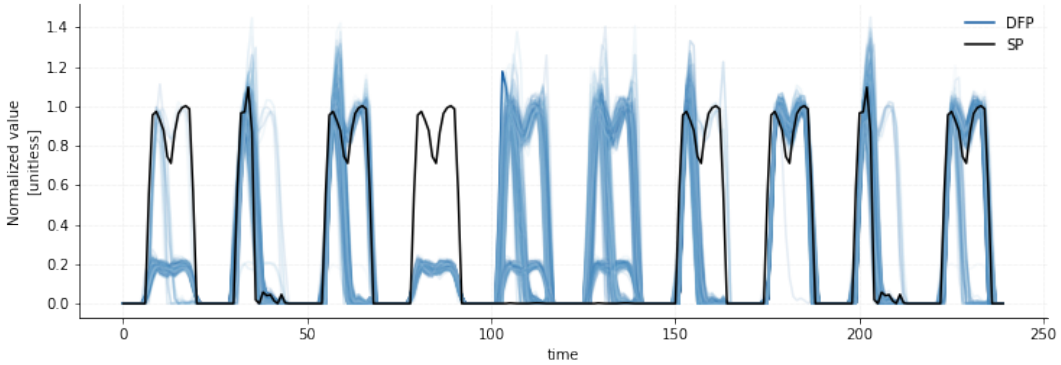
It is worthwhile noting that sometimes parameters HDI overlap, e.g. for parameter t_c, sometimes it seems to be shifted, e.g. for parameter h_e. When they do overlap, most of the time the DFP setting reduces the uncertainty on the said parameter, as can be expected since the functional variables add a degree of freedom per day for the emulator to adjust its response to observed data. In all cases, parameters HDI are not too different between both models.

4.3 Uncertainty reduction on functional variables

It is also worthwhile noting how the functional variables prior is updated with the metered data. Figure 4 displays the resulting posterior distribution after calibration along with the standardized profile from the simulation model, used in the SP model. While the prior distribution spans multiple kinds of profiles - half-days, full days, and holidays - the posterior distribution focuses much more on one type of profile, sometimes two, showing potential change in building occupancy from what would be expected. What is even more interesting is that the posterior seems to capture the fact that, during the calibration period, *Lycée Brequigny* was largely under-occupied due to the pandemic. However, while the school was under-occupied globally at this period, no record of individual days was kept, so it is difficult to assess whether or not the DFP accurately recovers the occupancy profiles.



(a) Functional prior for DFP model and SP values



(b) Functional posterior for DFP model and SP values

Figure 4: Functional prior (4a) and posterior (4b) for standardized occupancy functional variable. In black are the base values from the EnergyPlus model, that are taken in the SP model.

5. CONCLUSIONS & PERSPECTIVES

Our study introduces an extension of the usual calibration process exposed in (Kennedy and O'Hagan, 2001; Higdon et al., 2008a), that accounts for dynamic states of the building throughout the calibration period. The methodology is competitive with the current state-of-the-art, and seems to possess superior flexibility in fitting and explaining the observed data.

However, the methodology needs to be tested against a larger set of both real life and simulated experiments to be completely validated, in terms of parameter estimation and predictability. Besides, since the calibration period exhibits only weak variability on the functional variables, little improvement is shown on the case study. We expect our methodology to be much more valuable in cases where, for instance, occupancy would change more abruptly. Eventually, testing this methodology for continuous monitoring and diagnosis in an online setting seems promising to

update our knowledge of how the building is operated, and help diagnose faults at building level.

NOMENCLATURE

Built environment

AFDD	Automated Fault Detection and Diagnosis
AMI	Advanced Meter Infrastructure
BEM	Building Energy Model
BMS	Building Management System
EPC	Energy Performance Contract
HVAC	Heating, Ventilation and Air Conditioning

Computer model analysis

DoE	Design of Experiment
GSA	Global Sensitivity Analysis
LHD	Latin Hypercube Design

Math terms

DFP	Dynamic and Functional Prior calibration setting
GMM	Gaussian Mixture Model
PPL	Probabilistic Programming Languages
SP	Static Prior calibration setting
SPCA	Simultaneous functional Principal Component Analysis

Symbols

δ	Discrepancy model
η	Emulator
θ	Unknown and calibrated variables
D	Known variables, s.a. dry bulb temperature
t	Time index

REFERENCES

Auffray, Y., Barbillon, P., and Marin, J.-M. (2012). Maximin design on non hypercube domains and kernel interpolation. *Statistics and Computing*, 22(3):703–712.

Bayarri, M. J., Berger, J. O., and Liu, F. (2009). Modularization in bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1).

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007). A framework for validation of computer models. *Technometrics*, 49(2):138–154.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D. (2018). Pyro: Deep universal probabilistic programming.

Chong, A., Lam, K. P., Pozzi, M., and Yang, J. (2017). Bayesian calibration of building energy models with large datasets. *Energy and Buildings*, 154:343 – 355.

EnergyPlus (2019). *EnergyPlus Software, Version 9.1.0*.

Frank, S., Heaney, M., Jin, X., Robertson, J., Cheung, H., Elmore, R., and Henze, G. (2016). Hybrid model-based and data-driven fault detection and diagnostics for commercial buildings.

Frank, S., Lin, G., Jin, X., Singla, R., Farthing, A., and Granderson, J. (2019). A performance evaluation framework for building fault detection and diagnosis algorithms. *Energy and Buildings*, 192:84–92.

Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014). Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575–603.

Granderson, J., Touzani, S., Custodio, C., Sohn, M. D., Jump, D., and Fernandes, S. (2016). Accuracy of automated measurement and verification (m&v) techniques for energy savings in commercial buildings. *Applied Energy*, 173:296 – 308.

Hensman, J., Fusi, N., and Lawrence, N. D. (2013). Gaussian processes for big data.

Heo, Y., Choudhary, R., and Augenbroe, G. (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47:550 – 560.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008a). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583. cited By 356.

Higdon, D., Nakhleh, C., Gattiker, J., and Williams, B. (2008b). A bayesian calibration approach to the thermal problem. *Computer Methods in Applied Mechanics and Engineering*, 197(29-32):2431–2441.

Iooss, B. and Lemaître, P. (2015). *A Review on Global Sensitivity Analysis Methods*, pages 101–122. Springer US, Boston, MA.

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2014). Asymptotic normality and efficiency of two sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364.

Jin, R., Chen, W., and Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, 134(1):268 – 287.

Jradi, M., Arendt, K., Sangogboye, F., Mattera, C., Markoska, E., Kjærgaard, M., Veje, C., and Jørgensen, B. (2018). Obepme: An online building energy performance monitoring and evaluation tool to reduce energy performance gaps. *Energy and Buildings*, 166:196 – 209.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Kyzyurova, K. N., Berger, J. O., and Wolpert, R. L. (2018). Coupling computer models through linking their statistical emulators. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1151–1171.

Ling, Y., Mullins, J., and Mahadevan, S. (2014). Selection of model discrepancy priors in bayesian calibration. *Journal of Computational Physics*, 276:665 – 680.

Monari, F. (2016). *Sensitivity Analysis and Bayesian Calibration of Building Energy Models*. PhD thesis, University of Strathclyde.

Morris, M. D. and Mitchell, T. J. (1995). Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43(3):381 – 402.

Muehlenstaedt, T., Fruth, J., and Roustant, O. (2016). Computer experiments with functional inputs and scalar outputs by a norm-based approach. *Statistics and Computing*, 27(4):1083–1097.

Nanty, S. (2015). *Stochastic methods for uncertainty treatment of functional variables in computer codes : application to safety studies*. Thesis, Université Grenoble Alpes.

Nanty, S., Helbert, C., Marrel, A., Pérot, N., and Prieur, C. (2016). Uncertainty quantification for functional dependent random variables. *Computational Statistics*, 32(2):559–583.

Ramsay, J., Silverman, B., Media, S. S., and Silverman, H. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. MIT Press Ltd.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Johyn Wiley & Sons, Ltd.

Titsias, M. (2009). Variational learning of inducing variables in sparse gaussian processes. In van Dyk, D. and Welling, M., editors, *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 567–574, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA. PMLR.

Touzani, S., Ravache, B., Crowe, E., and Granderson, J. (2019). Statistical change detection of building energy consumption: Applications to savings estimation. *Energy and Buildings*, 185:123 – 136.