



**HAL**  
open science

## Assessment of multiple advanced Bayesian calibration algorithms in building energy models

Samih Akkari, Patrick Schalbart, Bruno Peuportier

► **To cite this version:**

Samih Akkari, Patrick Schalbart, Bruno Peuportier. Assessment of multiple advanced Bayesian calibration algorithms in building energy models. ConférenceConférence Francophone de l'International Building Performance Simulation Association ( IBPSA )France 2022, May 2022, Châlons en Champagne, France. hal-03801058

**HAL Id: hal-03801058**

**<https://hal.science/hal-03801058v1>**

Submitted on 6 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Assessment of multiple advanced Bayesian calibration algorithms in building energy models

Samih Akkari<sup>\*1</sup>, Patrick Schalbart<sup>1</sup>, Bruno Peuportier<sup>1</sup>

<sup>1</sup> MINES ParisTech, PSL Research University, CES - Centre for energy efficiency of systems, Z.I. Les Glaizes – 5 rue Léon Blum 91120 Palaiseau, France

[\\*samih.akkari@mines-paristech.fr](mailto:samih.akkari@mines-paristech.fr)

---

*RESUME. De nombreux chercheurs ont opté pour le calibrage bayésien des modèles énergétiques des bâtiments car il combine à la fois les données disponibles et les connaissances préalables sur le bâtiment, en prenant en compte l'incertitude des paramètres du modèle. Les méthodes bayésiennes peuvent être classées en approches dépendantes de la vraisemblance et sans vraisemblance. Dans cet article, nous tirons parti des deux classes et évaluons leurs comportements en termes de précision et d'efficacité de calcul en utilisant l'indicateur RMSE. À cet effet, des mesures in situ de la maison IBB de la plateforme INCAS ont été analysées. La campagne expérimentale est découpée en six scénarios différents dont deux ont été conservés pour la validation. Les paramètres les plus influents ont été sélectionnés via une analyse de sensibilité RBD-FAST. Il a été observé que tous les algorithmes fonctionnent bien, APMC conduisant au RMSE le plus bas. Globalement, il n'y a pas de grande différence entre les deux classes, ce qui ouvre la porte à d'autres applications du calcul bayésien approché dans le domaine des modèles énergétiques des bâtiments.*

*MOTS-CLÉS : Calibrage bayésien, Calcul bayésien approché, Modèles énergétiques des bâtiments*

---

*ABSTRACT. Bayesian calibration of building energy models has attracted many researchers since it combines both the data available and the prior knowledge about the building, and it considers uncertainty of the model's parameters. Bayesian methods can be classified into likelihood-dependent and likelihood-free approaches. In this article, both classes are assessed in terms of accuracy, and computational efficiency using RMSE indicator. For this purpose, in situ measurement of the IBB house of the INCAS platform were analysed. The experimental campaign is split into six different scenarios two of which were kept for validation. The most influential parameters were selected via RBD-FAST sensitivity analysis. It was observed that all algorithms performed well with APMC leading to the lowest RMSE. Globally, there is no large difference between both classes which opens the door for more applications of approximate Bayesian computation in the field of building energy models.*

*KEYWORDS: Bayesian calibration, approximate Bayesian computation, building energy models*

---

## 1. INTRODUCTION

Renovation of existing buildings can be studied using a Building energy model (BEM). Calibrating such a model aims at reducing uncertainties of predictions and fitting better to the actual behaviour. Normally, a sensitivity analysis is conducted prior to calibration in order to select the most influential parameters and to reduce computational burden. Various sensitivity methods are available in literature (Pannier et al. 2018). RBD-FAST is considered in this study due to its computational efficiency compared to other variance based methods.

Bayesian calibration benefits from prior knowledge and it naturally quantifies the uncertainties in the model predictions. Kennedy and O'Hagan (2001) proposed a general framework for the application of Bayesian calibration on computer models, and it was firstly introduced to the building sector by Heo et al. (2012). There are two families for Bayesian calibration: likelihood-dependent which are widely applied to BEM and approximate Bayesian computation (ABC) proposed by Pritchard et al. (1999) which approximates the likelihood with distance metric. .

Robillart (2015) firstly applied ABC-population Monte-Carlo algorithm (ABC-PMC) proposed by Beaumont et al. (2009) to calibrate a BEM generated by COMFIE. Zhu et al. (2020) applied the ABC approach based on Pritchard et al. (1999) framework with two different posterior post-processing techniques. The applied method suffers from computational burden and robustness issues since all the samples are drawn from the prior distribution. ABC-PMC applied by Robillart (2015) overcomes this problem by sampling from a distribution closer to the posterior rather than sampling from the prior.

Various updated more robust versions of ABC methods have been recently published. Still there is no sufficient applications of these methods to BEMs and both families of Bayesian calibration were not compared. In this paper, five algorithms from both families are selected, according to their popularity and their ability to be parallelised. A comparative analysis between those algorithms is done in terms of accuracy and computational efficiency.

## 2. METHODS

### 2.1. SENSITIVITY ANALYSIS

RBD-FAST is the method retained from the literature since it offers several advantages among the other variance-based methods. This method is less sensitive to the number of parameters compared to other methods as stated in the literature (Goffart and Woloszyn, 2021). It allows for the estimation of the sensitivity indices with only  $N$  model evaluations. In this regard, an input-output data set of 7000 samples was generated to ensure stability in ranking the most influential parameters given that the model comprises 127 parameters.

### 2.2. BAYESIAN CALIBRATION

Bayesian calibration relies on Bayes' theorem to infer some parameters given prior knowledge about them and measured data. In most cases, solving Bayes' theorem analytically is intractable, thus, a Markov Chain Monte-Carlo (MCMC) sampler is convenient to approximate the posterior distribution by generating samples from it.

The choice of the sampler has a large effect on the calibration convergence and its computation time. The chosen algorithms differ in terms of the use of the likelihood function and the adopted sampler type. Metropolis-Hastings and NUTS samplers are not considered in this study. NUTS requires numerous computations of the likelihood gradients, which is computationally inefficient given that it is un-parallelisable and that in this paper, no metamodel is used to replace the original one. The house is modelled as a single thermal zone, and thanks to model reduction applied to the BEM, the simulation is fast. Sequential Monte-Carlo samplers (SMC) are preferred since they can be easily parallelised. Two likelihood-dependent approaches – CATMIP (Minson, et al. 2013), and one used by Adams et al. (2020) – and three ABC approaches – ABC-PMC (Beaumont et al., 2009), APMC (Lenormand et al. 2013), ABC-RF (Raynal et al. 2017) – are selected. The hyper-parameters

corresponding to each algorithm are not tuned. Instead, they were kept at the default values recommended by the authors and in the literature.

It is worth mentioning that all these algorithms are different variants and applications of the SMC sampler except ABC-RF, which is based on using machine learning (random forest) in ABC context.

### 2.3. UNCERTAINTY PROPAGATION

To validate the performance and the prediction accuracy of the calibrated model, the priors and the posteriors of the calibrated parameters are propagated. This provides information regarding the uncertainties in the prediction before and after calibration. To quantify the model predictive performance, the temperature root mean square error (RMSE) between the measurements and the predictions is evaluated for each propagated sample and the average of all the samples was retained.

## 3. CASE STUDY

The studied building corresponds to the low energy I-BB house (Concrete construction) of the INES (National Institute of Solar Energy) "INCAS" platform, located in Le Bourget-du-Lac. The interior surface area is 89 m<sup>2</sup> with two floors (Munaretto et al., 2017).

The experimental campaign was designed to measure the temperature profile of the building. The considered period (January 1 to April 22, 2012) is subdivided into six scenarios (Figure 1), with separate or combined consideration of different physical phenomena. Parameters, such as the heating setpoint, the opening / closing of the shutters, or the mechanical ventilation speed and flow were modified.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
Heating power	20°C	Off	Off	24°C	Off	Off
Ventilation mode	BP	Off	Off	BP	BP	BP
Ventilation flow rate	Rate-a	0	0	Rate-a	Rate-a	Rate-b
Shutters	Closed	Closed	Closed	Closed	Open	Open
Separation GF/F1	Separated	Separated	Not separated	Separated	Not separated	Not separated
	01/01/2012	16/01/2012	09/02/2012	20/02/2012	15/03/2012	31/03/2012

Figure 1: Illustration of the different scenarios adapted from Munaretto et al. (2017)

In the following analysis, as in Robillart (2015), scenarios 1,2,3,and 5 were considered training scenarios on which calibration process was performed. Scenarios 4 and 6 were chosen as validation data because they are respectively close to the 1<sup>st</sup> and 5<sup>th</sup> scenarios. We can thus evaluate the behaviour of the calibrated model under relatively similar experimental conditions.

## 4. RESULTS

### 4.1. SENSITIVITY ANALYSIS

The six most influential parameters that were identified and were considered for calibration are: ventilation flow-rate, dissipated heat, heating power, heat capacity of concrete, solar albedo, and the conductivity of polystyrene as shown in Figure 2. The reason why the solar albedo and the heating power were found to be less influential is that there is no heating power in scenarios 2, 3, and 5, and

the shutters were closed in scenarios 1, 2 and 3, which affected the importance of the solar albedo. However, the parameters identified as influential are present in all the scenarios except for the ventilation that is only present in scenarios 1 and 5 (ventilation is off in scenarios 2 and 3).

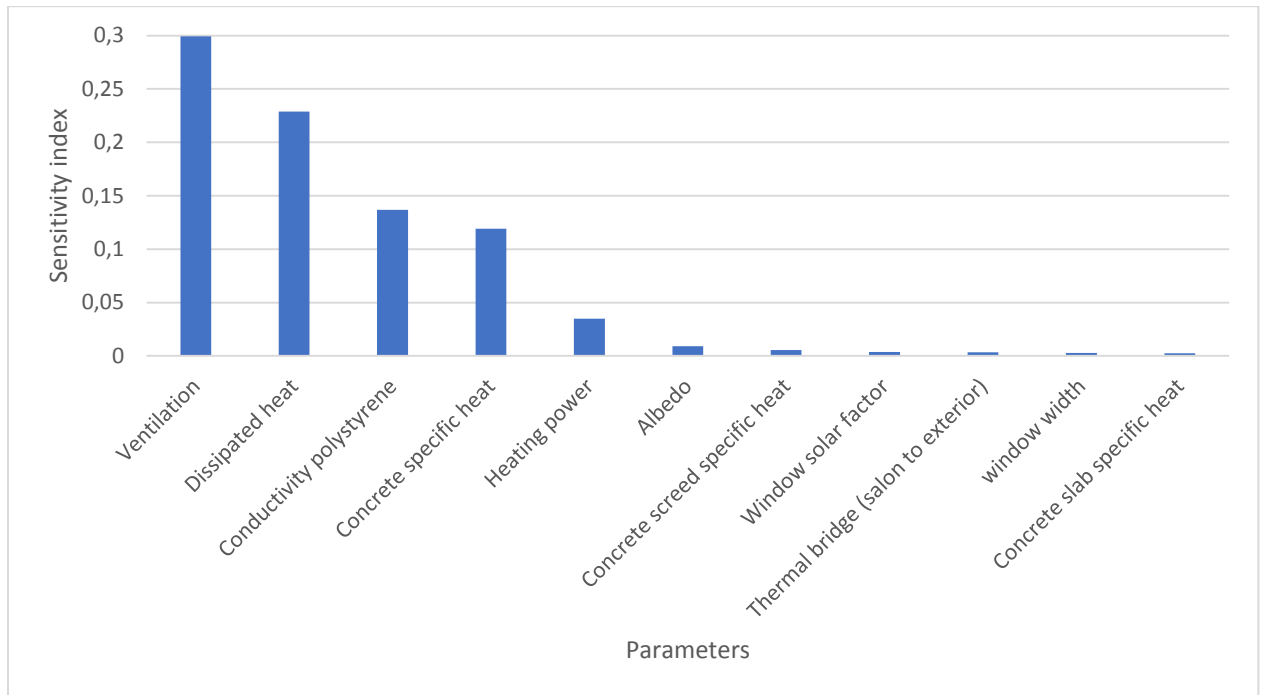


Figure 2: Parameters ranking with RBD-FAST

## 4.2. CALIBRATION

To perform the comparative analysis, the algorithms were firstly applied in controlled conditions where virtual data were generated from the building model based on pre-selected values for the calibration parameters. In a second step, actual on-site measurements were used to perform the inference. This allowed to undergo a comprehensive analysis on the algorithms performance with and without the influence of unknown disturbance and noise. For both cases, the priors considered by Robillart, (2015) were retained in this study. The true values from which the virtual data was generated were shifted by two standard deviations from the means of the selected priors.

### 4.2.1. Controlled conditions

Figure 3 shows the evolution of each algorithm against the number of model evaluations with a simulation budget of 30,000. APMC and Adams showed a similar performance and they converged better and faster to lower RMSE values than CATMIP and ABC-PMC: the latter being the slowest to converge. ABC-RF showed a different performance where, in contrary to the other algorithms, its accuracy with increasing model evaluations did not show a continuous increase: the RMSE indicator slightly decreases with higher model evaluations with significant variability.

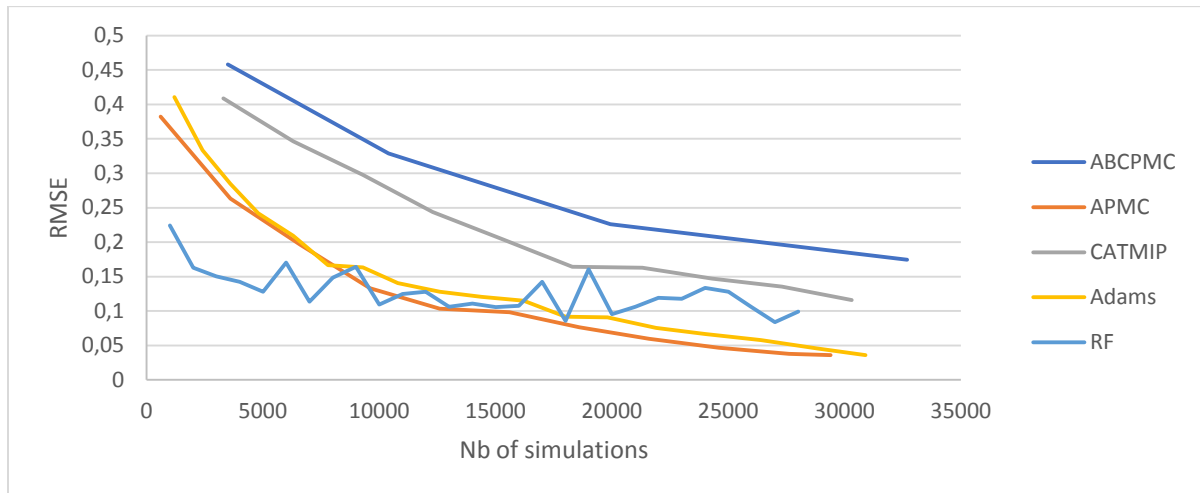


Figure 3: Model prediction accuracy for each algorithm (virtual data)

Another difference is that all the algorithms except ABC-RF were capable of reaching similar accuracies in finding the true values of the parameters and in model prediction performance but with different numbers of model evaluations. CATMIP and ABC-PMC required 54000 and 95000 simulations respectively to reach the same accuracies attained by Adams and APMC with only 30000 simulations. However, the model predictive accuracy of ABC-RF was less than the others even when trained with a data set size of 100000 samples: RF posterior yielded an RMSE of  $0.08^{\circ}\text{C}$  while the other algorithms reached  $0.035^{\circ}\text{C}$  with the mentioned number of model evaluations. Despite this disadvantage and the variability weighing on the method, it performs better than the other methods with small number of model evaluations: less than 10000 as depicted in Figure 3. It means that there is a potential in RF with small data size. The other methods can yield better accuracies with increasing model evaluations.

#### 4.2.2. Real measurements

The same methodology was applied on the real measurements of the case study. Figure 4 shows the RMSE of the posteriors and of the priors after a simulation budget of around 44000 model evaluations and Figure 5 shows the temperature profile in all the scenarios. All algorithms yielded a more precise model to adequately fit the data. There is a significant decrease in the RMSE value between the calibrated and the un-calibrated model, with APMC being the best and RF being the least accurate. The performance was good on the training scenarios (1,2,3 and 5) and on the testing scenarios (4 and 6) even though, some scenarios were better fitted than the others.

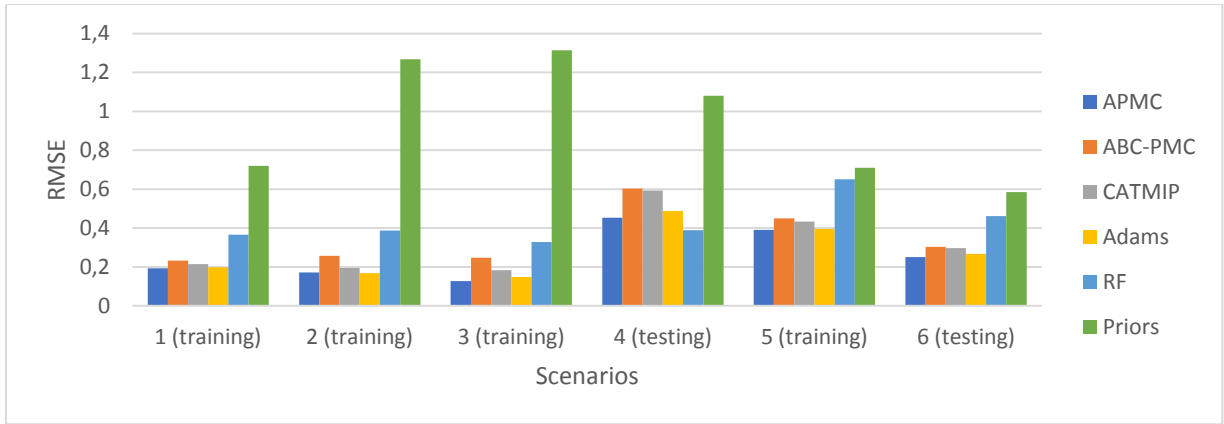


Figure 4: Model prediction accuracy after each algorithm

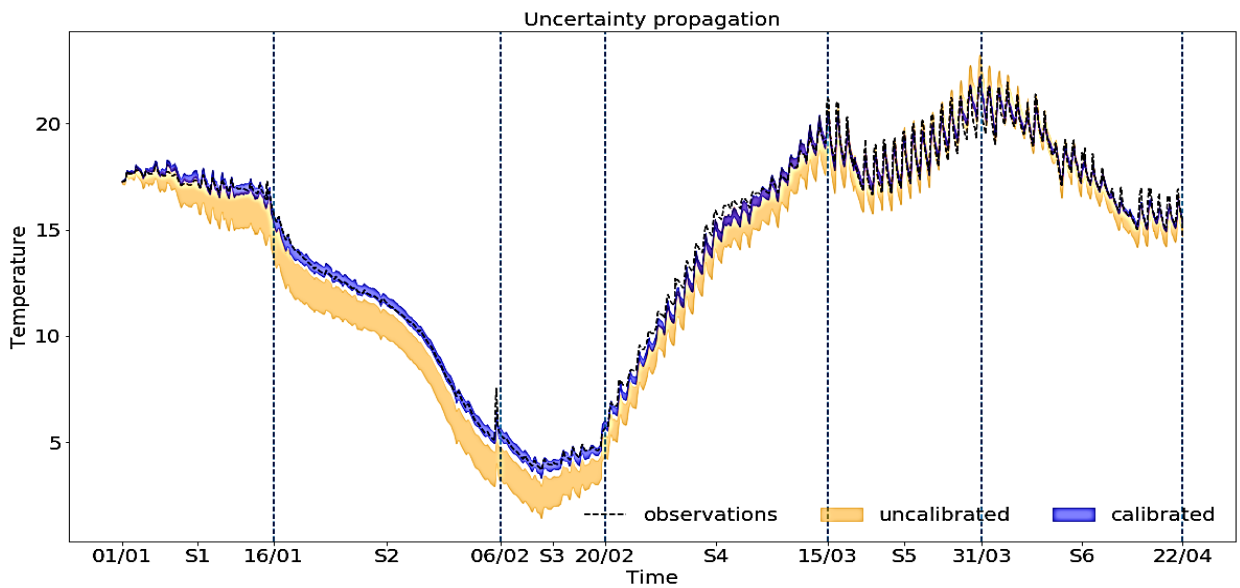


Figure 5: Uncertainty propagation on temperature profile

The accuracy of the algorithms is consistent with what was observed on the virtual data. Figure 6 shows the performance during the evolution of the algorithms. In this case, unlike previously, a plateau is observed where the minimum threshold defined cannot be attained. This means that the model is having difficulty in fitting the actual behaviour better. This could be related to different issues such as measurement bias and uncertainty, error in the specification of the uncalibrated parameters, or errors due to model assumptions. It can also be emphasised that the algorithms show similar relative computational efficiency as depicted previously.

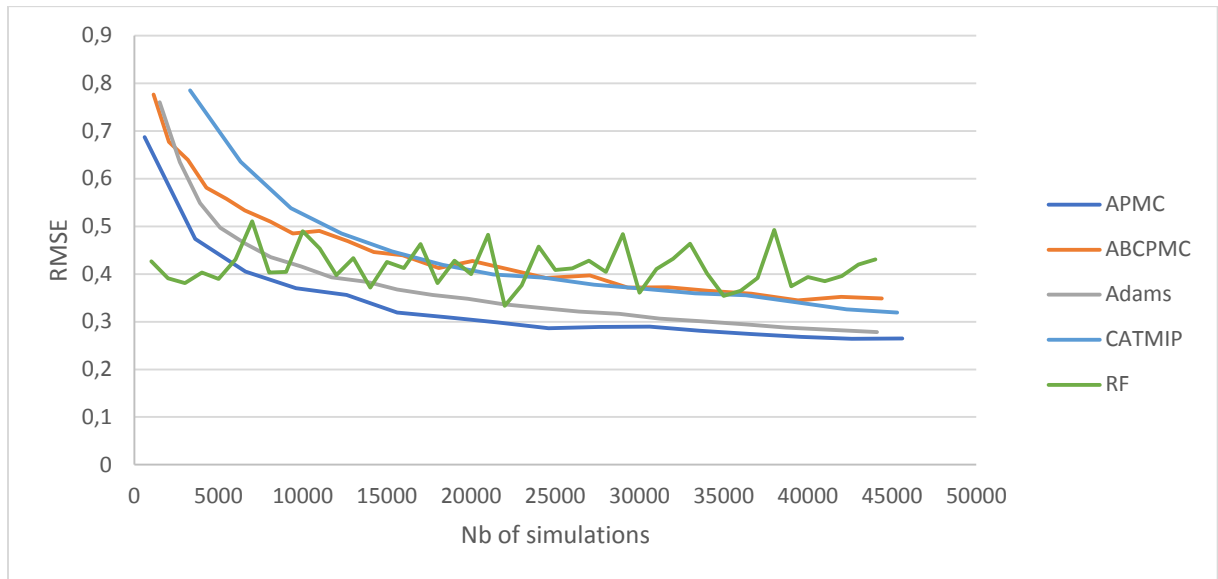


Figure 6: Model prediction accuracy for each algorithm (real measurements)

ABC-RF was also run with a data set size of 100000 samples and there were no major differences with the results obtained with smaller size. On top of that, it is important to mention the additional computational cost of training the random forests: for each parameter, a new random forest needs to be trained. One of the main aspects of RF is that it cannot predict values outside the range of the priors. Thus, if the true value of the process lies outside the boundaries of the priors, RF could only favour the samples that lie in the prior and are closer to the true value. That is to say that RF is less robust to the selection of the priors than the other algorithms. To clarify this finding, Figure 7 depicts the prior and the posterior of the heating power obtained with APMC and RF algorithms. We note that the posterior is quite similar for all algorithms except for RF.

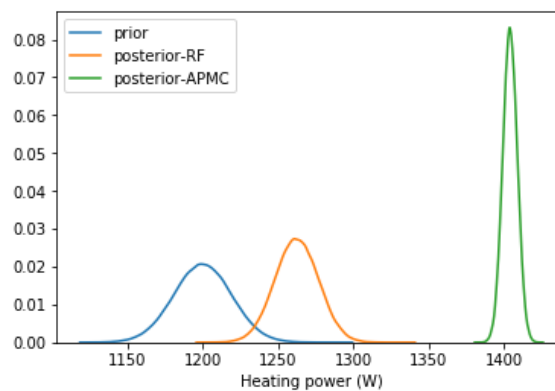


Figure 7: Prior vs posterior (heating power) APMC, RF

The mean of the posterior obtained with RF is 1262 W whereas the mean of the heating power posterior is around 1400W for the other algorithms. The same behaviour was also noticed with the heat capacity of the concrete wall. To improve the performance of this algorithm, it could be better to increase the ranges of the priors which would require a larger data set to explore well the parameter space. However, since it is observed that RF can yield, with a small data set, similar results as with a big data set, it could be better to increase the prior range with a similar data set size.



## 5. CONCLUSION

Different calibration algorithms were selected from the literature to assess their performance (APMC, ABC-PMC, RF, CATMIP, and Adams). The analysis was executed using data available for the IBB house of the INCAS platform and virtual data. In this case study, the likelihood-independent algorithms function as well as those that require the likelihood. That is to say that the approximation of the likelihood function was found to be adequate. APMC and Adams outperformed the other algorithms in terms of computational efficiency and precision. ABC-RF gave the least precise results with larger data sets and the most precise ones with small data sets.

For the virtual data, all the algorithms estimated well the true values of the parameters, while for the actual measurements, the parameters estimation followed a similar pattern for all the algorithms except for RF. This is most probably related to the ranges of the priors. If wider ranges were used in RF, improvements in the performance might be obtained.

Globally, all the considered algorithms enhanced the performance and the capability of the model to fit the data on the training and the testing data. Further improvements could have been attained if the building under investigation had been modelled as a multizone instead of monozone.

## 6. REFERENCES

- ADAMS, M. P., KOH, E. J. Y., VILAS, M. P., COLLIER, C. J., LAMBERT, V. M., SISSON, S. A., QUIROZ, M., MCDONALD-MADDEN, E., MCKENZIE, L. J. et O'BRIEN, K. R., 2020. Predicting seagrass decline due to cumulative stressors. In : *Environmental Modelling & Software*. août 2020. Vol. 130, p. 104717. DOI 10.1016/j.envsoft.2020.104717.
- BEAUMONT, M. A., CORNUET, J.-M., MARIN, J.-M. et ROBERT, C. P., 2009. Adaptive approximate Bayesian computation. In : *Biometrika*. 1 décembre 2009. Vol. 96, n° 4, p. 983- 990. DOI 10.1093/biomet/asp052.
- GOFFART, J. et WOLOSZYN, M., 2021. EASIRBD-FAST: An efficient method of global sensitivity analysis for present and future challenges in building performance simulation. In : *Journal of Building Engineering*. 1 novembre 2021. Vol. 43, p. 103129. DOI 10.1016/j.job.2021.103129.
- HEO, Y., CHOUDHARY, R. et AUGENBROE, G. A., 2012. Calibration of building energy models for retrofit analysis under uncertainty. In : *Energy and Buildings*. avril 2012. Vol. 47, p. 550- 560. DOI 10.1016/j.enbuild.2011.12.029.
- KENNEDY, M. C. et O'HAGAN, A., 2001. Bayesian calibration of computer models. In : *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. août 2001. Vol. 63, n° 3, p. 425- 464. DOI 10.1111/1467-9868.00294.
- LENORMAND, M., JABOT, F. et DEFFUANT, G., 2013. Adaptive approximate Bayesian computation for complex models. In : *Computational Statistics*. décembre 2013. Vol. 28, n° 6, p. 2777- 2796. DOI 10.1007/s00180-013-0428-3.
- MINSON, S. E., SIMONS, M. et BECK, J. L., 2013. Bayesian inversion for finite fault earthquake source models I—theory and algorithm. In : *Geophysical Journal International*. 1 septembre 2013. Vol. 194, n° 3, p. 1701- 1726. DOI 10.1093/gji/ggt180.
- MUNARETTO, F., RECHT, T., SCHALBART, P. et PEUPORTIER, B., 2017. Empirical validation of different internal superficial heat transfer models on a full-scale passive house. In : *Journal of Building Performance Simulation*. 6 juin 2017. Vol. 0, n° 0, p. 1- 22. DOI 10.1080/19401493.2017.1331376.

PANNIER, M.-L.,SCHALBART, P. et PEUपोर्टIER, B., 2018. Comprehensive assessment of sensitivity analysis methods for the identification of influential factors in building life cycle assessment. In : *Journal of Cleaner Production*. 20 octobre 2018. Vol. 199, p. 466- 480. DOI 10.1016/j.jclepro.2018.07.070.

PRITCHARD, J. K.,SEIELSTAD, M. T.,PEREZ-LEZAUN, A. et FELDMAN, M. W., 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. In : *Molecular Biology and Evolution*. 1 décembre 1999. Vol. 16, n° 12, p. 1791- 1798. DOI 10.1093/oxfordjournals.molbev.a026091.

RAYNAL, L.,MARIN, J.-M.,PUDLO, P.,RIBATET, M.,ROBERT, C. P. et ESTOUP, A., 2017. ABC random forests for Bayesian parameter inference. In : *Peer Community in Evolutionary Biology*. 17 novembre 2017. p. 100036. DOI 10.24072/pci.evolbiol.100036.

ROBILLART, M., 2015. *Etude de stratégies de gestion en temps réel pour des bâtiments énergétiquement performants* [en ligne]. phdthesis. S.l. : Ecole Nationale Supérieure des Mines de Paris. [Consulté le 9 avril 2020]. Disponible à l'adresse : <https://pastel.archives-ouvertes.fr/tel-01299525>.

ZHU, C.,TIAN, W.,YIN, B.,LI, Z. et SHI, J., 2020. Uncertainty calibration of building energy models by combining approximate Bayesian computation and machine learning algorithms. In : *Applied Energy*. 15 juin 2020. Vol. 268, p. 115025. DOI 10.1016/j.apenergy.2020.115025.