Quantifier l'apprentissage par les données en approche bayésienne

Sarah Juricic*[1], Simon Rouchier[1], Arnaud Jay[2], Jeanne Goffart[1], Gilles Fraisse[1]

[1] **Univ. Savoie Mont-Blanc, CNRS, LOCIE**
Chambéry, France
[2] **Univ Grenoble Alpes, CEA, LITEN, INES**
Grenoble, France
***sarah.juricic@gmail.com**

RESUME. Caractériser la performance thermique d'un bâtiment à partir de mesures in situ nécessite de pouvoir détecter les cas de non identifiabilité, i.e. indétermination des résultats. Si l'approche bayésienne s'affranchit de l'impossibilité numérique d'estimer cette performance, elle ne permet pas en soi de détecter des mesures insuffisantes pour l'apprentissage des propriétés recherchées. Or si les données sont peu informatives, comme quand l'échantillonnage temporel est inadapté, l'interprétation physique des résultats est compromise. Cet article se propose de mesurer l'apprentissage de la vraisemblance par les données au moyen de la divergence de Kullback-Leibler. Un cas d'étude montre l'usage de cette métrique en comparant l'apprentissage d'un modèle RC par un même jeu de données rééchantillonné à des pas de temps de 8 à 120 minutes. Les résultats suggèrent des pas de temps de moins de 30 minutes pour un apprentissage plus quantitatif.

MOTS-CLEFS. Identifiabilité pratique, Calibration bayésienne, Granularité temporelle

ABSTRACT. Upon estimating the thermal performance of a building from in situ measurements, non identifiability cases, i.e. undetermination of the results, must be efficiently uncovered. If a Bayesian approach is not concerned by the numerical issues raised by non identifiability, it does not indicate when there is no learning from the data. Yet, if the data is uninformative, as when temporal sampling is inappropriate, physical interpretation of the results is compromised. This paper proposes to measure learning from the likelihood, i.e. learning from the data only, by the means of a Kullback-Leibler divergence. A case study applies this metric to compare learning of an RC model from a single dataset resampled with time steps from 8 to 120 minutes. Results suggest time steps shorter than 30 minutes quantitatively enhance learning.

KEYWORDS. Practical identifiability, Bayesian calibration, Temporal granularity

1 Introduction

Accurate thermal diagnosis of buildings is a key to drive adapted retrofit strategies, necessary to reduce the energy use of the built environment. Diagnoses from in-situ non destructive measurements have recently gained a strong interest as they deliver insight on the actual thermal performance of the building.

Methods using RC thermal models and relying on controlled indoor conditions have been proven to be efficient (Madsen et al., 2015). By optimizing the heating input and taking advantage of the dynamic nature of the data, datasets of only a few days are sufficient to infer an estimation of the thermal performance. Success and accuracy of the numerical estimation will however be conditioned by the practical identifiability of the model : the ability of the model to infer unique solutions given some data. Practical identifiability is therefore a property that depends on both the model and the quality of information in the data.

The Bayesian approach presents then the advantage of directly accounting for existing expert knowledge of the system in the numerical resolution. It acts as a regularization of the inverse problem. Model parameters carry therefore in any case the prior information : there cannot be undefined parameters. But then, is it still possible to detect uninformative data ?

This paper proposes to measure the information solely gained from the data in a Bayesian calibration by the Kullback-Leibler divergence. In the perspective of thermal performance estimation, this metric is applied to assess the loss of information when data is inadequately sampled, from a temporal point of view.

2 The Bayesian approach to model identification

The estimation of a building thermal property requires three things : data, an appropriate thermal model and a numerical tool to infer the property of interest. The latter can be approached in a frequentist or a Bayesian way.

The Bayesian approach considers the unknown model parameters not as scalars but as random variables. Each parameter has then a probability density function : admissible values of that parameter are given a probability density. The probability density function of a parameter can be interpreted as the degree of knowledge of the *true* value of said parameter (Tarantola, 2005). With θ a parameter, the degree of knowledge is conditioned by prior information $p(\theta)$ an expert might have and by data purposely collected y. In the end, a parameter estimation in a Bayesian approach is merely estimating the probability density of θ given y. According to Bayes rule, equation (1) follows : the posterior distribution $p(\theta|y)$ is proportional to the prior knowledge $p(\theta)$ and to the likelihood $p(y|\theta)$. The likelihood represents the information gained from the data y.

$$p(\theta|y) = \frac{p(y|\theta)P(\theta)}{p(y)} \propto p(y|\theta)P(\theta) \tag{1}$$

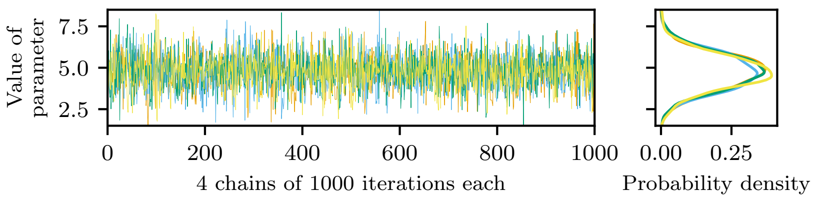


Figure 1. Bayesian calibration with the pySIP package (Raillon et al., 2019). (*Left*) : 4 independent chains of 1000 samples are drawn from the posterior distribution of each parameter of the calibrated model. (*Right*) : each chain constitutes an estimation of the posterior distribution.

The posterior distribution of a parameter is not analytically computable. It is therefore necessary to make a numerical estimation of it. To do so, an appropriate algorithm, such as Hamiltonian Monte-Carlo, draws samples from the posterior distribution (Tarantola, 2005; Betancourt, 2017). As shown in Fig. 1, with a sufficient number of draws (here the last 1000 out of 2000 iterations), the shape of the posterior distribution can be reconstructed. Good practice requires to produce independently 4 chains of iterations. When all 4 are significantly similar, as in the left hand side of Fig. 1, the estimation of the posterior distribution may be considered satisfactory.

3 Practical identifiability in a Bayesian framework

Fundamentally, Bayesian identifiability is not strictly speaking an issue : a Bayesian analysis with proper prior knowledge is always feasible. If there is no learning from the data, i.e. from the likelihood, prior and posterior distribution will be identical (Poirier, 1998). In this sense, identifiability in a Bayesian framework is actually a matter of learning from the likelihood (Xie et Carlin, 2006).

Considering $\theta = (\theta_1, \theta_2)$ a set of parameters, Dawid (1979) proposes a formal definition : the subset θ_2 is not identified by the data if the observation of the data does not increase our prior knowledge about θ_2 given θ_1 :

$$p(\theta_2|\theta_1, y) = p(\theta_2|\theta_1) \tag{2}$$

Sahu et Gelfand (1999) bring the definition of equation (2) into discussion and underline that it does not imply that there is no Bayesian learning (i.e. it does not imply that $p(\theta_2|y) = p(\theta_2)$), but just that there is no conditional learning. The authors then propose a looser yet less formal definition : if "for at least some parameters, the data provides little information" this subset of parameters are "weakly identified". Provided there were a suitable metric between both distributions, this definition would read as $p(\theta_2|\theta_1, y) \approx p(\theta_2|\theta_1)$.

To this purpose, Xie et Carlin (2006) propose a metric as defined in equation (3) based on the Kullback-Leibler (KL) divergence,. The KL divergence is a quantity used to measure the difference between two distributions. $D_{\theta_1,y}$ measures how much is left to learn given data y and is defined as :

$$D_{\theta_1,y} = KL(p(\theta_2|\theta_1), p(\theta_2|y)) = \int_{-\infty}^{\infty} p(\theta_2|\theta_1) log \frac{p(\theta_2|\theta_1)}{p(\theta_2|y)} d\theta_2 \tag{3}$$

Such a metric is in this case analytically not computable, a solution would be to estimate the divergence by a k-nearest-neighbor distance method from samples of the distribution (Wang et al., 2009; Hartland, 2018), as illustrated in Fig. 2.

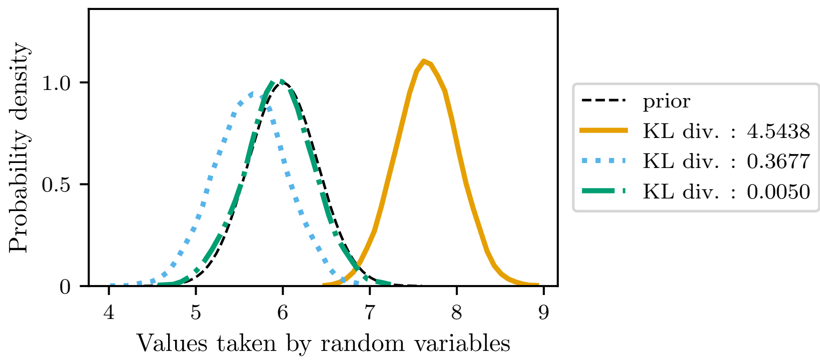


Figure 2. Illustration of the Kullback Leibler divergence for three posterior distributions : the less identifiable, the closer the distributions, the lower the KL divergence

Fig. 2 illustrates such application on three synthetic distributions. Their divergence to a reference "prior" distribution is calculated and displayed in the legend. Divergences close to 0 show similarity between two distributions : that would indicate poor learning from the data. On the contrary, the larger the divergence, the least resemblance prior and posterior have : that would indicate significant learning from the data. Noteworthy, the KL divergence measures the disagreement in both variance and mode difference.

4 Adequate time step for practical identifiability

A cause for practical non identifiability in the estimation of thermal performance is inadequate temporal granularity, i.e. too large time steps therefore overlooking high frequency phenomena.

This section proposes to use the Kullback-Leibler divergence to measure the information gained by a RC model from multiple datasets, each set having a different sampling frequency. This section lays out the issue of inadequate sampling of data, describes the house used as case study and shows finally how time steps influence how much the RC model learns from the data.

4.1 Some background on temporal granularity

Because data and models are not dealt with in continuous time but in a discretized form, RC models may suffer from aliasing. Aliasing happens when two different signals become indistinguishable when sampled. In indoor temperature measurements, faulty sampling will result in misreading or ignoring high frequencies significant variations of temperature (Madsen et al., 2015). Overlooking high frequency temperature variations may then result in practical non identifiability. RC model calibration should therefore necessarily use datasets with a minimal sampling frequency of twice the highest frequencies observed in the heat dynamics of the building, as stated by the Shannon theorem (Madsen, 2008). Madsen et al. (2015) suggest that the sampling time should ideally be kept under the hour.

This issue also relates to that of the representative characteristic times of the indoor air temperature behaviour. Sicard et al. (1985) performed a spectral analysis of the response of indoor air temperature to external solicitations on a numerical study case : outdoor temperature, heat flux indoor and heat flux on indoor walls. They showed that the indoor air temperature shows a strong mode at characteristic time 55 h. The value itself is related to the simulated building, but the order of magnitude is significant. In addition, they show how indoor air temperature has several significant modes around the characteristic time 0.1 h, corresponding to local heating of the air and of the surface of neighbouring walls. Although particularly large in magnitude compared to the others modes, they show that the 55 h mode is not sufficient to make accurate prediction. Predictions were more accurate when taking the two modes with the lowest frequency and in addition a higher frequency mode around 10 h^{-1} (time 0.1 h i.e. 6 min). These high frequency phenomena suggest then, according to Shannon's theorem, that the sampling time step should be maximum 3 minutes. Longer time steps will fail to provide informative data for 2^{nd} and higher order dynamic models.

4.2 Case study and experimental setup

The in-situ data, measured at a 1 minute time step, has been collected in a two storeys unoccupied house in Le Bourget du Lac (Savoie, France). The house is part of an experimental platform of 4 different houses, called the INCAS houses. The building envelope of this specific house is made of traditional shuttered concrete walls. At the time of the experiment, the building was being renovated. There were no thermal insulation on the vertical walls, but heavy insulation below the ground floor concrete slab and in the attics.

During the experiment, the heating power is controlled as to follow a *Pseudo-Random Binary Signal* (PRBS), as shown in Fig. 3. This type of signal takes two possible values, 100 W or 5600 W and ressembles a square wave signal with multiple frequencies. A PRBS signal is designed to cover both high and low frequencies to which a building might be sensitive.

As shown in Fig. 3, 48 h of data is used for the calibration. The outdoor temperature varies between 13 °C and 28 °C as is expected at the beginning of September in Le Bourget du Lac.

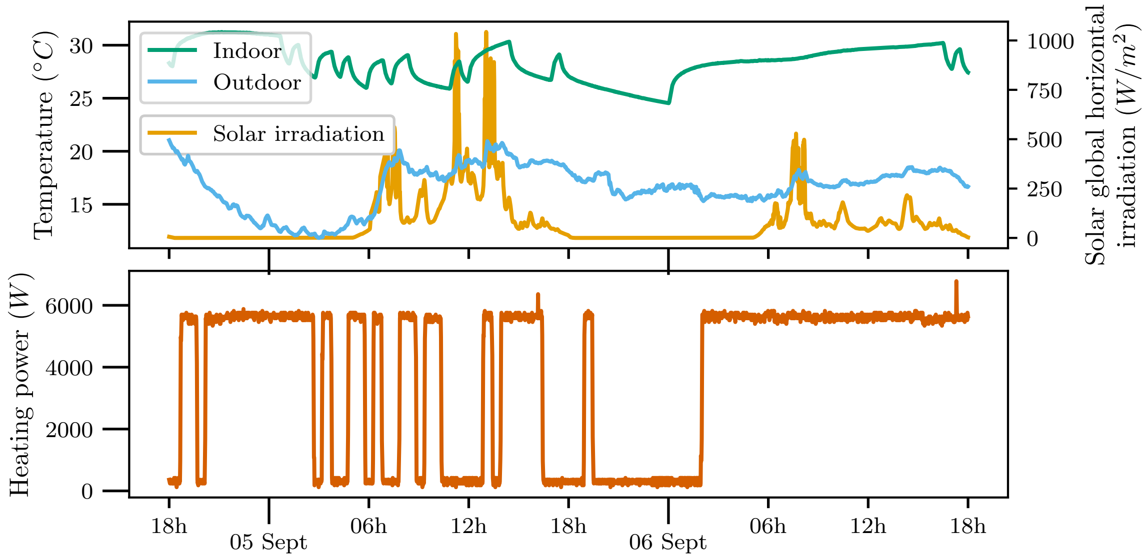


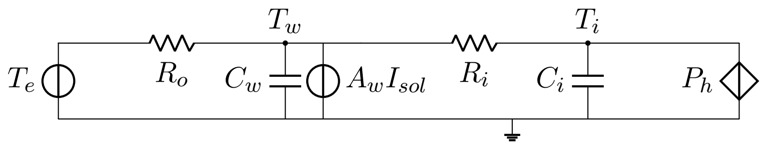
FIGURE 3. Boundary conditions in the IBB house during the PRBS test

The blind shutters were kept closed during the experiment as to limit the influence of the solar irradiation.

4.3 Calibration and learning from samplings of variable time steps

Even with suitable model structure, inadequate sampling might result in practical non identifiability : the data is insufficiently informative. In particular with RC models of the 2^{nd} order and higher, some parameters represent high frequency phenomena (Juricic et al., 2018). If one or more parameter(s) are practically not identifiable, their physical interpretability is not guaranteed.

The data collected in the INCAS house serves now as calibration data for the stochastic second order model shown in equation (4) (model R_oR_i C_wC_i A_w). The 1 minute time step measurements allow to resample the data at will. Time steps from 8 minutes up to 120 minutes are performed. Each subsequent dataset is used as data for a Bayesian calibration, performed with the Hamiltonian Monte-Carlo algorithm in the PySIP python package (Raillon et al., 2019). The estimated covariance of the stochastic term $\sigma d\omega$ of the model contributes to reflect the model error onto the parameter estimation uncertainty (Rouchier et al., 2018).



$$\begin{aligned}\begin{bmatrix}\dot{T}_w\\ \dot{T}_i\end{bmatrix} &= \begin{bmatrix}-\frac{1}{C_w}(\frac{1}{R_o}+\frac{1}{R_i}) & \frac{1}{C_wR_i}\\ \frac{1}{C_iR_o} & -\frac{1}{C_iR_i}\end{bmatrix}\begin{bmatrix}T_w\\ T_i\end{bmatrix}+\begin{bmatrix}\frac{1}{C_wR_o} & \frac{A_w}{C_w} & 0\\ 0 & 0 & \frac{1}{C_i}\end{bmatrix}\begin{bmatrix}T_e\\ I_{\text{sol}}\\ P_h\end{bmatrix}+\sigma d\omega \\ y &= \begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}T_w\\ T_i\end{bmatrix}+\epsilon\end{aligned} \quad (4)$$

Fig. 4 shows the posterior distributions of four parameters of the second order model R_oR_i C_wC_i A_w. There are roughly two cases : **(1)** the parameter has rather similar posterior distributions, regardless of the time granularity, such as parameter R_o **(2)** posterior distributions seem significantly different, as in particular for parameter C_i and to a lesser extent R_i and C_w.

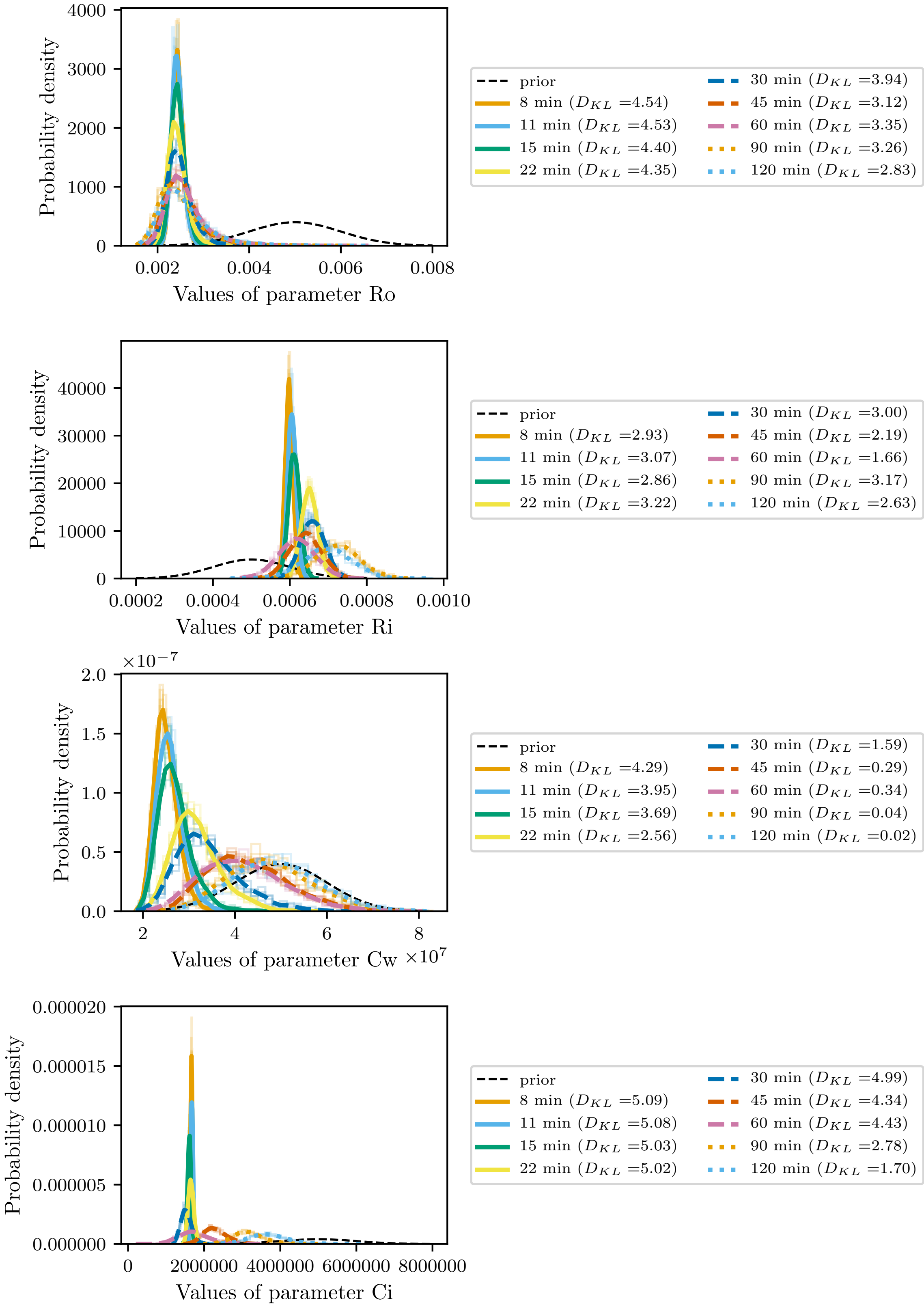


FIGURE 4. Posterior distributions of parameter R_o, R_i, C_w and C_i from model $R_oR_i\ C_wC_i\ A_w$, 48 hour data on different sampling time steps. The larger the KL-divergence, the more different posterior and prior distribution, the more the model has learnt from the data.

Parameter R_o presents distinct prior and posterior distributions. All posterior distributions looks relatively close. Confirmation is brought by the KL divergences, which show similar order of magnitudes. This would indicate that, to some extent, the estimation of R_o is insensitive to sampling with time steps shorter than 120 minutes. Shorter time steps however result in lower uncertainty in the estimation.

Contrastingly, estimations of parameter C_i or C_w show a different behaviour. Estimations from time step sampling larger than 15 minutes have significantly lower KL divergences than short time step samplings. Posterior distributions tend to show much narrower uncertainties with the 8, 11 and 15 minutes time step samplings, which contrasts with the 90 and 120 minutes sampling, closer to the prior. This all suggests that parameters C_w and C_i barely learns with low frequency sampled data.

Parameter R_i too shows a growing KL divergence with shorter time step samplings. The uncertainties are significantly narrower with the 8, 11 and 15 minutes time step samplings. There has been however in each estimation learning from the data. The amount of information gained seems just larger with a short time step sampling. Noteworthy, divergences from 90 and 120 minutes are very large, their posterior distributions are quite spread, and show higher values than the shorter time steps results. This might be the result of aliasing or also that the data is clearly insufficient for this second order model calibration. Model selection validation would make the latter clear and must in any case be performed (Madsen et al., 2015).

5 Discussion and conclusions

This paper has proposed a metric of similarity between the prior and the posterior distributions to assess the information solely gained from data in a Bayesian calibration. As an illustration of its use, the metric has been applied to assess the impact of different time granularity in the data.

An obvious criticism of using the KL divergence for that purpose is that the metric just measures the difference between two distributions. If the prior had been correctly guessed, very likely by pure and random chance or because there was high prior expert knowledge, the low KL divergence is not a sign of non identifiability. This means that no additional information was gained from the data, as the prior was already significantly informative. In this sense, the KL divergence is ***not*** an indicator of practical identifiability but indeed an indicator of additional information gained from the data alone. Interestingly, the KL divergences are not dependant on the order of magnitude of the parameters. Whether of order 10^6 or 10^{-3}, the metric seems to be consistent. This aspect is an argument in favour.

All in all, estimating the KL divergence could be used as a warning sign, provided the model had passed basic model selection. Indeed, a practically non identifiable parameter will result in a null KL divergence but a null KL divergence merely indicates no learning from the data. To waive the doubt, a second Bayesian calibration could be performed with less informative priors, in which case a practically non identifiable parameter would still yield null divergence. In practice, there is however little chance that a posterior distribution be randomly correctly guessed beforehand. It would mean that there were already high information about the parameter and that the data is ineffective to provide better knowledge about its value, which in itself is valuable insight into the data provided.

Now focusing on adequate time granularity, the results of Fig. 4 suggested that a sampling time lower than 15 minutes is preferable to reach practical identifiability of all parameters of the R_oR_i C_wC_i A_w model. This time granularity is larger than what was suggested in the work of Sicard et al. (1985), but is not surprising as adequate sampling dependent on the case study. It should still be inferred that the sampling time is fundamentally dependent on the thermal properties of the building under study but also that the risk of practical non identifiability grows with sampling of large time steps. Let us also underline the following. When shorter time step samplings imply better learning for parameters representing high frequency phenomena (R_i or

C_i), it is not detrimental to identifiability of parameters representing low frequency phenomena (R_o or to a lesser extent C_w).

Applying a similar experiment in buildings with different envelope structures and by calibrating 1^{st}, 2^{nd} and larger order RC models would certainly contribute to good practice among practitioners. At the same time, a proper physical interpretability study would bridge the gap between adequate sampling, adequate thermal models and accuracy to the actual thermal performance of a building.

6 Acknowledgments

The authors would like to thank the French National Research Agency (ANR) for funding this work through the BAYREB research project (ANR-15-CE22-0003) as much as the PROFEEL program coordinated by the Agence Qualité Construction (AQC) for providing for the experimental campaign in the INCAS houses through the SEREINE project.

References

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv preprint.*

Dawid, A. P. (1979). Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society : Series B (Methodological)*, 41(1).

Hartland, N. (2018). KL-divergence-estimators at https ://github.com/nhartland/KL-divergence-estimators.

Juricic, S., Rouchier, S., Foucquier, A., et Fraisse, G. (2018). Evaluation of the physical interpretability of calibrated building model parameters. In *7th IBPC*, Syracuse (NY).

Madsen, H. (2008). *Time series analysis.* Chapman & Hall / CRC.

Madsen, H., Bacher, P., Bauwens, G., Deconinck, A.-H., Reynders, G., Roels, S., Himpe, E., et Lethé, G. (2015). IEA EBC Annex 58 : Report of Subtask3, part 2. Technical report, IEA.

Poirier, D. J. (1998). Revising Beliefs in Nonidentified Models. *Econometric Theory*, 14(4).

Raillon, L., Rouchier, S., et Juricic, S. (2019). pySIP : an open-source tool for Bayesian inference and prediction of heat transfer in buildings. In *Congrès français de thermique*, Nantes.

Rouchier, S., Rabouille, M., et Oberlé, P. (2018). Calibration of simplified building energy models for parameter estimation and forecasting : Stochastic versus deterministic modelling. *Building and Environment*, 134.

Sahu, S. et Gelfand, A. (1999). Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association*, 94(445).

Sicard, J., Bacot, P., et Neveu, A. (1985). Analyse modale des échanges thermiques dans le bâtiment. *International Journal of Heat and Mass Transfer*, 28(1).

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics.

Wang, Q., Kulkarni, S. R., et Verdú, S. (2009). Divergence estimation for multidimensional densities via κ-nearest-neighbor distances. *IEEE Transactions on Information Theory*, 55(5).

Xie, Y. et Carlin, B. P. (2006). Measures of Bayesian learning and identifiability in hierarchical models. *Journal of Statistical Planning and Inference*, 136(10).