Bayesian approach to accounting for uncertain heating system efficiency in co-heating tests

Sarah Juricic¹, Simon Rouchier¹
¹ Univ. Savoie Mont-Blanc, CNRS, LOCIE, 73000 Chambéry, France

Abstract
Thermal characterisation of the building envelope from on-board monitoring, such as integrated co-heating, face the challenge of uncertain heating system efficiency. Whereas frequentist methods cannot account for its uncertainty when estimating heat transfer properties, this paper proposes a Bayesian approach to do so and illustrates it with data from a co-heating test under uncertainty. A Bayesian estimation of the Heat Transfer Coefficient shows a wider uncertainty than with a least squares approach, and better represents the actual state of knowledge. Bayesian modelling constitutes a promising, integrated and reasonably fast approach to deal with the many uncertainties of on-board thermal characterisation.

Key Innovations
- Bayesian framework is useful to solve an otherwise intractable non-identifiable problem
- Uncertainties are better reflected on the HTC estimation in a Bayesian approach
- Bayesian modelling very promising and reliable to exploit co-heating experiments under uncertainty

Practical Implications
Exploit co-heating experimental data in a Bayesian framework to better and easier account for uncertainties, in particular for integrated co-heating tests. HTC estimations then better reflect on the actual state of knowledge and are more trustworthy.

Introduction
To guide relevant energy conservation measures or to certify the as-built thermal performance of a building, an accurate estimation of the actual heat transfer properties of the building envelope, such as the overall Heat Transfer Coefficient (HTC, see Table 1 for nomenclature), is necessary. To alleviate the costs of many sensors and man-hours, estimations performed from on-board monitoring is a growing field of interest. The principle is to use built-in monitoring sensors for the indoor air temperature, for the heating system, etc. Data from on-board monitoring serve for the calibration of an appropriate model, which gives an estimation of the as-built HTC. The feasibility of HTC estimations from on-board monitoring has been assessed in, for example, Senave et al. (2019). The authors exploited on-board monitoring data with Auto-Regressive models with eXogeneous inputs and assessed from synthetic datasets the accuracy of the HTC estimations in different building types and with different temperature setpoint schedules. Farmer et al. (2016) present an integrated co-heating test methodology, where the heating power is not supplied by electric convectors but by the in-place heating system in the dwelling. In particular, Farmer et al. (2016) point out a major issue that has not been dealt with in

<table>
<thead>
<tr>
<th>Table 1: Nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTC</td>
</tr>
<tr>
<td>(\eta_{HS} )</td>
</tr>
<tr>
<td>(A_{sol} )</td>
</tr>
<tr>
<td>(N(\mu, \sigma))</td>
</tr>
<tr>
<td>(Inv\Gamma(\mu, \sigma))</td>
</tr>
<tr>
<td>(P_h )</td>
</tr>
<tr>
<td>(\Delta T)</td>
</tr>
<tr>
<td>(\Phi_{sol})</td>
</tr>
<tr>
<td>(\Phi_i)</td>
</tr>
<tr>
<td>(I_{sol})</td>
</tr>
</tbody>
</table>
Senave et al. (2019), who used electric heaters in their building energy model: using the heating system in the dwelling, and more generally any characterisation method that uses on-board monitoring faces the issue of accounting for the heating system efficiency.

Chambers and Oreszczyn (2019), for example, showed that it is impossible to separate the HTC and the efficiency $\eta_{HS}$, as evident in the energy balance equation in quasi steady-state conditions such as in Equation (1) (from Chambers and Oreszczyn (2019)). The total metered power demand $P_h$ is a function of the heating system efficiency $\eta_{HS}$, the HTC, the indoor outdoor temperature difference, the solar heat gains $\Phi_{sol}$ as well as other indoor heat gains $\Phi_i$. The efficiency $\eta_{HS}$ cannot therefore be overlooked because the total metered power $P_h$ is often not the power required to meet the heating demand power.

$$P_h = \frac{1}{\eta_{HS}}(HTC(T_{in} - T_{ea}) - \Phi_{sol} + \Phi_i) \quad (1)$$

Chambers and Oreszczyn (2019) addressed this issue by studying a novel indicator for energy performance called the Heat Power Loss Coefficient (HPLC) with $HPLC = HTC/\eta_{HS}$. Yet, if, for some reason, the HTC itself is the objective of the experiment, this solution is not satisfactory. A value for the heating system efficiency must therefore be provided. Farmer et al. (2016) used a measurement of the heating system efficiency by using a heat flow sensor, temperature sensors and an integrator. They found that the boiler’s efficiency of their case study A was 84% whereas 90% was expected. They added that some configurations could, however, lead to an uncertain measurement of the efficiency, such as low flow rates at small temperature differentials or in case the boiler casing is in the thermal envelope which creates additional and unmeasured heat gains. Farmer et al. (2016) did not quantify nor propagate the efficiency uncertainty on their results.

In summary, the HTC estimation from on-board monitoring faces the challenge of being fully dependent of a highly uncertain estimation of the heating system efficiency. In the worst case scenario, the heating system efficiency is not measured on-site and the data collected are exploited with an inaccurate efficiency value, which may reflect as a severely biased HTC estimation. In the best case scenario, the efficiency is measured. Although one may safely expect the measurement to be close to the actual value, some uncertainty remains and needs to be acknowledged in the estimation uncertainty.

The objective of this paper is to provide an integrated framework accounting for the uncertainty of the heating system efficiency when estimating the HTC. Instead of plainly propagating the uncertainty as a post-process, this paper shows how to consider the heating system efficiency as an unknown parameter in the problem formulation. By doing so, it concededly forms a non-identifiable and therefore intractable problem because, as stated by Chambers and Oreszczyn (2019), HTC and $\eta_{HS}$ cannot be estimated separately.

This paper demonstrates instead how data exploitation in a Bayesian framework alleviates the issue of non-identifiability. Xie and Carlin (2006) state indeed that, fundamentally speaking, any parameter with a proper posterior distribution is estimable. Poirier (1998) showed that non informative data is “inescapable” even in a Bayesian framework and that improper priors complicate the inference. Instead, with proper distributions to account for relevant prior knowledge, such as in Neath and Samaniego (1997) with a structurally non-identifiable binomial model, an inference is feasible. The Bayesian approach is therefore applied to the HTC estimation under uncertainty from an integrated co-heating test in a numerical study. The relevance of the approach is illustrated by a comparison with the common linear regression method.

**Methods**

**Overall methodology**

The target is to assess how the HTC estimation and its uncertainty are influenced by the value set for the heating system efficiency when on-board sensors are used to monitor an integrated co-heating test.

Two data processing scenarios are investigated: on the one hand under an ordinary least squares (OLS) approach and on the other hand under a Bayesian approach. Both approaches are described in the following sections.

Four cases are explored to compare both approaches and illustrate the advantages and limitations of a Bayesian approach. Table 2 summarises the four cases.

<table>
<thead>
<tr>
<th>Case (1)</th>
<th>Case (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_{HS}$ default to manufacturer value $\eta_{HS} = 0.91$</td>
<td>$\eta_{HS}$ measured $\eta_{HS} = 0.84$</td>
</tr>
<tr>
<td><strong>Case (1:LR)</strong></td>
<td><strong>Case (2:LR)</strong></td>
</tr>
<tr>
<td>Linear regression approach solved with OLS</td>
<td>Linear regression approach solved with OLS</td>
</tr>
<tr>
<td><strong>Case (1:B)</strong></td>
<td><strong>Case (2:B)</strong></td>
</tr>
<tr>
<td>Bayesian approach probability distribution $\eta_{HS}$ mode $\approx 0.91$ large variance</td>
<td>Bayesian approach probability distribution $\eta_{HS}$ mode $\approx 0.84$ narrow variance</td>
</tr>
</tbody>
</table>

The distinction between cases (1) and (2) follows the observation made by Farmer et al. (2016) when the heating system efficiency measured was found lower
than the manufacturer declared efficiency. The OLS approach uses therefore either the declared or the measured value. As for the Bayesian approach, as mentioned below, it declares the parameter $\eta_{HS}$ as a probability distribution with large variance when the efficiency is unknown. The variance is small but non-zero when the efficiency is supposed measured to account for the measurement uncertainties as underlined in Farmer et al. (2016) and shown in Nikula et al. (2016), who found a time dependent variability of the efficiency of a few percent.

**Co-heating protocol and data description**

Co-heating tests aim at estimating the HTC by assuming that the whole building energy balance has, under particular experimental conditions, a steady-state behaviour, as described in Equation (2) (Farmer et al., 2016).

$$HTC = (\eta_{HS}P_h + \Phi_{sol})/\Delta T$$  

(2)

The experimental protocol therefore requires quasi steady-state conditions. This is achieved by scheduling the indoor air temperature at a constant temperature, for example 25 °C. The particularity of an integrated co-heating test is to use the on-site heating system and temperature controllers (Farmer et al., 2016). In addition, the co-heating test is performed on a sufficient period of time, i.e. several weeks, to decrease the effect of thermal dynamics due to thermal storage and inertia. The data collected are averaged daily on 6 a.m.–6 a.m. intervals, in agreement with (Bauwens and Roels, 2014). This experimental protocol relies concededly on many assumptions and simplifications of the thermal physics in the building, even with quasi stationary induced indoor conditions. For a discussion, we refer to Bauwens and Roels (2014) who broadly elaborated on these assumptions.

The data processed in this paper are collected during a four-week co-heating test in a two-storey house of the INCAS platform located on the INES site, in Le Bourget du Lac (France). The wooden-framed house is highly insulated and airtight. This test was concededly not an integrated co-heating test but the data remain relevant to this methodology. As illustrated by Figure 1, the co-heating test was performed in late autumn 2020. Outdoor temperatures vary in a 0–12 °C range and the global solar irradiation indicates both sunny and cloudy days during the experiment. The indoor air temperature was set at 27 °C.

Figure 2 summarises the dataset when the data are averaged daily. There is one particular outlier visible on top of the Figure: the heating system had been down for a few hours after a power outage on November 30th morning. This day has therefore been excluded from the analysis.

**Data exploitation by least squares**

Co-heating test data are usually modelled by a multiple linear regression which considers $P_h$ as the dependent variable whereas $\Delta T$ and $I_{sol}$ are the independent variables (Bauwens and Roels, 2014). As such, two parameters are estimated: the HTC and the solar aperture $A_{sol}$. Then, Farmer et al. (2016) as well as Bauwens and Roels (2014) use a graphical trick and plot $P_h$ corrected by $A_{sol}I_{sol}$ which produces a scatter plot where the HTC is the slope. This trick is not used in this paper. Instead, the HTC and $A_{sol}$ estimates and their uncertainties obtained from the OLS approach will be under investigation.

The ordinary least squares approach minimises the sum of square residuals, where the residuals are the difference between prediction of $P_h$ by Equation (1) and the measured heating power. This approach is frequentist: it assumes no prior knowledge on the parameters and supposes a Gaussian posterior distribution of the parameters.

As mentioned earlier, the HTC estimation will directly depend on the value set for the heating system efficiency $\eta_{HS}$. Figure 3 illustrates this dependency by examining the sum of squared residuals (SSR) profile as a function of the heating system efficiency $\eta_{HS}$ and of the HTC value similarly to a profile likelihood (Kreutz et al., 2013). Each coloured point is the minimised SSR for all $A_{sol}$.
The co-heating protocol is designed to enhance the heating system efficiency \( \eta_{HS} \) from the linear regression model described by Equation (1) since this model is structurally non-identifiable; it has infinitely many solutions.

In a Bayesian approach, this problem is no longer non-identifiable. The principle is to consider all unknown parameters as probability distributions. The parameter estimates conditioned by the data observed are described by the posterior distribution, given Bayes rule here in Equation (3). The posterior distribution \( p(\theta | y) \) is a function of prior expert knowledge \( p(\theta) \) and knowledge gained from measurements \( p(y | \theta) \). This means that regardless of the informativeness of the data collected, all parameter values will be at least defined by their prior distribution. If there is no information to gain from the data collected, prior and posterior distributions will be identical (Poirier, 1998). This implies using a proper prior distribution for each parameter for the problem to be well defined.

\[
p(\theta | y) = \frac{p(y | \theta) \cdot p(\theta)}{p(y)}
\]

Equations (4). The measured heating power \( P_h \) is a function of two observed independent variables \( \Delta T \) and \( I_{sol} \) as well as three unknown parameters \( HTC, \ A_{sol} \) and \( \eta_{HS} \). Gelman et al. (2017) states that the choice of prior is all the more relevant that the data is rich and the samples abundant. The co-heating protocol is designed to enhance the relationship between \( P_h \) and \( \Delta T \). The data is then expected to be rich information about \( HTC \). A wide normal prior distribution seems relevant. For parameter \( A_{sol} \), however, informativeness is less certain. An inverse gamma prior distribution should avoid negative values while setting a poorly informative prior. Appropriateness of this prior will be discussed lateron. The problem formulation also accounts for a measurement error term \( \epsilon \) that also has a poorly informative inverse gamma prior.

For \( \eta_{HS} \) to remain physically representative, the prior distribution of \( \eta_{HS} \) should only take values between 0 and 1, should have a mode around the target efficiency but need not be symmetrical. Figure 4 shows three different prior distributions for \( \eta_{HS} \). From the figure, it can be inferred that a normal distribution is not an optimal choice, because it will set non-zero probabilities on efficiency values greater than 1. Conversely, the inverse gamma distribution is defined over the positive real line and the logit distribution over \([0,1]\). Both are asymmetrical, which is actually a better representation of a rational prior expert knowledge: if the manufacturer declares the efficiency to be 0.91, there is very little chance that the actual efficiency is higher and larger chance that it is somewhat lower. In this paper, the inverse gamma distribution was then chosen not only because it allows for probable values down to 0.7 but also because it has nice conjugate prior properties (Gelman, 2006).

\[
\begin{align*}
HTC & \sim N(50, 15) \\
A_{sol} & \sim Invg(\mu = 1.0, \sigma = 0.6) \\
\eta & \sim Invg(\mu = 0.11, \sigma = 0.2) \\
\eta_{HS} & = 1 - \eta \\
P_h & = \frac{HTC}{\eta_{HS}}\Delta T - \frac{A_{sol}}{\eta_{HS}}I_{sol} \\
\epsilon & \sim Invg(\mu = 50, \sigma = 30) \\
y & \sim N(P_h, \epsilon)
\end{align*}
\]

As the posterior distribution cannot be analytically computed, it needs to be numerically approached. To do so, an appropriate algorithm draws samples from the posterior distribution. Common algorithms, such as the Metropolis-Hastings algorithm, create a Markov chain which explores the parameter space until convergence towards the most credible parameter values, henceforth sampling the posterior distribution. The first exploratory iterations, in the burn-in period, are discarded to only keep the samples from the posterior distribution. Good
practice requires to provide several chains and to verify convergence towards a common posterior distribution. Finally, let us recall that in a Bayesian approach, the uncertainty is translated into a credible interval, and not a confidence interval such as with frequentist methods. This model was implemented with the pyme3 python package (Salvatier et al., 2016) and the code is available upon request to the authors. Four chains were run for each model, with 4000 iterations for each chain, among which 2000 iterations for burn-in to be on the safe side. The sampler used is the No U-Turn Sampler, proven to be efficient (Salvatier et al., 2016).

Results

This section presents and compares the estimations of parameters HTC, A{\text{sol}} and \( \eta_{HS} \) obtained with the OLS and Bayesian approaches. Table 3 in particular shows that the regression coefficient of both linear regressions is above 0.99 and can be considered satisfactory. This means that the data are rather well described by the regression model.

Table 3: Cases (1:LR) and (2:LR) linear regression results

<table>
<thead>
<tr>
<th></th>
<th>Case (1:LR)</th>
<th>Case (2:LR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta_{HS} )</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.994</td>
<td>0.994</td>
</tr>
<tr>
<td>HTC (W/K)</td>
<td>56.5</td>
<td>52.1</td>
</tr>
<tr>
<td>95 % C.I. (W/K)</td>
<td>[52.8,60.1]</td>
<td>[48.8,55.5]</td>
</tr>
<tr>
<td>A{\text{sol}} (m²)</td>
<td>0.72</td>
<td>0.67</td>
</tr>
<tr>
<td>95 % C.I. (m²)</td>
<td>[-0.74,2.19]</td>
<td>[-0.68,2.02]</td>
</tr>
</tbody>
</table>

Table 4 describes the diagnostics of both Bayesian models. The Gelman-Rubin metric \( R \) scores lower than 1.005 for all parameters in both cases, which gives great confidence that all 4 random walks in both cases have converged. The effective sample size (ESS) is in all cases greater than 1013, which indicates that the random walks have low auto-correlation and is therefore again an indicator of convergence. Had this value been lower, more samples might have been necessary. Noteworthy how the ESS is significantly lower for case (1:B) compared with case (2:B). This does not come as a surprise since the models are structurally non identifiable and case (1:B) propagates large uncertainties on \( \eta_{HS} \). It is therefore difficult for the sampler to efficiently explore the parameter space.

Table 4: Cases (1:B) and (2:B) calibration diagnostics

<table>
<thead>
<tr>
<th></th>
<th>Case (1:B)</th>
<th>Case (2:B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta_{HS} )</td>
<td>Prior mode 0.91</td>
<td>Prior mode 0.84</td>
</tr>
<tr>
<td>( R(HTC) )</td>
<td>1.004</td>
<td>1.000</td>
</tr>
<tr>
<td>( R(A{\text{sol}}) )</td>
<td>1.005</td>
<td>1.001</td>
</tr>
<tr>
<td>( R(\eta_{HS}) )</td>
<td>1.002</td>
<td>1.001</td>
</tr>
<tr>
<td>( R(\epsilon) )</td>
<td>1.001</td>
<td>1.000</td>
</tr>
<tr>
<td>ESS (HTC)</td>
<td>1153</td>
<td>3314</td>
</tr>
<tr>
<td>ESS (A{\text{sol}})</td>
<td>1013</td>
<td>3556</td>
</tr>
<tr>
<td>ESS (( \eta_{HS} ))</td>
<td>3153</td>
<td>4459</td>
</tr>
<tr>
<td>ESS (( \epsilon ))</td>
<td>2899</td>
<td>5146</td>
</tr>
</tbody>
</table>

The HTC estimations are shown in Figure 5. Because a linear regression is based on the hypothesis that the parameters have an asymptotic Gaussian behaviour, they are shown as such in the Figure. In yellow, the linear regression (1:LR) and the Bayesian (1:B) approaches have different outcomes. Their modes are significantly similar but the posterior distribution from the Bayesian modelling is much wider and shows a larger uncertainty. As expected, the posterior distribution is asymmetrical and reflects on the asymmetry of the prior distribution of the efficiency \( \eta_{HS} \).

This difference in uncertainty between both yellow-coloured cases (1:LR) and (1:B) is particularly relevant regarding the number of plausible values on the left hand side of the distributions. The OLS approach (1:LR) sets very low probability for HTC values lower than 52.8 W/K whereas the Bayesian approach (1:B) sets credible values down to 47 W/K. In other words, the common interval with the blue cases is much lower for the OLS approach than for the Bayesian approach, whereas the blue case estimations are much more likely to contain the actual HTC because they are based on a measured efficiency. The credible interval of the Bayesian approach is therefore much more likely to contain the actual HTC value too, which makes it a preferable option for data processing under uncertainty.

The comparison in Figure 5 of case (2:LR) with case (2:B), in blue, shows that the regression estimation and the Bayesian posterior distribution are almost identical. A slight shift towards higher HTC values is, however, visible, and is likely to be due to the parameterisation of the inverse gamma prior distribution, which mode is not perfectly centred around 0.84. The advantage of Bayesian modelling in this case to reflect on the state of uncertainty is less
evident and will be discussed in the following section.

Figure 6 shows the prior and posterior distributions of the heating system efficiency $\eta_{HS}$. In both cases (1:B) and (2:B), as was expected in line with Poirier (1998), the prior and posterior distributions are identical which confirms that there is no learning possible from the data. The only knowledge available after calibration is inherited from the prior distribution.

Figure 7 shows the linear regression-based estimations of the solar aperture parameter $A_{sol}$ and its posterior distributions when estimated in a Bayesian approach. In all cases, it is evident that the data are poorly informative for this parameter. The OLS results suggest that the parameter is unlikely to be significant, seen that $\theta$ is in the confidence interval. Estimations in a Bayesian approach show that there is little learning from the data: prior and posterior distributions are relatively close to one another. Both posterior distributions show a common mode towards the left of the mode of the prior and values above 2 are no longer considered likely. The uncertainty remains significantly large. In addition, although chosen to be poorly informative, it is evident that the prior influences the posterior by setting very low probability on values close to 0. It remains clear that there is almost no learning from the data since prior and posterior distributions are so close one to another. Yet, the parameterisation and choice of the prior distribution does not seem appropriate here and calls for further work. Overall, it can be concluded that parameter $A_{sol}$ is barely practically identifiable. This result is consistent with the data shown in Figure 2 where a clear linear dependency on the solar global irradiation would have been evident from a colour gradation from top to bottom. Instead, the data show a poor gradation and many outliers, hence a poorly identifiable solar aperture parameter.

Discussion

The uncertainty of not accurately knowing the heating system efficiency during an integrated co-heating test is not accounted for in the usual linear regression model solved by OLS, because it becomes an intractable numerical problem. The results of the HTC estimation have shown that basing a regression on a manufacturer value or on a measured value produces significantly different results. This means that a regression based on an erroneous efficiency value might produce an HTC estimation with a deceptive narrow confidence interval, which might not include the actual HTC value.

Conversely, a Bayesian approach makes it possible to account for the heating system efficiency in the model itself. This is done through a probability distribution, chosen to be representative of the actual prior information available on the efficiency parameter, whether it be measured or not. The results have shown that the uncertainty of the HTC estimation is wider and thus a better representation of the actual uncertainty in this experiment. This is a major outcome as different HTC estimations do no always share a common confidence intervals when the experiment is reproduced, as was the case in (Farmer et al., 2016).

The difference between linear regression and Bayesian modelling seemed less significant for the measured efficiency case (2). It merely means that the most important part of the final uncertainty is not caused by the uncertainty of the efficiency but more likely by measurement or model errors and possibly missing independent and unmeasured variables. In this sense, a Bayesian approach makes it easier to discern where the next efforts need to be made to lower the remaining uncertainties.

Least squares or any frequentist method results may naturally be a posteriori supplemented by an uncertainty analysis to propagate the parameter uncertainties. But this process demands an additional effort in modelling and coding whereas the Bayesian approach shown in this paper is an integrated framework to do so.

Noteworthy is the fact that Bayesian modelling and calibration is not concerned by the issue of structural non-identifiability, whereas it becomes for frequentist methods an intractable problem. Non-identifiable parameters will in any case not gain any information from the data collected, but their uncertainty will be reflected on the other identifiable parameters. A comparison between prior and posterior distributions then gives insight in the amount of information learnt from the data for each parameter, in particular when the prior is chosen non-informative.

These outcomes can be considered as a refinement of the findings of Raue et al. (2013), who found structural non-identifiability to be a major hinder for Bayesian calibration. They considered indeed that non-identifiable parameters have improper posterior distributions (non-integrable), given that their prior distributions had been defined as flat. This case indeed causes the sampling algorithm to very likely fail converging. In this paper however, the non-identifiability of parameter $\eta_{HS}$ is counterbalanced by its well-defined prior. Structural non-identifiability in a Bayesian approach should therefore not be considered as an issue per se, provided the non-identifiable parameters have a proper prior distribution, thanks to adequate prior information.

As for the solar aperture parameter $A_{sol}$, let us elaborate on reasons why it was found practically non-identifiable. First the building is very well isolated and all shading devices were closed during the experiment. In addition, the test was performed
in late autumn, thus during a period with low solar irradiation. The effect of sunny days on the heating power demand is therefore low and could explain why \( A_{\text{sol}} \) was barely identifiable. A deeper investigation on appropriate prior distributions for parameters susceptible to be non-identifiable should be conducted to provide more useful guidelines.

The results of this paper should encourage researchers and practitioners to adopt Bayesian modelling in their data exploitation process. In particular for the exploitation of co-heating experiments, the implementation in python is straightforward and the numerical cost acceptable. In this paper, one Bayesian calibration was based in total on 16 000 samples (4 chains of 4000 iterations each) and takes without parallelisation 30–45 seconds on a single i5-6200U CPU @ 2.30GHz core.

In the future, research could focus on integrating latent variables in a Bayesian approach to reflect on the uncertainties of the independent variable measurements. In the case of an integrated co-heating test, Farmer et al. (2016) showed the importance of using fans in the dwelling to enhance the indoor temperature homogeneity. Stating that indoor air temperatures are perfectly identical across a building during a co-heating test is indeed a strong hypothesis, which could easily be accounted for in a Bayesian model. The solar irradiation measurement too is considered accurate in a classical multivariate regression solved by OLS, whereas its uncertainty reaches up to 10 % error for daily values, and up to 20 % error for hourly values when devices with moderate performance are used (NREL, 2015). Furthermore, if the solar irradiation is not measured on site, as can be the case since it requires an expansive device, the uncertainty will be larger and needs to be accounted for in the model. Such additional uncertainties, along with
temperature dependent heating system efficiencies, can individually be accounted for by generalised least squares (NREL, 1988) or with generalised linear mixed models. Bayesian modelling can however simultaneously deal with loose assumptions, non-identifiable parameters and latent random variables in a unified approach, which is precisely why it is a promising option.

Another research prospect would be to investigate feasibility of Bayesian modelling of structurally non-identifiable dynamic models, such as RC models. On-board monitoring makes it indeed possible to collect data with a 5 minute time step. Time series modelling may thus rely on much more data, at much shorter time steps and can be tricky to calibrate when the data are poorly or only moderately informative. Accounting for the uncertainty of the heating system efficiency and that of the measured variables in an integrated framework becomes all the more relevant. The choice of prior distributions and the calibration convergence might, however, not be as straightforward as in this paper for regression modelling in a Bayesian framework and call therefore for further investigation.

Conclusion

This paper has identified the challenge of an unknown or uncertain heating system efficiency to exploit co-heating experiment data from on-board monitoring and has proposed an integrated Bayesian framework to account for its uncertainty.

The novelty of this work was therefore to tackle an otherwise intractable numerical problem in a Bayesian framework. The results show that this approach successfully provides estimates of the model parameters while reflecting appropriately on the uncertainty of the heating system efficiency.

Although the proposed method is only illustrated by the exploitation of an integrated co-heating test, it is expected to be equally applicable on other uncertainties in a co-heating experiment and on other methods for on-board characterisation. Applying Bayesian modelling more widely could significantly contribute to trustworthy on-board characterisation methods.

Code availability

The code embedded in a notebook with comments can be made available upon request to the authors.

References


