Continuous calibration of a digital twin; a particle filter approach

Rebecca Ward
Data-centric Engineering, The Alan Turing Institute
The British Library, 96 Euston Road
NW1 2DB, London UK
rward@turing.ac.uk

Ruchi Choudhary
Data-centric Engineering, The Alan Turing Institute
The British Library, 96 Euston Road
NW1 2DB, London UK
rchoudhary@turing.ac.uk

Alastair Gregory
Data-centric Engineering, The Alan Turing Institute
The British Library, 96 Euston Road
NW1 2DB, London UK
alsgregory12@gmail.com

Mark Girolami
Data-centric Engineering, The Alan Turing Institute
The British Library, 96 Euston Road
NW1 2DB, London UK
mgirolami@turing.ac.uk

Abstract

Assimilation of continuously streamed monitored data is an essential component of a digital twin. The assimilated data are then used to ensure the digital twin is a true representation of the monitored system; one way this is achieved is by calibration of simulation models, whether data-derived or physics-based. Traditional manual calibration is not time-efficient in this context; new methods are required for continuous calibration. In this paper, a particle filter methodology for continuous calibration of the physics-based model element of a digital twin is presented and applied to an example of an underground farm. The results are compared against static Bayesian calibration and are shown to give insight into the time variation of dynamically varying model parameters.

1 Introduction

The technological advancement and drop in price of monitoring equipment has led to a boom in availability of monitored data across industrial fields as diverse as aviation, manufacturing and the built environment. This facilitates the development of digital twin technology. The specification of what constitutes a digital twin is still evolving, but the essence comprises a monitored system together with a computational model of the system which demonstrably replicates the system behaviour in

Workshop on machine learning for engineering modeling, simulation and design @ NeurIPS 2020

the context of interest (Worden et al. (2020)). The computational model might be data-derived or, as is more usual in engineering systems, physics-based; the intention is that the computational model can give information that may not be easily accessible from the system directly and can be used to explore performance when it is impractical to run physical tests. The greatest potential for digital twinning perhaps lies in systems which are continuously operational generating live streamed data, in which case the computational model can be simulated in (close to) real-time to advise changes to operational parameters for improved efficiency (Madni et al. (2019)).

In order to simulate beyond the realm of the data it is essential to calibrate models to ensure that the model outputs match the available data where available. Calibration can be performed manually but manual calibration is time consuming and is not always practical, particularly in the situation where model parameters are not static but dynamic; instead an automated calibration process is required that forms an integral part of the continuously operational digital twin.

Bayesian calibration (BC) offers a formal way to combine prior knowledge and measured data with model predictions to improve the model. The Bayesian framework proposed by Kennedy and O'Hagan (KOH) (Kennedy and O'Hagan (2001)) has been explored in some depth for calibration of simulation models (Higdon et al. (2004); Heo et al. (2012)), specifically for the identification of static model parameters and the quantification of uncertainty. There has also been some exploration of its extension to dynamically varying parameters. For example, Chong et al. (2019) used the KOH formulation in conjunction with data from building energy management systems to continuously calibrate a building energy model, updating the model every month, demonstrating that prediction of future performance is improved with continuous calibration. The KOH formulation of Bayesian calibration is useful as it separates the different sources of errors between the model and the data and allows for them explicitly. However BC is computationally expensive, increasingly so as the numbers of calibration parameters and data points increase, as the parameter space is typically explored using some form of Monte-Carlo sampling where at each new point the relative posterior density of the old and new points is compared to determine whether the new point is accepted (Riddle and Muehleisen (2014)). This has implications for use in continuous calibration over short timescales - the run-time of the calibration must be shorter than the time interval between acquisition of new data points.

An alternative approach that potentially offers a more time-efficient solution to the problem of continuous calibration is Particle Filtering (PF). The aim of this study is to demonstrate that the PF approach can offer a calibration mechanism for a model that is simulating in real time. The model is a digital replica of an underground farm in London. A relatively simple physics-based simulation model of the farm has been developed that calculates temperature and relative humidity as a function of external weather conditions and farm operational strategies. An extensive programme of monitoring has also been carried out, so there are data available for calibrating the model to better represent the farm environment. Using both calibration approaches, we explore the potential to infer uncertain model parameter values and to track their dynamic variations in time.

2 Calibration approach

In both cases, the calibration approach make use of Bayes' formula.

$$P(Y|y(x)) = \frac{P(y(x)|Y).P(Y)}{P(y(x))} \tag{1}$$

The posterior probability of the parameter value Y, given the observed data point $y(x)$ is equal to the *likelihood* of the observed data point, $P(y(x)|Y)$, multiplied by the probability of Y before making the observation - the prior probability, $P(Y)$ - all normalised by the factor $P(y(x))$.

2.1 Bayesian calibration

In the Bayesian calibration formulation proposed by Kennedy and O'Hagan (Kennedy and O'Hagan (2001)), a numerical simulator $\eta(x, \theta^*)$ can be related to field observations $y(x)$ by the following equation:

$$y(x) = \eta(x, \theta^*) + \delta(x) + \epsilon \tag{2}$$

where x are observable inputs into the model, for example location of sensor or time of sensing, and θ^* represents the true but unknown values of the parameters θ which characterise the model. This formulation inherently expects a discrepancy, or bias, between the model and reality accounted for by $\delta(x)$. ϵ represents the observation error. Calibration of the model aims to identify the parameters θ^*; whereas traditional calibration approaches require multiple runs of the computer simulation with systematic variation of the input parameters and subsequent identification of the combination of parameters that gives the closest match to reality, the KOH framework offers a more efficient way to identify the uncertainty associated with the calibration parameters. Multiple runs of the computer simulation are still performed, but rather than using an exhaustive iterative approach, it is common practice to use an emulator to map the model inputs to the outputs. Gaussian process models are commonly used as the basis for the emulator, with separate models used for the simulator $\eta(x, \theta^*)$ and the discrepancy term $\delta(x)$ in the above equation.

The practical application of the Kennedy O'Hagan (KOH) approach to calibration of building energy simulation models is described in detail by Chong and Menberg (2018). The procedure is to compare observations against model outputs derived from computer simulations using the range of prior estimates of uncertain parameters. By exploring the likelihood of the observations given the simulation model outputs, the posterior distribution of each uncertain parameter is derived.

2.2 Particle filtering

Bayesian calibration has at its heart an assumption of stationarity i.e. the uncertain parameters, θ, are assumed to be constant over the range of the data. However, we desire to continuously calibrate the physics-based model so that the dynamic variation of parameters values can be inferred. For this, we propose using a particle filtering approach. This works sequentially by using the posterior distributions from one timestep as the prior distributions for the next, thereby updating estimates for each parameter based on the observations given. A shift in the posterior range of the parameter can be detected because the likelihoods of the particles is updated accordingly. Similar to the KOH method, the calibration process proceeds with multiple runs of the computer simulation as an initial step and an emulator of the computer model is used for the subsequent updates.

The procedure for the particle filtering approach is as follows: consider the computer model $\eta(\theta)$, where θ are parameters of the model. We conjecture that observed data are given by $y \sim \rho\eta + \delta + \epsilon$, where ρ is a scaling parameter, δ is a mean-zero Gaussian process representing the mismatch between the model and the data and ϵ is the measurement error. We assume that the smoothness of the emulator and model mismatch are determined by a lengthscale, l.

1. Start by sampling N different particles of the hyperparameters l, ρ and model parameters θ from the prior distributions $p(l), p(\rho)$ and $p(\theta)$. Denote these particles $\{l_j\}_{j=1:N}$, $\{\rho_j\}_{j=1:N}$ and $\{\theta_j\}_{j=1:N}$.
2. Obtain the simulated computer model outputs, d_i, at the coordinates, calibration parameters X_i^d, t_i. Also consider the observations Y_i at the (sensor) coordinates, (unknown) calibration parameters X_i^Y, θ.
3. Create the covariance matrices (for each of the particles), K_j, using the covariance functions k_η associated with the computer model emulator, η and k_Y associated with the model-data mismatch δ. These are conditioned on the lengthscale hyperparameter, l.
4. For each of the particles compute the marginal likelihoods $\xi_j = p([d_i, Y_i]^T; 0, K_j)$. This will be used to update the posterior of the parameters in the next step.
5. Compute the weights $w_j = \xi_j / \Sigma_{j=1}^N \xi_j$. Then resample the particle ensembles so that each resampled particle has even weight. This assigns a particle filtering step (works with non-Gaussian prior distributions) to finding an empirical approximation to the posterior of l, ρ and θ.
6. Use the N posterior approximations as prior particles for the next iteration and repeat.

3 Digital twin of an underground farm

The calibration concerns the simulation model of a hydroponic farm operating in a disused tunnel in London, UK. The farm grow salad crops using LED lighting while making use of the relatively stable thermal conditions in the tunnel. It is important to maintain environmental conditions at optimum levels for plant growth and to this end the space is monitored for temperature, relative humidity and CO_2 levels. There are two ventilation shafts along the tunnels which exchange the internal and external air and hence act to exchange heat and moisture; these are controlled using fans that are controlled via manual dials. The primary heat source is the LED lights which operate on a fixed daily regime.

3.1 Physics-based model

A simple physics-based model has been developed to represent a 1D slice through the central section of the farm. The model calculates temperature and relative humidity of the tunnel air as a function of time by solving heat and mass exchange equations pertinent to the tunnel geometry, subject to the temperature, moisture content and CO_2 concentration of the incoming ventilation air and the deep soil temperature. The moisture content and temperature are inextricably linked as changes to the moisture content of the air are associated with latent heat transfer. Hence the heat and mass balance equations are solved in parallel to determine the temperatures and relative humidities within the tunnel.

The model is also dependent on the operational conditions in the farm such as ventilation rate and internal air speed. These parameters are quite uncertain, partly because the farm has been constructed in a derelict space making use of the existing tunnel ventilation system for which there is only partial information available, and partly because these parameters are difficult to measure. In addition, the simplicity of the model means that parameters represent the combined effect of different physical components e.g. the combined effect of infiltration and controlled ventilation are simulated as a single ventilation rate. Calibration of the model and estimation of the parameters therefore becomes an important step in improving the model accuracy. But the parameters are not constant and so it becomes necessary to re-calibrate when parameters change. The problem is that it is not always obvious when parameters are changing - some changes, such as manual alterations to the ventilation settings are known, but the impact of the alteration in terms of the change to the ventilation rate is not a linear function of the dial settings.

This need to re-calibrate, together with the availability of continuously monitored data, makes development of a sequential calibration process desirable.

3.1.1 Sensitivity study

The most important quantities from the viewpoint of the growers are the air temperature and the relative humidity. Plants grow best in optimal temperature conditions and if the the relative humidity gets too high, growth of mould can occur affecting crop yield. So the calibration of the model focuses on these as the most important outputs.

For calibration of the model, the first step is to ascertain which parameters have the most significant impact on the model output and are therefore most important to calibrate. A sensitivity study has been carried out using the Morris method (Menberg et al. (2016)) to identify the relative significance of 8 input parameters. The results of the sensitivity study suggest that the most influential parameter for temperature is the ventilation rate, ACH, whereas the internal air speed, IAS, has the most significant impact on relative humidity. Relative humidity is itself a function of temperature, so rather than calibrating the model separately for both parameters, it suffices to calibrate for relative humidity alone.

3.2 Observations (monitored data)

In order to explore the application of the particle filter approach, a subset of monitored data has been selected for which it is known that there were no significant changes to operation except to the ventilation settings. This is important as we would like to understand to what extent the models are capable of identifying the change in the ventilation setting.

The relative humidity data are illustrated in Figure 1. As aforementioned, the primary heat input to the farm comes from operation of the LED lights, which are on and off for a fixed period of time each day, mimicking a diurnal cycle. While the data are monitored hourly we do not expect significant change during each period for which the lights are on or off, so we have reduced the data to 2 points per 24 hour cycle, one during the lights on period and one when the lights are off. Each black dot

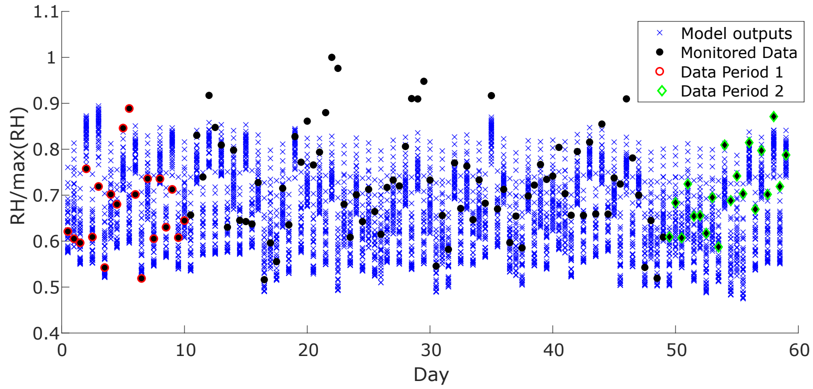

Figure 1: Observations (monitored data) plotted against model executions over the entire test period. Red and green dots indicate observations used for the 1st and 2nd KOH calibration respectively.

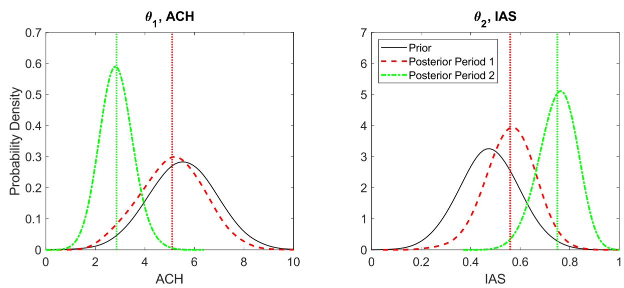

Figure 2: Prior versus posterior estimates of the uncertain model parameters estimated using KOH approach

in Figure 1 is a data point from the reduced dataset, and we have selected two separate periods of 20 data points shown in red and green for analysis using the static KOH BC approach. Period 1 (red) is chosen as it lies before the change in ventilation settings, and Period 2 (green) after the change. It is not feasible to use all 118 data points for the KOH approach as the run-time would be too great. Also shown in the figure are the outputs from the model runs executed for the calibration process (blue xs). There are several observations which lie outwith the range of the computer model executions. Specifically, observations of high relative humidity that are not predicted by the model. This is because the model does not represent the irrigation of the planted trays in great detail, and nor the process of washing the harvested produce. These operations are represented by a single mean saturation level that does not account for fluctuations that likely result in air moisture content.

As a first step, the KOH approach has been used to estimate the calibration parameters for Periods 1 and 2, using the relative humidity data as illustrated in Figure 1, together with information regarding the operational status of the lights - on/off - and the moisture content of the external air. Figure 2 shows the posterior distributions for both parameters for both time periods, with the first time period shown in red and the second in green, with the prior (normal) distribution in black. The posterior distributions suggest that the ventilation rate drops from a mean value of 5.1 in Period 1 to a mean value of 2.85 ACH in Period 2, whereas the internal air speed increases from 0.56 to 0.75 m/s. These values are plausible as lower ventilation rates and higher internal air speeds both give rise to a higher relative humidity, and the mean relative humidity increases by 5% from Period 1 to Period 2.

The particle filter approach has been run for the entire period, a total of 118 data points. The mean of the two uncertain parameters are shown in Figure 3 as a black solid line. Also shown are the mean parameter values inferred from the two static KOH calibrations, indicated on the figures as red dashed and green dashed lines over the periods of interest. Also indicated on Figure 3 are the settings of the fans of the two ventilation shafts over this period. Ventilation setting 2 is constant over Period 1 then drops from 2 to 0.5 ACH, whereas ventilation setting 1 is higher, at 4.5 ACH, during the initial period, and drops later to a value of 2 ACH.

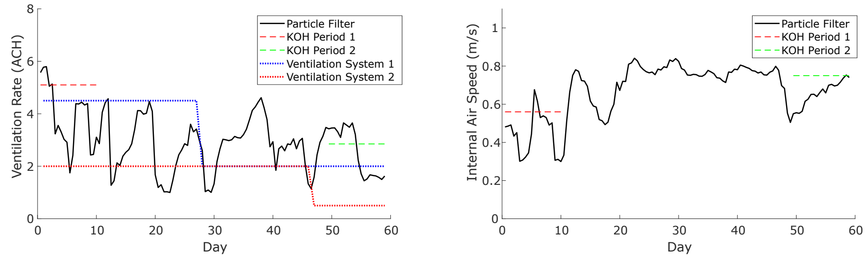

Figure 3: PF estimates of variation in mean parameter values over time compared against mean posterior estimates of KOH approach

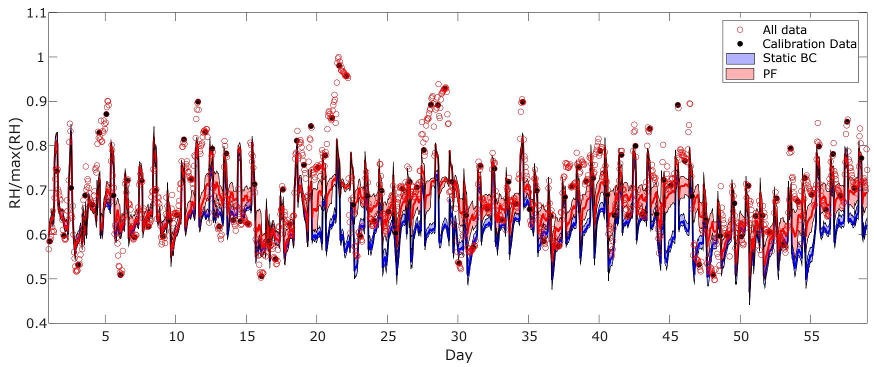

Figure 4: Relative Humidity simulation output for complete time period using PF posterior particle parameter values and KOH posterior parameter estimates for Period 1

The mean parameter values inferred from the particle filter suggest a much greater variation of the parameters than might be expected from the static settings of the ventilation system, although there is an observable downward trend to the ventilation rate estimated from the PF, in agreement with the KOH calibration and with the observed system settings. Equally, the PF approach suggests a higher internal air speed at the end of the test period than at the beginning, again in agreement with the KOH model. The PF does not identify step changes in ventilation rate corresponding to changes in the dial settings: this is due to the ventilation rate assumed in the model encompassing all air exchange, not just that due to the ventilation system.

What is particularly noticeable when comparing Figures 3 and 1 is that periods of low ventilation rate and high internal air speed, just before day 30 for example, correspond to high values of monitored relative humidity. At this point the mean values of the posterior distributions of the particles for ventilation rate and internal air speed are at the limits we have specified as being realistic; we do not expect ventilation rates lower than 1ACH or air speeds higher than 0.8m/s.

The aim in calibration is to identify parameter values that enable the simple model to simulate the real behaviour more accurately. As a final step, the model has been run using parameter values for ACH and IAS inferred from the PF process, and also from the KOH model calibration for Period 1. The results are shown in Figure 4. The entire set of monitored data is indicated in addition to the subset used for the calibration. The continuously varying parameters of the PF results show a better agreement with the monitored data, particularly as time progresses, as would be expected.

4 Conclusions

The above results have indicated that the particle filter approach is suitable for continuous calibration of the parameters of a simulation model. The static KOH calibration gives a good approximation to parameter values where these can be assumed to be constant over time, but the PF approach potentially can give more insight into the time variation of the parameters. The PF formulation used here, however, gives no indication of the structural errors in the model (more generally called the model bias), which is where the KOH approach is superior. Use of both approaches in tandem is beneficial for in depth understanding of the relationship between the model and reality.

But what exactly are we tracking? We are tracking the parameter values that give the best agreement of the model with the data. The relationship between the tracked parameters and reality depends on the extent to which the model represents reality. For example, here the ventilation rate assumed in the model encompasses all controlled and uncontrolled components of the ventilation and is not simply equal to the setting on the dial. While we know what the changes are to the ventilation settings, we do not know exactly what that means in terms of changes to the ventilation rate. We also do not know to what extent other operational conditions were in place that might have impacted on ventilation rate or air speed.

The model deficiencies are clear as highlighted above, in that the physics-based model is currently unable to account for the high relative humidities observed. This limitation will be explored in future studies.

Acknowledgments and Disclosure of Funding

This work was supported by AI for Science and Government (ASG), UKRI's Strategic Priorities Fund awarded to the Alan Turing Institute, UK(EP/T001569/1) and the Lloyd's Register Foundation programme on Data-centric Engineering.

References

Chong, A. and Menberg, K. (2018). Guidelines for the Bayesian calibration of building energy models. *Energy and Buildings*, 174:527–547.

Chong, A., Xu, W., Chao, S., and Ngo, N.-T. (2019). Continuous-time Bayesian calibration of energy models using bim and energy data. *Energy and Buildings*, 194:177–190.

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

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466.

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

Madni, A. M., Madni, C. C., and Lucero, S. D. (2019). Leveraging digital twin technology in model-based systems engineering. *Systems*, 7(7):328–351.

Menberg, K., Heo, Y., and Choudhary, R. (2016). Sensitivity analysis methods for building energy models: comparing computational costs and extractable information. *Energy and Buildings*, 133:433–445.

Riddle, M. and Muehleisen, R. T. (2014). A guide to Bayesian calibration of building energy models.

Worden, K., Cross, E., Gardner, P., Barthorpe, R., and D.J.Wagg (2020). On digital twins, mirrors and virtualisations. In Barthorpe, R., editor, *Model Validation and Uncertainty Quantification, Volume 3*. Springer.