Calibration of simplified building energy models for parameter estimation and forecasting: stochastic versus deterministic modelling

Simon Rouchier[a] **, Mickaël Rabouille[b], Pierre Oberlé[b]

[a]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LOCIE, 73000 Chambéry, France
[b]INES Plateforme Formation & Evaluation

Abstract

Due to the ill-posedness of many inverse problems, parameter estimates are often prone to a possibly large uncertainty, caused by a series of errors and approximations in the experimental and modelling work. Stochastic state-space models for time series modelling incorporate a term of process noise that represents system error; most studies on building thermal model calibration however employ deterministic models that overlook this error.

This paper investigates how accounting for modelling errors affects the results of model calibration. Several simplified models are defined to simulate the indoor temperature of an experimental test cell. Some models include process noise and others do not. The parameters of each model are then learned repeatedly by using several training datasets from the test cell. The MCMC algorithm is used for training. The robustness of parameter estimates between independent trainings is evaluated. Then, the forecasting ability of the deterministic and stochastic options are compared, in terms of accuracy and robustness. Results show that stochastic modelling considerably increases the uncertainty of parameter estimates, but ensures their consistency between separate trainings, whereas deterministic models are less robust and offer a less reliable forecasting.

Keywords

calibration; uncertainty; MCMC; Kalman filter

1 Introduction

The calibration of simplified building thermal models using in-situ measurements is now a widespread research topic [1]. It is a type of inverse problems, as the user attempts to identify the causes of a physical phenomenon by observing its consequences: typically, observing the evolution of indoor temperature leads to the estimation of external sollicitations or envelope properties. It is solved as an optimisation problem,

***Corresponding author. Email: simon.rouchier@univ-smb.fr

where the objective is the minimisation of the deviation between measurements and predictions from a model [2, 3]. Such calibration is commonly performed for two general types of applications: the characterisation of the intrinsic building performance [4, 5, 6, 7, 8, 9] or other physical values; the identification of a model for predictive purposes [10, 11, 12], for instance in the aim of model predictive control [13, 14, 15, 16]. In the first case, the model should be based on some physical representation of reality in order to assess physical parameter values. In the second, a black-box model is suitable as it is used primarily for predictive purposes.

Due to the ill-posedness of many inverse problems [17], parameter estimates are prone to a possibly large uncertainty, caused by a series of errors and approximations in the experimental and modelling work [18, 19]. First, the model is an approximation of the real system: this model discrepancy may result from missing physics, overlooked input variables, numerical approximations, erroneous hypotheses, etc. Secondly, experimental uncertainty arises from inaccurate or intrusive sensors. Accounting for these errors when solving an inverse problem allows guaranteeing the value of estimated parameters within certain bounds [1].

Most of the time, the inverse problem of parameter characterisation is formulated supposing an unbiased model [18]. According to this hypothesis, there exists a set of parameter values that will allow the model to accurately simulate reality, and the only deviation between its output and experimental observations is measurement noise. This hypothesis is exceedingly optimistic, especially when models used for the characterisation of building thermal behaviour are simplified resistor-capacitor (RC) structures [3]. Accounting for modelling approximations is essential for the legitimacy of calibrated models and the interpretability of their parameters. One possible way to do so is using stochastic differential equations, solved with a Kalman filter for the estimation of states [20]. Another option for the quantification of model uncertainty is to calibrate a discrepancy function in an iterative model updating procedure [21, 22].

The literature offers many applications of parameter estimation and forecasting with stochastic models [23, 24, 25, 26, 27], but no direct comparison with their deterministic counterpart. According to [26]: *stochastic models give more reproducible results and less bias, because random effects due to process and measurement noise are not absorbed into the parameter estimates but specifically accounted for by the noise terms.* Separately, [28] stated and demonstrated that *an analysis that does not account for model discrepancy may lead to biased and over-confident parameter estimates and predictions.* The target of the present paper is to show this effect, in the case of a simple building and lumped capacitance models, made of a network of resistors and capacitors (denoted RC models). The process noise is included in the formulation of some of these models (denoted stochastic models), and excluded in others (deterministic models). The parameters of each model are then learned repeatedly by using several training datasets from an experimental test cell. The robustness of parameter estimates between independent trainings is evaluated. Then, the forecasting ability of the deterministic and stochastic options are compared, in terms of accuracy and robustness.

Sec. 2 briefly recalls the theory of filtering, forecasting and inference in state space models, applied to building modelling. Sec. 3 extends the questioning of the paper shown above, and presents the experimental and numerical methodology to answer it. Sec. 4 then shows the results of this study.

2 Inference in state space models for building modelling

2.1 Linear state space models

The presented study considers the particular (yet quite widespread and flexible) case of linear, Gaussian, time-invariant, discrete-time state space models [20]:

$$\mathbf{x}_t = \mathbf{F}_\theta \, \mathbf{x}_{t-1} + \mathbf{G}_\theta \, \mathbf{u}_t + \mathbf{w}_t \tag{1}$$

$$\mathbf{y}_t = \mathbf{H}_\theta \, \mathbf{x}_t + \mathbf{v}_t \tag{2}$$

where t is a discrete time coordinate. The terms of this system are denoted as such:

- $\mathbf{x}_t$ is a p-dimensional vector of state variables;

- $\mathbf{y}_t$ is a q-dimensional vector of observations, or output variables;

- $\mathbf{u}_t$ is a r-dimensional vector of inputs variables;

- $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q})$ is the process noise;

- $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R})$ is the observation noise.

In addition to the covariances $\mathbf{Q}$ and $\mathbf{R}$, the system is defined by its matrices $\mathbf{F}_\theta$, $\mathbf{G}_\theta$ and $\mathbf{H}_\theta$. The subscript θ indicates that these matrices are functions of a vector of parameters θ: parameter estimation is the process of assessing θ from a set of observations $\mathbf{y}_{1:N} = \{\mathbf{y}_t, t \in 1 \ldots N\}$. The process noise $\{\mathbf{w}_t\}$ is a way to account for modelling approximations, unrecognized inputs or noise-corrupted input measurements [23]. The main target of this work is to show the importance of this term in the outcome of a parameter estimation problem.

[26] denote stochastic state-space models as *grey-box* models, as opposed to deterministic *white-box* models which do not account for process noise. This definition of grey-box versus white-box models is however not unanimous in the literature: in the present paper, both alternatives will be denoted as *stochastic* or *deterministic*.

2.2 Simplified building modelling

As a discrete-time model, Eq. 1 is not the direct expression of physical conservation equations. The present section describes how a simplified building model, written in continuous time, can be translated to this form in order to perform Kalman filtering, inference and forecasting.

This study uses RC models (or *lumped* models) for simplified building modelling. In the absence of non-linear phenomena (aeraulics, long-wave radiation, moisture transfer...), these models can be expressed as linear state space models in continuous time [23]:

$$\dot{\mathbf{T}}(t) = \mathbf{A}_\theta \, \mathbf{T}(t) + \mathbf{B}_\theta \, \mathbf{u}(t) + \mathbf{w}(t) \tag{3}$$
$$\mathbf{y}(t) = \mathbf{C}_\theta \, \mathbf{T}(t) + \mathbf{v}(t) \tag{4}$$

- $\mathbf{T}(t)$ is the p-dimensional vector of all temperatures calculated by the model;

- $\mathbf{y}(t)$ is the q-dimensional vector of output temperatures that will be compared to measurements (typically $q = 1$ and $\mathbf{y}(t)$ is the indoor temperature);

- $\mathbf{u}(t)$ is the r-dimensional vector of boundary conditions: prescribed heat input, solar radiation and outdoor temperature;

- $\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}_c)$ is the process noise in continuous time;

- $\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{R}_c)$ is the observation noise in continuous time.

The target of a model calibration problem is to fit the model output $\mathbf{y}(t)$ with measurements carried in an experimental setting, in order to estimate parameter values θ that constitute the terms of the system matrices $\mathbf{A}_\theta$, $\mathbf{B}_\theta$ and $\mathbf{C}_\theta$ (the latter is often a matrix of zeros and ones indicating which of the temperatures $\mathbf{T}(t)$ are observed). Measurements are classified according to their role with respect to the model: inputs $\mathbf{u}_{1:N} = \{\mathbf{u}_t, t \in 1 \ldots N\}$ are outdoor temperature, solar radiation and heating power; the observed output $\mathbf{y}_{1:N}$ is the indoor temperature. The equations for the specific RC models used in this study will be detailed below in Sec. 3.3.

The stochastic model described by Eq. 3 must be discretized in order to specify its evolution between discrete time coordinates, as in Eq. 1. Let us denote the sample interval length Δt and assume that the inputs $\mathbf{u}(t)$ are constant during this interval. Then the system made of Eq. 3 and 4 can be translated into the discrete system made of Eq. 1 and 2 through:

3

$$\mathbf{F}_\theta = \exp\left(\mathbf{A}_\theta \, \Delta t\right) \tag{5}$$

$$\mathbf{G}_\theta = \mathbf{A}_\theta^{-1} \left(\mathbf{F}_\theta - \mathbf{I}\right) \mathbf{B}_\theta \tag{6}$$

$$\mathbf{H}_\theta = \mathbf{C}_\theta \tag{7}$$

$$\mathbf{Q} = \int_0^{\Delta t} \exp\left(\mathbf{A}_\theta \, \Delta t\right) \mathbf{Q}_c \exp\left(\mathbf{A}_\theta^T \, \Delta t\right) \mathrm{d}t \tag{8}$$

$$\mathbf{R} = \frac{1}{\Delta t} \mathbf{R}_c \tag{9}$$

The typical workflow for calibrating an RC model is to first select a model structure (1R1C, 2R2C, etc.) and write its equations in matrix form 3 and 4. Measurements are then acquired at a sample rate of Δt and the model is discretized with Eq. 5 through 9 in order to obtain the system in the form of Eq. 1 and 2. Once the system is expressed as such, one can proceed to the next steps of filtering, parameter estimation and forecasting.

2.3 Kalman filter

Let us first suppose that the parameters θ of the system are known, and a sequence of output observations $\mathbf{y}_{1:N}$ and input variables $\mathbf{u}_{1:N}$ has been obtained.

Most building model calibration practitioners work with deterministic models, in which modelling errors are not explicitly expressed as it is in the previous section. In these circumstances, all states $\mathbf{x}_{1:N}$ of the system are predicted given some parameter values θ, and compared with observations in a single operation. The objective function of the parameter estimation problem is the sum of squared errors $\sum_{t=1}^{N} \left(\mathbf{y}_t - \mathbf{H}_\theta \mathbf{x}_t\right)^2$. In a stochastic setting however, the model is considered potentially wrong and its error covariance $\mathbf{Q}$ might also be unknown. If the model is linear, the estimation of the underlying states $\mathbf{x}_{1:N}$ given the observations $\mathbf{y}_{1:N}$ is accomplished by applying a Kalman filter.

Filtering produces $p\left(\mathbf{x}_t | \mathbf{y}_{1:N}, \theta\right)$, the probability distribution function of each state $\mathbf{x}_t$ given measurements and parameter values. In the following, definitions adapted from [20] are used: $\mathbf{x}_{t|s}$ is the expected state at time t given observations up to time s. $\mathbf{P}_{t|s}$ is the variance of the state $\mathbf{x}_t$, i.e. the mean-squared error.

$$\mathbf{x}_{t|s} = \mathrm{E}\left(\mathbf{x}_t | \mathbf{y}_{1:s}, \theta\right) \tag{10}$$

$$\mathbf{P}_{t|s} = \mathrm{Var}\left(\mathbf{x}_t | \mathbf{y}_{1:s,\theta}\right) = \mathrm{E}\left[(\mathbf{x}_t - \mathbf{x}_{t|s})(\mathbf{x}_t - \mathbf{x}_{t|s})^T | \mathbf{y}_{1:s}, \theta\right] \tag{11}$$

The Kalman filter algorithm is described here:

- Set the initial states $\mathbf{x}_{0|0}$ and their covariance $\mathbf{P}_{0|0}$

- for $t = 1...N$:

    1. **Prediction step**: given the previous state $\mathbf{x}_{t-1|t-1}$ and its covariance $\mathbf{P}_{t-1|t-1}$, the model estimates the one-step ahead prediction.

    $$\mathbf{x}_{t|t-1} = \mathbf{F}_\theta \, \mathbf{x}_{t-1|t-1} + \mathbf{G}_\theta \, \mathbf{u}_t \tag{12}$$

    $$\mathbf{P}_{t|t-1} = \mathbf{F}_\theta \, \mathbf{x}_{t-1|t-1} \, \mathbf{F}_\theta^T + \mathbf{Q} \tag{13}$$

    2. **Innovations** (prediction error) ε_t and their covariances $\mathbf{\Sigma}_t$ are then calculated, along with the Kalman gain $\mathbf{K}_t$:

    $$\varepsilon_t = \mathbf{y}_t - \mathbf{H}_\theta \, \mathbf{x}_{t|t-1} \tag{14}$$

    $$\mathbf{\Sigma}_t = \mathbf{H}_\theta \, \mathbf{P}_{t|t-1} \, \mathbf{H}_\theta^T + \mathbf{R} \tag{15}$$

    $$\mathbf{K}_t = \mathbf{P}_{t|t-1} \, \mathbf{H}_\theta^T \, \mathbf{\Sigma}_t^{-1} \tag{16}$$

4

3. **Updating step**: the new states at time t are updated from the prediction $\mathbf{x}_{t|t-1}$ and the innovation.

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t \, \varepsilon_t \tag{17}$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \, \mathbf{H}_\theta) \, \mathbf{P}_{t|t-1} \tag{18}$$

- The total (negative) log-likelihood can be calculated up to a normalizing constant:

$$-\ln L_y(\theta) = \frac{1}{2} \sum_{t=1}^{N} \ln |\mathbf{\Sigma}_t(\theta)| + \frac{1}{2} \sum_{t=1}^{N} \varepsilon_t(\theta)^T \, \mathbf{\Sigma}_t(\theta)^{-1} \, \varepsilon_t(\theta) \tag{19}$$

Roughly speaking, the Kalman filter applies Bayes' rule at each time step: the updated state $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_{t|t}, \mathbf{P}_{t|t})$ is a posterior distribution, obtained from a compromise between a prior output of the model $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_{t|t-1}, \mathbf{P}_{t|t-1})$ and the evidence brought by measurements $\mathbf{y}_t$. Their relative weight is expressed by the Kalman gain $\mathbf{K}_t$ that measures the relative confidence we put in both the model and the measurements.

This standard Kalman filter algorithm works for linear systems only. Non-linear systems require either the Extended Kalman Filter (used by [26]) or the Unscented Kalman Filter.

2.4 Parameter estimation by MCMC

The prediction of states by filtering supposes that parameter values θ are known. In many problems however, these values are not known. Parameter estimation, or inference, can be carried in either a deterministic or a stochastic framework. Maximum Likelihood Estimation (MLE) is the typical way to solve parameter estimation problems [26]: starting from initial values θ_0 for the parameters, a Newton-Raphson procedure is run with the negative log-likelihood (Eq. 19) as the objective function. One can also obtain an estimate of the uncertainty of the parameters, as the inverse of the Hessian matrix calculated at the ML estimator $\hat{\theta}_{\mathrm{ML}}$ [23].

Instead of MLE, the posterior distribution $p(\theta|\mathbf{y}_{1:N})$ is estimated here by Bayesian inference:

$$p(\theta|\mathbf{y}_{1:N}) \propto L_y(\theta) \, p(\theta) \tag{20}$$

where $L_y(\theta)$ is the likelihood function calculated by the Kalman filter algorithm, and $p(\theta)$ is the prior on the parameters. Bayesian inference describes parameters as probability density functions (PDF) and returns not only point and spread estimates of the likely solutions, but also a complete description of their uncertainty, conditioned by potential measurement noise and system noise.

The Metropolis-Hastings (MH) algorithm [29] is one of the Markov Chain Monte Carlo (MCMC) methods and is used here to approximate the posterior PDF. The algorithm is described extensively in the time series analysis literature [30] and in recent papers on building physics parameter identification. [31]. MCMC gives better approximations of the parameter confidence intervals in the case where the Jacobian of the system is not explicitly formulated. Examples of Bayesian inference in the calibration of building energy models [32, 6], and the characterisation of thermal properties of materials and components [33, 34, 31].

2.5 Forecasting with uncertain parameters

Forecasting is the estimation of future states $\{\mathbf{x}_t, t > N\}$ for time steps outside of the measurement range $\mathbf{y}_{1:N}$. The main reference for this section is [30].

Let us first fix the value of the parameter vector θ. Forecasting with a state space model is done by only iterating the prediction step equations 12 and 13, and bypassing the update equations 17 and 18, since no new observations are available. In other words, the distribution of $\mathbf{x}_{N+k}$ given the data $\mathbf{y}_{1:N}$ is Gaussian, with:

$$E\left(\mathbf{x}_{N+k}|\mathbf{y}_{1:N},\theta\right) = \mathbf{F}_\theta \, E\left(\mathbf{x}_{N+k-1}|\mathbf{y}_{1:N},\theta\right) + \mathbf{G}_\theta \, \mathbf{u}_{N+k} \tag{21}$$

$$\text{Var}\left(\mathbf{x}_{N+k}|\mathbf{y}_{1:N},\theta\right) = \mathbf{F}_\theta \, \text{Var}\left(\mathbf{x}_{N+k-1}|\mathbf{y}_{1:N},\theta\right) \mathbf{F}_\theta^T + \mathbf{Q} \tag{22}$$

Let us now suppose that θ is described by a probability distribution, such as the outcome of the MCMC algorithm applied in this study. The PDF of $\mathbf{x}_{N+k}$ given $\mathbf{y}_{1:N}$ is calculated by integrated the above PDF (fixed θ) over all possible values of θ, weighted by the posterior PDF:

$$\begin{aligned} p\left(\mathbf{x}_{N+k}|\mathbf{y}_{1:N}\right) &= \int_\Omega p\left(\mathbf{x}_{N+k},\theta|\mathbf{y}_{1:N}\right) \mathrm{d}\theta \\ &= \int_\Omega p\left(\mathbf{x}_{N+k}|\theta,\mathbf{y}_{1:N}\right) p\left(\theta|\mathbf{y}_{1:N}\right) \mathrm{d}\theta \end{aligned} \tag{23}$$

The first term within the integral of Eq. 23 is the same PDF as described by Eq. 21 and 22, and the second term of the integral is the posterior distribution of θ once it has been estimated from the data.

Using MCMC as a parameter estimation technique makes the computation of Eq. 23 particularly convenient, since the posterior distribution $p\left(\theta|\mathbf{y}_{1:N}\right)$ is directly available after inference. Forecasting uncertain states over this uncertain parameter space can be done by drawing a large enough sample $\theta_{1:n}$ from the posterior, and averaging the state expectancies and variances calculated by Eq. 21 and 22 over this sample:

$$E\left(\mathbf{x}_{N+k}|\mathbf{y}_{1:N}\right) \approx \frac{1}{n}\sum_{i=1}^{n} E\left(\mathbf{x}_{N+k}|\mathbf{y}_{1:N},\theta_i\right) \tag{24}$$

$$\text{Var}\left(\mathbf{x}_{N+k}|\mathbf{y}_{1:N}\right) \approx \frac{1}{n}\sum_{i=1}^{n} \text{Var}\left(\mathbf{x}_{N+k}|\mathbf{y}_{1:N},\theta_i\right) \tag{25}$$

3 Questioning and case study

3.1 The question

The calibration of building energy models, such as simplified RC structures, can be done either in the aim of physical parameter estimation or the identification of a system for predictive purposes. In both cases, the user starts by formulating the model in continuous time (Eq. 3 if the model is linear), which after discretisation resembles Eq. 1. This formulation can then be used for states estimation, parameter estimation and forecasting.

The process noise $\mathbf{w}_t$ is often not included in the system equation, although RC models are very simplified. This means that a potentially important source of error is neglected in an inverse problem, where each error may have important consequences on results.

The question this paper aims to answer is: **what are the effects of accounting for modelling uncertainty on the results of parameter estimation and indoor temperature forecasting?** In order to answer it, model calibration was carried with seven separate training datasets from the same building, operating in similar conditions (described below) from May to August 2017. The same RC model structure was thus calibrated seven times, resulting in seven estimates (posterior PDF's) for each parameter of the model (resistances, capacitances, effective solar aperture). Results will be presented in two parts:

- **Parameter estimation**: estimated parameter values and their uncertainty intervals are compared across all datasets, in order to show whether using a stochastic rather than deterministic model result in more robust (consistent) parameter estimates;

- **Forecasting**: each dataset is used alternatively as a training set or a test set, in a form of 7-fold cross-validation, in order to assess the predictive bias and variance of deterministic and stochastic alternatives.

In the stochastic modelling situation, the procedure for parameter estimation and forecasting follows the steps presented in Sec. 2. In the deterministic case, the system noise $\mathbf{w}_t$ is removed from Eq. 1 and 3: states $\mathbf{x}_{1:N}$ are single point values and no Kalman filtering is applied. The objective function of the parameter estimation problem is simply the sum of squared errors, as described above.

3.2 Case study

The experimental test cell used in this study is called the Armadillo Box. It is a demonstration building of 42 m² floor area, designed for the 2010 European Solar Decathlon by the ENSAG-GAIA-INES team. The envelope is a light wood framed construction with integrated insulation. Heating and cooling is performed by a "3 in 1" heat pump, and photovoltaic solar panels provide recharge for electric vehicles. A large glazing area on the southern facade ensures solar heat gain in winter, while shadings have been sized to reduce summer overheating. The building considered in this study, shown on Fig. 1, is a copy of the original Armadillo Box, built on the INES test facilities to investigate its performance on the long term. A technical room on the northern side hosts monitoring equipment.
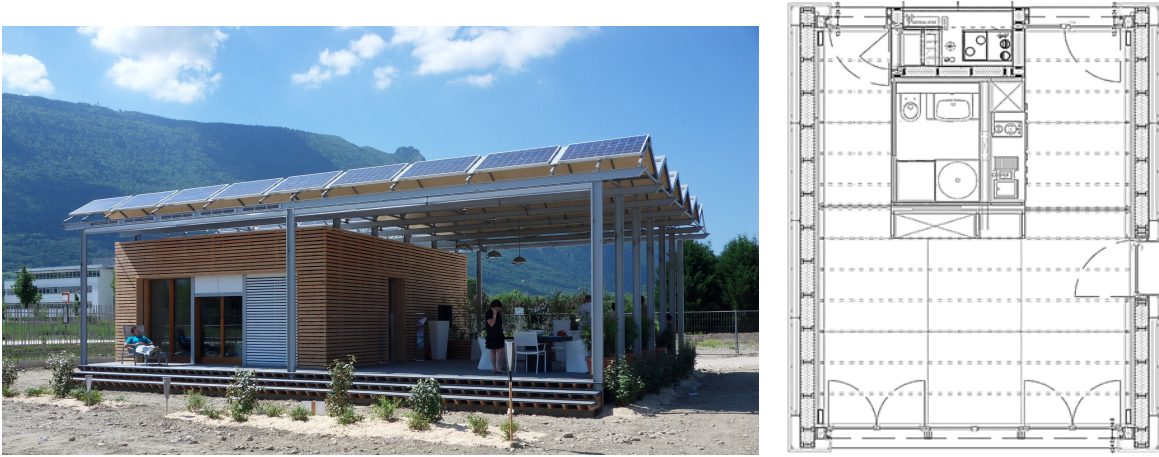

Figure 1: View of the southern facade and floor plan of the Armadillo Box

The building is monitored by a variety of sensors, but the present study only uses records of indoor temperature and prescribed heating power, in addition to weather data. The indoor temperature profiles used here have been averaged over several sensors distributed in the living space. Seven separate experimental sequences of four days each were used in this study. One of these test sequences is shown on Fig. 2: Fig. 2(a) shows indoor temperature T_{in}, adjacent room and outdoor temperature T_{out}; Fig. 2(b) shows the indoor heat input P_h and global solar irradiance on a southern surface I_{sol}.

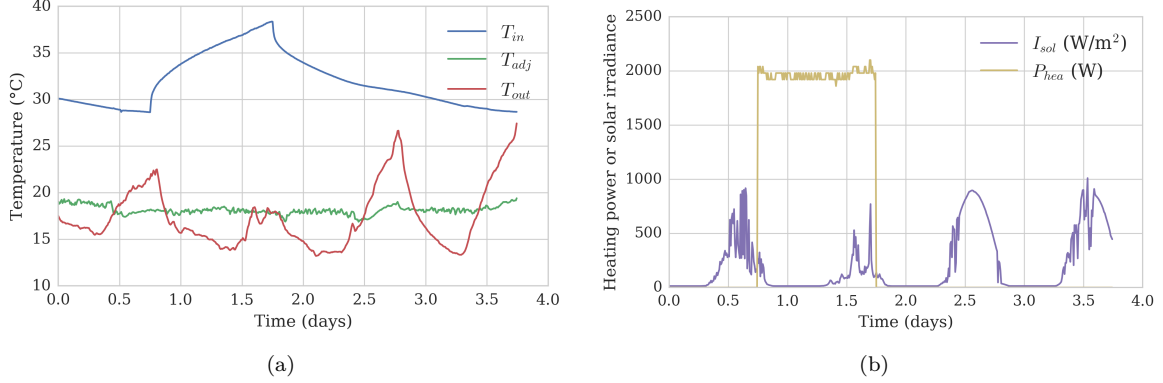| Sequence	Duration (h)	Heating (2 kW)		Weather	
		Start (h)	Duration (h)	$\overline{T}_{ext}$ (C)	$\overline{I}_{sol}$ (W)
D4	72	0	24	23.5	320.6
D5	90	18	24	21.5	256.5
D6	90	18	24	20.0	286.9
H1	100	3	48	16.3	225.4
H2	116	20	48	18.6	278.4
K1	90	18	24	17.4	207.3
K2	90	18	24	21.1	272.8

Table 1: Description of the test sequences

Figure 2: One measurement sequence in the Armadillo box

The test sequences resemble sollicitations imposed in similar studies [35]. Each sequence has a duration of 3 to 5 days, and includes a period of imposed indoor heating of 2 kW, for either 24 or 48 hours. Heating may start right after the beginning of the test, or after a period of free-floating temperature. Tab. 1 shows the conditions of each sequence, as well as its average outdoor temperature and solar irradiance. It is important to notice that all sequences are very similar: a single model, if calibrated correctly, *should* be able to reproduce all of them accurately. The point of the paper is to check if deterministic and/or stochastic models indeed show such a robustness.

3.3 Modelling

The questioning stated above supposes that a given RC model structure is trained twice from each data set: once in a deterministic formulation, once in a stochastic formulation. The robustness of each formulation is assessed by comparing parameter estimates arising from all datasets. An additional dimension is however added to the problem by applying the methodology to three types of RC models.

It seems reasonable to think that the lack of system uncertainty would be more problematic for simple models than for complex models: in an overly simplified model, such as one with a single thermal resistance R and capacitance C, modelling errors should be larger than in a model with close to no bias. As a result, we expect that if the system error $\mathbf{w}_t$ is neglected (deterministic setting), parameter estimates should be highly inconsistent across different training datasets. We expect this effect to be reduced with models of higher complexity and lower bias.

An additional dimension is added to the problem by investigating the above questioning with several levels of model complexity: 1R1C, 2R2C, 3R3C. The procedure begins by writing Eq. 3 and 4 for a given model structure (1R1C, 2R2C or 3R3C) and modelling type (deterministic or stochastic). Some variables of Eq. 3 and 4 are common to all situations:

$$\mathbf{u}(t) = \begin{bmatrix} T_{out} & I_{sol} & P_h \end{bmatrix}^T \tag{26}$$

$$\mathbf{y}(t) = \begin{bmatrix} T_{in} \end{bmatrix} \tag{27}$$

$$\mathbf{v}(t) \sim \mathcal{N}\left(0, r^2\right) \tag{28}$$

Other variables depend on the model structure and are listed on Tab. 2.

- Each model has a number of temperature nodes; the $\mathbf{C}_\theta$ matrix indicates the position of the observed (indoor) temperature. In each model, the temperature T_1 is observed.

- The 2R2C and 3R3C models respectively have one and two unobserved temperature states. Their initial states $\{T_{2,0}, T_{3,0}\}$ are unknown parameters of the problem.

8

Model	1R1C	2R2C	3R3C
$\mathbf{T}(t)$	$\begin{bmatrix} T_1 \end{bmatrix}$	$\begin{bmatrix} T_1 & T_2 \end{bmatrix}^T$	$\begin{bmatrix} T_1 & T_2 & T_3 \end{bmatrix}^T$
$\mathbf{A}_\theta$	$\begin{bmatrix} -\dfrac{1}{R_1 C_1} \end{bmatrix}$	$\begin{bmatrix} -\dfrac{1}{R_1 C_1} & \dfrac{1}{R_1 C_1} \\ \dfrac{1}{R_1 C_2} & -\dfrac{1}{R_1 C_2} - \dfrac{1}{R_2 C_2} \end{bmatrix}$	$\begin{bmatrix} -\dfrac{1}{R_1 C_1} & \dfrac{1}{R_1 C_1} & 0 \\ \dfrac{1}{R_1 C_2} & -\dfrac{1}{R_1 C_2} - \dfrac{1}{R_2 C_2} & \dfrac{1}{R_2 C_2} \\ 0 & \dfrac{1}{R_2 C_3} & -\dfrac{1}{R_2 C_3} - \dfrac{1}{R_3 C_3} \end{bmatrix}$
$\mathbf{B}_\theta$	$\begin{bmatrix} \dfrac{1}{R_1 C_1} & \dfrac{k_1}{C_1} & \dfrac{1}{C_1} \end{bmatrix}$	$\begin{bmatrix} 0 & \dfrac{k_1}{C_1} & \dfrac{1}{C_1} \\ \dfrac{1}{R_2 C_2} & \dfrac{k_2}{C_2} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & \dfrac{k_1}{C_1} & \dfrac{1}{C_1} \\ 0 & \dfrac{k_2}{C_2} & 0 \\ \dfrac{1}{R_3 C_3} & \dfrac{k_3}{C_3} & 0 \end{bmatrix}$
$\mathbf{C}_\theta$	$\begin{bmatrix} 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$
$\mathbf{Q}_c$	$\begin{bmatrix} q_1^2 \end{bmatrix}$	$\begin{bmatrix} q_1^2 & 0 \\ 0 & q_2^2 \end{bmatrix}$	$\begin{bmatrix} q_1^2 & 0 & 0 \\ 0 & q_2^2 & 0 \\ 0 & 0 & q_3^2 \end{bmatrix}$

Table 2: Matrices of the RC state-space models in continuous form

- C_i is the heat capacitance $(\mathrm{J.K^{-1}})$ of state T_i, k_i is its equivalent solar aperture $(\mathrm{m^2})$ and q_i is the standard deviation of the system error associated to it.

- The models are written with the assumption that the system error covariance matrix $\mathbf{Q}_c$ is diagonal. The standard deviations of system errors q_i and measurement error r are considered unknown, and will be inferred along with the other parameters of the models.

As an example, the lists of all unknown parameters of the 2R2C model structure, in the deterministic and stochastic cases respectively, are:

$$\theta_{2\mathrm{R2C}}^{det} = \begin{bmatrix} R_1 & R_2 & C_1 & C_2 & k_1 & k_2 & T_{2,0} & r \end{bmatrix} \tag{29}$$

$$\theta_{2\mathrm{R2C}}^{sto} = \begin{bmatrix} R_1 & R_2 & C_1 & C_2 & k_1 & k_2 & T_{2,0} & r & q_1 & q_2 \end{bmatrix} \tag{30}$$

Once written in continuous form, each system undergoes discretisation by Eq. 5 through 8. Each resulting discrete system is trained in a combination of the following settings:

- Deterministic or stochastic modelling;

- 1R1C, 2R2C or 3R3C model structure;

- One of 7 training data sets.

This results in 42 sets of posterior PDFs $p(\theta|\mathbf{y}_{1:N})$ that were calculated by the MCMC algorithm. The questioning of the paper is answered in two sections: Sec. 4.2 compares these PDFs to show the ability of stochastic and deterministic models for robust parameter estimation; Sec. 4.3 uses these estimated parameters to assess each model's ability to predict the indoor temperature in test datasets that are different from their training datasets.

4 Results and discussion

4.1 Model fit

Before showing the results of the robustness tests, let us first ensure that the selected model structures may capture the dynamics of the variable they aims at predicting (indoor air temperature).
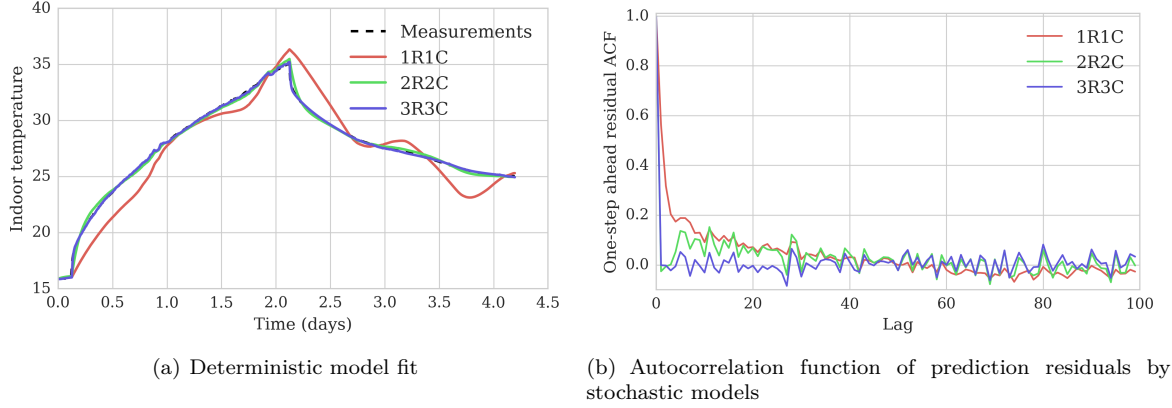

(a) Deterministic model fit

(b) Autocorrelation function of prediction residuals by stochastic models

Figure 3: Proof of sufficient model complexity

As an example, Fig. 3(a) displays the indoor temperature measurements of a dataset, and prediction by the three model structures, calibrated using this dataset for training, in a deterministic setting (the system uncertainty is neglected). It is clear that the 1R1C model is unable to reproduce the dynamics of the indoor air temperature, while the other two models have a sufficient number of degrees of freedom. The 3R3C is especially indistinguishable from measurements.

Showing the fit of stochastic models is less informative since the Kalman updating step always ensures that model predictions will fit measurements. Assessing whether a model may describe the dynamics of a system is done by examining prediction residuals or their auto-correlation function (ACF). Fig. 3(b) shows the ACF of residuals between observations and each of the predictions from the 3 models structures, in a stochastic setting. A low ACF is a necessary condition for parameter estimates to be reliable [27]: the 2R2C model has very few occurences of an ACF value above the 0.1 threshold, while the 3R3C model has none.

This analysis is not a careful model selection procedure: it merely shows that there is no need to add degrees of freedom to the models under investigation in this study. Examples of RC model selection procedures include for instance [27].

4.2 Parameter estimation

4.2.1 Heat loss coefficient

The first relevant indicator, that can be compared across all model structures, is the global heat loss coefficient (HLC) of the building. In each model structure, its PDF is obtained by reciprocating the sum of the PDFs of all resistances. The HLC is therefore comparable across all model structures: its estimate is expected to be consistent.

Fig. 4(a) compares the HLC estimated using deterministic models. Each box is the posterior PDF of the HLC from one model structure, trained with one data set. The left part shows the PDFs obtained using a 1R1C model, trained by each of the 7 training data sets. The last box of this series shows the HLC estimated by using all data sets altogether: this value is supposedly the most realistic of the series, since it uses seven times more information than each of the other estimates. The center of Fig. 4(a) compares HLC estimates from the deterministic 2R2C model, and the right part compares estimates from the 3R3C model.
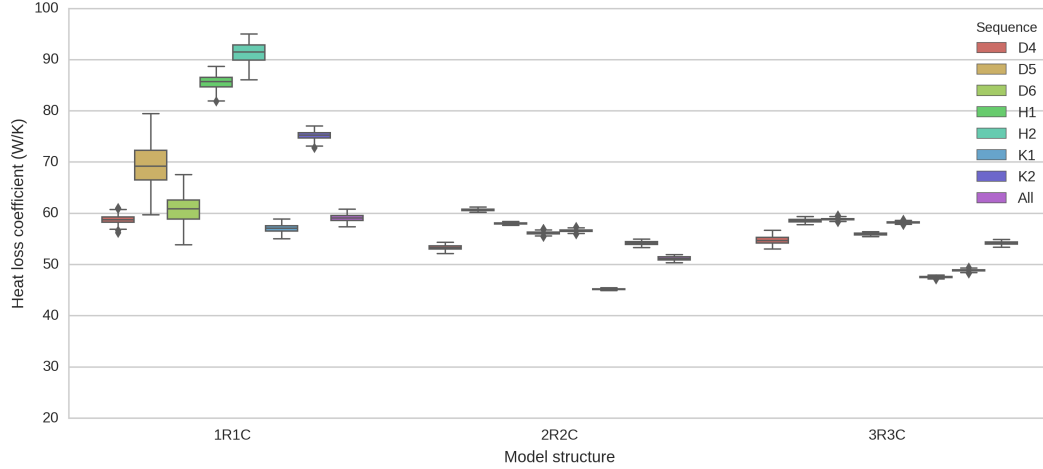
(a) HLC estimated with deterministic modelling
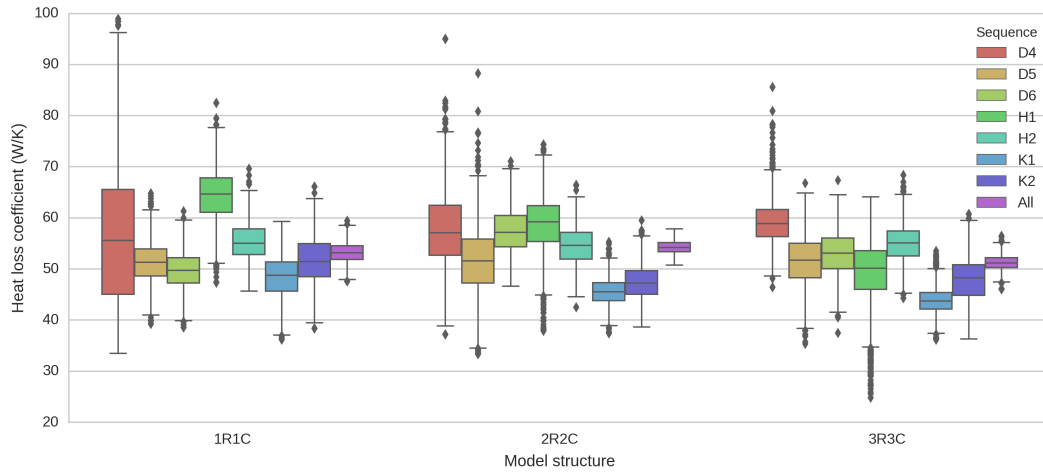
(b) HLC estimated with stochastic modelling

Figure 4: Estimation of the heat loss coefficient

Several observations can be drawn from Fig. 4(a) only, before comparing these results to the stochastic alternative. The first observation is a high inconsistency of the HLC parameter estimated with the same model but from different data. This is especially true in the case of the 1R1C structure: this model is overly simplified, and its inadequacy is not included in its formulation. The median value of posterior PDFs may span from 57 to 92 W/K. This means that using a single data set of a few days of observation offers no guarantee of accurate results for parameter estimation. This problem is mitigated, although still observable, in the case of the 2R2C and 3R3C models, which are able to mimic the behaviour of the building more appropriately. There is still a dispersion in the HLC estimation results.

The second observation is the narrowness of confidence intervals. By using a supposedly unbiased model for inference, the only admitted discrepancy between model output and observations is measurement noise. This leads to an overconfident parameter estimation, especially with a higher model complexity. More importantly, the confidence intervals of HLC do not overlap from one training data set to another. This means that these confidence intervals cannot be trusted, since the point of confidence intervals is to include all likely solutions. The issue is that they have been calculated by neglecting part of the uncertainty of the

problem.

Fig. 4(b) shows the same display of results, if the model used for inference is stochastic and employs a Kalman filter for the estimation of states. Each box is the posterior PDF of the HLC from one model structure, trained with one data set. The main difference with the deterministic setting is that confidence intervals are considerably larger. A likely explanation for this phenomenon lies in the principle of Bayesian filtering: at each time step, states are updated as a compromise between predictions and observations. The profile of one-step ahead predictions is more likely to match observations, resulting in a high likelihood for a larger range of parameter values. *Since the model is known to be wrong, the inference algorithm admits more parameter values as likely to be true.* As a result, confidence intervals are very conservative, especially for such simplified models (it should be noted that Fig. 4(b) shows no clear influence of the model complexity).

This should not be seen as an unsatisfactory result. The estimated HLC is now more reliable, and more importantly, its estimation is robust: indeed, there is an overlap between PDFs obtained from separate datasets. This means that the "true" value of the HLC may be included in the results of all trainings. Additionally, in each model structure, the PDF of the HLC estimated using all datasets overlaps each PDF from individual datasets. This is not the case when a deterministic model is used.

4.2.2 Heat capacitance

The global heat capacitance of the building is now investigated. It is the sum of the estimated values for all capacities of each model (they respectively have one, two or three). This variable is known to be model-dependent: we should not expect it to hold the same value across different model structures.

Fig. 5(a) and 5(b) show all PDFs of the global capacitance, estimated by using deterministic or stochastic models, respectively. The comparison allows the same discussion as the above investigation on the HLC:

- When neglecting system error in the model formulation (Fig. 5(a)), parameter estimation results are not consistent across training datasets. Important uncertainties are overlooked, preventing a robust model calibration. With a given model structure, the confidence intervals from separate trainings do not overlap: a single training offers no guarantee of proposing the correct value of the thermal capacity.

- When accounting for system error (Fig. 5(b)), the estimation of the thermal capacity by each model structure is more reproducible and reliable.

The latter observation is contradicted by two trainings of the stochastic 3R3C model, which resulted in capacity PDFs that hardly overlap other trainings. A possible explanation is a low identifiability of series of thermal capacities in RC models.

4.2.3 Measurement noise

The third indicator, used here to observe the relative robustess of deterministic and stochastic modelling in parameter estimation problems, is the standard deviation of observation noise r. Recall that observation noise appears in the measurement equation (Eq. 2), regardless of whether system noise is considered or not. Its standard deviation r is an unknown parameter of our models.

The advertised inaccuracy of temperature sensors used in the experimental study is $0.15°$C: the estimates of r should be lower than this value.

Fig. 6(a) shows the estimation of r by deterministic models. The deterministic 1R1C model estimates its value between 1 K and 2 K, which is much higher than expected. This observation may be given the following explanation: when system error is neglected, the only possible deviation between observations and model predictions is the measurement noise. In reality, this deviation is caused by modelling errors, ignored by our model: the inference algorithm attributes the entirety of the deviation to measurement errors and increases the estimated value of r. Deterministic 2R2C and 3R3C models overestimate r as well, although by a lower amount. Again, confidence intervals of this parameter do not overlap across separate trainings.

Fig. 6(b) shows the estimation of r by stochastic models. This estimation is more satisfactory in terms of the absolute value of r, and in terms of robustness: all trainings overlap within a realistic range.
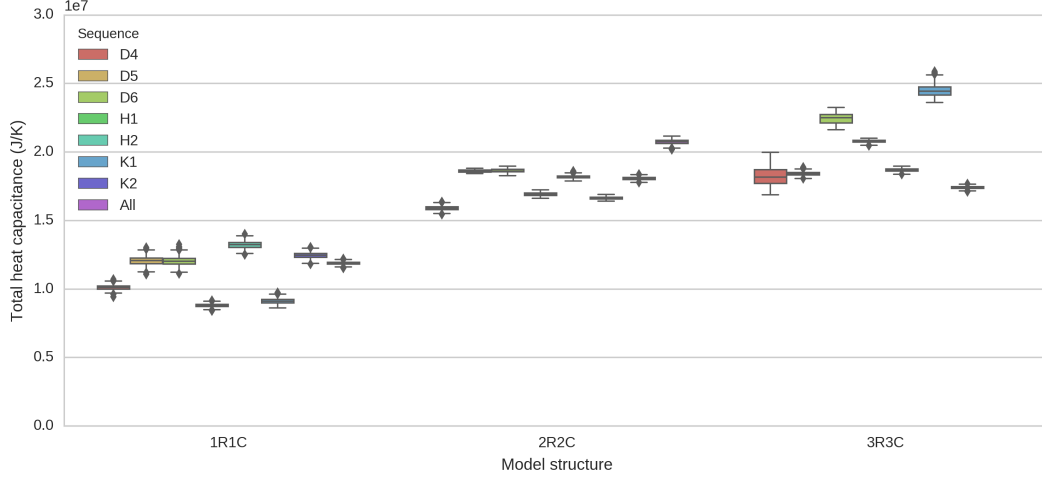
(a) Thermal capacitance estimated with deterministic modelling
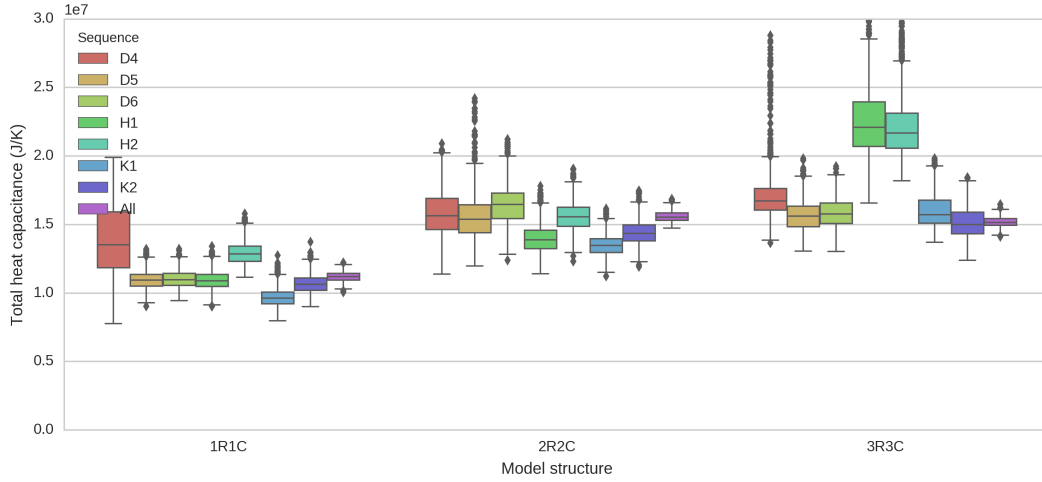
(b) Thermal capacitance estimated with stochastic modelling

Figure 5: Estimation of the total thermal capacitance

4.3   Forecasting

The decision of including system error in the formulation of a model considerably increases the confidence intervals of parameters estimated by inverse methods. The precision of parameter estimation is not increased, but its robustness is much higher. We now wish to observe if this decision has a similar impact on the forecasting capability of models.

Once a model has been calibrated with a training dataset, its forecasting accuracy and robustness may be assessed using another observation period as test set. Fig. 7 shows this test on two examples of training-test combinations: a successful prediction (Fig. 7(a)) and a less successful one (Fig. 7(b)). The predictions of indoor temperature by stochastic and deterministic 3R3C models, trained with a dataset, are compared to measurements of another set. Since parameter estimates are posterior PDFs, each prediction simulation is run $n = 100$ times by sampling the posterior: this results in confidence intervals on forecast profiles. Predictions by the stochastic model include the additional uncertainty on the states: Eq. 24 and 25 are used.
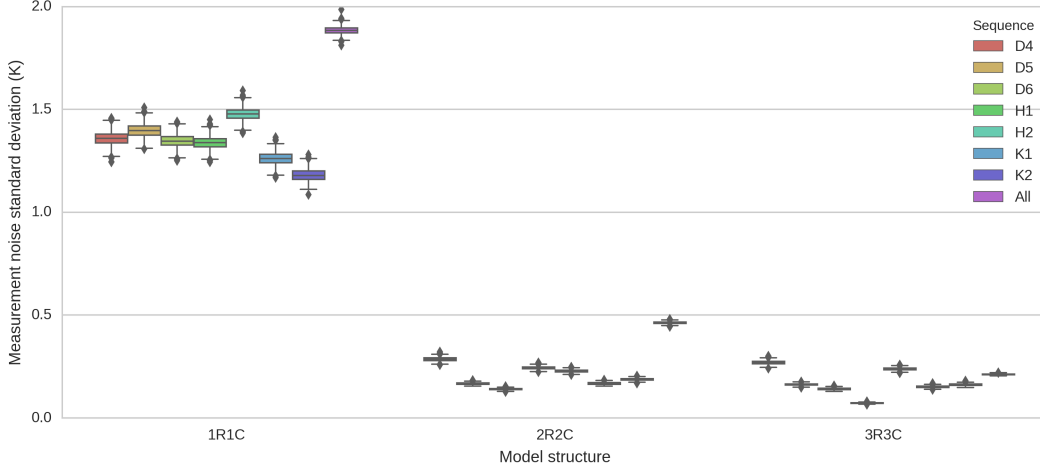
(a) R deterministic
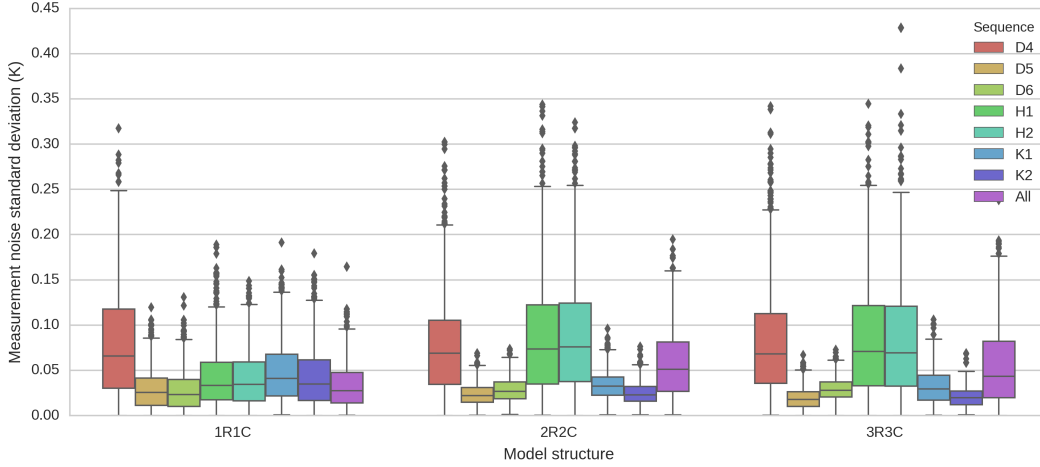
(b) R stochastic

Figure 6: Estimation of the measurement noise standard deviation

Fig. 7 shows that the stochastic model, where states are uncertain, predicts the indoor temperature with a much higher uncertainty than the deterministic model. In the latter, the only source of uncertainty is the variability of the parameter posterior. The former accounts for model inaccuracy in the forecast.

Next, a quantitative assessment of prediction accuracy and robustness is proposed, with a method inspired from k-fold cross validation.

- **Prediction accuracy** is measured by the cross-validation (CV) index, which is computed by averaging test errors [36]. Using a model trained with the learning dataset i, we denote MSE_i the averaged test error of this model over all test sets. The CV index is then the average of all MSE_i values over all

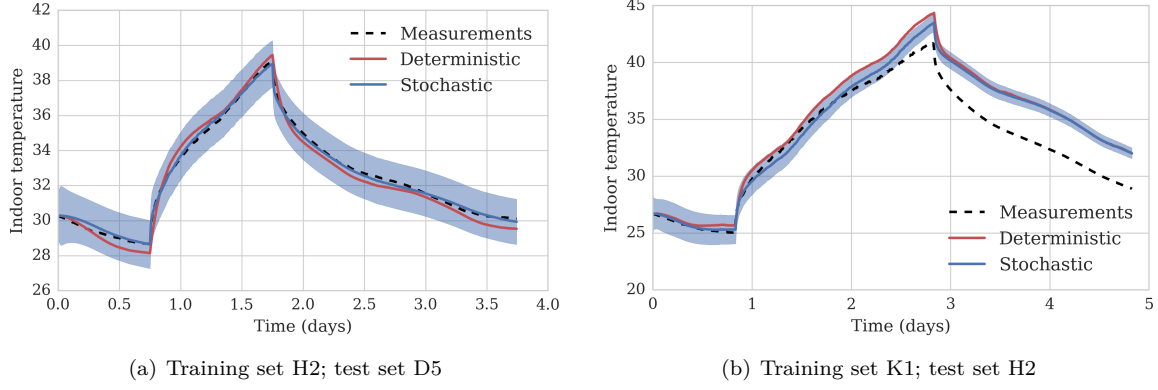(a) Training set H2; test set D5 (b) Training set K1; test set H2

Figure 7: Examples of forecast by the 3R3C model, either deterministic or stochastic

$D = 7$ datasets.

$$\text{MSE}_i = \frac{1}{D-1} \sum_{j \neq i} \left[\frac{1}{N} \sum_{t=1}^{N} (\mathbf{y}_t - \mathbf{C}\,\mathbf{x}_t) \right]_{(\text{train}=i;\text{test}=j)} \tag{31}$$

$$\text{CV} = \frac{1}{D} \sum_{i=1}^{D} \text{MSE}_i \tag{32}$$

- **Prediction robustness** is measured by an index denoted PR. Let us denote $\text{PR}_{i,j}$ the percentage of measurement points in the test dataset j, that fall inside the 95% confidence intervals of the predictions calculated by a model trained with dataset i. The global prediction robustness of one model is the averaged value of this indicator over all possible dataset combinations:

$$\text{PR} = \frac{1}{D} \sum_{i=1}^{D} \left(\frac{1}{D-1} \sum_{j \neq i} \text{PR}_{i,j} \right) \tag{33}$$

	Prediction accuracy: CV index ($\mathbf{K^2}$)	
	Deterministic	Stochastic
1R1C	6.54	5.00
2R2C	0.82	1.99
3R3C	0.86	1.10
	Prediction robustness: PR index	
	Deterministic	Stochastic
1R1C	6%	56%
2R2C	8%	63%
3R3C	8%	72%

Table 3: Assessment of prediction accuracy and robustness by each model structure; a lower CV score is better, a higher PR score is better

Results show that including the system error in the model formulation does not increase the prediction accuracy, but considerably increases its robustness: when forecasting with a stochastic model, the confidence that the real process is within confidence bounds of the prediction can be higher.

There is no apparent incidence of the test conditions on the accuracy of predictions. The duration of the training sets, duration of heating input, and weather conditions, do not seem to predetermine which

15

training sequence will yield the most robust model. These sequences are indeed very similar: the causes for prediction inaccuracy should be sought among phenomena that have not been observed, and not been considered by the simplified models.

4.4   Summary

Here is a summary of the results of both the parameter estimation and the forecasting investigations:

- Using a stochastic model in an inverse problem significantly increases the uncertainty of estimated parameters. The estimates are less confident, but more reliable as a consequence, since their confidence intervals overlap when using several training datasets separately.

- The uncertainty bounds of a parameter identified through a stochastic model give an insight of its identifiability. This allows pointing out which experimental training datasets are most informative, i.e. bring the most information into the parameter estimates.

- Using a deterministic model supposes that the model is unbiased. This hypothesis is especially inappropriate in the case of simpler models such as 1R1C. As a consequence, there is a larger variability of parameter estimates from separate training datasets. This problem is attenuated when the model complexity is increased.

- Predictions by stochastic models are not more precise, but more cautious: their confidence bounds are larger than from deterministic models. As a result, they are coherent with most of the validation data, whereas deterministic models are not.

5   Conclusion

This paper investigates the effects of system error in the calibration of building thermal models. State-space models are used to describe the behaviour of a test building. The process noise is included in the formulation of some of these models (denoted stochastic models), and excluded in others (deterministic models). The parameters of each model are then learned repeatedly by using several training datasets. The robustness of parameter estimates between independent trainings is evaluated. Then, the forecasting ability of the deterministic and stochastic options are compared, in terms of accuracy and robustness.

When neglecting system error, i.e. supposing that a building model is unbiased, the variability of parameter estimates between separate trainings is high. Their confidence intervals do not overlap, which means that they are not reliable. Accounting for system error considerably widens these confidence intervals: a stochastic model used in an inverse problem produces more robust parameter estimates. A similar observation can be done when comparing the forecasting capabilities of the deterministic and stochastic options: although using a stochastic setting does not increase the accuracy of predictions, it considerably improves their robustness.

The impact of different weather conditions on the variability of parameter estimates was not studied in this paper: all tests were carried in similar weather conditions. Introducing more weather variability would likely decrease the robustness of both deterministic and stochastic model calibrations. There are reasons to believe that stochastic models would keep an advantage over deterministic ones in such a case, although this point is worth investigating.

6   Acknowledgements

References

[1] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renewable and Sustainable Energy Reviews 37 (Supplement C) (2014) 123–141.

[2] T. A. Reddy, I. Maor, C. Panjapornpon, Calibrating Detailed Building Energy Simulation Programs with Measured Data-Part I: General Methodology, HVAC&R Research 13 (2) (2007) 221–241.

[3] A. Janssens, Report of Subtask 1b: Overview of methods to analyse dynamic data, in: IEA EBC Annex 58 - Reliable building energy performance characterisation based on full scale dynamic measurements, 2016.

[4] M. F. Fels, PRISM: An introduction, Energy and Buildings 9 (1-2) (1986) 5–18.

[5] A. Rabl, A. Rialhe, Energy signature models for commercial buildings: test with measured data and interpretation, Energy and Buildings 19 (2) (1992) 143–154.

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

[7] Y. Heo, V. M. Zavala, Gaussian process modeling for measurement and verification of building energy savings, Energy and Buildings 53 (2012) 7–18.

[8] G. Bauwens, S. Roels, Co-heating test: A state-of-the-art, Energy and Buildings 82 (2014) 163–172.

[9] Y. Zhang, Z. O'Neill, B. Dong, G. Augenbroe, Comparisons of inverse modeling approaches for predicting building energy performance, Building and Environment 86 (2015) 177–190.

[10] J. E. Braun, N. Chaturvedi, An Inverse Gray-Box Model for Transient Building Load Prediction, HVAC&R Research 8 (1) (2002) 73–99.

[11] B. Dong, C. Cao, S. E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy and Buildings 37 (5) (2005) 545–553.

[12] X. Li, J. Wen, Building energy consumption on-line forecasting using physics based system identification, Energy and Buildings 82 (2014) 1–12.

[13] J. A. Clarke, J. Cockroft, S. Conner, J. W. Hand, N. J. Kelly, R. Moore, T. O'Brien, P. Strachan, Simulation-assisted control in building energy management systems, Energy and Buildings 34 (9) (2002) 933–940.

[14] I. Hazyuk, C. Ghiaus, D. Penhouet, Optimal temperature control of intermittently heated buildings using Model Predictive Control: Part I - Building modeling, Building and Environment 51 (2012) 379–387.

[15] I. Hazyuk, C. Ghiaus, D. Penhouet, Optimal temperature control of intermittently heated buildings using Model Predictive Control: Part II - Control algorithm, Building and Environment 51 (2012) 388–394.

[16] Y. Lin, T. Middelkoop, P. Barooah, Issues in identification of control-oriented thermal models of zones in multi-zone buildings, in: 2012 IEEE 51st Annual Conference on Decision and Control (CDC), 2012, pp. 6932–6937.

[17] J. V. Beck, Inverse Heat Conduction: Ill-Posed Problems, James Beck, 1985.

[18] D. Maillet, Y. Jarny, D. Petit, Problèmes inverses en diffusion thermique - Outils spécifiques de conduction inverse et de régularisation, Techniques de l'ingénieur Transferts thermiques (2011) be8267.

[19] H. Madsen, Report of subtask 3b: Thermal performance characterisation using time series data - statistical guidelines, in: IEA EBC Annex 58 - Reliable building energy performance characterisation based on full scale dynamic measurements, 2016.

[20] R. Shumway, D. Stoffer, Time series analysis and its applications, Springer, 2016.

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

[22] P. D. Arendt, D. W. Apley, W. Chen, Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability, Journal of Mechanical Design 134 (10) (2012) 100908–100908–12.

[23] H. Madsen, J. Holst, Estimation of continuous-time models for the heat dynamics of a building, Energy and Buildings 22 (1) (1995) 67–79.

[24] K. K. Andersen, H. Madsen, L. H. Hansen, Modelling the heat dynamics of a building using stochastic differential equations, Energy and Buildings 31 (1) (2000) 13–24.

[25] J. N. Nielsen, H. Madsen, P. C. Young, Parameter estimation in stochastic differential equations: An overview, Annual Reviews in Control 24 (2000) 83–94.

[26] N. R. Kristensen, H. Madsen, S. B. Jorgensen, Parameter estimation in stochastic grey-box models, Automatica 40 (2) (2004) 225–237.

[27] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy and Buildings 43 (7) (2011) 1511–1522.

[28] J. Brynjarsdóttir, A. O'Hagan, Learning about physical parameters: the importance of model discrepancy, Inverse Problems 30 (11) (2014) 114007.

[29] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109.

[30] G. Petris, S. Petrone, P. Campagnoli, Dynamic Linear Models with R, Springer, 2009.

[31] S. Rouchier, T. Busser, M. Pailha, A. Piot, M. Woloszyn, Hygric characterization of wood fiber insulation under uncertainty with dynamic measurements and Markov Chain Monte-Carlo algorithm, Building and Environment 114 (2017) 129–139.

[32] C. Zayane, Identification d'un modèle de comportement thermique de bâtiment à partir de sa courbe de charge, Ph.D. thesis, École Nationale Supérieure des Mines de Paris (Jan. 2011).

[33] P. Biddulph, V. Gori, C. A. Elwell, C. Scott, C. Rye, R. Lowe, T. Oreszczyn, Inferring the thermal resistance and effective thermal mass of a wall using frequent temperature and heat flux measurements, Energy and Buildings 78 (2014) 10–16.

[34] J. Berger, H. R. B. Orlande, N. Mendes, S. Guernouti, Bayesian inference for estimating thermal properties of a historic building wall, Building and Environment 106 (2016) 327–339.

[35] E. Mangematin, G. Pandraud, D. Roux, Quick measurements of energy efficiency of buildings, Comptes Rendus Physique 13 (4) (2012) 383–390.

[36] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer, 2013.