Solving inverse problems in building physics: an overview of guidelines for a careful and optimal use of data

Simon Rouchier

Univ. Savoie Mont Blanc, CNRS, LOCIE, F-73000 Chambéry, France

Postprint: Rouchier S. (2018) Solving inverse problems in building physics: an overview of guidelines for a careful and optimal use of data, *Energy and Buildings* https://doi.org/10.1016/j.enbuild.2018.02.009

Abstract

Building physics researchers have benefitted from elements of statistical learning and time series analysis to improve their ability to construct knowledge from data. What is referred to here as inverse problems are actually a very broad field that encompasses any study where data is gathered and mined for information.

The purpose of the present article is twofold. First, it is a tutorial on the formalism of inverse problems in building physics and the most common ways to solve them. Then, it provides an overview of tools and methods that can either be used to assess or the reliability of inverse problem results, prevent erroneous interpretation of data, and optimise information gained by experiments. It provides an introduction, along with useful references, to the topics of estimation error assessment, regularisation, identifiability analysis, residual analysis, model selection and optimal experiment design. These concepts are presented in the context of building simulation and energy performance assessment: a simple RC model is used as a running example to illustrate each chapter.

Keywords

inverse problems; identifiability; validation; likelihood; model selection; optimal experiment design

Contents

1 Introduction **2**

1.1 General introduction 2

1.2 Running example: 2R2C model 4

2 Principle and formulation of inverse problems **4**

2.1 Types of problems and models 4

2.2 Formulation 6

2.2.1 General formulation 6

2.2.2 Linear models 8

2.2.3 Modelling errors 8

2.3 Parameter estimation in a deterministic setting 9

2.3.1 Least square estimation 9

2.3.2 Regularisation 10

2.4 Parameter estimation in a stochastic setting . . . 11
2.4.1 Kalman filter . . . 11
2.4.2 Maximum likelihood and maximum a posteriori estimation . . . 12
2.5 Example: 2R2C model . . . 13

3 Identifiability analysis 14
3.1 Structural identifiability . . . 14
3.2 Practical identifiability . . . 16
3.3 Example: 2R2C model . . . 17

4 Validation and diagnosis 18
4.1 Parameter confidence regions . . . 18
4.2 Residual analysis . . . 20
4.3 Cross-validation . . . 21
4.4 Example: 2R2C model . . . 22

5 Improving the identification 23
5.1 Model selection . . . 23
5.2 Optimal experiment design . . . 25

6 Conclusion 27

7 Acknowledgements 27

A Python code 33
A.1 Least-square estimation of the 2R2C model . . . 33
A.2 Sensitivity analysis . . . 35
A.3 Likelihood profiles . . . 36

1 Introduction

1.1 General introduction

According to the definition of [Beck and Woodbury, 1998], inverse techniques are a suite of methods which promise to provide better experiments and improved understanding of physical processes. Inverse problem theory can be summed up as the science of training models using measurements. The target of such a training is either to learn physical properties of a system by indirect measurements, or setting up a predictive model that can reproduce past observations.

In the last couple of decades, building physics researchers have benefitted from elements of statistical learning and time series analysis to improve their ability to construct knowledge from data. What is referred to here as inverse problems are actually a very broad field that encompasses any study where data is gathered and mined for information.

- **Material and component characterisation**: many material properties are not directly observable and must be estimated by indirect measurements. Inverse heat transfer theory [Beck, 1985] was developed as a way to quantify heat exchange and thermal properties from temperature sensors only, and has translated well into building physics: for instance, the characterisation of heat and moisture transfer properties of materials is an inverse problem under investigation [Künzel and Kiessl, 1996, Huang and Yeh, 2002, Rouchier et al., 2015, Berger et al., 2016, Rouchier et al., 2017] because of how time consuming traditional hygric characterisation methods are.
- **Building energy performance assessment**, from the original energy signature models [Fels, 1986, Rabl and Rialhe, 1992] to co-heating tests [Bauwens and Roels, 2014], is an inverse problem. It can be

used to formally estimate the energy savings after retrofit measures [Heo and Zavala, 2012, Zhang et al., 2015] or to point out faults in system or envelope performance [Yoshida et al., 2001, Heo et al., 2012].

- **Model predictive control** [Clarke et al., 2002, Hazyuk et al., 2012, Lin et al., 2012] requires models describing the thermal behaviour of the building, as well as the internal and external influences on its performance. Inverse problems thus include the identification of building energy performance models, weather forecast models [Oldewurtel et al., 2012, Dong and Lam, 2014], occupancy behaviour models [Dong and Andrew, 2009, D'Oca and Hong, 2015, Mirakhorli and Dong, 2016], that are reliable and computationally efficient.

These scientific challenges are gaining visibility due to the increasing availability of data (smart meters, building management systems...), the increasing popularity of data mining methods, and the available computational power to address them.

Many engineers and researchers however lack the tools for a critical analysis of their results. This caution is particularly important as the dimensionality of the problem (i.e. the number of unknown parameters) increases. When data are available and a model is written to get a better understanding of it, it is very tempting to simply run an optimisation algorithm and assume that the calibrated model has become a sensible representation of reality. If the parameter estimation problem has a relatively low complexity (i.e. few parameters and sufficient measurements), it can be solved without difficulty. In these cases, authors often do not carry a thorough analysis of results, their reliability and ranges of uncertainty. However, it is highly interesting to attempt extracting the most possible information from given data, or to lower the experimental cost required by a given estimation target. System identification then becomes a more demanding task, which cannot be done without proof of reliability of its results. One should not overlook the mathematical challenges of inverse problems which, when added to measurement uncertainty and modelling approximations, can easily result in erroneous inferences.

The purpose of the present article is twofold. First, it is a tutorial on the formalism of inverse problems in building physics and the most common ways to solve them. Then, it provides an overview of tools and methods that can either be used to **assess or the reliability** of inverse problem results, **prevent erroneous interpretation** of data and **optimise information** gained by experiments. It is focused on applications of these methods to building physics and energy simulation, but useful references to other fields are included where necessary. The paper does not mention all applications of inverse problems in building physics, but focuses on papers which addressed their challenges.

- Sec. 2 states the **formalism** of inverse problems as they are most commonly formulated in the theory of system identification and statistical inference. The different categories of models are then presented, along with the main paradigms for solving the estimation: **least square estimation** and **maximum likelihood estimation**. This section ends with a word on the main sources of errors in inverse problems and the need for regularisation.

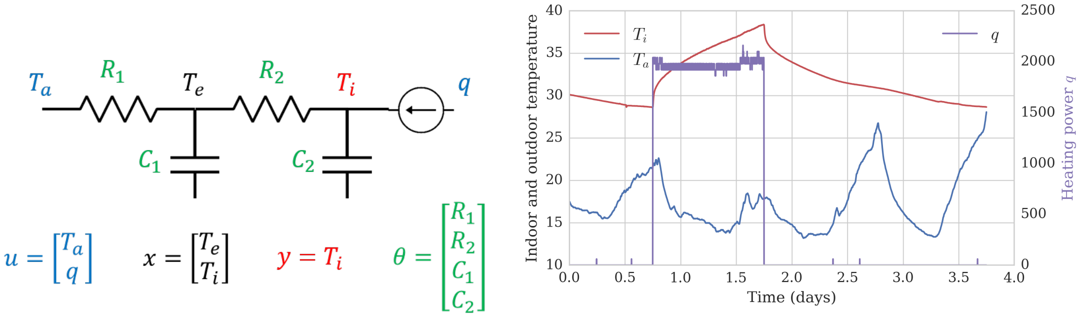
- Sec. 3 addresses the matter of **identifiability**. The parameters of a specific model, given measurement data, can be estimated with finite confidence intervals on two conditions: the model structure must allow parameters to be distinguishible, and the training dataset must be informative enough.

- Sec. 4 presents how calibrated models can be **validated**, and under what conditions the parameter estimates can be considered satisfactory. The presented methods allow **diagnosing deficiencies in the model formulation** if its complexity is insufficient or excessive.

- Sec. 5 presents two ways of ensuring that the most information is mined from the data. **Model selection methods** help pick the most appropriate model to explain a given dataset, and **optimal experiment design** is the search for the experimental setup that will maximise the information gained by a given model

1.2 Running example: 2R2C model

A picture is worth a thousand words. In an attempt to make methods more understandable and easier to apply, a running example illustrates chapter section of the paper. This example is a RC model shown on Fig. 1, the type of which is familiar to building physicists. Experimental measurements of the ambient (outdoor) temperature T_a, indoor temperature T_i and internal heat input q are available, and the target of the inverse problem is to identify the value of thermal resistors and capacitors that describe the building energy performance. These parameters form the unknown vector θ. The numerical model simulates T_i, along with an unobserved envelope temperature T_e. Other notations shown on Fig. 1(a) will be clarified below.



T_a
R_1
T_e
R_2
T_i
q
C_1
C_2
$u = \begin{bmatrix} T_a \\ q \end{bmatrix}$
$x = \begin{bmatrix} T_e \\ T_i \end{bmatrix}$
$y = T_i$
$\theta = \begin{bmatrix} R_1 \\ R_2 \\ C_1 \\ C_2 \end{bmatrix}$
Indoor and outdoor temperature
10
15
20
25
30
35
40
Time (days)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Heating power q
0
500
1000
1500
2000
2500
T_i
T_a
q

(a)

(b)

Figure 1: (a) 2R2C model and (b) dataset of the running example illustrating the paper

$$\begin{bmatrix} \dot{T}_e(t) \\ \dot{T}_i(t) \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{R_1 C_1} - \dfrac{1}{R_2 C_1} & \dfrac{1}{R_2 C_1} \\ \dfrac{1}{R_2 C_2} & -\dfrac{1}{R_2 C_2} \end{bmatrix} \begin{bmatrix} T_e(t) \\ T_i(t) \end{bmatrix} + \begin{bmatrix} \dfrac{1}{R_1 C_1} & 0 \\ 0 & \dfrac{1}{C_2} \end{bmatrix} \begin{bmatrix} T_a(t) \\ q(t) \end{bmatrix} \tag{1}$$

While each of the following chapters is documented with references pointing readers towards the appropriate literature, the paper may also be read from a more practical perspective by focusing on the “Example” sections (the last section of each chapter).

2 Principle and formulation of inverse problems

This chapter starts by an overview of the two main targets of data analysis in building sciences, and the categories of models typically used to solve them. Three main categories of models are identified in this classification: deterministic models, stochastic time series models and machine learning methods. The general formulation of inverse problems is presented in Sec. 2.2, then the solving methodology for deterministic and stochastic models are respectively described in Sec. 2.3 and 2.4. Machine learning is beyond the scope of this paper and is only briefly mentioned in Sec. 2.1.

2.1 Types of problems and models

Two types of inverse problems are mainly addressed here:

- **Parameter estimation** problems. The objective of the inverse problem is to find an estimate of physical properties. Parameters hold some physical meaning, which allows for their interpretation. Typical examples are the characterisation of material samples, or the detection of faults in building components. The validation of the results of the inverse problem focuses on the physical value of the parameter estimates and their covariance. Models used for characterisation purposes are typically more computationally demanding than those used solely for forecast.

- **System identification** problems. The objective is solely to establish a model that will be used for predictive purposes. Whether parameters hold physical meaning is irrelevant: the model can either be of grey-box or black-box type. A typical application is the forecasting of building energy performance, weather and occupant behaviour, in the aim of model predictive control. In this category of problems, the experiment can be replaced by a reference model, and the target is to find a lower-order, less computationally expensive model, which reproduces its input-to-output behaviour: these are problems of model reduction or surrogate modelling [Hazyuk et al., 2012]. Models used for prediction/forecast are typically either linear time-series analysis models (ARMA, ARX, etc.), or black-box models from the field of machine learning (neural networks, support vector machines, etc.). The calibrated model must be validated by predicting the output on a new data set, independent from the one used for the calibration.

The terms of **characterisation** and **model calibration** are often used to describe either type of problem. Other types of inverse problems may be mentioned, although they have little applications in building physics, namely problems whose objective is the reconstruction of an input signal u or of the initial conditions x_0.

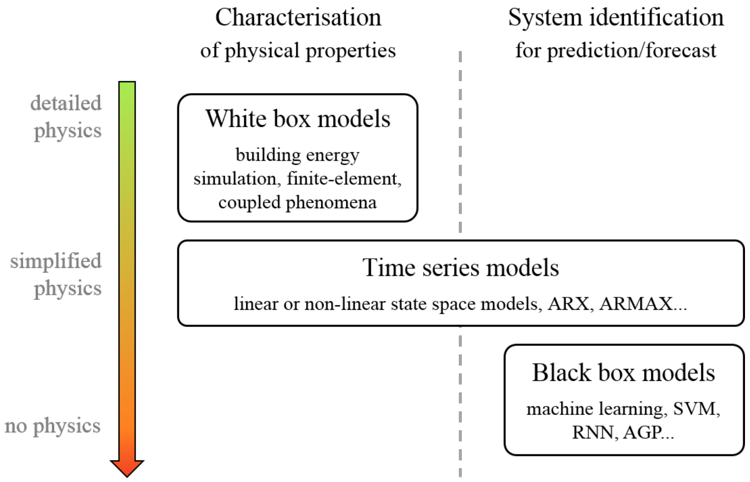


Figure 2: Classes of models and their typical field of application

Once the target of the inverse problem is set, the user should choose a model structure that will fit their requirements. These structures are presented on Fig. 2 in three categories.

1. Deterministic models, defined by the knowledge of the observed physical phenomena, so that measurements may bring an estimate of some physical parameters. These models are often non-linear and will be presented as such in the general case.

2. Stochastic time series models, including state-space models, that can be used either for parameter estimation or system identification. These models are often linear, which brings additional possibilities for their identifiability analysis and the calculation of their sensitivity and information matrices.

3. Black-box models with no link to physics, calibrated by a training dataset then used for prediction. These models originate from machine learning and statistical inference methods, and can either have a pre-defined structure (parametric models) or not (non-parametric models).

The first type of model is typically encountered when modelling coupled phenomena, such as hygrothermal or thermo-aeraulic models with field-dependent variables, or when the output y is the outcome of a Building Energy Simulation (BES) software which equations are not explicitely available for derivation. They are almost exclusively used for the characterisation of physical properties: there is rarely a point in using a non-linear white-box model for prediction/forecast, since linear time-series analysis methods and machine learning tools can perform significantly better.

When calibrating a model purely for predictive purposes, it is not necessary for it to be defined by physical laws. Statistical models and machine learning provide flexible solutions to the problem of identifying reproducible patterns in data. The goal is to understand the structure of the data in order to reproduce it. Statistical modelling relies on a mathematically proven theory behind the model but sometimes require strong assumptions on the data (typically stationarity in the weak sens). Machine learning attempts to understand the structure of the data in a less theoretical way.

Machine learning is possibly one of the fastest-growing tools for data analysis, and is increasingly applied to understanding and predicting building energy performance. The field encompasses regression, classification and clustering applied to many applications: predicting time series, understanding occupant behaviour, weather forecast, energy performance assessment, etc. The most common examples are Artificial Neural Networks (ANN) and Support Vector Machines (SVM), two subsets of artificial intelligence (AI) that are applicable to either classification or regression. A recent review of some applications of AI for building energy performance assessment was made by [Chou and Bui, 2014], along with an application of ANN and SVM for heating and cooling load prediction. Non-parametric time series analysis methods can also analyse data to capture a description of the process that generated it, without an a priori specified model structure. They are highly flexible due to not being constrained to a pre-defined model structure: this is especially useful when no prior information about the model structure is assumed. They include kernel estimation methods, splines, nearest neighbor, gaussian process models (a.k.a. kriging), etc.

The scope of the present paper does not cover the topic of machine learning, which could span over several articles by itself. Instead, here are a few references to recent comparative studies of their performance for further reading:

- Energy performance prediction, or performance comparison between pre- and post-retrofit by Gaussian Process Modeling [Heo and Zavala, 2012], Gaussian Mixture Regression [Zhang et al., 2015], ANN [Karatasou et al., 2006], etc.

- Weather forecast: each weather data type (outdoor temperature, wind speed, solar radiation) has its own uncertainty and time scale of fluctuations, and a single model cannot be suited to all. A comparison of time series analysis and ANN performance for short-term forecast was made by [Florita and Henze, 2009]. [Dong and Lam, 2014] used Adaptive Gaussian Processes for wind speed prediction and Hammerstein-Wiener models for temperature and solar radiation.

- Occupant presence and behaviour is represented by stochastic models (Hidden Markov Models, Semi-Markov Models) that are trained with data from occupancy sensors, occupant interaction with the appliances, etc. [Wang et al., 2005, Dong and Lam, 2011, Dong and Lam, 2014, Virote and Neves-Silva, 2012]

2.2 Formulation

2.2.1 General formulation

The general principle of solving a system identification problem is to describe an observed phenomenon by a model allowing its simulation. Measurements $\mathbf{z} = (\mathbf{u}, \mathbf{y})$ are carried in an experimental setup. A model is defined as a mapping between some of the measurements set as input $\mathbf{u}$ (boundary conditions, weather data) and some as output $\mathbf{y}$. The model equations are parameterised by a finite set of variables θ. Parameter estimation is the process of assessing θ from a discrete set of N data points $\mathbf{y}_{1:N} = \{\mathbf{y}_k, k \in 1 \dots N\}$.

The output of the ideal, undisturbed physical system is noted $\mathbf{y}^*$, which is the hypothetical outcome of an ideal, non-intrusive sensor. Under the hypothesis of additive measurement noise $\varepsilon(t)$, the observed output sequence is:

$$\mathbf{y}_k = \mathbf{y}_k^* + \varepsilon_k \tag{2}$$

The most common situation is that of additive white gaussian noise, i.e. $\varepsilon_k \sim \mathcal{N}(0, \sigma)$ is a sequence of independent and identically distributed (i.i.d.) variables, where the k index denotes data points and the measurement uncertainty σ may or may not be known.

The aim of the inverse problem is to approximate the system with a mathematical formulation of the outputs $\hat{y}(t, \theta)$ that will allow the estimation of θ. Ideally, the model is unbiased: it accurately describes the behaviour of the system, so that there exists a true value θ^* of the parameter vector for which the output $\hat{y}$ reproduces the undisturbed value of observed variables.

$$\mathbf{y}^*(t) = \hat{\mathbf{y}}(t, \theta^*) \tag{3}$$

Eq. 3 is written in continuous time: the discrete system output from Eq. 2 is the series of values taken by the continous process $\mathbf{y}^*(t)$ at the time coordinates $\{t_k, k \in 1 \dots N\}$. In the following, the continous and discrete notations of each variable may be used alternatively.

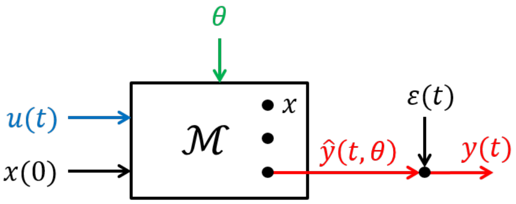


Figure 3: Sketch of the general model formulation in the hypothesis of an unbiased model

In practice, θ^* will never be reached exactly, but rather approached by an estimator $\hat{\theta}$, which may hold different values according to the criteria it follows. The unbiased model hypothesis overlooks inevitable modelling errors that occur when translating physics into equations, and solving these equations numerically. When solving an inverse problem in such a deterministic setting, it is possible to prevent these errors from interfering with inference results by applying regularisation techniques. These techniques will be introduced in Sec. 2.3.2.

Alternatively, the state-space representation of dynamic models is very common and allows explicitely formulating model inadequacy. It is written here in a continuous-discrete form:

$$\dot{\mathbf{x}}(t, \theta) = \mathbf{f}(\mathbf{x}(t, \theta), \mathbf{u}(t), \theta) + \mathbf{w}(t) \tag{4}$$

$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k, \mathbf{u}_k, \theta) + \varepsilon_k \tag{5}$$

Eq. 4 is the state-space representation of a dynamic model, possibly non-linear, written in continuous time. It describes the temporal evolution of all states of the system $\dot{\mathbf{x}}(t, \theta)$ by a user-defined function $\mathbf{f}(\mathbf{x}(t, \theta), \mathbf{u}(t), \theta)$ of the inputs $\mathbf{u}(t)$ and the parameters θ. This formulation is general and does not presume a given structure: the model can be a custom set of differential equations chosen by the user to describe a small number of states. Its last term $\mathbf{w}(t)$ is often not included, which leads to a deterministic system of Ordinary Differential Equations (ODEs). It is the system noise, which was introduced in building thermal models by [Madsen and Holst, 1995] as a way to account for modelling approximations, unrecognized inputs or noise-corrupted input measurements. Adding this term turns the system into a set of Stochastic Differential Equations (SDEs): in this setting, a filter (generally a Kalman filter) is applied for the estimation of a discrete vector of states $\mathbf{x}_k$ from observations $\mathbf{y}_k$, and the maximum likelihood or maximum a posteriori estimation methods are used for parameter estimation.

Eq. 5 is the measurement equation: it relates the states $\mathbf{x}$ to the observed output $\mathbf{y}$ and includes a term of measurement noise ε_k, similar to Eq. 2. It is written in discrete time, since observations $\mathbf{y}_{1:N}$ are a finite vector of data points. This means that Eq. 4 needs discretization before solving: an example will be shown below.

Note on notations: the nomenclature is summarised on Tab. 1. In the rest of the article, each variable may be called in either a continous or a discrete notation, and in either a vector or a scalar notation.

Table 1: Nomenclature of the inverse problem

Output		Parameters	
$\mathbf{y}$	Observations	θ	Unknown parameters
$\mathbf{y}^*$	Real process	θ^*	True value of θ
$\hat{\mathbf{y}} = \mathbf{g}(\mathbf{x}_k, \mathbf{u}_k, \theta)$	Model output	$\hat{\theta}$	Estimator
Errors		**Estimator indices**	
$\varepsilon_k = \mathbf{y}_k - \mathbf{y}_k^*$	Measurement noise	LS	Least-square
$\mathbf{r}(t, \theta) = \mathbf{y}(t) - \hat{\mathbf{y}}(t, \theta)$	Residual	ML	Maximum likelihood
$\sigma_\theta = \hat{\theta} - \theta^*$	Estimation error	MAP	Maximum a posteriori

For instance, a discrete-vector notation $\mathbf{y}_k$ indicates that measurements at the time coordinate k may be vector-valued (if several sensors are used).

2.2.2 Linear models

Some disambiguation is necessary when designating a model as *linear*: this term may denote either a linear parameter-to-output relation, or a linear input-to-output relation. In order to avoid confusion, the notions of linearity to the parameters and linearity to the inputs are separated here. The first one is used here to designate a direct linear relation between unknown parameters θ and output $\mathbf{y}$:

$$\mathbf{y} = \mathbf{S}\,\theta \tag{6}$$

where the sensitivity matrix $\mathbf{S}$ is of size $N \times p$, with N the number of data points (i.e. the size of the observation sequence $\mathbf{y}$) and p the number of parameters of the model. This type of model may be encountered in heat transfer simulations at the material or component scale. A typical example is the 1D heat conduction problem [Beck, 1985]: the reconstruction of the boundary heat flow imposed on a wall from transient temperature measurements can be formulated in the form of Eq. 6 [Maillet et al., 2011a].

Alternatively, following the conventions of control theory and system identification [Ljung, 1998], the term of linear system is used here to designate a linear input-to-output relation. The continuous-time state-space representation of such a system can be written as:

$$\dot{\mathbf{x}}(t) = \mathbf{A}(\theta)\mathbf{x}(t) + \mathbf{B}(\theta)\mathbf{u}(t) + \mathbf{w}(t) \tag{7}$$

$$\mathbf{y}_k = \mathbf{C}(\theta)\mathbf{x}_k + \mathbf{D}(\theta)\mathbf{u}_k + \varepsilon_k \tag{8}$$

The system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are functions of the parameters and are generally considered time-invariant. Another convenient way to represent a linear (deterministic) system is the transfer function form that can be easily derived from the state-space representation:

$$H(s) = \frac{Y(s)}{U(s)} = \mathbf{C}\left(s\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{B} + \mathbf{D} \tag{9}$$

The class of linear systems is a very recurrent formulation in building physics, because it includes all RC models and the family of autoregressive models used in time series analysis: ARX, ARMAX, Box-Jenkins models, etc. Since a lot of the content of the inverse problems theory originates from control theory [Walter and Pronzato, 1997], this class of system offers many possibilities for analysing the feasibility of the inverse problem, validating and improving its results. The class of autoregressive models for the statistical analysis of time series are black-box models by definition, but their parameters can be given physical interpretation by comparing them with equivalent physical models [Jiménez et al., 2008].

2.2.3 Modelling errors

The hypothesis of an unbiased model $\hat{y}$ (Eq. 3) states that there exists a parameter value θ^* for which the model output is separated from the observations y only by a zero mean, Gaussian distributed measurement noise. It means that the model perfectly reproduces the physical reality, and the only perceptible error is due to the imperfection of sensors. This is exceedingly idealistic because all models are wrong to some extent. Building energy simulation is a multi-physics, multi-scale topic that cannot accurately portray all phenomena of heat and mass transfer: forward problems are always simplified to some extent. The identification procedure is a series of experimental and numerical steps along which lay several sources of errors [Maillet et al., 2011b]: the forward problem is an approximation of the modelled physical process, with a given spatial discretisation; a hypothesis on the model may be excessively simplifying or the parametrization of a function may be wrong; the intrusiveness of a sensor may be overlooked; measurements are affected by noise and depend on sensor calibration, etc.

Modelling errors are caused by all hypotheses and shortcuts taken while translating physical phenomena into equations (a model), and translating these equations in a form that can be solved numerically (discretization, linearization, etc.) After finding a parameter estimate $\hat{\theta}$ with one of the inverse techniques listed below, it is more accurate to disaggregate the deviation between its prediction $\hat{y}(\hat{\theta})$ and the experiment y in two terms: a model error and a measurement error [Kaipio and Somersalo, 2005].

$$y(t) = \hat{y}(\hat{\theta}, t) + \left[y^*(t) - \hat{y}(\hat{\theta}, t)\right] + \varepsilon(t) \tag{10}$$

The term in brackets is the model error, i.e. the deviation between the ideal system y^* and its approximation $\hat{y}$. A technique to treat this error was presented by [Arridge et al., 2006, Kaipio and Somersalo, 2007, Nissinen et al., 2008] in a Bayesian paradigm: it is given a prior probability density, as would be the measurement error $\varepsilon(t)$. The approximation error can also be estimated after the parameter estimate $\hat{\theta}$ has been found, by filtering the measurement noise out of the residuals with a Fourier transform, assuming it has a different frequency than the approximation error.

$$r(t, \hat{\theta}) = y(t) - \hat{y}(\hat{\theta}, t) = \underbrace{\varepsilon(t)}_{\text{white noise}} + \underbrace{\left[y^*(t) - \hat{y}(\hat{\theta}, t)\right]}_{\text{approximation error}} \tag{11}$$

Modelling approximations are problematic because inverse problems are typically ill-posed [Beck and Woodbury, 1998]: their solution is highly sensitive to noise in the measured data and approximation errors. This can be quantified by the condition number of the sensivitity matrix of the forward problem: it will be shown below how much a nearly singular information matrix may affect the uncertainty of the estimates. A global optimum of the inverse problem may then be found with unrealistic physical values for the material properties as a consequence of seemingly moderate errors made when setting up the problem.

The two following sections respectively describe the procedure for parameter estimation in a deterministic and in a stochastic paradigm, and address how modelling approximations have been dealt with in building physics applications in each setting: either by using regularisation, or by including process noise in a state-space model.

2.3 Parameter estimation in a deterministic setting

2.3.1 Least square estimation

The most intuitive path to solving a parameter estimation problem is to find the set of parameters θ that minimises the squared deviation between observations y and model prediction $\hat{y}$. The least square estimator (LSE) $\hat{\theta}_{LS}$ is the global optimum of the least squares criterion:

$$r^2(\theta) = \sum_{k=1}^{N} \left(y_k - \hat{y}_k(\theta)\right)^2 = \left[\mathbf{y} - \hat{\mathbf{y}}(\theta)\right]^T \left[\mathbf{y} - \hat{\mathbf{y}}(\theta)\right] \tag{12}$$

In the model is linear to the parameters (see Eq. 6), a direct evaluation of the least square estimate $\hat{\theta}_{LS}$ is possible without iteration [Maillet et al., 2010]:

$$\hat{\theta}_{LS} = \left(\mathbf{S}^T \mathbf{S}\right)^{-1} \mathbf{S}^T \mathbf{y} \tag{13}$$

The solution only exists if $\mathbf{S}^T\mathbf{S}$ is non singular, that is if $\mathbf{S}$ is of full rank [Maillet et al., 2010].

If the model cannot be written in the form of Eq. 6, an iterative scheme is needed to find the LSE. These iterative optimization schemes fall within two categories: gradient-based and gradient-free methods. Gradient-based methods rely on the value of the derivatives of the objective function (here the sum of squared residuals $r^2(\theta)$), calculated at the current iteration, to propose the next value of θ. The gradient of the cost function can for instance be approached by the adjoint state method [Brouns et al., 2013, Brouns et al., 2017], parameter-perturbation methods or sensitivity-equation methods [Palomo Del Barrio and Guyon, 2003]. Most of the built-in curve fitting methods available in scientific programming softwares include an implementation of the Gauss-Newton or Levenberg-Marquardt algorithm and can perform a numerical approximation of the Jacobian. Alternatively, gradient-free methods only use the current value of the objective function $r^2(\theta)$ to update the proposal of θ. Examples of their implementation in building sciences include: the calibration of a mono-zone building model including convection and radiation effects with a genetic algorithm [Lauret et al., 2005]; the calibration of a series of increasingly large thermal and hygric models with a genetic algorithm [Kramer et al., 2013]; the characterisation of hygrothermal properties of a porous material with the covariance matrix adaptation algorithm [Rouchier et al., 2015].

Solving an inverse problem by returning only point estimates $\hat{\theta}$ is not very informative. Due to the ill-posedness of inverse problems, small measurements errors and seemingly reasonable model hypotheses may add up to very large errors on the parameter estimates. Interactions between parameters in the forward problem also translate as correlations in the multivariate probability function of the estimate (more on the topic of identifiability follows in Sec. 3). When presenting the results of an inverse problem, it is important to illustrate the estimate uncertainty with its covariance matrix. The covariance matrix of the LSE can be derived as the inverse of the information matrix $\mathbf{I} = \mathbf{S}^T\mathbf{S}/\sigma^2$ [Cai and Braun, 2015, Maillet et al., 2011a]:

$$\mathrm{cov}\left(\hat{\theta}_{LS}\right) = \sigma^2 \left(\mathbf{S}^T\mathbf{S}\right)^{-1} \tag{14}$$

where σ is the measurement uncertainty and $\mathbf{S}_k$ is the sensitivity matrix calculated locally:

$$S_{ij}(\theta) = \left[\frac{\partial \hat{y}_i(\theta)}{\partial \theta_j}\right] \tag{15}$$

The standard deviation on the individual parameter estimates $\sigma_{\hat{\theta}}$ and the corresponding correlation matrix can then be obtained by decomposing the covariance matrix.

The least-square estimation is the typical framework for calibrating deterministic systems, i.e. systems described by equations which do not include a stochastic term. Some criteria are similar to the sum of squared residuals defined in Eq. 12, and are sometimes used for an easier interpretation of model fitness: the root mean square error [Joe and Karava, 2017], mean absolute percentage error [Dong and Lam, 2014], etc.

2.3.2 Regularisation

Regularisation aims at reducing the effect of data inaccuracy on the identification. The first possible approach for regularisation is to reduce the degrees of freedom of the problem by restricting the search to a set of admissible solutions. It is the principle of the truncated singular value decomposition technique [Hansen, 1990] and the future information method [Beck, 1985]. The second approach, known as Tikhonov regularisation [Tikhonov and Arsenin, 1977], is another way to introduce a constraint by penalizing the fitness value of unrealistic solutions. The objective function $r^2(\theta)$ (Eq. 12) is modified after this principle. A quadratic term is introduced, adding a convex component to the search space and orienting the search towards a prior estimate θ_p of the expected solution vector.

$$\hat{\theta}_{LS} = \mathrm{argmin}_\theta \left\{ \sum_{i=1}^{N} \left(y_i - \hat{y}_i(\theta)\right)^2 + \alpha \|\theta - \theta_p\|^2 \right\} \tag{16}$$

The regularisation parameter $\alpha \geq 0$ balances the evaluation of individuals between the optimization of the least square criterion, and the agreement with a range of physically admissible solutions. A low value of α implies an insufficient regularisation of the problem, while a high value imposes too much of a constraint and forces the solution to match the prior. Guidelines exist for the correct choice of α, such as the L-curve method [Hansen, 1992]. This method states that several runs of the search algorithm with different values of α result in an L-shaped graph when displaying the solutions $\|\theta - \theta_p\|$ versus their residuals $\|y - \hat{y}\|$, and that the optimal choice for α is near the corner of this L-curve. This method is used in [Rouchier et al., 2015] to tune the regularisation parameter and facilitate the identification of the hygrothermal properties of a material. It was shown by [Wang and Zabaras, 2004] that the LSE with Tikhonov regularisation and the MAP estimator have similar mathematical forms.

2.4 Parameter estimation in a stochastic setting

2.4.1 Kalman filter

In deterministic circumstances, modelling errors are not explicitally expressed : all states $\mathbf{x}_{1:N}$ of the system are single-point values and entirely specified by the model structure and parameter values θ. In a stochastic setting however, the model is considered potentially wrong. Each vector of states is defined by a probability distribution function $p\left(\mathbf{x}_k|\mathbf{y}_{1:N},\theta\right)$, given a sequence of measurements $\mathbf{y}_{1:N}$ and parameter values θ.

If the model is linear, the estimation of this PDF is accomplished by applying a Kalman filter at each time step. In the following, definitions adapted from [Shumway and Stoffer, 2016] are used: $\mathbf{x}_{k|s}$ is the expected state at time t given observations up to time s. $\mathbf{P}_{k|s}$ is the variance of the state $\mathbf{x}_k$, i.e. the mean-squared error.

$$\mathbf{x}_{k|s} = \mathrm{E}\left(\mathbf{x}_k|\mathbf{y}_{1:s},\theta\right) \tag{17}$$

$$\mathbf{P}_{k|s} = \mathrm{Var}\left(\mathbf{x}_k|\mathbf{y}_{1:s}\right) = \mathrm{E}\left[(\mathbf{x}_k - \mathbf{x}_{k|s})(\mathbf{x}_k - \mathbf{x}_{k|s})^T|\mathbf{y}_{1:s},\theta\right] \tag{18}$$

The Kalman filter algorithm is described here, applied to the discrete linear state-space model shown on Eq. 34 and 35. We suppose that the observation noise ε_k follows a zero-mean normal law of covariance matrix $\mathbf{R}$.

- Set the initial states $\mathbf{x}_{0|0}$ and their covariance $\mathbf{P}_{0|0}$
- for $k = 1...N$:

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

$$\mathbf{x}_{k|k-1} = \mathbf{F}\,\mathbf{x}_{k-1|k-1} + \mathbf{G}\,\mathbf{u}_k \tag{19}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}\,\mathbf{x}_{k-1|k-1}\,\mathbf{F}^T + \mathbf{Q} \tag{20}$$

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

$$\varepsilon_k = \mathbf{y}_k - \mathbf{H}_\theta\,\mathbf{x}_{k|k-1} \tag{21}$$

$$\mathbf{\Sigma}_k = \mathbf{C}\,\mathbf{P}_{k|k-1}\,\mathbf{C}^T + \mathbf{R} \tag{22}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\,\mathbf{C}^T\,\mathbf{\Sigma}_k^{-1} \tag{23}$$

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

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k\,\varepsilon_k \tag{24}$$

$$\mathbf{P}_{k|k} = \left(\mathbf{I} - \mathbf{K}_k\,\mathbf{C}\right)\,\mathbf{P}_{k|k-1} \tag{25}$$

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

$$-\ln L_y(\theta) = \frac{1}{2}\sum_{k=1}^{N} \ln|\mathbf{\Sigma}_k(\theta)| + \frac{1}{2}\sum_{k=1}^{N} \varepsilon_k(\theta)^T\, \mathbf{\Sigma}_k(\theta)^{-1}\, \varepsilon_k(\theta) \tag{26}$$

Roughly speaking, the Kalman filter applies Bayes' rule at each time step: the updated state $p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_{k|k}, \mathbf{P}_{k|k})$ is a posterior distribution, obtained from a compromise between a prior output of the model $p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k|k-1}, \mathbf{P}_{k|k-1})$ and the evidence brought by measurements $\mathbf{y}_k$. Their relative weight is expressed by the Kalman gain $\mathbf{K}_k$ 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 [Kristensen et al., 2004]) or the Unscented Kalman Filter.

2.4.2 Maximum likelihood and maximum a posteriori estimation

Maximum likelihood estimation (MLE) is a standard approach to parameter estimation in statistics. It is a prerequisite for many statistical inference methods [Myung, 2003] used for model selection criteria, parameter significance tests, Bayesian methods, etc. and provides techniques for pointing out model deficiencies. The basic idea is to construct a function of the data and the unknown parameters called the likelihood function [Åström, 1980]. The maximum likelihood estimate $\hat{\theta}_{ML}$ is the parameter value that maximizes this criterion.

The likelihood function L_y of the parameter θ given the observed data y is equal to the probability of observing data y given the parameter θ:

$$L_y(\theta) = p(\mathbf{y}_{1:N}|\theta) \tag{27}$$

This function measures how likely the observed sample $\mathbf{y}_{1:N}$ is, as a function of a parameter value θ. The point of MLE is to find the estimator $\hat{\theta}_{ML}$ that maximizes it (or, more frequently, that minimizes the negative log-likelihood $-\ln L_y(\theta)$), i.e. the value which is the most likely to have generated the data.

Supposing that the measurement noise at each time step t_k is independent, identically distributed and Gaussian of covariance matrix $\mathbf{R}$, the likelihood is deduced from the Kalman filter algorithm in Eq. 26. It can then be used as the objective function of a minimization routine.

The central limit theorem states that the unbiased maximum likelihood estimator is asymptotically gaussian with mean $\hat{\theta}_{ML}$ and a covariance matrix $\mathrm{cov}(\hat{\theta})$. The precision of the estimator has a lower bound given by the Cramér-Rao theorem by:

$$\mathrm{cov}\left(\hat{\theta}_{ML}\right) = E\left[\left(\hat{\theta}_{ML} - \theta^*\right)\left(\hat{\theta}_{ML} - \theta^*\right)^T\right] \geq \mathbf{F}(\hat{\theta}_{ML})^{-1} \tag{28}$$

where $\mathbf{F}(\theta)$ is the observed Fisher information matrix:

$$\mathbf{F}(\theta) = E\left[\left(\frac{\partial \ln L_y(\theta)}{\partial\theta}\right)^T \left(\frac{\partial \ln L_y(\theta)}{\partial\theta}\right)\right] \tag{29}$$

or alternatively it is equal to the negative of the Hessian matrix of the log-likelihood function:

$$F_{i,j}(\theta) = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln L_y(\theta)\right] \tag{30}$$

Note that this definition of the covariance matrix of the error of the estimate supposes that the estimator is unbiased [Emery and Nenarokomov, 1998] as defined in Eq. 3. The marginal uncertainty on the individual estimates $\sigma_{\hat{\theta}}$ and the corresponding correlation matrix can then be obtained by decomposing the covariance matrix.

The likelihood approach to parameter estimation is a statistical paradigm, which differs from the typical view of least square estimation. If the system is modelled by stochastic differential equations, the evaluation of the likelihood function can be solved as a general nonlinear filtering problem: [Madsen and Holst, 1995] proposed a method based on the extended Kalman filter, for the identification of thermal models of buildings. This work was then followed by many noteworthy contributions on solving inverse problems described by SDEs in the MLE framework [Andersen et al., 2000, Kristensen et al., 2004, Bacher and Madsen, 2011].

Alternatively, the Maximum a Posteriori (MAP) approach is very similar to MLE and is applicable when some prior information on the parameters $p(\theta)$ is available. The MAP estimator $\hat{\theta}_{MAP}$ is the argument which maximises the posterior $p(\theta|\mathbf{y})$ distribution defined by Bayes' theorem:

$$p(\theta|\mathbf{y}) \propto L_y(\theta) p(\theta) \tag{31}$$

where $p(\theta)$ is the prior distribution of the parameters, i.e. the formulation of initial expert knowledge that may be known before collecting observations. If there is little prior information about parameter values, i.e. the prior $p(\theta)$ may be a uniform distribution on a large interval, the MAP and MLE estimates are equivalent. The incorporation of expert knowledge in parameter estimation puts the problem in the framework of Bayesian inference. Bayes' theorem updates the probability for a hypothesis $p(\theta)$ as more information y becomes available. Bayesian parameter estimation is typically solved by Markov Chain Monte-Carlo methods (MCMC) that generate a sequence $\{\theta_n, n = 0, 1, 2...\}$ approximating the posterior PDF $p(\theta|\mathbf{y})$. It is a non-parametric description of the posterior from which it is simple to extract an approximation of the MAP estimate and its covariance matrix $\mathrm{cov}(\hat{\theta}_{MAP})$. Here are some examples of building physics inverse problems solved in the Bayesian framework: the inverse heat conduction problem [Wang and Zabaras, 2004, Kaipio and Fox, 2011]; the estimation of the thermal properties of a wall using temperature sensors [Berger et al., 2016] and heat flux measurements [Biddulph et al., 2014]; the analysis of building energy savings after retrofit and payback times [Heo et al., 2012]; the characterisation of hygric properties of porous materials from relative humidity and weight measurements [Rouchier et al., 2017]; the calibration of building thermal models [Schetelat and Bouchié, 2014, Zayane, 2011].

2.5 Example: 2R2C model

Let us recall the application example shown in Sec. 1.2: a 2R2C model has been defined (Fig. 1(a)) in order to estimate the characteristics of a building from a dataset holding three measured time series: indoor and outdoor temperature, and heating power (Fig. 1(b)). Its physical equation (Eq. 1) can be formulated as a continuous-time state-space system similar to Eq. 7 and 8:

$$\underbrace{\begin{bmatrix} \dot{T}_e(t) \\ \dot{T}_i(t) \end{bmatrix}}_{\dot{\mathbf{x}}(t)} = \underbrace{\begin{bmatrix} -\dfrac{1}{R_1 C_1} - \dfrac{1}{R_2 C_1} & \dfrac{1}{R_2 C_1} \\ \dfrac{1}{R_2 C_2} & -\dfrac{1}{R_2 C_2} \end{bmatrix}}_{\mathbf{A}(\theta)} \underbrace{\begin{bmatrix} T_e(t) \\ T_i(t) \end{bmatrix}}_{\mathbf{x}(t)} + \underbrace{\begin{bmatrix} \dfrac{1}{R_1 C_1} & 0 \\ 0 & \dfrac{1}{C_2} \end{bmatrix}}_{\mathbf{B}(\theta)} \underbrace{\begin{bmatrix} T_a(t) \\ q(t) \end{bmatrix}}_{\mathbf{u}(t)} + \mathbf{w}(t) \tag{32}$$

$$\mathbf{y}_k = \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix}}_{\mathbf{C}(\theta)} \begin{bmatrix} T_e \\ T_i \end{bmatrix}_{t=t_k} + \varepsilon_k \tag{33}$$

The vector of states $\mathbf{x}(t) = [T_e(t), T_i(t)]^T$ includes the observed indoor temperature T_i and unobserved envelope temperature T_e. The $\mathbf{C}$ matrix points to which of the states is observed.

Eq. 32 needs to be translated from continuous to discrete time, since observations are only available at a finite set of time coordinates $\{t_k, k \in 1 \dots N\}$ separated by a time step Δt. The discretisation process of linear state space models is described by [Madsen and Holst, 1995] and briefly summarized here. The goal is to turn the continuous-discrete set of equations 32 and 33 into the following one:

$$\mathbf{x}_k = \mathbf{F}\,\mathbf{x}_{k-1} + \mathbf{G}\,\mathbf{u}_k + \mathbf{w}_k \tag{34}$$

$$\mathbf{y}_k = \mathbf{C}\,\mathbf{x}_k + \varepsilon_k \tag{35}$$

where the $\mathbf{F}$ and $\mathbf{G}$ matrices of the discrete equation result from the $\mathbf{A}$ and $\mathbf{B}$ matrices of the continuous equation, and the process noise in discrete time $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_d)$ has a covariance matrix $\mathbf{Q}_d$ that can be estimated from the covariance matrix of the process noise in continuous time $\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}_c)$:

$$\mathbf{F} = \exp(\mathbf{A}\,\Delta t) \tag{36}$$

$$\mathbf{G} = \mathbf{A}^{-1}\,(\mathbf{F} - \mathbf{I})\,\mathbf{B} \tag{37}$$

$$\mathbf{Q}_d = \int_0^{\Delta t} \exp(\mathbf{A}\,\Delta t)\;\mathbf{Q}_c \exp\left(\mathbf{A}^T\,\Delta t\right) \mathrm{d}t \tag{38}$$

This model takes the outdoor temperature T_a and indoor heating q as inputs, and returns a predicted time series of indoor temperature T_i. The inverse problem is the estimation of 4 parameters $\theta = \{R_1, R_2, C_1, C_2\}$ by fitting this prediction on the observed profile of T_i.

Table 2: Parameter estimation results in the 2R2C example

Parameter	Initial guess	$\hat{\theta}_{LS}$	$\sigma_{\hat{\theta}}$	t-statistic	p-value
R_1 (W/K)	1×10^{-2}	2.13×10^{-2}	5.45×10^{-5}	392	< 0.01
R_2 (W/K)	1×10^{-2}	2.37×10^{-3}	2.27×10^{-5}	104	< 0.01
C_1 (J/K)	1×10^{7}	1.56×10^{7}	8.93×10^{4}	175	< 0.01
C_2 (J/K)	1×10^{7}	1.93×10^{6}	5.22×10^{4}	37	< 0.01
$T_e(0)$ (C)	20	30.27	2.36×10^{-2}	1281	< 0.01

The Python code for running a simple Least-Square estimation of θ by the Levenberg-Marquardt algorithm is given in Appendix A. Running this code requires a data file available within the folder of execution. In this example, the initial condition on the unobserved envelope temperature T_e is considered unknown, and estimated along with the other parameters. The results are shown on Tab. 2

3 Identifiability analysis

The present section is concerned with the a priori feasibility of parameter estimation. An appropriate model structure and a sufficiently rich data set are two necessary (though not sufficient) conditions for a satisfactory parameter estimation. It is possible to check for these conditions before running the often costly estimation algorithm, in order to adapt either the model or the data.

The usual definition of identifiability originates from [Bellman and Åström, 1970]. This notion was originally predominantly developed to help understanding complex biological systems, each of which is modelled by a specific set of differential equations with unobservable parameters. The question of identifiability is whether the input-output relation of the system may be explained by a unique parameter combination θ.

$$y(\theta) = y(\tilde{\theta}) \Rightarrow \theta = \tilde{\theta} \tag{39}$$

Two conditions are required for the parameter estimates to be identifiable: the model structure must allow for parameters to be theoretically distinguishible from one another, with no redundancy; the data must be informative so that parameter uncertainty is not prohibitively high after identification. These conditions are respectively denoted structural and practical identifiability.

3.1 Structural identifiability

Structural identifiability relates the possibility of finding parameter estimates to the structure of the model, independently from measurements.

A priori identifiability of linear systems

Let us illustrate this question in the particular case of linear, time invariant systems. This includes RC models for buildings where not all temperature nodes are observed.

$$\dot{\mathbf{x}}(t) = \mathbf{A}(\theta)\mathbf{x}(t) + \mathbf{B}(\theta)\mathbf{u}(t) \tag{40}$$

$$\mathbf{y}_k = \mathbf{C}(\theta)\mathbf{x}_k \tag{41}$$

It was shown by [Grewal and Glover, 1976] that two sets of parameter values are indistinguishable if and only if they both yield the same impulse responses and transfer functions. This leaves two possibilities for the assessment of the identifiability of the system on Eq. 40: either taking a Laplace transform of the system and check whether the same input-output relation implies an unique parameter set [Walter and Pronzato, 1997], or expressing its impulse response with its Markov parameters. The first method will be illustrated on the 2R2C example in Sec. 3.3 below. The latter method was done by [Agbi et al., 2012] for the identification of a multi-zone thermal model, as a preliminary step to study the impact of experimental data quality. The system on Eq. 40 is identifiable if and only if the equality of Markov parameters implies the equality of parameters. As shown by [Dötsch and Van Den Hof, 1996], this is equivalent to a rank test of the Jacobian matrix, locally defined around $\theta = \theta_0$ by:

$$\left.\frac{\partial S_m(\theta)}{\partial \theta}\right|_{\theta=\theta_0} = \left[\frac{\partial h^{(i)}(\theta)}{\partial \theta_j}\right]_{\theta=\theta_0} \tag{42}$$

where the impulse response of the system h and its derivatives are expressed by the first m Markov parameters of the model:

$$h^{(k)}(\theta) = \mathbf{C}(\theta)\mathbf{A}^{k-1}(\theta)\mathbf{B}(\theta) \tag{43}$$

This is equivalent to checking whether the following structural information matrix M is of full rank:

$$M = \left.\left(\frac{\partial S_m(\theta)}{\partial \theta}\right)^T \left(\frac{\partial S_m(\theta)}{\partial \theta}\right)\right|_{\theta=\theta_0} \tag{44}$$

A priori identifiability of non-linear systems

The issue of structural identifiability however applies to all classes of models, and not only RC networks. The identifiability of non-linear models is analysed from the same theoretical basis [Bellman and Åström, 1970, Grewal and Glover, 1976]: proving that the input-output relation of the model can only be explained by a single set of parameters. Recent overview articles [Raue et al., 2014, Grandjean et al., 2017] provide a list of *a priori* structural identifiability analysis methods. Particularly, [Grandjean et al., 2017] give a particularly clear explanation of the following alternatives, and apply them linear and non-linear models close to those used in building simulation.

- The Taylor series expansion approach was theorised by [Pohjanpalo, 1978]. It relies on the uniqueness of the coefficients of a Taylor series expansion of the output with respect to time. This philosophy is therefore similar to the impulse response method, except that the model is dealt with in continuous time. There exists and order of differentiation for this series expansion, which coefficients form a non-linear algebraic system of equations in the parameters and from which the structural identifiability may be pronounced. Its solvability is checked by the rank of the Jacobian matrix. As underlined by [Sedoglavic, 2001], the order of differenciation in this method is not bount, which can lead to highly complex calculations as the models grow large. The author circumvents the exponential complexity by the use of differential algebra for the series expansion. For this purpose, [Sedoglavic, 2001] developed an algorithm available on Maple and later [Karlsson et al., 2012] on Mathematica.

- Based on the theory from [Ritt, 1950] for differential algebra, a global identifiability analysis can be performed for dynamic models described by polynomial or rational equations [Raue et al., 2014]. The characteristic set of the differential ideal from the model structure can be used to define a normalized exhaustive summary of the model, which is in essence an implicit description of its input-output behaviour. Showing the injectivity of the exhaustive summary proves the identifiability of the model. Later, [Pia Saccomani et al., 2003] developed an algorithm made available by [Bellu et al., 2007] as the DAISY algorithm. Its easiness of use makes it an interesting tool although it quickly becomes prohibitive for large systems.

To the author's knowledge, there has however been no application of these methods to the field of building energy simulation. A comparative study has however been applied to civil engineering problems by [Chatzis et al., 2015].

3.2 Practical identifiability

Structural identifiability analysis methods mentioned above do not account for the richness of measurements (or lack thereof), and therefore do not guarantee that parameters can be properly estimated in practice. It is possible that the current experimental settings do not offer enough richness of information for the identification of some parameters, despite these parameters being theoretically distinguishable in the model structure. Furthermore, the assessment of structural identifiability does not account for measurement uncertainty either. This uncertainty may have serious consequences on parameter uncertainty, especially concerning parameters which have little influence on the system output.

Practical identifiability relates the parameter estimation possibilities to the experimental design (type and amount of measurements), the richness of available data and its accuracy, in addition to accounting for the type of model used. A parameter within a model is identifiable in practice if the data brings enough information to estimate it with finite confidence intervals.

It is possible to filter out parameters that are unlikely to be learned from the data by running a preliminary sensitivity analysis. Parameter estimation algorithms are often computationally expensive and their cost quickly rises with the number of parameters of the model. This is a motivation for excluding parameters with little influence on the output, especially since their estimates are bound to have wide confidence intervals. Sensitivity analysis is the main mathematical tool for the purpose of identifying the physical phenomena that can be really tested on the available experimental data. It measures the effects of parameter variations on the behaviour of a system and allows two things: ranking parameters by their significance so that non-influencial parameters may be filtered out, and identifying correlations between parameters which may prevent their estimation. Many local and global sensitivity analysis methods are applicable, providing first-order and total-order sensitivity indices from which correlations can be assessed: differential sensitivity analysis [Lomas and Eppel, 1992] calculates the sensitivity of the model output to each parameter locally as defined by Eq. 15. It was used by [Palomo Del Barrio and Guyon, 2003] to calculate indices for individual parameter influence and correlations, and by [Berger et al., 2016] as a preliminary step before model calibration. Sampling-based methods (variance-based and one-at-a-time methods) allow a global sensitivity analysis but are seldom used in preparation of an inverse problem, due to their computational cost [Lomas and Eppel, 1992, Rabouille, 2014].

Principal component analysis (PCA) is a way to move the problem into a less correlated parameter space. It was applied by [Palomo Del Barrio and Guyon, 2003] before the calibration of a thermal model [Palomo del Barrio and Guyon, 2004]. Similarly, [Cai and Braun, 2015] use a significance vector defined as the square root of diagonal elements of the information matrix $\mathbf{S}^T\mathbf{S}$, then apply a method based on principal component analysis (PCA) to remove the most correlated parameters from it. The criterion for measuring parameter correlations is the condition number of the information matrix.

Practical identifiability is a local property: the main identifiability check must therefore be run after identification, once parameter estimates have been found. This is part of the steps one should follow to validate the estimates, by observing their confidence regions and quantifying the information gained by the model from the data.

3.3 Example: 2R2C model

Structural identifiability analysis is illustrated here with the 2R2C model example (Fig. 1). It is a linear model for which the transfer function method is suitable [Walter and Pronzato, 1997]. The system can be written in Laplace form as:

$$s\,\mathbf{X}(s) = \mathbf{A}\,\mathbf{X}(s) + \mathbf{B}\,\mathbf{U}(s) \tag{45}$$

$$\mathbf{Y}(s) = \mathbf{C}\,\mathbf{U}(s) \tag{46}$$

The transfer function of this system is then a [1 × 2] matrix:

$$\mathbf{H}(s,\theta) = \frac{\mathbf{Y}(s)}{\mathbf{U}(s)} = \mathbf{C}\left(s\mathbf{I}_2 - \mathbf{A}\right)^{-1}\mathbf{B} \tag{47}$$

$$= \frac{1}{s^2 + \dfrac{C_1R_1 + C_2R_1 + C_2R_2}{C_1C_2R_1R_2}s + \dfrac{1}{C_1C_2R_1R_2}} \begin{bmatrix} \dfrac{1}{C_1C_2R_1R_2} & \dfrac{1}{C_2}s + \dfrac{R_1+R_2}{C_1C_2R_1R_2} \end{bmatrix} \tag{48}$$

Note that this derivation can be done manually due to the simplicity of the 2R2C model. In case of a more complicated linear model, a symbolic computation software is preferable.

The system is structurally identifiable iff the unicity of transfer function implies the unicity of parameters:

$$\mathbf{H}(s,\theta) = \mathbf{H}(s,\tilde{\theta}) \Rightarrow \theta = \tilde{\theta} \tag{49}$$

This is solved by checking for unicity of each term of the transfer function for two parameter sets θ and $\tilde{\theta}$:

$$\frac{C_1R_1 + C_2R_1 + C_2R_2}{C_1C_2R_1R_2} = \frac{\tilde{C}_1\tilde{R}_1 + \tilde{C}_2\tilde{R}_1 + \tilde{C}_2\tilde{R}_2}{\tilde{C}_1\tilde{C}_2\tilde{R}_1\tilde{R}_2} \tag{50}$$

$$\frac{1}{C_1C_2R_1R_2} = \frac{1}{\tilde{C}_1\tilde{C}_2\tilde{R}_1\tilde{R}_2} \tag{51}$$

$$\frac{1}{C_2} = \frac{1}{\tilde{C}_2} \tag{52}$$

$$\frac{R_1+R_2}{C_1C_2R_1R_2} = \frac{\tilde{R}_1+\tilde{R}_2}{\tilde{C}_1\tilde{C}_2\tilde{R}_1\tilde{R}_2} \tag{53}$$

One can quickly check that the unicity of two transfer functions $\mathbf{H}(s,\theta)$ and $\mathbf{H}(s,\tilde{\theta})$ indeed implies the equality of each individual parameter: the condition of structural identifiability is satisfied.

Practical identifiability is another necessary condition for the results of the inverse problem to be relevant. Running a sensitivity analysis on the model parameters is a way to estimate their relevance in the problem, although it is not a sufficient condition for identifiability.

Table 3: FAST sensitivity analysis of the 2R2C parameters to the least-square criterion

Parameter	First order index	Total order index
R_1 (W/K)	0.467	0.753
R_2 (W/K)	0.012	0.049
C_1 (J/K)	0.001	0.030
C_2 (J/K)	0.001	0.005
$T_e(0)$ (C)	0.207	0.467

An example of a simple local FAST sensitivity analysis on the 2R2C model is shown here. The influence of each parameter on the least-square residual criterion is shown on Tab. 3. The code for running this analysis is available in Appendix A.2. Results suggest that thermal capacitances have little influence on the solution: their estimation should be validated with caution after solving the inverse problem.

4 Validation and diagnosis

Let us suppose that the user has gathered measurement data $\mathbf{z} = (\mathbf{u}, \mathbf{y})$, chosen a numerical model and its parameterisation θ to depict the observed phenomena, checked for theoretical identifiability, and run a parameter estimation algorithm in either the least-squares or maximum likelihood framework, to obtain an estimate $\hat{\theta}$ and its covariance matrix $\mathrm{cov}(\hat{\theta})$. Let us now address how the results of an inverse problem solved with one model type and one data set can be validated.

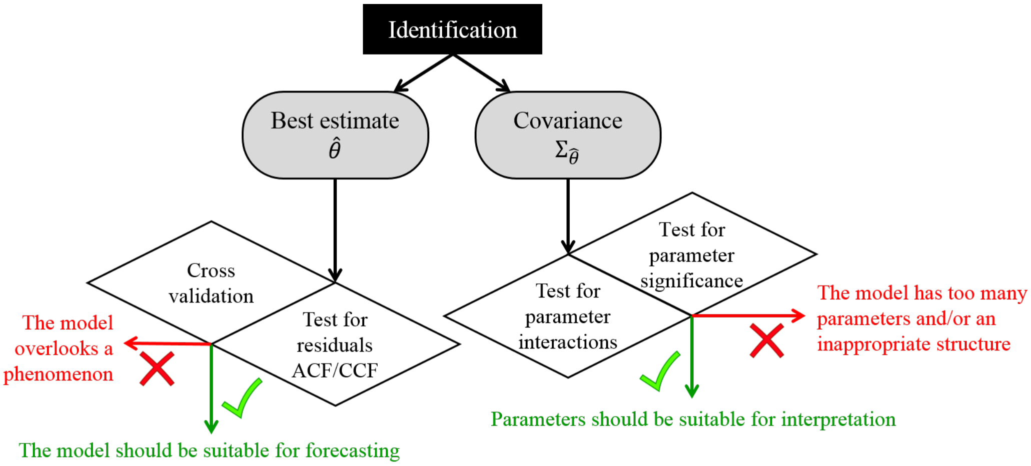

Identification
Best estimate $\hat{\theta}$
Covariance $\Sigma_{\hat{\theta}}$
Cross validation
Test for residuals ACF/CCF
The model overlooks a phenomenon
The model should be suitable for forecasting
Test for parameter interactions
Test for parameter significance
The model has too many parameters and/or an inappropriate structure
Parameters should be suitable for interpretation

Figure 4: Validation steps

Once the parameter estimation is complete, several steps should be followed to make sure the results are usable. These steps are listed by [Madsen, 2016] for both model selection and validation, and fall within three categories.

- **Tests on the parameter estimates**. This is to make sure that the chosen model structure is appropriate and does not include redundant parameters. It is especially important in characterisation studies, where the parameter values are the sole target of the identification. Practical identifiability is a measure of the information truly gained by the model from the experiment, and helps identify how much each parameter has been updated by observations.
- **Tests on the model output** and residuals. This is to make sure that the model complexity is sufficient to replicate the observations, and can be trusted to simulate the reality with different initial and boundary conditions. It is especially important if the target of the system identification study is to establish a predictive model.
- **Out-of-sample validation** of the predicted output using a different dataset than the one used for training.

In addition to checking model validity, these steps allow establishing a diagnosis of which improvements can be brought to the model.

4.1 Parameter confidence regions

In this part, the analysis concerns mostly the covariance matrix of the estimate $\mathrm{cov}(\hat{\theta})$, which can be calculated by Eq. 14 in the least-squares estimation, or Eq. 28 in the maximum-likelihood estimation. The covariance matrix is related to the correlation matrix $\mathbf{R}_{\hat{\theta}}$ by:

$$\mathrm{cov}\left(\hat{\theta}\right) = \sigma_{\hat{\theta}}\, \mathbf{R}_{\hat{\theta}}\, \sigma_{\hat{\theta}} \tag{54}$$

where $\sigma_{\hat{\theta}}$ is a diagonal square matrix containing the standard deviation $\sigma_{\hat{\theta}i}$ of each individual parameter $\hat{\theta}_i$. The matrices $\mathbf{R}_{\hat{\theta}}$ and $\sigma_{\hat{\theta}}$ are the basis for testing for superfluous parameters in the model.

A preliminary sensitivity analysis performed before parameter estimation does not guarantee that the confidence intervals and regions of estimates are finite. Such an analysis is performed either globally (for Monte-Carlo and sampling-based methods) or locally near an initial guess value of the parameters. In either case, its outcome cannot precisely depict the uncertainty ranges of parameters near the estimate $\hat{\theta}$ obtained by the identification.

The first criterion for validating parameter estimates is their individual significance. A low influence of a parameter on the model output results in low values of the sensitivity matrix $\mathbf{S}$ (Eq. 15) or information matrix $\mathbf{F}$ (Eq. 29), which translates into high values in the diagonal of the covariance matrix (Eq. 54). It is important to note that the calculation of the covariance matrix depends on the data: it only measures whether parameters are significant **in the conditions of the experiment**. The marginal significance of a parameter is evaluated by comparing its absolute value $\hat{\theta}_i$ with its standard deviation $\sigma_{\hat{\theta},i}$. It can be done by a simple comparison of both, or with a t-test for statistical significance: [Ljung, 1998, Madsen, 2007] use the value of the standard deviation $\sigma_{\hat{\theta},i}$ to test the hypothesis H_i that $\hat{\theta}_i$ is statistically significant, against the hypothesis H_0 that it is not. Alternatively, the confidence interval of a single parameter can be approached by the value of its diagonal term in the covariance matrix.

The second criterion for validating parameters is the lack of serious correlations between estimates. The correlation matrix $\mathbf{R}_\theta$ has coefficients between -1 and 1, indicating pairwise coupled effects of parameters on the model output. A high correlation between two parameter estimates means that the model structure should be revised or that one of the parameters should be fixed to an assumed value. A statistically insignificant parameter may disturb the estimation of more important parameters it interacts with: it is generally stated that if a parameter is found to be either insignificant or strongly correlated with another, it should be removed from the model and the estimation should be conducted once more [Palomo Del Barrio and Guyon, 2003, Madsen, 2016].

Perhaps the most informative way to assess the practical identifiability of a model is the display of confidence regions and intervals for its parameter estimates. A likelihood-based method of setting these regions is described by [Meeker and Escobar, 1995, Raue et al., 2009] as the likelihood ratio test and is briefly summarised here. Suppose a model of p parameters θ which exhibit some interaction. We want to draw the confidence regions for a subset θ_1 of the parameters of length p_1, with the remaining parameters denoted θ_2, in order to see if this region is finite and the model identifiable. If the maximum likelihood estimator $\hat{\theta}_{ML}$ has been identified, the likelihood ratio function is defined by:

$$R(\theta_1) = \max_{\theta_2} \left[\frac{L_y(\theta_1, \theta_2)}{L_y(\hat{\theta})} \right] \tag{55}$$

A property of the likelihood ratio test is that $-2\ln\left[R(\theta_1)\right]$ asymptotically follows a χ^2 distribution with p_1 degrees of freedom [Meeker and Escobar, 1995]. An approximate $100(1-\alpha)\%$ likelihood-based confidence region for θ_1 is the set of all values such that:

$$-2\ln\left[R(\theta_1)\right] < \Delta^2_{1-\alpha,p_1} \tag{56}$$

where $\Delta^2_{1-\alpha,p_1}$ is the $1-\alpha$ quantile of the χ^2 distribution with p_1 degrees of freedom. Note that this test can be performed after a deterministic parameter estimation, by using the sum of squared residuals instead of the likelihood function:

$$-2\ln\left[\frac{L_y(\theta)}{L_y(\hat{\theta})}\right] = \frac{1}{\sigma^2}\left(r^2(\theta) - r^2(\hat{\theta})\right) \tag{57}$$

From this theory of asymptotic likelihood-based confidence regions, [Raue et al., 2009] proposed the definition of the profile likelihood function $\chi^2_{\mathrm{PL}}(\theta_i)$ of a single parameter θ_i as an a posteriori way to check its structural and practical identifiability. This function is defined as the likelihood ratio in the particular case of a single explanatory parameter:

$$\chi^2_{\mathrm{PL}}(\theta_i) = \max_{j\neq i}\left[\frac{L_y(\theta_i,\theta_j)}{L_y(\hat{\theta})}\right] \tag{58}$$

As written by [Raue et al., 2009]: structurally non-identifiable parameters are characterized by a flat profile likelihood. The profile likelihood of a practically non-identifiable parameter has a minimum, but is not excessing a threshold $\Delta_{1-\alpha}$ for increasing and/or decreasing values of θ_i (here, $\Delta_{1-\alpha}$ is the $1-\alpha$ quantile of the χ^2 distribution with one degree of freedom). As an example, the 95% confidence interval of a single parameter θ_i is the interval of values so that $\chi^2_{\mathrm{PL}}(\theta_i)$ does not exceed the threshold $\Delta_{1-95\%} = 3.84$.

A comprehensive application of this theory in a building physics application was proposed recently by [Deconinck and Roels, 2017] to measure the identifiability of parameters of several RC models describing the thermal characteristics of a building component. Results reveal large differences in the practical identifiability of models between winter and summer conditions: this underlines the importance of the richness of excitation data on the results of an inverse problem, independently from the model structure. Two-dimensional confidence regions can also be shown by applying Eq. 56 to pairwise parameter combinations: these regions can be plotted in the form of a correlation matrix enriched with precise confidence thresholds [Raue et al., 2009, Feroz et al., 2011].

4.2 Residual analysis

The most straightforward way to check for the validity of a calibrated model is a visual comparison between measurements and simulations that takes into account both the measurements noise and the model input data uncertainties. The agreement between model and reality is stated to be good when a significant overlapping is observed between simulations and measurements uncertainty bands [Palomo Del Barrio and Guyon, 2003]. Note that this statement also holds for a white-box model where parameters are not the outcome of an inverse problem: the validation of such a model should meet these standards as well.

However, a more systematic analysis of the residuals is often preferable. Statistical tools to compare measurements and simulations can be used to assess the validity of the model after calibration. The following steps were recommended for all model validation procedures during the PASSYS project by [Palomo Del Barrio et al., 1991]. Let us focus on the definition of the residuals calculated from the calculation of any of the estimates $\hat{\theta}$ defined above:

$$r(t) = y(t) - \hat{y}(t,\hat{\theta}) \tag{59}$$

Residuals from an ideal unbiased model should behave like white noise, i.e. a stochastic process that approaches a stationary normal distribution with zero mean. In time series analysis, the stationarity of this stochastic process can be checked if its mean and variance do not vary over different time periods. Another method is to use the normalized autocorrelation function (ACF) of residuals [Godfrey, 1980]:

$$\mathrm{ACF}_r(\tau) = \frac{1}{\sigma_r^2} E\left[(r(t)-\mu_r)\,(r(t+\tau)-\mu_r)\right] \tag{60}$$

where μ_r and σ_r^2 are the mean and variance of the process (the residual). The ACF measures the average correlation between points separated by a lag τ within the time series. The ACF of a true white noise signal is zero for all lags other than zero. The whiteness test on the ACF is the first statistical test in residual analysis [Palomo Del Barrio et al., 1991]. The second test is the independence test described below.

The criterion of white noise residuals is very difficult to meet in practice [Kramer et al., 2013] as all models include hypotheses and approximations in addition to measurement uncertainty. Furthermore, numerically high values of the ACF do not precisely point out the source of model inadequacies. Enter the cross-correlation function (CCF) [Godfrey, 1980]:

$$\mathrm{CCF}_{r,u}(\tau) = \frac{1}{\sigma_r\sigma_u} E\left[(r(t)-\mu_r)\,(u(t+\tau)-\mu_u)\right] \tag{61}$$

The CCF checks is the residuals are correlated with any of the input processes u. Should a significant cross-correlation with one of the inputs be found, then it is likely that this input is improperly accounted for in the model [Madsen, 2007]. Ideally, a correct model structure should not yield any cross-correlation between residual and input signals. Statistical tests to meet the independence criterion are described by [Palomo Del Barrio et al., 1991].

A strong residual ACF does not necessarily imply that an input is being overlooked. An example of model which checks the independence test (CCF) but fails the whiteness test (ACF) is shown by [Kramer et al., 2013]. The satisfying low values of the CCF suggest than no input is missing in the model. By adding an error model to the model structure, the authors then saw the ACF fit within a reasonable bandwidth. According to [Ljung, 1998], the ACF obtained from a model missing an error model is less likely to meet the whiteness test. Other examples of residual analysis for model validation include:

- [Jiménez et al., 2008] apply residual analysis to the validation of different ARX and ARMAX models in a numerical study: these models include error models.
- [Reynders et al., 2014] identify reduced-order models of buildings on reference simulations from a detailed physical model. The reduced-order models are described by stochastic differential equations. The authors analyse the cumulated periodograms (this is equivalent to analysing the ACF) to pick the necessary complexity of reduced models.
- [Joe and Karava, 2017] draw ACF and CCF profiles after estimation in a deterministic agent-based framework and found relatively high values for these functions.

The overall conclusion is that deterministic grey-box modelling is unlikely to meet the standards for the whiteness test of residuals. In a deterministic context, a good overlap between confidence regions of predictions and observations is often a sufficient criterion for judging that a model structure is appropriate.

4.3 Cross-validation

The result of an inverse problem is the set of parameters $\hat{\theta}$ with which a specific model $\hat{y}$ offers the closest fit to the training data. The goal of system identification is however to build a model which accurately predicts the outcome of new input conditions. The generalization performance of a model relates to its prediction capability on independent test data [Hastie et al., 2001]. Estimating this prediction error is the first step towards the selection of the appropriate model complexity to represent the reality. A complex model will make more use of the training data than a simple model: its average training error will be lower, but the covariance of the parameter estimates $\mathrm{cov}(\hat{\theta})$ will be higher, hence so will the variance of output predictions over the test dataset. There is a model complexity threshold over which decreasing the training error means increasing the generalization error: such a model is overfitted and has a poor prediction performance.

Expected prediction errors $E\left[(y-\hat{y}(u))^2\right]$ can be decomposed into two main components: the squared bias and the variance [Hastie et al., 2001]. As a reminder, the hypothesis of additive Gaussian measurement noise still holds: $y(t) = y^*(t) + \varepsilon(t)$ with $\epsilon \sim \mathcal{N}(0, \sigma)$.

$$\underbrace{E\left[(y-\hat{y}(u))^2\right]}_{\text{Expected prediction error}} = \underbrace{\left(E\left[\hat{y}(u)\right]-y^*\right)^2}_{\text{Bias}^2} + \underbrace{E\left[\left(\hat{y}(u)-E\left[\hat{y}(u)\right]\right)^2\right]}_{\text{Variance}} + \sigma^2 \tag{62}$$

The squared bias is the deviation between the average estimation $E\left[\hat{y}(u)\right]$ and the real, noise-free value of the output y^*. This term should be equal to zero under the hypothesis of an unbiased model $y^*(t) = \hat{y}(t, \theta^*)$. This hypothesis is however very optimistic in practice and requires a model with a large number of degrees of freedom (parameters). The variance is the expected deviation of the prediction $\hat{y}(u)$ around its mean. It can be evaluated by propagating the parameter uncertainty $\mathrm{cov}(\hat{\theta})$ into an output uncertainty. The last term is an irreducible error due to the measurement noise. Fig. 5 illustrates the bias-variance tradeoff in the search of the lowest prediction error.

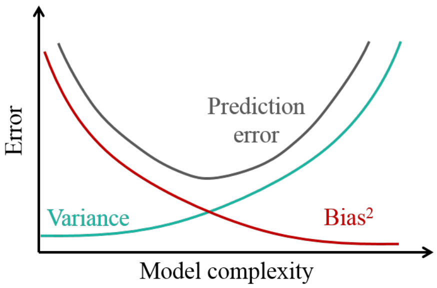
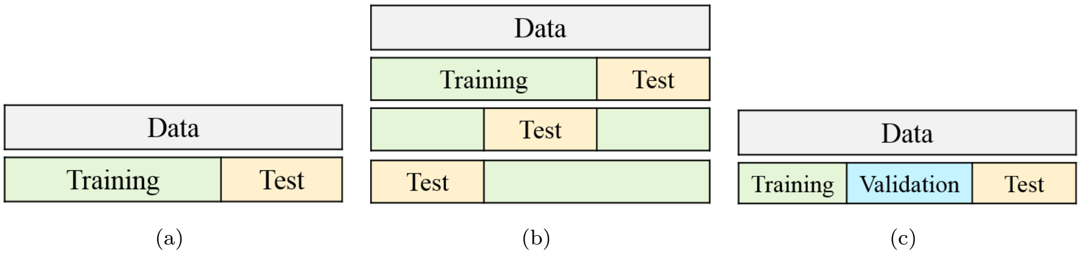


Figure 5: Bias-variance trade off



(a)

(b)

(c)

Figure 6: Splitting data for cross-validation

Evaluating the exact prediction error requires a separate dataset from the one used for training purposes. Cross-validation is a convenient way to assess the generalization ability of a model by splitting the original dataset in several samples. The most intuitive approach is the holdout method which splits the original data in two sets (Fig. 6(a): a training set used to fit the model (typically two thirds of the original data) and a test set for its validation. Alternatively, k-fold cross-validation (Fig. 6(b)) splits the original sample into k subsamples. $k-1$ samples are used as training data while the remaining sample is used as validation data. The process is then repeated k times by using each subsample once for validation. This method can give estimates of the variability of the true estimation error. In a data-rich situation, the best approach is to split the dataset into three parts [Hastie et al., 2001] as shown on Fig. 6(c): a training set used to fit the models; a validation set used to estimate prediction error for model selection; a test set to assess the generalization error of the selected model.

4.4 Example: 2R2C model

The Levenberg-Marquardt used for least-square estimation of the 2R2C model (see code in Appendix A.1) returns not only point values of the parameters, but also an estimation of the covariance matrix cov $\left(\hat{\theta}_{LS}\right)$. Using this matrix in Eq. 54, as is shown in the code, returns the standard deviation of estimated parameters and their correlation matrix.

- Results of the test for parameter significance are shown on Tab. 2: all parameters are considered relevant according to this analysis.
- The correlation matrix $\mathbf{R}_{\hat{\theta}}$ is shown on Tab. 4 for the assessment of parameter interactions.

Results show some strong interactions between parameters. The highest correlation coefficient is between R_2 and C_2. This can be illustrated by likelihood-based confidence regions. A grid of $R_2 - C_2$ values is first defined. On each point of the grid, the inverse problem is solved on all remaining parameters and the likelihood ratio function (Eq. 55) is calculated. This allows drawing two-dimensional likelihood-based

Table 4: Correlation matrix of the parameter estimates in the 2R2C example

	R_1	R_2	C_1	C_2	$T_e(0)$
R_1	1	-0.22	0.09	-0.41	-0.59
R_2	-0.22	1	0.37	0.79	-0.43
C_1	0.09	0.037	1	0.13	-0.24
C_2	-0.41	0.79	0.13	1	-0.10
$T_e(0)$	-0.59	-0.43	-0.24	0.10	1

confidence regions according to Eq. 56. This is illustrated by Fig. 7, which results from the code written in Appendix A.3.

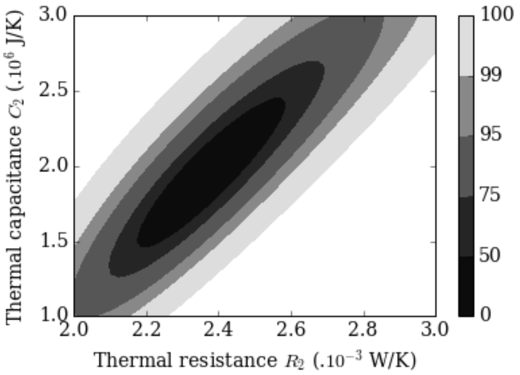


Figure 7: Likelihood-based confidence regions of the $R_2 - C_2$ parameter pair. Grey levels show quantiles of the χ^2 distribution

Although all parameters of the model are theoretically identifiable, there is a strong interaction between some of them which was not shown by the prior identifiability analysis. In this specific case, the data seems insufficient to properly identify the R_2 and C_2 parameters simultaneously, as is shown by the size of their likelihood-based confidence regions.

5 Improving the identification

What is meant here by *improving the identification* is the set of techniques that allow an optimal use of data, so that we learn the most of it without making the mistake of overfitting.

5.1 Model selection

When given a measurement data set from which to learn parameter values, it is tempting to choose a comprehensive model that attempts to explain all variations of the data. As already mentioned above, this might not be the wisest choice because of the risk of overfitting and poor performance prediction. If the goal of the inverse problem is not prediction, but only to characterise thermophysical properties from an experiment, a complex model will yield large uncertainty intervals for the parameter estimates.

The goal of model selection is to find exactly to what extent the data can be interpreted without coming to erroneous conclusions. The user is given the choice of several model structures, each representing the reality with a different level of detail, and must choose one that will accurately explain the observed phenomena without running into issues of practical identifiability and overfitting. Selecting the most appropriate model for prediction purposes is a matter of bias-variance tradeoff.

A comparison of several model selection criteria and statistical tests was proposed by [Posada and Buckley, 2004].

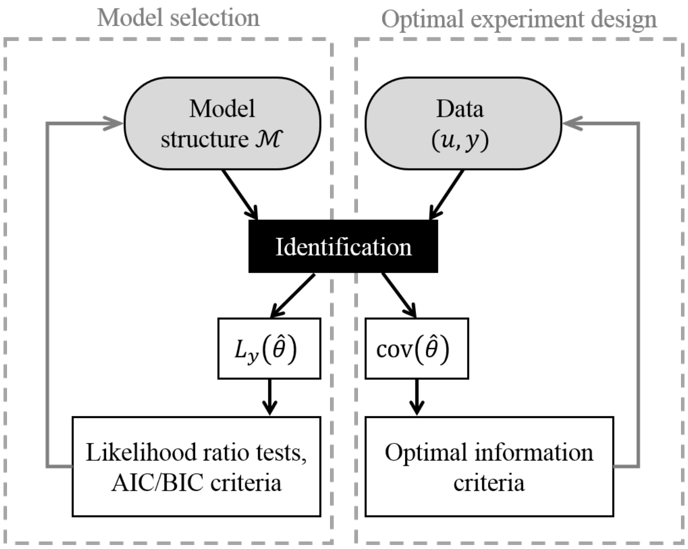

Model selection
Optimal experiment design
Model structure $\mathcal{M}$
Data (u, y)
Identification
$L_y(\hat{\theta})$
$\mathrm{cov}(\hat{\theta})$
Likelihood ratio tests, AIC/BIC criteria
Optimal information criteria

Figure 8: Iterative processes for model selection and optimal experiment design

Prediction error

The first possibility is to use the results of cross-validation as a criterion for model selection. The predictive capability of each model is assessed by one of the methods mentioned on Fig. 6, by using a different dataset from the training data. Whichever model yields the lowest prediction error (Eq. 62) is preferred. The disadvantage of this method is that is requires enough data to split it in separate sets. Moreover, the error estimates may tend to be biased to be conservative, indicating higher error than in reality [Hastie et al., 2001]

Information criteria

Model selection can also be done without using a separate dataset for cross-validation, by estimating the in-sample prediction error [Hastie et al., 2001]. Information criteria are an attempt to quantify the loss of information between an ideal model that generated the data, and a candidate model. They are applicable to parametric models that enable the calculation of likelihoods. Akaike's Information Criteria (AIC) is a function of the maximized likelihood of a model and of its number of parameters p. The Bayesian Information Criterion (BIC) is similar, but accounts for the number of observations N.

$$\mathrm{AIC} = -2 \ln L_y(\theta) + 2\,p \tag{63}$$

$$\mathrm{BIC} = -2 \ln L_y(\theta) + p \ln N \tag{64}$$

These criteria can be seen as a slight modification of the maximum likelihood criterion, with an additional term to penalize high model complexity. AIC tends to be preferred if the model is to be used for forecasting or control, and BIC is advised is the model is used for identification of physical properties. BIC tends to penalize complex models more and give preference to simpler models in selection [Hastie et al., 2001].

It is very simple to rank several calibrated candidate models by either their AIC or BIC: model selection simply comes down to choosing the model that minimizes them. The criteria are applicable in settings where the fitting is carried out by maximization of a log-likelihood. However, they may also be applied to least-square fitting, since it is equivalent to maximum likelihood estimation under the hypothesis of additive Gaussian noise.

Likelihood ratio

The third category of model selection is applicable to nested models: a set of hierarchically ranked submodels built by removing parts of a larger model. The hierarchical likelihood ratio tests consist in performing pairwise likelihood ratio tests in order of increasing complexity, until it is found that adding parameters is statistically irrelevant. A detailed application of this technique to building thermal models is shown by [Bacher and Madsen, 2011]. Its performance has been compared to other statistical tests by [Prívara et al., 2012].

Let us suppose a model with parameters $\theta \in \mathbb{R}^p$ and a simpler submodel $\theta_0 \in \mathbb{R}^{p_0}$ with $p_0 < p$. The likelihood ratio function is the ratio of the maximum likelihood values obtained with a dataset y over the respective search spaces of both models:

$$\lambda(y) = \frac{\max_{\theta_0} L_y(\theta_0)}{\max_{\theta} L_y(\theta)} \tag{65}$$

The maximum likelihood of the larger model is always higher than that of its submodel, but the question is whether this performance is significantly better in a statistical sense. The test statistic $-2\ln\lambda(y)$ is used to test the hypothesis H_0 that the data is better explained by the submodel, versus the hypothesis H_a that it is better explained by the larger model. It converges asymptotically to a χ^2 distributed variable with $(p - p_0)$ degrees of freedom [Madsen and Thyregod, 2010].

A detailed methodology of model description, fitting, selection and validation was shown by [Bacher and Madsen, 2011]. The authors defined seventeen RC models to describe the thermal behaviour of a test building with different numbers of states and parameters. They applied the likelihood ratio test to find the sufficient complexity and validated the selected model with the criteria of residuals autocorrelation function.

5.2 Optimal experiment design

While a model selection procedure searches among candidate models which one is most suited to explain one dataset, Optimal Experiment Design (OED) is the search for the dataset that will give the most informative training to a predefined model. This is particularly necessary when a model is defined by the needs of a project and cannot be simplified (material characterisation, fault detection, disaggregation of heat losses, etc.). For instance, an inverse problem of material characterisation must include the target physical properties in its parameters, and the possibilities of model selection are limited. The preferable way of improving parameter identifiability is through the enrichment of data.

Optimal experiment design can be defined as the search for the experimental setup and conditions of data acquisition that will maximise the information gain by a given model, at the lowest possible price. The "price" of data is measured in terms of the duration of the identification experiment, the perturbation induced by the excitation signal, the number of sensors, or any combination of these [Bombois et al., 2006, Gevers et al., 2009]. The goal of optimal experimental design is to minimise the estimate covariance $\mathrm{cov}(\hat{\theta})$ [Emery and Nenarokomov, 1998], and thus to improve the practical identifiability without redefining the model structure. Note that no design of experiment can make up for a lack of structural identifiability.

There are essentially three ways to improve the richness of measurements for identification purposes: performing longer experiments or adding sensors, i.e. acquiring more data points; carefully selecting sensor placement; imposing an enriched input signal. These options, especially the first one, result in an increase cost of the experiment. The third option requires an experiment in controlled conditions. In most of the optimal experiment design studies, the number of sensors and duration of the experiment are fixed settings, so that the point is to search for the optimal conditions (input signal and sensor positions) at a constant experiment cost. OED principles and optimality criteria have been reviewed by [Fedorov, 2010].

General guidelines

The first option is to define guidelines by expert knowledge for a more or less empirical enrichment of the inputs. Several guidelines are given in [Madsen, 2016] to make controllable inputs more informative:

- The excitation signal should be periodic with a frequency matching the time constants of the system. These time constants can be easily approximated in the case of linear systems, by writing them in the form of transfer functions.
- The amplitude of the fluctuations should be high, while remaining within realistic values.
- If the system has several inputs, they should be uncorrelated. For instance, if the system is subjected to solar radiation and a controlled excitation (such as indoor heating), then this controlled input should not present periodic fluctuations of 24 hour.

Several highly informative excitation signals have been designed for the identification of linear systems, such as the pseudo-random binary signal (PRBS) and harmonic signals [Godfrey, 1980].

Optimisation

If the system is complex, non-linear or black-box with little prior information on its time constants and expected response, it might be difficult to design an informative excitation for it from the above general guidelines. Another possibility is then to perform a dedicated optimisation of the information gained by the model from the experiment. As a reminder, this amount of information is quantified by the Fisher Information Matrix:

$$\mathbf{F}(\theta) = E\left[\left(\frac{\partial \ln L_y(\theta)}{\partial \theta}\right)^T \left(\frac{\partial \ln L_y(\theta)}{\partial \theta}\right)\right] \tag{66}$$

Under the usual hypotheses of additive white noise on the measurements, an equivalent information matrix can also be written without using the likelihood function [Walter and Pronzato, 1997, Agbi et al., 2012, Cai et al., 2016]:

$$\mathbf{F}(\theta) = \frac{1}{\sigma^2}\mathbf{S}^T\mathbf{S} = \sum_{i=1}^{N} \left[\frac{\partial \hat{y}_i(\theta)}{\partial \theta}\right]^T \frac{1}{\sigma^2} \left[\frac{\partial \hat{y}_i(\theta)}{\partial \theta}\right] \tag{67}$$

As already mentioned, the central limit theorem states that an unbiased maximum likelihood estimator is asymptotically gaussian. Its covariance has a lower bound given by the Cramer-Rao theorem [Walter and Pronzato, 1997]:

$$\operatorname{cov}\left(\hat{\theta}\right) \geq \mathbf{F}(\hat{\theta})^{-1} \tag{68}$$

As underlined by [Gevers et al., 2009], the existence of a finite covariance matrix relies on a positive definite information matrix. Increasing the size of the FIM thus reduces the uncertainty of the parameter estimates. In order to quantify the size of the FIM, several scalar indicators are available [Bastogne, 2008]. For instance, the D-optimality criterion is the determinant of the FIM, and the goal of OED is to maximize it:

$$\Psi_D = \det \mathbf{F} \tag{69}$$

This is a quantitative measure of practical identifiability. Other criteria include the minimum eigenvalue or the trace of the information matrix. It is preferable to bring all parameters within the same order of magnitude through normalizing constant before calculating the FIM and the optimality criteria.

A controllable input signal can be parameterised as a weighted sum of elementary functions or a harmonic signal, which allows reconstructing custom forms from a finite number of parameters. The experiment is described by these parameters and some additional experiment settings such as sensor positions. OED seeks the set of these experiment parameters that maximizes a scalar measure of information gain. It is an iterative procedure, since the optimal inputs are determined by supposing that the sought system parameters are known. It was found by [Gevers et al., 2009] that the solutions of optimal experiment design problems are most easily expressed in the form of multisines, i.e. input signals that have a discrete spectrum.

To this day, OED has seen more applications to the optimal estimation of heat transfer properties than to the field of building sciences. As a few examples: [Artyukhin and Budnik, 1985] researched the optimal sensor placement in the inverse heat conduction boundary problem. [Nenarokomov and Titov, 2005] estimate the emissivity of insulated materials and analyze the influence of the external heat flux on the accuracy of the solution of the inverse problem. [Karalashvili et al., 2015] use OED to design an optimal falling film experiment for the identification of mass transport coefficients. The D-optimality criterion is used to maximize the FIM. This paper is an interesting example of an incremential workflow implying OED and model selection, for a maximized utilization of data. [Berger et al., 2017] is an example of OED applied to a non-linear problem: the estimation of heat and moisture transfer properties of building materials; [Cai et al., 2016] proposed generating an optimal training data set for zone temperature setpoints to maximise the accuracy of parameter estimates in an RC building model.

6 Conclusion

What is referred to here as inverse problems are actually a broad field that encompasses any study where data is gathered and mined for information: material and component characterisation, building energy performance assessment from in-situ measurements, system identification for model predictive control... These scientific challenges are gaining visibility due to the increasing availability of data (smart meters, building management systems...), the increasing popularity of data mining methods, and the available computational power to address them. Many engineers and researchers however lack the tools for a critical analysis of their results.

A scientist who wishes to gain knowledge from measurements should follow a workflow that allows them to get the most information from data without making mistakes. The paper is an overview of guidelines to reach this aim:

- The identifiability of the model structure can be checked before solving the inverse problem, in order to avoid parameter interactions and redundant parameters.
- The impact of approximation errors cannot be overlooked. All models are biased to some extent, which adds up to the effect of measurement uncertainty, and may result in disastrous estimation errors if the proper validation steps are not carried.
- Validation steps should include residual analysis in order to diagnose unaccounted phenomena, and an estimation of the confidence regions of parameter estimates. Note that these confidence regions should not necessarily be trusted if the model bias is important.
- Choosing the appropriate model structure is a bias-variance tradeoff: a detailed model offers a better fit with data than a simple one, but an excessive complexity will offer poor prediction accuracy.
- The information gained by a model from an experiment can be measured by information indicators, and maximised by tuning the experiment design.

7 Acknowledgements

The author would like to thank the National Research Agency (ANR) for funding the present work through the BAYREB research project.

References

[Åström, 1980] Åström, K. J. (1980). Maximum likelihood and prediction error methods. *Automatica*, 16(5):551–574.

[Agbi et al., 2012] Agbi, C., Song, Z., and Krogh, B. (2012). Parameter identifiability for multi-zone building models. In *2012 IEEE 51st Annual Conference on Decision and Control (CDC)*, pages 6951–6956.

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

[Arridge et al., 2006] Arridge, S. R., Kaipio, J. P., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., and Vauhkonen, M. (2006). Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems*, 22(1):175.

[Artyukhin and Budnik, 1985] Artyukhin, E. A. and Budnik, S. A. (1985). Optimal planning of measurements in numerical experiment determination of the characteristics of a heat flux. *Journal of engineering physics*, 49(6):1453–1458.

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

[Bastogne, 2008] Bastogne, T. (2008). Experimental Modeling of Dynamical Systems - Applications in Systems Biology. Habilitation à diriger des recherches, Université Henri Poincaré - Nancy I.

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

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

[Beck and Woodbury, 1998] Beck, J. V. and Woodbury, K. A. (1998). Inverse problems and parameter estimation: integration of measurements and analysis. *Measurement Science and Technology*, 9(6):839.

[Bellman and Åström, 1970] Bellman, R. and Åström, K. J. (1970). On structural identifiability. *Mathematical Biosciences*, 7(3):329–339.

[Bellu et al., 2007] Bellu, G., Saccomani, M. P., Audoly, S., and D'Angiò, L. (2007). DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine*, 88(1):52–61.

[Berger et al., 2017] Berger, J., Dutykh, D., and Mendes, N. (2017). On the optimal experiment design for heat and moisture parameter estimation. *Experimental Thermal and Fluid Science*, 81:109–122.

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

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

[Bombois et al., 2006] Bombois, X., Scorletti, G., Gevers, M., Van den Hof, P. M. J., and Hildebrand, R. (2006). Least costly identification experiment for control. *Automatica*, 42(10):1651–1662.

[Brouns et al., 2013] Brouns, J., Nassiopoulos, A., Bourquin, F., and Limam, K. (2013). State-parameter identification for accurate building energy audits. In *Building Simulation 2013 conference*, page 7p, France.

[Brouns et al., 2017] Brouns, J., Nassiopoulos, A., Limam, K., and Bourquin, F. (2017). Heat source discrimination in buildings to reconstruct internal gains from temperature measurements. *Energy and Buildings*, 135:253–262.

[Cai and Braun, 2015] Cai, J. and Braun, J. (2015). An inverse hygrothermal model for multi-zone buildings. *Journal of Building Performance Simulation*, 9(5):510–528.

[Cai et al., 2016] Cai, J., Kim, D., Braun, J. E., and Hu, J. (2016). Optimizing Zone Temperature Setpoint Excitation to Minimize Training Data for Data-driven Dynamic Building Models. In *Proceedings of the American control conference*, pages 6–8, Boston.

[Chatzis et al., 2015] Chatzis, M. N., Chatzi, E. N., and Smyth, A. W. (2015). On the observability and identifiability of nonlinear structural and mechanical systems. *Structural Control and Health Monitoring*, 22(3):574–593.

[Chou and Bui, 2014] Chou, J.-S. and Bui, D.-K. (2014). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings*, 82:437–446.

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

[Deconinck and Roels, 2017] Deconinck, A.-H. and Roels, S. (2017). Is stochastic grey-box modelling suited for physical properties estimation of building components from on-site measurements? *Journal of Building Physics*, 40(5):444–471.

[D'Oca and Hong, 2015] D'Oca, S. and Hong, T. (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, 88:395–408.

[Dong and Andrew, 2009] Dong, B. and Andrew, B. (2009). Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. In *Proceedings of 11th IBPSA International Conference*, Glasgow, UK.

[Dong and Lam, 2011] Dong, B. and Lam, K. P. (2011). Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network. *Journal of Building Performance Simulation*, 4(4):359–369.

[Dong and Lam, 2014] Dong, B. and Lam, K. P. (2014). A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation*, 7(1):89–106.

[Dötsch and Van Den Hof, 1996] Dötsch, H. G. M. and Van Den Hof, P. M. J. (1996). Test for local structural identifiability of high-order non-linearly parametrized state space models. *Automatica*, 32(6):875–883.

[Emery and Nenarokomov, 1998] Emery, A. F. and Nenarokomov, A. V. (1998). Optimal experiment design. *Measurement Science and Technology*, 9(6):864.

[Fedorov, 2010] Fedorov, V. (2010). Optimal experimental design. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):581–589.

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

[Feroz et al., 2011] Feroz, F., Cranmer, K., Hobson, M., Austri, R. R. d., and Trotta, R. (2011). Challenges of profile likelihood evaluation in multi-dimensional SUSY scans. *Journal of High Energy Physics*, 2011(6):1–23.

[Florita and Henze, 2009] Florita, A. R. and Henze, G. P. (2009). Comparison of Short-Term Weather Forecasting Models for Model Predictive Control. *HVAC&R Research*, 15(5):835–853.

[Gevers et al., 2009] Gevers, M., Bazanella, A. S., Bombois, X., and Miskovic, L. (2009). Identification and the Information Matrix: How to Get Just Sufficiently Rich? *IEEE Transactions on Automatic Control*, 54(12):2828–2840.

[Godfrey, 1980] Godfrey, K. R. (1980). Correlation methods. *Automatica*, 16(5):527–534.

[Grandjean et al., 2017] Grandjean, T. R. B., McGordon, A., and Jennings, P. A. (2017). Structural Identifiability of Equivalent Circuit Models for Li-Ion Batteries. *Energies*, 10(1):90.

[Grewal and Glover, 1976] Grewal, M. and Glover, K. (1976). Identifiability of linear and nonlinear dynamical systems. *IEEE Transactions on Automatic Control*, 21(6):833–837.

[Hansen, 1990] Hansen, P. (1990). Truncated Singular Value Decomposition Solutions to Discrete Ill-Posed Problems with Ill-Determined Numerical Rank. *SIAM Journal on Scientific and Statistical Computing*, 11(3):503–518.

[Hansen, 1992] Hansen, P. C. (1992). Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review*, 34(4):561–580.

[Hastie et al., 2001] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *Elements of Statistical Learning: data mining, inference, and prediction. 2nd Edition.* Springer series in statistics. Springer.

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

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

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

[Huang and Yeh, 2002] Huang, C.-H. and Yeh, C.-Y. (2002). An inverse problem in simultaneous estimating the Biot numbers of heat and moisture transfer for a porous material. *International Journal of Heat and Mass Transfer*, 45(23):4643–4653.

[Jiménez et al., 2008] Jiménez, M. J., Madsen, H., and Andersen, K. K. (2008). Identification of the main thermal characteristics of building components using MATLAB. *Building and Environment*, 43(2):170–180.

[Joe and Karava, 2017] Joe, J. and Karava, P. (2017). Agent-based system identification for control-oriented building models. *Journal of Building Performance Simulation*, 10(2):183–204.

[Kaipio and Somersalo, 2007] Kaipio, J. and Somersalo, E. (2007). Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504.

[Kaipio and Fox, 2011] Kaipio, J. P. and Fox, C. (2011). The Bayesian Framework for Inverse Problems in Heat Transfer. *Heat Transfer Engineering*, 32(9):718–753.

[Kaipio and Somersalo, 2005] Kaipio, J. P. and Somersalo, E. (2005). *Statistical and Computational Inverse Problems.* Applied Mathematical Science. Springer Verlag, New York.

[Karalashvili et al., 2015] Karalashvili, M., Marquardt, W., and Mhamdi, A. (2015). Optimal experimental design for identification of transport coefficient models in convection-diffusion equations. *Computers & Chemical Engineering*, 80:101–113.

[Karatasou et al., 2006] Karatasou, S., Santamouris, M., and Geros, V. (2006). Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, 38(8):949–958.

[Karlsson et al., 2012] Karlsson, J., Anguelova, M., and Jirstrand, M. (2012). An Efficient Method for Structural Identifiability Analysis of Large Dynamic Systems. *IFAC Proceedings Volumes*, 45(16):941–946.

[Kramer et al., 2013] Kramer, R., van Schijndel, J., and Schellen, H. (2013). Inverse modeling of simplified hygrothermal building models to predict and characterize indoor climates. *Building and Environment*, 68:87–99.

[Kristensen et al., 2004] Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237.

[Künzel and Kiessl, 1996] Künzel, H. M. and Kiessl, K. (1996). Calculation of heat and moisture transfer in exposed building components. *International Journal of Heat and Mass Transfer*, 40(1):159–167.

[Lauret et al., 2005] Lauret, P., Boyer, H., Riviere, C., and Bastide, A. (2005). A genetic algorithm applied to the validation of building thermal models. *Energy and Buildings*, 37(8):858–866.

[Lin et al., 2012] Lin, Y., Middelkoop, T., and Barooah, P. (2012). 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)*, pages 6932–6937.

[Ljung, 1998] Ljung, L. (1998). *System Identification: Theory for the User, 2nd Edition.* Prentice Hall., 2nd edition.

[Lomas and Eppel, 1992] Lomas, K. J. and Eppel, H. (1992). Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings*, 19(1):21–44.

[Madsen, 2007] Madsen, H. (2007). *Time Series Analysis.* CRC Press.

[Madsen, 2016] Madsen, H. (2016). 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.*

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

[Madsen and Thyregod, 2010] Madsen, H. and Thyregod, P. (2010). *Introduction to General and Generalized Linear Models.* CRC Press.

[Maillet et al., 2010] Maillet, D., Jarny, Y., and Petit, D. (2010). Problèmes inverses en diffusion thermique - modèles diffusifs, mesures, sensibilités. *Techniques de l'ingénieur*, Transferts thermiques:be8265.

[Maillet et al., 2011a] Maillet, D., Jarny, Y., and Petit, D. (2011a). Problèmes inverses en diffusion thermique - Formulation et résolution du problème des moindres carrés. *Techniques de l'ingénieur*, Transferts thermiques:be8266.

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

[Meeker and Escobar, 1995] Meeker, W. Q. and Escobar, L. A. (1995). Teaching about Approximate Confidence Regions Based on Maximum Likelihood Estimation. *The American Statistician*, 49(1):48–53.

[Mirakhorli and Dong, 2016] Mirakhorli, A. and Dong, B. (2016). Occupancy behavior based model predictive control for building indoor climate - A critical review. *Energy and Buildings*, 129:499–513.

[Myung, 2003] Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100.

[Nenarokomov and Titov, 2005] Nenarokomov, A. V. and Titov, D. V. (2005). Optimal experiment design to estimate the radiative properties of materials. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 93(1âĂŞ3):313–323.

[Nissinen et al., 2008] Nissinen, A., Heikkinen, L. M., and Kaipio, J. P. (2008). The Bayesian approximation error approach for electrical impedance tomography-experimental results. *Measurement Science and Technology*, 19(1):015501.

[Oldewurtel et al., 2012] Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., Lehmann, B., and Morari, M. (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45:15–27.

[Palomo Del Barrio and Guyon, 2003] Palomo Del Barrio, E. and Guyon, G. (2003). Theoretical basis for empirical model validation using parameters space analysis tools. *Energy and Buildings*, 35(10):985–996.

[Palomo del Barrio and Guyon, 2004] Palomo del Barrio, E. and Guyon, G. (2004). Application of parameters space analysis tools for empirical model validation. *Energy and Buildings*, 36(1):23–33.

[Palomo Del Barrio et al., 1991] Palomo Del Barrio, E., Marco, J., and Madsen, H. (1991). Methods to compare measurements and simulations. In *Proceedings of the conference on Building Simulation IBPSA*, Nice, France.

[Pia Saccomani et al., 2003] Pia Saccomani, M., Audoly, S., and D'Angiò, L. (2003). Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*, 39(4):619–632.

[Pohjanpalo, 1978] Pohjanpalo, H. (1978). System identifiability based on the power series expansion of the solution. *Mathematical Biosciences*, 41(1):21–33.

[Posada and Buckley, 2004] Posada, D. and Buckley, T. R. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology*, 53(5):793–808.

[Prívara et al., 2012] Prívara, S., Váňa, Z., Žáčeková, E., and Cigler, J. (2012). Building modeling: Selection of the most appropriate model for predictive control. *Energy and Buildings*, 55:341–350.

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

[Rabouille, 2014] Rabouille, M. (2014). *Recherche de la performance en simulation thermique dynamique : application à la réhabilitation des bâtiments*. phdthesis, Université de Grenoble.

[Raue et al., 2014] Raue, A., Karlsson, J., Saccomani, M. P., Jirstrand, M., and Timmer, J. (2014). Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*, 30(10):1440–1448.

[Raue et al., 2009] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.

[Reynders et al., 2014] Reynders, G., Diriken, J., and Saelens, D. (2014). Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals. *Energy and Buildings*, 82:263–274.

[Ritt, 1950] Ritt, J. F. (1950). *Differential Algebra*. American Mathematical Soc.

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

[Rouchier et al., 2015] Rouchier, S., Woloszyn, M., Kedowide, Y., and Béjat, T. (2015). Identification of the hygrothermal properties of a building envelope material by the covariance matrix adaptation evolution strategy. *Journal of Building Performance Simulation*, 9(0):101–114.

[Schetelat and Bouchié, 2014] Schetelat, P. and Bouchié, R. (2014). ISABELE: a Method for Performance Assessment at Acceptance Stage using Bayesian Calibration. In *9th international conference on system simulation in buildings*.

[Sedoglavic, 2001] Sedoglavic, A. (2001). A Probabilistic Algorithm to Test Local Algebraic Observability in Polynomial Time. In *Proceedings of the 2001 International Symposium on Symbolic and Algebraic Computation*, ISSAC '01, pages 309–317, New York, NY, USA. ACM.

[Shumway and Stoffer, 2016] Shumway, R. H. and Stoffer, D. S. (2016). *Time series analysis and its applications*. Springer Texts in Statistics. Springer.

[Tikhonov and Arsenin, 1977] Tikhonov, A. N. and Arsenin, V. Y. (1977). Solutions of ill-posed problems.

[Virote and Neves-Silva, 2012] Virote, J. and Neves-Silva, R. (2012). Stochastic models for building energy prediction based on occupant behavior assessment. *Energy and Buildings*, 53:183–193.

[Walter and Pronzato, 1997] Walter, E. and Pronzato, L. (1997). Identification of parametric models. *Communications and Control Engineering*, 8.

[Wang et al., 2005] Wang, D., Federspiel, C. C., and Rubinstein, F. (2005). Modeling occupancy in single person offices. *Energy and Buildings*, 37(2):121–126.

[Wang and Zabaras, 2004] Wang, J. and Zabaras, N. (2004). A Bayesian inference approach to the inverse heat conduction problem. *International Journal of Heat and Mass Transfer*, 47(17-18):3927–3941.

[Yoshida et al., 2001] Yoshida, H., Kumar, S., and Morita, Y. (2001). Online fault detection and diagnosis in VAV air handling unit by RARX modeling. *Energy and Buildings*, 33(4):391–401.

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

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

A Python code

A.1 Least-square estimation of the 2R2C model

This is the Python code that runs the 2R2C parameter estimation described in Sec. 2.5. It relies on a datafile called data.csv, available upon contacting the author. Alternatively, the user can change the data importing section according to their needs. They can also use this template to test other types of models.

```
#==============================================================================
# Various imports and definitions
#==============================================================================
import numpy as np
from scipy.linalg import expm
from numpy.linalg import inv

def dot3(A,B,C):
    return np.dot(A, np.dot(B,C))

def stack4(A,B,C,D):
    return np.vstack((np.hstack((A,B)), np.hstack((C,D))))
```

```
#==============================================================================
# Importing the data
#==============================================================================
”””
In this example, data is contained in a file called data.csv with labeled columns
You should of course adapt this section to your case
”””
import pandas
dataset = pandas.read_csv('data.csv')

time_ = np.array(dataset['Time'])
T_in  = np.array(dataset['T_int'])  # indoor temperature
T_ext = np.array(dataset['T_ext'])  # outdoor temperature
q     = np.array(dataset['q'])      # indoor prescribed heat

delta_t = time_[1] - time_[0] # time step size
u = np.vstack([T_ext ,q]).T

#==============================================================================
# Definition of the model
#==============================================================================
”””
This is the simulation function of the very simple 2R2C model
This class is extendable to any other RC model structure
”””
def RC_model_simulation(time_, R1, R2, C1, C2, xe_0):

    # Matrices of the system in continuous form
    Ac = np.array([[-1/(C1*R1)-1/(C1*R2), 1/(C1*R2)],
                   [1/(C2*R2), -1/(C2*R2)]])
    Bc = np.array([[1/(C1*R1), 0],
                   [0, 1/C2]])

    # Matrices of the discretized state-space model
    F = expm(Ac*delta_t)
    G = dot3(inv(Ac), F-np.eye(2), Bc)
    H = np.array([[0, 1]])

    # Initialisation of the states
    x = np.zeros((len(time_), 2))
    x[0] = np.array((xe_0, T_in[0]))

    # Simulation
    for i in range(1,len(time_)):
        x[i] = np.dot(F, x[i-1]) + np.dot(G, u[i-1])

    # This function returns the second simulated state only
    return np.dot(H, x.T).flatten()

#==============================================================================
# Curve fitting
#==============================================================================
”””
This section evaluates the parameters of the model using observations T_int
Note that the initial condition on the unobserved state is an unknown parameter

You should provide an initial guess for the parameters
initial guess for resistances: 1e-2 W/K
initial guess for capacitances: 1e7 J/K
initial guess for the initial envelope temperature: 30 C
”””

from scipy.optimize import curve_fit

theta_init = [1e-2, 1e-2, 1e7, 1e7, 20]
popt, pcov = curve_fit(RC_model_simulation,
                       xdata = time_,
                       ydata = T_in,
                       p0 = theta_init,
                       method='lm')

# Calculating the indoor temperature predicted with the optimal parameters
T_in_opt = RC_model_simulation(time_, popt[0], popt[1], popt[2], popt[3], popt[4])
# Least square criterion for the optimal parameters
r_opt = np.sum((T_in_opt-T_in)**2)

”””
Test for parameter significance and correlation
”””
# Standard deviation of the parameter estimates
stdev = np.diag(pcov)**0.5
# Correlation matrix
R = dot3( np.linalg.inv(np.diag(stdev)), pcov, np.linalg.inv(np.diag(stdev)))
# t-statistic
t_stat = popt / stdev
```

A.2 Sensitivity analysis

This is the Python code that runs the sensitivity analysis which is part of the identifiability analysis of the 2R2C model, described in Sec. 3.3. The SALib[1] Python library is required. It is advised to run the code in the previous section first.

```
# SALib requires the 'problem' to be defined in a dictionary first
problem = {'num_vars': 5,
           'names': ['R1', 'R2', 'C1', 'C2', 'xe_0'],
           'bounds': [[1e-2, 3e-2],
                      [1e-3, 3e-3],
                      [1e7, 2e7],
                      [1.5e6, 2.5e6],
                      [25, 35]]}
```

[1] https://salib.readthedocs.io/en/latest/

```
# Create a matrix of model inputs for the FAST method
from SALib.sample import fast_sampler
X = fast_sampler.sample(problem, 5000, M=4)
# Evaluate the output on each parameter of the sample
Y = np.zeros(len(X))
for i, theta in enumerate(X):
    y    = RC_model_simulation(time_, theta[0], theta[1], theta[2], theta[3], theta[4])
    Y[i] = np.sum((y-T_in)**2)
# Analyse and return sensitivity coefficients
from SALib.analyze import fast
sa_results = fast.analyze(problem, Y, M=4)
```

A.3 Likelihood profiles

This is the Python code for calculating and drawing the likelihood-based confidence regions for the R_2 and C_2 parameters of the 2R2C model.

```
# Definition of a grid of values for R2 and C2
R2_vec = np.linspace(2e-3, 3e-3, num=50)
C2_vec = np.linspace(1e6, 3e6, num=50)
Ri, Ci = np.meshgrid(R2_vec, C2_vec)

# Series of optimizations where the R2 and C2 parameters are fixed
theta_init = [1e-2, 1e7, 20]
residuals = np.zeros_like(Ri.ravel())
for i in range(len(Ri.ravel())):

    def RC_model_simulation_R2C2fixed(time_, R1, C1, xe_0):
        return RC_model_simulation(time_, R1, Ri.ravel()[i], C1, Ci.ravel()[i], xe_0)

    popt, pcov = curve_fit(RC_model_simulation_R2C2fixed,
                           xdata = time_,
                           ydata = T_in,
                           p0 = theta_init,
                           method='lm')

    y     = RC_model_simulation(time_, popt[0], Ri.ravel()[i], popt[1],
                                Ci.ravel()[i], popt[2])
    residuals[i] = np.sum((y-T_in)**2)

# This is equivalent to the Likelihood ratio function
profile_likelihood = np.reshape((residuals-r_opt), (50,50))
# Quantiles of the chi2 distribution
percentiles = [0, 50, 75, 95, 99, 100]
levels = [ np.percentile(np.random.chisquare(2, size = 9000), q=_)
           for _ in percentiles ]

# Plotting
fig = plt.figure()
cax = plt.contourf(Ri*1000, Ci/1e6, profile_likelihood,
                   levels = levels, cmap = 'Greys_r')
```

```
plt.xlabel('Thermal resistance $R_2$ (.$10^{-3}$ W/K)')
plt.ylabel('Thermal capacitance $C_2$ (.$10^6$ J/K)')
cbar = fig.colorbar(cax, ticks=levels)
cbar.ax.set_yticklabels(percentiles)  # vertically oriented colorbar
```