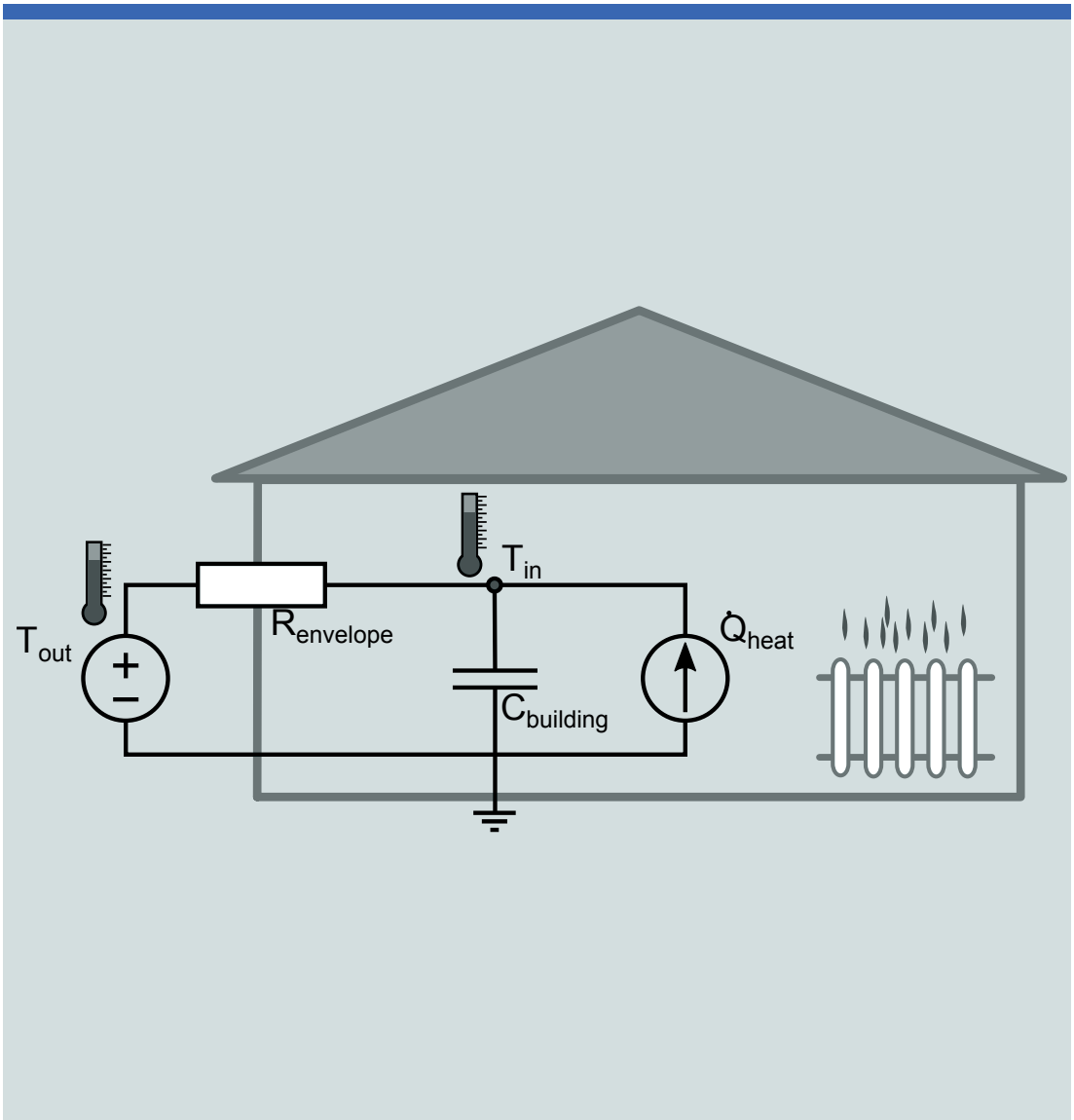Ole Magnus Hamre Brastein

Parameter estimation and analysis for grey-box models of building thermal behavior

USN

Ole Magnus Hamre Brastein

Parameter estimation and analysis for grey-box models of building thermal behavior

A PhD dissertation in
Process, Energy and Automation Engineering

In loving memory of my grandfather Idar Hamre (27.11.1927 - 12.05.2019). Thank you for inspiring my journey as an engineer.

I dedicate this thesis to my children; Idar Andreas, Halvor Kristoffer, Viktoria Marie and Maria Sofie.

Preface

This thesis is submitted to the University of South-Eastern Norway (USN) in partial fulfilment of the requirements of the degree of Philosophiae Doctor (PhD) in the Process, Energy and Automation Engineering program. The work has been conducted as part of the Self-Monitoring, Analysis and Reporting Technologies (SMART) research group at USN. The PhD project work has been conducted under the supervision of Professor Nils-Olav Skeie, with co-supervision by Professor Carlos F. Pfeiffer and Associate Professor Roshan Sharma.

The thesis consists of five scientific papers, four journal papers and one conference paper. The thesis is divided in two main parts. The first part presents the background and research objectives, an extensive literature review and some additional notes on the experimental setup. Further, a summary of relevant methods and a summary of the five scientific papers is given. Finally, the main results from all five papers are summarised and viewed together before the work is concluded with some observations of what has been achieved towards the research objectives and what could be further achieved in continuation of the presented thesis.

Porsgrunn, September 13, 2020

Ole Magnus Hamre Brastein

Acknowledgment

I would like to express my sincere thanks to my main supervisor Nils-Olav Skeie. I am forever grateful to you for your expert tutelage; for being both supportive and patient, for challenging me to work towards the research goals and for keeping me on track when my ever widening interests in different topics challenged my focus. I would also like to thank my co-supervisors Carlos F. Pfeiffer and Roshan Sharma for their insight, advice and many fruitful discussions. A further sincere thank you also to my other co-authors; Bernt Lie, Ali Ghaderi and D.W.U. Perera, for their insights and contributions on the presented papers. Finally, thanks to my PhD student colleges, in particular Dr. Liubomyr Vytvytskyi, for the discussions, presentations and debates over topics of common interest. Thank you all for sharing your knowledge and insights, and for pushing me towards becoming a better engineer and scientist.

A PhD project is a challenging undertaking in the best of circumstances, and without the support of those closest to me the research work presented in this thesis would not have been possible. I would therefore like to express my heartfelt gratitude to my family, in particular my wife Elena and my children; Idar Andreas, Halvor Kristoffer, Viktoria Marie and Maria Sofie, for their continued patience, support and understanding.

Summary

Reduction of anthropogenic CO_2 emissions is one of the most important scientific endeavours of our time. Space heating of buildings is responsible for a considerable portion of the worlds total energy consumption. The Energy Performance of Buildings Directive, issued by the European Union, estimates that approximately 20% of the energy consumption within the EU is the result of heating, ventilation and air conditioning in buildings. Consequently, the reduction of energy consumption in buildings has received significant scientific attention. Towards this goal, methods for creating *models* of building thermal behaviour is an important subtask.

The first of two main goals within building thermal behaviour modelling research is to create models that can accurately predict *future* thermal behaviour of buildings. The second, but equally important, goal is construction of models that can be used as *classification tools* to evaluate the thermal performance based on data collected from a specific building. The former of these goals aims to reduce energy consumption by improved control of temperature thus reducing the amount of energy required to maintain comfortable living conditions. The latter approach is useful towards understanding energy demands of individual buildings, such that the building occupants and owners can make qualified decisions on what energy conserving measures to implement, and also for the authorities to compose taxation schemes based on energy efficiency.

Modelling building thermal behaviour is challenging due to the complex nature of buildings, i.e., use of a wide variety of materials and different building geometries. Further, the physical buildings often does not match the building specifications and blueprints, due to workmanship issues or continued modification and renovation of existing buildings. Additionally, weather conditions and occupant demands makes experimental design difficult. Because of these inherent uncertainties involved in building thermal modelling it is useful to formulate such models as stochastic differential equations. This type of models, often called grey-box models, allows the combination of prior expert knowledge with parameters that are calibrated to fit a specific building. This approach produces models that tends to provide good prediction accuracy for future behaviour while also being interpretable by humans. The stochastic modelling framework has a strong mathematical foundation which provides a framework that can be used to estimate parameters, analyse estimation uncertainty, and to perform model selection and validation. The grey-box modelling

framework also fits naturally with the Bayesian statistics framework and the Markov Chain Monte Carlo methods, which has gained popularity over the recent years.

Grey-box models of building thermal behaviour are typically simplified description of the physics involved. Since the models are constructed using prior system specific knowledge, the parameters are often cognitively connected to the thermal properties of the physical buildings, i.e., the model parameters are used as *soft-sensors*. However, interpreting model parameters as representative of the physical properties of the building requires a careful analysis of the parameter identifiability to ensure that the calibrated parameters are unambiguous, and to estimate the uncertainty of the obtained parameters.

In this thesis, the stochastic modelling framework is combined with Kalman filter implementations that does not require differentiable models. This allows estimation of parameters for externally simulated models which facilitates experimentation with model structures. Further, the grey-box parameter estimation uncertainty is analysed using several different methods, including the Profile Likelihood framework, and the extended Profile Posterior method. Both profiling methods are extended to create 2D profiles which allows more detailed identifiability analysis of the parameter space. The Profile Posterior method is compared to the results obtained using Markov Chain Monte Carlo methods. The combination of model formulation as stochastic differential equations with Markov Chain Monte Carlo methods offers a particularly powerful and efficient model calibration framework, which can be utilised also for calibration of external software simulations.

The use of stochastic model formulations is applicable to a wide range of modelling challenges. Given that almost every conceivable model is in some way an approximation of the real system, the stochastic differential equation parameter estimation framework has been argued as a natural framework for modelling dynamic system models in general. The benefits of performing model calibration utilising a framework with a solid statistical foundation that provides tools for model validation and parameter identifiability analysis well out-ways the complexities of the methods involved.

Contents

Part I

Overview

Nomenclature

Symbol	Explanation
ADMM	Alternating direction of methods of multipliers
ANN	Artificial Neural Network
ARMAX	Autoregressive moving average with exogenous input
ARX	Autoregressive with exogenous input
BEMS	Building Energy Management System
CI	confidence interval
CP	cumulative periodogram
CTSM	Continuous Time Stochastic Modelling
DAQ	data acquisition
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
EPBD	Energy Performance of Buildings Directive
EU	European Union
GA	Genetic algorithm
HLC	Heat-loss coefficient
HVAC	Heating, ventilation and air conditioning
KF	Kalman Filter
KS	Kolmogorov-Smirnov
LLM	Local linear model
LoLiMoT	Local linear model tree
MC	Monte Carlo
ML	Machine Learning
MPC	Model Predictive Control
N2SID	Nuclear Norm Subspace Identification
NARX	Non-linear ARX
NSGA-II	Non-dominated sorting genetic algorithm
PDE	partial differential equation
PL	Profile Likelihood
PP	Profile Posterior
RMSE	Root mean square error
SDE	Stochastic Differential Equation
SSE	Sum of square error
SSID	Subspace Identification
TN	Thermal Network
TS	Takagi-Sugeno fuzzy model
UKF	Unscented Kalman Filter
ZC	zero crossing

1 Introduction

1.1 Background

Reduction of anthropogenic CO_2 emissions is perhaps the most important scientific endeavour of our time. According to the Energy Performance of Buildings Directive (EPBD) [1] the energy consumed by buildings accounts for 40% of the total energy consumption within the European Union (EU). Heating, ventilation and air conditioning (HVAC) account for 50% of the total energy consumption in buildings [2, 3]. Hence, reducing energy consumption in buildings has received significant scientific attention. While modern construction techniques can produce energy efficient *new* buildings [1, 4], renewal rates of buildings are low; around 0.4% to 1.2% according to EPBD [1]. Hence, it is of interest to study methods for energy reduction that can be applied to *existing* buildings. A promising solution is the use of *models* that can *predict* the future thermal behaviour of a building, including the *heating* and *cooling times* under some expected weather and occupancy conditions. By application of Model Predictive Control (MPC), as part of a Building Energy Management System (BEMS), energy consumption can be minimised based on predicted weather conditions and building usage patterns [5]. The *first* use-case for models of building thermal behaviour of interest in this work is therefore the *prediction* use-case of forecasting future temperatures in buildings.

Another important tool in the fight against climate change is the use of *classification* schemes, typically used to determine the level of *taxation* based on the idea that *excessive* consumption of energy and/or generation of pollution, with respect to some predefined norm, should be penalised by a higher cost. For buildings, energy classification schemes are typically based on building *specifications*, i.e., technical documentation of buildings including dimensions of the building, material choices and the use of HVAC systems [4]. However, physical buildings can differ significantly from their documented specifications, due both to workmanship issues and the typically continuous process of building modifications [6]. The possibility of using *calibrated* models to estimate the *actual* thermal properties of buildings based on *in situ* measurements is therefore a *second* use-case for building thermal behaviour models [6–9]. A model that is used to estimate a *property* of a physical system is often denominated as a *soft-sensor*.
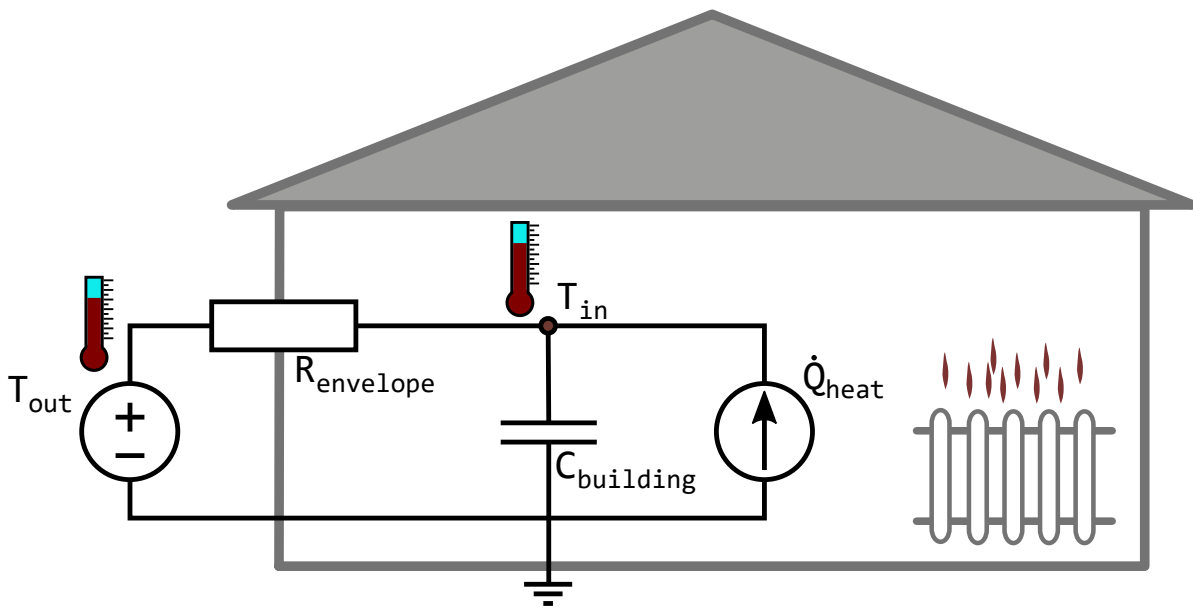
Figure 1.1: The figure shows an example of the simple 1R1C model, in which the entire building envelope is modelled by a single parameter R_{envelope}. The thermal capacitance is similarly described by the single parameter C_{building}. The model inputs are the outdoor temperature T_{out} and the supplied heat energy $\dot{Q}_{\text{heat}}$. The model predicts the indoor temperature T_{in}.

Models can be created based on two distinctly different methodologies. One possibility is the creation of *white-box* models based on human *expert* knowledge of building physics typically obtained from detailed building specifications. This approach offers the advantages of utilising prior knowledge of the building in the model, and also producing models with parameters that have a *prescribed* physical definition, e.g. thermal properties of building materials. However, the creation of detailed white-box models are time-consuming and require considerable human effort. An alternative type of models, often denominated *black-box* models, are created based on some *general* mathematical structure with a number of coefficients or parameters that are *calibrated.* Black-box model parameters are optimised such that the model is *fitted* to a set of measured input/output responses recorded from the physical building. The black-box approach has the advantage of being calibrated to a *specific* building, without use of possibly erroneous buildings specifications, which typically results in improved prediction accuracy. However, since a generalised mathematical structure is used, it is typically difficult to *analyse* the parameters of a black-box model to gain any insight into the *thermal properties* of the physical building.

A third, intermediate approach is the use of a simplified *thermal network (TN)* model, an example of which is shown in Fig. 1.1. TN models are constructed by interconnecting temperature *point nodes,* distributed throughout the building, by *lumped* thermal resistances and capacitances [10]. Heat-energy supplied by active

White-box   Grey-box   Black-box

None   Calibration   Full

Full   Physical information   None
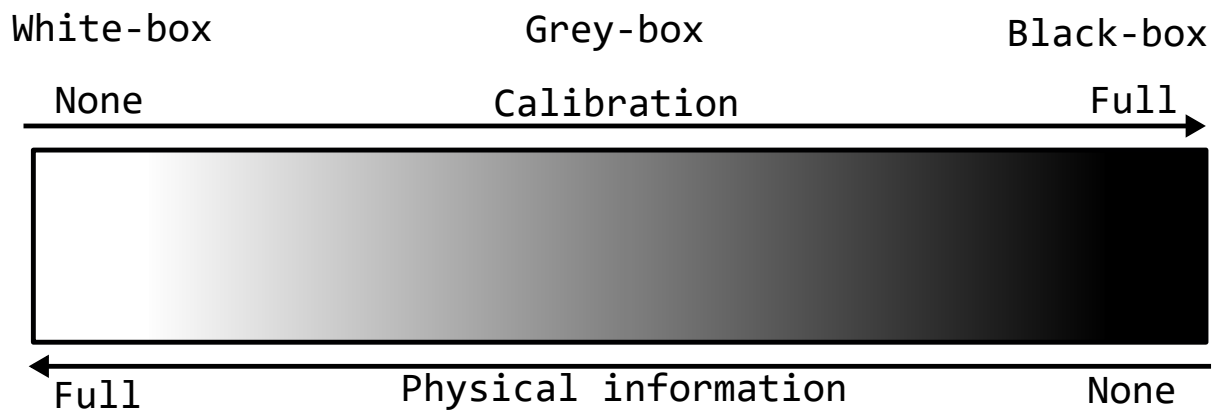
Figure 1.2: The figure illustrates the relationship between white-, grey- and black-box models. A pure white-box model depends only on physical information, while a pure black-box requires no physical information thus requiring full parameter calibration. Grey-box models are combinations of these extremes, utilising both prior physical information and parameter calibration.

heating systems is models as a current or energy flow source, while temperatures that are considered unaffected by the building, e.g. the outdoor temperature, is modelled as a voltage potential source. The TN structure is typically based on a *cognitive* description of the thermodynamics of the building [11]. The lumped thermal resistance and capacitance parameters, denominated as the vector $\boldsymbol{\theta}$, are calibrated from in situ measurements. Typically, these networks are expressed as an electric resistance/capacitance circuit *analogue*. The TN models offer a way to combine *some* a priori physical insight into the thermal behaviour of the building with *calibrated* parameters. For the *prediction* use-case, a wide-spread application of MPC technology in BEMS also for *residential* buildings requires a modelling methodology that does *not* depend on expensive human expertise. For the *soft-sensor* use-case, a general classification scheme requires *comparison* of models for different buildings, something which is arguably easier to accomplish with the simplistic TN model structures than with complex detailed white-box models. Hence, given the need for both facilitating *energy reduction* for HVAC and *improved classification* schemes based on in situ measurements, TN models have received significant scientific interest [6–9, 12–15].

TN models belong to a class of models denominated as *grey-box*, since they are derived by a combination of white- and black-box techniques, as illustrated in Fig. 1.2. As is the case for TN models, grey-box models are typically based on a *simplified* description of the system, hence they contain significant *epistemic* uncertainty. Parameter estimation for grey-box models is therefore a challenging task, since *both* the epistemic uncertainty in the model and the *aleatoric* uncertainty induced by *measurement noise* in the recorded calibration data should be accounted for. A convenient framework for formulation and parameter estimation in grey-box models is given in [16]. Based on expressing the model as a *stochastic differential equation*

(SDE) and an *algebraic* measurement equation*,* which facilitates the expression of both sources of uncertainty, [16] presents a statistically well founded framework for estimating the grey-box parameters. It is interesting to note that most white-box models contain *some* approximations to the real system and are therefore in some sense grey. This observation has lead to an argument for the grey-box approach as a general framework for the modelling of dynamic systems [17].

The TN grey-box model approach has shown promising results for both the *prediction* and the *soft-sensor* use-cases. By *learning* the building thermal behaviour from in situ measurements, appropriately designed TN models tends to *generalise* well and therefore provide *realistic* predictions of future behaviour [6, 12, 14, 15]. There are also several examples in the literature that utilise TN models to estimate thermal *properties* of buildings [6, 7, 9]. However, the structure of a TN grey-box model is greatly simplified which challenges the *interpretation* of the lumped parameters as *representative* estimates of the *thermal properties* of the physical building. Further, TN models are developed *cognitively [11],* often over many iterations that gradually increase in model complexity [13]. This process leads to developed models that can become *over-parametrised* resulting in ambiguous parameter estimates [18]. Finally, calibration of building thermal behaviour models is typically performed using *observational* data of the weather, and measurements of building temperatures with constrained variability due to occupant requirements. Given only limited experimental control, the acquired data may not be sufficiently informative on the building thermal dynamics. Because of these challenges the interpretation of lumped TN parameters as physical properties of the building benefits from an *in-depth* analysis of the estimated parameters [6]. Manual calculation of the *lumped* TN parameters is usually infeasible, hence validating the *physical interpretation* of a calibrated TN model should be based on statistical testing of the estimated parameters. Arguably, a *necessary* requirement for interpretation of estimated parameters in the *soft-sensor* use-case is that the optimal estimate is both *unambiguous* and reasonably *consistent* across different data-sets.

1.2 Objectives and scope

The current scientific literature on TN models provides a solid framework for *calibration* of model parameters and *comparison* of different model structures [19, 20]. However, much of the existing literature assumes the TN model parameters to be physically determined constants [6]. The previously discussed challenges of epistemic uncertainty induced by strong simplifications, possible over-parametrisation, and lack of dynamic information content in observational data, can result in parameters for which the assumption of physical interpretation may not hold. This topic has, with a few recent exceptions, received limited attention for TN models [6]. The main research objective of this work is therefore analysing the use of TN models

Scope / Papers	1 Simplifying model structure experimentation	2a Identifiability and uncertainty of θ	2b Consistency of θ for different data	3 Resolving ambiguity of θ	4 Prediction uncertainty for new data
A					
B					
C					
D					
E					

Figure 1.3: An overview of the papers shows how the papers (A-E) are distributed over the numbered points (1-4) in the defined scope. Colours are used to separate the different elements of the defined scope.

as soft-sensors for building thermal properties by studying ambiguity of estimated parameters, and also potential remedies to resolve parameter ambiguity. Methods to test for *structural* and *practical identifiability [21]* of estimated parameters is of particular interest [6]. Additionally, methods for validating the consistency of the estimated parameters across different calibration data is studied to further verify the soft-sensor use-case.

A second research objective of this work is to find methods that can simplify TN model evaluation during parameter calibration. Deriving *differential equations* from a TN model is somewhat labour intensive [18], and the iterative model development process typically results in several model structures that must be calibrated [13]. Therefore, methods that can simulate TN models without the need to develop the differential equations may facilitate model building, and could potentially allow automated model construction.

The scope of the work is divided on four areas as illustrated in Fig. 1.3, where points 1 and 2 are the most important.

1. Develop a method for simulating a TN model with a given set of parameters, *without* manually obtaining the differential equations, that is usable in parameter estimation. Given that TN models can be expressed as electric circuit analogues, it is interesting to consider how simulation of circuits is performed in the *electronics* field.

2. Validate if estimated parameters of a TN model can be assumed *representative constants* of the buildings thermal properties. This topic is divided into two separate areas:

a) Analysing the *uncertainty* and *ambiguity* of the estimated parameters, and in particular diagnosing *structural* and *practical identifiability.*

b) Testing the *consistency* of the estimated TN parameters for *different* data acquired from the same physical building.

3. Investigate what remedies may be applied if the parameters for a specific TN model is found *not* to fulfil the criteria required for interpretation as physical properties.

4. TN model parameters are estimated under significant uncertainty. It is therefore interesting to study how parameter uncertainty is propagated to the model predictions for new independent data.

1.3 Contributions

The contributions of this work are distributed over five scientific papers, labelled A to E. The papers cover the objectives and scope as defined in Fig. 1.3 and outlined in Section 1.2 as follows:

1. In Paper B [22], the grey-box parameter estimation framework used in [16] is *adapted* to using an *external* simulation software tool to simulate TN models described as a *list* of interconnected resistances and capacitances.

2. Parameter uncertainty, ambiguity and consistency is covered in all papers as follows:

    a) The ambiguity of the estimated parameters is studied by use of *randomisation* in Paper A [23] and by application of the *Profile Likelihood* (PL) method to analyse parameter *identifiability* [21, 24, 25] in papers B to E [22, 26–28]. In particular, the PL method is extended to create two-dimensional profiles (PL2D) which can be used to analyse parameter inter-dependence [22]. The PL2D method is further refined for improved computational efficiency in Paper C [26]. In Paper E [28] this method is compared with the Markov Chain Monte Carlo (MCMC) method.

    b) The consistency of estimated parameters is studied in Paper A [23] by comparing results from different independent data-sets. In Paper C [26], data recorded *simultaneously* from two different sensors is analysed to show the effect of sensor placement on TN model calibration. The use of *non-parametric block-based bootstrapping* [29–31] methods is investigated in Paper D [27], where also the *consistency* of dynamic information *over time* is analysed by application of the PL method to a *sliding window.*

3. Resolving ambiguous parameter estimates typically requires adding new information to the estimation problem. In Paper A [23], ambiguity is resolved by fixing one of the parameters to a value precomputed from building specifications. An alternative solution is presented in Paper E [28] where the application of a *prior* distribution determined from the building specifications is used to resolve the ambiguity.

4. Paper E [28] discusses the MCMC method and the Bayesian interpretation of parameter estimation. The representation of parameter uncertainty obtained by MCMC is used to compute a *stochastic forecast* that estimates the *uncertainty* of building temperature *predictions*.

1.4 Outline of thesis

This thesis consist of two main parts. The first part consists of seven chapters, where the Chapter 1 presents the background for the project, the research objectives and the scope of the work. Chapter 2 provides an extensive literature review of the relevant fields. Chapter 3 presents additional information on the experimental setup used to acquire data for calibrating grey-box models. Chapter 4 is a summary of selected methods relevant to the defined scope and used in the papers, with some overviewing remarks as to how the methods can be applied together. Chapter 5 presents short summaries of each of the five scientific papers, while Chapter 6 discusses the main results from the papers. Finally, the thesis is concluded in Chapter 7.

2 Literature Review

Given the ongoing effort to reduce energy consumption in buildings, a large body of literature exists that applies many different methods towards the creation and calibration of building thermal behaviour models. Particular focus has been given to the *prediction* use-case, towards application of MPC [5], but there are also a significant body of literature which takes interest in the estimated parameter values for the *soft-sensor* use-case. The following review of building thermal behaviour models provides an overview of relevant methods, and how the scientific papers and contributions of this work is connected to existing literature. While the primary focus is grey-box models, an overview of selected applications of white- and black-box modelling is also given. It is particularly interesting to observe that the demarcation between white- and grey-box models is not always obvious. The term grey-box can be applied in the specific sense of an SDE model, as defined in [16], but also the wider sense of any model constructed from a description of the system physics but with calibrated parameters. For further literature reviews on modelling of building thermal behavior see e.g. [19, 32, 33].

2.1 White-box models

A white-box model is developed as a detailed *mathematical description* of the physics of a particular system, usually developed by starting from *first principles* such as *energy and mass balances*. A recent example of this approach applied to building thermal behaviour is found in [34], where a *multi-floor* building model is developed in both Modelica, by use of the graphical interface Dymola and based on the Modelica Buildings Library developed by Lawrence Berkeley Library, and by implementation of state-space differential equations in MATLAB. The equations where inherited from a previous paper by the authors [35], where a *multi-zone* building model is developed based on *mass- and energy-balances.* Particular attention is given to potential heat transfer by *convection,* due to the flow of air both between zones and floors of the building and between building and environment from ventilation [34]. Another interesting feature of these papers, developed in yet another previous paper by the same authors [36], is the use of *spherical* analogue simplified model for the furniture described by a *partial differential equation* (PDE) in spherical coordinates that is developed to have *equivalent average thermal diffusivity* with the furniture

in the building. The combined work in these and other papers by the same authors presents a detailed framework for the construction and implementation of white-box models of multi-zone multi-floor buildings. The developed models tends to be complex, with a large number of states and parameters that must be identified from specifications of the buildings, as is typical of the white-box modelling paradigm. In order to obtain acceptable model fit certain parameters require *tuning*, which is typically achieved by manual adjustment of parameters during repeated simulations [34].

It is interesting to note that both the use of an *equivalent spherical model approximation* for the furniture, and the tuning of certain model parameters to obtain acceptable model fit, arguably shifts the model classification towards that of grey-box models. Indeed, parameter *calibration* of white-box models may in general benefit from the application of grey-box parameter estimation, as pointed out by [17]. Some of the models in [34–36] are implemented in *simulation tools* without use of explicit equations, which is typically for white-box building models [37, 38]. In such cases, a grey-box model treatment for parameter estimation of external software, as outlined in point 1 of the scope defined in Section 1.2 and similar to that presented in Paper B [22], may be beneficial.

2.2 Black-box models

Black-box modelling is a popular approach to modelling of dynamic systems, as evident by the large body of literature, where models are developed by calibrating the parameters of a general mathematical structure. Many different model structures, often with specialised algorithms for calibration, exist in the literature. Naturally, black-box modelling is mostly applicable to the *prediction* use-case. Some classical approaches [39, 40] include the use of *autoregressive moving average with exogenous input (ARMAX)* models and *subspace identification (SSID)*. An example of SSID, i.e., the use of *"orthogonal* and *oblique* projections to obtain the Kalman state sequence"* [41, 42], of building thermal behaviour is found in [41]. The model structure is typically the standard stochastic *linear time invariant* (LTI) system description used in the Kalman filter literature [43]. The typical subspace identification approach stacks measurements into *Hankel* matrices [40], for both system inputs and outputs separately and combines the matrices of the LTI model into a *Toeplitz* matrix [40]. The resulting system of equations can then be solved as a *least squares* problem to obtain the LTI system matrices [41]. The approach typically uses an estimate of the Kalman gain K in order to obtain the *residuals* of both the state transition and the measurement equation [40, 44]. In [41] the *Autocovariance Least-Squares (ALS) method* [45] to estimate the process and measurement noise covariance matrices, $\mathscr{W}$ and $\mathscr{V}$, which is subsequently used to compute K [41, 42]. Estimating noise covariance matrices $\mathscr{W}$ and $\mathscr{V}$ is generally a difficult problem. An

alternative adaptive method of estimation is presented in [46]. A further alternative is the use of numerical optimisation of some appropriately defined objective function, such as the approach used in Papers C to E [26–28] and also in [20]. A similar, more recent, application of SSID for building thermal behaviour is found in [47], in which the LTI system model is identified by use of the *Nuclear Norm Subspace Identification (N2SID)* method [48]. The N2SID method seeks to utilise two *structural properties* of typical SSID methods; the low rank obtained by multiplying the *extended observability matrix* with the *state sequence matrix,* and the *block structure* of the Toeplitz matrices [48]. This is achieved by utilising the nuclear norm $\|\cdot\|_*$ as a heuristic approximation of the $\mathbf{rank}\,(\cdot)$ operator [49], to minimise the system order, while also optimising the model fit $\|\hat{y} - y\|_2^2$, as a combined objective function. A weighting parameter λ is tuned to obtain a reasonable balance between model complexity and model fit [48, 50]. The optimisation problem is solved in [47] by application of *alternating direction methods of multipliers* (ADMM) algorithm [50].

In recent years, the black-box modelling field of *machine learning (ML)* has received enormous scientific and popular interest, in particular the use of *artificial neural networks (ANN)* and *decision trees (DT)* [51–54], due in part to the application of ML methods to complex problems such as playing games [55] and driving a car [56]. Both SSID and ARMAX models are *linear* model structures which limits their usefulness for complex non-linear systems. ML methods has found application in modelling of building thermal behaviour as well. An example of the use of a DT method is found in [57], in which a *local linear model tree (LoLiMoT)*, consisting of several *local linear models (LLM)* identified for separate partitions of the sample space, is used to model building thermal behaviour. The resulting model is equivalent to a *Takagi-Sugeno (TS)* fuzzy model, in which local ARX models are used to describe non-linear system behaviour by local linear approximation [57]. The presented model structure is identified by the specialised LiLoMoT algorithm to minimise the SSE between simulation and measurement over a calibration data-set from a step test [57]. The model is validated by predicting an independent data-set recorded during normal building occupancy [57]. The approach shows RMSE well within acceptable performance for the predictive control use-case [57]. The resulting model is also fast in forward execution and is therefore considered a good choice for use in MPC [5].

Another ML based approach towards non-linear modelling of building thermal behaviour is given in [58], where an ANN based on radial base functions is used to model the predicted mean vote (PMV) index. The PMV index is used to predict the average vote of a large group of people on a thermal sensation scale [58, 59]. One such scale is prosed by the American Society of Heating Refrigerating and Air Conditioning engineers (ASHRAE) taking the values $[-3,\ 3]$ to represents human qualitative thermal sensation from *cold* to *hot* [58, 60]. Since the computation of the PMV index is time consuming, hence difficult to utilise in MPC under strict

time response demands for the control loop, [58] proposes to train an ANN as a *surrogate* model to estimate the PMV index. The developed MPC algorithm is used to control the HVAC system by a discrete set of control actions to obtain the target PMV index using a *branch and bound* [61] search for optimisation [58].

2.3 Grey-box models

Buildings are constructed using a large number of materials with different thermal properties, typically combined in complex geometries to form individual segments such as walls, roofs and floors. This makes detailed physical white-box modelling of individual parts challenging, as illustrated in the review in Section 2.1. Additionally, the available specifications of the physical building structures are often inaccurate or insufficient for detailed white-box model construction. Hence, the use of *calibrated* models has received significant scientific attention in the field of building thermal behaviour modelling. Given that there is often *some* physical information available, at least the overall layout of walls, doors and windows, the use of grey-box models in the form of *thermal networks (TN)* has been an especially popular modelling approach [10].

2.3.1 TN models expressed as stochastic differential equations

Since grey-box models are in general strong *approximations* of the real system, it is convenient to formulate them as *stochastic differential equations (SDE)*. A seminal work on the formulation of TN models as SDEs, with a statistically solid framework for parameter estimation, is found in [12]. By introducing the *diffusion* term in the state transition model and maximising the likelihood function to obtain calibrated parameters, a statistically solid framework for model calibration and selection is obtained [12, 16]. The diffusion term is used to model a combination of errors caused by model approximations, unmodelled and unrecognised disturbances and noise on the input measurements $u_{[N]}$ [12, 16]. It is interesting to note that the framework and software used in [12] was an earlier *linear* version of the *Continuous Time Stochastic Modelling (CTSM)* framework later presented in [16, 62, 63]. In CTSM, a *Kalman Filter (KF)* [64–66] is used to estimate *one-step-ahead* predictions and compute $\varepsilon_{k|k-1} = y_k - \mathbb{E}\left[y_k|y_{[k-1]}, \mathcal{M}, \theta\right]$, i.e., the *residual* between model prediction and observation [67]. The computed residuals $\varepsilon_{k|k-1}$ are used to evaluate the likelihood function $L\left(\theta; y_{[N]}\right)$, under an assumption of normality which is later verified by statistical testing [67, 68]. Papers B through E in this work [22, 26–28] uses a *specialised implementation* of this methodology, adopted from the CTSM framework.

Note that in the KF and system identification literature [16, 40, 43, 67] $\varepsilon_{k|k-1}$ is denominated as the *innovations* process, i.e., the part of the observed signal that cannot be estimated. Some authors e.g. [46] refer to the error between measurement and *aposteriori* estimate $\varepsilon_{k|k} = y_k - \mathbb{E}\left[y_k | y_{[k]}, \mathcal{M}, \theta\right]$ as the *residual,* a quantity not typically computed in the KF since it is not needed to compute the Kalman gain K, to distinguish from the *innovation* $\varepsilon_{k|k-1}$. In this work and in the papers B to E [22, 26–28] the term *residual* is used to describe $\varepsilon_{k|k-1}$. The exact notation and mathematical definition of the involved terms should prevent any confusion as to what is being computed.

Another paper using the CTSM framework presents three increasingly complex models of building integrated photovoltaic modules, derived from the system physics but with lumped parameters that are calibrated from data [69]. Two of the presented models are *non-linear*, by inclusion of *wind* and *long wave radiation* effects on the system [69]. In order to handle non-linear models, the current version of CTSM [16, 62] utilises the *Extended Kalman Filter (EKF)* [70, 71]. The standard KF applies well known matrix equations to propagate the covariances of the state estimates that are only applicable to linear models [43, 72]. The working principle of the EKF, also part of the well established KF literature, is the use of linearisation by computing the Jacobian to *approximate* the first moments of the multi-dimensional *Fokker-Planck* equation by a Taylor series expansion truncated after the first term [43, 72–74]. Hence, the *linear* equations for propagation of *covariance* can be applied also in the EKF [43]. Two more recent adaptions of the KF to non-linear system is the *Unscented Kalman Filter (UKF)* [75, 76] and the *Ensemble Kalman Filter (EnKF)* [74]. Both of these adaptations are based on the idea that it may be more beneficial to approximate the state *distribution*, rather than the *non-linear transformation* constituted by the state transition equation for a non-linear model [74, 76]. In addition to improved performance for highly non-linear systems, both the UKF and EnKF offer the practical advantage of handling *non-differentiable* models, i.e., models for which obtaining the *linearised* system matrices is infeasible. When estimating parameters for an *externally simulated* grey-box model, with no *explicitly* described model equations for the state propagation, the ability to handle non-differentiable models in the KF implementation is of crucial importance. This possibility is especially interesting for point 1 of the scope in this work, and was explored in Paper B [22] where both the UKF and the EnKF are used to estimate the residuals of an externally simulated TN model.

Another example of the CTSM framework applied to buildings is found in [13], where *fifteen* different models, constituting variations of a complex TN structure, are calibrated and compared using the likelihood ratio test [77]. Starting from the simplest model, the complexity is iteratively increased until no further *significant* improvement in log likelihood is obtained, thus identifying the minimum complexity model with optimal performance [13]. Another example using CTSM is [15] where a set of 5 TN models of different levels of complexity is calibrated against data

obtained by *simulation* of a white-box model, using the IDEAS library from KU Leuven together with some observational data on solar irradiation. One point of interest in the paper is the discussion of noise influence on the parameters, which can be studied due to use of virtual measurements on the simulated white-box model [15]. The paper concludes that the addition of *unbiased* measurement noise does not significantly affect the optimal parameter estimate, but it does increase the uncertainty of the parameter estimate [15]. This conclusion is similar to that of Paper E [28], where the estimated parameters from a data-set recorded from a *physical* building is compared with the estimated parameters from the *same* data-set with *added* zero-mean Gaussian noise. Further, according to [15], bias introduced by inaccurate sensor placement *will* affect the estimated parameter values, since the sensor location in the physical or virtual building directly *determines* the physical *interpretation* of the temperature nodes of the TN model [15]. This is again similar to the conclusion in Paper C [26], which explored the same challenge but with observed data recorded *simultaneously* from two sensors in a physical building. One more, recent, paper utilising the CTSM framework, and also comparing the results to the MATLAB System Identification toolbox, is found in [78] which focus on the control use-case with a first order model developed for a super-insulated residential house.

2.3.2 Other grey-box approaches

In addition to the SDE formulation reviewed in Section 2.3.1, there are also other parameter estimation methods that have been applied towards building thermal behaviour models that are classified by the authors as grey-box. Many of these approaches use the deterministically computed *simulation error* from a ballistic simulation compared with the measurements, i.e., the output error method (OEM) [40, 67]. However, model implementation, the choice of *optimisation method* and the formulation of the *objective function* typically differs. One approach, used in Paper A [23], is the straight forward numerical minimisation of the mean square error (MSE) between the simulated and measured building temperatures. An example of the use of numerical optimisation is [14], where the *Interior Point* optimisation method [79, 80] is used to calibrate four different TN models. The objective function is defined as the *product* of the MSE for heating power and indoor temperature [14]. Models are validated by forecasting on independent data [14]. An interesting point in the paper is the Sobol sensitivity analysis [81] which is used to diagnose parameters that does not sufficiently influence the objective functions when perturbed [14]. Another example of grey-box modelling, where the parameters of an otherwise white-box model developed from application of mass- and energy-balances to the main components of an HVAC system is given in [37]. The paper presents detailed state-space thermal models, which are implemented in Simulink in MATLAB using the Control and Estimation Tools Manager, and combined into a complete model for

the HVAC system. As typical for the white-box paradigm, the resulting models is complex with a large number of parameters. Although the model structures are developed as typical white-box models based on balance laws, the parameters are *estimated* from measured data by use of non-linear least squares optimisation of the simulation error, which results in the models being classified as *grey-box* according to the authors definition [37]. The use of the term grey-box is further supported by the addition of a *constant error term* to account for unmodelled effects in most of the presented state transition models [37]. The authors report without further discussion that different initial conditions for the optimisation problem resulted in different locally optimal solutions [37], a phenomenon which is often observed when the parameters are ambiguous due to *identifiability* issues, as discussed in Paper A [23] and D [27].

An early review of different approaches to modelling building thermal behaviour, written over 30 years ago, is provided in [18], which compares the use of ARMAX and TN models. The paper includes an interesting discussion of TN models, which points out the difficulties in choosing between the typically large number of possible TN structures for any given building. A solution for this model selection challenge has since been proposed by [13, 17, 69], using the likelihood ratio test based on Wilks theorem [77]. Further, the author of [18] points out that TN structures can easily become *over-parameterised* which leads to *ambiguous* parameter estimates, a difficulty which was observed for the R3C2 model used in all five papers of this thesis as well [22, 23, 26–28]. Potential remedies for parameter non-identifiability due to over-parameterisation are discussed in Papers A, C, D and E [23, 26–28], as it relates to point 3 of the scope defined in Section 1.2. Finally, [18] points out that the process of developing the differential equations from TN circuit models is somewhat labour intensive. This challenge motivated the definition of point 1 in the scope of this work, and the use of an *external circuit simulator* presented in Paper B [22].

It is interesting to note that the TN model equations in [18] are given without a *diffusion* term, i.e., as an ordinary differential equation (ODE) and not the SDE formulation discussed in Section 2.3.1. Further, the author argues in *favour* of using the ballistic simulation error [18]. This explains the lack of an error term in the state transition, since the ballistic simulation error formulation is *deterministic*. Paper D [27] and E [28] discuss the conditions under which the deterministic simulation error and the grey-box SDE formulation is the same, specifically that the process noise $\mathscr{W} \equiv 0$. The choice of error formulation is also a topic in many other papers and books on grey-box models and on parameter estimation in general, see e.g. [67, 82]. Another paper, which also discusses both ARMAX models and TN models , concludes with preference on the *prediction error*, due to its favourable availability of diagnostic and validation tools [83]. In this work, the simulation error was used in Paper A [23], while the SDE formulation was used in the remaining papers [22, 26–28].

One further commonality between [18] and [83] is the application of ARMAX models to building thermal behaviour. In [83] the MATLAB IDENT toolbox is used to obtain the ARMAX model parameters. The paper further discusses how the resulting ARMAX model parameters can be used to obtain parameters of a grey-box TN model that allows *physical interpretation* of the model parameters [83]. The *interpretation* of ARMAX model parameters in relation to the physical properties of the building is also a central topic in [18]. A final point of interest covered by both [18] and [83] is the choice of *dependent* variable, i.e., the model output to be fitted to data [18, 83]. Specifically, if the indoor temperature is controlled and therefore close to constant, it may be beneficial to fit the model to the *heating energy demand* Q, rather than the *indoor temperature* of the building T_b, since the former will have more dynamic variations and therefore produce better estimates of the buildings physical properties [18, 83].

Another approach to parameter estimation, commonly applied in the field of cybernetics, is the use of *dual estimation* in an Extended Kalman Filter (EKF), i.e., the estimation of both state and parameters [43]. By augmenting the state vector with the unknown parameters, the resulting model usually becomes non-linear and therefore requires a non-linear KF implementation [43]. An example of this approach applied to building thermal models is given in [84], where the simple 1R1C model is used in a *parameter augmented model* together with an EKF to *simultaneously* estimate both the temperature state and the parameters [84]. Further, the heat gain disturbance introduced by the building occupants is estimated by the EKF [84]. Intuitively, one would expect humans to *release* body heat, and therefore *contribute* to the heating of a building. However, the authors observe that the estimated heat gain from occupants is in fact *negative*, which they argue is explainable by assuming the dominating contribution of the occupants is through the *loss* of heat by convection through opening windows and doors [84]. The paper also applies calibration of model parameters by using the Nelder-Mead simplex algorithm [85] on the MSE simulation error for four TN structures. In conclusion, the paper reports that through the online estimation of both parameters and heat gain disturbance by occupants satisfactory performance w.r.t. prediction accuracy and computation time is achieved, with the outlook of utilising the approach for MPC [84].

The recent increase in popularity of machine learning (ML) methods has led to widespread application in a number of fields, including modelling of building thermal behaviour. While ML methods typically are considered black-box, some hybrid models combining ML methods and physical information is classified as grey-box models in the wider sense. One example is the application of the *genetic algorithms (GA),* a type of optimisation or search method inspired by the natural evolution of genomes [86, 87], applied to parameter estimation. In [38], a Building Energy Simulation (BES) model derived from physical specifications of the test building, implemented in a specialised simulation tool called EnergyPlus [88] is calibrated by with a GA. A sensitivity analysis [89] is performed to determine the most important parameters

of the BES model. The parameters are subsequently *discretised* prior to application of GA and fitted to measurement data [38] using the *non-dominated sorting genetic algorithm (NSGA-II)* algorithm [90]. Even after excluding non-sensitive parameters, the authors report the resulting discretised parameter space requires an infeasible 2.08×10^{11} simulations for an exhaustive search [38]. Hence, the use of GA optimisation is attractive given the large search space [38]. Another application of GA optimisation of parameters is found in [91], where the parameters of simplified TN models for different building segments is obtained. Rather than fitting the temperature directly to observations, the objective function is defined from the frequency and phase response of the TN model, with respect to theoretically computed values obtained from building specifications [91]. However, the paper also uses a simplified TN model for the internal thermal mass which is fitted to measured cooling load [91].

The use of an ANN to predict a variable relevant for building thermal behaviour, the input solar gains, is found in [92]. A *non-linear ARX (NARX)* model is constructed by utilizing an ANN to *regressively* predict future solar irradiation. The proposed model uses a set of physics based white-box models to compute predictions of solar gains, which is subsequently used as inputs for the NARX-ANN. The result is a model that uses both *physical information* about the nature of solar irradiation and established models, while also allowing some case specific learning to be applied by calibration in the ANN [92]. The approach can therefore be considered another example of the grey-box methodology.

2.3.3 Parameter identifiability for grey-box models

White- and grey-box models are based on prior information of the physical system $\mathscr{S}$. Hence, the parameters tends to represent some *real* physical quantities of $\mathscr{S}$, e.g., the thermal resistance of walls, doors or windows. It is often assumed that these parameters have some *unambiguous, constant,* but *unknown* true value $\boldsymbol{\theta}^*$, such that $\mathscr{M}(\boldsymbol{\theta}^*) = \mathscr{S}$. However, when calibrating the model $\mathscr{M}(\boldsymbol{\theta})$, the estimate $\hat{\boldsymbol{\theta}}$ may be *ambiguous*, even if, intuitively, the parameters of the physical system is well defined constants. The objective function $g(\boldsymbol{\theta})$ calculates the *model fit* based on errors between model predictions and measured data. Hence, if perturbations of some sub-set $\boldsymbol{\theta}_s \in \boldsymbol{\theta}$ is *not* reflected by the objective function, the resulting estimate $\hat{\boldsymbol{\theta}}$ will be ambiguous. If the cause of the ambiguous parameter estimates is the model *structure*, i.e., the model output is *not* affected by some subset of parameters $\boldsymbol{\theta}_s$, those parameters are denominated *structurally non-identifiable* [93]. Subsequently, the parameters in $\boldsymbol{\theta}_s$ does not affect $g(\boldsymbol{\theta})$ which results in an equipotential *manifold* on the objective functions *hyper-surface* in the parameter space Θ. Similarly, parameters can be *inter-dependent*, such that the effect of perturbations in one parameter on $g(\boldsymbol{\theta})$ can be *cancelled* by corresponding perturbations of another

parameters, which also results in similar equipotential manifolds. Another potential cause of non-identifiable parameters is *insufficient dynamic information content* in the calibration data. Since $g(\theta)$ compares model predictions to data, a lack of *variation* in the data may also cause perturbations of certain parameters to produce only *statistically insignificant* changes in $g(\theta)$. Such parameters are denominated *practically* non-identifiable [21].

The use of grey-box models is motivated by the possibility of the *interpretation* of the estimated parameters as derived from the physical properties of the system. However, as pointed out by [6], although the majority of papers in the field treat parameters as representative of the thermal properties of the building, a careful analysis of the parameter *identifiability* is usually not performed. The limited number of publications in this field which treats of parameter identifiability was also pointed out in the thorough review of inverse methods presented in [19]. A notable exception is [6], where both *practical* and *structural identifiability* of the parameters for a TN model is analysed using the Profile Likelihood (PL) method [21, 24, 25]. When using likelihood maximisation to estimate parameters the PL method is a natural choice of identifiability analysis method, as was noted by the authors behind the CTSM framework [94]. Given the limited number of papers that discuss identifiability of TN models [6], the application of the PL method is one of the central topics of this work, as illustrated by point 2a in the scope definition of Section 1.2. In addition to the application of the 1D PL method to TN models, the extension to creating 2D projected profiles, which allows the analysis of parameter interactions in addition to testing parameter identifiability, is discussed in Papers B to E [22, 26–28]. It should be noted that a similar extension was concurrently developed in [19].

The challenge of estimated parameter ambiguity is mostly relevant to parameters for which the assumption of a true physical value is reasonable. In the black-box modelling paradigm there is usually no assumption of an unambiguously optimal parameter vector that has some specific physical interpretation [40, 68]. Despite having multiple solutions, and being subject to local minima in the calibration, black-box models are well suited for *predicting* system behaviour [40, 95]. Similarly, the ability of a TN model to accurately predict system behaviour *despite* having non-identifiable parameters was shown in Paper E [28]. Hence, the challenge of parameter identifiability is arguably restricted to the *soft-sensor* use-case where the parameter values themselves are of interest.

2.3.4 Bayesian parameter estimation for building models

There are—unbeknown to most non-statisticians—two fundamentally different ways of looking at statistical estimation of parameters, namely *frequentist* and *Bayesian*, which has led to a 250-year long argument about which perspective is better [96]. Although Bayesian statistics is older than the frequentist [96], the latter has been,

and arguably still is, predominant within engineering disciplines. However, Bayesian statistics has gained significantly in popularity over the last, say, 20 to 30 years, to some extent driven by the improvement in computing power and the development of the Markov Chain Monte Carlo (MCMC) algorithms [97, 98]. There are many, both explicit and subtle, differences between frequentists and Bayesians, a detailed review of which is well beyond the scope of this work. However, one interesting and relevant distinction is in the way probabilistic statements of accuracy can be applied to estimated parameters. Papers A to D [22, 23, 26, 27] are all based on a *frequentist* interpretation of parameter estimation, while Paper E [28] discusses and compares *both* approaches.

As discussed in Section 2.3.3, parameters are often estimated under the assumption that there exist some true parameter $\boldsymbol{\theta}^* \to \mathscr{M}(\boldsymbol{\theta}^*) = \mathscr{S}$. The uncertainty of the resulting estimate $\hat{\boldsymbol{\theta}}$ is described by a *confidence interval* (CI), with prescribed confidence $\boldsymbol{\alpha}$, that relates to the true parameter $\boldsymbol{\theta}^*$. This approach is typical of *frequentist* statistics, which based on the definition of likelihood $L\left(\boldsymbol{\theta}; y_{[N]}\right) = p\left(y_{[N]}|\boldsymbol{\theta}\right)$ treats the *data* as a random variable and the *parameter* is assumed to be a fixed but unknown constant. Interestingly, the membership of $\boldsymbol{\theta}^*$ in a *specific* CI is *not* a question of probability as clearly stated in [99], despite common claims to the contrary. Since the parameters are considered to be unknown *constants*, and the limits of a *particular* CI is clearly known constants as well, the only probabilistic values that *can* be assigned to the question of membership in a CI is the trivial values 0 and 1 [99]. The confidence level $\boldsymbol{\alpha}$ is rather defined as the a priori *expected* probability of capturing $\boldsymbol{\theta}^*$ in a CI, i.e., the expectation $\mathbb{E}\left[p\left(\underline{\boldsymbol{\theta}} \leq \boldsymbol{\theta}^* \leq \overline{\boldsymbol{\theta}}\right)\right] = \boldsymbol{\alpha}$ [100]. The *average* probability of capturing $\underline{\boldsymbol{\theta}} \leq \boldsymbol{\theta}^* \leq \overline{\boldsymbol{\theta}}$ is $\boldsymbol{\alpha}$ in the sense that the *frequency* of correct CI's approaches $\boldsymbol{\alpha}$ in the limit of infinite trials [100]. It is interesting to note that the bootstrapping concept of using *randomisation* to create a large number of pseudo data-set [29, 31], with subsequent parameter estimation, is closely related to the fundamental idea of repeated experiments on which CIs are based. This topic was discussed in Paper D [27] as it relates to the point 2b in the scope defined in Section 1.2.

The Bayesian framework, however, considers the *posterior* distribution $p\left(\boldsymbol{\theta}|y_{[N]}\right) \propto p\left(y_{[N]}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$, which unlike the likelihood function $p\left(y_{[N]}|\boldsymbol{\theta}\right)$ is a probability distribution on the parameter $\boldsymbol{\theta}$ that takes its values over the *parameter* space [24, 101]. Since the parameter is now considered a *random variable*, it is justified to give probabilistic statements about the estimate $\hat{\boldsymbol{\theta}}$, including use of a *prior probability distribution* $p\left(\boldsymbol{\theta}\right)$ to describe any a priori *physical knowledge* or *subjective belief* that may inform the estimation of $\boldsymbol{\theta}$ in addition to the likelihood function. Obtaining the posterior distribution analytically is in many cases intractable, but requires the use of estimation algorithms such as MCMC methods [101, 102], e.g. the Metropolis-Hastings algorithm [97, 98]. It is also interesting to note that there is a close relationship between the resulting marginal distributions from MCMC, and the projected posterior distributions obtained by the Posterior Profile (PP) method [24] which is

discussed in Paper E [28]. For further discussion on the relationship and difference between frequentist and Bayesian statistics see e.g. [96, 103].

Given the inherent uncertainties in grey-box models and their parameters, the Bayesian framework for parameter estimation is of particular interest. By *inferring* parameter *posterior* distributions it is possible to gain better insight into the model behaviour, by better accounting for the *uncertainty* in the parameter estimates in simulations, as discussed in Paper E [28]. The usefulness of Bayesian parameter estimation for grey-box models of building thermal behaviour is further evident by the recent increase in publications on this topic. In [104] three different MCMC methods are compared for the purpose of calibrating parameters of a white-box model implemented in EnergyPlus [88]. After an initial sensitivity analysis, the posterior distribution of a sub-set of parameters is inferred, together with the parameters of three Gaussian Processes (GP) used to describe the model uncertainty and measurement noise [104], following the approach in [105]. It is interesting to note that the calibration of models implemented as *software code* given in [105] shares some similarities with point 1 in the scope of this work, as discussed in Section 1.2. Specifically, calibrating parameters for a model that only exists as a black-box piece of code or software, is the central topic of [105]. In Paper B [22], the same challenge of calibrating *software* is solved by adopting the methodology of CTSM [16, 62, 63], but using an UKF to estimate the residuals $\varepsilon_{k|k-1}$.

An example of using Bayesian parameter calibration for TN models is given in [20], where three different model structures where calibrated for seven different data-sets, using both the stochastic estimation framework [20] as reviewed in Section 2.3.1 and the deterministic simulation error. The results show that by MCMC calibration of the parameters of a TN model formulated as an SDE, the resulting posterior samples $\theta_{[K]}$ can be used to reliably compute overall heat loss coefficient and capacitance for the building, including a reasonable uncertainty estimate for the computed thermal properties [20]. The estimated overall heat loss coefficient (HLC) and total thermal capacitance are compared over seven different data-sets. Further, the SDE based stochastic model calibration is compared to the ballistic simulation error approach, which shows that the consistency of the results between the seven data-sets is significantly more reasonable using the SDE approach. This approach is similar to Paper E [28], which also focuses on the consistency of estimated parameters across different data and noise levels, as outlined in scope element 2b. While both [20] and Paper E [28] focuses on the use of *different* data-sets, and alternative approach to testing consistency based on non-parametric block-based bootstrapping is given in Paper D [27]. One further similarity between [20] and Paper E [28] is the use of stochastic forecasting to estimate the effect of uncertain parameters on the model predictions, a topic which is central to scope element 4 in Section 1.2.

Another similar example of Bayesian inference with MCMC on TN models is found in [7], which studies the thermal properties of two different wall constructions, rather

than the complete building envelope. The parameter posterior distributions are reported as *corner plots*, a typical combination of 1D and 2D marginal posterior distributions, rather than as point-wise MAP estimates with credibility intervals given in [20]. The method of [7] was developed in previous papers, see e.g. [8] and [9] where the proposed framework is explained in detail. Of particular interest is the use of the *odds ratio* test to perform Bayesian model selection [8, 106]. The methodology of [7–9] is similar to [20] and to that used in Paper E [28]. However, Paper E [28] includes the novelty of of comparing the MCMC results with the typically frequentist profiling methods PL and PP. Another topic of Paper E [28] that is not covered by [7–9, 20] is the study of uncertainty in the estimated *measurement* and *process noise* covariance parameters

3 Experimental setup

3.1 Building and sensors

All data used in the papers A to E [22, 23, 26–28] are collected from an experimental building located on the Porsgrunn campus the University of South-Eastern Norway (USN). The building was constructed in 2014 and has been the subject of several previous scientific studies [107–109]. The building is a small single zone structure with a total volume of approximately $9.4\,m^3$. Further, the building is situated on concrete support structures such that the floor is not in contact with the ground [107–109]. The walls are constructed using different types of insulation materials such as wood, cement chipboard, particle board, cardboard, cladding, glass wool insulation, and polyethylene vapour barriers [108], in three sections indicated as *Solid*, *Standard* and *Sustainable*, in Fig. 3.1. The roof is covered by asphalt sheets with an inner construction of wood. Given the diverse use of materials and construction techniques, the heat transfer characteristics vary significantly between different sections of the building [108]. The building is sealed and does not have any type of ventilation [107–109]. The building has three small windows, each of size $60 \times 90\,cm^2$ located on the south, east and west wall, and a door of size $90 \times 200\,cm^2$. The building is situated between several large office buildings. Given the significant shading, the small window size and the typical limited solar irradiance on site during Norwegian winters, the solar contribution to heating has typically been ignored in the previous studies of this building [108, 109], as it is in the papers A to E [22, 23, 26–28]. Similarly, the cooling effect of wind is also assumed to be negligible for this particular structure and location. The building has a simple active heating system consisting of an electric resistive heating element with rated power of $370\,W$, regulated by a bi-metal thermostat. An additional $\sim 100\,W$ of thermal energy is continuously supplied from the DAQ system and a connected logging computer.

The building is outfitted with a large amount of sensors connected to two Data Acquisition (DAQ) module (NI USB-6218) that are continuously logged by a connected PC to provide detailed thermal measurements throughout the building structure [108, 109]. The sensors include TMP36 semi-conductor temperature sensors, PT1000 resistive thermic elements, Light Dependant Resistors (LDR) and Honeywell 4000 humidity sensors, located on the interior surfaces of the building, under the outside wooden cladding exposed to outdoor temperature and inside the building envelope, e.g., in the insulation layers [108, 109]. The supplied electric power to
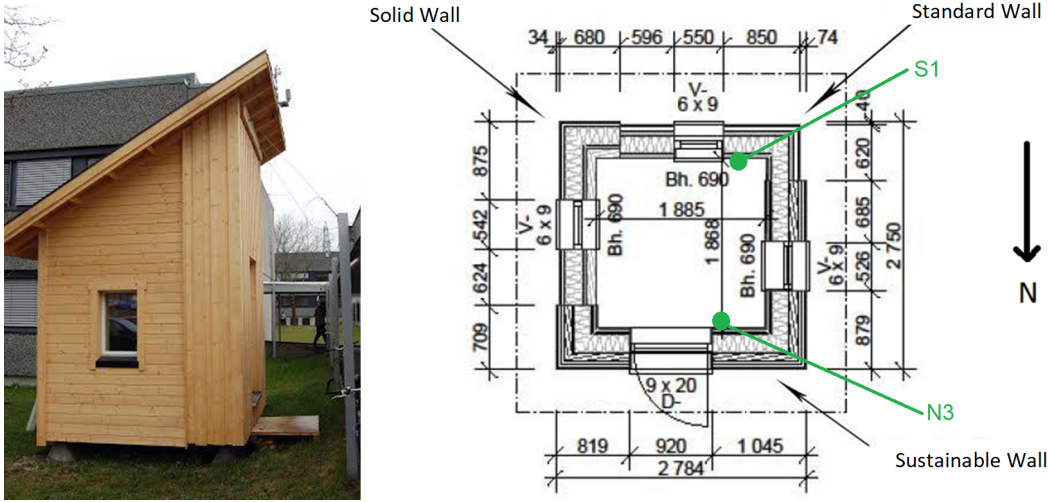
Figure 3.1: Experimental building and floor plan.

Table 3.1: Overview of data-sets

Data-set	Sensor systems in use	Used for papers
Nov1 - 2015	NI-DAQ sensors + Weather station	A
Nov2 - 2015	NI-DAQ sensors + Weather station	A, B
Dec1 - 2015	NI-DAQ sensors + Weather station	A
Dec2 - 2015	NI-DAQ sensors + Weather station	A
Feb1 - 2018	NI-DAQ sensors	C, E
Feb2 - 2018	NI-DAQ sensors	D, E
Feb3 - 2018	NI-DAQ sensors	E

the building is measured by a Honeywell CSLA1CD Hall effect sensor. Additionally, a commercially available domestic use weather station, which measures both indoor and outdoor temperature, rainfall and wind-speed, is installed [108, 109]. The weather station has a logging frequency of **30 min** while the DAQ system operates on a **5 min** interval for thermal, humidity and light measurements and a **10 s** interval for power measurements [108, 109]. The two systems are not synchronised, such that some data pre-processing is required for temporal alignment of data [22, 23]. And overview of the collected data-sets, and the papers they are used for, is given in Table 3.1.

Given that the primary focus of this work is the *soft-sensor* use-case of parameter estimation, the simplicity of the structure, the large amount of sensor data available, and the possibility to conduct experiments without regards for occupant needs, makes the building an ideal experimental platform for conducting *benchmark* tests on parameter estimation and analysis methods. Since the building is exposed to on site weather conditions, the quality of the collected data are realistic w.r.t. what one may obtain from in situ measurements on a residential building in Norway. Note that
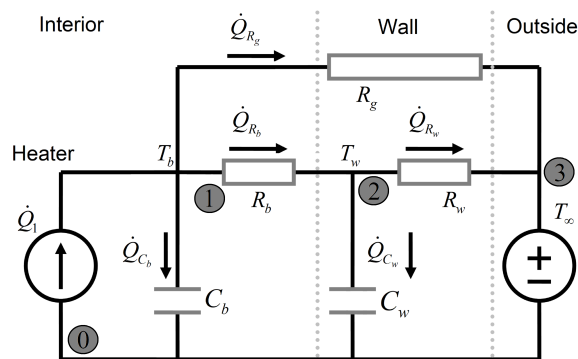
Figure 3.2: Simplified thermal network model with three resistors and two capacitors.

for most of the data-sets, the indoor temperature is driven by a step in the supplied heat energy. Such step responses are not compatible with human occupancy, but are usefull for acquisition of sufficiently rich dynamic information content in the recorded data. It is a reasonable assumption that similar step-response tests can be conducted while a residential home is unoccupied.

3.2 Grey-box thermal network model

The focus of all five papers A to E [22, 23, 26–28] are the calibration of parameters of thermal network (TN) models. A possible choice of TN model of the experimental building, partially based on the R4C2 model presented in [14], is shown in Fig. 3.2. The simplified TN model, illustrated as an Resistor-Capacitor circuit, consists of five components: the thermal resistance between room air and wall R_b, the building envelope R_w, and the thermal resistance of windows and doors R_g. The two capacitances C_b and C_w represent the thermal capacitance of the building interior and envelope, respectively. The model has two outputs: the room temperature T_b and the wall surface temperature T_w, and two inputs: the supplied electric power $\dot{Q}$ and the outside temperature T_∞. A variation of this R3C2 model, is the R2C2 model used in papers [26–28] where the thermal resistance R_g is removed, equivalent to setting $R_g \equiv \infty$ in the R3C2 model.

An important note on the use of simplified TN models is that the choice of sensor location in the physical building directly determines the physical interpretation of the lumped parameters. Given that the experimental building is constructed with three different sets of material choice, the thermal characteristics of each segment differs. Hence, although the same model structure is used in all papers, the physical interpretation of the individual parameters may be slightly different. This point is in particular highlighted in Paper C [26].

4 Methods

4.1 Estimating parameters of externally simulated models (Scope 1)

Most dynamic system models contains some form of uncertainty w.r.t. the real physical system and can therefore with advantage be formulated using a stochastic differential equations (SDE) grey-box framework [16, 17]. The following formulation, adopted from [16], is used as the starting point for parameter estimation in papers B to E [22, 26–28]:

$$dx_t = f(x_t, u_t, t, \boldsymbol{\theta}) dt + \sigma(u_t, t, \boldsymbol{\theta}) d\omega_t \qquad (4.1)$$

$$y_k = h(x_k, u_k, t_k, \boldsymbol{\theta}) + v_k \qquad (4.2)$$

The state transition model in Eq. (4.1) consist of the *drift* term f, which describes the *deterministically* known apriori physical knowledge of the system [17], parametrised by $\boldsymbol{\theta}$, and the *diffusion* term which describes the uncertainty of the state transition as a stochastic process [16, 17, 72]. The measurement Eq. (4.2) consist of the function h, which describes a *deterministically* known mapping on the model states and inputs parametrised by $\boldsymbol{\theta}$, and the measurement noise $v_k \sim \mathcal{N}(0, V)$. In the context of thermal network (TN) models of buildings, the deterministic drift term is a set of ODE's derived from the RC circuit analogue, typically by use of *Kirchoff's node balance laws* [110]. As pointed out in [18], the development of equations is somewhat labour intensive, in particular for complex RC circuit models. Hence, point 1 of the scope presented in Section 1.2 is to obtain a method that allows estimation of the parameters of a TN model *without* deriving the equations.

A possible solution is to replace the deterministic drift term of Eq. (4.1) and the measurement equation in Eq. (4.2) with an *external simulator* that can propagate the model state without deriving equations. The simulator can be implemented as a memory less system that takes as input the previous state estimate $\hat{x}_{k-1|k-1}$ and system input u_k and estimates the next state $\hat{x}_{k|k-1}$ and output $\hat{y}_{k|k-1}$. Specifically:

$$\hat{x}_{k|k-1} = f_{\text{sim}}\left(\hat{x}_{k-1|k-1}, u_k, t_k, \boldsymbol{\theta}\right) \qquad (4.3)$$

$$\hat{y}_{k|k-1} = h_{\text{sim}}\left(\hat{x}_{k|k-1}, u_k, t_k, \boldsymbol{\theta}\right)$$

Further, by including the *diffusion* term to absorb model errors, as discussed in Section 2.3.1, the parameters of the external simulator can be estimated in a grey-

box framework adopted from CTSM. While the method is only considered for TN models, the estimation of parameters for more general simulated models [105], e.g. using EnergyPlus [88], discussed in Section 2.3, suggests that the developed approach may be of more general interest.

4.1.1 Component list representation of TN models

The proposed solution for TN models, presented in Paper B [22], is inspired by the tools available for circuit analysis in the electronics field, e.g. SPICE [111], and based on describing the TN models as a *list of interconnected components*. The R3C2 model shown in Fig. 3.2 can be described as:

Table 4.1: Component list description of the R3C2 model.

Component	Type	Input	Output
$\dot{Q}_1$	Power Source	0	1
T_∞	Temperature Potential Source	0	3
C_b	Capacitance	1	0
C_w	Capacitance	2	0
R_b	Resistance	1	2
R_w	Resistance	2	3
R_g	Resistance	1	3
T_b	Output Potential	0	1
T_w	Output Potential	0	2

Each component in the R3C2 model shown in Table 4.1 is described by a component name, identifying the *parameter* to be estimated or the *measured signal* to be associated with a model input or output. The last two components in Table 4.1 are used to instruct the simulator which of the nodes should be reported back to the estimation algorithm as estimated measurements. Each component has two terminals, *Input* and *Output*, which are connected to a *numbered* node, as shown by the circle inscribed node numbers of Fig. 3.2. By describing a TN model on the form of Table 4.1 experimentation with model structure can be performed simply by *adding* or *removing* lines from the table, thus eliminating the need for rewriting model equations. Although beyond the scope of the present work, the suggested method may potentially be used to *automatically* construct the TN model structures by e.g. *combinatoric* search methods [112] that can manipulate the component list with subsequent parameter estimation, model comparison and validation.

Paper B [22] presents a simple simulator capable of propagating the state of a model on this form, by summing the energy flows into and out of each node [110]. The capacitors, which are the only dynamic elements and represents the model *states*, are discretised using a *backward Euler* scheme $\left(\frac{dx}{dt}\right)_{t_k} \approx \frac{x_k - x_{k-1}}{\Delta t}$. Voltage sources are

implemented as *constraints* on the difference between the states of the two connected nodes. The system of discrete time node equations for the state x_k can be written on the form:

$$Ax_k + A_m x_{k-1} + Bu_k = 0 \qquad (4.4)$$

The contributions from all components are summed, such that the balance equation for node i constitutes row i in A, A_m and B . The state of the TN model can then be propagated by solving Eq. (4.4) at each time-step. The simulator is validated in Paper B [22] by comparing the resulting state trajectories and estimation residuals with the linear ODE's of the R3C2 model.

Given the general task of estimating parameters for a model that exists only in computer code [105], the suggested approach could be of more general use than for TN models. The RC electric circuit analogue models could be extended with an inductor element L, thus allowing representation of *general* linear model by use of an appropriately designed RLC circuit analogue. Combining the potential of creating the RLC circuit analogue model structure *automatically* with the grey-box parameter estimation scheme may be a promising machine learning approach for the creation of grey-box dynamic system models.

4.1.2 Evaluating likelihood of parameters for external simulations as SDE's

Although the simulator model in Eq. (4.4) is *linear*, it exists in an *external software* which the parameter estimation algorithm can only access by supplying the current state and input, and retrieving the state and output at next time-step. The parameter estimation algorithm does not make any assumptions of linearity, or indeed assumption of any kind as to how the simulator propagates the state. Since the state transition is computed by simulation without any explicit equations the use of *linearisation* to obtain standard linear form models used in the KF and the EKF is not feasible. Hence the use of a KF implementation that can utilize *non-differentiable* models is required. In Paper B [22], both the UKF and the EnKF is reviewed for the purpose of computing one-step ahead prediction residuals $\varepsilon_{k|k-1}$ of an externally simulated TN model.

4.2 Analysing parameter uncertainty and identifiability (Scope 2a)

Estimated grey-box model parameters are influenced by both *aleatoric* uncertainty, such as measurement noise, discretisation and sampling in the *data acquisition* (DAQ) system [113], and the *epistemic* uncertainty introduced by simplifications

and unmodelled effects in the physical system. These uncertainties are propagated to the parameter estimates. Hence, an important part of any statistically solid parameter estimation framework is to provide some measure of the *uncertainty* of the estimated parameters.

4.2.1 Confidence intervals and regions from the Hessian

Arguably, the most prolific description of estimation uncertainty in the literature is the use of *confidence intervals* (CI) [99]. The typical CI for a scalar point estimate is expressed as

$$\hat{\theta}_i \pm \sqrt{\Delta_\alpha \Sigma_{i,i}} \rightarrow \hat{\theta}_i \pm z\sigma_i \tag{4.5}$$

where σ_i is the standard deviation, with variance $\Sigma_{i,i} = \sigma_i^2$, of the i-th parameter estimate $\hat{\theta}_i$, and $z = \sqrt{\Delta_\alpha}$ is determined by the desired confidence level α. The interval in Eq. (4.5) is known as an *asymptotic* CI, which is computed under the assumption of a symmetric parameter distribution [21]. Extending to the multivariate case, *ellipsoid* confidence *regions* for multiple parameters can be described as a *subset* of the parameter space Θ for which the following holds:

$$\left\{ \theta : \left(\theta - \hat{\theta} \right)^T \Sigma_\theta^{-1} \left(\theta - \hat{\theta} \right) < \Delta_\alpha \right\} \subseteq \Theta \tag{4.6}$$

As in the scalar case, the asymptotic confidence region in Eq. (4.6) is based on a quadratic assumption, which in the multivariate case results in an ellipsoid confidence region centred on a presumed optimal estimate $\hat{\theta}$. The weighting matrix Σ_θ is the *covariance* of the estimated parameters, and can be obtained by estimating the *Hessian* of the log likelihood $\Sigma_\theta^{-1} = H_{\mathrm{L}} = \nabla^T \nabla \ln p \left(y_{[N]} | \theta \right) \big|_{\theta = \hat{\theta}}$ [16, 21].

Rather than relying on the quadratic assumption, a more general expression of a confidence region, or interval in the scalar single parameter case, can be obtained by placing a *threshold* on the *log likelihood* function $\ell_{\mathrm{L}}(\theta) = \ln p \left(y_{[N]} | \theta \right)$:

$$\left\{ \theta : \ell_{\mathrm{L}}(\theta) - \ell_{\mathrm{L}}\left(\hat{\theta} \right) < \Delta_\alpha \right\} \subseteq \Theta \tag{4.7}$$

The result is denominated as a *likelihood based* confidence region [21, 25]. The threshold Δ_α can be drawn from the $\chi^2_{\alpha, n_{\mathrm{df}}}$ distribution by *Wilks' Theorem [21, 77]*. Note that likelihood based CI's on the form of Eq. (4.7) can still be computed for individual parameters, similarly to asymptotic CI's, even if the parameter space is multidimensional, as discussed in Section 4.2.2. The advantage of Eq.(4.7) is that the resulting region or interval can be *asymmetric* and therefore *unbounded* in some direction. This asymmetry allows diagnosis of *structural* and *practical parameter identifiability,* according to the definitions of [21], by analysing the bounds of the likelihood based CI and checking for the presence of a well defined optimum despite obtaining an unbounded CI. However, obtaining the members of such a set is computationally demanding, requiring the evaluation of $\ell_{\mathrm{L}}(\theta)$ for a large number of

parameters $\boldsymbol{\theta}$. Hence, confidence regions on the form of Eq. (4.6) are useful as an *estimate* of uncertainty when identifiability of parameters and the assumption of a symmetric distribution is reasonable. The analysis and representation of uncertainty is discussed in further detail in Paper D [27].

4.2.2 Profile likelihood

Based on the definition of a likelihood based CI, the PL method computes *profiles* of the log likelihood by projecting the high dimensional hyper-surface in parameter space onto single parameters [21]. The projected *likelihood profile* $\ell_{\text{PL}}(\boldsymbol{\theta}_i)$ is obtained by a two-step optimisation procedure, where one parameter $\boldsymbol{\theta}_i$ is held at a *fixed* value while optimising the remaining parameters $\boldsymbol{\theta}_{j\neq i}$ [21]. The process is repeated for multiple values of $\boldsymbol{\theta}_i$. Let

$$\ell_{\text{PL}}(\boldsymbol{\theta}_i) = \min_{\boldsymbol{\theta}_{j\neq i}} \ell\left(\boldsymbol{\theta}_{j\neq i}; \mathcal{M}, \mathcal{K}, \boldsymbol{\theta}_i\right) \tag{4.8}$$

The resulting $\ell_{\text{PL}}(\boldsymbol{\theta}_i)$ can be plotted as a function of $\boldsymbol{\theta}_i$ by varying parameter $\boldsymbol{\theta}_i$ in some way, prior to optimising the remaining $\boldsymbol{\theta}_{j\neq i}$, and subsequently analysed according to the definitions of structural and practical identifiability [21]. A straightforward solution, if the log likelihood $\ell_{\text{L}}(\boldsymbol{\theta})$ is well behaved within the constraints of Θ, is to use a *brute force* approach with an even sampling of $\boldsymbol{\theta}_i$. Alternatively, a two-sided gradient decent algorithm, using a freely optimized parameter vector as a starting point, can be applied [21, 114]. This method is used in papers B to E [22, 26–28].

4.2.2.1 2D Profile Likelihood

The PL method is a powerful tool for analysing both identifiability and estimating the uncertainty of parameters [6, 21]. However, one potential drawback of the method is the projection of a high dimension hyper-surface onto single parameters [21]. If there are parameter interactions, e.g. inter-dependency, this information is lost in the projections. Hence, an improvement on the PL method is to project the hyper-surface onto a plane $\Theta_{i,j}$ of two parameters. These 2D projections can be plotted as topological heat-maps that highlight parameter interactions as well as identifiability and estimation accuracy. The projection $\ell_{\text{PL2D}}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)$ is obtained by

$$\ell_{\text{PL2D}}\left(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j\right) = \min_{\boldsymbol{\theta}_{k\neq i,j}} \ell\left(\boldsymbol{\theta}_{k\neq i,j}; \mathcal{M}, \mathcal{K}, \boldsymbol{\theta}_i, \boldsymbol{\theta}_j\right) \tag{4.9}$$

The method is explained in detail in Paper B [22] with a similar idea contemporarily suggested in [19]. Projecting to a parameter plane $\Theta_{i,j}$ is, in theory, a straight forward extension to the established profile likelihood method. However, due to the

increase in computational complexity, the implementation of this method is challenging. Further, when using a brute force approach, a high resolution discretisation of the plane $\Theta_{i,j}$ is needed to capture the most significant regions; the equipotential manifolds. The projections of the equipotential manifolds can in some cases, e.g., if some parameters are interdependent, present as *narrow* valleys in Θ that can be difficult to obtain if the brute force grid resolution of $\Theta_{i,j}$ is to coarse. Efficient implementation of the method is therefore crucial for its applicability. Towards that goal, a further refinement of the proposed method is presented in Paper C [26], which uses previously obtained projections from the discrete points in $\Theta_{i,j}$ as initial conditions for estimating the remaining parameter $\theta_{k \neq i,j}$. The idea is based on the assumption that the optimal values of $\theta_{k \neq i,j}$ for a point (θ_i, θ_j) in $\Theta_{i,j}$ is a reasonable initial guess for *other* points in the neighbourhood, e.g. $(\theta_i + \Delta_i, \theta_j + \Delta_j)$ when Δ_i, Δ_j are small. The new version of the algorithm was in [26] shown to reduce computation time by 4-10 times. The drawback of the modification is that the optimisation of parameters $\theta_{k \neq i,j}$ depend on the estimation of a neighbouring point. This dependence between neighbouring solutions requires a more careful implementation to ensure efficient parallelisation of the method on multiple CPU cores. The suggested algorithm in Paper C [26] solves this by using deterministic chains that explores $\Theta_{i,j}$ in an ordered scheme that allows efficient parallelisation.

4.2.3 Posterior Projections and MCMC

As discussed in Section 2.3.4, the *Bayesian* framework, focused on analysing the posterior distribution, is an alternative to analysing the likelihood, i.e. the typical *frequentist* approach. The posterior distribution $p\left(\theta | y_{[N]}\right)$ is obtained by *Bayes Theorem* [16] as

$$p\left(\theta | y_{[N]}\right) = \frac{p\left(y_{[N]} | \theta\right) p\left(\theta\right)}{p\left(y_{[N]}\right)} \tag{4.10}$$

which combines the likelihood with a priori information of the parameters through the *prior* distribution $p\left(\theta\right)$. Since the *evidence* scaling constant $p\left(y_{[N]}\right)$ is independent of θ, the posterior is proportional to the *numerator* of Eq. (4.10), i.e., $p\left(\theta | y_{[N]}\right) \propto p\left(y_{[N]} | \theta\right) p\left(\theta\right)$. Often priors are chosen uniform *unbounded*, i.e. $p\left(\theta\right) = 1$ for $\theta \in \mathbb{R}^{n_\theta}$, or as a uniform *bounded* window where $p\left(\theta\right) = 1$ for $\theta \in \Theta$ and $p\left(\theta\right) = 0$ for $\theta \notin \Theta$. In both cases, the posterior is also proportional to the *likelihood*, i.e., $p\left(\theta | y_{[N]]}\right) \propto p\left(y_{[N]} | \theta\right)$, over the *support* of the prior where $p\left(\theta\right) \neq 0$. In the frequentist framework, as a consequence of their omission, the uniform unbounded prior can be considered a *default* choice. This proportionality between posterior and likelihood allows some interchangeable use of methods and illustrates the similarities between the two perspectives on parameter estimation, as discussed in Section

2.3.4 and in Paper E [28]. There are also other fundamental *differences* between frequentist and Bayesian statistics, e.g. treatment of parameters as *unknown constants* or *random variables,* and the use of *subjective* priors [96, 102, 103, 115], a thorough discussion of which is beyond the scope of this text.

Since the posterior is a hyper-surface in high dimensional parameter space Θ, visualisation requires some form of processing into a 2D or 3D representation that can be plotted. One possibility is the use of *projection,* in the same way that the PL method projects the likelihood onto individual parameter axis. This approach was presented in [24]. The 1D projection of the posterior is obtained as

$$\ell_{\text{PP}}(\theta_i) = \min_{\theta_{j \neq i}} \ell_{\text{P}}\left(\theta_{j \neq i}; y_{[N]}, \theta_i\right) \tag{4.11}$$

This is equivalent to Eq. (4.8), except that the projection is performed on the log posterior $\ell_P(\theta) = -2\ln p\left(\theta | y_{[N]}\right)$. The *Posterior Projection* (PP) method can also be extended to project 2D topologies onto a plane $\Theta_{i,j}$, as discussed in detail in Paper E [28].

An alternative approach is the use of a class of methods named *Markov Chain Monte Carlo* (MCMC) [97, 98, 115, 116] that *approximate* the posterior by random *sampling* of the parameter space Θ. The MCMC algorithms are based on a chain of successive samples, where each new sample is drawn proportionally to the posterior $p\left(\theta | y_{[N]}\right)$ such that the posterior distribution and its parameters can be approximated from the empirical distribution of the drawn samples [97, 98, 115, 116]. This sampling is realized by drawing candidates θ_k^c for the next sample in the chain according to a proposal distribution $\theta_k^c \sim q\left(\theta_k | \theta_{k-1}\right)$, combined with a statistical test that either *accepts* or *rejects* the new candidate sample θ_k^c, based on a specific statistical criterion. In the Metropolis-Hastings algorithm the acceptance test is defined by a uniform probability $p_a \sim U(0,1) < \min(1,\alpha)$, where α is the *ratio* of the posterior probability of the candidate sample w.r.t. the current sample [97, 98]. The MCMC class of methods and the Metropolis-Hastings variation is the central topic of Paper E [28].

It is interesting to note that both the PP and the MCMC method seek to obtain *estimates* of the posterior distribution hyper-surface, as discussed in Paper E [28] which presents a comparison of these two methods. The most obvious difference is in the way the two methods explore the parameter space Θ. The MCMC uses *stochastic* exploration, focused on the regions of the highest posterior density assuming proper chain mixing and convergence. This approach provides computational efficiency at the expense of a somewhat noisy approximation of the posterior hyper-surface, particularly in the lower density regions. In contrast, the projection based PP method explores Θ deterministically and exhaustively within the resolution of the brute force discretisation. Hence, the projected profiles tends to be smooth and captures the global optima within the bounds of Θ, at the cost of a significantly increased computational load. The 1D PP method is reasonably fast to compute,
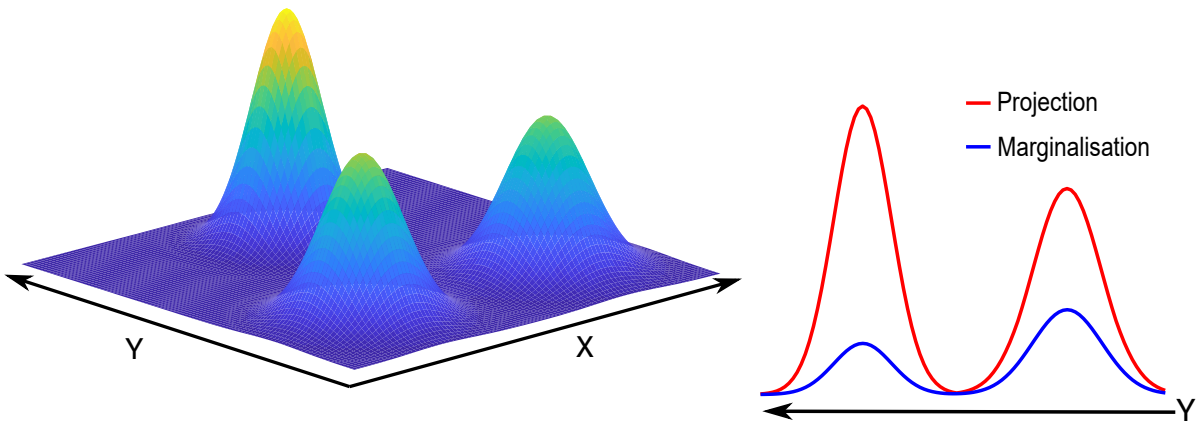
Figure 4.1: *The Dromedarius and the Camel*: illustration of the difference between marginalisation and projection of the optimum.

and requires only projections onto each parameter, i.e. n_θ projections. However, the 2D PP method is computationally demanding to the point of being impractical if the number of parameters is large. However, as discussed in Paper E [28], the *combination* of both MCMC and the projection methods can provide a more detailed analysis of the posterior hyper-surface, or indeed the likelihood hyper surface when uniform priors are used.

It should be noted that, due to the popularity of MCMC, many variations exist, e.g. Gibbs Sampling, Hamiltonian Monte Carlo , etc. [115, 116], beyond the Metropolis-Hastings which was used in Paper E [28]. The Metropolis-Hastings is however easy to implement, which makes it a convenient choice for the comparison with other methods, as is the goal of Paper E [28].

4.2.3.1 Projections vs marginalisation

Another, arguably more *subtle*, difference between the MCMC and the projection methods is the way in which the results are *visualised*. Given that the posterior hyper-surface is typically multi-dimensional, the results obtained by both methods must be *processed* in some way for visualisation in 2D or 3D. The PP method *projects* the surface onto a single parameter axis, or a plane of two parameters in the PP2D case. This projection is achieved by numerically computing the *optimum* along some direction in Θ, e.g., a line normal to the parameter axis θ_i that intersects the axis in a specific point (θ_i) or a *normal* to the plane $\Theta_{i,j}$ that intersects the plane in a specific point (θ_i, θ_j), as described by Eq. (4.9) and Eq. (4.8). In contrast, the MCMC method represents the posterior by its *marginal distributions*, i.e., the distribution over a *sub-set* of the parameters with the remaining parameters *integrated out* [115, 116]. This process of *integration* over the remaining parameters $\theta_{k \neq i,j}$ can be seen as computing the *volume* along the same *normal* for each point

(θ_i) on the axis of parameter θ_i or point (θ_i, θ_j) in the plane $\Theta_{i,j}$. The similarity and difference between these two approaches is illustrated in Fig. 4.1. The left panel shows an entirely artificial 2D distribution over parameters *X* and *Y*, aptly named "The Dromedarius and the Camel", with three *modes*, i.e., peaks, of different heights. The right panel of Fig. 4.1 shows the *projection* onto the *Y* parameter (red) and the *marginal* distribution over *Y* (blue). Observe that the *projections*, which takes the *optimum* for each value of *Y* with *X* as a free parameter, always returns the *highest* value, while the marginal distribution *accumulates* the *volume* over *X* for each value of *Y*. Hence, the optimal *point estimate* of *Y* is *different* for the two methods. The maximum *volume* over *X* is found for the two lower peaks, while the maximum *height* of *Y* is found for the single peak. It is also worth noting that a point estimate $\hat{\theta}_{MAP}$, typically used in a frequentist setting, would by numerical optimisation return the highest peak. The practice of computing point estimates is not typically used in the Bayesian paradigm, which tends to utilise the full marginal posterior instead. If one where to compute the optimum of the *marginal posterior* it would in some cases be *different* than than the MAP estimate.

It is also interesting to note from inspecting the projection and marginalisation of *only* the highest peak, that both of the 1D representations of the 2D distribution are in fact *proportional* [24]. Hence, the MCMC and the PP method respectively returns marginal distributions and projections that are in general two different representations of the posterior hype-surface, but for some *specific* shapes of parameter distributions, e.g., the Normal distribution, they are of *identical* shape but *shifted* in log space *[24, 28]*.

4.3 Parameter consistency (Scope 2b)

The consistency of the estimated parameters across different calibration data can also be considered a form of uncertainty measurement. As discussed in Section 2.3.4, the foundation of the frequentist confidence interval (CI) [99] is the *repetition* of experiments. By collecting a large number of experimental data-sets and estimating parameters from each set, one could in principle *measure* the accuracy of the resulting estimated parameters [100]. In practice, it is rarely the case that one has access to data from a large number of repeated experiments, hence the CI methodology is useful in order to *estimate* the uncertainty bounds from a single experiment. However, it is always useful to compare results from at least several different data-sets, a practice which is common also in the literature on TN models of building thermal behaviour [20, 23, 26–28].

4.3.1 Bootstrapping for time-series data

Repeating building thermal behavior experiments a large number of times is in most cases infeasible, due to the cost and time required for performing experiments. It is therefore of interest to consider alternative methods to test *consistency* of estimated parameters w.r.t. *changes* in the calibration data. One popular approach is the use of *bootstrapping* techniques. Originally suggested in [31], bootstrapping creates multiple *pseudo data-sets* by *simulating* repeated experimentation using *random selection with replacement.* By randomly drawing a *new* set of samples from a single original data-set, one can create a data-set that has *somewhat* different information content compared with the original. The process can be repeated to create a large number of such pseudo data-sets. These data-sets are naturally *not* independent, since they are all derived from the same original, but for appropriate choices of bootstrap method and hyper-parameters they may exhibit enough variation to provide a reasonable estimate of the parameter consistency. This process closely fits the intuition of the CI, i.e., estimating parameters from different data-sets to compute the variation of the resulting parameter estimates [99, 100]. Bootstrapping is a popular method, and has formed the basis for many statistical and machine learning methods, such as *Random Forrest* [53] and *XGBoost* [54].

However, drawing samples randomly with replacement, as outlined in [31], presumes that the samples are *independently and identically distributed* (i.i.d), which is clearly *not* the case for time-series data. Hence, using the bootstrapping framework on time-series data requires some adaptions of the method. These revised bootstrapping methods are typically divided in two classes: *parametric* and *non-parametric* bootstrapping [29]. The parametric bootstrapping method seeks to create i.i.d. data by *de-trending* through calibrating a model, e.g., an ARMAX structure, that captures the behaviour of the underlying system in the data, leaving only *residuals* that are presumed i.i.d. for a sufficient model [29]. New data-sets can then be created by bootstrapping on the residuals and adding them to the trend derived from simulating the calibrated model [29].

In contrast, the non-parametric class of bootstrapping methods for time-series data does *not* rely on de-trending by model calibration, but instead applies bootstrapping directly on the time-series data by various forms of *segmentation* [29]. By randomly drawing *blocks* of data, rather than individual samples, the time-series dependency information is maintained while still allowing the creation of pseudo data-sets that randomly contain different parts of the information content of the original data [29]. The demarcation of the non-parametric bootstrapping methods is in *how* the segmentation of one original data-set into randomly selected blocks is performed. Both methods are compared and discussed in further detail in Paper D [27] with a short overview given in the sequel. Paper D [27] also discusses how parameter estimation can be performed on a set of small data-blocks, similarly to the use of multiple data-sets used in the CTSM framework [16].
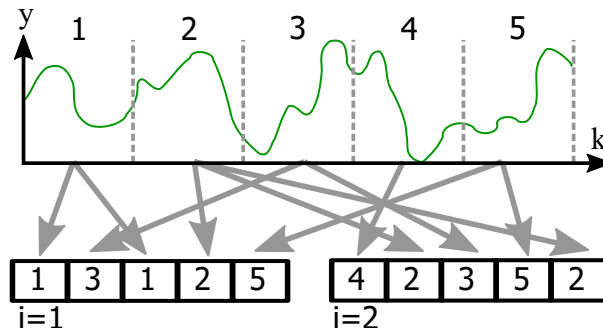
Figure 4.2: An example of simple block bootstrapping with 5 blocks, creating 2 pseudo data-sets. (Figure from Paper D [27])

4.3.1.1 Simple block-based bootstrapping

The *non-overlapping block method,* arguably the simplest form of non-parametric bootstrapping, divides the data-set into blocks of fixed length, where each block is constructed by taking a consecutive sequence of samples from the original data [117, 118]. A pseudo data-set is then created by drawing **K** blocks randomly with replacement, as illustrated in Fig. 4.2. The random selection with replacement used in bootstrapping gives that some blocks may be chosen multiple times in a specific pseudo data-set, while others are not chosen at all. This is shown in Fig. 4.2 where, e.g., block 1 appears twice in the first pseudo data-set, as the 1$^{\text{st}}$ and 3$^{\text{rd}}$ block, while block 4 is not included.

4.3.1.2 Stationary bootstrapping

The *stationary bootstrapping* method, presented in [30], is a somewhat more sophisticated block-based bootstrapping method, which creates bootstrapped data-sets that are themselves *stationary* series [29, 30]. This is achieved by constructing blocks of *random length* that follows a geometric distribution [29]. The stationary bootstrap method is realised by use of a probability test, and consists of two steps. The starting point of each block is drawn uniformly across all *N* original samples. Next, with probability $1 - p$, the consecutive sample from the original series is added to the current block. Alternatively, with probability *p*, a new block is started by again uniformly drawing a new starting point. This test on *p* is repeated until the combined length of all blocks is approximately *N*. An illustration of this algorithm is given in the left panel of Fig. 4.3 The resulting blocks length will follow a geometric distribution such that the expected length of each block is $\frac{1}{p}$ and the expected total number of blocks is $\frac{N}{p}$. An illustrative example of two iterations of stationary bootstrapping is shown in the right panel of Fig. 4.3.
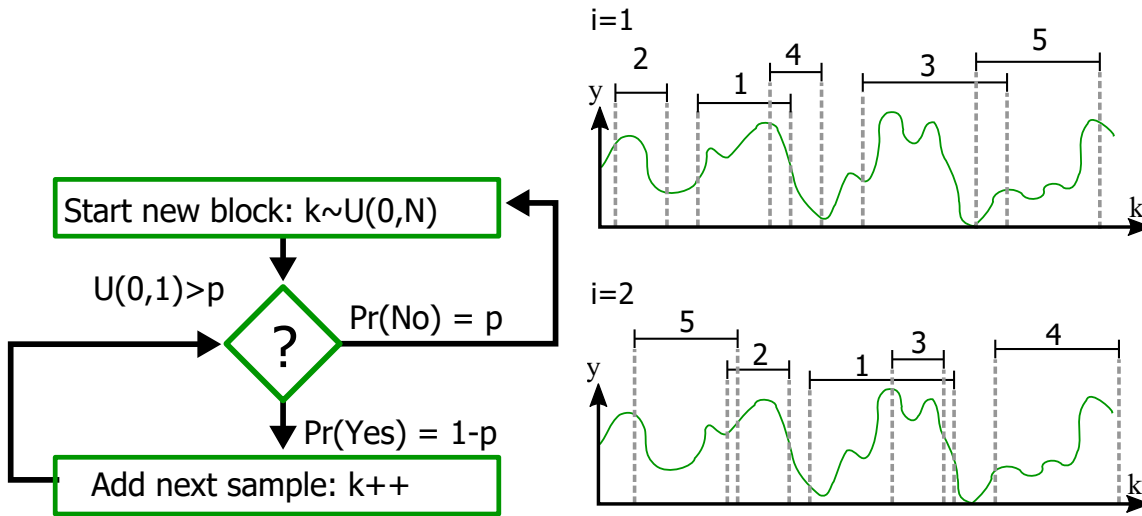
Figure 4.3: Left: Simplified block diagram of Stationary Bootstrapping.
Right: Illustrative example of two iterations of stationary bootstrapping, resulting in five blocks of data, possibly overlapping, with uniformly drawn starting point and geometric length distribution. (Figure from Paper D [27])

4.3.2 1D Profile Likelihood for a moving window

As discussed in Sections 2.3.3 and 4.2, there are situations where parameters cannot be *unambiguously* identified due to a lack of *dynamic information content* in the data. When block-based bootstrapping methods are used, this concern is exacerbated by the fact that information content can *vary* across a data-set, i.e., there can be segments of a data-set that has significantly less dynamic information content than the rest of the data-set. If, by random change, a bootstrapped pseudo data-set predominantly contains repeated segments from *non-informative* sections of the original data-set, the resulting pseudo data-set may be non-informative as a whole, hence resulting in *non-identifiable* parameters. Estimating parameters for a non-informative pseudo data-set produces ambiguous results that can present as *outliers* in the resulting set of bootstrapped parameter estimates. These outliers will significantly affect the estimated *consistency* for the parameters. Hence, it is interesting to diagnose variation of dynamic information content prior to application of bootstrapping methods.

One possible approach is the novel method presented in Paper D [27], which utilised the PL method computed for a *moving window* over the original data. This method tests the consistency of *both* the optimal parameter estimate and the parameter *identifiability* over time. If the optimal estimate varies significantly, this indicates inconsistent dynamic information in the data. More importantly, the method will highlight segments of data that are non-informative for parameter estimation specifically for the model structure of interest. Rather than testing dynamic information content based on statistical information theory, the method uses the *model structure*

and the parameters $\boldsymbol{\theta}$ directly with the PL method to test information content in data by diagnosing identifiability.

4.3.3 Random initial conditions

Another potential source of inconsistency of point estimates for parameters, which is not related to the data itself, is the *initial guess* $\boldsymbol{\theta}_0$ used as a starting point for numerical optimisation. If there are flat manifolds in the likelihood / posterior hyper-surface the initial guess $\boldsymbol{\theta}_0$ can influence the obtained optimal estimate significantly, due to ambiguity of the optimal parameter estimate. A simple yet intuitive and effective method to test for such inconsistencies was presented in Paper A [23] and further discussed in Paper D [27]. By numerically optimising a point estimate for the parameters $\hat{\boldsymbol{\theta}}$ with the initial guess drawn uniformly at random $\boldsymbol{\theta}_0 \sim U(\Theta)$, and subsequently repeating this procedure a reasonably large number of times, say, 50 to 100, it is possible to *cognitively* evaluate the degree of influence from the initial guess to the presumed optimal estimate. If such an influence is found, it may indicate parameter non-identifiability. At present, Paper A [23] has received 18 citations, many of which has taken an interest in the fact that the optimal estimate obtained by numerical optimisation can in some cases be influenced by the initial guess.

4.4 Resolving ambiguous parameter estimates (Scope 3)

Ambiguity of estimated parameter occurs due to insufficient information in the calibration data w.r.t. the model structure, e.g., insufficient measurands recorded from the system resulting in parameters that does not affect model output estimates, or inadequate experimental design resulting in insufficient dynamic information content in the data. Hence, resolving ambiguous parameter estimates requires the addition of new information, either measurements taken from other properties of the system with corresponding changes to the model equations, or experimental re-design for improved system excitation. Alternatively, if the source of ambiguity or non-identifiability is structural the model structure can be revised to better fit the available measurements. A fourth possibility is to introduce a priori information about the parameters derived from the specifications of the physical system, by introducing non-flat priors $p(\boldsymbol{\theta})$.

4.4.1 Application of priors

Applications of physical information to determine priors is typically associated with the *Bayesian* framework for parameter estimation. However, priors are used also

in a *frequentist* framework, e.g., to utilise previous estimation results [16]. What separates the Bayesian paradigm from the frequentist in this regard is the *interpretation* of priors as a *subjective degree of belief* about the distributed probability of the parameters. Further, as discussed in in Section 4.2.3, the use of bounds on the parameters, which is typical of a *frequentist* numerical optimisation approach, is equivalent to the addition of a bounded uniform prior.

As discussed in Paper E [28], the interpretation of *bounds* on estimated parameters can be seen from the Bayesian perspective as *reshaping* the likelihood hyper-surface. This perspective is particularly interesting when parameters are inter-dependent, such that *active* bounds on one parameter can limit the likelihood profile of other parameters, as also discussed in Paper E [28]. The log posterior $\ell_{\mathrm{P}}\left(\boldsymbol{\theta}; y_{[N]}\right)$ can be written

$$\ell_{\mathrm{P}}\left(\boldsymbol{\theta}; y_{[N]}\right) = \ell_{\mathrm{L}}\left(\boldsymbol{\theta}; y_{[N]}\right) - 2\ln p\left(\boldsymbol{\theta}\right) \tag{4.12}$$

By taking the Hessian of Eq. (4.12):

$$H_{\mathrm{P}} = H_{\mathrm{L}} - 2\ln p\left(\boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{4.13}$$

Where $H_{\mathrm{L}} = \nabla^T \nabla \ell_{\mathrm{L}}\left(\boldsymbol{\theta}; y_{[N]}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ and $H_{\mathrm{P}} = \nabla^T \nabla \ell_{\mathrm{P}}\left(\boldsymbol{\theta}; y_{[N]}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. Hence, the effect of introducing a non-flat prior, say a Normal distribution around some nominal value $\boldsymbol{\theta}^0$, is to *reshape* the curvature of the log posterior hyper-surface, by introducing another source of *sensitivity* to perturbations of $\boldsymbol{\theta}$. It is interesting to note that the use of priors on the parameters can be considered as a form of *regularisation* to improve the *generalisation* of a calibrated model [19, 40].

Ambiguity of estimated parameters, which is caused by perturbations of $\boldsymbol{\theta}$ resulting in no, or only statistically insignificant, changes in the posterior $p\left(\boldsymbol{\theta}|y_{[N]}\right)$, can be reduced by application of a non-flat prior. How to determine these priors depends on both available information of the physical system and to some degree on the type of parameter estimation paradigm used, i.e. frequentist or Bayesian. However, a detailed discussion of the differences between *subjective* and *objective* priors is beyond the scope of the present text [24, 96].

4.5 Stochastic predictions (Scope 4)

Given that parameters of TN grey-box models are estimated with uncertainty, the Bayesian framework is particularly interesting. By obtaining the full posterior distribution of the parameters, described by a set of *samples* $\boldsymbol{\theta}_{[K]}$ drawn in proportion to $p\left(\boldsymbol{\theta}|y_{[N]}\right)$, it is possible to carry this uncertainty into the model predictions. It is well know that for a linear system uncertainty of the estimated state can be carried by propagating the covariance matrices, as is done in the Kalman filter [43]. When computing forecasts, i.e., simulating future behaviour of a system, a similar

propagation of uncertainty can be used to estimate the prediction uncertainty for the forecast, given the measurement and process noise covariance matrices. However, for non-linear models, or models for which obtaining the standard linear model matrices is intractable, e.g., an external simulator as described in Section 4.1, a Monte Carlo (MC) simulation is typically used. Rather than propagating the uncertainties using measurement and process noise *covariance matrices* in the linear equations, the measurement and process errors w_k and v_k are *drawn* by using a random number generator (RNG). The model is simulated a large number of times to obtain a set of different state trajectories that can be used to estimate the *mean* and *covariance* of the forecast at each future time-step. Given that the MCMC method produces a set of samples $\boldsymbol{\theta}_{[K]}$ with a distribution proportional to the posterior $p\left(\boldsymbol{\theta}|y_{[N]}\right)$, a possible extension of the MC simulation is to generate separate trajectories for all, or a sub-set of, the parameters in $\boldsymbol{\theta}_{[K]}$, where the errors w_k and v_k are draw by an RNG as before. The resulting set of trajectories can then be used to compute forecast uncertainty which also includes the *parameter uncertainty* propagated to the model output. This approach was used in Paper E [28], and also in other works on stochastic parameter estimation and forecasting, e.g, [20].

5 Summary of papers

5.1 Paper A - Parameter estimation for grey-box models of building thermal behaviour

Paper A analyses the *ambiguity* of the estimated parameters for a typical thermal network model [14]. The parameters are obtained by minimising the root mean square error (RMSE) between the temperature trajectory from a *deterministic* simulation and the measured temperatures of the experimental building. Hence, the paper treats the R3C2 model as a grey-box model in the somewhat wider sense, i.e., a simplified physical model with calibrated parameters, and not in the stochastic differential equation (SDE) sense discussed in Section 2.3.1.

The main contribution of this paper is using two methods to *visually* obtain the *shape* of the objective function. The first method, similar to the stochastic method *Uninformed Random Picking* (URP) [112], draws uniform *random* values for the parameters and plots the resulting parameters and RMSE as scatter plots. While simple, the method identifies whether or not the objective function is flat around the minimum, and therefore will result in *ambiguous* parameter estimates. The second method is also a simple visualisation scheme that uses direct numerical optimisation to minimise RMSE combined with a uniformly random initial starting guess $\theta_0 \sim \mathscr{U}(\theta_{\min}, \theta_{\max})$. The final estimate $\hat{\theta}$ is for some model structures and data-sets dependent on the random initial guess and therefore ambiguous. This method repeats the optimisation multiple times with different initial guess θ_0. By plotting the initial guess and the final estimate together, the ambiguity of the estimated parameters can be visualised. Further, the paper shows how the ambiguity can be resolved for the R3C2 model, reducing the degrees of freedom (d.o.f.) by assigning a constant value, pre-computed from physical specification of the building, to one of the parameters. Both methods are used to illustrate how the reduction in d.o.f. obtains an objective function that is *convex* with a well defined minima.

Paper A utilises a validation scheme for the optimised parameters based on cross-validation over four separate data-sets. Each data-set is used in turn to calibrate parameters, while the remaining three data-sets are used as independent validation data. With four data-sets, a total of sixteen experiments are performed, which together shows that the simple R3C2 TN model is capable of generalisation and

therefore predict independent data with reasonable accuracy. The latter is important mainly for the prediction use-case, but can also be used as a *validation* of the estimated parameters in the soft-sensor use-case. The results show that even though there is significant uncertainty on the estimated parameters, the model is still able to predict the system behaviour. However, due to the use of deterministic RMSE as an objective function, it is difficult to quantify the influence of parameter uncertainty on predictions, a topic which was discussed further in Paper E [28].

5.2 Paper B - Parameter estimation for externally simulated thermal network models

The focus of Paper B is to facilitate experimentation and identification of appropriate model structures for TN models by utilising an *external simulator*, as discussed in Section 4.1. Paper B uses the same model as in Paper A, but treats it as a *stochastic differential equation* (SDE), and subsequently applies a method for evaluating the likelihood function $L\left(\boldsymbol{\theta}; y_{[N]}\right)$ based on the *Continuous Time Stochastic Modelling (CTSM)* framework [16]. The first contribution of this paper is the use and comparison of the *Unscented Kalman Filter (UKF)* and the *Ensemble Kalman Filter (EnKF)* as potential substitutes for the *Extended Kalman Filter (EKF)* used in the original CTSM framework. The UKF and EnKF have some advantages for non-linear models, but their main feature of interest in this context is their ability to handle *non-differentiable* models. Hence, the use of UKF and EnKF to estimate one-step ahead prediction residuals facilities the use of an *external circuit simulator* to simulate the thermal network RC model. The use of a circuit simulator to compute state transitions of the TN allows the model to be described simply as a list of interconnected resistors and capacitors, similar to a circuit node-list, which significantly simplifies experimentation with different model structures. The simulator is discussed in Section 4.1. The accuracy of the simulator is compared with the developed model equations in a linear KF, which shows only negligible approximation errors incurred when the external software simulator is used. Further, the results show that the UKF w.r.t. the EnKF is both faster and provides estimates that are closer to what is obtained using the linear KF.

The second contribution of this paper is an extension of the Profile Likelihood (PL) method [21], from the typical use in *one* parameter dimension, to instead project the likelihood function into a plane of *two* parameters $\Theta_{i,j}$. While considerably more computationally demanding, the 2D likelihood profiles provide a more detailed view of the likelihood function and can be used to diagnose parameter interactions. The method is explained in Section 4.2.2.1. The results of applying the proposed PL2D method to the R3C2 structure shows that there are inter-dependant parameters in the model, which explains the ambiguity of the parameters noted in Paper A

and why the issue was resolved by reducing the d.o.f. Hence, the PL2D method is shown to provide diagnostic information on both parameter inter-dependency and identifiability which can be used to refine the structure of a TN model.

5.3 Paper C - Sensor placement and parameter identifiability in grey-box models of building thermal behaviour

Paper C uses the PL2D method presented in Paper B to analyse the likelihood hyper-surface in parameter space using calibration data from two different sensor locations for the temperature T_w in the R3C2 model. The data from both sensors are collected simultaneously over a period of 10 days, together with other system inputs and the reference measurement for the state T_b. The main contribution of the paper is using the PL2D method to show how the different heat transfer characteristics of two of the distinct insulation sections of the experimental building, discussed in Chapter 3, affects the parameter identification. The resulting 2D profiles shows how the increased dynamic information content obtained from a sensor mounted on a low grade insulated wall produces *less* uncertainty in the estimated parameters.

The paper further discusses how the parameter inter-dependency produces functional relations between the parameter such that the model dynamics is the same irrespective of the parameters within a certain range. By simplifying the TN model structure the resulting ambiguity of the parameter are resolved, hence the R2C2 model is found to have identifiable parameters.

The second contribution of this paper is a modification of the computation scheme used in the PL2D method, in which the estimation of free parameters for each point in the prescribed grid in the plane $\Theta_{i,j}$ uses the estimate from a *neighbouring* point as a *warm start*, thus reducing computation time by a factor of 4 to 10, depending on model complexity. Finally, the paper shows that if the likelihood hyper-surface is close to a Gaussian distribution the confidence region obtained by interpolating isolines on the PL2D profile is similar to the ellipsoid bound obtained from the inverse Hessian $\Sigma_{\boldsymbol{\theta}} = H\left(\boldsymbol{\theta}\right)^{-1}\Big|_{\hat{\boldsymbol{\theta}}}$, thus further substantiating the PL2D method results.

5.4 Paper D - Estimating uncertainty of model parameters obtained using numerical optimisation

Paper D is written as a *tutorial* and a comparison of several different methods used in the previous papers, including the randomisation based methods of Paper A, and

the one and two-dimensional PL methods. The paper compares these methods on both a simple first order model, using simulated data with varying forms of input excitation, and on the R2C2 model using data from the experimental building. The comparison also discusses practical aspects of the methods, such as computation time, scalability to higher dimensional parameter space and ease of implementation. Additionally, the paper discusses different ways of representing uncertainty, e.g., as asymptotic or likelihood based confidence intervals and regions.

The paper further discusses the parameter estimation scheme used in CTSM and in papers B to E [22, 26–28], and compares it with the ballistic simulation error approach used in Paper A [23]. The latter is shown mathematically to be a special case of the former on the condition that the model is *identical* to the data generating system such that the state transition uncertainty has covariance $\mathscr{W} \equiv 0$.

Next, the paper discusses the use of *block-based non-parametric bootstrapping* in the context of SDE formulated TN models. These methods are discussed in Section 4.3.1. The use of bootstrapping is shown to occasionally produce *outlier* parameter estimates for some data-sets and model structures. If there are segments in the original data that contain little dynamic information, some of the randomly generated data-sets will *by chance* contain samples only from the *non-informative* parts of the original data. Hence, some of the generated pseudo data-sets will be *non-informative* w.r.t. parameter estimation. The main novel contribution of this paper is the application of the PL method to a *sliding window*. This method tests the dynamic information content by evaluating the practical identifiability of the parameters for a *specific* model. This approach can be used to verify the *consistency* over the data-set of the dynamic information content, i.e., parameter identifiability, uncertainty and the optimal parameter estimate.

5.5 Paper E - Analysing uncertainty in parameter estimation and prediction for grey-box building thermal behaviour models

Paper E follows the work of [24] which extends the PL method by instead projecting the posterior distribution $p\left(\theta | y_{[N]}\right) \propto p\left(y_{[N]} | \theta\right) p(\theta)$ to obtain the Posterior Profile (PP) method. Further, the paper discusses how the PP method can also be extended to create 2D profiles, similarly to the PL2D method. The resulting *projections* of the posterior hyper-surface are similar to the *marginal distributions* obtained by the Markov Chain Monte Carlo (MCMC) method. Hence, the focus of the Paper E is comparing the PP 1D and 2D methods with the MCMC method. Both methods *approximate* the posterior hyper-surface by drawing samples, either

deterministically by a discretised grid or *stochastically* by the sample and accept/reject scheme of MCMC. The difference between the deterministic grid based sampling in the projection methods and the stochastic guided random walk sampling of the MCMC method is mostly visible in the resulting plots for the *low* posterior density regions. Additionally, the two method differ in how the n_θ dimensional posterior hyper-surface is represented, by either projections or marginal distributions, discussed in Section 4.2.3.1. As Paper E explains, both methods obtain essentially the same information, but the MCMC method are *significantly* faster, typically by an order of magnitude or more. However, there are in some cases advantages to the deterministic sampling, since it does not depend on convergence towards the high posterior density region but rather obtains the maximum posterior *globally* in Θ, within the accuracy allowed by the discretisation grid. The main contribution of Paper E is showing the similarities and differences between these methods, and how these methods can be used together to produce a thorough analysis of the estimated parameter.

Further, Paper E discusses some of the differences between the *frequentist* statistics view of parameter estimation which is used in the previous four papers, and the *Bayesian* statistics framework that is the foundation of the MCMC method, and how these differences pertain to the PL, PP and MCMC methods. By treating the parameters as random variables, the *empirical* distribution over the samples obtained by MCMC can be used to produce a *stochastic* forecast of the building temperatures that includes realistic uncertainty estimates for future temperature predictions.

Finally, the paper includes a discussion on how, in a Bayesian framework, the use of a prior distribution $p(\theta)$ obtained from *building specifications* can be used to resolve parameter ambiguity in an otherwise non-identifiable model structure by introducing more information into the estimation problem.

6 Discussion of results

6.1 External simulators for dynamic models

Most published works on the use of thermal networks to model buildings investigate several different model structures in order to determine the *minimum model complexity* that adequately captures the behaviour of the physical building [13–15, 20]. While designing different TN model structures as RC circuits cognitively based on the heat transfer characteristics of a particular building is relatively straight forward, as pointed out by [18], the process of developing equations from the TN models is somewhat labour intensive and error prone. To facilitate experimentation with a higher number of different TN model structures it is useful to simplify the simulation and parameter estimation of these models. Hence, one of the main research objectives of this thesis was to investigate methods for calibrating the parameters of Resistor-Capacitor (RC) circuit analogue models *without* deriving the equations.

The suggested approach, as discussed in Section 4.1, simulates TN models described as a simple list of interconnected components. This is similar to how circuit simulators, e.g. SPICE [111], are used in the electronics field. By utilising a model description such as the one shown in Table 4.1 experimenting with different TN model structures is as simple as combinatorially adding or removing lines to a table. The suggested approach may be useful for applications other than building thermal behaviour modelling, to the extent that RC circuit equivalent models are used in other fields, e.g., to describe mechanical resonator systems such as piezoelectric elements [119].

Since the use of electric circuit analogue models is almost always a simplification of the physical system, it is natural to treat these models in a stochastic differential equations (SDE) grey-box framework [16]. By modifying the parameter estimation method used in the Continuous Time Stochastic Modelling (CTSM) framework [16], specifically by replacing the Extended Kalman Filter (EKF) with either the Unscented Kalman Filter (UKF) [75] or the Ensemble Kalman filter (EnKF) [74], the likelihood function $L\left(\boldsymbol{\theta}; y_{[N]}\right)$ can be evaluated also for *non-differentiable* models such as an external RC circuit simulator. If a Bayesian interpretation of the parameter estimation problem is used, e.g., combining the work presented in Paper B [22] and Paper E [28], the suggested approach follows similar work on calibrating computer software [105].

The comparison between the UKF and the EnKF presented in Paper B [22] strongly favours the UKF. Given that the EnKF was mainly developed for systems with a large number of states [74], e.g., meteorological models, it is not surprising that UKF has better performance on the comparatively simple TN models. It could also be noted that the EKF, although excluded from the comparison in Paper B [22] due to its inherent dependence on a linearised model representation for covariance propagation, could potentially be used also for externally simulated models in combination with some form of numerical linearisation scheme [120].

6.2 Parameter identifiability

Using either the likelihood function $p\left(y_{[N]}|\boldsymbol{\theta}\right)$, or by inclusion of a prior, the posterior distribution $p\left(\boldsymbol{\theta}|y_{[N]}\right) \propto p\left(y_{[N]}|\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}\right)$, as an objective function to evaluate *model fit* of a parameter estimate $\hat{\boldsymbol{\theta}}$ can be viewed as defining a *hyper-surface* $g\left(\boldsymbol{\theta}\right)$ in the parameter space $\boldsymbol{\Theta} \in \mathbb{R}^{n_{\boldsymbol{\theta}}}$. Given a particular model structure $\mathcal{M}$, the model fit of any *proposed* parameter $g\left(\hat{\boldsymbol{\theta}}\right)$ is evaluated on this hyper surface. Hence, the *shape* of the hyper-surface determines how *different* values of $\boldsymbol{\theta}$ compares w.r.t. their ability to *explain* the observed behaviour of the physical system. Since the likelihood/posterior hyper-surface exists in $n_{\boldsymbol{\theta}}$-dimensional space, analysis of its shape typically require some form of computational approximation or visualisation into 1D or 2D representations for plotting. In the literature on TN models for buildings, uncertainty is typically estimated from a local sensitivity analysis, often based on asymptotic confidence intervals derived from the covariance $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \mathbb{V}\left(\boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$. The inverse covariance can be approximated by the Hessian of the *log likelihood* $\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1} = \nabla^T \nabla \ln p\left(y_{[N]}|\boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$ [16, 21], i.e., the *curvature* of the likelihood hyper-surface around the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$. Alternatively, sensitivity can be obtained from using some other form of heuristic perturbation scheme around the $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ [121].

In the five papers presented in Part II of this thesis the parameter *identifiability* is analysed by inspecting the shape of the hyper-surface $g\left(\boldsymbol{\theta}\right)$. As illustrated by the comparison of the simple Uninformed Random Picking (URP) method and the Profile Likelihood (PL) method in Paper D [27], the PL2D and the Hessian based confidence ellipsoid in Paper C [26], and the 2D Posterior Profile (PP2D) and Markov Chain Monte Carlo (MCMC) in Paper E [28], all these methods are essentially different ways to obtain and visualise the *same* likelihood/posterior hyper-surface $g\left(\boldsymbol{\theta}\right)$. These papers joins a recent trend in the field, as illustrated by the reviews in Section 2.3.3. The advantage of these methods over the Hessian based *asymptotic confidence interval* is their ability to represent *asymmetric* likelihood functions, which in turn can be used to obtain *likelihood based confidence intervals* [21, 25]. These likelihood

based CI's are used to diagnose both *practical* and *structural* parameter identifiability, based on the definition presented in [21]. With the extension to 2D profiling, as presented in Paper B [22], these tools permit a detailed analysis of the parameter space which can be used to revise both model structures and experimental design for data collection. Note that the construction of likelihood based CI's are derived from Wilks theorem [77] and the likelihood ratio (LR) [21]. Hence, if the posterior hyper-surface is analysed for a non-uniform prior, a different framework should be used to create *credibility* regions [9, 106].

The concept of parameter identifiability in the PL framework is defined from the existence of, possibly asymmetric, confidence limits around the optimal point estimate $\hat{\theta}_{\mathrm{MLE}}$, i.e., whether the data is sufficient to determine a *bounded* CI for a given model structure $\mathcal{M}$. Structural non-identifiability is detected by the presence of equipotential manifolds in Θ. In an unbounded parameter space $\Theta = \mathbb{R}^{n_\theta}$ the resulting CI will be unbounded in both directions. The introduction of a bounded parameter space $\Theta \subset \mathbb{R}^{n_\theta}$, equivalent to introducing a uniform window prior $p(\boldsymbol{\theta})$ as discussed in Section 4.2.3, may however reshape the hyper-surface $g(\boldsymbol{\theta})$ such that the functional relation of the structurally non-identifiable parameters produce bounded, but abnormally wide, likelihood based CI's. This effect is observed for the R3C2 model. One of the benefits of the PL2D scheme is the ability to *detect* these artefacts more clearly than what is possible in the 1D implementation.

In contrast to structural non-identifiability, *practically* non-identifiable parameters have a well defined optimum, but due to insufficient dynamic information content in the calibration data it is not possible to distinguish different parameters along some direction in Θ, thus producing a CI that is unbounded in at least one direction. It is interesting to note that although the definitions of parameter non-identifiability is based on the CI definition [25, 99], which is a typical frequentist concept that treats parameters as unknown constants and data as the random variable, the definition of practical identifiability also requires to inspect the likelihood profile itself for the presence of a well defined optimum. Arguably, studying the projected log likelihood in this way conceptually approaches the Bayesian mindset of treating probability as a degree of belief, i.e., asking the question; given the data, what can be inferred or learned about the unknown parameters. The possibility to extend the PL method to project the posterior distribution instead, as discussed in Paper E [28] and [24], further strengthens the sense that these methods, at least in some practical sense, utilises a Bayesian interpretation of the parameter estimation problem.

6.2.1 Resolving ambiguous parameter estimates

Ambiguity of the estimated parameters may be resolved by introducing more information into the estimation problem. In the case of practical non-identifiability,

experimental redesign may be applied to obtain more *informative* data. For structurally non-identifiable parameters one possibility, as pointed out by [21] is the introduction of *qualitatively new* measurements, e.g., measuring more properties of the building. Another possibility, as discussed in Paper E [28], is the introduction of *subjective* priors $p(\boldsymbol{\theta})$ derived from the physical specifications of the building. In the Bayesian setting, priors can be used to introduce a priori information with some *specified uncertainty*. In contrast to the approach of Paper E [28], the more traditional *numerical optimisation* approach in Paper A [23] resolved the ambiguity due to parameter inter-dependence by assigning a pre-computed *constant* value to the parameter R_{g}. Adopting the view that the application of a non-uniform prior distribution effectually *reshapes* the likelihood hyper-surface by "bending" it, the application of a constant can be considered as taking a $n_{\boldsymbol{\theta}} - 1$ dimensional *slice* of Θ, which is equivalent to assigning a prior $p\left(R_{\mathrm{g}} = c\right) = 1$ and $p\left(R_{\mathrm{g}} \neq c\right) = 0$ or a normal distribution with zero standard deviation $p\left(R_{\mathrm{g}}\right) = \mathscr{N}\left(c, 0\right)$. As such, the approach in Paper A [23] can be seen as a special case of the use of priors in Paper E [28].

As shown in Paper E, the presence of non-identifiable parameters may not be problematic if one is interested in only derived parameters, e.g., total thermal resistance or eigenvalues of the system, since these can be identifiable, even though individual RC parameters are not. Hence, the discussion on identifiability related to the soft-sensor use-case may be confined to the parameters of interest. If only the total thermal resistance R_{tot} is needed, it is the identifiability of R_{tot} that must be verified.

6.2.2 Prediction with non-identifiable parameters

If the ambiguity of the estimated parameters is caused by *structural* non-identifiability, i.e., the parameter optimisation problem has many equally good solutions due to the model structure having *redundant* parametrisation, the ambiguous parameter estimates *may* also be equally applicable for prediction on independent data. Intuitively, a model with structurally non-identifiable parameters may still be able to *learn* the system behaviour through calibration and therefore successfully predict future system behaviour. If, however, the ambiguity is caused by a lack of dynamic information in the calibration data, resulting in practically non-identifiable parameters, prediction performance is likely to be adversely affected. Estimated parameters calibrated on *non-informative* data, i.e., data which does not contain sufficient information of the dynamic behaviour of the system, cannot be reasonably expected to perform adequately in a prediction setting.

In Paper A [23] the results for both the R3C2 and R2C2 model presented in Table 5, respectively in columns RMSE5 and RMSE4, show that assigning a constant value to one parameter, thereby reducing the dimension of the parameter space, does not

affect the models ability to fit the calibration data. This is reasonable, since the ambiguity in the R3C2 model is caused by parameter inter-dependency which is a structural problem. Paper A [23] did not investigate the prediction power of the models any further. This topic was therefore revisited in Paper E, where *stochastic forecasting* was used with parameters from both the full model, with or without a prior, and the reduced model. The results in Paper E [28] showed that all the analysed experimental cases obtained *similar* prediction accuracy for an *independent* test set.

The fact that a grey-box model with *structurally* non-identifiable, hence ambiguous, parameters can still provide good predictions of system behaviour on independent data is arguably not surprising. Indeed, in the *black-box* paradigm, ambiguous *optimal* parameter estimates is the norm rather than a special case. For example, in system identification (SID), a change of *basis* for the state-space results in different state transition, input and measurement matrices [40]. Another example of ambiguous parameters is the training weights in an artificial neural network (ANN). The number of training weights is often very high, on the order of thousands to millions, especially for a multilayer deep learning network. Hence, a *trained*, i.e., parameter optimised, network is never a unique solution, but rather a local minima that shows satisfactory prediction performance [51]. The parameters of both the SID and ANN models are inherently ambiguous due to structural redundancy. Hence, interpretation of the parameters is usually infeasible, perhaps with the exception of certain *derived* parameters like the *eigenvalues* of the state transition matrix which are invariant under change of basis/rotation. Let by eigen-decomposition $A = M^{-1}\Lambda M$ and $\tilde{A} = T^{-1}AT = \tilde{M}^{-1}\tilde{\Lambda}\tilde{M}$, where T is a rotation matrix, M is the eigenvectors as columns and Λ is a diagonal matrix of eigenvalues. Then $\tilde{A} = T^{-1}M^{-1}\Lambda MT \rightarrow \tilde{M} = MT$ and $\tilde{\Lambda} = \Lambda$. However, these models are still perfectly suitable for the *prediction* use-case [40].

6.2.3 Information consistency for block-based bootstrapping

One of the novel adaptions of the Profile Likelihood (PL) method presented in this thesis is the use of 1D PL, together with a sliding window, for the purpose of evaluating the *consistency* of dynamic information content a the data-set. The method is particularly useful in combination with block-based bootstrapping, since the presence of non-informative segments of data can produce pseudo data-sets that result in *practically* non-identifiable parameters that present as outliers in the set of bootstrapped parameter estimates. The use of PL 1D for a sliding window to test information consistency may be considered a generally applicable method, and not restricted to building thermal behaviour. It can be seen as a diagnostic tool to analyse why the application of bootstrapping has resulted in some unexpected

Table 6.1: Selected parameter estimates from multiple papers.

#	Model	$R_{\mathrm{b}}\left[\frac{K}{W}\right]$	$R_{\mathrm{w}}\left[\frac{K}{W}\right]$	$C_{\mathrm{b}}\left[\frac{J}{K}\right]$	$C_{\mathrm{w}}\left[\frac{J}{K}\right]$	$R_{\mathrm{tot}}\left[\frac{K}{W}\right]$
Paper A Nov1	R3C2 4DOF	0.057	0.066	1.271×10^6	1.673×10^6	0.081
Paper A Nov2	-	0.041	0.079	1.107×10^6	1.321×10^6	0.080
Paper A Dec1	-	0.033	0.072	1.033×10^6	1.519×10^6	0.073
Paper A Dec2	-	0.029	0.087	0.852×10^6	1.241×10^6	0.078
Paper C S1	R2C2	0.040	0.048	1.267×10^6	0.419×10^6	0.088
Paper C N3	-	0.035	0.051	1.137×10^6	2.735×10^6	0.086
Paper E Case 2	R3C2 w/prior	0.072	0.084	1.444×10^6	0.293×10^6	0.093
Paper E Case 3	R2C2	0.043	0.051	1.446×10^6	0.481×10^6	0.094

dispersion in estimated parameters, or as a precursory test to verify that block-based bootstrapping methods can be applied with a reasonable expectation of success.

6.3 Interpretation of parameters

TN models reduce the somewhat complex heat transfer characteristics of a building to a simplified flow of thermal energy between a few *point nodes*. When calibrating the parameters of such a model, the choice of *reference* signal for the temperature at each of these *nodes* will determine the physical interpretation of the estimated parameters. Although the components of a model, e.g., the thermal resistance R_{w} in the R2C2 and R3C2 models, are intended to represent a specific property of the building, the thermal resistance of the envelope in the case of R_{w}, the estimated values will be adapted to the measured thermal variation at the actual sensor location. Hence, the placement of sensors significantly influences the interpretation of the estimated parameters as representative of the heat transfer characteristics of the building.

Table 6.1 compares selected parameter estimates from different cases in Papers A, C and E [23, 26, 28]. The results for Paper A and Paper E Case 2 are obtained by calibrating the R3C2 model with $R_{\mathrm{g}} = \mathcal{N}\left(0.24, \sigma_{\mathrm{g}}^2\right)$ where $\sigma_{\mathrm{g}} = 0$ in Paper A. It is interesting to note that despite the data being collected from the *same* experimental building, admittedly using partially different data acquisition systems, the variation for some of the parameters is significant. In particular, the values of C_{w} are heavily dependent on the choice of sensor location, as discussed in Paper C [26]. This could also explain the deviations in the C_{w} estimates from Paper A [23] given that a different sensor setup was used. Further, the estimates in Paper A were obtained using the *ballistic simulation* approach and may therefore be affected by the estimation procedure as well as the sensor location. However, the parameters are in most cases relatively consistent for each set of calibration data / paper, in particular for

the overall thermal resistance R_{tot}, even though different model configurations where used. Hence, it is reasonable to assume that the *main* contribution to the variations between the different data-sets are driven by unmodelled disturbances. However, the overall thermal resistance R_{tot}, which is probably the estimated parameter that would be of most significance in terms of energy efficiency classification, is relatively consistent across the papers.

6.4 Grey-box SDE modelling as a general framework

Arguably, the most common approach to model fitting is the use of direct numerical optimisation of parameters in a ballistic simulation with measured inputs. Although models are almost always an approximation of the real system, hence subject to epistemic and aleatoric uncertainty, the use of the SDE formulation is apparently less common. In that respect, it is interesting to observe that the parameter estimation scheme adopted from CTSM can be *reduced* to the ballistic simulation approach, as discussed in Papers D [27] and E [28], by setting the process noise covariance $\mathcal{W} \equiv 0$, and also assuming perfect initial information of the state such that $X_0 \equiv 0$. Both of these assumptions is inherent in the ballistic simulation approach. In this limit case, the Kalman gain $K_k = 0$ thus ignoring measurements and the state trajectory is computed *only* from model predictions as a ballistic simulation. The interpretation of a model with zero process noise covariance is a fully deterministic model that *perfectly* predicts the state transitions in the data generating system, and therefore is only subject to *aleatoric* measurement uncertainty. While this may be a reasonable approximation in some cases, particularly for simulated data, it seems reasonable that the use of the full SDE formulation, with subsequent application of the parameter estimation scheme, may be beneficial for most real systems [16, 17].

The other limit case occurs when measurement noise covariance is zero $\mathcal{V} \equiv 0$. As discussed in Paper E [28], in the linear case, this results in the parameter being fitted to predict $\hat{y}_k = f(y_{k-1}, u_k, \theta)$ and therefore ignoring any *previous* measurements $y_{[k-2]}$. The Kalman gain $K_k = \tilde{C}^{-1}$ and therefore the a posteriori state update $\hat{x}_{k|k} = \tilde{C}^{-1} y_k$ ignores the apriori state estimate completely. In the more typical cases where both process and measurement noise are non-zero, the Kalman gain balances the influence of both error sources. The parameters are fitted such that the *estimated* state trajectory best explains the data with minimal squared error. Hence, the parameter scheme used in CTSM, and adopted in Papers B to E [22, 26–28], can be seen as a general approach that balances the error terms. Returning to the discussion on prediction error versus ballistic simulations in [18], the KF based approach can be seen as an intermediate solution.

7 Conclusions and future work

The main research objective of this thesis is the study of ambiguity in estimated parameters for thermal network (TN) models. This topic is discussed in all five papers, with particular focus on the Profile Likelihood (PL) method [21, 25, 122], its Profile Posterior (PP) variation [24], and the suggested 2D variations of these methods. Combining these *projection* based methods with the popular Markov Chain Monte Carlo (MCMC) method and the *randomisation* based methods presented in Paper A [23] permits a detailed analysis of the identifiability of the TN model parameters, as well as flexibility in choice of analysis methods. Applying the PL method to a *sliding window* provides a diagnostic method for verification of the consistency of dynamic information over time. This is particularly useful when block-based bootstrapping methods are used to test consistency of the estimated parameters against variations in the calibration data. Together, these methods form a diagnostic framework that can be used to determine if parameters of a model can be uniquely identified, detect structural and practical non-identifiability, and to suggest appropriate action in the cases where identifiability is not observed.

The second research objective is to develop a framework that simplifies experimentation with TN model structures. This is achieved by utilising an *external simulator* that describes a TN model as a table of interconnected components. Given that TN models are in general strong simplifications of the physical building, with significant epistemic and aleatoric uncertainty in their formulation, the use of stochastic differential equations (SDE) provides a beneficial framework for parameter estimation. Consequently, the parameter estimation scheme based on the popular Continuous Time Stochastic Modelling (CTSM) framework was adapted to handle *non-differentiable* models such that it can be utilised for parameter calibration of the external simulator. By evaluating the likelihood function, or the posterior distribution by inclusion of a prior distribution, the resulting parameter estimation method provides a statistical foundation for parameter estimation and analysis.

7.1 Future work

Electric circuit analogue models is also used in other engineering fields, besides thermal modelling, e.g., to describe mechanical resonator systems such as piezo-electric elements [113, 119]. As is the case with TN models, the choice of circuit

representation is often somewhat flexible, and the parameters of the circuit analogue model requires calibration [119]. Hence, the combination of a grey-box SDE parameter estimation scheme and a circuit analogue simulator may be of use in applications other then building thermal modelling.

The simulator developed in Paper B is limited to resistor and capacitor elements, since these are the building blocks needed for thermal modelling. A further extension could be to include inductive elements L, such that the resulting RLC circuits could be used to model any linear system as an equivalent electric circuit. Since the UKF and the EnKF also handles non-linear state transitions, the external simulator could be extended to include non-linear components as well, which may or may not have an actual electric equivalent implementation. Two examples could be a resistive elements which creates a drop in potential proportional to the *square* of the flow, i.e., $p = R \cdot q^2$, similar to a *friction* term in fluid mechanics [123], or a variable resistor $R = f(u)$ which could be used to model the variable heat convection due to wind or ventilation [14].

Another interesting possibility is the *automation* of model structure identification. When a TN model is described on the form of Table 4.1, the experimental identification process of the TN structure is simply a question of combinatorial changes to the table. This process could possibly be automated, by repeated modifications to the model structure followed by parameter optimisation, model validation and possibly an identifiability analysis to detect redundancy. One requirement for automatic manipulation of model structures is a well defined validation scheme that can compare and rank different model structures with parameters that may be non-identifiable, including a way to penalise excessively complex model structure that are likely to over-fit calibration data.

Bibliography

[1] Recast, E.P.B.D., Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), Official Journal of the European Union 18 (06) (2010) 2010.

[2] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy and Buildings 40 (3) (2008) 394 – 398. `doi:https://doi.org/10.1016/j.enbuild.2007.03.007`.

[3] A. Kaminska, Impact of heating control strategy and occupant behavior on the energy consumption in a building with natural ventilation in poland, Energies 12 (22). `doi:10.3390/en12224304`.

[4] Kommunal- og regionaldepartementet, Byggteknisk forskrift (TEK17), Lovdata, Jun.

[5] M. Killian, M. Kozek, Ten questions concerning model predictive control for energy efficient buildings, Building and Environment 105 (2016) 403–412.

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

[7] V. Gori, C. A. Elwell, Characterization of the thermal structure of different building constructions using in-situ measurements and bayesian analysis, Energy Procedia 132 (2017) 537–542.

[8] V. Gori, V. Marincioni, P. Biddulph, C. A. Elwell, Inferring the thermal resistance and effective thermal mass distribution of a wall from in situ measurements to characterise heat transfer at both the interior and exterior surfaces, Energy and Buildings 135 (2017) 398–409.

[9] V. Gori, A novel method for the estimation of thermophysical properties of walls from short and seasonally independent in-situ surveys, Ph.D. thesis, UCL (University College London) (2017).

[10] R. C. Sonderegger, Dynamic models of house heating based on equivalent thermal parameters, Ph.D. thesis, Princeton Univversity (1977).

Bibliography

[11] R. R. McMurchy, R. G. Payne, R. L. Dotts, Thermal network modeling handbook, Tech. Rep. NASA-CR-144418, TRW-14690-H003-RO-00, National Aeronautics and Space Administration, MannedSpacecraft Center (1972).

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

[13] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy and Buildings 43 (7) (2011) 1511 – 1522. `doi:https://doi.org/10.1016/j.enbuild.2011.02.005`.

[14] T. Berthou, P. Stabat, R. Salvazet, D. Marchio, Development and validation of a gray box model to predict thermal behavior of occupied office buildings, Energy and Buildings 74 (2014) 91–100.

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

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

[17] T. Bohlin, S. F. Graebe, Issues in nonlinear stochastic grey box identification, International journal of adaptive control and signal processing 9 (6) (1995) 465–490.

[18] A. Rabl, Parameter estimation in buildings: methods for dynamic analysis of measured energy use, Journal of solar energy engineering 110 (1) (1988) 52–66.

[19] S. Rouchier, Solving inverse problems in building physics: An overview of guidelines for a careful and optimal use of data, Energy and Buildings 166 (2018) 178–195.

[20] S. Rouchier, M. Rabouille, P. Oberlé, Calibration of simplified building energy models for parameter estimation and forecasting: Stochastic versus deterministic modelling, Building and Environment 134 (2018) 181–190.

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

[22] O. M. Brastein, B. Lie, R. Sharma, N.-O. Skeie, Parameter estimation for externally simulated thermal network models, Energy and Buildings`doi:https://doi.org/10.1016/j.enbuild.2019.03.018`.

[23] O. M. Brastein, D. Perera, C. Pfeiffer, N.-O. Skeie, Parameter estimation for grey-box models of building thermal behaviour, Energy and Buildings 169 (2018) 58 – 68. `doi:https://doi.org/10.1016/j.enbuild.2018.03.057`.

[24] A. Raue, C. Kreutz, F. J. Theis, J. Timmer, Joining forces of bayesian and frequentist methodology: a study for inference in the presence of non-identifiability, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371 (1984) (2013) 20110544.

[25] W. Q. Meeker, L. A. Escobar, Teaching about approximate confidence regions based on maximum likelihood estimation, The American Statistician 49 (1) (1995) 48–53.

[26] O. M. Brastein, R. Sharma, N.-O. Skeie, Sensor placement and parameter identifiability in grey-box models of building thermal behavior, in: Proceedings of The 60th Conference on Simulation and Modelling (SIMS 60), 13-16 August 2019, Västerås, Sweden, Linköping University Electronic Press, 2009, p. tbd. `doi:DOI:10.3384/ecp2017051`.

[27] O. M. Brastein, B. Lie, C. F. Pfeiffer, N.-O. Skeie, Estimating uncertainty of model parameters obtained using numerical optimisation, Modeling, Identification and Control 40 (4) (2019) 213–243. `doi:10.4173/mic.2019.4.3`.

[28] O. M. Brastein, A. Ghaderi, C. F. Pfeiffer, N.-O. Skeie, Analysing uncertainty in parameter estimation and prediction for grey-box building thermal behaviour models, Energy and Buildings`doi:https://doi.org/10.1016/j.enbuild.2020.110236`.

[29] D. N. Politis, The impact of bootstrap methods on time series analysis, Statistical Science (2003) 219–230.

[30] D. N. Politis, J. P. Romano, The stationary bootstrap, Journal of the American Statistical association 89 (428) (1994) 1303–1313.

[31] B. Efron, Bootstrap Methods: Another Look at the Jackknife, The Annals of Statistics 7 (1) (1979) 1–26.

[32] A. Afram, F. Janabi-Sharifi, Review of modeling methods for hvac systems, Applied Thermal Engineering 67 (1-2) (2014) 507–519.

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

[34] D. Perera, D. Winkler, N.-O. Skeie, Multi-floor building heating models in matlab and modelica environments, Applied energy 171 (2016) 46–57.

[35] W. Perera, C. F. Pfeiffer, N.-O. Skeie, Modeling and simulation of a multizone building for better control, in: Proceedings of the 55th Conference on Simulation and Modelling (SIMS 55), Modelling, Simulation and Optimization, 21-22 October 2014, Aalborg, Denmark, no. 108, Linköping University Electronic Press, 2014, pp. 268–276.

[36] D. Perera, C. F. Pfeiffer, N.-O. Skeie, Modelling the heat dynamics of a residential building unit: Application to Norwegian buildings, Modeling, Identification and Control 35 (1) (2014) 43–57. `doi:10.4173/mic.2014.1.4`.

[37] A. Afram, F. Janabi-Sharifi, Gray-box modeling and validation of residential HVAC system for control system design, Applied Energy 137 (2015) 134–150.

[38] G. R. Ruiz, C. F. Bandera, T. G.-A. Temes, A. S.-O. Gutierrez, Genetic algorithm for building envelope calibration, Applied Energy 168 (2016) 691–705.

[39] T. Söderström, P. Stoica, System identification, Prentice-Hall, Inc., 1988.

[40] L. Ljung, System Identification: Theory for the User, Prentice Hall information and system sciences series, Prentice Hall PTR, 1999.

[41] J. Cigler, S. Prívara, Subspace identification and model predictive control for buildings, in: 2010 11th International Conference on Control Automation Robotics & Vision, IEEE, 2010, pp. 750–755.

[42] H. Kwakernaak, R. Sivan, Linear optimal control systems, Vol. 1, Wiley-interscience New York, 1972.

[43] D. Simon, Optimal state estimation: Kalman, H infinity, and nonlinear approaches, John Wiley & Sons, 2006.

[44] D. Di Ruscio, Combined Deterministic and Stochastic System Identification and Realization: DSR - A Subspace Approach Based on Observations, Modeling, Identification and Control 17 (3) (1996) 193–230. `doi:10.4173/mic.1996.3.3`.

[45] B. J. Odelson, M. R. Rajamani, J. B. Rawlings, A new autocovariance least-squares method for estimating noise covariances, Automatica 42 (2) (2006) 303–308.

[46] S. Akhlaghi, N. Zhou, Z. Huang, Adaptive adjustment of noise covariance in kalman filter for dynamic state estimation, in: 2017 IEEE Power & Energy Society General Meeting, IEEE, 2017, pp. 1–5.

[47] B. Telsang, J. Dong, M. Olama, T. Kuruganti, S. Djouadi, Nuclear-norm-based subspace identification of multi-zone building hvac system, in: 2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC), IEEE, 2019, pp. 165–169.

[48] M. Verhaegen, A. Hansson, Nuclear norm subspace identification (n2sid) for short data batches, IFAC Proceedings Volumes 47 (3) (2014) 9528–9533.

[49] M. Fazel, H. Hindi, S. P. Boyd, et al., A rank minimization heuristic with application to minimum order system approximation, in: Proceedings of the American control conference, Vol. 6, Citeseer, 2001, pp. 4734–4739.

[50] A. Hansson, Z. Liu, L. Vandenberghe, Subspace system identification via weighted nuclear norm optimization, in: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, 2012, pp. 3439–3444.

[51] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, Vol. 1, MIT press Cambridge, 2016.

[52] M. Kuhn, K. Johnson, Applied predictive modeling, Vol. 26, Springer, 2013.

[53] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[54] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM, 2016, pp. 785–794.

[55] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., Mastering the game of go with deep neural networks and tree search, nature 529 (7587) (2016) 484.

[56] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, et al., Stanley: The robot that won the darpa grand challenge, Journal of field Robotics 23 (9) (2006) 661–692.

[57] M. Killian, B. Mayer, M. Kozek, Effective fuzzy black-box modeling for building heating dynamics, Energy and Buildings 96 (2015) 175–186.

[58] P. Ferreira, A. Ruano, S. Silva, E. Conceicao, Neural networks based predictive control for thermal comfort and energy savings in public buildings, Energy and buildings 55 (2012) 238–251.

[59] P. O. Fanger, et al., Thermal comfort. analysis and applications in environmental engineering., Thermal comfort. Analysis and applications in environmental engineering.

[60] ANSI ASHRAE, Standard 55-2010: Thermal environmental conditions for human occupancy, ASHRAE, Atlanta USA.

[61] A. H. Land, A. G. Doig, An automatic method of solving discrete programming problems, Econometrica 28 (3) (1960) 497–520.

[62] R. Juhl, J. K. Møller, H. Madsen, ctsmr-Continuous Time Stochastic Modeling in R, arXiv preprint arXiv:1606.00242.

[63] N. R. Kristensen, H. Madsen, Continuous time stochastic modelling, Mathematics Guide (2003) 1–32.

[64] R. E. Kalman, A new approach to linear filtering and prediction problems, Journal of basic Engineering 82 (1) (1960) 35–45.

[65] R. E. Kalman, Contributions to the theory of optimal control, Bol. soc. mat. mexicana 5 (2) (1960) 102–119.

[66] R. E. Kalman, R. S. Bucy, New results in linear filtering and prediction theory, Journal of basic engineering 83 (1) (1961) 95–108.

[67] H. Madsen, Time series analysis, Chapman and Hall/CRC, 2007.

[68] R. Johansson, System Modeling and Identification, Information and system sciences series, Prentice Hall, 1993.

[69] M. Jiménez, H. Madsen, J. Bloem, B. Dammann, Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules, Energy and Buildings 40 (2) (2008) 157–167.

[70] G. L. Smith, S. F. Schmidt, L. A. McGee, Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle, Tech. rep. (1962).

[71] B. A. McElhoe, An assessment of the navigation and course corrections for a manned flyby of mars or venus, IEEE Transactions on Aerospace and Electronic Systems (4) (1966) 613–623.

[72] A. H. Jazwinski, Stochastic processes and filtering theory, Dover Publications, Inc, 1970.

[73] E. A. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in: Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000, IEEE, 2000, pp. 153–158.

[74] G. Evensen, The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Systems 29 (3). `doi:10.1109/MCS.2009.932223`.

[75] S. J. Julier, J. K. Uhlmann, New extension of the kalman filter to nonlinear systems, in: Signal processing, sensor fusion, and target recognition VI, Vol. 3068, International Society for Optics and Photonics, 1997, pp. 182–194.

[76] S. J. Julier, J. K. Uhlmann, A general method for approximating nonlinear transformations of probability distributions, Tech. rep., Technical report, Robotics Research Group, Department of Engineering Science (1996).

[77] S. S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, The Annals of Mathematical Statistics 9 (1) (1938) 60–62.

[78] P. J. C. Vogler-Finck, J. Clauß, L. Georges, I. Sartori, R. Wisniewski, Inverse model identification of the thermal dynamics of a norwegian zero emission house, in: D. Johansson, H. Bagge, Å. Wahlström (Eds.), Cold Climate HVAC 2018, Springer International Publishing, Cham, 2019, pp. 533–543.

[79] R. H. Byrd, J. C. Gilbert, J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, Mathematical programming 89 (1) (2000) 149–185.

[80] R. A. Waltz, J. L. Morales, J. Nocedal, D. Orban, An interior algorithm for nonlinear optimization that combines line search and trust region steps, Mathematical programming 107 (3) (2006) 391–408.

[81] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and computers in simulation 55 (1-3) (2001) 271–280.

[82] Y. Tomita, A. A. Damen, P. M. V. D. HOF, Equation error versus output error methods, Ergonomics 35 (5-6) (1992) 551–564.

[83] M. Jiménez, H. Madsen, K. K. Andersen, Identification of the main thermal characteristics of building components using matlab, Building and Environment 43 (2) (2008) 170–180.

[84] S. F. Fux, A. Ashouri, M. J. Benz, L. Guzzella, EKF based self-adaptive thermal model for a passive house, Energy and Buildings 68 (2014) 811–817.

[85] J. C. Lagarias, J. A. Reeds, M. H. Wright, P. E. Wright, Convergence properties of the nelder–mead simplex method in low dimensions, SIAM Journal on optimization 9 (1) (1998) 112–147.

[86] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Artificial Intelligence, Addison-Wesley Publishing Company, 1989.

[87] M. Affenzeller, S. Wagner, S. Winkler, A. Beham, Genetic algorithms and genetic programming: modern concepts and practical applications, Chapman and Hall/CRC, 2009.

[88] D. B. Crawley, L. K. Lawrie, C. O. Pedersen, F. C. Winkelmann, Energy plus: energy simulation program, ASHRAE journal 42 (4) (2000) 49–56.

[89] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity analysis in practice: a guide to assessing scientific models, Chichester, England.

[90] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: Nsga-ii, IEEE Transactions on Evolutionary Computation 6 (2) (2002) 182–197. `doi:10.1109/4235.996017`.

[91] S. Wang, X. Xu, Simplified building model for transient thermal performance estimation using ga-based parameter identification, International Journal of Thermal Sciences 45 (4) (2006) 419–432.

[92] Z. Boussaada, O. Curea, A. Remaci, H. Camblong, N. Mrabet Bellaaj, A nonlinear autoregressive exogenous (narx) neural network model for the prediction of the daily direct solar radiation, Energies 11 (3) (2018) 620.

[93] R. Bellman, K. J. Åström, On structural identifiability, Mathematical biosciences 7 (3-4) (1970) 329–339.

[94] R. Juhl, J. K. Møller, J. B. Jørgensen, H. Madsen, Modeling and prediction using stochastic differential equations, in: Prediction Methods for Blood Glucose Concentration, Springer, 2016, pp. 183–209.

[95] L. Ljung, Prediction error estimation methods, Circuits, Systems and Signal Processing 21 (1) (2002) 11–21. `doi:10.1007/BF01211648`.

[96] B. Efron, Bayesians, frequentists, and scientists, Journal of the American Statistical Association 100 (469) (2005) 1–5.

[97] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, The journal of chemical physics 21 (6) (1953) 1087–1092.

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

[99] J. Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 236 (767) (1937) 333–380.

[100] S. Kullback, A Note on Neyman's Theory of Statistical Estimation, The Annals of Mathematical Statistics 10 (4) (1939) 388–390.

[101] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical recipes in C++, Vol. 3, Cambridge University Press, 2007.

[102] R. E. Kass, B. P. Carlin, A. Gelman, R. M. Neal, Markov chain monte carlo in practice: a roundtable discussion, The American Statistician 52 (2) (1998) 93–100.

[103] M. J. Bayarri, J. O. Berger, The interplay of bayesian and frequentist analysis, Statistical Science (2004) 58–80.

[104] A. Chong, K. P. Lam, A comparison of mcmc algorithms for the bayesian calibration of building energy models, in: Proceedings of the 15th IBPSA Building Simulation Conference, Vol. 4, 2017.

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

[106] D. J. MacKay, D. J. Mac Kay, Information theory, inference and learning algorithms, Cambridge university press, 2003.

[107] D. W. U. Perera, Mathematical models for real-time estimation of space heating in buildings, Ph.D. thesis, University College of Southeast Norway (2016).

[108] D. W. U. Perera, N.-O. Skeie, Estimation of the heating time of small-scale buildings using dynamic models, Buildings 6 (1). `doi:10.3390/buildings6010010`.

[109] D. W. U. Perera, A. Perera, C. Pfeiffer, N.-O. Skeie, Structural observability analysis and ekf based parameter estimation of building heating models, Modeling, Identification and Control 37 (3) (2016) 171–180. `doi:10.4173/mic.2016.3.3`.

[110] J. W. Nilsson, Electric circuits, Pearson Education India, 2008.

[111] L. W. Nagel, SPICE2: A computer program to simulate semiconductor circuits, Ph. D. dissertation, University of California at Berkeley.

[112] H. H. Hoos, T. Stützle, Stochastic local search: Foundations and applications, Elsevier, 2004.

[113] J. P. Bentley, Principles of measurement systems, Pearson education, 2005.

[114] T. Maiwald, J. Timmer, Dynamical modeling and multi-experiment fitting with PottersWheel, Bioinformatics 24 (18) (2008) 2037–2043.

[115] S. Brooks, A. Gelman, G. Jones, X.-L. Meng, Handbook of markov chain monte carlo, CRC press, 2011.

[116] J. Kruschke, Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, Academic Press, 2014.

[117] H. Lodhi, D. Gilbert, Bootstrapping parameter estimation in dynamic systems, in: International Conference on Discovery Science, Springer, 2011, pp. 194–208.

[118] H. R. Kunsch, The jackknife and the bootstrap for general stationary observations, The Annals of Statistics (1989) 1217–1241.

[119] S. H. Kim, S. Ju, C. H. Ji, S. J. Lee, Equivalent circuit model of an impact-based piezoelectric energy harvester, Journal of Physics: Conference Series 557 (2014) 012094. `doi:10.1088/1742-6596/557/1/012094`.
URL `https://doi.org/10.1088%2F1742-6596%2F557%2F1%2F012094`

Bibliography

[120] T. S. Schei, A finite-difference method for linearization in nonlinear estimation algorithms, Automatica 33 (11) (1997) 2053 – 2058. `doi:10.1016/S0005-1098(97)00127-1`.

[121] C. S. Ferrero, Q. Chai, M. Dueñas Díez, S. H. Amrani, B. Lie, Systematic analysis of parameter identifiability for improved fitting of a biological wastewater model to experimental data, Modeling, Identification and Control 27 (4) (2006) 219.

[122] S. A. Murphy, A. W. Van der Vaart, On profile likelihood, Journal of the American Statistical Association 95 (450) (2000) 449–465.

[123] G.-O. Kaasa, Ø. N. Stamnes, O. M. Aamo, L. S. Imsland, et al., Simplified hydraulics model used for intelligent estimation of downhole pressure for a managed-pressure-drilling control system, SPE Drilling & Completion 27 (01) (2012) 127–138.
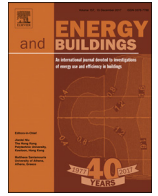
Part II

Scientific Publications

Article A

Parameter estimation for grey-box models of building thermal behaviour

Authors O. M. Brastein, D.W.U Perera, C. Pfeiffer and N.-O. Skeie

Parameter estimation for grey-box models of building thermal behaviour☆

O.M. Brastein*, D.W.U. Perera, C. Pfeifer, N.-O. Skeie

Department of Electrical engineering, Information Technology and Cybernetics, University Collage of Southeast Norway, 3918 Porsgrunn, Norway

ABSTRACT

Good models for building thermal behaviour are an important part of developing building energy management systems that are capable of reducing energy consumption for space heating through model predictive control. A popular approach to modelling the temperature variations of buildings is grey-box models based on lumped parameter thermal networks. By creating simplified models and calibrating their parameters from measurement data, the resulting model is both accurate and shows good generalisation capabilities. Often, parameters of such models are assumed to be a combination of different physical attributes of the building, hence they have some physical interpretation. In this paper, we investigate the dispersion of parameter estimates by use of randomisation. We show that there is significant dispersion in the parameter estimates when using randomised initial conditions for a numerical optimisation algorithm. Further, we claim that in order to assign a physical interpretation to grey-box model parameters, we require the estimated parameters to converge independently of the initial conditions and different datasets. Despite the dispersion of estimated parameters, the prediction capability of calibrated grey-box models is demonstrated by validating the models on independent data. This shows that the models are usable in a model predictive control system.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

A large part of the world's energy production is used for heating and cooling buildings. The fraction of total energy production consumed for utilities in commercial and residential buildings has been estimated at 32% by the International Energy Agency (IEA), according to [1]. Even though modern building techniques are able to reduce the energy used for heating, the renewal rate of buildings is low. Berthou et al. [2] reports renewal rates of 1% per year in France. This illustrates the need for good building energy management systems (BEMS) in existing buildings as well.

A model predictive control (MPC) system is an attractive solution for use in a BEMS. Models of building thermal behaviour can be used to predict the heating and cooling time of a building. In a MPC system, a model is used to simulate the system ahead in time in order to find a sequence of inputs that controls the system to the desired state. In a BEMS, the use of MPC will allow for improved tracking of the temperature setpoint as well as minimization the energy consumption [2,3]. Predictions of future system in-

puts are readily available from weather forecasts, which helps to facilitate the use of MPC.

There have been several publications studying the use of MPC for building thermal control. In [4] the authors use both active heating and passive solar blinds to control indoor air temperature. The paper also gives a thorough introduction to the various MPC control methods, such as deterministic and stochastic MPC. In [5] a complete building model is developed as a set of layered models and used in an MPC. The authors report an energy saving of 63% in thermal energy and 29% in HVAC electric energy, for a four-month test period. These examples show the potential benefits of using MPC for building thermal control. They also show the importance of a good prediction model for MPC to be feasible.

There are a number of different modelling approaches that can be used to model the thermal behaviour of a building in an MPC system [6]. In Perera et al. [1], a white-box model based on mass and energy balance is derived and calibrated for specific buildings. This type of model gives a set of ordinary and/or partial differential equations (ODE/PDE) that must be discretised and solved. For complex models, a large number of parameters are required that can be difficult to identify. Another approach to modelling is the use of black-box models, which relies solely on measurement data without any prior knowledge of the building, e.g. ARMAX [7,8] or PLS-R [9,10] models. These types of models show high prediction accu-

racy, but do not usually allow the application of physical knowledge to define the model. This approach also produces models with low generalisation between different buildings, which makes building-to-building comparisons of models difficult [11]. Comparing the thermal behaviour of buildings can be of interest for the purposes of energy consumption classification.

Another approach to modelling thermal behaviour of buildings is the use of grey-box models [2,3,12,13]. A grey-box model is based on a simplified structure derived from a cognitive understanding of the physics involved. For the heating of buildings, the model structures may consist of thermal networks [14], i.e. resistor–capacitor circuit equivalent lumped parameter models. Rather than deriving the full model as in [1], the simplified model structure is developed from an understanding of the heat transfers involved in a building, which provides directly a reduced order model. This process can be referred to as 'cognitive' model development [14]. The parameters of such models are lumped parameters, i.e. each parameter represents a combination of multiple physical quantities. Such parameters must be identified from measurement data, since they are generally difficult to compute based on technical building specifications. A grey-box model therefore uses a combination of the white- and black-box approaches [15].

It is often assumed that the parameters of such models can be assigned physical meaning. The identified parameters are compared to the physical properties of the building [16,17]. For interpretation of model parameters to be justified, we suggest that the results of the parameter estimation process must show a low degree of dispersion, e.g. be independent of the initial guess parameter vector for the estimation algorithms. Estimation of parameters is required to give similar results when using different datasets from the same building.

The estimation of parameters requires the measurement data to contain enough dynamic information about the system to accurately calibrate the model [16,18–20]. Since the subject of this work is physical buildings, the experimental design is challenging. The outdoor weather conditions acts as a model input, particularly the outdoor temperature. Further, it is of interest to estimate the parameters under realistic conditions for an occupied building. Hence the choice of excitation of the system is limited. Lack of dynamic information in the data is known to give problems with practical identifiably [19].

Since all the parameters of a grey-box model must be estimated, an additional challenge with calibrating grey-box model parameters is over-parameterisation [16]. This is known to give non-convergent parameter estimates, since an over-parameterised model has undetermined optimal parameters, i.e. infinitely many solutions exist.

While challenges caused by practical identifiably and/or over-parameterisation may give reason to question the physical interpretation of the estimated parameters the models may still be usable in an MPC. In this work, the dispersion of parameter estimates under different experimental conditions is investigated using multiple sets of experimental data from a real building. Further, calibrated models are validated on independent data to show that they are capable of predicating the thermal behaviour of the test building, hence rendering them usable in an MPC system.

2. Model, methods and measurements

A common approach to parameter estimation is the use of numerical optimisation [19], either directly [2] or in the form of a maximum likelihood (ML) method [17,21]. When using numerical optimisation, it is of interest to investigate the dispersion in the estimated optimal parameters under different experimental condi-
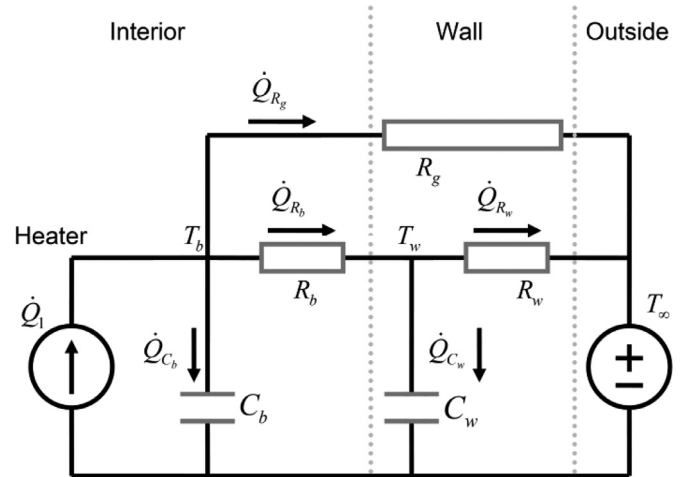

Fig. 1. The R3C2 thermal network model.

tions. In particular, it is interesting to study if the initial guess for the optimisation affects the estimated parameters.

2.1. Model and parameters

The model used in this paper is a thermal network model of a building [3,14,16,17,20,22], presented using an electrical circuit equivalent model. Thermal resistance is modelled as resistors and thermal capacitance as capacitors. The resulting model is a circuit where the temperature is used as the driving potential, and the flow through the circuit is the heat flow. This approach has been used in a number of published papers on modelling building thermal behaviour, e.g. [3,13].

The focus of this paper is estimation of the model parameters. For simplicity, only one model is investigated, and the model structure is chosen as a minimalistic representation of the experimental building from which the calibration data is collected. The model is shown in Fig. 1. This model is similar to the R3C2 model used in [2], but the resistance for ventilation is removed since there is no ventilation system installed in the test building.

The model consists of two states T_b and T_w, which correspond to the interior temperature of the building and the wall temperature respectively. Wall temperature is measured on the inner surface of the wall. For each state there is an associated capacitance, C_b and C_w. These capacitances represent the building's ability to store thermal energy in the interior and the building envelope, e.g. walls, floor and ceiling. The remaining three model components are resistances. R_b represents the thermal resistance between the building interior and the wall. R_w is the resistance to heat flow through the wall, i.e. between the state T_w and the outside temperature. The third resistance R_g represents the resistance to heat flow through the parts of the building envelope that are not included in the state T_w, such as windows and the door. The driving forces of the system are $\dot{Q}$ and T_∞, where $\dot{Q}$ is a heat flow source, e.g. an electric heater. The outside temperature is modelled as a potential source T_∞.

Deriving equations from a thermal network model can be done with, for example, Kirchhoff's node potential law [23,24]. Each state in the circuit, T_b and T_w, is assigned to a circuit node and the flow into and out of each node is balanced. The model can be written in state-space form as a set of ordinary differential equations (ODEs) [18]:

$$\frac{dT_b}{dt} = -\left(\frac{1}{C_b R_b} + \frac{1}{C_b R_g}\right)T_b + \left(\frac{1}{C_b R_b}\right)T_w + \left(\frac{1}{C_b}\right)\dot{Q}_1 + \left(\frac{1}{C_b R_g}\right)T_\infty \quad (1)$$

$$\frac{dT_w}{dt} = \left(\frac{1}{C_w R_b}\right)T_b - \left(\frac{1}{C_w R_w}\right)T_w + \left(\frac{1}{C_w R_w}\right)T_\infty \tag{2}$$

The parameter vector is then defined as:

$$\theta = \begin{bmatrix} R_g & R_b & R_w & C_b & C_w \end{bmatrix}^T \tag{3}$$

The output from the model is the states themselves, and both states are directly measurable. The measurement equations are therefore found by including the measurement noise terms v_b and v_w with the system states:

$$\begin{aligned} y_{T_b} &= T_b + v_b \\ y_{T_w} &= T_w + v_w \end{aligned} \tag{4}$$

The measurement noise v_b and v_w is assumed zero mean Gaussian with standard deviation r_b and r_w:

$$v_b \sim N(0, r_b) \quad , \quad v_w \sim N(0, r_w) \tag{5}$$

Both measurements are temperatures from the same Data Acquisition (DAQ) system, hence it is reasonable to expect that they have similar noise characteristics. This is verified by computing the standard deviation for both measurements over a range of measurements where the temperatures are approximately constant.

2.2. Deterministic approach to parameter estimation

The state transition equations in the model are an approximation of the real system. Hence, there will be some unknown influence from modelling errors in the simulations. The state transition equations in Eqs. (1) and (2) are written as ODE's. These could be extended to a set of stochastic differential equations (SDE) by the addition of a process noise term w [20]. The probability density function (pdf) of the process noise term w is not easily obtained, however it is often assumed Gaussian. This assumption can be checked during model validation [20,21] by use of e.g. autocorrelation to verify that the residuals are indeed white noise This method can confirm the assumption of Gaussian process noise [20]. While the use of SDE's is a more statistically solid approach [16,17,20,21], it is also possible to treat the estimation problem as deterministic without the process noise term w [6,25]. The model is then a set of deterministic ODE's. This reduces the estimation of model parameters to a least squares curve fitting problem which can be solved directly by numerical optimization [2]. In this work, the deterministic curve fitting approach is used.

2.3. Parameter ranges and nominal values

A nominal parameter vector is used as a starting point for the estimation methods. These parameter values are based on trial and error experiments, together with prior knowledge of the approximate range where reasonable values may be obtained. The physical insight required as a starting point for these trial and error experiments is limited to the approximate order of magnitude of the parameters. It is assumed that this can be obtained for most practical buildings. It is not required that the nominal values themselves give a good prediction model for the building, only that they are approximately in the correct range. They are mainly used as normalisation constants, such that parameter estimation can be performed in unit scale, and to restrict the search space to a region of interest where reasonable parameter values may be obtained.

Table 1 gives a summary of the nominal parameter values with min./max. ranges used for the R3C2 model. The nominal values are chosen such that any region of interest in the parameter space is contained within +/−70% of each nominal value. Hence $\theta_{\min} = 0.3 \times \theta_0$ and $\theta_{\max} = 1.7 \times \theta_0$. In the following, discussions on the estimation of parameters and the shape of the objective function in parameter space are valid only within the ranges specified in Table 1.

Table 1
Nominal parameter values and min./max. range.

	R_g [K/W]	R_b [K/W]	R_w [K/W]	C_b [J/K]	C_w [J/K]
Nom. value (θ_0)	0.160	0.060	0.100	1200 k	1200 k
Min. ($\theta_{\min}$)	0.048	0.018	0.03	360 k	360 k
Max. ($\theta_{\max}$)	0.272	0.102	0.170	2040 k	2040 k

2.4. Numerical optimisation

The subject of numerical optimisation used for parameter identification is covered in literature, e.g. [2,19,26]. Optimisation algorithms are used to find a minimum point of an objective function. In parameter estimation, this objective function is typically the model fit, i.e. the square error over a set of reference data compared to model simulations. In this work, the simulation error is computed over the whole calibration period, rather than using the one-step ahead prediction errors traditionally used in statistics [8,20]. This results in a least squares curve fitting approach to parameter estimation, as discussed in Section 2.2. This gives the objective function as the mean square error (MSE) between simulated and measured temperatures computed over the whole dataset. A standard quadratic norm [18] is chosen as the error function, i.e.:

$$J = \sum_{i=1}^{N}\sum_{k=1}^{nx} e_k^2 = \sum_{i=1}^{N}\sum_{k=1}^{nx}\left(T_k^i - T_k^{i,ref}\right)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{J}{nx \cdot N}} \tag{7}$$

Here, nx is the number of states (temperatures) and N is the number of samples in the dataset. The sum of squared error of each of the two temperature states is added together and the errors of all states are weighted equally. An alternative approach is to weight the errors by their uncertainty, e.g. the covariance of the measurements, as is typically done in the statistical approach to parameter estimation [20]. Since the temperatures are measured by the same DAQ system it is reasonable to assume similar uncertainty in both measurements and hence use equal weights on the error for both states. All samples in the calibration data Z^N are assumed to have equal uncertainty, such that the weighting of the errors is also uniform in time.

To simplify the evaluation of results, the objective score J is divided by the number of samples and the square root taken, thus giving the root mean square error (RMSE). This quantity has the same unit (°C) as the states. The optimisation algorithm works directly on the sum of square error objective J.

The optimisation algorithm of choice in this paper is the Constrained Optimization BY Linear Approximation (COBYLA), which was developed by M.J.D. Powell [26]. The implementation is taken from the Accord.Net project [27]. There are two important settings of the algorithm, RhoEnd and MaxIttr. The first controls the accuracy of the optimised variables, while the second is a loop break to ensure the algorithm completes even if no optimal solution is found. In this project, RhoEnd is set to 1e-4 and MaxIttr is 1e4. This allows COBYLA to find the solution with sufficient accuracy. In all following cases, the algorithm completes without reaching the maximum number of iterations.

2.5. Experimental setup and data

The building of interest in this paper is an experimental setup built in 2014 at the campus of University College of Southeast Norway, in Porsgrunn [28].

Fig. 2 shows a view of the test building from the outside, together with the floor plan. This building has support structures
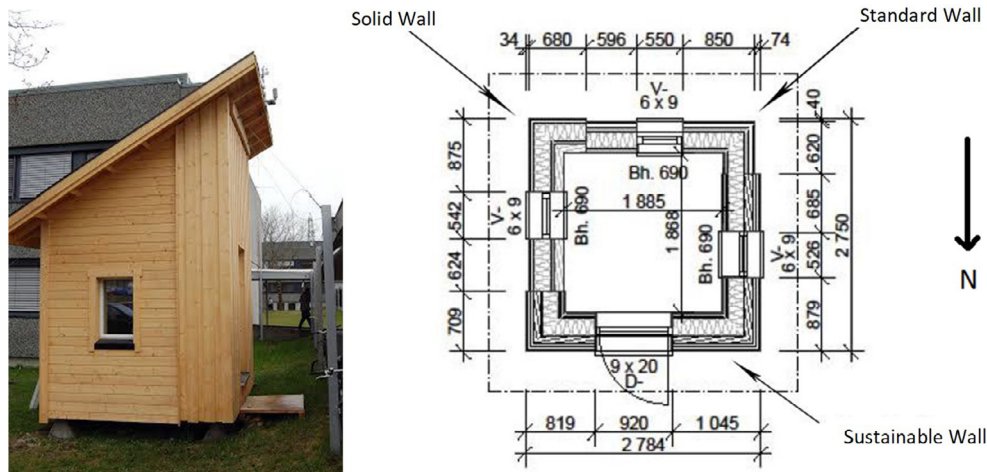
Fig. 2. A view of the experimental building with floorplan.

made out of concrete so it is not in contact with the ground. The inner volume is approximately 9.4 m³ and sealed such that neither natural ventilation nor mechanical ventilation is provided. The building has three small windows of each 60×90 cm², one in each of the south, east and west directions, and a door of 90×120 cm² in the north direction. Three buildings surrounding the experimental building limit the solar irradiation through the windows.

The building envelope is constructed using layers of diverse materials, including wooden cladding, glass wool, air fill, polyethylene vapour barriers, wood, cement chipboard, particleboard and cardboard. There are three different types of walls each having unique construction by compounding different materials stated above. Roof and floor of the building also have different composition in each.

The experimental building has an electrical heater of 375 W consisting of a thermostat controller, a measurement system, and a logging computer that uses approximately 100 W. The measurement system consists of sensors for measuring the inside and outside temperatures and humidity, air pressures, rainfall, and wind speed and wind direction, and total power usage. The power usage is logged every 10 s, the temperatures and humidity are logged every 10 min, and the weather data is logged every 30 min. All data is stored in CSV (Comma Separated Values) files. The temperature sensors are mostly of the silicon type (TMP 36) with an accuracy of $+/-1$ °C. Data was collected using multiple data from these CSV files, and pre-processed. During the pre-processing steps the data was filtered, combined and re-sampled by linear interpolation. This gives a combined dataset where all the variables of Z^N are known at the same time instants t_i, regardless of the initial sampling times for the different CSV files. In the final dataset, a sampling time of d$t = 10$ min is used.

Measurement data was collected from the building during November and December 2015. Since buildings are exposed to weather, experimental design options are limited. This poses a challenge in ensuring that the calibration data contains sufficient dynamic information for parameter estimation. If buildings are occupied, this poses further restrictions on the experimental design. The data used here is derived from an empty building, but under realistic experimental conditions, i.e. exposed to weather and other unmeasured disturbances.

Since the window area exposed to solar irradiation is limited and the data was collected during the winter in Norway, heat gain from solar irradiation is assumed negligible. Hence, in order to limit the model complexity, the solar irradiation is not included in the dataset or the model.

A list of four datasets, including identifying names, is given in Table 2. Estimation of parameters, i.e. fitting a model structure to a particular physical system, requires information about that system in the form of a set of measurements Z^N [18]. This set contains both reference data in the form of system outputs y and information about the excitations for the system in the form of system inputs u, i.e.:

$$Z^N = \{y_1, u_1, y_2, u_2, \ ..., \ y_N, u_N\} \tag{8}$$

If the states are directly measurable, as is the case with building temperatures, the outputs are simply equal to the states with the addition of measurement noise. The measured states y_i in Z^N are temperatures T_b and T_w. The inputs u_i consist of the outside temperature T_∞ and power consumption $\dot{Q}$ of the building.

The contents of each of the four datasets are plotted in Fig. 3. These datasets were generated by introducing arbitrarily chosen steps in the thermostat setting of the electric heater in order to simulate realistic variations in setpoint for a building in use. While this approach may be inferior to, for instance, a pseudo random binary sequence (PRBS) [22], it is more realistic since it mimics temperature changes of a building in use. Changing temperature setpoints from a low to a high setting (approximate step sizes of 10 °C–20 °C are used here) takes a significant time, typically around 12−48 hours. This is reflected in the slow response to the step in supplied power, shown in the plots of data.

A constant minimum power consumption of about 100 W is observed in Fig. 3. This power is consumed by computer equipment used for data logging in the test facility. When the on/off thermostat control turns the heater on, an additional 350 W of power is drawn by the heater.

When the heater is regulating the temperature at a setpoint requiring less than full capacity (<100% on-time), there are spikes in the power consumption data caused by the thermostat switching the heater on and off. These spikes are present in the November data, but not in the December data. In November, the heater was set at a maintainable temperature setting of approximately 25 °C, i.e. the system is under on/off control. In December, the heater setpoint was turned to maximum, which is higher than the achievable temperature at full power under the given experimental conditions. Hence, the heater was constantly on for long periods of time.

For the parameter estimation, all the supplied power is included in $\dot{Q}$. Hence the input to the model is taken directly from the data as shown here. This is a simplification, since some power is consumed by other equipment in the building. However, it is assumed that all the supplied energy is converted to heat, either by

Table 2
List of datasets.

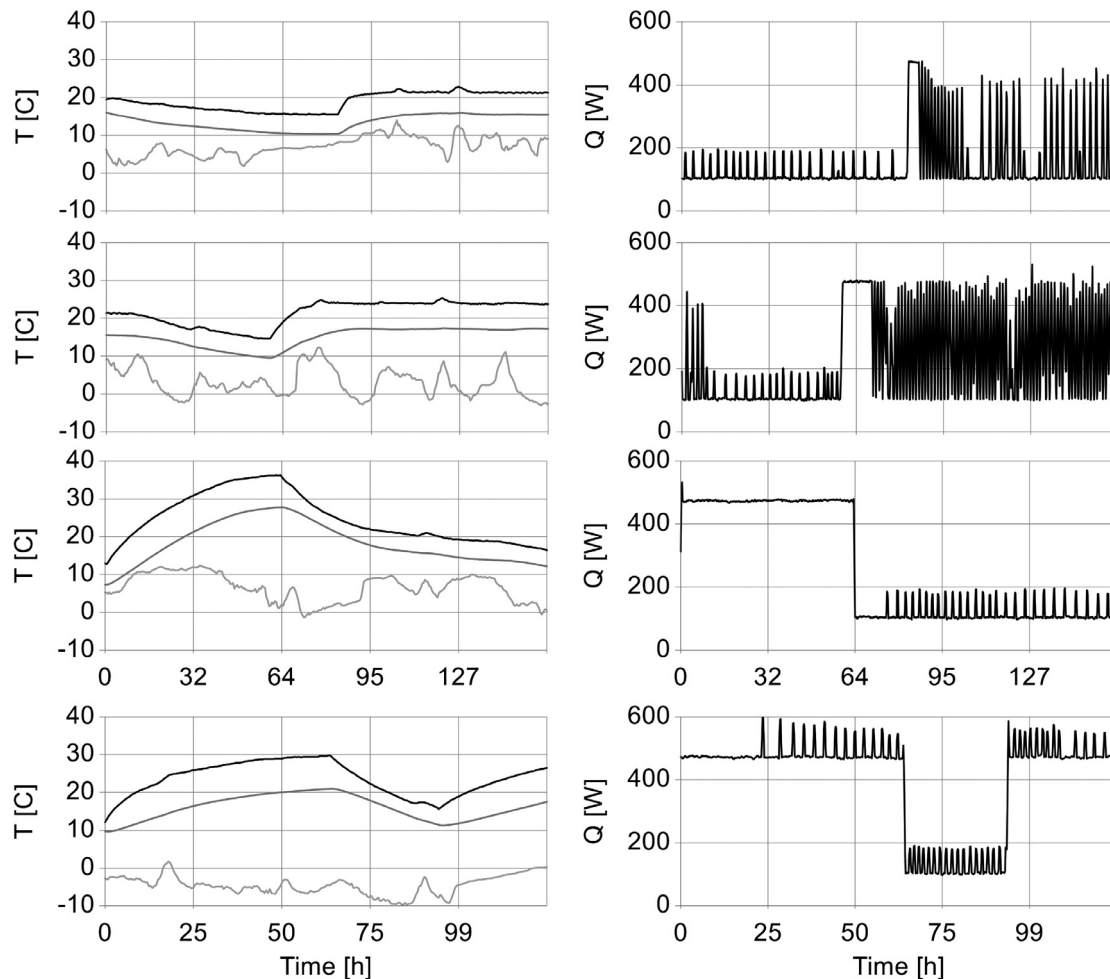| Dataset name	Start time	End time	Length [h]	N
Nov1	02.11.2015 17.18.00	10.11.2015 00.08.00	175	1050
Nov2	10.11.2015 00.18.00	17.11.2015 07.08.00	175	1050
Dec1	04.12.2015 18.00.00	11.12.2015 08.50.00	159	954
Dec2	11.12.2015 19.40.00	17.12.2015 00.00.00	124.5	747

Fig. 3. Plots of all datasets. On the left side, temperatures for building interior T_b (black), building wall T_w (grey) and outdoor temperature T_∞ (light grey) are shown. On the right side, the power consumed in the building is plotted (black). The datasets Nov1, Nov2, Dec1 and Dec2 are plotted from top to bottom respectively.

the electric heater or the other equipment. For simplicity, the total heat generation from all appliances is denoted as a single heater.

3. Results

The R3C2 model is based on the physical structure of a building and the parameters are estimated using measurement data. The thermal parameters of buildings are assumed to be time invariant. Hence, it is expected that there exists a single, true, parameter vector for the grey-box model determined by the physical properties of the building. Convergence of parameter estimates towards such a parameter vector can be considered a requirement for physical interpretation of the final model. The results presented here are computed mainly from the *Nov1* dataset.

The denomination *degrees of freedom* (DOF) is used to describe the number of free parameters in the estimation problem. The R3C2 model has five parameters, constituting five DOF.

3.1. Maximum degrees of freedom, five estimated parameters

The first method used to investigate the parameter space of the model is a simple Monte Carlo approach. In the following results, $M = 50,000$ simulations of the model are executed. Model parameters are randomly and independently drawn from the range $\theta_{\min} \leq \theta \leq \theta_{\max}$, given in Table 1. This is similar to a simple random search, except that we are interested in the collection of all results, not just the one with the lowest RMSE. The resulting plots consist of points, one for each simulation, which gives a view of the shape of the objective function J in parameter space. The abscissa of the plots represents parameter values and the ordinate axis represents the resulting root mean square error (RMSE) for each simulation.

In this section, all five parameters of the R3C2 model are free, i.e. they are estimated as opposed to kept constant.

As shown in Fig. 4, most of the randomly drawn parameter vectors are not optimal since the RMSE is higher than the minimum. However, some solutions fall close to the minimum RMSE of
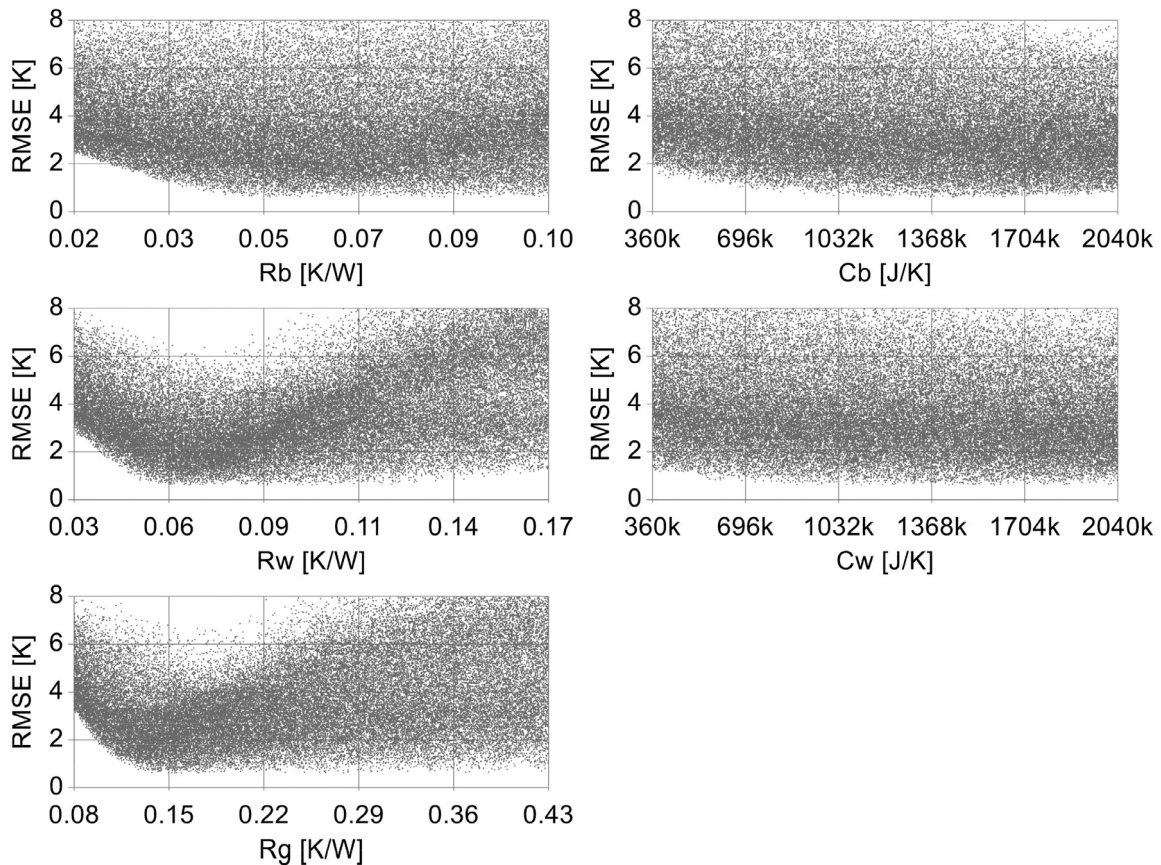
Fig. 4. Monte Carlo simulations with 5 degrees of freedom. This figure shows the resulting scatter plots for the parameters.

0.591 °C for the R3C2 model and Nov1 dataset. When all five parameters are estimated, the objective function is flat in the minimum RMSE region. Since there is no clear minimum, the estimated parameter from an optimisation algorithm will not converge to a single parameter vector but rather produce vectors in an optimal range, all with the same minimum RMSE.

The next method is also based on Monte Carlo (MC) simulations. Randomly drawn parameters are used as the initial guess of a numerical optimisation algorithm (COBYLA). The initial guess and the subsequent parameter estimate are plotted for all iterations of the optimisation algorithm. The points are interconnected by a line in order to identify which particular initial guess corresponds to each solution. Each plot consists of 50 iterations, i.e. repeated randomised initial guesses followed by execution of the optimisation algorithm.

As shown in Fig. 5, the optimisation algorithm finds optimal solutions, i.e. RMSE $\sim$ 0.6 °C, from all 50 starting points, but the resulting optimal solution varies significantly with the initial guess. The parameter estimates do not converge. Based on the plots in Fig. 4, this is expected. Fig. 5 further shows that the RMSE objective function is flat around the minima, since many equally good solutions exist for a large range of parameter values.

3.2. Reduced degrees of freedom, 4 free parameter

As discussed, we expect the estimated parameters to converge towards a true, physically determined, parameter vector. Since the results so far indicate that this is not the case, an explanation for the dispersion in parameter estimates is needed.

Over-parameterization [16], i.e. too many degrees of freedom (DOF) in the estimation problem, is one plausible explanation for the observed flatness of the objective function. The DOF is reduced

Table 3
Building specification according to manufacturer of windows and door.

	U [W/m²K]	A [m²]	UA [W/K]	R [K/W]
Door	1.2	1.76	2.1	0.48
Windows	1.3	1.57	2.0	0.50
Total	–	–	4.1	0.24

by fixing one parameter at a constant value, thus reducing the number of free parameters to estimate. Reduction in freedom is only applied to the estimation problem, by the reduction of free variables, and does not affect the model structure. All five parameters are used in the simulations and the model remains the same.

The parameter chosen to be kept as constant is R_g. This parameter represents the thermal resistance of windows and doors, i.e. the part of the building envelope directly exposed to both interior and exterior temperatures. Because R_g represents doors and windows, which usually have known 'UA' values [1], it is assumed that R_g would be the easiest to compute based on building technical specifications. UA values are computed as the product of U and A, where U is the reciprocal of thermal resistance per area and A is the area. Hence, knowing R or U for all windows and doors, as well as their area A allows for computation of $R_g = 1/(U_g A_g)$.

As shown in Table 3, the specifications give a theoretical R_g of 0.24. However, this includes only one door and three windows. Comparing the R3C2 model to the building, R_g is also expected to include any other element of the building that gives a direct influence between outside temperature and indoor air temperature without affecting the wall temperature. Hence, the theoretical R_g can be considered an upper bound of the true R_g, as any contribution by remaining building elements can only lower R_g. Note there-
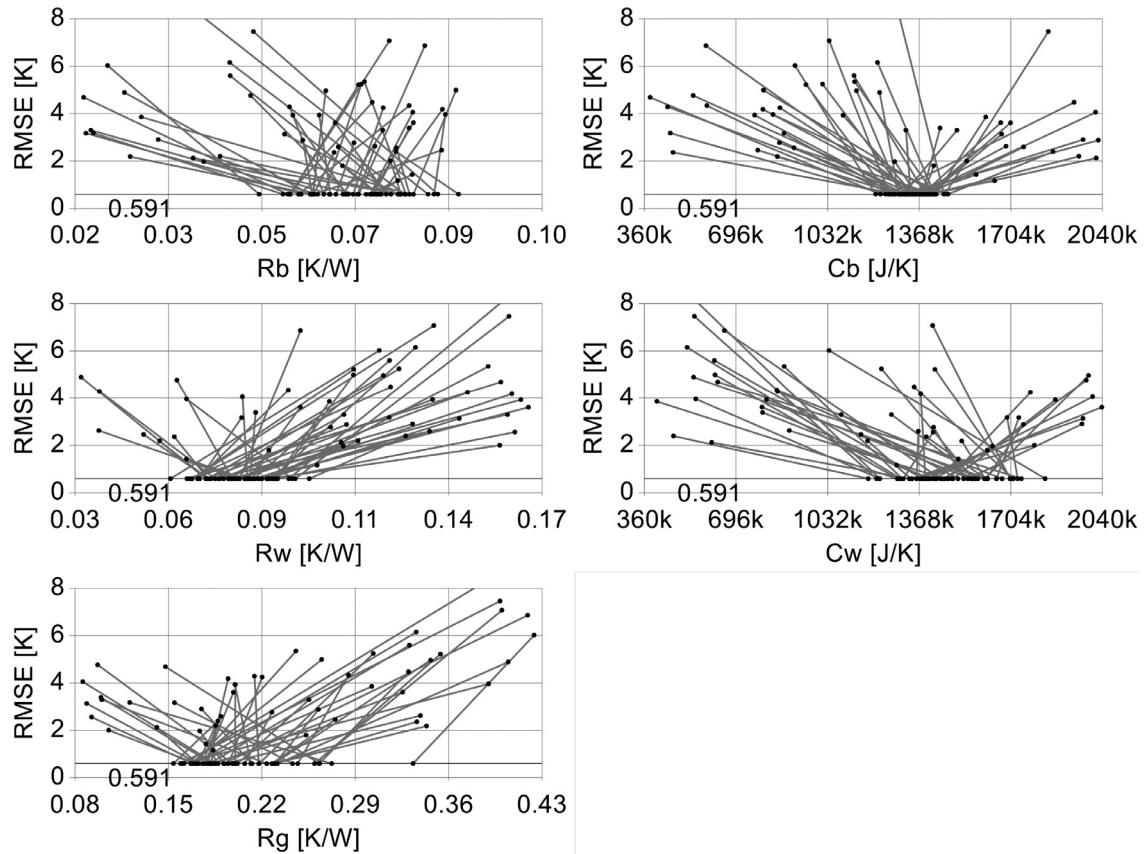
Fig. 5. Randomised initial guess and estimated solution is plotted together. This figure shows the results of applying the bespoke method to the R3C2 model simulated in the Nov1 dataset.
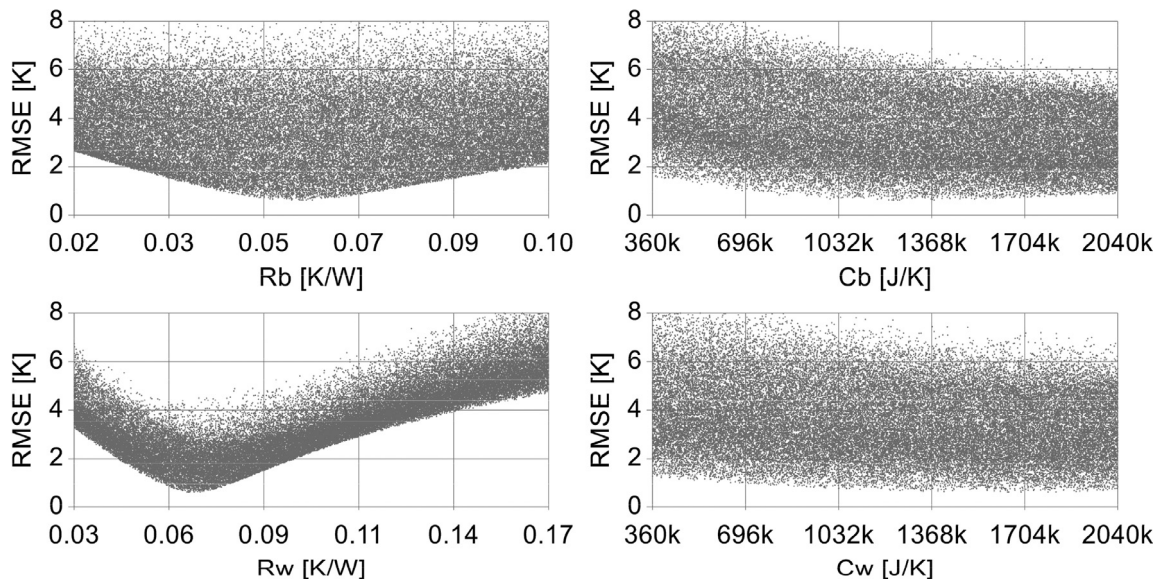

Fig. 6. Scatter plots of MC simulation in parameter space. This figure shows the results of applying MC simulation of the parameter space to the R3C2 model simulated on the Nov1 dataset, with a fixed parameter $R_g = 0.24$ [K/W].

fore that it is considered acceptable that the theoretical R_g is in the upper region of the nominal range defined in Table 1.

First, the parameter space for the reduced DOF case is observed by using MC simulation of parameter space.

As shown in Fig. 6, the MC simulations of the parameter space now show a significantly different shape of the objective function, which now has a well defined minimum. Subsequently, optimisa-

tion is now expected to converge towards a single parameter vector.

Comparing Fig. 7 with Fig. 5, it is evident that the dispersion in estimated parameters is significantly decreased. Based on the MC simulations of parameter space shown in Fig. 6, it is reasonable to conclude that the objective function has more pronounced minima when one DOF is eliminated from the estimation problem. With a
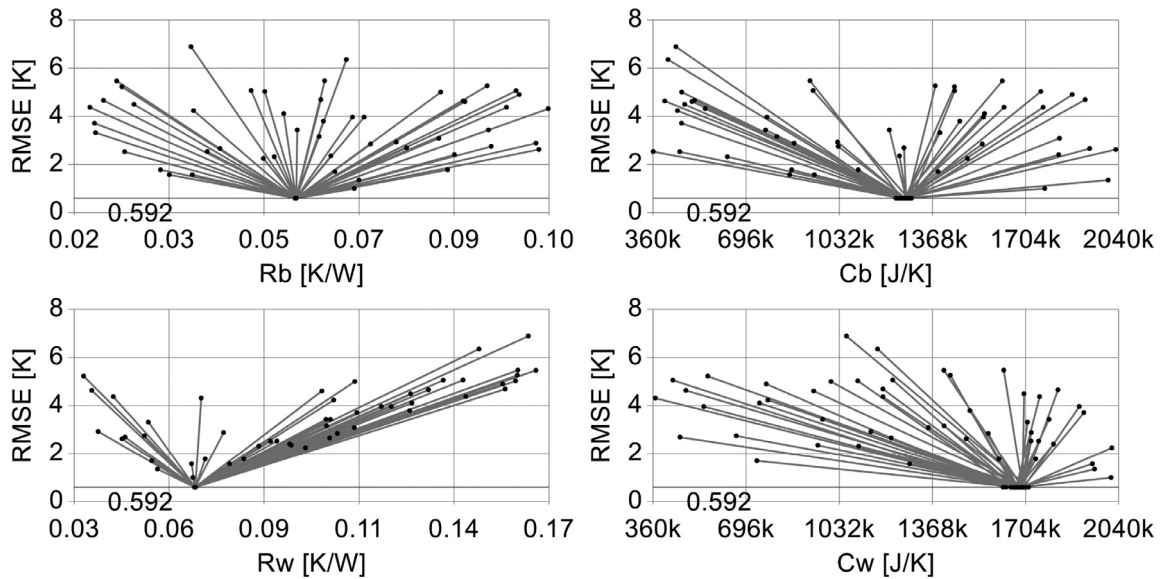
Fig. 7. Randomised initial guess and optimal solution is plotted together. This figure shows the results of applying the bespoke method to the R3C2 model simulated on the Nov1 dataset. With reduced freedom in the estimation problem, the trajectories of the optimisation now show a distinct convergent pattern, independent of the initial random guess.

Table 4
Sample standard deviation of parameter estimates, as a percentage of nominal value.

	R_g [%]	R_b [%]	R_w [%]	C_b [%]	C_w [%]
5 DOF	19.23	14.43	10.05	5.093	12.18
4 DOF	–	0.066	0.079	1.230	2.208

distinctly convex objective function, the optimiser is able to find the same optimal solution independent of initial guess.

Multimodal objective functions are a typical challenge of numerical optimisation. This has not been discussed here, since the plots from the MC method are used to show that the objective function is either flat or unimodal in the region of interest.

3.3. Quantitative comparison of results from DOF reduction

In order to quantitatively compare the cases presented in Sections 3.1 and 3.2, the standard deviation (SD) of estimated parameter values is computed.

Table 4 lists the SD as a percentage of the nominal value, as a measure of the dispersion of estimated parameters from both the cases with five and four DOF. By reducing the number of free parameters, the dispersion of the parameter estimates is significantly reduced, in particular for the resistance parameters. Slightly more variation is observed in the two capacitance parameters.

3.4. Comparing results from multiple datasets

Until now, all results are taken from a single dataset, Nov1. As demonstrated, too many DOF will lead to a flat objective function, which in turn gives a large dispersion in estimated parameters. Next, it is of interest to introduce more data to the parameter estimation, namely the datasets Nov2, Dec1 and Dec2. Thermal building behaviour parameters are assumed to change slowly with time, unless modifications to the building structure are introduced. For the datasets presented in Section 2.5, no such modifications occurred. Hence, it is reasonable to assume that the parameter estimation methods will give similar results for the four datasets.

Fig. 8 shows the plots from randomised initial guess to estimated parameter for the parameter R_b estimated on all four

datasets using the four DOF model. As previously demonstrated for dataset Nov1, the parameter estimates converge independently of the initial guess per iteration of the method. This also holds for the other three datasets. However, the parameter value that gives minimum RMSE is different for each dataset, i.e. the solution for each dataset is not the same parameter vector.

Fig. 9 shows the results for the parameter C_b. The plots show that the estimated parameter values for C_b also vary between the four datasets, same as for R_b, although this variation is smaller for C_b than R_b. Similar results are obtained for the other two parameters. Next, it is useful to quantitatively compare the estimated parameter values and their corresponding variance, for all four datasets.

Estimated values for each parameter for all datasets are given in Table 5 and plotted in Fig. 10. Columns RMSE4 and RMSE5 give the root mean square error (RMSE) of the simulations using estimated parameter vectors, from four and five DOF cases respectively. There are significant variations in the estimated parameters when comparing multiple datasets, e.g. R_b for Nov1 is twice as large as for Dec1 and Dec2. It is unlikely that the thermal resistance between room interior temperature and wall temperature is halved in just four weeks. As such, these variations are probably not caused by physical effects. A plausible explanation is insufficient dynamic information content in the data. As discussed, this can lead to problems with practical identifiably [19].

An important observation from Table 5 is the similarity between RMSE values for the five and four DOF cases, as shown in the two last columns. While fixing R_g to a constant value – thereby removing it from the parameter estimation problem – decreases the dispersion of estimated parameters, this does not affect the final RMSE of the model with estimated parameters.

The standard deviation of parameter estimates, as a percentage of nominal values, is given in Table 6. As shown, the dispersion of parameter estimates is similar for all four datasets, even though the estimated parameter values are different. These results could indicate that the spread of the parameter estimates is problem specific, independent of the datasets, while the actual value of each parameter depends on the calibration data.
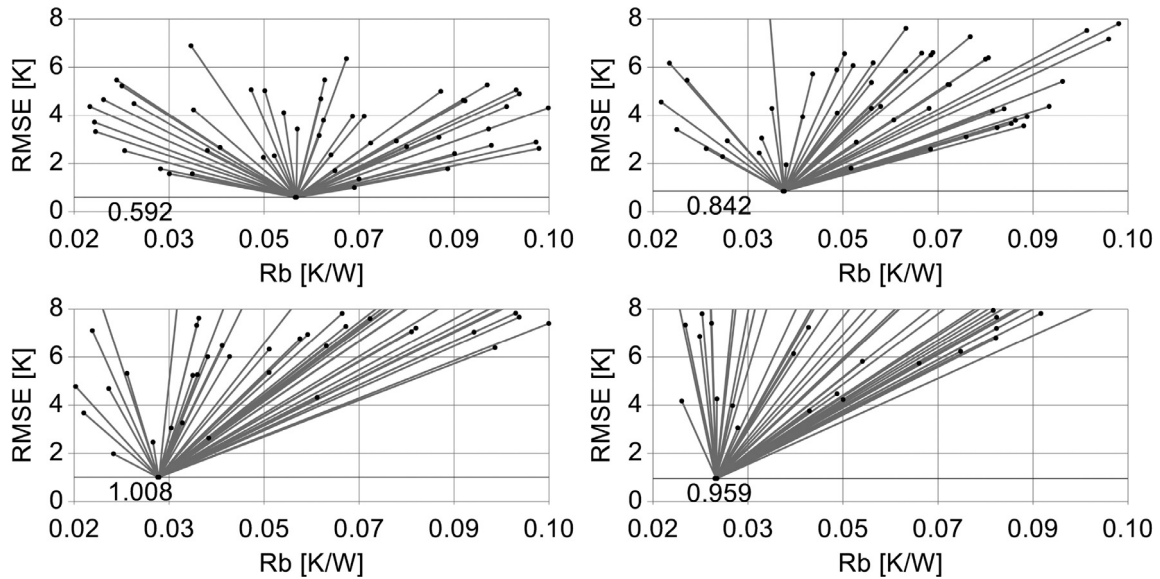
Fig. 8. Comparing optimisation results and convergence for R_b over four datasets. The top left plot is for Nov1, top right for Nov 2, lower left for Dec1 and lower right for Dec2.
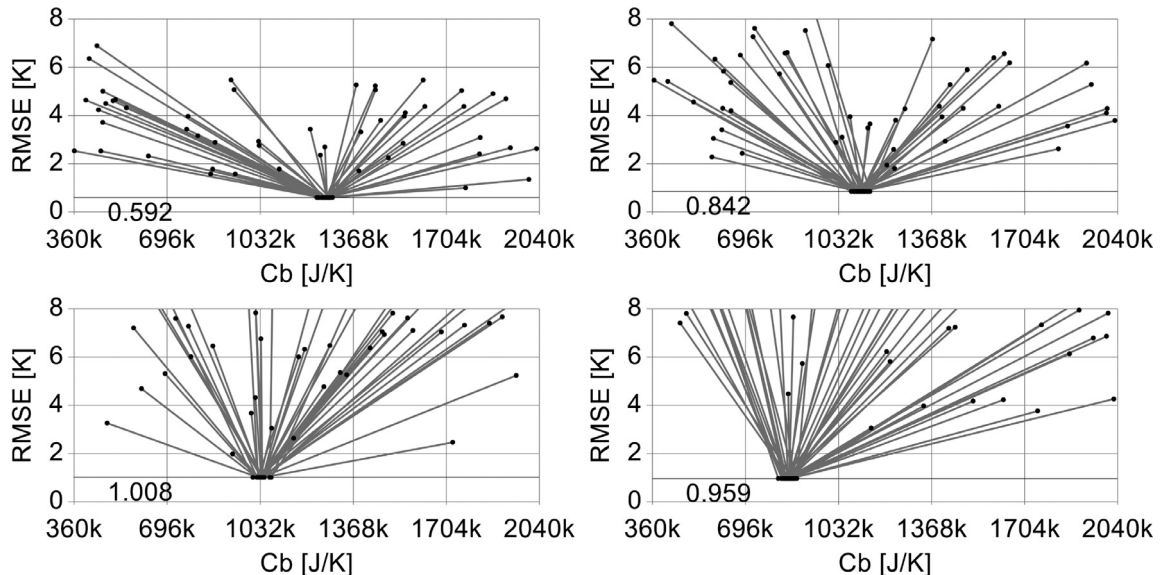

Fig. 9. Comparing optimisation results and convergence for C_b over all four parameter sets. The top left plot is for Nov1, top right for Nov 2, lower left for Dec1 and lower right for Dec2.

Table 5
Summary of average parameter estimates for all four datasets.

Dataset	R_b [K/W]	R_w [K/W]	C_b [J/K]	C_w [J/K]	RMSE4 [K]	RMSE5 [K]
Nov1	0.057	0.066	1271 k	1673 k	0.592	0.591
Nov2	0.041	0.079	1107 k	1321 k	0.842	0.835
Dec1	0.033	0.072	1033 k	1529 k	1.008	1.010
Dec2	0.029	0.087	852 k	1241 k	0.959	0.963

3.5. Model validation

In order to show that the identified model parameters could be used for predicting the thermal behaviour of the building, it is required to validate the model using new data, independent of the data used for parameter estimation. So far, the results are based on model fit, i.e. how well the model is able to fit a particular set of data. A superior measure of model accuracy is the models ability to predict the thermal behaviour using a new dataset. This validates the model performance in similar conditions as those of a Model Predictive Control (MPC) system. Hence it gives a good indication of the models performance used for the purpose of controlling the building temperature. The results in this section are based on the four DOF model.

The results for T_b from validating the model are shown in Fig. 11. The columns represent the dataset used for validation while the rows show the dataset used for parameter estimation. This gives that the diagonal plots are model fit result, while the
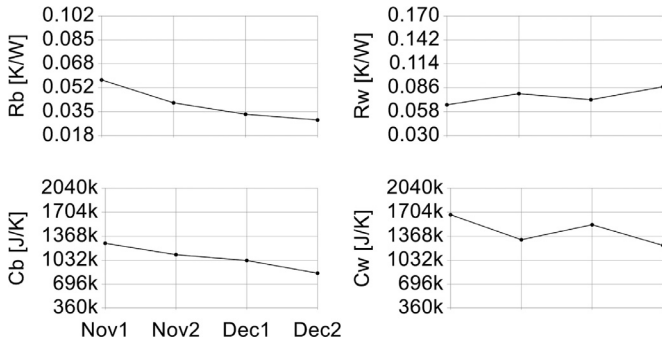
Fig. 10. Estimated parameters for each dataset.

Table 6
Summary of parameter estimates' sample standard deviation for all four datasets.

Dataset	R_b [%]	R_w [%]	C_b [%]	C_w [%]
Nov1	0.065	0.079	1.230	2.208
Nov2	0.079	0.027	1.252	1.995
Dec1	0.066	0.070	0.854	1.250
Dec2	0.105	0.027	1.078	2.144

Table 7
Simulation errors (RMSE) in T_b for all four datasets and parameter sets.

Identification dataset	Input and reference dataset			
	Nov1	Nov2	Dec1	Dec2
Nov1	0.330	0.596	1.899	1.174
Nov2	0.562	0.513	1.692	0.861
Dec1	1.247	1.463	0.588	1.803
Dec2	0.936	0.857	1.586	0.807

Table 8
Simulation errors (RMSE) in T_w for all four datasets and parameter sets.

Identification dataset	Input and reference dataset			
	Nov1	Nov2	Dec1	Dec2
Nov1	0.492	2.095	2.494	5.605
Nov2	1.326	0.668	1.119	2.528
Dec1	1.068	0.833	0.822	2.958
Dec2	2.155	1.471	2.460	0.526

off-diagonal elements are model validation results. The model fit results on the diagonal are plotted on grey background to separate them from the validation results, and are included in the figure for comparison with validation results. The measured temperatures are plotted in black, and the simulated in grey.

The RMSE results for all 16 combinations of dataset and parameter sets are given in Tables 7 and 8. These results, together with the plots in Fig. 11, show that the model is capable of giv-ing good prediction accuracy also for unknown data, although the RMSE varies for different combinations. As for Fig. 11, the diago-nal elements on grey background are the model fit results, while off-diagonal elements in both tables are results from validating the model on independent data. The prediction error for T_b is around 0.5 °C–1.5 °C for most cases, with some combinations approaching 2 °C. Considering that MPC uses feedback, model prediction errors around 1C is likely adequate for the intended purpose. As evident from Table 8 the errors are significantly larger for T_w. However, in a building energy management system, it would be the interior temperature T_b that is of importance for the controller.

Input and reference dataset
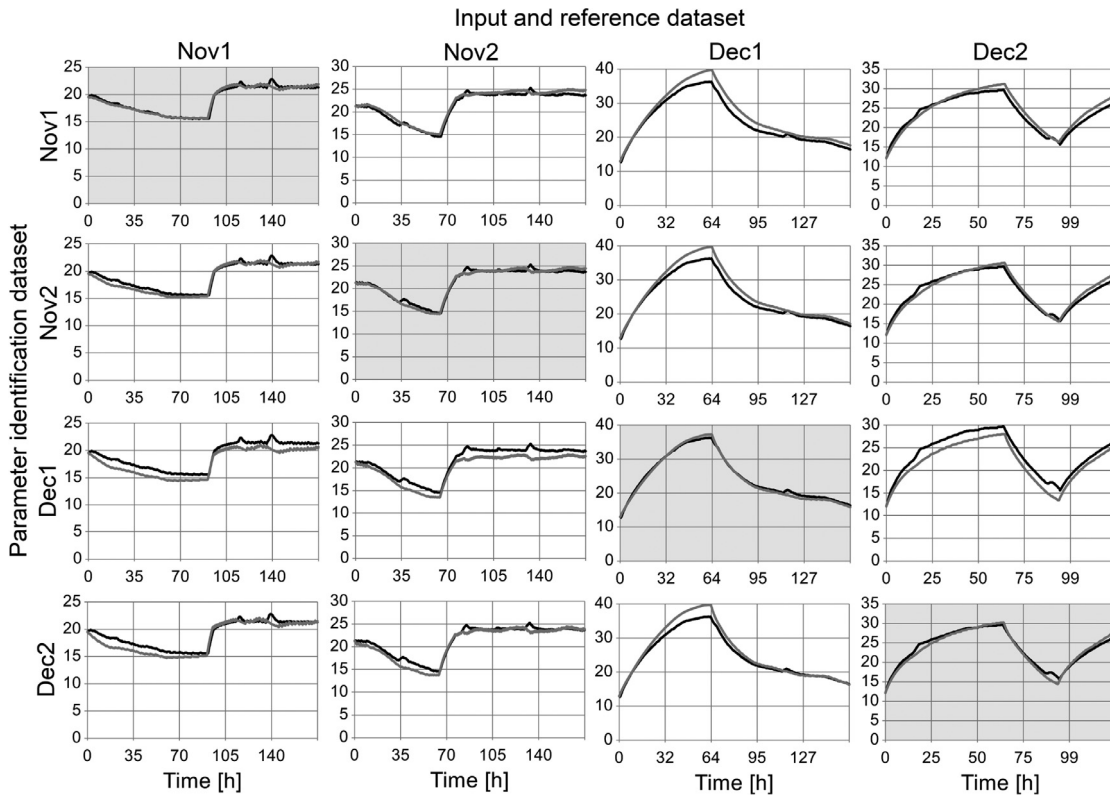
Fig. 11. Results from model simulation are plotted for all four datasets using the four parameter sets shown in Table 5. The results on the diagonal of this figure are model fit results, while the off-diagonal elements are validation results. For the validation plots the parameters are estimated using a dataset that is independent from the one used for inputs and measurement references. The model fit results are plotted on grey background to separate them from the validation results. The simulation results are plotted in grey while the reference data are plotted in black.

4. Discussion

It could be argued that physical interpretation of grey-box parameters for a building model requires that the parameter estimates converge towards a single point in parameter space. If the parameter estimates show a large dispersion in their description of a predominantly time-invariant physical system, they cannot be directly correlated to the specific physical properties of the building.

However, the model as a whole may still be able to accurately predict the behaviour of the system, even if the parameters cannot be assigned a physical interpretation. This is illustrated in the model validation in Section 3.5. The calibrated models where shown to give mean square errors of prediction of around 0.5 °C–1.5 °C for a ~7 day test period for the internal building temperature T_b. Models with prediction errors of this magnitude may be considered usable in MPC systems.

Despite acceptable model validation results, it is of interest to study the reasons for the observed dispersion of estimated parameters, since this gives reason to question the physical interpretability of the estimated parameters. The use of Monte Carlo (MC) sampling of parameter space together with scatter plots of the resulting root mean square error (RMSE) from simulations compared to measurement data was shown to provide a view of the objective function in parameter space. This allows visualisation of the convexity of the objective function, which in turn facilitates cognitive evaluation of the expected optimisation algorithm behaviour.

A randomised initial starting point and repeated execution of optimisation algorithms were used to show that the optimal solution can depend on the starting point. Further, this dispersion in estimated parameters was shown to depend to some extent on the degrees of freedom (DOF) in the estimation problem, since reducing the DOF in the estimation decreases the dispersion of estimated parameters.

Despite reducing the number of estimated parameters, the results still do not converge to a single parameter vector when multiple calibration datasets are used. It is plausible that the datasets used in this work do not contain sufficient dynamic information to identify the model parameters. It is well known that lack of dynamic information in calibration data, caused by insufficient excitation of the physical system during data acquisition, may give rise to problems of practical identifiably [16,18,19]. Due to the experimental conditions encountered when collecting data on building thermal behaviour, ensuring that data contain sufficient dynamic information can be challenging.

Another possible explanation for the observed behaviour of the parameters is that the simplicity of the model allows the estimation procedure to use model parameters to account for unmodelled effects in the physical system. Unknown disturbances, such as variations in humidity or wind, may not be similar between the four investigated datasets.

5. Conclusion

The use of multiple datasets for parameter estimation, as well as the use of MC methods, was shown to give insight into the dispersion of estimated parameters. The model with identified parameters was further shown to give good predictions of building thermal behaviour, and as such would be suitable for model predictive control. However, if physical interpretation of individual parameters is of interest, the dispersion in the parameter estimates needs to be eliminated. This can be done by addressing the challenges of practical identifiably through improved excitation of the physical system. Further, the over-parameterisation of the grey-box model can be reduced by limiting the number of free parameters in the estimation problem.

References

[1] D.W.U. Perera, C.F. Pfeiffer, N.-O. Skeie, Modelling the heat dynamics of a residential building unit: Application to Norwegian buildings, Model. Identif. Control 35 (2014) 43–57.
[2] T. Berthou, P. Stabat, R. Salvazet, D. Marchio, Development and validation of a gray box model to predict thermal behavior of occupied office buildings, Energy Build. 74 (2014) 91–100 5//.
[3] S.F. Fux, A. Ashouri, M.J. Benz, L. Guzzella, EKF based self-adaptive thermal model for a passive house, Energy Build. 68 (Part C) (2014) 811–817 1//.
[4] F. Oldewurtel, A. Parisio, C.N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, et al., Use of model predictive control and weather forecasts for energy efficient building climate control, Energy Build. 45 (2012) 15–27 2012/02/01/.
[5] T. Hilliard, L. Swan, Z. Qin, Experimental implementation of whole building MPC with zone based thermal comfort adjustments, Build. Environ. 125 (2017) 326–338 2017/11/15/.
[6] S. Prívara, J. Cigler, Z. Váňa, F. Oldewurtel, C. Sagerschnig, E. Žáčeková, Building modeling as a crucial part for building predictive control, Energy Build. 56 (2013) 8–22 2013/01/01/.
[7] L. Ljung, Prediction error estimation methods, Circuits Syst. Signal Process. 21 (2002) 11–21 2002/01/01.
[8] M.J. Jiménez, H. Madsen, K.K. Andersen, Identification of the main thermal characteristics of building components using MATLAB, Build. Environ. 43 (2008) 170–180 2//.
[9] K. Esbensen, Multivariate Analysis in Practice, CAMO, Oslo, 1998.
[10] D.W.U. Perera, M. Halstensen, N.-O. Skeie, Prediction of space heating energy consumption in cabins based on multivariate regression modelling, Int. J. Model. Optim. 5 (2015).
[11] A. Afram, F. Janabi-Sharifi, Review of modeling methods for HVAC systems, Appl. Therm. Eng. 67 (2014) 507–519 6//.
[12] A. Afram, F. Janabi-Sharifi, Gray-box modeling and validation of residential HVAC system for control system design, Appl. Energy 137 (2015) 134–150 1/1/.
[13] G. Reynders, J. Diriken, D. Saelens, Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals, Energy Build. 82 (2014) 263–274 10//.
[14] K.K. Associates, Thermal Network Modeling Handbook, 2000.
[15] M.J. Jiménez, H. Madsen, J.J. Bloem, B. Dammann, Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules, Energy Build. 40 (2008) 157–167 //.
[16] H. Madsen, J. Holst, Estimation of continuous-time models for the heat dynamics of a building, Energy Build. 22 (1995) 67–79 3//.
[17] N.R. Kristensen, H. Madsen, S.B. Jørgensen, Parameter estimation in stochastic grey-box models, Automatica 40 (2004) 225–237 2//.
[18] L. Ljung, System Identification - Theory For the User, Prentice Hall PTR, Upper Saddle River, NJ 07458, 1999.
[19] R. Johansson, System Modeling and Identification, Prentice Hall, Englewood Cliffs, 2012.
[20] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy Build. 43 (2011) 1511–1522 7//.
[21] N.R. Kristensen, H. Madsen, Continuous Time Stochastic Modelling, CTSM 2.3 - Mathematcs Guide, Technical University of Denmark, 2013.
[22] K.K. Andersen, H. Madsen, L.H. Hansen, Modelling the heat dynamics of a building using stochastic differential equations, Energy Build. 31 (2000) 13–24 1//.
[23] J.W. Nilsson, S.A. Riedel, Electric Circuits, 6th ed, Prentice Hall, 2001.
[24] T.L. Floyd, Electronic Devices, Prentice Hall, Upper Saddle River, N.J, 2002.
[25] C. Verhelst, F. Logist, J. Van Impe, L. Helsen, Study of the optimal control problem formulation for modulating air-to-water heat pumps connected to a residential floor heating system, Energy Build. 45 (2012) 43–53 2012/02/01/.
[26] M.J.D. Powell, S. Gomez, J.-P. Hennart, A Direct Search Optimization method that models the objective and constraint functions by linear interpolation, in: Advances in Optimization and Numerical Analysis, Springer Netherlands, Dordrecht, 1994, pp. 51–67. Eds., ed.
[27] C.R. Souza, The Accord.NET Framework, Sao Carlos, Brazil, 2014 ed.
[28] D. Perera, N.-O. Skeie, Estimation of the heating time of small-scale buildings using dynamic models, Buildings 6 (2016) 10.

Article B

Parameter estimation for externally simulated thermal network models

Authors O. M. Brastein, B. Lie, R. Sharma and N.-O. Skeie
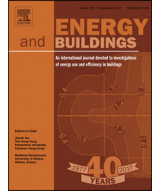
Parameter estimation for externally simulated thermal network models☆

O.M. Brastein*, B. Lie, R. Sharma, N.-O. Skeie

Department of Electrical Engineering, Information Technology and Cybernetics, University of South-Eastern Norway, Porsgrunn N-3918, Norway

ABSTRACT

Obtaining accurate dynamic models of building thermal behaviour requires a statistically solid foundation for estimating unknown parameters. This is especially important for thermal network grey-box models, since all their parameters normally need to be estimated from data. One attractive solution is to maximise the likelihood function, under the assumption of Gaussian distributed residuals. This technique was developed previously and implemented in the Continuous Time Stochastic Modelling framework, where an *Extended Kalman Filter* is used to compute residuals and their covariances. The main result of this paper is a similar method applied to a thermal network grey-box model of a building, simulated as an electric circuit in an *external tool*. The model is described as a list of interconnected components without deriving explicit equations. Since this model implementation is not differentiable, an alternative Kalman filter formulation is needed. The *Unscented* and *Ensemble Kalman Filters* are designed to handle non-linear models without using Jacobians, and can therefore also be used with models in a non-differentiable form. Both Kalman filter implementations are tested and compared with respect to estimation accuracy and computation time. *The Profile Likelihood* method is used to analyse structural and practical parameter identifiability. This method is extended to compute two-dimensional profiles, which can also be used to analyse parameter interdependence by providing insight into the parameter space topology.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Background

The heating and cooling of buildings consumes a significant part of the world's total energy production. While new building materials and techniques may reduce the energy consumption of buildings, the renewal rate of buildings is low [1]. Hence, it is important to study methods that can also reduce energy consumption in existing buildings.

Building Energy Management Systems (BEMS) utilising advanced model-based control methods [2] to forecast the temperature variations of a building in order to predict an optimal sequence of control inputs is a promising method for the reduction of energy consumption. Since the model's prediction accuracy directly influences the efficiency of such methods, it is important to develop accurate models of building thermal behaviour. In addition to describing the time evolution of the system states and outputs, a good model must accommodate descriptions of both measurement noise and process noise [3,4]. This requires a statistically solid framework for estimating unknown parameters [5].

Thermal network models are often used to model the thermal behaviour of buildings [1,6–8]. Implemented as Resistor–Capacitor equivalent circuits, these models offer an intuitive model design based on a cognitive understanding of the thermal physics involved. Since, typically, all parameters of such models must be identified from data, it is important to investigate parameter identifiability prior to assuming physical interpretation of the estimated parameter values [8].

1.2. Previous work

1.2.1. Modelling of dynamic systems

Models are sometimes classified based on the level of physical insight used in their derivation. If the model is mechanistic, i.e., based purely on physical equations, it is classified as *white-box*. Such models excel at describing non-linear state transitions and measurements. They also tend to generalise well between similar systems [5,9]. An alternative approach is the use of *system identification* (SID) methods [3,4,10–12], where a predetermined model structure with unknown coefficients is calibrated using measure-

ments of the system inputs and outputs. This results in a *black-box* model in which no prior physical insight is used, except in the choice of input and output measurements, sample time, and the approximate model complexity. These models tend to have better prediction accuracy, but less capability to generalise [5,9]. SID methods tend to provide better statistics on the model uncertainty, which are typically computed during the calibration process [3–5].

A third, intermediate, possibility is the *grey-box* model, which is based on a simplified model structure constructed using *naive* physical knowledge of the system. Model parameters are calibrated from measurements of the system, similarly to black-box models. Grey-box models are often treated in a stochastic framework [5]. It could be argued that most white-box models include some approximations and/or need calibration of certain parameters. Hence, they can benefit from the application of stochastic grey-box calibration methods. This approach has indeed been claimed as a natural framework for modelling dynamic systems in general [13].

1.2.2. The CTSM framework

Estimation of parameters is essentially an optimisation problem, which requires a well-defined objective function. Several alternatives are used in the literature, such as the deterministic simulation error approach [1]. A statistically solid alternative for stochastic grey-box models is found in [5,14], which is based on maximising the likelihood function evaluated by computing residuals in a Kalman Filter. This method has been previously developed in a number of publications [5,14–16] and implemented in the Continuous Time Stochastic Modelling (CTSM) framework [15]. In CTSM, the residuals needed to evaluate the likelihood function are computed using an *Extended Kalman Filter* (EKF) with subsampling of the *state transition* equations to improve response to non-linear models [5,15]. The EKF is based on linearising the state transitions and/or measurement equations, which requires that the model equations are differentiable [17–19].

1.2.3. Identifiability

Since thermal network building models are partially based on physical knowledge, it is often suggested that the parameters can be assigned a physical interpretation [1,5,6]. This assumption should, however, be verified in the context of parameter identifiability [3,20]. It is well known that models can contain parameters that are structurally non-identifiable [3,20]. Further, lack of proper excitation of the system during data acquisition may lead to practical non-identifiability [3,8,20–22]. While the model structure may be designed such that the parameters are *intended* to have a specific physical meaning, it is not certain that the estimated parameters support this assumption. A good tool for identifiability analysis is the profile likelihood method [8,21,22].

1.3. Overview of paper

In this paper, a resistor-capacitor equivalent thermal network model of a building is expressed as a list of interconnected electrical components. The model is simulated in an external tool without deriving explicit model equations, hence the model cannot be differentiated. This is motivated by the need to simplify experimentation with different model structures in a way that could potentially be automated. The parameter estimation method from the CTSM framework is adapted to non-differentiable models, which requires an alternative to the EKF for computing residuals. Both *the Unscented Kalman Filter* (UKF) [18] and *Ensemble Kalman Filter* (EnKF) [23] are compared and considered for the estimation of residuals. The explicit model equations are also derived on standard linear form, and used with a standard Kalman Filter as a baseline for comparison. Observe that while the model used here is

linear, the method is not restricted to linear models; the externally simulated state transitions could well be non-linear.

A profile likelihood approach is used [22] to analyse parameter identifiability. The method is extended to create two-dimensional profiles in the form of *topological heat maps*. These 2D plots are computed for all combinations of parameters. In addition to diagnosing the identifiability of the parameters, these plots allow detection of parameter interdependence.

The paper is organised as follows. The theoretical basis is discussed in Section 2. The model, external simulator and experimental set-up is presented in Section 3, and the results are presented and discussed in Section 4.

2. Theoretical basis

2.1. Stochastic model parameter estimation

Estimation of parameters for a known model structure [17] can be defined as solving the optimisation problem:

$$\hat{\theta} = \arg\min_{\theta} g(\theta; \mathcal{M}, \mathcal{K}, \mathcal{A}) \qquad (1)$$

$s.t.\ \theta \in \Theta$

Here, $\mathcal{M}$ is a predetermined model structure, which is parametrised by $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^{n_\theta}$ is a set of feasible values for the model parameters that form inequality constraints for the optimisation problem in Eq. (1). $\mathcal{K}$ represents the experimental conditions, including a set of measurements of system inputs and outputs. These measurements are used to evaluate the objective function g when θ is varied over the feasible set Θ by a numerical optimisation algorithm $\mathcal{A}$. In the sequel, the algorithm *Constrained Optimisation By Linear Approximation* (COBYLA) [24] is used. This algorithm is gradient free, hence ideal for solving Eq. (1). COBYLA also supports inequality constraints which can be used to impose the limits of the feasible region Θ on the parameter estimates.

Since the model structure $\mathcal{M}$ is a representation of a system $\mathcal{S}$, it is often assumed that $\mathcal{S} \in \mathcal{M}(\Theta)$ and that consequently there exists a true parameter vector θ^* such that $\mathcal{M}(\theta^*) = \mathcal{S}$. However, this is rarely the case, especially for simplified grey-box models based on a *naive* physical understanding of the system $\mathcal{S}$. Typically, the estimate $\hat{\theta}$ depends on the amount of dynamic information in $\mathcal{K}$, the choice of objective function g, and to some extent on the optimisation algorithm $\mathcal{A}$. Hence, it is necessary to analyse the identifiability of the estimated parameters. This topic is further discussed in Section 2.4.

Next, define the continuous time input $u_t \in \mathbb{R}^{n_u}$ and output $y_t \in \mathbb{R}^{n_y}$, and the corresponding ordered sequences of discrete time measurements u_k and y_k taken from the system $\mathcal{S}$:

$$y_{[N]} = [y_0, y_1, \ldots, y_N] \qquad (2)$$

$$u_{[N]} = [u_0, u_1, \ldots, u_N] \qquad (3)$$

Here, the integer subscripts $k = 0, 1, \ldots, N$ denote the discrete time sampling instants, and the subscript enclosed in $[\cdot]$ is used to indicate an ordered sequence.

A grey-box model can be expressed as a continuous time stochastic differential equation (SDE) with a discrete time measurement equation; adopting the notation of [5]:

$$dx_t = f(x_t, u_t, t, \theta)dt + \sigma(u_t, t, \theta)d\omega_t \qquad (4)$$

$$y_k = h(x_k, u_k, t_k, \theta) + e_k \qquad (5)$$

where $t \in \mathbb{R}$ is the time variable and $x_t \in \mathbb{R}^{n_x}$ is the continuous time state vector. The first and second terms in the *state transition* equation, given in Eq. (4), are commonly called the *drift* and *diffusion* term, respectively [5,25]. The diffusion term expresses the

process noise as the function σ multiplied with the differential of a standard Wiener process ω_t. The discrete time *measurement equation* is given in Eq. (5).

2.2. Maximum likelihood

This section gives a summary of the theoretical basis adopted from the CTSM framework [5,14,15]. The objective function g in Eq. (1) can be derived from the likelihood function, which is defined as the probability of observing the measurement sequence $y_{[N]}$ when θ and $\mathcal{M}$ are known, i.e.:

$$L\big(\theta; y_{[N]}, \mathcal{M}\big) = p\big(y_{[N]}|\theta, \mathcal{M}\big) \tag{6}$$

In the sequel, the model structure $\mathcal{M}$ is implicitly assumed known and omitted from the condition. By application of the rule $P(A \cap B) = P(A|B)P(B)$[25], Eq. (6) can be expanded such that:

$$L\big(\theta; y_{[N]}\big) = \left(\prod_{k=1}^{N} p\big(y_k|y_{[k-1]}, \theta\big) \right) p(y_0|\theta) \tag{7}$$

The diffusion term in Eq. (4), which is assumed to be additive and independent of the state x, is driven by a *Wiener process* whose differential is Gaussian distributed [5]. Hence, it is reasonable to assume that the conditional probabilities in Eq. (7) can

be approximated by Gaussian distributions [5,15]. This assumption can be checked during model validation by testing the residuals for normality [3,5]. The likelihood can then be expressed as a multivariate Gaussian distribution [5],

$$L\big(\theta; y_{[N]}\big) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\epsilon_k^T \mathcal{E}_{k|k-1}^{-1}\epsilon_k\right)}{\sqrt{\det\big(\mathcal{E}_{k|k-1}\big)}\big(\sqrt{2\pi}\big)^{n_y}} \right) p(y_0|\theta) \tag{8}$$

A Kalman Filter may be used to estimate the quantities

$$\hat{y}_{k|k-1} = \mathbb{E}\big[y_k|y_{[k-1]}, \theta\big] \tag{9}$$

$$\epsilon_k = y_k - \hat{y}_{k|k-1} \tag{10}$$

$$\mathcal{E}_{k|k-1} = \mathbb{E}\big[\epsilon_k \epsilon_k^T\big] \tag{11}$$

In the CTSM framework, an EKF is used. In Section 2.3 the alternative use of UKF and EnKF is discussed.

Eq. (8) can further be simplified by taking the negative of the logarithm; defining the log likelihood function $\ell(\theta; y_{[N]})$:

$$\ell\big(\theta; y_{[N]}\big) = -\ln\big(L\big(\theta; y_{[N]}\big)\big) \tag{12}$$

The solution to the optimisation problem is not affected since

$$\arg\max_{\theta \in \Theta} L\big(\theta; y_{[N]}\big) = \arg\min_{\theta \in \Theta} \ell\big(\theta; y_{(N)}\big) \tag{13}$$

Table 1
Comparing equations for UKF (left) and EnKF (right).

Definitions and initialisation	
$\zeta_m^{(0)} = \frac{\lambda}{\lambda+n_x}$	$w_k^{(i)} \sim \mathcal{N}(\bar{w}_k, \mathcal{W}_k), \quad i \in \{1, \ldots, n_p\}$
$\zeta_c^{(0)} = \frac{\lambda}{\lambda+n_x} + \big(1 - \alpha^2 + \beta\big)$	$v_k^{(i)} \sim \mathcal{N}(\bar{v}_k, \mathcal{V}_k), \quad i \in \{1, \ldots, n_p\}$
$\zeta_m^{(i)} = \zeta_c^{(i)} = \frac{1}{2(\lambda+n_x)}, \quad i \in \{1, \ldots, 2n_x\}$	$x_{0|0}^{(i)} \sim \mathcal{N}(\bar{x}_0, X_0), \quad i \in \{1, \ldots, n_p\}$
$\lambda = \alpha^2(n_x + \kappa) - n_x$	
$\hat{x}_{0|0} = \mathbb{E}[x_0] = \bar{x}_0$	$\hat{x}_{0|0} = \frac{1}{n_p}\sum_{i=1}^{n_p} x_{0|0}^{(i)}$
$X_{0|0} = \mathbb{V}\big[x_0 - \hat{x}_{0|0}\big] = X_0$	$X_{0|0} = \frac{1}{n_p-1}\sum_{i=1}^{n_p}\big(x_{0|0}^{(i)} - \hat{x}_{0|0}\big)(\ldots)^T$
State propagation	
$x_{k-1|k-1}^{\{2n_x+1\}} = \varsigma\big(\hat{x}_{k-1|k-1}, X_{k-1|k-1}\big)$	
$x_{k|k-1}^{(i)} = f\big(x_{k-1|k-1}^{(i)}, u_{k-1}, \bar{w}_k\big) \quad i \in \{0, \ldots, 2n_x\}$	$x_{k|k-1}^i = f\big(x_{k-1|k-1}^{(i)}, u_{k-1}, w_{k-1}^{(i)}\big) \quad i \in \{1, \ldots, n_p\}$
$\hat{x}_{k|k-1} = \sum_{i=0}^{2n_x} \zeta_m^{(i)} x_{k|k-1}^{(i)}$	$\hat{x}_{k|k-1} = \frac{1}{n_p}\sum_{i=1}^{n_p} x_{k|k-1}^{(i)}$
$^{a)}X_{k|k-1} = \sum_{i=0}^{2n_x} \zeta_c^{(i)}\big(x_{k|k-1}^{(i)} - \hat{x}_{k|k-1}\big)(\ldots)^T + \mathcal{W}_k$	$^{b)}X_{k|k-1} = \frac{1}{n_p-1}\sum_{i=1}^{n_p}\big(x_{k|k-1}^{(i)} - \hat{x}_{k|k-1}\big)(\ldots)^T$
Measurement estimate	
$x_{k|k-1}^{\{2n_x+1\}} = \varsigma\big(\hat{x}_{k|k-1}, X_{k|k-1}\big)$	
$y_{k|k-1}^{(i)} = h\big(x_{k|k-1}^{(i)}, u_{k-1}, \bar{v}_k\big) \quad i \in \{0, \ldots, 2n_x\}$	$y_{k|k-1}^{(i)} = h\big(x_{k|k-1}^{(i)}, u_{k-1}, v_{k-1}^{(i)}\big) \quad i \in \{1, \ldots, n_p\}$
$\hat{y}_{k|k-1} = \sum_{i=0}^{2n_x} \zeta_m^{(i)} y_{k|k-1}^{(i)}$	$\hat{y}_{k|k-1} = \frac{1}{n_p}\sum_{i=1}^{n_p} \hat{y}_{k|k-1}^{(i)}$
Innovation and cross covariance	
$Z_{k|k-1} = \sum_{i=0}^{2n_x} \zeta_c^{(i)}\big(x_{k|k-1}^{(i)} - \hat{x}_{k|k-1}\big)\big(y_{k|k-1}^{(i)} - \hat{y}_{k|k-1}\big)^T$	$Z_{k|k-1} = \frac{1}{n_p-1}\sum_{i=1}^{n_p}\big(x_{k|k-1}^{(i)} - \hat{x}_{k|k-1}\big)\big(y_{k|k-1}^{(i)} - \hat{y}_{k|k-1}\big)^T$
$^{a)}\mathcal{E}_{k|k-1} = \sum_{i=0}^{2n_x} \zeta_c^{(i)}\big(y_{k|k-1}^{(i)} - \hat{y}_{k|k-1}\big)(\ldots)^T + \mathcal{V}_k$	$\mathcal{E}_{k|k-1} = \frac{1}{n_p-1}\sum_{i=1}^{n_p}\big(y_{k|k-1}^{(i)} - \hat{y}_{k|k-1}\big)(\ldots)^T$
$K_k = Z_{k|k-1}\mathcal{E}_{k|k-1}^{-1}$	$K_k = Z_{k|k-1}\mathcal{E}_{k|k-1}^{-1}$
Aposteriori update$^{c)}$	
$\epsilon_{k|k-1} = y_k - \hat{y}_{k|k-1}$	$x_{k|k}^{(i)} = x_{k|k-1}^{(i)} + K_k(y_k - y_{k|k-1}^{(i)}) \quad i \in \{1, \ldots, n_p\}$
$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k\epsilon_{k|k-1}$	$^{b)}\hat{x}_{k|k} = \frac{1}{n_p}\sum_{i=n_p}^{n_p} x_{k|k}^{(i)}$
$X_{k|k} = X_{k|k-1} - K_k\mathcal{E}_{k|k-1}K_k^T$	$^{b)}X_{k|k} = \frac{1}{n_p-1}\sum_{i=1}^{n_p}(x_{k|k}^{(i)} - \hat{x}_{k|k})(\ldots)^T$

$^{a)}$ Assuming affine noise. (See Remark 3).
$^{b)}$ Can be omitted (See Remark 5).
$^{c)}$ Mathematically equivalent but *not* interchangable (See Remark 6).

Finally, by conditioning on knowing y_0, and eliminating the scaling constants $\frac{1}{2}$ from $\ell(\theta; \theta; y_{[N]})$, the objective function from Eq. (1) is given as:

$$g(\theta; \mathcal{M}, \mathcal{K}) = \sum_{k=1}^{N} \epsilon_k^T \mathcal{E}_{k|k-1}^{-1} \epsilon_k + \ln\left(\det\left(\mathcal{E}_{k|k-1}\right)\right) \tag{14}$$

where the constant term $c = N \cdot n_y \cdot \ln(2\pi)$ is dropped.

2.3. Alternative KF formulations

The popularity of the Kalman Filter has led to a number of adaptions. The *Extended Kalman Filter* (EKF) is perhaps the most common such adaption and is used in [5]. In the sequel, two other well known KF variations are outlined; the *Unscented Kalman Filter* (UKF) [18] and the *Ensemble Kalman Filter* (EnKF) [23]. In addition to better approximations for non-linear models, UKF and EnKF dispense with the computation of Jacobians and therefore do not require the model to be differentiable [18]. Both filters are listed and compared in Table 1.

Given the SDE for the state transition as in Eq. (4), the time evolution of the probability density function (pdf) of the state, $p(x, t)$, is described by the *Fokker–Planck equation* [23], also known as the *Kolmogorov forward equation* [5]. The multi-dimensional Fokker–Planck equation [25] can be expressed as

$$\frac{\partial p(x,t)}{\partial t} + \sum_i \frac{\partial}{\partial x_i} (f_i(x_t, u_t, t, \theta) p(x,t))$$
$$= \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} p(x,t) \left(\sigma W \sigma^T\right)_{ij} \tag{15}$$

where f_i is the ith component of the state transition model.

In the EKF, the linearised model is used to approximate the first moments of this pdf [23] by a Taylor series expansion truncated after the first term [17,19]. In both UKF and EnKF, the Fokker–Planck equation is instead solved by approximating the solution to Eq. (15) using a set of state realisations. The key difference between the UKF and EnKF is in how that set is constructed. The UKF draws its state realisation set, called *sigma points*, using the unscented transform (UT). The UT of an expected state $\bar{x}$ with covariance X deterministically computes a set of sigma points $x^{\{N\}} = \left\{ x^{(i)} : \quad i = 0, 1, \ldots, N \right\}$, where the shorthand $\{\cdot\}$ superscript indicates a set and a superscript $(\cdot)$ denotes a member. For convenience of notation, a UT operator $\varsigma(\bar{x}, X)$ that returns a set of $N = 2n_x + 1$ sigma points is defined as

$$x^{(0)} = \hat{x} \tag{16}$$

$$x^{(i)} = \hat{x} + \left(\sqrt{(n_x + \lambda)X}\right)_i, \; i \in \{1, \ldots, n_x\} \tag{17}$$

$$x^{(n_x+i)} = \hat{x} - \left(\sqrt{(n_x + \lambda)X}\right)_i, \; i \in \{1, \ldots, n_x\} \tag{18}$$

The square root is often implemented using a Cholesky decomposition, and the subscript i denotes the i-th column [17,18]. Note that there are different versions of the UT [3,19], where the one presented in Eqs. (16)–(18) is used in the sequel. For a Gaussian random variable (GRV), the UT is known to approximate the pdf $p(x, t)$ to third order accuracy, and to the second order for non-Gaussian random variables [17]. The introduction of $\lambda = \alpha^2(n_x + \kappa) - n_x$ in Eqs. (16)–(18) gives a set of tuning parameters that can improve approximations of higher order moments [17–19].

In contrast to the deterministic UT, the EnKF represents the state pdf using a Monte Carlo (MC) sampling method [17,18,23]. The pdf is approximated as $p(x, t) = \frac{dN}{n_p}$, where dN is the number of state realisations in some small unit volume and n_p is the total

number of realisations [23]. The set of realisations, i.e., the ensemble, is initially drawn at random using the mean and covariance of the initial state. Subsequently, each realisation is propagated as a distinct trajectory, thus making the EnKF equivalent to using a Markov Chain Monte Carlo (MCMC) method to solve the Fokker–Planck equation [23].

2.3.1. Remarks to Table 1

Remark 1. Initialisation for both filters is equivalent if n_p is "large", since the computed ensemble values based on MC sampling converge to the expectation values $\bar{x}_0$ and X_0.

Remark 2. In the UKF, the sigma transform is applied twice to compute the sigma points for both apriori and aposteriori state and covariance estimates. In the EnKF, the realisations are drawn only in the initialisation, and subsequently propagated independently.

Remark 3. The process noise $w_k \sim \mathcal{N}(\bar{w}_k, \mathcal{W}_k)$ and measurement noise $v_k \sim \mathcal{N}(\bar{v}_k, \mathcal{V}_k)$ enter the UKF and EnKF in different ways. The model in Eqs. (4) and (5) assumes affine noise, hence the noise covariances are added to the respective propagation equations in the UKF. For non-affine noise, there are other adaptions of the UKF, e.g., estimating noise by augmenting the state vector, that can be used [18]. In the EnKF, a random number generator (RNG) is used to draw instances of the noise which is subsequently used in the state transition and measurement equations for propagation of the ensemble.

Remark 4. If $\zeta_m^{(i)} = \frac{1}{n_p}$ and $\zeta_c^{(i)} = \frac{1}{n_p-1}$ in the UKF formulation, the corresponding equations for estimating mean and covariance from the realisation set would be identical to EnKF (except for the iteration index) when n_p is large and $\lambda = 0 \leftrightarrow \alpha = 1, \kappa = 0$.

Remark 5. In order to show the similarity of UKF and EnKF, both filters are formulated with expressions for computing apriori and aposteriori covariance for the state estimate. Observe that for the UKF these are needed in order to compute new sets of sigma points, while in the EnKF this computation can be omitted. Indeed, a fundamental advantage of the EnKF is that it does not require explicit computation of the apriori and aposteriori state estimate covariance matrices, but rather propagates them as approximations in the ensemble. This is an advantage of the EnKF for models with a high number of states.

Remark 6. The EnKF aposteriori update of state realisations and covariance can be shown to be equivalent to the corresponding aposteriori update in the UKF. However, since EnKF treats the set of realisation as independent state trajectories, the ensemble must be updated from apriori to aposteriori state estimates. Hence, the two formulations are not interchangeable, despite being mathematically equivalent.

Remark 7. UKF has three hyper parameters, α, κ and β; default tunings are suggested for standard noise models in the UKF literature. The EnKF has only one hyper parameter: the number of realisations n_p.

2.4. Profile likelihood

Parameter estimates are often reported as a point in the parameter space Θ, or as a *confidence interval* [26] with some stated confidence α. An alternative solution is to present the distribution of the parameters over the feasible range Θ. Since the estimation of parameters is based on the *likelihood function* in Eq. (6), one attractive choice for creating parameter distributions is the *profile likelihood* (PL) method presented in [8,21,22]. This approach was also suggested by the authors of CTSM [27,28]. The PL method explores the parameter space by optimising the parameters in two steps,

rather than simultaneously as in Eq. (1). For simplicity of notation, the dependence on $y_{[N]}$ is omitted from the log likelihood function $\ell(\theta; y_{[N]})$ in the sequel. The *profile likelihood* $\ell_{PL}(\theta_i)$ is defined as the minimum log likelihood for θ_i when the remaining parameters are freely optimised [22,29]:

$$\ell_{PL}(\theta_i) = \min_{\theta_{j \neq i}} g\left(\theta_{j \neq i}; \mathcal{M}, \mathcal{K}, \theta_i\right) \tag{19}$$

Values of θ_i must be chosen prior to optimising the remaining $\theta_{j \neq i}$ [22]. A straightforward solution, if the objective function g is well behaved within the constraints of Θ, is to use a *brute force* approach with an even sampling of θ_i. Alternatively, a two-sided gradient decent algorithm, using a freely optimised parameter vector as a starting point, can be applied [22,30]. The resulting likelihood distribution can be plotted as a function of θ_i and subsequently analysed according to the definitions of structural and practical identifiability for *likelihood-based confidence intervals* [8]. Unlike the asymptotic confidence interval, which is based on the curvature of the likelihood function by computation of the Hessian [8,22], the likelihood-based confidence interval is computed by applying a *threshold* to the likelihood function to compute a confidence region [22,29]. Let

$$\left\{\theta : \ell(\theta) - \ell\left(\hat{\theta}\right) < \Delta_\alpha\right\}, \quad \Delta_\alpha = \chi^2(\alpha, n_{df}) \tag{20}$$

where $\hat{\theta}$ is a freely estimated, presumed optimal, parameter vector, and the threshold Δ_α is the α percentile of the χ^2-distribution with n_{df} degrees of freedom. It follows from *Wilks' theorem* [31] that the logarithm of the likelihood ratio Λ test statistic

$$2\ln(\Lambda) = 2\ln\left(\frac{L(\theta)}{L(\hat{\theta})}\right) = \ell(\theta) - \ell(\hat{\theta}) \tag{21}$$

can be used to compare two models. The difference in log likelihood $\ell(\theta) - \ell\left(\hat{\theta}\right)$ is asymptotically χ^2-distributed [22,32], with n_{df} equal to the difference in the number of free parameters between θ and $\hat{\theta}$. Hence, the PL method uses a χ^2 threshold with $n_{df} = 1$. This form of confidence interval allows interpretation of structural and practical identifiability by inspection of the upper and lower confidence boundaries [22]. If $\ell(\theta)$ is lower than the threshold in both directions, i.e., the interval at the stated confidence level is unbounded ($\pm\infty$), the parameter is classified as structurally non-identifiable [22]. If $\ell(\theta)$ is bounded in one direction, this indicates practical non-identifiability [22,29]. Profile likelihood plots are interpreted similarly. If the plot is lower than the confidence threshold in both directions or only one, this indicates structural or practical non-identifiability, respectively.

2.4.1. Two-dimensional profile likelihood

The PL method essentially projects the n_θ dimensional space Θ onto the single parameter θ_i, by freely estimating the remaining parameters. Hence, if parameters are not independent, the PL method tends to overestimate the width of the likelihood-based confidence interval. A step towards remedying this issue is to modify the PL method to hold out *two* parameters rather than one, i.e.,

$$\ell_{PL2}(\theta_i, \theta_j) = \min_{\theta_{k \neq i,j}} g\left(\theta_{k \neq i,j}; \mathcal{M}, \mathcal{K}, \theta_i, \theta_j\right) \tag{22}$$

This results in a two-dimensional distribution which can be analysed in a similar way to the one-dimensional PL [22], using the definition in Eq. (20). The PL2 results are plotted as topological surfaces [22]. This projects the parameter space Θ onto the plane of θ_i and θ_j. In addition to diagnosing identifiability issues, these plots can be used to diagnose parameter interdependence. Observe that since $\hat{\theta}$ has n_θ free parameters while the PL2 estimate has

$n_\theta - 2$, this gives $n_{df} = 2$ for the computation of Δ_α from the χ^2-distribution in Eq. (20).

Applying a confidence threshold to the PL2 method produces *confidence regions* in the (θ_i, θ_j) plane, rather than *intervals* in a single parameter. Based on confidence thresholds computed from the χ^2 distribution, a similar interpretation of these two-dimensional topologies can be applied to diagnose identifiability by requiring that the region is bounded in all directions. If there is an unbounded equipotential *valley* with a log likelihood below the Δ_α threshold, the parameter is structurally non-identifiable. If the interval or region is unbounded only in one direction, this indicates a practically non-identifiable parameter. Examples of two-dimensional PL plots are given in Section 4. If parameter interdependence is observed, re-parametrisation of the model such that the interdependency is resolved, may be advisable in order to obtain a model with tighter confidence bounds on the estimated parameters.

2.4.2. Interpretation of wide confidence regions

It can be argued that a *wide* confidence region is indicative of an identifiability issue even if the region is bounded. If the range of acceptable parameter values is large, the interpretation of the estimated parameters as being determined by the physical properties of the system, i.e., $\mathcal{S} \in \mathcal{M}(\Theta) \rightarrow \mathcal{M}(\hat{\theta}) \approx \mathcal{S}$, is questionable.

One possible cause of wide confidence bounds on the estimated parameters is the presence of nuisance parameters, i.e., parameters whose value is insignificant for the model estimates.

2.4.3. Effect of constrained parameters

Observe that solving the two-step optimisation problem in Eq. (19) subjected to the constraint $\theta \in \Theta$ imposes a restriction on the identified profile $\ell_{PL}(\theta_i)$. This constraint may skew the results, since the remaining parameters $\theta_{i \neq j}$ are only considered within the region Θ. If parameters are not independent, the profile of one parameter may be influenced by the constraints of another. In the PL2 method, the effect of constrained optimisation of parameters is easier to diagnose, since dependent parameters can be identified from the topology plots.

2.5. Model validation

The CTSM method requires evaluation of the residuals to verify that the assumption of Gaussian distributed residuals is justified [5,15]. In the CTSM literature, the autocorrelation function (ACF) is used to test for normality of residuals in the time-domain, while a cumulative periodogram (CP) is used in the frequency domain [5,8,15]. There are also a number of alternative tests for normality that can be applied, such as the *zero-crossings* test or the *Kolmogorov–Smirnov* test [3].

3. Case study model and simulation

3.1. Model

A thermal network model of a building can be expressed as a resistor-capacitor (RC) circuit. These models are based on a *naive* physical understanding of temperature variations in the building structure, which entails simplifications that necessarily introduce modelling errors. The result is a simplified, lumped parameter model, which should be treated in the framework of grey-box modelling, and hence formulated as stochastic differential equations (SDE) as in Eq. (4) [5].

Fig. 1 shows an example of a candidate RC model which was developed to approximate the thermal behaviour of the experimental building discussed in Section 3.2, partially based on the
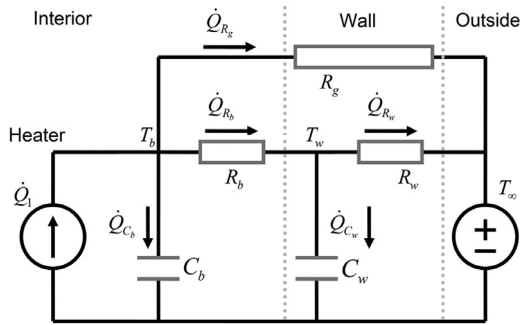
Fig. 1. The R3C2 thermal network model of an experimental building can be expressed as a resistor–capacitor equivalent circuit containing three resistors and two capacitors.
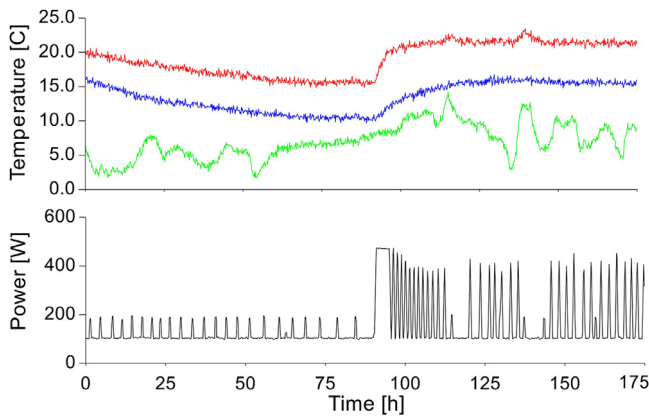

Fig. 2. Calibration data for the R3C2 model. The model outputs T_b (red) and T_w (blue) are plotted together with the outdoor temperature input T_∞ (green). The input power $\dot{Q}$ is plotted separately. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Nominal parameter values and min/max limits for resistances [K/W] and capacitances [J/K].

	R_b	R_w	R_g	C_b	C_w
θ_0	0.100	0.100	0.250	1200k	1200k
θ_{min}	0.030	0.030	0.075	360k	360k
θ_{max}	0.170	0.170	0.425	2040k	2040k

R4C2 model presented in [1]. The model has two outputs: the room temperature T_b and the wall surface temperature T_w, and two inputs: the consumed power by an electric heating element $\dot{Q}$ and the outside temperature T_∞. Five components form the model structure: the thermal resistance between room air and wall R_b, the building envelope R_w, and the thermal resistance of windows and doors R_g. The two capacitances C_b and C_w represent the thermal capacitance of the building interior and envelope, respectively. A nominal parameter vector θ_0, listed in Table 2, is used as the initial value for parameter estimation. Additionally, the feasible values region Θ is limited by θ_{min} and θ_{max}, which are chosen as $0.3 \times \theta_0$ and $1.7 \times \theta_0$, respectively.

3.2. Calibration data

The calibration data used for parameter estimation was obtained from an experimental building located at Campus Porsgrunn of the University of South–Eastern Norway (USN). The data was collected by multiple data acquisition systems, each producing a separate data subset, and combined into a consistent dataset in the preprocessing ste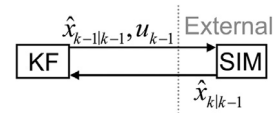p. The data was first filtered to remove noise and subsequently resampled into a uniform temporal scale. In order to maintain measurement uncertainty after preprocessing, a random noise component of covariance 0.1 was added to the temperature measurements. The resulting data is presented in Fig. 2.

3.3. RCSimulator

The choice of model structure for a thermal network model, i.e., the RC circuit, usually involves significant experimentation [1,7,16]. To simplify, and possibly automate, the process of finding appropriate model structures, it is useful to simulate such models without requiring explicit model equations. Since the thermal networks are modelled as RC circuits, it is natural to look to the electronics field where circuits are often simulated using tools such as SPICE [33]. A circuit simulator can be used to propagate the state, hence replacing the *drift* term of Eq. (4), as illustrated in Fig. 3. Using this set-up with the parameter estimation method in Section 2.2 requires a KF implementation that can handle non-differentiable models, such as UKF and EnKF.

A simple circuit simulator is constructed, named *RCSimulator* for reference in the sequel. Circuit simulators typically define the circuit model as a list of interconnected components, which can be taken directly from the schematic in Fig. 1. By convention, all components have two terminals named *in* and *out*. Each node is assigned an integer index which is used to configure the connections of the components as a circuit. For example, letting node T_b have index 1 and T_w index 2, the component R_b would have input/output assignment (1,2). For each node in the circuit, *Kirchoff's node current law* is used to balance the flow in and out of the node [34]. The system of node equations can be written in difference form:

$$A x_k + A_m x_{k-1} + B u_k = 0 \qquad (23)$$

The contributions from all components are summed together, such that rows i in A, A_m, and B constitute the balance equation for node i. Eq. (23) is solved for x_k at each time-step in order to propagate the state. The only dynamic element is the capacitor, which is implemented using an implicit Euler discretisation, $\left(\frac{dx}{dt}\right)_{t_k} \approx \frac{x_k - x_{k-1}}{\Delta t}$, by contributing to both the A and A_m matrices. Voltage sources are implemented as constraints on the difference between the states of the two connected nodes. The measurement Eq. (5) can be implemented as measuring the potential between selected nodes in the RC circuit.

The simulation scheme, and in particular the discretisation of the capacitive elements, could be extended with more accurate approximations such as the *Runge–Kutta 4th order* (RK4) scheme [35]. It is also possible to introduce non-linear components, such as variable resistors. Observe that while the test case model used here is linear, the method of estimating residuals with UKF or EnKF for externally simulated models has no such restriction.

3.4. Discrete time linear model

For comparison, the model is also expressed in a standard linear state space form

$$\frac{dx}{dt} = A x_t + B u_t + G w_t \qquad (24)$$

Fig. 3. Illustration of Kalman Filter (KF) with externally simulated (SIM) state propagation.

where

$$x_t = \begin{bmatrix} T_b \\ T_w \end{bmatrix}, \; u_t = \begin{bmatrix} \dot{Q}_1 \\ T_\infty \end{bmatrix}, \; B = \begin{bmatrix} \frac{1}{C_b} & \frac{1}{C_b R_g} \\ 0 & \frac{1}{C_w R_w} \end{bmatrix}$$

$$A = \begin{bmatrix} -\frac{1}{C_b R_b} - \frac{1}{C_b R_g} & \frac{1}{C_b R_b} \\ \frac{1}{C_w R_b} & -\frac{1}{C_w R_b} - \frac{1}{C_w R_w} \end{bmatrix}$$

and $w_t \sim N(0, W)$ is the process noise (model error), W is the spectral density of w_t and G is a noise distribution matrix which is assumed constant. Hence $Gw_t dt$ is equivalent to the stochastic diffusion term $\sigma(u_t, t, \theta)d\omega$ in Eq. (4) such that $GWG^T = \sigma\sigma^T$ [5].

The model in Eq. (24) is written in continuous time and must be discretised for use in a computer KF implementation. Assuming zero order hold (ZOH) on the inputs [36] gives

$$\frac{d}{dt}\begin{bmatrix} x \\ u \end{bmatrix}_t = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ u \end{bmatrix}_t + \begin{bmatrix} G \\ 0 \end{bmatrix}w_t \tag{25}$$

The discrete time equivalent system, again assuming ZOH on inputs, gives the difference equation

$$\begin{bmatrix} x \\ u \end{bmatrix}_k = \begin{bmatrix} \tilde{A} & \tilde{B} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ u \end{bmatrix}_{k-1} + \begin{bmatrix} \tilde{G} \\ 0 \end{bmatrix}w_k \tag{26}$$

where $w_k \sim N(0, \mathcal{W})$ and $\mathcal{W}$ is the process noise covariance. On the interval $[t_{k-1}, t_k[\to \Delta t = t_k - t_{k-1}$, Eq. (25) has the known solution

$$\begin{bmatrix} x \\ u \end{bmatrix}_{t_k} = \exp\left(\Delta t \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}\right)\begin{bmatrix} x \\ u \end{bmatrix}_{t_{k-1}} + \begin{bmatrix} \Delta x_w \\ 0 \end{bmatrix} \tag{27}$$

which by direct comparison with Eq. (26) gives the discrete model matrices $\tilde{A} = e^{A\Delta T}$ as the upper left sub-matrix and $\tilde{B}$ as the corresponding upper right sub-matrix. The diffusion of the states driven by the process noise w_t is here expressed by the term Δx_w which is yet undetermined. $\tilde{G}$ is obtained from the expectation

$$E\left(x_{k|k-1}x_{k|k-1}^T\right) = X_{k|k-1}$$

$$= \tilde{A}X_{k-1|k-1}\tilde{A}^T + \int_{t_{k-1}}^{t_k} e^{A\tau}Q_c e^{A^T\tau}d\tau \tag{28}$$

where $Q_c = GWG^T$. This integral can be solved using the Van Loan method in [37]. Let

$$D = \exp\left(\Delta t \begin{bmatrix} -A & Q_c \\ 0 & A^T \end{bmatrix}\right) = \begin{bmatrix} D_{11} & D_{12} \\ 0 & D_{22} \end{bmatrix} \tag{29}$$

Then

$$Q(\Delta t) = \int_{t_{k-1}}^{t_k} e^{A\tau}Q_c e^{A^T\tau}d\tau = \tilde{G}\mathcal{W}\tilde{G}^T = D_{22}^T D_{12} \tag{30}$$

From Eq. (30) it is possible to compute *either* the process noise covariance $\mathcal{W}$ or the discrete time distribution matrix for the process noise $\tilde{G}$ when the other is known or assumed. Setting $\tilde{G} = I$ gives $\mathcal{W} = Q(\Delta t) = D_{22}^T D_{12}$ which is equivalent to the result in [5]. This is also the form typically used in derivation of Kalman Filters [18,19], as shown in Table 1.

While the primary focus of this paper is on the externally simulated model, the model in Eq. (26) is used as a baseline for comparison with other combinations of Kalman Filter implementations and model discretisation methods. Other approximations for the discrete time model can be found using, e.g., explicit Euler or the RK4 scheme [35].

4. Results and discussion

4.1. Test cases

Multiple combinations of KF implementations and model state propagation methods are tested and compared. The various cases, listed in Table 3, are compared with a baseline (BL) consisting of the linear explicit model from Section 3.4 and a standard KF.

Table 3
Model representation as list of interconnected components.

#	Kalman Filter	Model
BL	KF	Eq. (26) (exact)
1	KF	Eq. (24) (exp. Euler)
2	UKF	RCSimulator
3	UKF	Eq. (24) (RK4)
4	EnKF ($n_p = 50$)	RCSimulator,
5	EnKF ($n_p = 500$)	RCSimulator
6	EnKF ($n_p = 2000$)	RCSimulator
7	EnKF ($n_p = 5000$)	RCSimulator

Table 4
Freely estimated parameters for each case.

#	R_b	R_w	R_g	C_b	C_w	$\Delta\ell$
BL	0.097	0.114	0.136	1642k	1272k	-
1	0.100	0.118	0.134	1653k	1238k	0.22
2	0.101	0.119	0.133	1643k	1220k	−0.44
3	0.100	0.118	0.134	1651k	1228k	−0.11
4	0.075	0.103	0.133	1969k	1493k	121
5	0.093	0.122	0.201	1633k	1571k	32.5
6	0.080	0.100	0.217	1779k	1277k	12.5
7	0.076	0.091	0.190	1961k	1663k	4.51

4.2. Tuning

The covariance matrix for the measurement noise $\mathcal{V} = \text{diag}(0.1, 0.1)$ is obtained from the calibration data in $\mathcal{K}$. The process noise covariance $\mathcal{W}$ is often difficult to estimate, hence it is typically treated as a tuning parameter. By using the model validation results, i.e., testing the normality of the residuals as discussed in Section 4.4, $\mathcal{W} = \text{diag}(0.004, 0.002)$ is found by trial and error. An alternative solution is to include the elements of $\mathcal{W}$ in θ and let the optimisation algorithm $\mathcal{A}$ determine them [5], in accordance with Eq. (14).

The UKF and EnKF have some additional tuning parameters. The default UKF settings $\alpha = 10^{-3}$, $\kappa = 0$, $\beta = 2$ as suggested in [17] are used. For the EnKF, the only tuning parameter is the number of realisations n_p for which four different values are used, depending on the test cases given in Table 3.

4.3. Estimated parameters

The parameters are estimated for each test case by minimising Eq. (14) using the residuals and their covariance as obtained from the KF in each case. The results are shown in Table 4. The rightmost column lists the difference in log likelihood, $\Delta\ell$, between each case and the BL.

The estimated parameters $\hat{\theta}$ and corresponding relative log likelihood, $\Delta\ell$, for Cases 1–3 closely match the results computed for the baseline, which indicate that the UKF, combined with either RK4 discretisation or the RCSimulator, correctly estimates the residuals for evaluation of the likelihood function in Eq. (8).

The results for the EnKF Cases 4 to 7 differ significantly more from the BL case, as shown in Table 4. The parameters are different from the BL and also the log likelihood values are significantly higher than for the BL case. They tend towards the BL case as n_p is increased, but at $n_p = 5000$ in Case 7, there is still a significant difference. The EnKF is based on a Monte Carlo approximation of the state distribution $p(x, t)$. Hence, the computed residuals and covariance are also an approximation. This leads to a significant deviation from the BL results. The UKF is known to be exact for linear systems [18], hence no such deviation from the BL is observed for Cases 2 and 3, beyond some small deviations resulting from differences in state propagation approximations.
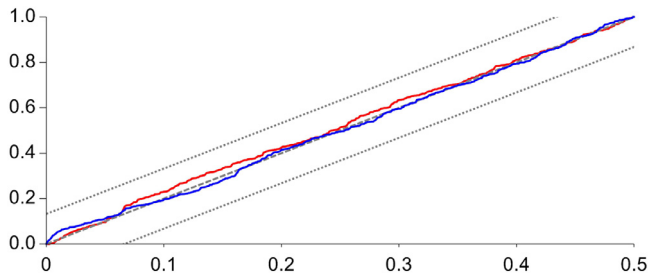
Fig. 4. Cumulative periodogram with 95% confidence bands for the residuals of the outputs T_b (red) and T_w (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Deterministic simulation. Measured values (grey) with T_b (red) and T_w (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
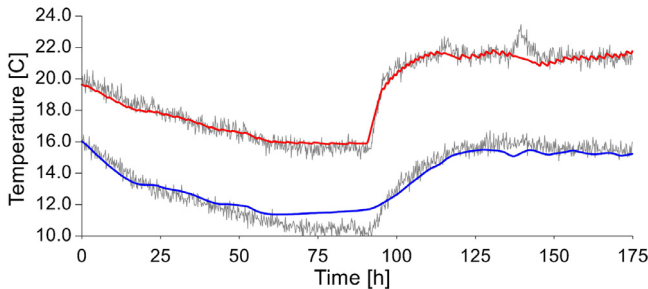
Table 5
Testing distribution of residuals for normality.

Output	T_b	T_w
Zero crossing	530	526
Kolmogorov–Smirnov	0.0232	0.0126

4.4. Model validation and normality tests

The use of the likelihood function for parameter estimation is based on the assumption of Gaussian distributed residuals, hence the residuals should be tested for normality in order to verify that assumption [5]. In the sequel, residuals from Case 2 are used, but near identical results are obtained for Cases 1, 3 and BL. Cases 4 to 7 are not considered further in this section. There are several tests that can be used to validate the normality assumption, such as the cumulative periodogram (CP) presented in Fig. 4. Other normality tests are the *zero-crossing count* (ZC) and the *Kolmogorov–Smirnov* (KS) test [3] as listed in Table 5. The *zero-crossing* test for $N = 1050$ samples gives a 95% confidence interval [3] of $(493, 556)$ which covers the test results for both outputs. Similarly, the KS test gives a critical value of 0.0417, which is higher than the score for both outputs. Finally, the CP in Fig. 4 indicates that the residuals are evenly distributed in the frequency domain, hence giving an approximately linear CP within the 95% confidence band [8,14,16,38]. After having tuned the process noise covariance matrix $\mathcal{W}$ by trial and error, these tests indicate that the residuals are well approximated by a Gaussian distribution.

Another method for validation of a model with estimated parameters is to perform a deterministic simulation using measured system inputs [3]. If the parameters are reasonable, the model predictions should approximately agree with the observed output measurements. The results of such a simulation are presented in Fig. 5. As shown, the simulation reproduces the system behaviour from the calibration data reasonably well, which further validates the model and the estimated parameter values.
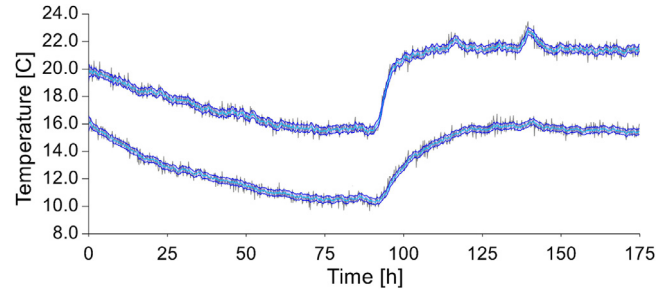

Fig. 6. Results from Case 2, using the UKF with RCSimulator. Observed measurements are plotted in grey, and estimated outputs with two sd error bands in blue: T_b (top) and T_w (bottom). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
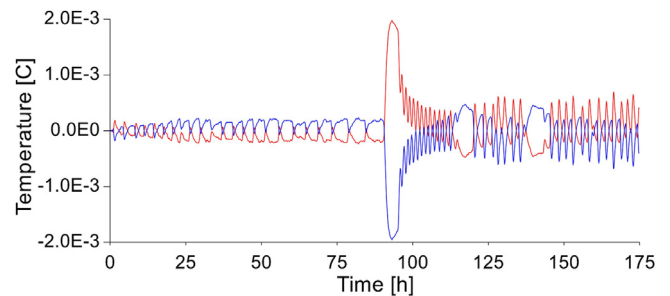

Fig. 7. Difference in output T_b between Case 1 (red) and Case 2 (blue) and the baseline linear KF with exact model discretisation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Comparing difference with baseline for each case.

#	$RMSE(\Delta\epsilon)$	$RMSE(\Delta\mathcal{E})$	Runtime
BL	–	–	1.4 ms
1	7.91×10^{-4}	9.49×10^{-6}	1.7 ms
2	7.79×10^{-4}	9.35×10^{-6}	10.4 ms
3	5.5×10^{-9}	5.77×10^{-13}	9.9 ms
4	1.39×10^{-1}	4.88×10^{-2}	0.092 s (0.078 s)
5	4.13×10^{-2}	1.56×10^{-2}	0.88 s (0.75 s)
6	2.14×10^{-2}	7.76×10^{-3}	3.5 s (3.1 s)
7	1.40×10^{-2}	4.70×10^{-3}	9.2 s (7.9 s)

4.5. Comparing Kalman filters

In this section, the results for all cases in Table 3 are compared in detail. First, Fig. 6 shows the estimated output for Case 2 plotted with an error band of two standard deviations together with the measured output. A visual inspection of Fig. 6 shows that the estimated output ± 2 standard deviations (sd) captures most of the variation in the measurements, indicating that the UKF correctly estimates the covariance of the estimated output.

Next, the root mean square error (RMSE) is used to compare the difference between each case and the BL, for both the residuals, $\Delta\epsilon$, and their covariance, $\Delta\mathcal{E}$. The results are presented in Table 6, together with the runtime of each case. For the EnKF, the computation time was measured both with and without (in brackets) computation of apriori and aposteriori covariance.

Several interesting observations can be made from Table 6. First, the quantified RMSE results show that Cases 1 to 3 give similar results to the BL case. In particular, Case 3 is near identical to the BL, with an RMSE around 10^{-9}. This result shows that the UKF gives near optimal estimates, with the deviations mostly explained by approximations introduced by the model discretisation method. Further, the results from EnKF differ significantly, even
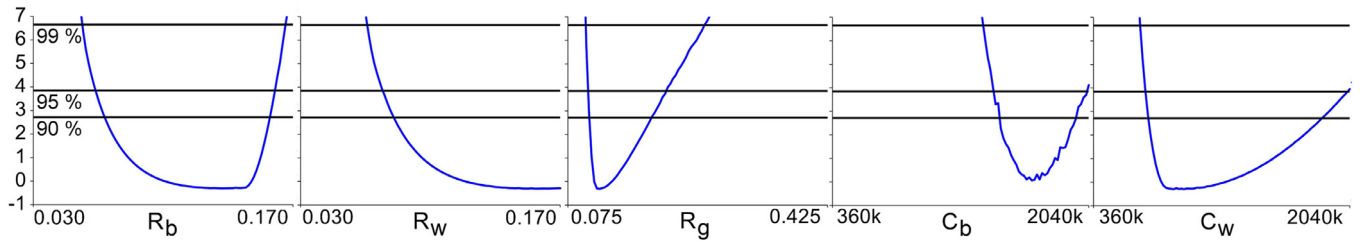
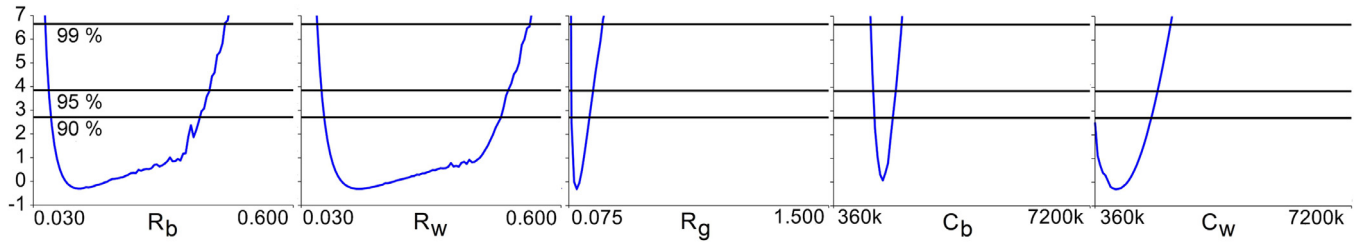Fig. 8. PL1 method for diagnosing parameter identifiability.

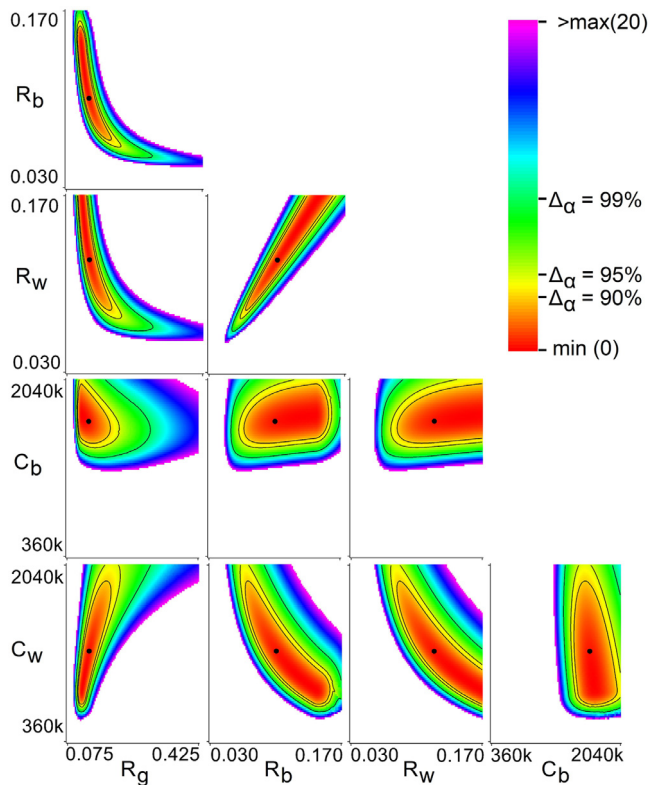Fig. 9. PL1 method with extended feasible region Θ.

Fig. 10. Heat map with isolines at confidence levels 90%, 95% and 99%, from the PL2 method. The freely optimised solution $\hat{\theta}$ is marked by a black dot in all plots. All plots cover the entire sub-region of Θ for their respective parameters. For reduced clutter, the axis labels are only included on the left-most/lower-most plots. The figure legend shows that the colour red indicates a relative likelihood close to 0, while purple indicates a relative log likelihood of 20. Values above 20 are not plotted, thus highlighting the most interesting region in the parameter space Θ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with $n_p = 5000$, from those obtained by KF and UKF, especially for low n_p.

To verify the implementation of the RCSimulator, Cases 1 and 2 are compared in Fig. 7. The error, compared to BL, in residuals of output T_b is similar for Cases 1 and 2, but interestingly with the opposite sign. This is reasonable since Case 1 uses explicit Eu-

ler discretisation of the continuous time model, while the RCSimulator, discussed in Section 3.3, uses implicit Euler approximation for the capacitance elements. The output T_w shows the same behaviour.

4.5.1. Computation time

Based on experience with the test case, the objective function is known to be well behaved in Θ, hence, a simple brute force algorithm is used to draw θ_i, and θ_j. With a resolution of 100 steps per parameter, the PL2 method requires 10,000 repeated executions of the parameter estimation algorithm per combination of parameters. With five parameters, there are ten different combinations of parameters which gives a total of 100,000 required executions. Hence, computational time for this method is significant. The evaluation of Eq. (22) can be efficiently parallelised, which reduces computation time. For the test case used in this paper the computation time per PL2 plot is around 30–60 minutes when the UKF is used with the RCSimulator on a six-core CPU. For large number of parameters, the overall computational load may be unpractical. However, if the objective function is smooth in Θ, a lower resolution plot may be sufficient to diagnose parameter identifiability and interdependence issues, or at least identify which parameter combinations may warrant further study with higher resolution. Further improvements in computation time may be achieved by applying faster maximum likelihood estimation algorithms [39].

The runtime of the EnKF, as shown in Table 6, are orders of magnitude slower than the UKF cases, while the UKF cases are only about six times slower than the standard KF cases. Observe also that the overhead of using the external simulator is only around 1 ms. Since the EnKF uses a MC sampling of the state distribution $p(x, t)$ in order to solve the Fokker–Planck Eq. (15), a larger number of realisations is required compared with the UKF [23]. The unscented transform used in UKF requires $2n_x + 1$ state realisations. Observe that the computational times in Table 6 are approximately linear in the number of realisations used for both EnKF and UKF. This is expected due to the relative similarities of the two KF implementations as indicated in Table 1. For a simple model such as the thermal network used here, with $n_x = 2$, it is not surprising that UKF far outperforms EnKF in terms of computational efficiency. EnKF was after all designed for large scale systems where n_x is very high.

In the sequel, only the UKF will be considered for further study. As discussed, the PL method requires a large number of repeated parameter optimisations, which in turn require an even larger
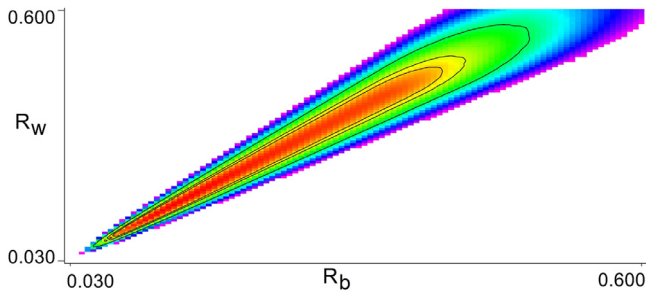
Fig. 11. PL2 topology of R_b vs R_w with extended feasible region Θ.

number of KF executions. Hence, runtime is of critical importance, which excludes the use of EnKF in this case.

4.6. Profile likelihood for Case 2

The theoretical foundation for the *profile likelihood* (PL) method [8,22] is discussed, and extended to simultaneously analysing two parameters in Section 2.4. In the sequel, both PL methods are applied to the R3C2 model from Section 3. Results are obtained using the configuration in Case 2 from Table 3. The log likelihood obtained by freely estimating all parameters is used as a reference, i.e., the plots present $\ell_{\text{PL}}(\theta_i) - \ell(\hat{\theta})$.

Fig. 8 shows the parameter profiles from the PL1 method in Eq. (19), with confidence thresholds at stated confidence levels from the χ^2 distribution with $n_{\text{df}} = 1$. The topological surfaces from the PL2 method in Eq. (22) are plotted as heat maps, with isolines at specified confidence levels from the χ^2 distribution with $n_{\text{df}} = 2$, in Fig. 10.

As discussed in Section 2.4, the definition of identifiability used in [22] requires the *likelihood-based confidence interval/region* to be bounded for a prescribed confidence level. Observe from Fig. 8 that four of the parameters have bounded profiles within the feasible region Θ. However, parameter R_w is indicated as practically non-identifiable, since the profile is below the prescribed thresholds in the positive direction. However, the PL analysis is computed subject to the constraint $\theta \in \Theta$. It is not clear from the results if extending the feasible values for R_w would result in a bounded PL also for this parameter. Hence, further examination of R_w by widening the feasible region Θ is required.

The results from PL2 in Fig. 10 show the same pattern of identifiability as for the one-dimensional PL plots. At first glance, the PL2 method confirms that parameter R_w is classified as practically non-identifiable, while R_b has a wide but bounded profile. However, the PL2 method provides additional insight into the parameter domain Θ. Observe from the plot for R_b vs R_w that the shape of the log likelihood function of parameter space $\ell(\Theta)$ projected onto the (R_b, R_w) plane indicates that there is interdependence between these two parameters. By comparing the PL1 plot for R_b to the PL2 plot for R_b vs R_w and projecting the PL2 results onto the R_b axis, it can be observed that the sharp bend in the PL1 Fig. 8 plot is actually a consequence of the constraint on R_w. A similar constraint artefact, a sharp bend near the lower end of the feasible range, can be seen in the PL1 plot for C_w. It is likely that the observed topology from Fig. 10 would be extended if the parameter domain Θ was wider, hence allowing optimal values for R_b to be obtained above the constrained profile observed in Fig. 8.

Further, the topology obtained from PL2 shows that the width of the profile from PL1 is significantly overestimated due to the parameter interdependence. Once either R_b or R_w is obtained, the profile of the other parameter is much narrower than what the PL1 method suggests.

4.6.1. Extending the parameter space

Next, the parameter space Θ is widened by letting the parameters extend to six times the nominal value, e.g., $\theta_{\max} = 6 \cdot \theta_0$. The PL1 results in Fig. 9 now show that all five parameters are identifiable, i.e. the likelihood profiles cross the confidence threshold in both directions. Further, the PL2 result for the parameter combination (R_b, R_w) in Fig. 11 shows that the topology observed in Fig. 10 indeed extends beyond the initial limited parameter space Θ.

5. Conclusion

In this paper, the *Unscented Kalman Filter* (UKF) and the *Ensemble Kalman Filter* (EnKF) have been compared for the purpose of estimating residuals and covariance for evaluation of the likelihood function. The state transition model was implemented in a non-differentiable external simulation tool, hence requiring a Kalman filter implementation other than the *Extended Kalman Filter* (EKF). The results from applying both filters to the parameter estimation problem show that the UKF outperforms the EnKF in both accuracy and computational time for this particular model. Since the UKF requires fewer realisations for few states, this result is expected. The EnKF was developed to handle large scale dynamic systems with a high number of model states and relatively few measurements. Since the underlying case study model is linear, the UKF is optimal, limited only by approximations to the state propagation method in the external simulation tool. For models with a larger number of states the EnKF may well be a better choice.

The use of a UKF allowed the likelihood function to be evaluated even though the thermal network building model was implemented as a non-differentiable external simulator. Comparing the results of the externally simulated component list model with simulations using the explicitly expressed linear equations showed that the results are near identical. While the use of the simulator and a UKF is about six times slower in computational time compared to the explicit model and a standard KF, the external model allows for simple manipulation of the component list model structure. This could potentially be used to automate the construction of thermal network model structures for the thermal behaviour of a specific building prior to parameter estimation.

The profile likelihood (PL) method was applied to the thermal network model to create one- and two-dimensional parameter profile plots. The PL2 plots were used to show that R_b and R_w are interdependent, which caused the PL1 method to overestimate the width of their respective profiles. Further, the interdependence of the parameters also skewed the results of the PL1 method due to the constraints of the feasible parameter region. Based on the PL2 plots, the parameter region was extended, which resulted in improved likelihood profiles also from the PL1 method. The combination of one- and two-dimensional likelihood profiles was shown to provide valuable insight into the parameter domain. These plots show that all of the parameters are identifiable, but with large confidence regions. This indicates lack of dynamic system excitation in the calibration data, which could be remedied by improved experimental design, e.g., use of Pseudo Random Binary Sequence as actuation.

References

[1] T. Berthou, P. Stabat, R. Salvazet, D. Marchio, Development and validation of a gray box model to predict thermal behavior of occupied office buildings, Energy Build. 74 (2014) 91–100, doi:10.1016/j.enbuild.2014.01.038.

[2] M. Killian, M. Kozek, Ten questions concerning model predictive control for energy efficient buildings, Build. Environ. 105 (2016) 403–412, doi:10.1016/j.buildenv.2016.05.034.

[3] R. Johansson, System Modeling and Identification, Information and System Sciences Series, Prentice Hall, 1993.

[4] L. Ljung, System Identification: Theory for the User, Prentice Hall information and System Sciences Series, Prentice Hall PTR, 1999.

[5] N.R. Kristensen, H. Madsen, S.B. Jørgensen, Parameter estimation in stochastic grey-box models, Automatica 40 (2) (2004) 225–237, doi:10.1016/j.automatica.2003.10.001.

[6] S.F. Fux, A. Ashouri, M.J. Benz, L. Guzzella, EKF based self-adaptive thermal model for a passive house, Energy Build. 68 (2014) 811–817, doi:10.1016/j.enbuild.2012.06.016.

[7] G. Reynders, J. Diriken, D. Saelens, Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals, Energy Build. 82 (2014) 263–274, doi:10.1016/j.enbuild.2014.07.025.

[8] A.-H. Deconinck, S. Roels, Is stochastic grey-box modelling suited for physical properties estimation of building components from on-site measurements? J. Build. Phys. 40 (5) (2017) 444–471, doi:10.1177/1744259116688384.

[9] A. Afram, F. Janabi-Sharifi, Gray-box modeling and validation of residential HVAC system for control system design, Appl. Energy 137 (2015) 134–150, doi:10.1016/j.apenergy.2014.10.026.

[10] L. Ljung, Prediction error estimation methods, Circuits Syst. Signal Process. 21 (1) (2002) 11–21, doi:10.1007/BF01211648.

[11] D. Di Ruscio, Combined deterministic and stochastic system identification and realization: DSR–a subspace approach based on observations, Model. Identif. Control 17 (3) (1996) 193–230, doi:10.4173/mic.1996.3.3.

[12] R. Ergon, D. Di Ruscio, Dynamic system calibration by system identification methods, in: Proceedings of the European Control Conference (ECC), IEEE, 1997, pp. 1556–1561.

[13] T. Bohlin, S.F. Graebe, Issues in nonlinear stochastic grey box identification, Int. J. Adapt. Control Signal Process. 9 (6) (1995) 465–490.

[14] H. Madsen, J. Holst, Estimation of continuous-time models for the heat dynamics of a building, Energy Build. 22 (1) (1995) 67–79, doi:10.1016/0378-7788(94)00904-X.

[15] N.R. Kristensen, H. Madsen, Continuous Time Stochastic Modelling-CTSM 2.3., Technical University of Denmark, Lyngby, Denmark, 2003.

[16] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy Build. 43 (7) (2011) 1511–1522, doi:10.1016/j.enbuild.2011.02.005.

[17] E.A. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in: Proceedings of the IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium. AS-SPCC., IEEE, 2000, pp. 153–158.

[18] S.J. Julier, J.K. Uhlmann, New extension of the Kalman filter to nonlinear systems, in: Proceedings of the Signal Processing, Sensor Fusion, and Target Recognition VI, 3068, International Society for Optics and Photonics, 1997, pp. 182–194.

[19] D. Simon, Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches, John Wiley & Sons, 2006.

[20] C.S. Ferrero, Q. Chai, M. Dueñas Díez, S.H. Amrani, B. Lie, Systematic analysis of parameter identifiability for improved fitting of a biological wastewater model to experimental data, Model. Identif. Control 27 (4) (2006) 219–238, doi:10.4173/mic.2006.4.2.

[21] S.A. Murphy, A.W. Van der Vaart, On profile likelihood, J. Am. Stat. Assoc. 95 (450) (2000) 449–465, doi:10.2307/2669386.

[22] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, J. Timmer, Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, Bioinformatics 25 (15) (2009) 1923–1929, doi:10.1093/bioinformatics/btp358.

[23] G. Evensen, The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Systems 29 (3) (2009) 83–104, doi:10.1109/MCS.2009.932223.

[24] M.J. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in: Advances in Optimization and Numerical Analysis, Springer, 1994, pp. 51–67.

[25] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Dover Publications, Inc, 1970.

[26] E.L. Lehmann, G. Casella, Theory of Point Estimation, Springer Science & Business Media, 2006.

[27] R. Juhl, J.K. Møller, H. Madsen, ctsmr-Continuous Time Stochastic Modeling in R. preprint arXiv:1606.00242, 2016.

[28] R. Juhl, J.K. Møller, J.B. Jørgensen, H. Madsen, Modeling and prediction using stochastic differential equations, in: Prediction Methods for Blood Glucose Concentration, Springer, 2016, pp. 183–209.

[29] D. Venzon, S. Moolgavkar, A method for computing profile-likelihood-based confidence intervals, Appl. Stat. (1988) 87–94, doi:10.2307/2347496.

[30] T. Maiwald, J. Timmer, Dynamical modeling and multi-experiment fitting with potterswheel, Bioinformatics 24 (18) (2008) 2037–2043, doi:10.1093/bioinformatics/btn350.

[31] S.S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Math. Stat. 9 (1) (1938) 60–62, doi:10.1214/aoms/1177732360.

[32] W.Q. Meeker, L.A. Escobar, Teaching about approximate confidence regions based on maximum likelihood estimation, Am. Stat. 49 (1) (1995) 48–53, doi:10.2307/2684811.

[33] L.W. Nagel, SPICE2: a computer program to simulate semiconductor circuits, Ph.D. dissertation, University of California at Berkeley, 1975.

[34] J.W. Nilsson, Electric Circuits, Pearson Education India, 2008.

[35] C. Runge, Ueber die numerische auflösung von differentialgleichungen, Mathematische Annalen 46 (2) (1895) 167–178, doi:10.1007/BF01446807.

[36] R.A. DeCarlo, Linear Systems: A State Variable Approach with Numerical Implementation, Prentice-Hall, Inc., 1989.

[37] C. Van Loan, Computing integrals involving the matrix exponential, IEEE Trans. Autom. Control 23 (3) (1978) 395–404, doi:10.1109/TAC.1978.1101743.

[38] H. Madsen, Time Series Analysis, Chapman and Hall/CRC, 2007.

[39] D. Boiroux, R. Juhl, H. Madsen, J.B. Jørgensen, An efficient UD-based algorithm for the computation of maximum likelihood sensitivity of continuous-discrete systems, in: Proceedings of the IEEE Fifty fifth Conference on Decision and Control (CDC), IEEE, 2016, pp. 3048–3053.

Article C

Sensor placement and parameter identifiability in grey-box models of building thermal behaviour

Authors O. M. Brastein, R. Sharma and N.-O. Skeie

Sensor placement and parameter identifiability in grey-box models of building thermal behaviour

O. M. Brastein*, R. Sharma, N.-O. Skeie

Department of Electrical Engineering, Information Technology and Cybernetics
University of South-Eastern Norway, N-3918 Porsgrunn,
*(ole.m.brastein@usn.no)

Abstract

Building Energy Management systems can reduce energy consumption for space heating in existing buildings, by utilising Model Predictive Control. In such applications, good models of building thermal behaviour is important. A popular method for creating such models is creating Thermal networks, based cognitively on naive physical information about the building thermal behaviour. Such models have lumped parameters which must be calibrated from measured temperatures and weather conditions. Since the parameters are calibrated, it is important to study the identifiability of the parameters, prior to analysing them as physical constants derived from the building structure. By utilising a statistically founded parameter estimation method based on maximising the likelihood function, identifiability analysis can be performed using the Profile Likelihood method. In this paper, the effect of different sensor locations with respect to the buildings physical properties is studied by utilising likelihood profiles for identifiability analysis. The extended 2D profile likelihood method is used to compute two-dimensional profiles which allows diagnosing parameter inter-dependence, in addition to analysing the identifiability. The 2D profiles are compared with confidence regions computed based on the Hessian.

Keywords: building energy management systems, thermal behavior, parameter estimation, parameter identifiability, Profile Likelihood

1 Introductions

1.1 Background

A significant portion of the worlds total energy production is consumed by heating and cooling of buildings (Perera et al., 2014). Building Energy Management Systems (BEMS) is therefore an important part of the ongoing effort to reduce anthropogenic CO_2 emissions. In particular, Model Predictive Control (MPC) has been shown to reduce energy consumption in buildings by utilizing models to predict the thermal behaviour of a building (Fux et al., 2014; Killian and Kozek, 2016). Hence, the development of models of building thermal behaviour has received considerable interest by the scientific community in recent years.

1.2 Previous work

A common approach to the modelling of building thermal behaviour is the use of *thermal network* models (Berthou et al., 2014; Reynders et al., 2014). These models are often described using electric analogues Resistor-Capacitor (RC) circuits. Based on a *naive* understanding of the thermodynamics involved, these RC circuits constitute *simplified lumped parameter* models. Parameters are estimated from measurements of temperature inside the building, weather conditions and input power consumed for space heating. As simplified models based on both physical insight and measurement data, thermal network models constitute a compromise between fully physics based white-box and purely data-driven black-box models. This type of model, often called grey-box models, allows use of prior knowledge of the system while also allowing calibration of parameters to adapt the model to a particular building. This approach offers improved prediction accuracy while also allowing use of prior physical information to be injected into the model (Madsen and Holst, 1995; Bacher and Madsen, 2011; Kristensen et al., 2004).

Since the model structure is designed based on knowledge of a particular building, it is often assumed that the parameters are determined by the physical properties of that building. However, since the parameters are identified from data, this assumption needs to be verified in the context of *parameter identifiability* (Reynders et al., 2014; Deconinck and Roels, 2017; Ferrero et al., 2006). In particular, testing of practical identifiability (Raue et al., 2009), i.e., if sufficient dynamic information about the underlying system is contained in the calibration data (Ferrero et al., 2006), is of importance.

1.3 Overview of paper

Since weather is part of the experimental conditions, and, typically, the acceptable range of indoor temperatures, as well as input heater power, is limited, model calibration must usually be performed on sub-optimal data. One element which, to some degree, is open to experimentation is the location of the sensors. As simplified models, thermal networks reduce large indoor spaces and objects, such as the building envelope, to point nodes in the RC circuit. How these nodes correspond to the physical building is

determined by the sensor location. In this work, we study different sensor placements in an experimental building to show how sensor location, with respect to the physical properties of the building, affects the dynamic information contained in the data and hence the practical identifiability of model parameters. Parameter identifiability is analysed using the *Profile Likelihood* method, both in the single parameter dimension and in two parameter dimensions by projecting the profile onto a plane in parameter space. The latter method allows improved insight into the parameter domain, including analysing parameter inter-dependence and the effects of a constrained parameter space.

2 Theoretical basis

The framework presented in (Kristensen et al., 2004), named Continuous Time Stochastic Modelling (CTSM), is a statistically well founded approach to parameter estimation. The theoretical basis is briefly summarised below. For a more detailed discussion see (Kristensen et al., 2004). Consider the estimation problem:

$$\hat{\theta} = \arg\min_{\theta} g\left(\theta; \mathcal{M}, \mathcal{K}, \mathcal{A}\right) \qquad (1)$$
$$s.t. \ \theta \in \Theta$$

Here, $\mathcal{M}$ is a predetermined model structure parametrised by $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^{n_\theta}$ is a set of feasible values for the model parameters that form inequality constraints for the optimisation problem in Eq. (1). Parameters in θ are varied over the feasible set Θ by a numerical optimisation algorithm $\mathcal{A}$. The experimental conditions $\mathcal{K}$ include measurements for the continuous time input $u_t \in \mathbb{R}^{n_u}$ and output $y_t \in \mathbb{R}^{n_y}$. The corresponding ordered sequences of discrete time measurements u_k and y_k taken from the system $\mathcal{S}$ are $y_{[N]} = [y_0, y_1, \ldots, y_N]$ and $u_{[N]} = [u_0, u_1, \ldots, u_N]$, where the integer subscripts $k = 0, 1, \ldots, N$ denote the discrete time sampling instants, and the subscript enclosed in $[\cdot]$ is used to indicate an ordered sequence.

The likelihood function, i.e., the probability of observing the measurement sequence $y_{[N]}$ when θ and $\mathcal{M}$ are given, is defined:

$$L\left(\theta; y_{[N]}, \mathcal{M}\right) = p\left(y_{[N]} | \theta, \mathcal{M}\right) \qquad (2)$$

By assuming that the residuals are Normal distributed, and applying the product rule to expand the probability in Eq. (2), we obtain (Kristensen et al., 2004):

$$L\left(\theta; y_{[N]}\right) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\varepsilon_k^T \mathscr{E}_{k|k-1}^{-1} \varepsilon_k\right)}{\sqrt{\det\left(\mathscr{E}_{k|k-1}\right)} \left(\sqrt{2\pi}\right)^{n_y}} \right) p\left(y_0 | \theta\right) \qquad (3)$$

The quantities $\hat{y}_{k|k-1}$, ε_k and $\mathscr{E}_{k|k-1}$, which can be obtained using a Kalman Filter (KF) (Kristensen et al., 2004), is needed for evaluation of the multivariate Gaussian in Eq. (3). By taking the negative logarithm, and eliminating the factor $\frac{1}{2}$, the result $\ell\left(\theta\right) = -\ln L\left(\theta\right)$, where dependency on $y_{[N]}$ and $\mathcal{M}$ is omitted for simplicity, can be used as the objective g in Eq. (1).

2.1 Profile likelihood

Since the model structure $\mathcal{M}$ is a representation of a system $\mathcal{S}$, it is often assumed that $\mathcal{S} \in \mathcal{M}\left(\Theta\right)$ and that consequently there exists a true parameter vector θ^* such that $\mathcal{M}\left(\theta^*\right) = \mathcal{S}$. However, this is rarely the case, especially for simplified grey-box models based on a *naive* physical understanding of the system $\mathcal{S}$. Typically, the estimate $\hat{\theta}$ depends on several factors, such as the amount of dynamic information in $\mathcal{K}$, the choice of objective function g, and to some extent on the optimisation algorithm $\mathcal{A}$. Hence, prior to interpretation of parameters as physical constants of $\mathcal{S}$, it is necessary to perform an identifiability analysis. Since the parameters are estimated using the Likelihood function, the *Profile Likelihood* (PL) method (Raue et al., 2009; Deconinck and Roels, 2017) is a natural choice. The *likelihood profile* $\ell_{\text{PL}}\left(\theta_i\right)$ is defined as the minimum log likelihood for θ_i when the remaining parameters are freely optimised (Raue et al., 2009; Venzon and Moolgavkar, 1988):

$$\ell_{\text{PL}}\left(\theta_i\right) = \min_{\theta_{j \neq i}} g\left(\theta_{j \neq i}; \mathcal{M}, \mathcal{K}, \theta_i\right) \qquad (4)$$

Values of θ_i must be chosen prior to optimising the remaining $\theta_{j \neq i}$ (Raue et al., 2009). The resulting likelihood profile can be plotted as a function of θ_i and subsequently analysed according to the definitions of structural and practical identifiability for *likelihood-based confidence intervals* (Deconinck and Roels, 2017). The likelihood-based confidence interval obtains a confidence region by applying a *threshold* to the likelihood function (Raue et al., 2009; Venzon and Moolgavkar, 1988). Let

$$\left\{\theta : \ell\left(\theta\right) - \ell\left(\hat{\theta}\right) < \Delta_\alpha\right\} \quad , \quad \Delta_\alpha = \chi^2\left(\alpha, n_{\text{df}}\right) \qquad (5)$$

where $\hat{\theta}$ is a freely estimated, presumed optimal, parameter vector, and the threshold Δ_α is the α percentile of the χ^2-distribution with n_{df} degrees of freedom.

Profile likelihood in two parameter dimensions

By freely estimating all but one parameter, the PL method essentially projects the n_θ dimensional space Θ onto the single parameter θ_i. This projection is known to overestimate the width of the likelihood-based confidence interval if there are inter-dependent parameters. A step towards remedying this issue is to modify the PL method to hold out *two* parameters (PL2) rather than one, i.e.;

$$\ell_{\text{PL2}}\left(\theta_i, \theta_j\right) = \min_{\theta_{k \neq i,j}} g\left(\theta_{k \neq i,j}; \mathcal{M}, \mathcal{K}, \theta_i, \theta_j\right) \qquad (6)$$

This projects the parameter space Θ onto the plane $\Theta_{i,j} = \left(\theta_i, \theta_j\right)$ s.t. $\theta_i, \theta_j \in \Theta$. In addition identifiability issues, these profiles can also diagnose *parameter interdependence* by inspecting the shape of the confidence regions. The resulting two-dimensional profiles can be analysed similarly to the one-dimensional profiles (Raue et al., 2009), using the definition in Eq. (5). These profiles are computed for all possible combinations of parameters. A *confidence region* in the $\Theta_{i,j}$ plane is obtained by applying the Δ_α threshold. Observe that since $\hat{\theta}$ has n_θ free

Figure 1. PL2 improved *warm start* algorithm

parameters while the PL2 estimate has $n_\theta - 2$, this gives $n_{df} = 2$ for the computation of Δ_α from the χ^2-distribution in Eq. (5). The free estimate $\hat{\theta}$ may with advantage be chosen as the minimum $\ell_{PL2}(\theta_i, \theta_j)$ obtained from all profiles. This search procedure approximates, since it is subject to the brute force discretisation performed in PL2, a free optimisation of all parameters using the already computed ℓ_{PL2} results. Since the PL2 profiles covers the entire parameter space Θ, this procedure is less affected by local minima than a direct numerical optimisation. Parameter identifiability is obtained if the region is bounded in all directions and the size and shape of this region determines the accuracy of the parameter estimates. If the region contains an unbounded equipotential *valley* in the log likelihood space, the parameter is considered structurally non-identifiable. If the profile has a well defined minima, but is unbounded in one direction, i.e., the log likelihood is below the Δ_α threshold, this indicates a practically non-identifiable parameter (Raue et al., 2009).

Implementation and computation time

In (Brastein et al., 2019), a brute force method was used, running individual optimisations for each predetermined combination of θ_i, θ_j, each iteration starting from the nominal parameter vector θ_0. Here, the profiles are constructed by a set of chained optimisations where each new point uses a previously optimised $\hat{\theta}_{k \neq i,j}$ from a near-by point in $\Theta_{i,j}$ as a *warm start*, working from the centre of the plane $\Theta_{i,j}$ towards the edges. This process is illustrated in Fig. 1. This modification reduced the computation time by approximately 4-10 times, since for each computation of $\ell_{PL2}(\theta_i, \theta_j)$, the initial guess for the free parameters are taken from a near-by, previously optimised, solution and hence are already close to optimal.

2.2 Parameter estimation uncertainty

An estimate of the uncertainty of the estimated parameters can be obtained by computing the covariance of the estimated parameters $\Sigma = 2H^{-1}$ where $H = \nabla^T \nabla \ell(\theta)|_{\theta = \hat{\theta}}$ is the *Hessian* of $\ell(\theta)$, whose elements are approximated as (Kristensen et al., 2004; Raue et al., 2009):

$$h_{i,j} \approx \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta) \right) \Bigg|_{\theta = \hat{\theta}} \tag{7}$$

The partial derivatives of $\ell(\theta)$ can be numerically obtained using *the central difference approximation*. From the covariance matrix, asymptotic *point-wise* confidence limits on the estimated parameters can be computed (Raue et al., 2009)

$$\hat{\theta}_i \pm \sqrt{\chi^2(\alpha, n_{df}) \Sigma_{i,i}} \tag{8}$$

where for *point-wise* intervals $n_{df} = 1$. Alternatively, confidence ellipsoids of dimension n_θ as a set of θ can be defined from the inequality:

$$\left\{ \theta : (\theta - \hat{\theta})^T \Sigma^{-1} (\theta - \hat{\theta}) \leqslant \Delta_\alpha \right\} \quad , \quad \Delta_\alpha = \chi^2(\alpha, n_{df}) \tag{9}$$

where the scale of the ellipsoid is determined by the factor Δ_α computed as in Section 2.1 (Johnson and Wichern, 2007). Given that the covariance matrix is symmetric and positive definite, the boundary of an ellipsoid can be obtained by the Cholesky decomposition $\Sigma = LL^T$, hence (Press et al., 1992):

$$(\theta - \hat{\theta})^T \Sigma^{-1} (\theta - \hat{\theta}) = \Delta \Rightarrow |L^{-1}(\theta - \hat{\theta})|^2 = \Delta_\alpha \tag{10}$$

Next, suppose x is a point on a unit hypersphere, then the ellipsoid boundary is obtained by the affine transformation

$$\theta = \hat{\theta} + \sqrt{\Delta} L x \tag{11}$$

Elliptic regions in the plane $\Theta_{i,j}$ can be computed by projecting the n_θ dimensional ellipsoid onto $\Theta_{i,j}$. With the χ^2 confidence bound given with $n_{df} = 2$, assuming all uncertainty on the parameters in the plane $\Theta_{i,j}$, these elliptic confidence regions are comparable to the *PL2* method presented in Section 2.1. Observe that the confidence region based on applying a threshold on the likelihood function as in Eqs. (4) and (6) are often considered superior to the Hessian method, since the Hessian assumes symmetric distributions and therefore cannot be used to identify practical identifiability (Raue et al., 2009). However, the Hessian approach is much faster to compute and gives a reasonable estimate of the estimated parameter uncertainty if the parameters are approximately Normal distributed.

3 Experimental setup

3.1 Experimental building

The experimental building, which is located at Campus Porsgrunn of the University of South-Eastern Norway (USN), is shown in Fig. 2. The building is constructed with three different types of walls. As shown in Fig. 2, the North wall is constructed using materials with high insulation quality, which is typically used in modern sustainable buildings. The South wall is constructed using traditional building materials, with lower thermal insulation capabilities.
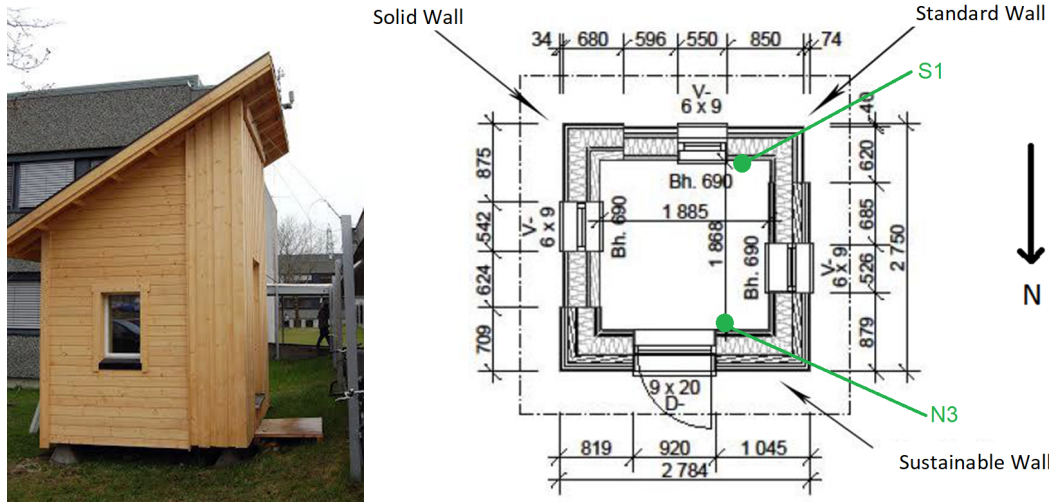
Figure 2. The experimental building has walls constructed using different techniques for insulation. Sensors are located on all walls at different height above floor, and in different insulation layers. The sensors used in this project, N3 and S1, are measuring the wall temperature inside the building on the north and south wall, respectively.
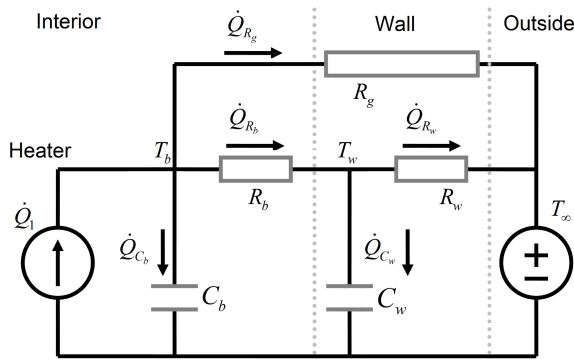

Figure 3. RC circuit model of the building.

3.2 Model

Figure 3 shows a possible model structure, which was developed to approximate the thermal behaviour of the experimental building, partially based on the R4C2 model presented in (Berthou et al., 2014). The RC circuit consists of five components: the thermal resistance between room air and wall R_b, the building envelope R_w, and the thermal resistance of windows and doors R_g. The two capacitances C_b and C_w represent the thermal capacitance of the building interior and envelope, respectively. The model has two outputs: the room temperature T_b and the wall surface temperature T_w, and two inputs: the consumed power by an electric heating element $\dot{Q}$ and the outside temperature T_∞. The parameter vector θ holds the value of each of the five components. By applying Kirchoff's balance laws to the circuit, the model can be expressed as a linear stochastic differential equation

$$\frac{dx}{dt} = Ax_t + Bu_t + w_t \quad (12)$$

$$y_t = x_t + v_t \quad (13)$$

Table 1. Nominal parameter values and min/max limits for resistances [K/W] and capacitances [J/K].

	R_g	R_b	R_w	C_b	C_w
θ_0	0.15	0.20	0.30	$1000k$	$200k$
θ_{min}	0.03	0.03	0.03	$800k$	$1k$
θ_{max}	0.25	2.00	2.00	$1800k$	$1000k$

where

$$x_t = \begin{bmatrix} T_b \\ T_w \end{bmatrix}, u_t = \begin{bmatrix} \dot{Q} \\ T_\infty \end{bmatrix}, B = \begin{bmatrix} \frac{1}{C_b} & \frac{1}{C_b R_g} \\ 0 & \frac{1}{C_w R_w} \end{bmatrix}$$

$$A = \begin{bmatrix} -\frac{1}{C_b R_b} - \frac{1}{C_b R_g} & \frac{1}{C_b R_b} \\ \frac{1}{C_w R_b} & -\frac{1}{C_w R_b} - \frac{1}{C_w R_w} \end{bmatrix}$$

and $w_t \sim N(0, W)$ is the process noise (model error), W is the spectral density of w_t. All states are measurable, hence Eq. (13) with measurement noise $v_t \sim N(0, V)$. Observe that the model equations are expressed in continuous time, and discretised by the estimation software using a Runge-Kutta 4th (Runge, 1895) order approximation. Observe also that while the model is linear, the algorithm is not restricted to linear models. The choice of Kalman Filter implementation is determined by the type of model being used (Brastein et al., 2019).

Table 1 lists a set of experimentally obtained nominal parameters, which are used as initial guesses for model calibration, and min/max limits which corresponds to the bounds of the constrained parameter space Θ.

3.3 Calibration data

Figure 4 shows a set of calibration data, which consist of four temperature measurements and one measurement of supplied input power. The data was recorded in February 2018. Originally, the data was collected at 1 minute intervals but has been downsampled to 30 min time-step, by
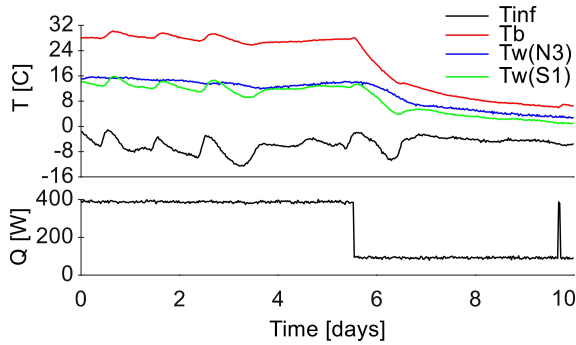
Figure 4. Data recorded from sensor at different locations in the building

Table 2. Estimated covariance matrices with corresponding KS test result (critical value for 95% conf. is 0.062) .

#	$\mathscr{W}_{T_{\mathrm{b}}}^{\frac{1}{2}}$	$\mathscr{W}_{T_{\mathrm{w}}}^{\frac{1}{2}}$	$\mathscr{V}_{T_{\mathrm{b}}}^{\frac{1}{2}}$	$\mathscr{V}_{T_{\mathrm{w}}}^{\frac{1}{2}}$	$KS_{T_{\mathrm{b}}}$	$KS_{T_{\mathrm{w}}}$
S1	0.115	0.104	0.028	0.037	0.054	0.042
N3	0.117	0.077	0.019	0.145	0.046	0.035

extracting every 30th measurement. Two of the temperatures correspond to model state T_{b} and the model input outdoor temperature T_{∞}. The remaining two measurements correspond to different alternative sensor locations for T_{w}, one on the north wall (sensor N3) and one on the south wall (sensor S1). Figure 4 shows that there is significant differences between these two measurements in dynamic content, due to the different construction materials used in the North and South wall, which will lead to differences in the identifiability analysis of the estimated parameters. In the sequel, two different cases S1 and N3 are analysed, distinguished by the choice of reference measurement for the output T_{w}.

Optimisation algorithm

In (Brastein et al., 2019) COBYLA (Powell, 1994), based on linear approximations, was used as the optimisation algorithm $\mathscr{A}$ in Eq. (1). In this work, further experimentation with other optimisation algorithms showed that a quadratic approximation algorithm, such as BOBYQA (Powell, 2009), gives significantly faster convergence, by approximately a factor of 5, as well as more consistent results by improved ability to avoid local minima. BOBYQA is therefore used in the sequel.

4 Results and discussion

A requirement for using Kalman Filters to obtain residuals for subsequent evaluation of the likelihood function in Eq. (3) is obtaining reasonable estimates for process and noise covariance matrices, respectively $\mathscr{W}$ and $\mathscr{V}$. In (Brastein et al., 2019) $\mathscr{V}$ was obtained from data, while $\mathscr{W}$ was found by manual experimentation. A better approach is to estimate them from data, by including them in θ.

In order to reduce the number of free parameters, both covariance matrices are assumed diagonal. Further, the
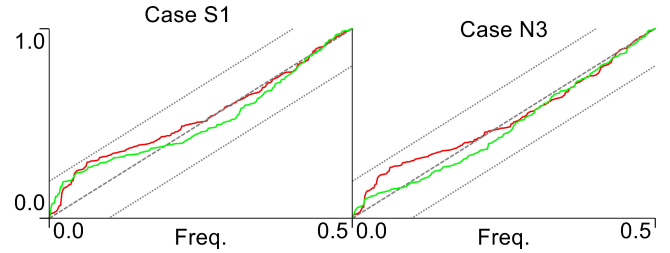
Figure 5. CP diagram of residuals for outputs T_{b}(red) and T_{w}(green)
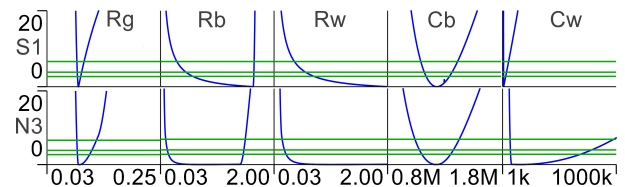
Figure 6. PL results for Cases S1 and N3. Green lines indicate in increasing order 90%, 95% and 99% confidence limits.

square root of the diagonal elements are added to the parameter vector θ with some appropriate bounds in Θ and subsequently estimated by numerical optimisation of Eq. (1). The resulting covariances are shown in Table 2. Observe that the results corresponding to the state/measurement T_{b} are similar for both cases, while the results for T_{w} differ significantly. This is expected due to the differences in noise characteristics and dynamic information content in the data collected from the two sensors. For Case S1 the estimates for $\mathscr{W}$ is similar for both states, where as for Case N3 the differences between measurements of T_{b} and T_{w} results in different estimates for the corresponding elements in $\mathscr{W}$.

The residuals obtained after optimising all parameters must be analysed for normality, in order to justify the use of the multivariate Gaussian in Eq. (3) for evaluation of the likelihood function (Kristensen et al., 2004). Figure 5 shows a *cumulative periodogram* (CP), with 95% confidence bounds obtained from the Kolmogorov-Smirnov criterion (Madsen, 2007; Madsen and Holst, 1995; Deconinck and Roels, 2017; Bacher and Madsen, 2011). The CP plot shows that the residuals are well approximated by a normal distribution. Additionally, the Kolmogorov-Smirnov normality-test results are listed in Table 2. After calibration of the parameters, including the noise covariance matrices, residuals are found to pass the normality tests.

4.1 Profile likelihood

Once the covariance parameters have been determined, the remainder of this paper is focused on analysing the parameter space Θ by use of the Profile Likelihood (PL) (Raue et al., 2009) method, first in a single parameter dimension, and next in two parameter dimensions. The PL results in Fig. 6 show, as expected, that some of the parameters have narrower profiles for the S1 Case compared with
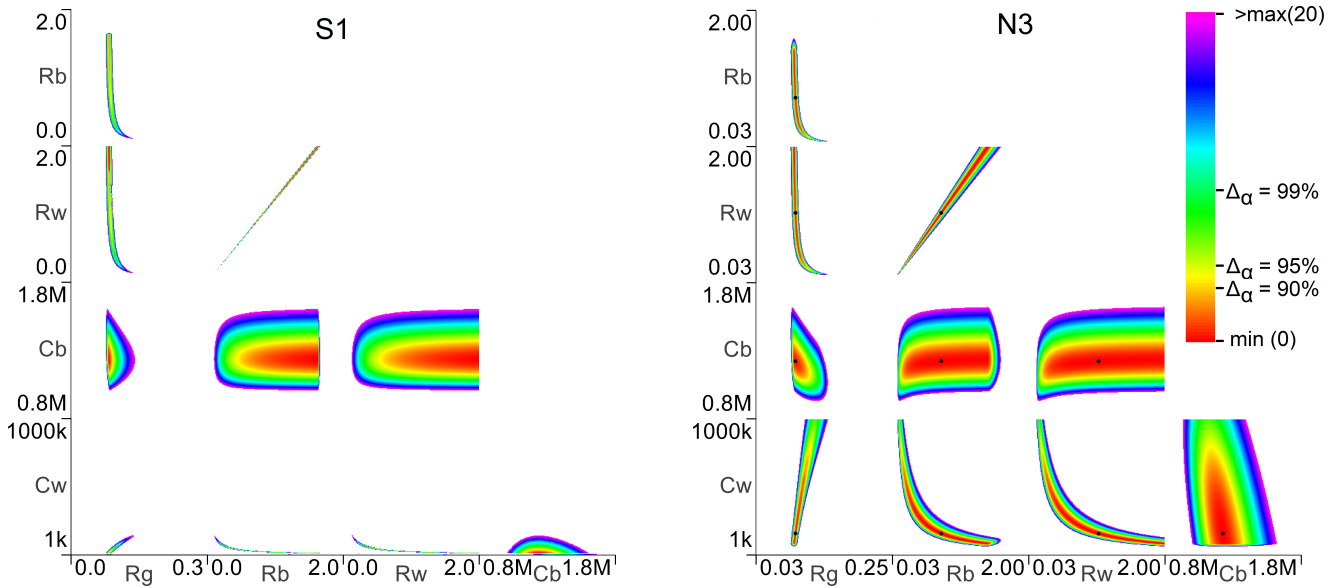
Figure 7. PL2 results for Case S1 (left) and Case N3 (right). Confidence limits, base on the χ^2 distribution with 2 degrees of freedom is indicate on the figure legend to the right.

the N3 Case. Since the excitation in T_w is much larger for Case S1, it is expected that the identifiability analysis reflects this by computing tighter confidence regions for the parameters most affected by T_w. Observe in particular how the profiles for C_w and R_g indicates considerably improved identifiability of these parameters for Case S1. The profiles for parameter C_b is almost identical, which is expected, since this parameter is not influenced by T_w.

A second observation from Fig. 6 is that R_w is diagnosed as practically non-identifiable, since the profile extend towards infinity in the positive direction. Observe also that R_b follows a similar trend, but with an abrupt break in the profile, which leads to a bounded profile for R_b. However, if parameters R_b and R_w are inter-dependent, the projection of the likelihood function for the parameter space Θ onto R_b will be affected by R_w, and subsequently by the constraint imposed by Θ. This type of constraints, in the presence of parameter inter-dependence, is known to produce such breaks in the computed profiles (Brastein et al., 2019), as discussed in Section 2.1.

4.2 Profile likelihood in 2D

Next, the two-dimensional profile likelihood (PL2) method is applied in order to investigate parameter inter-dependency. The result is shown in Fig. 7. By projecting the ML function in Eq. (3) onto a plane of two parameters, rather then a single parameter axis as in PL, it becomes possible to diagnose parameter identifiability by observing the shape of the confidence regions. First, observe that the PL2 results show a similar improvement in identifiability for Case S1 over Case N3 for the same parameters. The confidence regions for C_w and R_g are significantly reduced for Case S1, whereas the region for C_b is similar for both cases. Hence, the results of the PL analysis is confirmed

by the PL2 results.

Further, the PL2 results also show that R_b and R_w are highly inter-dependent, a fact which was not easily observed in Fig. 6. The projected topology of these two parameters shows a near linear relationship between them. This explains why the PL profile for R_b contains an abrupt break, caused by the constraint on R_w and their inter-dependence. In this context, it is interesting to consider whether this lack of identifiability for parameters R_b and R_w are of practical, i.e. related to information content in data, or structural nature. Parameter inter-dependency is clearly caused by the model structure, not the data. However, the PL2 profile shows that while the parameters are linearly dependant, e.g. $R_b = cR_w$, neither parameter is identifiable, since the profile is unbounded in one direction in Θ. Hence, it is accurate to claim that these two parameters are practically non-identifiable, but also that there is a structural problem of parameter inter-dependency. The latter may be eliminated by re-parametrising the model, say, by introducing the relationship $R_b = cR_w$ with the constant c pre-determined based on Fig. 7. However, there is no physical reason to assume that the thermal resistance between the inside wall and the building interior should be depending linearly on the thermal resistance of the wall itself, hence this modification of the model equations seems somewhat *ad hoc*. A better alternative is to modify the RC circuit model *structure* such that the parameter inter-dependency is resolved.

4.3 Reduced order model for Case S1

From Fig. 6 and 7 for Case S1 it appears that the value of C_w tends towards zero as R_b and R_w increases. This could indicate that removing C_w, and replacing the state T_w by a measurement $T_w = \frac{T_\infty R_b + T_b R_w}{R_b + R_w}$, is an appropriate modification. However, after calibrating the reduced model, the re-
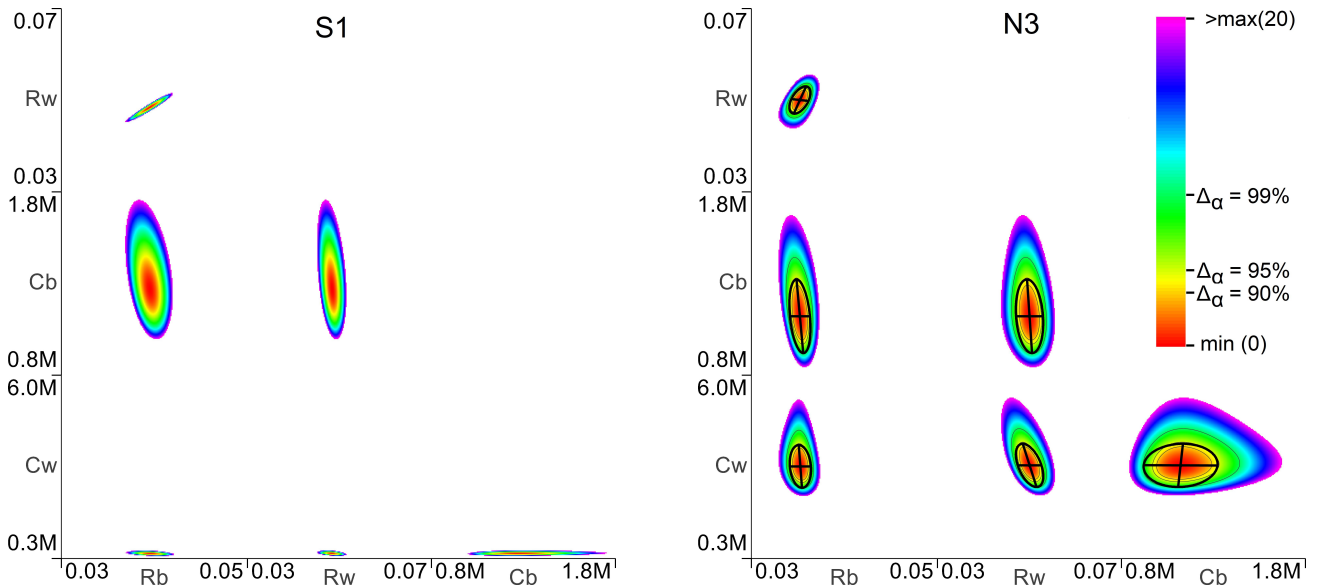
Figure 8. PL2 results after removing R_g from the model, for Case S1 (left) and Case N3 (right). Confidence limits, base on the χ^2 distribution with 2 degrees of freedom is indicate on the figure legend to the right. Likelihood threshold isolines and a 95% confidence error ellipse boundary, based on the inverse Hessian, has been added to the plot for N3.

sidual analysis for output T_w, based on a CP diagram, does not support the normality assumption, which in turn indicates that the model order is to low. Hence, removing C_w is not an acceptable modification to the model structure.

4.4 Parameter inter-dependence

Observe from Fig. 6 and 7 that for Case N3 there is a significant flat region in the profiles of both R_b and R_w. Table 3 contains a selection of values from within this region, which has been computed by keeping R_b constant while optimising the remaining parameters. This experiment shows that by varying the parameters within this optimal region, the total thermal resistance between building interior and the outside temperature, $R_{tot} = R_g || (R_b + R_w)$, where $||$ indicates a parallel connection of resistors, is constant. Also, the time constant for the wall capacitor $\tau_w = (C_w R_b || R_w)$ is approximately constant for the same experiments. Since the total resistance $R_{tot} < R_g$ it follows that R_b and R_w can grow large, by compensating with R_g and C_w, without affecting the models predictions. Combined with the PL2 analysis results, this indicates an *over-parametrised* model.

4.5 PL analysis of model without Rg

A natural next step is to reduce the number of parameters by removing the presumed redundant parameter R_g from

Table 3. Optimised values of R_w, R_g and C_w with fixed values of R_b from within the flat region observed in the PL results.

R_b	R_w	R_g	C_w	R_{tot}	τ_w
0.700	0.993	0.093	163k	0.088	67.0k
0.900	1.277	0.092	126k	0.088	66.5k
1.300	1.844	0.091	88k	0.088	67.0k

the model circuit. Repeating the PL2 analysis of the model with R_g removed gives the plots shown in Fig. 8. The profiles for all parameter combinations are now approximately elliptical, which indicates parameter independence. Observe also from Fig. 8 that the min/max limits which constitutes the bounds Θ has been changed to comply with the reduced model structure. Further, all parameters now have bounded profiles, which indicates identifiability. By comparing Case S1 and Case N3, the effect of low excitation in T_w for Case N3 is observed also for the reduced model. In addition to obtaining a different optimal value for C_w, as expected, since the sensor is mounted on a different wall, the profile is much wider for Case N3. This indicates a wider confidence region for this parameter, hence a more uncertain estimate.

Table 4. Optimal parameters with R_g removed.

	R_b	R_w	C_b	C_w
S1	0.040	0.048	1267k	419k
N3	0.035	0.051	1137k	2735k

With identifiable parameters, it is interesting to compare the optimal parameter estimates, listed in Table 4, for each case. Observe first that for both cases, the total thermal resistance between building interior and outside temperature $R_b + R_w \approx 0.088$, which was the value obtained for the total resistance in Table 3. Observe also that both resistances and the interior capacitance C_b is similar for both cases, while the value obtained for C_w is much larger for the N3 case, as expected, due to the high grade insulation used in the North wall.

Hessian vs Profile Likelihood

Observe from Fig. 8 that the super-imposed Hessian based error ellipse at $\alpha = 95\%$ are similar to the 95% confidence regions computed by thresholds on the likelihood profile. Observe especially for the profile R_b vs R_w that the two confidence region methods produce almost identical regions, since the projected likelihood profile is almost symmetric. For e.g. C_b vs C_w the Hessian ellipse and likelihood thresholds are of similar size, but the Hessian has an offset due to the non-symmetric likelihood profile. This shows the advantage of the profile likelihood based regions, in that they can produce accurate results for asymmetric parameter distributions.

5 Conclusion

In this paper, two different sensor locations, giving different dynamic information in the recorded calibration data, was used to estimate and analyse parameters of a thermal network grey-box model of building thermal behaviour. The sensor locations differ with respect to the physical properties of the building, with one sensor fitted to a high insulation sustainable wall, and the other to a standard insulation wall. The profile likelihood method was used, projecting the likelihood function in both one and two parameter dimensions, to show the difference in confidence regions produced by lack of excitation in the calibration data. Confidence regions computed by applying a threshold to the 2D profiles were compared with error ellipses computed based on the Hessian, which shows that while the two confidence region methods give similar results, the PL method better represents the uncertainty when the parameter distribution is asymmetric. The two-dimensional likelihood profile results were used to diagnose parameter non-identifiability, and the model structure was subsequently modified to resolve the problem, thus obtaining identifiable parameters.

References

Peder Bacher and Henrik Madsen. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43(7):1511 – 1522, 2011. ISSN 0378-7788. doi:https://doi.org/10.1016/j.enbuild.2011.02.005.

Thomas Berthou, Pascal Stabat, Raphael Salvazet, and Dominique Marchio. Development and validation of a gray box model to predict thermal behavior of occupied office buildings. *Energy and Buildings*, 74:91–100, 2014.

Ole Magnus Brastein, Bernt Lie, Roshan Sharma, and Nils-Olav Skeie. Parameter estimation for externally simulated thermal network models. *Energy and Buildings*, 2019. ISSN 0378-7788. doi:10.1016/j.enbuild.2019.03.018.

An-Heleen Deconinck and Staf Roels. 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, 2017.

Cristina Sarmiento Ferrero, Qian Chai, Marta Dueñas Díez, Sverre H Amrani, and Bernt Lie. Systematic analysis of parameter identifiability for improved fitting of a biological wastewater model to experimental data. *Modeling, Identification and Control*, 27(4):219, 2006.

Samuel F Fux, Araz Ashouri, Michael J Benz, and Lino Guzzella. EKF based self-adaptive thermal model for a passive house. *Energy and Buildings*, 68:811–817, 2014.

R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007. ISBN 9780131877153.

M Killian and M Kozek. Ten questions concerning model predictive control for energy efficient buildings. *Building and Environment*, 105:403–412, 2016.

Niels Rode Kristensen, Henrik Madsen, and Sten Bay Jørgensen. Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237, 2004.

Henrik Madsen. *Time series analysis*. Chapman and Hall/CRC, 2007.

Henrik Madsen and Jan Holst. Estimation of continuous-time models for the heat dynamics of a building. *Energy and buildings*, 22(1):67–79, 1995.

D.W.U. Perera, Carlos F. Pfeiffer, and Nils-Olav Skeie. Modelling the heat dynamics of a residential building unit: Application to Norwegian buildings. *Modeling, Identification and Control*, 35(1):43–57, 2014. doi:10.4173/mic.2014.1.4.

Michael JD Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis*, pages 51–67. Springer, 1994.

Michael JD Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, pages 26–46, 2009.

William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes in C++*, volume 2. Cambridge University Press, 1992.

Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

Glenn Reynders, Jan Diriken, and Dirk Saelens. Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals. *Energy and Buildings*, 82:263–274, 2014.

C. Runge. Ueber die numerische Auflösung von Differentialgleichungen. *Mathematische Annalen*, 46(2):167–178, Jun 1895. ISSN 1432-1807. doi:10.1007/BF01446807.

D. J. Venzon and S. H. Moolgavkar. A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 37(1): 87–94, 1988. doi:10.2307/2347496.

Article D

Estimating uncertainty of model parameters obtained using numerical optimisation

Authors O. M. Brastein, R. Sharma and N.-O. Skeie

Estimating uncertainty of model parameters obtained using numerical optimisation

O. M. Brastein [1]   B. Lie [1]   C. Pfeiffer [1]   N.-O. Skeie [1]

[1] *Department of Electrical Engineering, Information Technology and Cybernetics, University of South-Eastern Norway, N-3918 Porsgrunn, Norway.*

Abstract

Obtaining accurate models that can predict the behaviour of dynamic systems is important for a variety of applications. Often, models contain parameters that are difficult to calculate from system descriptions. Hence, parameter estimation methods are important tools for creating dynamic system models. Almost all dynamic system models contain uncertainty, either epistemic, due to simplifications in the model, or aleatoric, due to inherent randomness in physical effects such as measurement noise. Hence, obtaining an estimate for the uncertainty of the estimated parameters, typically in the form of confidence limits, is an important part of any statistically solid estimation procedure. Some uncertainty estimation methods can also be used to analyse the practical and structural identifiability of the parameters, as well as parameter inter-dependency and the presence of local minima in the objective function. In this paper, selected methods for estimation and analysis of parameters are reviewed. The methods are compared and demonstrated on the basis of both simulated and real world calibration data for two different case models. Recommendations are given for what applications each of the methods are suitable for. Further, differences in requirements for system excitation are discussed for each of the methods. Finally, a novel adaption of the Profile Likelihood method applied to a moving window is used to test the consistency of dynamic information in the calibration data for a particular model structure.

Keywords: Parameter estimation, Uncertainty analysis, Bootstrapping, Profile Likelihood

1 Introduction

1.1 Background

1.1.1 Dynamic system models

Dynamic system models are important for a large range of scientific and industrial applications, such as *model predictive control* of dynamic systems Killian and Kozek [2016], Wang [2009] or creating *digital twins* of chemical process plants for monitoring or operator training Rosen et al. [2015]. Typically, the performance of the overall system depends on the accuracy of the model predictions. Often, models contain parameters that are difficult to obtain from system specifications. Hence, calibration of model parameters is an important part of developing good quality dynamic models. Additionally, the model parameters are sometimes used as *soft-sensors* for system variables that are otherwise difficult to measure. This requires a specific physical interpretation of the estimated parameters, which places further requirements on the model calibration process.

For many real world processes, models can be created based on balance laws and application of detailed knowledge about the physics and chemistry involved in the process. This approach often includes approximations in order to keep the model's detail level manageable. Such models are classified as *mechanistic,* or *white-box*, since they describe detailed physical information about the system mechanisms, using a mathematical language, in a way that is interpretable by a

human expert. For this type of models, parameters tend to be derived from physical specifications of the system. It is also common for such models to include parameters that require estimation from measurement data.

An alternative approach to creating dynamic system models is the use of *System Identification* (SID) methods Ergon and Di Ruscio [1997], Ljung [1999], where models are created by calibrating parameters of a *predetermined* mathematical structure in order to capture the relevant dynamic system behaviour. SID models are created without explicit use of prior physical information, hence, they are often classified as *black-box,* or *data-driven,* models Kristensen et al. [2004]. One advantage of the SID approach is that it captures the process behaviour directly from measurement data, which avoids modelling errors caused by *insufficient specification* of the system. To properly capture the system behaviour, SID methods require a complete set of basis functions Farrell and Polycarpou [2006]. If the applied set of basis functions is insufficient, the identified model may still *approximate* the system behaviour, but with model errors, e.g., non-linear system identified using a linear model structure. Further, SID methods obtain all system information from data, hence the quality of the *calibration* data, in particular the level of *dynamic information*, directly influences the quality of the obtained model. Finally, the SID approach tends to provide better statistics on the model accuracy, produced as part of the calibration procedure Johansson [1993], Ljung [1999].

A third, intermediate, possibility is to combine *cognitively* constructed model structures, based on *naive* prior physical knowledge, with parameter calibration, to create a *simplified* lumped parameter model. The resulting model, often classified as *grey-box*, tends to have most, if not all, its parameters unknown, which requires full model calibration Berthou et al. [2014], Bohlin and Graebe [1995], Kristensen et al. [2004]. Due to the significant approximations applied in the creation of grey-box models, they should be treated in a stochastic framework, using *Stochastic Differential Equations (SDE)* to describe the dynamic system behaviour. These models are *approximations* by design, using only limited physical insight, which introduces significant *epistemic* uncertainty. However, since the models are based on, at least, a *naive* physical understanding of the underlying system, the parameters are often *assumed* to be physical constants.

Arguably, most white-box models contain some uncertainty in the formulation, which gives rise to model errors, and can therefore benefit from application of grey-box modelling methods for parameter estimation. This approach has indeed been claimed as a natural framework for modelling dynamic systems in general Bohlin and Graebe [1995], Kristensen et al. [2004].

1.1.2 Identifiability

Parameters of models derived, at least partially, from prior knowledge of the underlying physical system are often *assumed* to be constants of the system. Subsequently, a globally optimal value, which can be obtained *unambiguously* by optimisation, is assumed to exist. This assumption should, however, be verified in the context of *parameter identifiability* Ferrero et al. [2006], Johansson [1993], Juhl et al. [2016a], Raue et al. [2009]. This is especially important for *grey-box* models, which contain large *epistemic* uncertainty due to the strong approximation applied in their construction.

It is well known that models can contain parameters that are *structurally* non-identifiable due to overparametrisation, which leads to parameter redundancy, or parameters for which perturbations of the parameter values have no observable effect on the model output Ferrero et al. [2006], Johansson [1993], Raue et al. [2009]. Additionally, lack of *sufficient excitation* of the system during data acquisition may lead to *practical* non-identifiability of certain parameters Deconinck and Roels [2017], Ferrero et al. [2006], Johansson [1993], Murphy and Van der Vaart [2000], Raue et al. [2009]. If the measured inputs and outputs of the physical system do not contain the necessary dynamic information, the influence of some parameters on the error function used for parameter optimisation may be negligible, thus leading to non-identifiability. While *structural* identifiability is independent of the experimental conditions, *practical* identifiability is a function of the dynamic information content in the data-set, and subsequently depends on the experimental configuration Raue et al. [2009].

Due to these potential challenges with parameter identifiability, a model structure may be *designed* with parameters that are *intended* to have a specific physical meaning, but it is not certain that the *estimated* parameters support this assumption Deconinck and Roels [2017]. While the parameters of *physical* systems are clearly *constants* of the system, the *estimated* parameters of a model are always subject to uncertainty and potential non-identifiability.

1.2 Previous work

1.2.1 Parameter estimation and the CTSM framework

Estimation of parameters requires a well defined objective function which *adequately* describes the model fit. Several alternatives are used in the literature, such

as the shooting/ballistic simulation error approach, based on deterministic simulations Berthou et al. [2014], Brastein et al. [2018], or the *maximum likelihood* approach used in the Continuous Time Stochastic Modelling (CTSM) framework Kristensen et al. [2004], Madsen and Holst [1995]. CTSM is based on maximising the *likelihood function Akaike [1998], Rossi [2018]* evaluated by computing residuals, which are assumed to be Normal distributed, in a Kalman Filter. This method has previously been developed in a number of publications Bacher and Madsen [2011], Juhl et al. [2016b], Kristensen and Madsen [2003], Kristensen et al. [2004], Madsen and Holst [1995] and implemented in the CTSM framework Kristensen and Madsen [2003]. This approach offers the advantage of an objective function with a solid statistical framework, which enables use of statistical tools for model validation and selection Kristensen et al. [2004].

1.2.2 Profile likelihood

While *structural* identifiability is well defined in the literature, the *practical* identifiability of parameters is less clearly defined Raue et al. [2009]. Although there are several methods that can identify structural non-identifiability, e.g., Power Series Expansion Pohjanpalo [1978], it is desirable to have a method that can identify *both* types of parameter identifiability. A good choice is the *Profile Likelihood* (PL) method, which creates profiles or distributions of the parameter likelihood, and subsequently can produce likelihood based confidence intervals Deconinck and Roels [2017], Murphy and Van der Vaart [2000], Raue et al. [2009], Venzon and Moolgavkar [1988]. These intervals can be used to diagnose parameter identifiability Raue et al. [2009].

1.2.3 Bootstrapping for time-series data

The idea of Bootstrapping was first introduced in Efron [1979], as a method of estimating the *variance*, i.e., uncertainty, of a statistic. The method has become popular, due in part to its simplicity. The fundamental idea in bootstrapping is to estimate *properties*, such as the *uncertainty* of an estimated parameter, by *randomly drawing samples with replacement* from the *original* data, thus obtaining multiple *different* data-sets. These different data-sets will produce *slightly* different parameter estimates, which allows estimation of the *uncertainty* of the estimated parameter by computing the *mean* and *covariance* of the bootstrapped estimates. Data-sets generated by bootstrapping are often called *pseudo* data-sets to emphasise the fact that they are all re-combinations of the original data, and *not, new, independent* data-sets collected from the physical system. An interesting property of the bootstrapping

method is its intuitive similarity to the basis of the confidence interval (CI) as presented in Kullback [1939], Neyman [1937]; running *multiple* experiments to compute the uncertainty of results.

A fundamental requirement of the bootstrap method, as presented in Efron [1979], is that the samples in the original data must be *independently and identically distributed* (i.i.d), which is a property not usually observed for time-series data Kunsch [1989]. Hence, there has been several adaptions of bootstrapping for time-series data. One solution is to fit a parametric Auto Regressive Moving Average (ARMA) model to the data, and bootstrap the residuals, which are presumed i.i.d, to create new data-sets Kunsch [1989], Lie [2009], Politis [2003]. However, this approach is limited to systems which can be adequately described by such model structures and thus produce i.i.d. residuals. Hence, *non-parametric* approaches to bootstrapping for time-series data has been receiving significant interest in research Kunsch [1989], Lodhi and Gilbert [2011], Politis [2003], Politis and Romano [1994]. In particular, various forms of block based bootstrapping, i.e., methods that segment the data into blocks, and draw randomly with replacements from the blocks, rather than the samples themself, has shown promising results Kunsch [1989]. Examples include overlapping and non-overlapping block bootstrap Kunsch [1989], Lodhi and Gilbert [2011], moving block bootstrap Kunsch [1989], and stationary bootstrapping Politis and Romano [1994]. For a detailed review of bootstrapping for time-series data, see Politis [2003].

1.3 Overview of paper

In this paper, selected methods for estimating uncertainty of estimated parameters are presented and demonstrated on two separate test cases. The methods discussed in this paper are based on the use of *numerical optimisation*, using a well defined objective function to evaluate the fit of a parameter set. The focus in this paper is on analysing the *parameters* themselves, rather than the prediction accuracy of the calibrated model.

The theoretical foundation of parameter estimation and analysis is presented in Section 2. Sections 2.1 and 2.2 detail the foundation of parameter estimation and the representation of uncertainty, respectively. Section 2.3 presents the theoretical foundation of each of the discussed methods. Results of applying the methods is presented in Section 3. Finally, recommendations as to what applications each of the methods is most suitable for are given in Section 3.3 before the paper is concluded in Section 4.
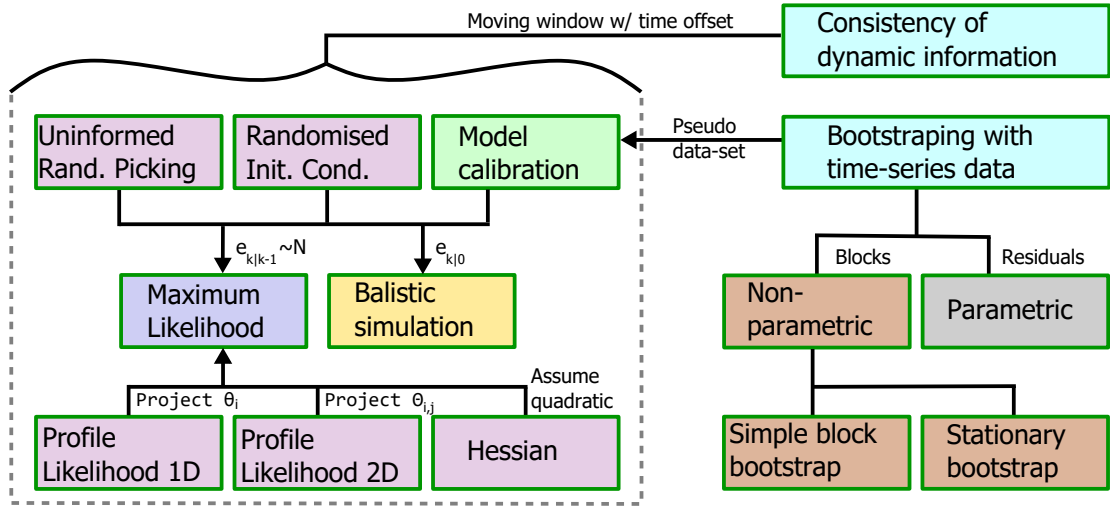
Figure 1: An overview of all the methods discussed in this paper, and how they are related. On the left-hand side, six optimisation based methods are shown, and related with the two types of objective functions that are discussed. The top three methods: Uninformed Random Picking, Randomised Initial Conditions and straight forward Model Calibration, can be used with both types of objective function. The lower three methods: Profile Likelihood 1D and 2D, and the Hessian curvature method, all require a Maximum Likelihood objective function. The top right block in the figure shows the moving window method, which can be used with all the six different methods to test the results for consistency in time. Next, the discussed variations of bootstrapping for time-series data are shown, and associated with model calibration, since the working principle of bootstrapping methods requires separate calibration of parameters for each generated pseudo data-set. Further, the two types of bootstrapping for time-series data, parametric vs non-parametric, are shown. Only non-parametric bootstrapping is discussed in this paper, specifically simple block bootstrapping and stationary bootstrapping, as illustrated in the lower right part of the figure.

2 Theoretical basis

This section discusses in detail several methods for estimating parameters of dynamic models, and, in particular, methods that also estimate the *uncertainty* and *identifiability* of the estimated parameters. An overview of the methods is presented in Fig. 1.

2.1 Parameter estimation

Estimation of parameters θ for a known model structure $\mathcal{M}$ can formally be defined as solving the optimisation problem:

$$\hat{\theta} = \arg\min_{\theta} g\left(\theta; \mathcal{M}, \mathcal{K}, \mathcal{A}\right) \tag{1}$$

$$\text{s.t. } \theta \in \Theta$$

Here, $\Theta = \{\theta : \theta_{\min} < \theta < \theta_{\max}; \theta \in \mathbb{R}^{n_\theta}\}$ is a continuous space of feasible values for the model parameters, where $\theta_{\min}$ and $\theta_{\max}$ are the lower and upper bounds. Hence, the space Θ forms inequality constraints for the optimisation problem in Eq. (1). $\mathcal{K}$ are the experimental conditions, including a set of measurements of

system inputs and outputs, which are defined in continuous time as input $u_t \in \mathbb{R}^{n_u}$ and output $y_t \in \mathbb{R}^{n_y}$, and the corresponding ordered sequences of discrete time measurements u_k and y_k:

$$y_{[N]} = [y_0, y_1, \ldots, y_N] \tag{2}$$
$$u_{[N]} = [u_0, u_1, \ldots, u_N]$$

Here, the integer subscripts $k = 0, 1, \ldots, N$ denote the discrete time sampling instants and the *subscript* enclosed in $[\cdot]$ is used to indicate an ordered sequence. These measurements are used to evaluate the objective function $g(\theta)$ when θ is varied over the feasible set Θ by a numerical optimisation algorithm $\mathcal{A}$.

2.1.1 Uncertainty of estimated parameters

A common assumption is that $\mathcal{S} \in \mathcal{M}(\Theta)$, where $\mathcal{S}$ is the true system, and that there exists a true parameter vector θ^* such that $\mathcal{M}(\theta^*) \equiv \mathcal{S}$. However, if the data used for parameter estimation, i.e., $y_{[N]}$ and $u_{[N]}$, is collected as *measurements* from the physical system $\mathcal{S}$, it will contain *aleatoric* uncertainty

due to the inherent randomness of *measurement noise* Bentley [2005], Ljung [1999]. Additionally, most dynamic system models contain approximations with respect to $\mathcal{S}$, which also introduces some *epistemic* uncertainty in the model equations. Hence, the assumption $\mathcal{M}(\theta^*) \equiv \mathcal{S}$ is almost always questionable. Further, the estimate $\hat{\theta}$ depends on the amount of dynamic information in $\mathcal{K}$, the choice of model fit objective function $g(\theta)$, and to some extent on the optimisation algorithm $\mathcal{A}$. Hence, even if there exists a well defined, globally optimal, set of parameters θ^*, it may not be possible to obtain an *unambiguous* parameter *estimate*. Therefore, prior to interpreting a set of estimated parameters as *determined* by the physical properties of $\mathcal{S}$, and subsequently assuming a physical interpretation to the *estimated* parameter values, it is necessary to analyse the estimation uncertainty and *identifiability* of the parameters Ljung [1999].

2.1.2 Stochastic parameter estimation

Dynamic system models are typically formulated with uncertainty in both the *state transition* and *measurement* equations Bohlin and Graebe [1995], Kristensen et al. [2004], Simon [2006]. Such models can conveniently be expressed as a continuous time stochastic differential equation (SDE) for the state transition Jazwinski [1970]. Since the data used for calibration is typically available only at discrete time instants, a discrete time measurement equation is a convenient formulation. Adopting the notation of Kristensen et al. [2004]:

$$dx_t = f(x_t, u_t, t, \theta)\, dt + \sigma(u_t, t, \theta)\, d\omega_t \quad (3)$$

$$y_k = h(x_k, u_k, t_k, \theta) + e_k \quad (4)$$

where $t \in \mathbb{R}$ is the time variable and $x_t \in \mathbb{R}^{n_x}$ is the continuous time state vector. The first and second terms in Eq. (3) are commonly called the *drift* and *diffusion* term, respectively Jazwinski [1970], Kristensen et al. [2004]. The diffusion term expresses the process noise as the function σ multiplied with the differential of a standard Wiener process ω_t Jazwinski [1970], Kristensen et al. [2004]. The discrete time *measurement equation* is given in Eq. (4). The CTSM framework Kristensen and Madsen [2003], Kristensen et al. [2004], Madsen and Holst [1995] presents a statistically solid approach to estimating parameters in such stochastic models. A *Maximum Likelihood* estimate of θ can be obtained by deriving the objective $g(\theta)$ in Eq. (1) from the likelihood $L(\theta)$, defined as the *joint* probability $\mathrm{P_r}(\cdot)$ of observing the measurement sequence $y_{[N]}$ when θ and $\mathcal{M}$ are known, i.e., $L(\theta; y_{[N]}, \mathcal{M}) = \mathrm{P_r}(y_{[N]}|\theta, \mathcal{M})$. Note that while $L(\theta; y_{[N]}, \mathcal{M})$ is defined using probability, the resulting likelihood function is *not* a probability *distribu-*

tion, since the integral of the likelihood over all possible parameters does *not* equal 1.

For simplicity of notation, the model structure $\mathcal{M}$ is implicitly assumed known and omitted from the condition. The likelihood can be expanded to conditional probabilities by the chain rule:

$$L(\theta; y_{[N]}) = \left(\prod_{k=1}^{N} \mathrm{P_r}(y_k|y_{[k-1]}, \theta)\right) \mathrm{P_r}(y_0|\theta) \quad (5)$$

Equation (3) assumes the diffusion term to be additive and independent of the state x, and driven by a *Wiener process* Kristensen et al. [2004]. Hence, it is reasonable to assume that the conditional probabilities in Eq. (5) can be approximated by Gaussian distributions Kristensen and Madsen [2003], Kristensen et al. [2004]. The likelihood can then be expressed as a multivariate Gaussian distribution Kristensen et al. [2004],

$$L(\theta; y_{[N]}) =$$
$$\left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\epsilon_{k|k-1}^T \mathcal{E}_{k|k-1}^{-1} \epsilon_{k|k-1}\right)}{\sqrt{\det(\mathcal{E}_{k|k-1})}\left(\sqrt{2\pi}\right)^{n_y}}\right) \mathrm{P_r}(y_0|\theta) \quad (6)$$

To ensure that Eq. (6) is justified, the normality assumption on the residuals can, and *should*, be checked during model validation Johansson [1993], Kristensen et al. [2004]. Model validation is further discussed in Section 2.1.4

The residuals $\epsilon_{k|k-1}$ and their covariance $\mathcal{E}_{k|k-1}$ are needed to evaluate Eq. (6). These quantities can be obtained by use of a Kalman Filter (KF):

$$\hat{y}_{k|k-1} = \mathbb{E}\left[y_k|y_{[k-1]}, \theta\right] \quad (7)$$

$$\epsilon_{k|k-1} = y_k - \hat{y}_{k|k-1} \quad (8)$$

$$\mathcal{E}_{k|k-1} = \mathbb{E}\left[\epsilon_{k|k-1}\epsilon_{k|k-1}^T\right] \quad (9)$$

The choice of KF implementation, either the standard linear KF for linear models, or a non-linear variant such as the *Extended KF* (EKF) or the *Unscented KF* (UKF), depends on the model equations Brastein et al. [2019a].

Equation (6) is further simplified by conditioning on knowing y_0, taking the negative logarithm, and eliminating the factor $\frac{1}{2}$. Finally, the objective $g(\theta)$ in Eq. (1) is defined as $g(\theta; \mathcal{M}, \mathcal{K}) = \ell(\theta)$ where the log likelihood function $\ell(\theta)$, omitting the dependency on $y_{[N]}$ for simplicity of notation, is given as

$$\ell(\theta) = \sum_{k=1}^{N} \epsilon_{k|k-1}^T \mathcal{E}_{k|k-1}^{-1} \epsilon_{k|k-1} + \ln\left(\det(\mathcal{E}_{k|k-1})\right) \quad (10)$$

The constant term $c = N \cdot n_y \cdot \ln(2\pi)$ has also been omitted.

2.1.3 Deterministic parameter estimation

Dynamic system models typically contain both *aleatoric* and *epistemic* uncertainty caused by the inherent randomness of measurements and the use of approximations in the model equations, respectively. Despite the well understood stochastic nature of such models, it is a common practice to treat all uncertainty as *aleatoric* and present at the model output. This results in a *deterministic*, sometimes called a *shooting* or *ballistic,* simulation, approach, in which the simulated state trajectory is *completely determined* by the given parameter vector θ, the initial conditions, and the measured system inputs. Essentially, the parameter estimation problem is then formulated as a *curve fitting* of the state trajectory transformed through the measurement equation. Rewriting the model from Eqs. (3) and (4) in discrete time *without* the diffusion term, let

$$\hat{x}_{k|0} = \tilde{f}\left(\hat{x}_{k-1|0}, u_k, \theta\right) \tag{11}$$
$$\hat{y}_{k|0} = \tilde{h}\left(\hat{x}_{k|0}, u_k, \theta\right) + e_k$$

where the estimated state $\hat{x}_{k|0}$, and subsequently the estimated output $\hat{y}_{k|0}$, at time k are computed using *only* information available at *initial* time. The *Ordinary Least Squares* (OLS) estimate of the parameters is obtained by minimising the sum of square errors (SSE):

$$\tilde{g}\left(\theta\right) = \sum_{k=1}^{N} \tilde{\epsilon}_{k|0}^T Q \tilde{\epsilon}_{k|0} \tag{12}$$

where Q is a weighting matrix. Here, the estimation error $\tilde{\epsilon}_{k|0} = y_k - \hat{y}_{k|0}$ depends only on information at initial time t_0, which is in contrast to the residual obtained by *the one-step ahead predictions* in Eq. (8).

It is interesting to observe that the estimate obtained by minimising Eq. (12) corresponds to the *maximum likelihood estimate* (MLE) obtained from minimising Eq. (6) if, and only if, $\tilde{\epsilon}_{k|0} = \epsilon_{k|k-1}$ and the innovation covariance $\mathcal{E}_{k|k-1}$ is constant such that $Q = \mathcal{E}^{-1}$. Hence, minimising Eq. (12) gives an MLE estimate of the parameters only if the state transition model is *exact* w.r.t. the data generating system $\mathcal{S}$, i.e., the uncertainty associated with the diffusion term in Eq. (3) is zero and the measurement noise distribution is *stationary* with zero mean. Note also that if all measurements have the same variance, i.e., $\mathcal{E}^{-1} = Q = c \cdot I$ in Eq. (12), the weighting matrix can be taken outside the summation and subsequently eliminated, thus obtaining the *unweighted* least squares estimate.

While this can be a reasonable approximation, it is rarely exactly true, except when the calibration data is generated by simulations of the same model structure $\mathcal{M}$. Observe further that, assuming affine noise, this corresponds to obtaining the quantities in Eqs. (7) to

(9) in a Kalman Filter with the process noise covariance $\mathcal{W} = 0$ and constant measurement noise covariance $\mathcal{V}$. Hence, the deterministic shooting error approach to parameter estimation may be seen as a special case of the scheme used in the CTSM framework and outlined in Section 2.1.2.

An interesting observation from comparing the two types of error calculation, e.g., $\tilde{\epsilon}_{k|0} = y_k - \hat{y}_{k|0}$ and $\epsilon_{k|k-1} = y_k - \hat{y}_{k|k-1}$, is that the SSE objective computed based on $\tilde{\epsilon}_{k|0}$ in Eq. (12) will have a gradient that is strongly non-linear in the parameters, due to the recursive predictor used in Eq. (11), i.e., $\hat{y}_{k|0} = f\left(\hat{y}_{k-1|0}, u_{k-1}, \theta\right)$. In contrast, the one-step-ahead prediction based likelihood objective in Eq. (10) will have a gradient that is linear in the parameters, since the predictor for the output is a function of measurements at previous time-steps, i.e., $\hat{y}_{k|k-1} = f\left(y_{[k-1]}, u_{k-1}, \theta\right)$.

2.1.4 Model validation

Since the objective function $\ell\left(\theta\right)$ in Eq. (10) depends on an assumption of normally distributed residuals, computed from one-step ahead predictions in a KF, it is necessary to verify the normality assumption subsequent to estimating model parameters. The literature detailing the CTSM framework specifically calls for evaluation of the residuals to verify the normality assumption Kristensen and Madsen [2003], Kristensen et al. [2004]. A practical test for normality can be applied by computing and plotting a cumulative periodogram (CP) of the residuals Deconinck and Roels [2017], Kristensen and Madsen [2003], Kristensen et al. [2004], where the *Kolmogorov-Smirnov criterion* can be used to place confidence bounds on the CP test Madsen [2007]. There are also a number of alternative tests for normality that can be applied such as the *zero-crossings* test or the *Kolmogorov-Smirnov* test Johansson [1993].

The possibility of validating a dynamic system model by testing the residuals for normality is a distinct *advantage* of the stochastic parameter estimation framework. For a *deterministic shooting simulation* approach, in which there may be *bias* errors that carry over from the state estimate at the previous time-step as shown in Eq. (11), there can be no reasonable assumption of normality for the estimation error $\tilde{\epsilon}_{k|0}$, unless the state transition model is *exact* Madsen [2007].

2.2 Expressing uncertainty of estimated parameters

A convenient way of describing the uncertainty of estimated parameters is by defining a sub-region in Θ, with some specific statistical criteria quantifying the

uncertainty of the parameters in the sub-region relative to the true, but unknown, parameters θ^*. One possible choice is the use of a *confidence region* with stated confidence α Neyman [1937], Raue et al. [2009]. In general, a region of arbitrary shape in Θ can be defined as a set, based on the *difference* in the objective function *relative* to a presumed optimal estimate $\hat{\theta}$:

$$\left\{ \theta : g\left(\theta\right) - g\left(\hat{\theta}\right) < \Delta \right\} \tag{13}$$

where the threshold Δ is defined by some appropriate *statistical criterion*. The definition of the threshold depends on how the objective function g is defined, e.g., for a likelihood objective the thresholds can be computed from the χ^2 distribution as shown in Section 2.2.2. The set in Eq.(13), which contains any parameters θ for which the objective differs from the optimum by less than Δ, can be of any shape, including multimodal. However, the computation of a free-form set will require a large number of evaluations of the objective function for different θ in order to determine the set members. Therefore, a common approximation is to assume an ellipsoid, rather than free-form, region, defined as

$$\left\{ \theta : \left(\theta - \hat{\theta}\right)^T \Sigma^{-1} \left(\theta - \hat{\theta}\right) < \Delta \right\} \tag{14}$$

where the *size* of the ellipsoid is determined by the threshold Δ, again computed by some appropriate statistical criterion. The weighting matrix Σ, typically the covariance of the estimated parameters, determines the *rotation* and *relative length* of the ellipsoid axes. Regions defined as in Eq. (14) also define a set of θ based on relative deviation compared to $\hat{\theta}$, but by assuming a quadratic approximation, the ellipsoid region is much faster to compute.

The points on the ellipsoid surface can be obtained by utilising the Cholesky decomposition $\Sigma = LL^T$, assuming Σ is positive definite Press et al. [2007]:

$$\left(\theta - \hat{\theta}\right)^T \Sigma^{-1} \left(\theta - \hat{\theta}\right) = \Delta \Rightarrow \left| L^{-1} \left(\theta - \hat{\theta}\right) \right|^2 = \Delta \tag{15}$$

Next, suppose x is a point on a unit hypersphere, then the ellipsoid surface boundary is obtained by the affine transformation

$$\theta = \hat{\theta} + \sqrt{\Delta} L x \tag{16}$$

2.2.1 Asymptotic confidence regions

Two common ways of expressing uncertainty is by defining a region in Θ, either a *univariate, point-wise,* confidence interval (CI), or a *multivariate, simultaneous,* confidence region, both defined by their prescribed confidence level α Neyman [1937].

Asymptotic CIs are based on the curvature of the objective function, which can be computed by utilising the covariance Σ_θ of the estimated parameters around the optimum $\hat{\theta}$ Deconinck and Roels [2017], Raue et al. [2009] to define a region on the form in Eq. (14). The threshold is then $\Delta = \Delta_\alpha$, where Δ_α is computed from the $\chi^2_{\alpha, n_{df}}$ distribution, with degrees of freedom n_{df} equal to the number of parameters in the *simultaneous* confidence region Press et al. [2007]. Observe that for *point-wise* confidence intervals of single parameters, Eq. (16) with $x \in \{\cos(0), \cos(\pi)\} = \{1, -1\}$ reduces to the familiar confidence interval for a scalar variable, where $\Sigma_{i,i} = \sigma_i^2$ Raue et al. [2009], i.e.;

$$\hat{\theta}_i \pm \sqrt{\Delta_\alpha \Sigma_{i,i}} \tag{17}$$

For *point-wise* intervals, Δ_α is drawn from the $\chi^2_{\alpha, n_{df}}$ distribution with $n_{df} = 1$ which is equivalent to the Normal c.d.f. with $\alpha/2$ confidence in each tail. The use of asymptotic confidence regions is widespread in all branches of science, particularly due to their ease of computation. If the *parameters* are in fact Gaussian distributed, the ellipsoid confidence regions are exact which further strengthens their popularity.

2.2.2 Likelihood based confidence regions

Unlike the asymptotic confidence interval in Eq. (17), a *likelihood based confidence interval* is computed by applying a *threshold* on the likelihood function to compute a confidence region in the form Eq. (13) Meeker and Escobar [1995], Raue et al. [2009]. Let

$$\left\{ \theta : \ell\left(\theta\right) - \ell\left(\hat{\theta}\right) < \Delta_\alpha \right\} \quad , \quad \Delta_\alpha = \chi^2_{\alpha, n_{df}} \tag{18}$$

where $\hat{\theta}$ is a freely estimated parameter vector, which is presumed optimal, and the threshold Δ_α is the α percentile of the $\chi^2_{\alpha, n_{df}}$-distribution with n_{df} degrees of freedom. It follows from *Wilks' theorem* Wilks [1938] on the logarithm of the likelihood ratio Λ that the test statistic

$$2\ln\left(\Lambda\right) = 2\ln\left(\frac{L\left(\theta\right)}{L\left(\hat{\theta}\right)}\right) = \ell\left(\theta\right) - \ell\left(\hat{\theta}\right)$$

can be used to compare two models. The difference in log likelihood $\ell\left(\theta\right) - \ell\left(\hat{\theta}\right)$ is asymptotically χ^2-distributed Meeker and Escobar [1995], Raue et al. [2009], with n_{df} equal to the difference in number of free parameters between θ and $\hat{\theta}$ Press et al. [2007].

Arguably, likelihood based confidence intervals are conceptually simpler than asymptotic CIs due to their thresholded set definition. However, determining the set members is computationally intensive. An advantage of the likelihood based CI is that, due to its set

form definition, it does not assume a symmetric distribution of the parameters, and can in fact take on *any* shape, including multi-modal. Hence, likelihood based CIs are often considered superior to asymptotic CIs Raue et al. [2009].

2.2.3 Parameter profiles or distributions

An alternative to presenting the uncertainty of the estimated parameters as regions in Θ is to present the parameters as a *distribution* in Θ. Typically, a statistical quantity is used to create the profile, such as a *probability density function* or the *log-likelihood*. Profiles can be created over the entire Θ, or a subset of Θ as projections to single parameters θ_i, or planes $\Theta_{i,j}=\{\theta_i,\theta_j\}$, such that $\Theta_{i,j} \subset \Theta$. A parameter profile is more *descriptive* than a confidence region, since it shows how the uncertainty is *distributed* across the parameter space Θ. Since parameter profiles can be converted to confidence regions by applying some statistically defined threshold, they may be considered a superior form of uncertainty description. An example of this approach is the *Profile Likelihood* method presented in Section 2.3.4.

Another method of obtaining distributions of parameters, which are in fact probability distributions for the parameter θ, is through the use of Bayesian statistics and *Markov Chain Monte Carlo* (MCMC) methods. However, these methods are beyond the scope of this paper.

2.2.4 Interpretation of confidence regions

An interesting observation relating to the *interpretation* of the computed regions is that, while quite often assumed in published literature, the *confidence* of the computed regions is *not* a statement on the *probability* of said region containing the true parameters θ^*, as clearly stated in Neyman [1937]. Both θ^* and the computed confidence region are constants, not random variables. Hence, their relationship is not a question of probability, except for the trivial values of *zero* and *one*, which simply state whether or not the true parameter is a *member* of the computed confidence region. However, what *can* be stated in probabilistic terms is the *expected* probability of capturing θ^* in the CI, *prior* to performing the experiment and computing the interval, which is equal to the confidence α Kullback [1939]. This *expected* probability of capturing θ^* is called the *coverage probability*. If multiple experiments are carried out, with subsequent computations of CIs, the *ratio* of intervals that successfully captures the true parameter θ^* to the total number of experiments performed will be equal to the *coverage probability* Kullback [1939].
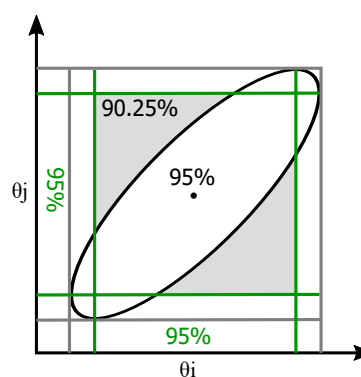
Figure 2: Comparing a simultaneous confidence ellipse for two parameters with the point-wise CIs (green) for each parameter, shows that the projections of the ellipse (grey) is wider than the point-wise CIs. Also, the combined confidence $(0.95\times0.95=0.9025)$ of both point-wise CIs together is shaded in grey. Note the difference between two combined point-wise CIs (shaded square) and the simultaneous confidence ellipse.

2.2.5 Simultaneous and point-wise confidence regions

As discussed in Sections 2.2.1 and 2.2.2, the uncertainty of estimated parameters can be expressed as confidence regions in Θ. However, it is often of interest to make statements about the uncertainty of *individual* parameters, rather than *simultaneous* statements about multiple parameters together. In this context it is important to distinguish between *simultaneous* and *point-wise*, i.e., one-at-a-time, intervals. For a simultaneous CI, the uncertainty of the estimated parameters is stated for multiple parameters together, i.e., the computed confidence *region* captures the true parameters θ^* with *coverage probability* α Kullback [1939], Neyman [1937]. In comparison, a *point-wise* CI holds for that parameter alone, i.e., the *coverage probability* for capturing the single parameter is α.

To create *simultaneous scalar intervals* for each parameter, the higher dimension region can be *projected* onto each parameter, as illustrated in Fig. 2. Such projected simultaneous intervals should not be confused with *point-wise* CIs, nor should their *combined* confidence be stated as α Johnson and Wichern [2007], i.e., the coverage probability of all *projected* intervals holding is *not* α. The projected shadow of a higher order simultaneous confidence region is larger than the point-wise intervals Johnson and Wichern [2007], Press et al. [2007], as illustrated in Fig. 2.

Since Θ typically has more than two dimensions,

graphical presentation of confidence regions requires projections of some form. In such cases, care should be taken to clearly state the resulting confidence level. Just as for the elliptic region projected onto a single parameter axis in Fig. 2, a higher dimension ellipsoid projected onto a plane will give a larger elliptic *shadow* projection than a confidence ellipse computed for just two parameters in the plane.

2.2.6 Diagnosing identifiability by analysing uncertainty

Determining if a parameter is *structurally* or *practically* identifiable is important if the parameter value itself is of interest, i.e., if a *physical interpretation* of the parameter is assumed. A link between the uncertainty of a parameter, in the form of a likelihood based confidence interval as presented in Section 2.2.2, and the *structural* and *practical identifiability*, was given in Raue et al. [2009].

Structural non-identifiability is caused by *redundant* parametrisation of the model equations, such that a sub-set of the parameters θ_s has no effect on the observable outputs y, and is therefore independent of the experimental conditions $\mathcal{K}$ Raue et al. [2009]. Hence, there exists a manifold in the parameter space Θ where the objective function $\ell(\theta)$ has a constant value. Further, it is possible to obtain a functional relation between the parameters θ_s which describes this equipotential manifold in the objective function. Consequently, a likelihood based confidence interval will be unbounded in both direction, i.e., $[-\infty, +\infty]$, for the *structurally non-identifiable* parameters in θ_s Raue et al. [2009].

In contrast, *practical* non-identifiability is caused by a lack of dynamic information about the system in $\mathcal{K}$ and hence a direct result of the *experimental design* and data acquisition process. Unlike structural identifiability, practical identifiability is not clearly defined in the literature Raue et al. [2009]. However, an elegant definition is found in Raue et al. [2009], where practical non-identifiability is diagnosed if the corresponding likelihood based confidence region, as in Eq. (18), is extended to infinity in decreasing and/or increasing direction , i.e., the objective function stays below a specific threshold Δ_α in at least one direction, despite the presence of a well defined optimum $\hat{\theta}$. Observe that the use of *likelihood* based confidence regions is necessary for determination of practical non-identifiability, since the *asymptotic* CI will always be symmetric and also finite if $\Sigma_{i,i} > 0$, and therefore cannot present the necessary characteristics for diagnosing practical non-identifiability Raue et al. [2009].

In Raue et al. [2009] the definition of parameter identifiability is presented as a true/false question. For the
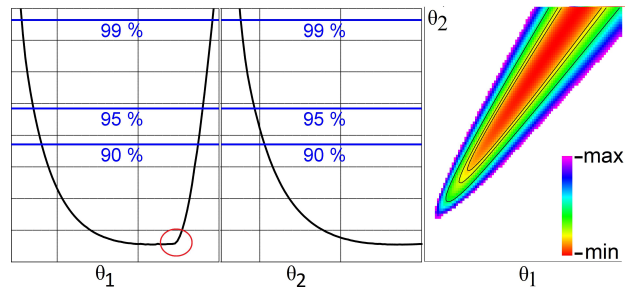

Figure 3: The log-likelihood profile of two interdependent parameters is plotted in the plane of both parameters (right panel), with corresponding projections to each of the two parameters.

case of structural identifiability, this is clearly appropriate. However, practical identifiability is a function of the system excitation during data acquisition, and hence the dynamic information content in $\mathcal{K}$. Therefore it may be appropriate to treat practical identifiability as a *quantity,* rather than a *true/false* property. If the computed confidence region for a parameter is *wide*, that parameter may arguably be considered *less* identifiable than a parameter with smaller confidence region. In particular, comparing parameters estimated from two different data-sets of the same system with different levels of excitation, the resulting CIs of the same parameter may have different widths. Hence, it is reasonable, and intuitively satisfactory, to relax the definition of practical non-identifiability given in Raue et al. [2009] to also include parameters with *abnormally wide* CIs. Unfortunately, relaxing the diagnostic criteria in this way leads to a *cognitive judgment* on what width of a CI is *abnormal* for any specific parameter. Resolving this question requires using system specific knowledge, and is further complicated by variations in scale of the parameters which makes normalisation a prerequisite for comparing CIs for different parameters.

2.2.7 Inter-dependent parameters and the effect of constraints on projections

When projecting a higher order region onto a single parameter θ_i or a plane $\Theta_{i,j}$, it is important to consider inter-dependent parameters. A projection of a higher dimensional region in Θ will, due to parameter inter-dependence, result in a projection that is wider than any cross-section, since the dependency information in the higher order structure is lost in the projection Johnson and Wichern [2007]. Further, if the parameter space Θ is constrained or bounded, inter-dependent parameters can introduce *artefacts* in the

projection of one parameter, caused by the constraints on another, inter-dependent, parameter.

An example of these phenomena is shown in Fig. 3. The *right* panel shows a log-likelihood profile of two parameters as a heat-map in the plane of both parameters. The *left* and *centre* panel show the same profile, projected onto each parameter axis. First, observe that the 2D profile in the right panel shows that the parameters are inter-dependent, since there is a clear linear relationship between the two parameters. Next, observe that the one-dimensional profiles, which are projections of the two-dimensional surface onto each parameter, are wider than any cross-section taken from the 2D profile. Finally, observe that for parameter θ_1 the profile contains a sharp *bend*, highlighted by a red circle in Fig. 3. When comparing to the full 2D profile, it is clear that this *bend* is actually an *artefact* in the θ_1 profile, introduced by the constraint on θ_2, i.e., $\theta_{2,\min} < \theta_2 < \theta_{2,\max}$, and the inter-dependence between the parameters.

2.3 Uncertainty estimation and analysis methods

In this section, a selection of methods for estimation of uncertainty and identifiability analysis is presented with some illustrative examples. More extensive examples of these methods are given in Section 3.

2.3.1 Uninformed Random Picking

It is often helpful to visualise the shape of the objective function g in the parameter space Θ. Initially, the optimal parameter estimate, the existence of a well defined optimum, and/or the number of optima, is typically unknown. Hence, a method which requires no assumptions about the objective function $g(\theta)$ and the parameter space Θ, is desirable. An intuitive approach is to evaluate the objective $g(\theta)$ for some selected *set* of parameters $\theta_{\{K\}} = \{\theta_k : k \in 1, ..., K\}$ and plot the resulting θ_k vs $g(\theta_k)$ as a *scatter plot* for each parameter. A simple way of selecting $\theta_{\{K\}}$ is by use of randomisation: drawing the parameters uniformly across Θ such that $\theta_k \sim \mathcal{U}(\theta_{min}, \theta_{max})$ for $k \in \{1, ..., K\}$. The resulting scatter plots will show that there exists an optimal *front* in Θ which corresponds to the projection of the objective $g(\theta)$ onto each parameter axis. Of course, a large number of the randomly drawn points in $\theta_{\{K\}}$ are not located near the optimal front. However, by randomised selection with *large K*, typically on the order of 10.000 to 500.000 or higher depending on the dimensionality of Θ, the plots will contain *enough* data near the optimal front to *visually* inspect the shape of the objective function. Subsequently, the existence of a well defined optimum, presence of flat regions, and the
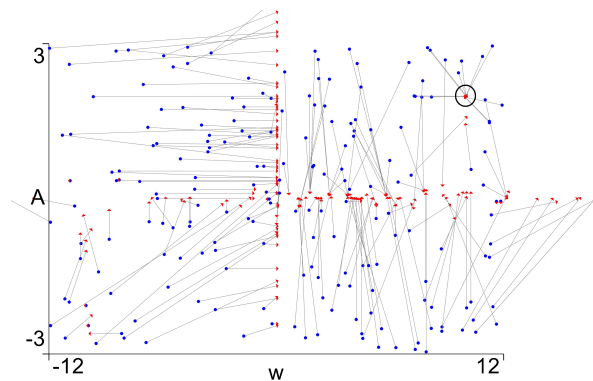

Figure 4: The Random initial guess method is used to test for local minima when estimating the amplitude (A) and frequency (w) of a simulated sine wave with added noise. Out of $K = 200$ repeated optimisations, only 12 correctly find the true parameters θ^*(marked by a black circle in the figure). Blue dots mark the randomised initial guess and red dots mark the parameters obtained after optimisation.

number of modes, can be ascertained. This method is often named *Uninformed Random Picking (URP)* Hoos and Stützle [2004], where the term *uninformed* reflects the fact that no prior assumptions is used in the choice of θ_k.

The resulting plots will be a projection of K parameter vectors onto each of the n_θ parameter axis in Θ. Hence, the method suffers from the challenges related to such projections, as discussed in Section 2.2.7. Examples of the use of the URP method is given in Section (3.1.1).

2.3.2 Randomised initial guess

A variation of the *Uninformed Random Picking* method from Section 2.3.1 is to use randomisation to uniformly draw the initial guess θ^0 and subsequently optimise all parameters, i.e., repeatedly solve the optimisation problem in Eq. (1) K times, with a randomised initial guess $\theta_k^0 \sim \mathcal{U}(\theta_{\min}, \theta_{\max})$ for $k \in \{1, ..., K\}$. This method, although simple, can be a good test to check the convergence of the parameter estimation method. If repeated executions of the optimisation algorithm $\mathcal{A}$ returns significantly different optimal estimates $\hat{\theta}_k$ depending on the initial guess θ_k^0, a *physical interpretation* of the estimated parameters as constants given by the system $\mathcal{S}$ must be considered *questionable*. This may indicate a problem with parameter identifiability, which should be analysed further. Since the optimisation algorithm $\mathcal{A}$ allows *directed* exploration

of the objective function, the number of iterations K can be much lower than required for the URP method, say, 10 to 500. As for the URP method, the choice of K depends on the dimensionality of the parameter space Θ.

Additionally, since optimisation is performed from a number of different starting points in Θ, this method can be useful to identify local minima in the parameter space, provided the number of iterations K is large enough to cover the parameter space with reasonable density.

The resulting optimal estimates $\hat{\theta}_k$ are plotted together with the initial guesses θ_k^0 to indicate the trajectories of the optimisation algorithm $\mathcal{A}$. The results can either be plotted for two parameters against each other, forming a projection of the corresponding optimisation trajectories onto a plane of two parameters, or they can be plotted for each parameter; θ_k^0, $\hat{\theta}_k$ vs $g\left(\theta_k^0\right)$, $g\left(\hat{\theta}_k\right)$. The resulting plots will give a good visualisation of the projected shape of the objective function.

An example of this method is shown in Fig. 4. By simulating a sine wave $y(t) = A\sin(wt)$, where the parameters are amplitude $A = 2$ and frequency $w = 10$, and adding Gaussian noise of standard deviation 0.5, a calibration data-set is created. When estimating the parameters of this simple model, a large number of local optima are found, especially for $A = 0$ or $w = 0$. Hence, the estimated solution $\hat{\theta}$ is highly dependent on the initial guess θ^0. Only 12 of the $K = 200$ repeated optimisations correctly obtain $\hat{\theta} = \theta^*$. This example shows the importance of considering local minima when estimating parameters. It is also interesting to observe that there are different patterns of trajectories in each of the four quadrants of the plot. These variations are caused by the optimisation method $\mathcal{A}$, and shows that also the chosen algorithm for optimisation can have a strong influence on the parameter estimate.

2.3.3 Hessian of the likelihood function

A commonly used method for estimating the uncertainty of the estimated parameters is to utilize the shape of the objective function $g(\theta)$ directly by calculating the curvature around the optimal estimate $\hat{\theta}$, by computing the *Hessian* of $\ell(\theta)$; $H = \left.\nabla^T\nabla\ell(\theta)\right|_{\theta=\hat{\theta}}$. The covariance of the estimated parameters can be computed as $\Sigma_\theta = 2H^{-1}$, where the factor 2 is included to compensate for previously dropping the factor $\frac{1}{2}$ in the definition of $\ell(\theta)$ in Eq. (10). The elements of H are approximated as Kristensen et al. [2004], Raue et al. [2009]:

$$h_{i,j} \approx \left.\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell(\theta)\right)\right|_{\theta=\hat{\theta}} \tag{19}$$

which can be numerically computed using, e.g., *central difference approximation*. Observe that the Hessian is by definition *symmetric*, which is a drawback if the *shape* of the objective function is *non-symmetric* around the optimum. Observe also that, from general optimisation theory, while the curvature of any objective function g could be considered an approximation of uncertainty Nocedal and Wright [2006], the estimation of parameter covariance Σ_θ from the Hessian requires that a log likelihood objective $g(\theta) = \ell(\theta)$ is used. This method obtains directly the parameter covariance Σ_θ which can be used to construct an asymptotic confidence region as in Eq. (14).

2.3.4 Profile likelihood

As discussed in Section 2.2.2, *likelihood based* CIs are often considered superior to *asymptotic* CIs Raue et al. [2009]. Further, parameter *distributions* are arguably a more descriptive representation of uncertainty than confidence regions. Hence, obtaining parameter distributions based on the likelihood function is an attractive tool for parameter analysis. An elegant method for computing such distributions is the *profile likelihood* (PL) method presented in Deconinck and Roels [2017], Murphy and Van der Vaart [2000], Raue et al. [2009]. The PL method explores the parameter space by optimising the parameters in *two* steps, rather than simultaneously as in Eq. (1). The *profile likelihood* $\ell_{\mathrm{PL}}(\theta_i)$ is defined as the minimum log likelihood for a given θ_i when the remaining parameters are freely optimised Raue et al. [2009]:

$$\ell_{\mathrm{PL}}(\theta_i) = \min_{\theta_{j\neq i}}\ell(\theta_{j\neq i};\mathcal{M},\mathcal{K},\theta_i) \tag{20}$$

Values of θ_i must be chosen prior to optimising the remaining $\theta_{j\neq i}$ Raue et al. [2009]. A straightforward solution, if the objective function g is well behaved within the constraints of Θ, is to use a *brute force* approach with an even sampling of θ_i. Alternatively, a two-sided gradient decent algorithm, using a freely optimized parameter vector as a starting point, can be applied Maiwald and Timmer [2008], Raue et al. [2009]. The resulting likelihood distribution can be plotted as a function of θ_i and subsequently analysed according to the definitions of structural and practical identifiability for *likelihood based confidence intervals* Deconinck and Roels [2017], as discussed in Section 2.2.2. A threshold can be applied to the constructed profile, as described in Section 2.2.2, where, by *Wilks'*

Theorem Wilks [1938], the threshold Δ_α can be drawn from the $\chi^2_{\alpha,n_{\text{df}}}$ distribution. The freely estimated $\hat{\theta}$ has n_θ degrees of freedom (d.o.f.), while the PL estimate has $n_\theta - 1$ d.o.f., hence the threshold Δ_α is computed with $n_{\text{df}} = 1$.

Observe that since the PL method essentially projects the n_θ dimensional space Θ onto the single parameter θ_i, by freely estimating the remaining parameters, the PL method tends to overestimate the width of the likelihood based confidence interval if parameters are not independent, as discussed in Section 2.2.7

2.3.5 Two-dimensional profile likelihood

In order to improve the PL methods projections under the influence of inter-dependent parameters, the method can be modified to hold out *two* parameters rather than one, i.e.:

$$\ell_{\text{PL2}}\left(\theta_i, \theta_j\right) = \min_{\theta_{k \neq i,j}} \ell\left(\theta_{k \neq i,j}; \mathcal{M}, \mathcal{K}, \theta_i, \theta_j\right) \quad (21)$$

This projects the parameter space Θ onto the plane of θ_i and θ_j; $\Theta_{i,j}$, which results in a two-dimensional distribution that can be analysed in a similar way to the one-dimensional PL Raue et al. [2009], using the definition in Eq. (18) and discussed in Section 2.2. The PL2 results can be plotted as topological surfaces Raue et al. [2009], which can be used to diagnose parameter *inter-dependence*, since the two-dimensional projections are capable of representing *relationships* between parameters. These projections constitute an exhaustive search over the plane $\Theta_{i,j}$. Hence, both local and global optima can be obtained from inspection of the projected profiles.

Applying a confidence threshold to the PL2 method produces *confidence regions* in the $\Theta_{i,j}$ plane. Based on confidence thresholds computed from the χ^2-distribution, a similar interpretation of these two-dimensional topologies can be applied to diagnose identifiability by requiring that the region is bounded in all directions Raue et al. [2009]. If there is an unbounded equipotential *valley* with a *constant optimal* log likelihood, the parameter is structurally non-identifiable. If the interval or region is unbounded in some direction but still has a well defined optimum, this indicates a practically non-identifiable parameter Raue et al. [2009]. Observe that since $\hat{\theta}$ has n_θ free parameters while the PL2 estimate has $n_\theta - 2$, this gives $n_{\text{df}} = 2$ for the computation of Δ_α from the χ^2-distribution in Eq. (18).

While the extension of the PL method to create projections in the plane $\Theta_{i,j}$ is intuitive, and the resulting plots exhibit some interesting characteristics as tools for analysing parameter identifiability and inter-dependence, this modification strongly increases the computation time of the method. To create the projections of $\ell\left(\theta\right)$ onto $\Theta_{i,j}$, a large number of objective function evaluations must be performed. Using a brute force sampling of $\Theta_{i,j}$ with N steps for each parameter returns N^2 pairs of parameter values, each of which requires optimisation of the remaining parameters; $\theta_{k \neq i,j}$. This process must be repeated for each combination of parameters, which further increases the computational burden. Hence, the method requires careful use of parallelisation and software engineering to be computationally feasible. Of particular importance is utilising the fact that neighbouring points in $\Theta_{i,j}$ are likely to have similar optimal values for $\theta_{k \neq i,j}$. Hence, using previously optimised free parameters as a *warm-start* for computing new $\ell_{\text{PL2}}\left(\theta_i, \theta_j\right)$ points significantly reduces computation time.

Due to the extensive computation time for this method, it is advisable to initially perform exploratory analysis with relatively low number of steps N, with subsequent higher resolution analysis in specific regions of interest. However, the initial analysis must use a discretisation resolution sufficiently detailed to find the regions of interest. The number of resolution steps for each parameter which is required for successful application of the PL2 method depends on the problem, and should be found by experimentation.

Finally, observe that when a brute force discretisation of $\Theta_{i,j}$ is used, the resulting set of optimised parameters constitutes an exhaustive search of the discretised parameter space Θ. Hence, an estimate $\hat{\theta}$, which is *globally optimal* within the accuracy and bounds allowed by the brute force discretised Θ, can be obtained by taking the minimum from all the $\ell_{\text{PL2}}\left(\theta_i, \theta_j\right)$ profiles.

2.3.6 Bootstrapping for dynamic models

The data samples of a time-series are not independent. Hence, the traditional bootstrapping method of randomly drawing individual samples with replacement is not applicable, because the sample to sample dependency information would be lost in the generated pseudo data-set Kunsch [1989], Politis [2003]. A popular modification of the bootstrapping method is to divide the original data into blocks, either overlapping or non-overlapping, with uniform or randomly chosen lengths and/or starting points Politis [2003]. In this paper, two versions of block based bootstrapping for time series data is considered; *non-overlapping block bootstrap* Lodhi and Gilbert [2011] and *stationary bootstrap* Politis and Romano [1994]. The difference between these two approaches is in how the data is separated into blocks. Each method is outlined below.

The idea behind all bootstrap methods is to generate multiple pseudo data-sets, in order to estimate the
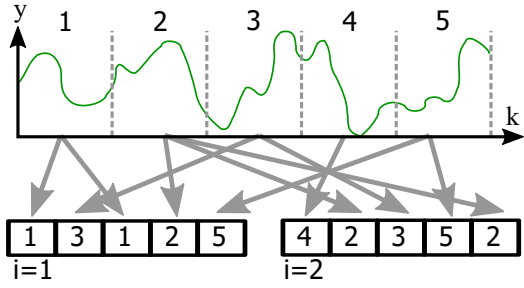
Figure 5: An example of simple block bootstrapping with $K = 5$ blocks, creating $M = 2$ pseudo data-sets.

Figure 6: Simplified block diagram of Stationary Bootstrapping.

variance of some estimate, e.g., dynamic model parameters. Hence, the bootstrapping procedure must be repeated M times, such that each iteration produces a *different* pseudo data-set, and hence a *different* parameter estimate $\hat{\theta}_j$. Note that any objective function could potentially be combined with this type of bootstrapping, e.g., the ballistic SSE approach in Section 2.1.3 or the likelihood $\ell(\theta)$ in Section 2.1.2.

Based on these M estimated parameter vectors, the *mean* parameter estimate, and the *covariance* of that mean estimate can be computed, i.e:

$$\hat{\theta} = \frac{1}{M} \sum_{j=1}^{M} \hat{\theta}_j$$

$$\Sigma_\theta = \frac{1}{M-1} \sum_{j=1}^{M} \left(\hat{\theta}_j - \hat{\theta} \right)^2 \tag{22}$$

Confidence regions on the form of Eq. 14 can then be constructed for the *mean parameter estimate*, where the threshold Δ is drawn from the *F-distribution* Johnson and Wichern [2007].

Additionally, the resulting M parameter estimates can be plotted as scatter plots or as histograms, either for individual parameters or for combinations of two parameters. Observe that these plots suffer from the same limitations related to the projection of high dimensional parameter space Θ onto single parameter axis as discussed in Section 2.2.7.

The first bootstrap method, *non-overlapping block bootstrapping*, is achieved by dividing the data-set $y_{[N]}$ and $u_{[N]}$ into K blocks of length l. Let $y_{[l]}^{(i)}$ and $u_{[l]}^{(i)}$ be block i of measured system outputs and inputs, respectively, where $i \in \{1, ..., K\}$. Each block is constructed by taking a consecutive sequence of samples from the original data, $y_{[N]}$ and $u_{[N]}$, starting from

Figure 7: Illustrative example of two iterations of stationary bootstrapping, resulting in five blocks of data, possibly overlapping, with uniformly drawn starting point and geometric length distribution.

sample $l \cdot (i-1)$, such that:

$$y_{[l]}^{(i)} = \left[y_{l \cdot (i-1)+k} \ : \ k \in [1, ..., l] \right] \tag{23}$$

$$u_{[l]}^{(i)} = \left[u_{l \cdot (i-1)+k} \ : \ k \in [1, ..., l] \right] \tag{24}$$

A pseudo time-series data-set is then created by drawing K blocks randomly with replacement, as illustrated in Fig. 5. As with traditional bootstrapping, some blocks will not be drawn, while others may be drawn multiple times. Hence, some data points will not appear in the new pseudo data-set, while other data points will appear multiple times. This is shown in Fig. 5, where for the first pseudo data-set block 1 appears twice, as the 1st and 3rd block, while block 4 is not included.

An alternate method for drawing random blocks of data is the *stationary bootstrapping* method Politis and Romano [1994], where blocks of data are constructed with a random length of geometric distribution Politis [2003]. The advantage of this approach is to create

bootstrapped data-sets that are themself stationary series Politis [2003], Politis and Romano [1994].

The stationary bootstrap method is realised by use of a probability test, and consists of two steps. First, the starting point of each block is drawn uniformly across all N original samples, $k_0 \sim \mathcal{U}(0, N)$. Next, with probability $1-p$, add the next consecutive sample from the original series, or with probability p, start a new block by again uniformly drawing a new starting point. This test on p is repeated until the combined length of all blocks is approximately N. The resulting blocks length will follow a geometric distribution such that the probability of block i having length m is $P_r(l_i = m) = (1-p)^{m-1} p$. This process is illustrated in Fig. 6. The expected length of each block is $E(l) = \frac{1}{p}$ and the expected total number of blocks is $E(K) = \frac{N}{p}$. An illustrative example of two iterations of stationary bootstrapping is shown in Fig. 7. Comparing Fig. 7 to the non-overlapping block bootstrapping in Fig. 5 shows the difference in the two methods, in that the first method has non-overlapping blocks of uniform length, which is randomly recombined to create the pseudo data-set, while the stationary bootstrap method uses randomisation to choose both the start and length of each block.

Since both these approaches, indeed, all block based bootstrapping methods for time-series data, involve dividing the original time-series data into blocks and recombining them to form new pseudo data-sets, the question of how to join together multiple randomly selected blocks into a new complete data-set arises Kunsch [1989]. For estimation of parameters for dynamic system models using a data-set that is essentially segmented non-consecutive blocks, an intuitive solution is to compute the objective function $g(\theta)$ for each block and aggregate the results. If the objective is defined on summation form as in Eq. (6), the overall objective function for a block segmented data-set of K blocks can be defined:

$$g_{\mathrm{B}}\left(\theta; \mathcal{M}, y_{[N]}, u_{[N]}\right) = \sum_{i=1}^{K} g^{(i)}\left(\theta; \mathcal{M}, y_{[l_i]}^{(i)}, u_{[l_i]}^{(i)}\right)$$

The initial conditions for evaluating the objective for each block, $g^{(i)}$, such as the initial state, must be determined for each block, rather than for the whole data-set as in Eq. (1). If the states are measurable, the choice of initial state for each block can be obtained from the measurements. Alternatively, the initial state can be treated as an unknown parameter and estimated for each randomly drawn block.

An important consideration when performing block based bootstrapping on time series data for dynamic model parameter estimation, is the *consistency* of the dynamic information in the data. If certain segments

of the data contain significantly *less* dynamic information than the rest, e.g., if the system is in *steady state* for parts of the original data-set, some iterations of the bootstrap procedure may return pseudo data-sets that are *less informative* w.r.t. parameter estimation. These pseudo data-sets may produce *practically non-identifiable* parameters Raue et al. [2009], which manifest as *outliers* among the M bootstrap estimates. Such outliers will significantly effect the computed covariance in Eq. (22). Hence, it is important to consider the *consistency* of dynamic information in the original data, prior to applying bootstrapping methods.

2.3.7 Consistency of dynamic information in calibration data

An intuitive method for testing the *consistency* of dynamic information content in data is to draw a set of overlapping, consecutive, data segments, taken equidistant across the data-set. Each segment is of length l, and extracted from starting points $w \cdot (i - 1)$, where w is the step length;

$$y_{[l]}^{(i)} = \left[y_{w \cdot (i-1)+k} : k \in [1, ..., l]\right] \tag{25}$$

$$u_{[l]}^{(i)} = \left[u_{w \cdot (i-1)+k} : k \in [1, ..., l]\right] \tag{26}$$

The approach constitutes a *moving window* that travels across the data-set with step length w. The segment length l and the step length w are considered tuning parameters and should be determined experimentally. For each segment, a parameter vector $\hat{\theta}^{(i)}$ is estimated by minimising the objective $g^{(i)}\left(\theta; \mathcal{M}, y_{[l]}^{(i)}, u_{[l]}^{(i)}\right)$. Note that this is fundamentally different from the bootstrapping approach, since no randomisation is used to combine multiple segments and the parameter estimation is performed separately for each consecutive segment. As for the block based bootstrapping methods in Section 2.3.6, the initial conditions needed to evaluate $g^{(i)}$ must be obtained for each segment, either as estimated parameters or directly from observations if the states are measurable.

For *each* segment, some appropriate method of uncertainty estimation, e.g., the *Hessian* method of Section 2.3.3 or the *Profile Likelihood* method in Section 2.3.4, is used to evaluate the *uncertainty* and/or *identifiability* of the estimated parameters. By plotting the results as a function of the segment starting point $w \cdot (i - 1)$, and observing how the parameter uncertainty and/or identifiability changes with time as the window is moved, the consistency of dynamic information in the data can be evaluated. Observe also that if the PL method is used, the results should be plotted as the relative log likelihood $\ell(\theta) - \ell\left(\hat{\theta}_i\right)$, since the optimal log likelihood will be different for each segment.

If parameter calibration from different segments of the data produce significantly different uncertainty estimates, this indicates an inconsistency in dynamic information, which subsequently can influence uncertainty estimation methods based on block bootstrapping. Observe that since a small subset of the calibration data is used, the uncertainty estimates for each segment will be larger than what is obtained using the complete original data-set.

In addition to test the consistency of dynamic information by estimating the *uncertainty* for each step, the method also produces an estimate of the optimal *parameters* $\hat{\theta}^{(i)}$ for each segment. These estimates can be used to test if the optimal model parameters change over time for a specific data-set. If the parameters are interpreted as constants of the physical system, time variation of θ can be an indication of unsatisfactory calibration data. Arguably, *unexpected* time variation of parameters may also indicate an oversimplified model structure, such that the calibrated parameters are affected by unmodelled time-varying disturbances, resulting in variations in the parameter estimates over time.

2.4 Summary

Section 2 of this paper has presented the theoretical foundation for a number of methods that can be used to analyse the parameter estimation problem for dynamic models, in particular the *identifiability* and *uncertainty* of the estimated parameter. Which method is best suited for a particular application largely depends on the application, and what type of analysis is of interest. The aim of these methods is to obtain *accurate* dynamic system models, but also to *validate* the estimated parameters in the presence of *aleatoric* and *epistemic* uncertainty. In many applications the choices for experimental design is limited. Hence, parameters must be estimated under less than ideal conditions. It is especially important in these cases to carefully analyse the resulting parameter estimates in the context of *identifiability* and *uncertainty*. In engineering applications, parameters are often assumed, quite reasonably from a detailed physical understanding of the underlying system, to be constants of the physical system. However, due to the effects of measurement noise, unmodelled disturbances, insufficient dynamic information, modelling errors and simplifications, etc., it may not be possible to obtain an unambiguous *estimate* of the parameters. Hence, the methods presented in this section can be valuable engineering tools for providing a thorough analysis of the parameter estimation problem. In the sequel, examples of the application of these methods to two experimental cases are presented. These examples illustrate the kind of insight that can be gained by applying the methods to practical parameter estimation problems.

3 Experimental cases

In this section, the methods presented in Section 2 are demonstrated on two test cases. The first case is a simple first order model with a single input. The parameters of the model are calibrated using data obtained by simulating the same model, with added, randomly generated, measurement noise. Hence, there is no *epistemic* uncertainty in the parameter estimation, only the *aleatoric* uncertainty of the output measurement noise. The second case is an example of a grey-box model, specifically a thermal network model of a building, which aims to predict temperature variations. These models are particularly interesting from a parameter estimation and analysis perspective, since they are constructed *cognitively* based on *naive* physics, and hence have significant *epistemic* uncertainty in them.

3.1 First order dynamic model

A first order model with input is defined as:

$$\dot{x} = -ax + bu \tag{27}$$

$$y = x + v_k \tag{28}$$

where $v_k \sim \mathcal{N}(0, \mathcal{V})$ is the Gaussian distributed measurement noise, u is the model input and the parameters are $\theta = [a, b]$. By Laplace transformation, the transfer function from input to output is obtained as:

$$H(s) = \frac{y(s)}{u(s)} = \frac{b}{s+a} = \frac{K}{\tau s + 1}, \tag{29}$$

which is a low-pass filter with exogenous input, with gain $K = \frac{b}{a}$ and time constant $\tau = \frac{1}{a}$. The output y is hence simply the low-pass filtered input u plus the measurement noise.

The model is excited with six different input signals, each a total of 10 second of data at $\Delta t = 0.01$, presented in Fig. 8. The six data-sets are chosen to demonstrate the effect of different types of excitation on the various methods to be tested. As shown in Fig. 8, the first three data-sets are a step (STP), a square wave (SQR), and a sine wave (SIN), where short-hand names are given to simplify tabulating results in the sequel. The square and sine wave have a signal period $T = 2s$. The remaining three excitation signals are *pseudo random binary sequences* (PRBS), i.e. signals generated by drawing a sequence of *random* binary numbers and transforming those into a non-uniform square wave signal. The length in time of each bit, i.e., bit-length, is $0.1s$ (PR1), $0.2s$ (PR2) and $0.5s$ (PR5) for the last
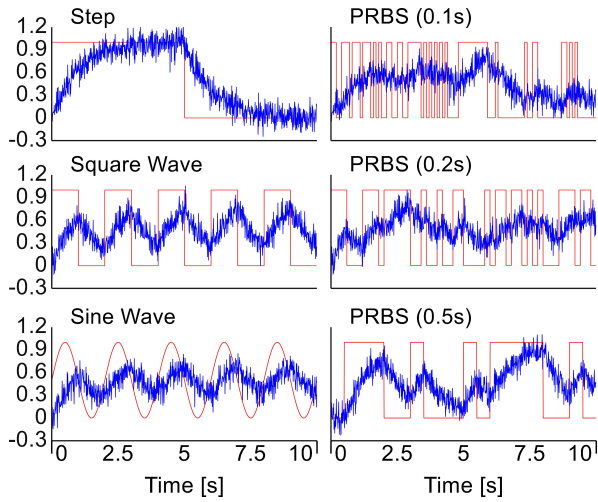
Figure 8: Six different input signals (red) are used to excite the first test case model. For each signal, the model is simulated deterministically to obtain the output (blue). The three left panels show the Step, Square wave and Sine wave signals. The Square wave and Sine wave both have a period of $T = 2s$. The right panels show the three pseudo random binary sequence (PRBS) signals, which differ in what length in time each bit in the sequence represents ($0.1s$, $0.2s$ or $0.5s$).

Figure 9: Cumulative Periodograms for all six datasets show that the residuals are close to Gaussian distributed, well within the 95% confidence bounds.

three data-sets. Hence, a value of true/false in the PRBS indicates input $u = 1/u = 0$ for $0.1s$, $0.2s$ or $0.5s$, respectively.

The model is simulated for each of the six input signals, with parameters $a = 1$, $b = 1$ and added measurement noise $v_k \sim \mathcal{N}\left(0, 0.1^2\right)$, to obtain an output. Hence, for this model, the true parameter vector θ^* is known. The resulting input-output data-sets are used as $y_{[N]}$ and $u_{[N]}$ in the following tests. Since there is no diffusion term in the state transition in Eq. (27), and the calibration data is simulated with the same model for which parameter analysis is performed, the model is exact, hence $\mathcal{W} = 0$. As expected, due to the simplicity of the model, and the simulated data-set, the residuals are close to Gaussian, as shown by the CP diagrams in Fig. 9.

3.1.1 Uninformed Random Picking and Profile likelihood

When starting to analyse the parameter space, a good first step is to visualise the shape of the objective function $g(\theta)$ in the parameter space Θ. Usually, the existence of a well defined, *unambiguous*, optimum $\hat{\theta}$ is not known initially. Hence, a good starting method is the *Uninformed Random Picking (URP)* method described

in Section 2.3.1. The result of using URP as an exploratory tool on the first order model is shown in Fig. 10. Additionally, the *Profile Likelihood* (PL1) method is applied to show that both methods obtain the same optimal *front* across Θ. Observe from Fig. 10 that the grey dots correspond to each of $K = 50.000$ randomly drawn θ_j, each simulated to compute $g(\theta_j)$, while the red line is the PL1 profile. The PL1 profile corresponds closely to the optimal front obtained by the URP.

Plotting the results together with the likelihood profile, shows that the same information, the shape of the objective, is obtained by both methods. Hence, it is interesting to compare the methods on computation time and implementation. For this simple model, the execution time of URP ($K = 50.000$) and PL1(500 steps in θ_i) are $12s$ vs $17s$, hence the computation time is short enough to be insignificant. However, for larger models, there may be significant differences. The PL1 method requires optimisation of $n_\theta - 1$ parameters for each step in θ_i, hence, a large number of parameters significantly increases the load on the optimisation algorithm. In contrast, the URP method requires no optimisation, but is affected by the dimensionality of Θ due to the dispersion of the randomly drawn points. With large number of parameters, K must be chosen large enough that the randomly drawn parameters reasonably covers the whole space Θ, which results in longer computation time.

An interesting observation when comparing PL1 and URP, since they both give essentially the same result, is that URP is significantly easier to implement, since it does not require an optimisation algorithm. For some applications, this may be a distinct advantage.

Next, observe that both the URP and the PL1 method rely on projections to plot the results as functions of a single parameter. These projections are known to overestimate the width of the pro-
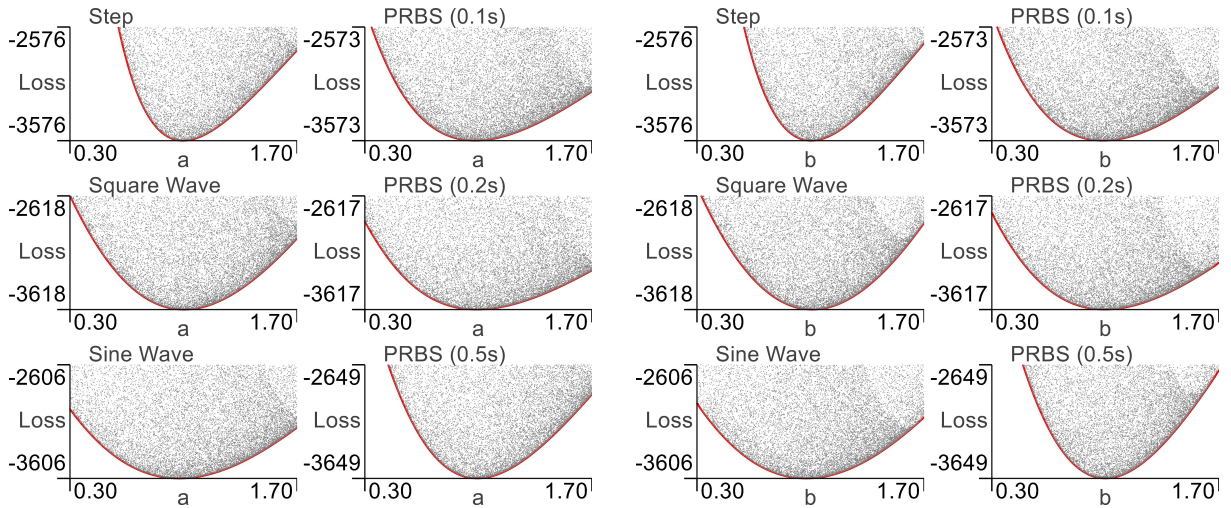
Figure 10: Exploratory analysis of the $\ell(\theta)$ objective using the PL1 (red) and the URP (grey) methods. Results for parameter a (left) and b (right) show that both parameters are unambiguously identifiable for all six data-sets.
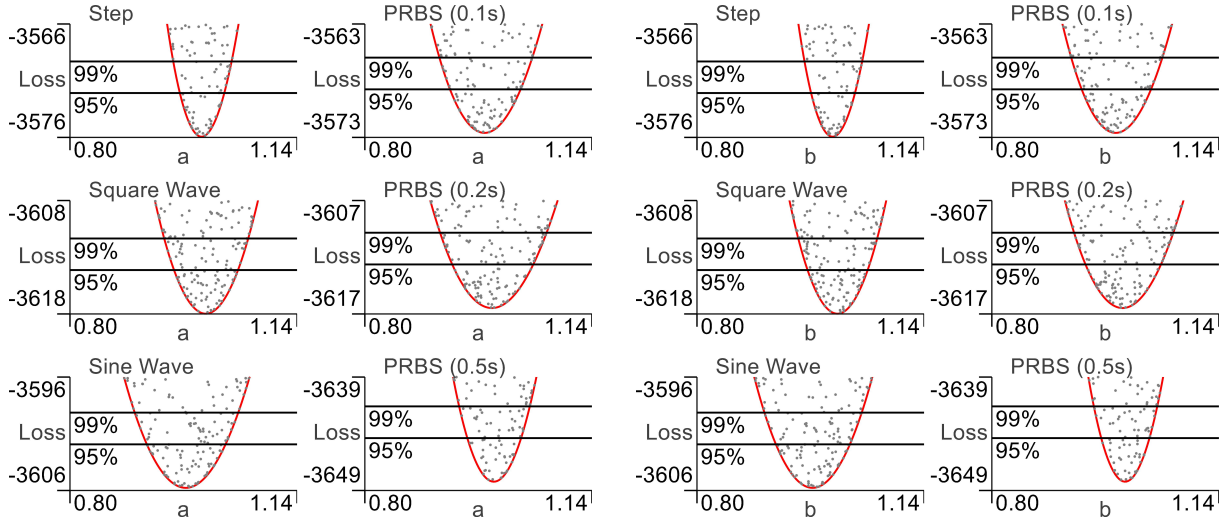

Figure 11: PL1 and URP results, zoomed in around the optimum, for parameter a (left) and b (right).

files/intervals, as discussed in Section 2.2.7.

Finally, observe that the plots in Fig. 10 are obtained as a form of exploratory analysis, hence with wide bounds on Θ and subsequently with a large range on the objective score axis; relative log likelihood $\ell(\theta) - \ell\left(\hat{\theta}\right) < 1000$ is used here. These plots are interesting as a first step, but for the purpose of estimating uncertainty of the optimal estimate $\hat{\theta}$, only the immediate neighbourhood of $\hat{\theta}$ is of interest. Hence, Fig. 11 shows the same results but with different scaling on the axis. Here, the width of Θ is significantly reduced, and also the range in objective score is reduced to a more reasonable 10. This likelihood range allows for adding confidence thresholds at α equal to 90% and 95%. From Fig. 11 it is immediately apparent that

the Step and PRBS (0.5s) data-sets produce narrower shapes around $\hat{\theta}$ than the other four data-sets, which indicates better estimation accuracy, i.e., tighter confidence bounds from the applied thresholds.

3.1.2 Randomised initial conditions

Another useful method, especially as an initial exploratory analysis tool, is the use of randomised initial conditions with subsequent optimisation, discussed in Sec. 2.3.2. The result of applying this method is shown in Fig. 12, where the results are plotted as a vs b, i.e., both parameters against each other. Hence, Fig. 12 shows the whole parameter space Θ for this model. As shown, the optimum $(1, 1)$ is obtained for
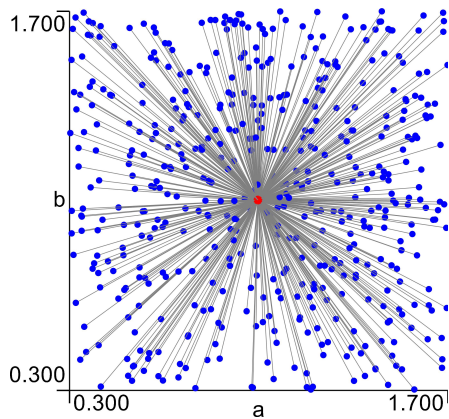
Figure 12: Randomised initial conditions shows that the optimum is globally unambiguous in Θ and obtained independent of the initial guess θ_0. Results are shown only for data-set Step (other five sets show the same behaviour).
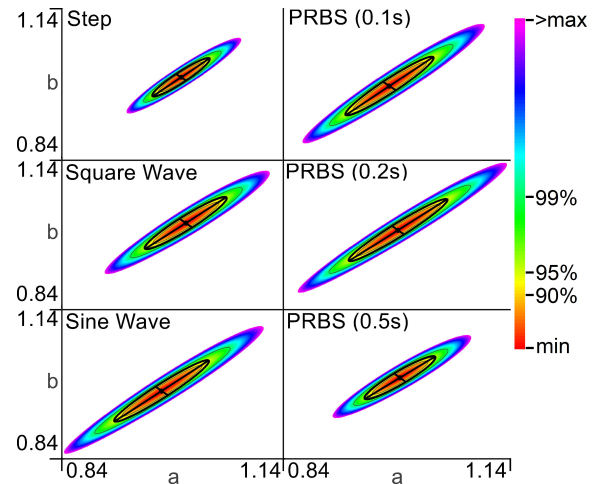

Figure 13: PL2 heat-maps with 90%, 95% and 99% confidence iso-lines, for all six data-sets, with added 95% confidence ellipses (thick line), computed from the Hessian of $\ell(\theta)$, for comparison.

all $K = 500$ randomly drawn initial guesses, which shows that the optimum is unambiguously obtainable in Θ, and not influenced by the initial guess θ_j^0. For comparison, see Section 2.3.2 where another example of this method is shown in Fig. 4 in which there are a large number of local minima. Together with the results in Section 3.1.1, Fig. 12 shows that the objective function has a well defined single global optimum. The major difference between results from the six different data-sets is the shape of the objective around the optimum, and subsequently the accuracy of the obtained parameter estimate, which will be further discussed in the sequel.

3.1.3 Profile Likelihood 2D and Hessian

A natural next step is to perform a detailed analysis of the neighbourhood around $\hat{\theta}$, i.e., the parameter ranges obtained from the PL1 analysis shown in Fig. 11. To analyse the parameter space, the two-dimensional Profile Likelihood (PL2) method from Section 2.3.5 is used. The results, shown in Fig. 13, use the same range for all six data-sets in order to directly compare the obtained profiles. For comparison, ellipses as in Eq. (14), computed by using the Hessian method from Section 2.3.3, are added to the PL2 plots. Observe first from Fig. 13 that the parameter distributions in Θ are very well approximated by the Hessian based ellipses. This is expected, due to the simplicity of the model and the simulated data with added Gaussian noise. Both methods use the same objective function $\ell(\theta)$, with the only difference being that the Hessian method assumes a quadratic distribution to compute elliptic regions.

Table 1: Standard deviations of parameters computed with the Hessian method.

Data	a	b	σ_a	σ_b	$\sqrt{\sigma_a \sigma_b}$
STP	0.997	1.003	0.016	0.016	0.015
SQR	1.003	1.010	0.022	0.021	0.021
SIN	0.974	0.973	0.026	0.025	0.025
PR1	0.980	0.986	0.024	0.024	0.023
PR2	0.991	0.996	0.027	0.025	0.025
PR5	0.993	1.000	0.019	0.017	0.018

Next, observe that the elliptic confidence regions in Fig. 13 are rotated at an approximately 45 degree angle, or equivalently from Table 1 that the covariance $\sigma_a \sigma_b$ between the two parameters is significant, compared to the variance of each variable. This indicates that the parameters are *dependent*, which is expected from Eq. (29), since $K = \frac{b}{a}$. The parametrisation of the model in Eq. (27) was chosen specifically to demonstrate this point. Subsequently, as discussed in Section 2.2.7, the PL1 projections from Fig. 11 are too wide. Indeed, by projecting the PL2 results in Fig. 13 onto each of the two parameter axes, the resulting profiles would be exactly the results from the PL1 method. Hence, it can be observed that the PL1 method significantly over-estimates the width of the parameter profiles due to parameter inter-dependence. Note that it is recommended to attempt resolving parameter inter-dependence by choosing a different parametrisation in Eq. (27), e.g. choosing the parameters K and τ such that the state transition equation becomes $\dot{x} = \frac{1}{\tau}(-x + Ku)$.
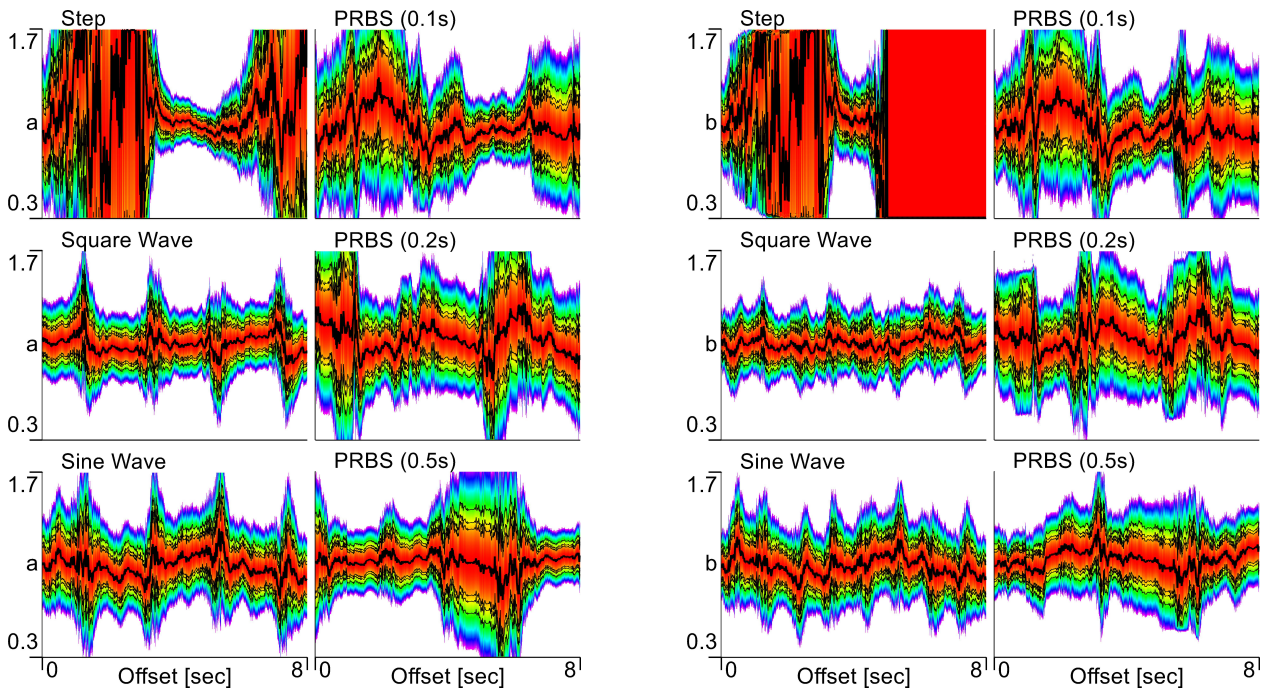
Figure 14: Consistency of dynamic information for identification parameters a (left) and b (right) is examined using a window length of $l = 200$ samples, equivalently 2 seconds. The window offset is varied on the interval $t_0 \in [0,\,8[$ seconds in steps of $w = 1$, or equivalently $\Delta t = 0.01$. At each step, the Profile Likelihood method is used to evaluate identifiability of the parameters using the information in the window. The results are plotted as a heat-map with time offset on the horizontal axis, and the thick line represents the optimal parameter estimate for each window. The parameters are examined on the interval $[0.3,\,1.7]$.

Finally, observe from inspection of Fig. 13 and the corresponding quantified standard deviations and covariance of the parameters in Table 1, that the Step and PRBS (0.5s) data-sets provide slightly more accurate estimates of the parameters, compared with the other four. The variations in parameter estimation uncertainty, and correspondingly the shape of the neighbourhood around $\hat{\theta}$ in Θ, are caused by the use of different excitation signals. Hence, the differences between the results illustrate the well known fact that the choice of excitation signal during experiments directly influences the parameter estimation uncertainty.

3.1.4 Consistency of dynamic information

Next, it is of interest to assess the consistency of dynamic information in the data-sets, using the Moving Window method described in Section 2.3.7. The results of applying the Profile Likelihood (PL), described in Section 2.3.4, to data segments of length l taken equidistantly across the original data with step length w, is shown in Fig. 14. The PL method provides better estimates of the uncertainty and *identifiability* for the data in each step of the moving window, compared with

the Hessian, since it can represent *asymmetric* distributions. The results, which for this method is a function of parameter θ_i and the time offset $w \cdot (i - 1)$, are plotted as heat-maps with confidence iso-lines at 90%, 95% and 99%. Figure 14 shows that there is a significant difference between the Step data-set and the other five sets in that the Step data-set has large segments where the parameter uncertainty is high, i.e., large equipotential bands in the parameter direction. This indicates that there is insufficient dynamic information in these segments of the data to obtain good parameter estimates. Observe also that for the Square and Sine wave data-sets, the results are the least affected by the window offset, hence, these data-sets contain the most consistent dynamic information. Similarly, the optimal estimate, marked by a thick black line in Fig. 14, is showing significant fluctuations for the Step data-set, while for the Square and Sine Wave data-sets, the estimates are mostly consistent w.r.t. the time window offset.

These considerations will be especially important in the sequel, when block based bootstrapping methods are used, but the results are also interesting in themselves, as a way to test the dynamic information con-

Table 2: Bootstrap results ($M = 200$ iterations), Case A: Simple ($K = 10$), Case B: Simple ($K = 5$), Case C: Stationary ($p = 0.005$).

Data	#	a	b	σ_a	σ_b	$\sqrt{\sigma_a \sigma_b}$
STP	A	0.958	0.985	0.080	0.073	0.074
	B	1.009	1.018	0.063	0.058	0.060
	C	1.013	1.015	0.065	0.070	0.061
SQR	A	1.049	1.033	0.026	0.015	0.019
	B	1.005	1.025	0.014	0.007	0.008
	C	1.001	1.014	0.018	0.016	0.015
SIN	A	0.975	0.965	0.017	0.025	0.015
	B	0.970	0.970	0.018	0.021	0.016
	C	0.977	0.977	0.028	0.025	0.026
PR1	A	0.976	1.006	0.068	0.058	0.059
	B	0.979	0.983	0.024	0.028	0.025
	C	0.949	0.951	0.033	0.034	0.032
PR2	A	1.063	1.040	0.051	0.039	0.043
	B	1.026	1.032	0.031	0.018	0.023
	C	0.999	1.007	0.029	0.026	0.027
PR5	A	0.991	0.997	0.046	0.040	0.038
	B	0.994	1.011	0.025	0.036	0.030
	C	1.014	1.027	0.020	0.023	0.020

Figure 15: Histogram (top) and scatter plot (bottom) showing how data with poor dynamic information content (left panels, Step dataset) induces outliers in the results. Plots are obtained using Stationary bootstrap with tuning parameter probability $p = 0.005$ and $M = 10000$ iterations.

tent of different excitation signals, especially when using calibration data obtained from physical systems with limited choices in the experimental design.

3.1.5 Bootstrapping

The use of bootstrapping methods for dynamic data to estimate the uncertainty of estimated parameters, is discussed in Section 2.3.6. Here, the simple block bootstrap, with block lengths $l = 100$ and $l = 200$, respectively 1 and 2 seconds of data, is tested and compared with the Stationary Bootstrap method using $p = 0.005$. The results, after $M = 200$ iterations, are presented in Table 2. First, observe that, as expected based on the results in Section 3.1.4, the Step data-set shows considerably higher covariance than the other data-sets. Since large segments of the Step data-set does not contain sufficient dynamic information for parameter estimation, some of the randomised pseudo-data-sets created by bootstrapping does not contain enough information to estimate parameters, hence the higher covariance. This is further illustrated by Fig. 15, which shows how the Step data-set produces significantly larger spread in parameter estimates, compared with the Square Wave data-set. Note that a much higher number of iterations, $M = 10.000$, was used for Fig. 15 in order to obtain a good histogram illustration. From Section 3.1.4, the Square and Sine Wave data-sets are known to have significantly better consistency of dynamic information across the data-set
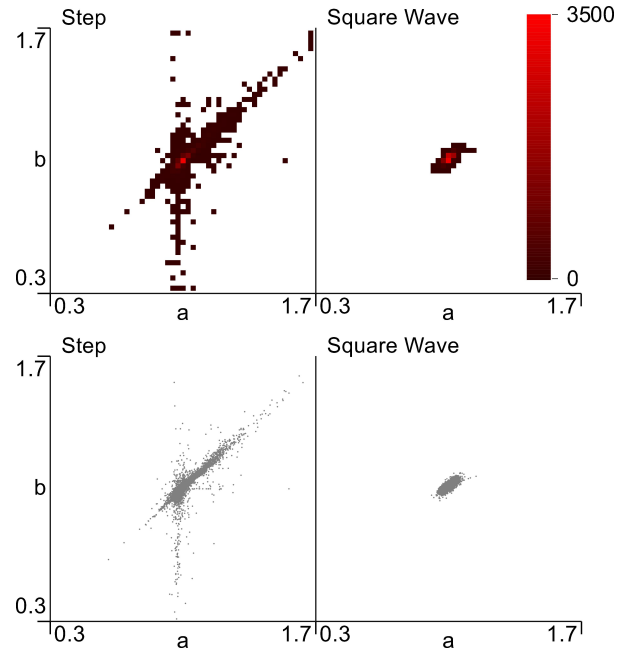
compared with the Step data-set. Hence, more consistent parameter estimates with lower variance is obtained from the block-based bootstrapping methods.

Next, observe that the Stationary Bootstrap method produces approximately the same results as the block based bootstrapping with $l = 200$, in this case. This may be explained by the Stationary Bootstrap using $p = 0.005$ which gives an expected block length also of 200.

Finally, observe that for the datasets for which the dynamic information is of sufficient consistency, the estimates of the parameter uncertainty Σ_θ for cases B and C in Table 2 are similar to those obtained from the Hessian method in Table 1.

3.1.6 Frequency information in input and output

A commonly used method of examining dynamic information content in data is to compute a frequency spectrum using the Fast Fourier Transform (FFT) algorithm. Due to its widespread use and popularity, computationally efficient implementations exist, which makes this an easily accessible tool for analysing data. Applying FFT to the six data-sets, both the input signal and the measured output with noise, gives the res-

ults shown in Fig. 16. Comparing the FFT results to those obtained by the parameter analysis methods presented previously provides some interesting insight into the *differences* in results obtained from each of the six excitation signals.

First, comparing the Sine and Square wave data-sets, observe that the Sine wave has only one frequency component at $f = 0.5\,Hz$, excluding the mean signal level component at $0\,Hz$, while the Square Wave contains also higher order harmonics of the base frequency. However, since the model is essentially a low-pass filter with a critical frequency of $\frac{1}{2\pi}$, these higher order harmonics are damped, thus having only limited effect on the model output. Note however that despite having almost identical frequency information in the output y, the spectra for the input u differ significantly. The fact that these higher order harmonics in the input spectra are damped, thus not significantly infuencing the output y, is also informative w.r.t. the input-output relationship of the model, thus producing slightly smaller confidence regions for the Square Wave data-set.

A similar observation can be made from comparing the three PRBS data-sets. For the sets with bit-length $0.1s$ and $0.2s$ more of the input signal power is located in the damped frequency region of the model. Hence, the PRBS signal with bit-length $0.5s$ produces better parameter identification results, since more of the frequency information is passed through the model.

Finally, comparing the Step and Square Wave data-sets shows why the Step data-set produces the narrowest confidence regions from parameter estimation. Observe that the Step data-set contains the most signal strength in the frequency pass-band of the model. Since more of the information in the input u affects the output y, the parameter estimation methods produce estimates with lower uncertainty.

3.1.7 Comparing excitation signals

An interesting observation can be made from comparing the results of the various methods for all six data-sets. While the Step data-set gives the *best* estimation accuracy for the Likelihood based methods, such as the Hessian curvature method and the Profile Likelihood method, it also produces the *worst* estimation accuracy when *bootstrapping* is used. The reason for this can be observed from the consistency plot in Fig. 14. While the Step data-set contains segments of data that are largely *uninformative* for the purpose of estimating dynamic model parameters, the segments that *do* contain sufficient information produce the highest accuracy estimates. The PL confidence bands produced when the step change in the input is included in the moving window are the *tightest* among all the results produced, hence, give the lowest estimation uncertainty. How-
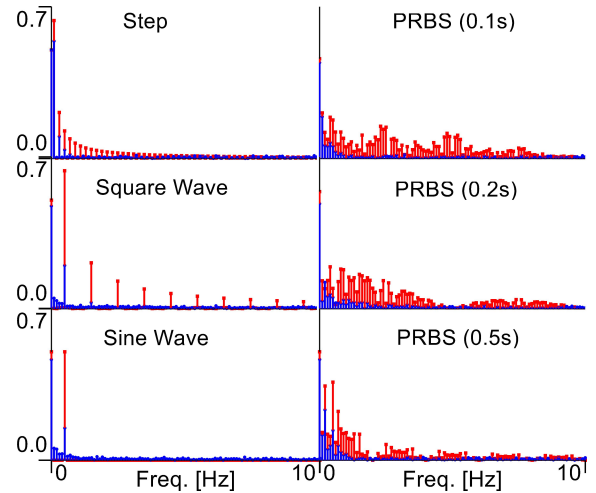

Figure 16: The Fast Fourier Transform (FFT) can be used to obtain a frequency domain representation of the dynamic information content in both input and simulated output. Input u (red) and output, including measurement noise, y (blue) is shown.

ever, since bootstrapping randomly selects segments of data, there will be some bootstrapped pseudo data-sets that do not contain data from the informative segment of the Step data-set and therefore produce outlier parameter estimates such as the ones shown in Fig. 15.

This example shows that assessing dynamic information content for model calibration is not straightforward, even in this simple case. Hence, it is useful to apply a method for evaluating the consistency of dynamic information, such as the one presented in Section 2.3.7.

3.1.8 Computation time

An important consideration for any numerical estimation method is the computation time it takes to obtain results. Typically, computation time depends on hyper parameters of the method, such as the resolution of likelihood profiles or the number of iterations for bootstrapping and randomisation based methods. Further, computation time also depends on the dimensionality n_θ of the parameter space Θ. Some selected examples of computation times for the previously presented results are given in Table 3. Since all the parameter estimation methods are based on a large number of simulations of a known model structure with varying parameters θ and a set of measurement data $y_{[N]}$ and $u_{[N]}$, the computation time is also influenced by the complexity of the model, the choice of Kalman Filter implementation for computation of residuals and the size N of the calibration data-set. Further, since

233

Table 3: Computation time for the LP model from the Step data-set. The other data-sets produce comparable execution times.

Method	Time
PL1 (resolution 500)	$\sim 13s$
PL1 (resolution 5000)	$\sim 90s$
URP ($K = 50.000$)	$\sim 19s$
URP ($K = 500.000$)	$\sim 160s$
Rand. Initial Conditions ($K = 50$)	$\sim 7s$
Rand. Initial Conditions ($K = 500$)	$\sim 57s$
Moving Wnd. (res. 200, $w = 1$, $l = 200$)	$\sim 575s$
Moving Wnd. (res. 200, $w = 1$, $l = 100$)	$\sim 317s$
Moving Wnd. (res. 100, $w = 1$, $l = 200$)	$\sim 311s$
Moving Wnd. (res. 200, $w = 10$, $l = 200$)	$\sim 57s$
PL2 (400×400 resolution)	$\sim 2800s$
Hessian	$\sim 1s$
Bootstrap A ($M = 200$)	$\sim 14s$
Bootstrap B($M = 200$)	$\sim 23s$
Bootstrap C($M = 200$)	$\sim 24s$
Bootstrap C($M = 10.000$)	$\sim 990s$

most of the presented methods use numerical optimisation, computation times can be influenced by optimisation related effects, such as variations in computation time due to obtaining estimates at different local minima. Hence, it is interesting to compare and discuss the computation time for a known model and data-set.

First, observe that for the PL1 and randomisation based methods, i.e., URP and Random Initial Conditions, the computation time is approximately linear in the resolution/randomised iterations. However, the number of iterations of the randomisation methods required to adequately explore the parameter space Θ depends on the dimensionality. The PL1 method however projects the likelihood function of the parameter space onto each parameter axis. Hence, the effect of high dimension parameter spaces will be more significant for the randomisation based methods than the PL1 method.

Next, comparing the PL2 exhaustive search of Θ with the ellipsoid approximation obtained from computing the Hessian, the difference in computation time is around three orders of magnitude. Further, the PL2 method projects the likelihood function of the entire parameter space Θ onto each possible plane. Hence, the computational time increases exponentially with the number of parameters n_θ.

The computation time of the Moving Window analysis, which applies the PL1 method on a moving window sub-set of the data, is shown to be approximately linear in the step length w. This is expected since the step length directly determines for how many windows of data the PL1 method is executed. Further, the computation time is also linear in the window length l. The number of window data sub-sets is $n_\mathrm{w} = \frac{N-l}{w}$ which is only somewhat affected by l. However, the PL1 method is approximately linear in the length of the data used, which results in the computation time for the Moving Window analysis being also approximately linear in l. Finally, the computation time is linear in the PL1 *resolution*, which was previously shown for the PL1 method applied to the full data-set. Naturally, the same applies when the method is used on a small sub-set of the data.

Finally, the computation time is approximately the same for the bootstrap methods in cases B and C. This is expected, since Case B uses blocks of fixed length $l = 200$ while the Stationary Bootstrap method in Case C uses $p = 0.005$, which gives the average block length $E(l) = \frac{1}{p} = 200$. Comparing this to Bootstrap Case A shows that both bootstrap methods are approximately linear in the expected block length. Additionally, since bootstrapping must be repeated M times in order to simulate running M experiments, the computation time is also approximately linear in M.

3.2 Thermal network model of a building

The second test case consists of a thermal network model of a small experimental building located at the Porsgrunn Campus of the University of South-Eastern Norway (USN). Thermal network models are created *cognitively* based on *naive* physical descriptions of the thermodynamics of the buildings, and can be expressed as Resistor-Capacitor (RC) circuit *analog* models Berthou et al. [2014], Deconinck and Roels [2017], Fux et al. [2014], Madsen and Holst [1995]. Specifically, the R3C2 model, partially based on the R4C2 model presented in Berthou et al. [2014], is created by ignoring heat convection and radiation. Due to the strong simplification used in these models, they contain significant *epistemic* uncertainty, in addition to the *aleatoric* measurement uncertainty induced by acquiring data from a physical building. Due to the simplified nature of the model, the assumption $\mathcal{S} \in \mathcal{M}(\Theta)$ is clearly unjustified here. However, it may still be possible to obtain $\hat{\theta}$ such that $\mathcal{M}(\hat{\theta})$ is a good approximation of $\mathcal{S}$. Hence, it is interesting to analyse the parameter space Θ of this model to evaluate the identifiability and estimation uncertainty of $\hat{\theta}$.

The model circuit equivalent is shown in Fig. 17. The model has two outputs: the room temperature T_b and the wall surface temperature T_w, and two inputs: the consumed power by an electric heating element $\dot{Q}$ and the outside temperature T_∞. Five components form the model structure: the thermal res-
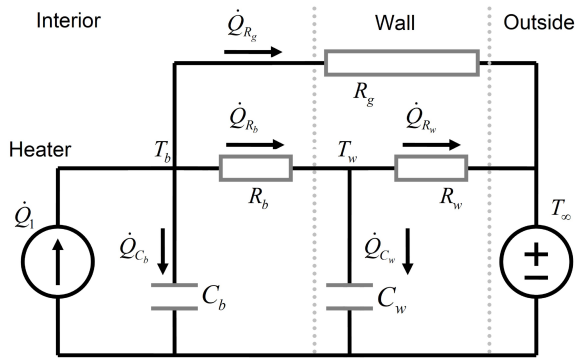
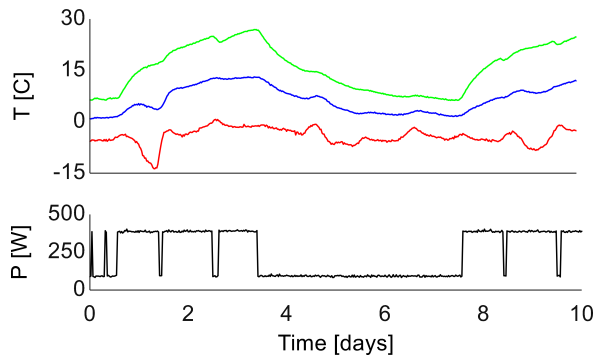Figure 17: Simplified thermal network model with three resistors and two capacitors.

Figure 18: Calibration data for the R3C2 model. Temperatures T_∞ (red), T_w (blue) and T_b (green), and also the power consumption Q, was recorded in February 2018.

istance between room air and wall R_b, the building envelope R_w, and the thermal resistance of windows and doors R_g. The two capacitances C_b and C_w represent the thermal capacitance of the building interior and envelope, respectively. Additionally, the process and measurement noise covariances W and V are also estimated as model parameters, since they are needed in the Kalman filter. Both covariance matrices are assumed diagonal, adding a total of four noise related parameters to the vector θ.

A calibration data-set for this model is shown in Fig. 18. The data was collected from the experimental building during February 2018, using a pre-installed data acquisition system and set of sensors Brastein et al. [2018].

3.2.1 Profile Likelihood of R3C2 model

Initially, both PL1 and PL2 methods were used to perform an exploratory analysis of the parameter space of the R3C2 model. The results of these analyses, presented in Figs. 19 and 20, show that there is a problem

with the parameter space of this model, particularly that the parameter R_b and R_w are inter-dependent. Observe from Fig. 20 that the R_b vs. R_w plot shows a linear relationship. Hence, the PL1 results for these two parameters in Fig. 19, which can be considered a projection of the PL2 result onto the individual parameter axis, show a large equipotential flat region which extends up to at least $5\frac{K}{W}$. Observe also that for R_b the PL1 profile makes a sharp bend at around $4.5\frac{K}{W}$, such that the profile is bounded for R_b. However, as discussed in Section 2.2.7, inter-dependent parameters can cause *artefacts* in the PL plots, due to the bounds on one parameter having a limiting effect on the other dependent parameters. The bend in the profile of R_b is an example of such an *artefact*.

Subsequently, the R3C2 model is found to be over-parameterised. After some experimentation, based on previous experience with the model Brastein et al. [2019b], the resistor R_g is removed from the circuit model in Fig. 17, in an attempt to make the remaining parameters identifiable. The resulting model, named R2C2, is used in the sequel and further analysed.

3.2.2 Profile Likelihood of R2C2

The first analysis performed on the reduced R2C2 model is a combination of the PL1 method and the URP method. The results, presented in Fig. 21, show that all four parameters are now identifiable, since the likelihood based confidence intervals are bounded with a clearly defined minima. Secondly, comparing URP to PL1 shows that although the URP method successfully captures the *general shape* of the objective function around $\hat\theta$ using $K = 500.000$ randomly drawn parameters, it is not enough to properly capture the optimal front. Hence, there is some small difference between the PL1 and URP results. By its use of numerical optimisation, the PL1 method successfully finds the optimal profile in likelihood space for each parameter. The main result from the application of PL1 is to obtain reasonable bounds θ_{min} and θ_{max} on Θ for further analysis, something for which the PL1 method is ideally suited.

Next, the PL2 method is used to further analyse the parameter space Θ, in particular to test for inter-dependency of parameters and further study the identifiability. For comparison, the Hessian method from Section 2.3.3 is used to compute the covariance of the estimated parameters Σ_θ, and subsequently compute an elliptic confidence region for the true parameters θ^*. The Hessian ellipses are superimposed on the PL2 heat-maps in Fig. 22. Two interesting observations can be made from these results. First, the results show that after removing R_g, all parameters are identifiable, i.e., the confidence regions are bounded, given the
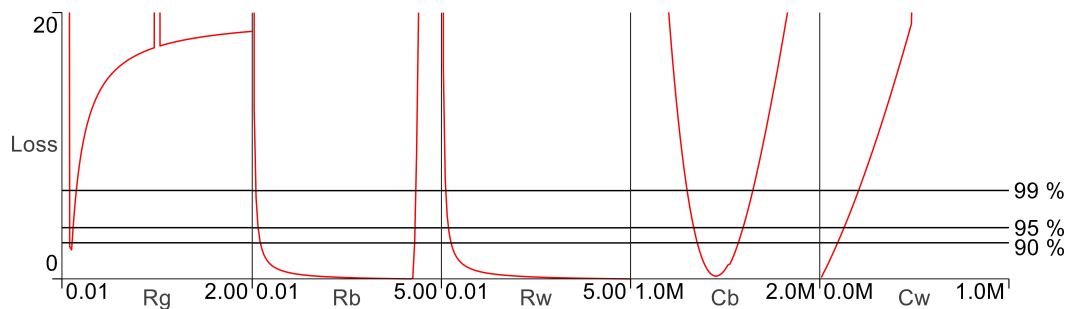
235

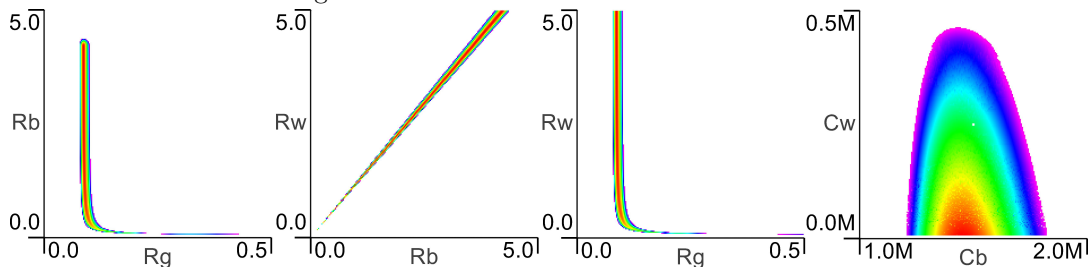Figure 19: PL1 results for the R3C2 model.

Figure 20: Selected PL2 results for the R3C2 model shows that the parameters R_w and R_b are inter-dependent.
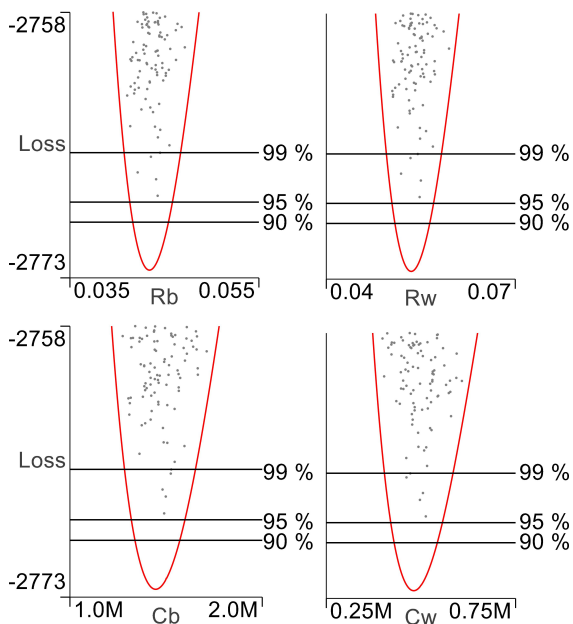

Figure 21: PL1 and URP results for the reduced R2C2 model show that even with $K = 500.000$ randomly drawn parameter vectors, the coverage is not good enough, since the optimal front from the PL method is not the same as that of the URP method. However, the shape of the objective is still approximated by the URP method, indicating its usefulness also for higher dimensional parameter spaces.
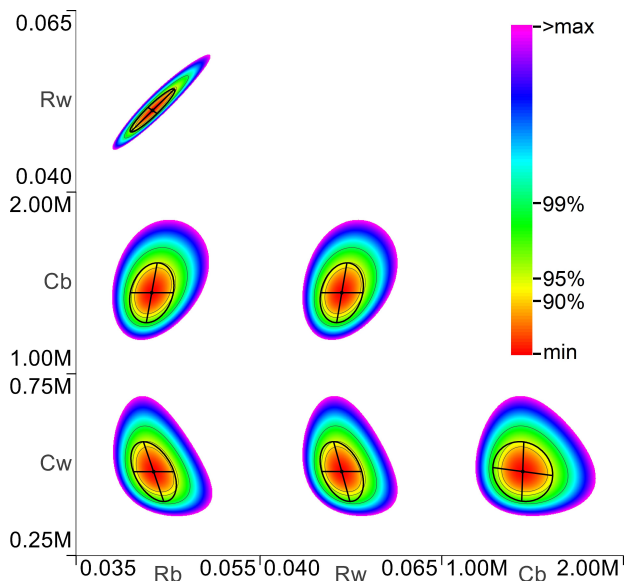
Figure 22: PL2 and Hessian ellipses (thick black) for the R2C2 model. Iso-lines trace the 90%, 95% and 99% confidence bounds computed from the PL2 results, based on the $\chi^2_{n_{df}}$-distribution with $n_{df} = 2$. The Hessian method is used to compute Σ_θ and superimpose an elliptic approximate confidence region at $\alpha = 95\%$.
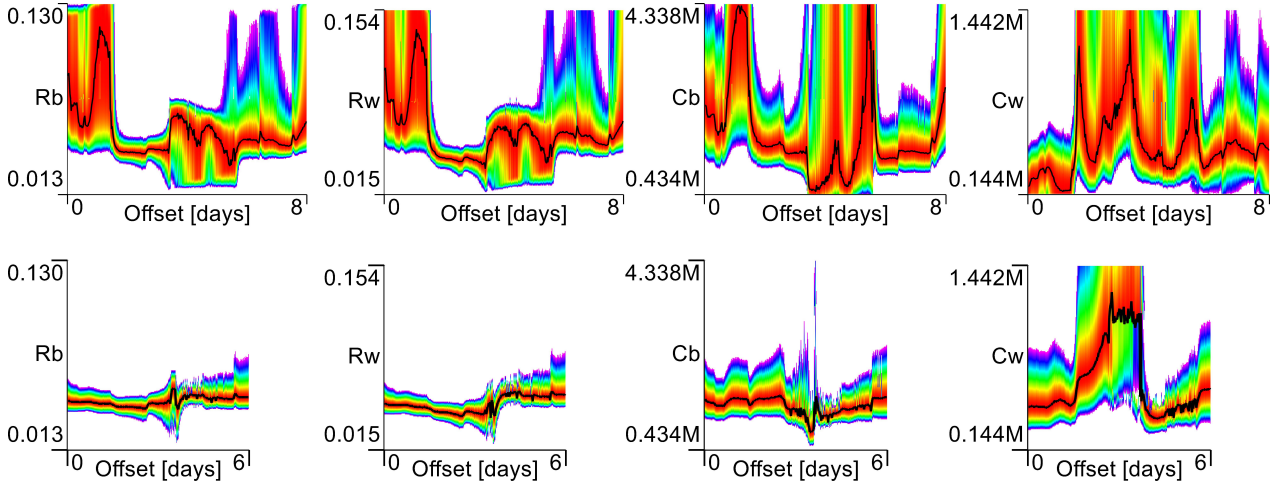
Figure 23: Consistency of dynamic information for the R2C2 model. The top four plots show the results of applying PL1 to a moving window of length 96 (2 days), while the lower four plots use a window length of 192 (4 days).

data in Fig. 18. The parameters R_b and R_w are still inter-dependent, as illustrated by the rotation of the approximately elliptic PL2 profile, but there is still a clearly defined optimum. Second, the Hessian method produces a reasonable approximation of the 95% iso-line confidence bounds in all the projected parameter planes. Where the PL2 method and Hessian method differ, it can be observed from Fig. 22 that the PL2 method, which by brute force computation captures the true projection of $\ell(\theta)$ onto $\Theta_{i,j}$, finds profiles that are not quite elliptic. The discrepancies observed visually therefore seem reasonable w.r.t. the shape of the PL2 profile. Observe for example that the C_b vs C_w profile is elongated in the increasing direction of both parameters, hence the discrepancy between PL2 and Hessian ellipse is mostly located towards the decreasing parameter directions.

Table 4: Optimal parameters with normalised standard deviations computed with the Hessian method for the R2C2 model.

	R_b	R_w	C_b	C_w
$\hat{\theta}_i$	0.0434	0.0512	1.446×10^6	0.481×10^6
$\frac{\sigma_i}{\hat{\theta}_i}$	0.0233	0.0210	0.0467	0.0702

The optimal parameters, which are the same for both PL2 and Hessian methods, are shown in Table 4 together with the standard deviations computed from inverting the Hessian, normalised over the optimal parameters.

3.2.3 Consistency of dynamic information

Since it is of interest to test bootstrapping methods also on the R2C2 model, a verification of the dynamic information content is first needed. A typical challenge for building thermal behaviour models is the restrictions on experimental design, since weather, including outside temperature, is a model input. Additionally, there are limitations to acceptable ranges of indoor temperature and limited available input power for heating, which further complicates the experimental design for this type of models. Therefore, model calibration must often be performed on *low informative* data. Hence, methods that can evaluate the quality of the dynamic information in the data is of interest. By using the PL1 method for a moving window of data, as discussed in Section 2.3.7, it is possible to obtain a visual diagnosis of estimation accuracy and parameter identifiability for segments of the data.

As shown in Fig. 23, the estimation accuracy of parameters in a window of length 96 samples (2 days) is somewhat poor for significant segments of data, in particular for the first part of the data-set. The parameter C_w is particularly difficult to identify, even for the 192 sample (4 days) window. From inspecting the calibration data in Fig. 18 this result is expected, due to the limited variation observed in temperature T_w. For the parameters R_b, R_w and C_b, the consistency test shows that the uncertainty is mostly consistent in time, with only minor variations, for the 4 day window case. The results also show that the optimal value for these three parameters do not vary significantly over time, for the window length of 4 days. However, the parameter C_w is estimated with significant time vari-
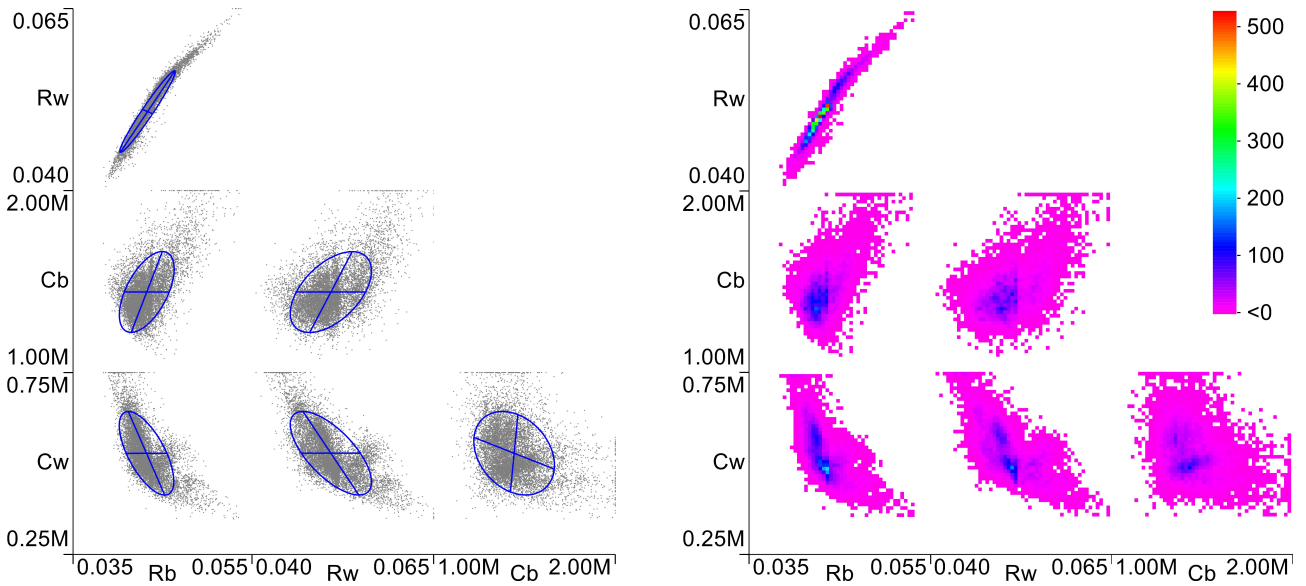
Figure 24: Bootstrap results for the R2C2 model, with $M = 10.000$ iterations represented as scatter plots with 95% confidence ellipses, simultaneous for the projected parameters, for the mean estimate (left) and corresponding 2D histograms (right). Both the scatter plots and the histograms are presented as two-dimensional projections onto each possible parameter combination plane $\Theta_{i,j}$.

ations also for the longest window length, as shown in the lower right panel of Fig. 23. This indicates poor identifiability of C_w, and may result in unsatisfactory results if block-based bootstrapping methods are used to estimate uncertainty.

3.2.4 Bootstrapping

The parameters of the R2C2 model is next analysed using Stationary bootstrapping with $p = 0.005$, which gives expected, i.e., average, block length $E(l) = 200$, since a window length of 192 samples (4 days) appears to be an acceptable choice based on the results in Fig. 23. The resulting mean parameters and normalised standard deviations, after running bootstrapping for $M = 10.000$ iterations, is shown in Table 5.

Table 5: Optimal parameters with normalised standard deviations computed with the Bootstrapping method for the R2C2 model.

	R_b	R_w	C_b	C_w
θ_i	0.0432	0.0509	1.443×10^6	0.528×10^6
$\frac{\sigma_i}{\theta}$	0.043	0.067	0.093	0.131

Comparing the results in Table 5 with Table 4, the estimated *mean* of the M bootstrapped iterations agrees well with the result obtained by optimisation and PL2 brute force exhaustive search. The normalised covariances obtained by bootstrapping, i.e., the covari-

ance of M iterations of repeated generation of pseudo data-sets with subsequent parameter estimation, are approximately two times larger than those obtained by inverting the Hessian of the likelihood function. Considering the significantly different theoretical foundation of these two methods of uncertainty estimation, a difference of a factor of two or three may be considered a reasonable agreement between the two methods, in particular since the consistency test in Fig. 23 showed that the calibration data contains some low informative regions which can cause outliers in the bootstrapped parameter estimates. A histogram over all M iterations is shown in Fig. 24. Since the parameter space is of a dimension higher than two, the histograms are plotted as projections onto parameter planes $\Theta_{i,j}$, similar to the projected profiles obtained from the PL2 method. Interestingly, the shape of the histograms is similar to the PL2 profiles obtained in Fig. 22. However, due to the effect of outliers caused by some of the randomised pseudo data-sets being significantly less informative than the full data-set, the spread of the histogram. i.e., the covariance of the mean estimate, is larger than the covariance obtained from the Hessian in Table 4. Observe also the *clustering* of parameter estimates at the *edges* of the histogram plots, which indicates that for certain iterations of the bootstrap methods, the obtained parameters are located at the constraints of the parameter space Θ. This is a further indication that some pseudo data-sets are non-informative w.r.t. parameter estimation, since

the resulting parameters at the bounds of Θ deviate significantly from those obtained when the full data-set is used.

3.2.5 Computation time

Computation times for the various analysis methods applied to the R2C2 model are shown in Table 6. First, observe that the computation times are considerably longer than those found for the simple first order model in Table 3, e.g., the URP method with $K = 500.000$ randomly drawn parameters was completed in $\sim 160s$ for the first order model but took $\sim 0.15h = 540s$ for the R2C2 model. Despite using a dataset with only approximately half the number of samples, 480 vs 1000, the computation time for the R2C2 model is approximately 3.4 times longer. This extended computation time is caused by increased model complexity. First, the model has two states rather than one. Additionally, the R2C2 model uses a UKF rather than a standard KF, which further increases computational time. When analysing the R2C2 model, the software evaluates the model's equations ~ 540.000 times per second for a total of ~ 1100 simulations per second. Comparably, the simpler first order model's equations are evaluated $\sim 3.100.000$ times per second, for a total of ~ 3100 simulations per second. Since the URP method does not use optimisation, model complexity, length of the data-set and the number of URP iterations K are the main factors that influence computation time, hence the results can be directly compared. Accounting for differences in data-set length, the increased model complexity of the R2C2 model, including its use of UKF with Runge-Kutta 4th order discretisation Runge [1895] of the state equation, increases computation time by approximately $\frac{540}{160}\frac{1000}{480} = 7$ times.

Next, observe that the stationary bootstrap, which shows similar results to the PL2 method, is about 40 times faster. This increased computation speed is obtained at the cost of inducing outlier estimates, caused by Bootstrapped pseudo data-sets that are less informative w.r.t. parameter estimation than the full data-set. Hence, due to these outliers, the uncertainty estimate is somewhat inflated compared to that obtained when computing the Hessian of the Likelihood function over the whole data-set.

Finally, observe from Table 6 that the Moving Window analysis computation time is only approximately linear in the window length l. The analysis using a longer window length of 192 is finished with a 1.62 times longer computation time, compared with the window length of 96. While this method is theoretically linear in window length l, the shorter window is less informative w.r.t. parameter estimation, as Fig. 23 shows. Hence, the task of the numerical optimiser

Table 6: Computation time for the LP model from the Step data-set. The other data-sets produce comparable execution times.

Method	Time
PL1 (resolution 500)	$\sim 4.35h$
URP ($K = 500.000$)	$\sim 0.15h$
Moving Wnd. (res. 200, $w = 1$, $l = 96$)	$\sim 3.60h$
Moving Wnd. (res. 200, $w = 1$, $l = 192$)	$\sim 5.77h$
PL2 (400×400 resolution)	$\sim 15.18h$
Stationary Bootstrap ($M = 10.000$)	$\sim 0.35h$

is more challenging, which increases the computation time slightly for the shorter window. This example illustrates that calculating computation time for complex analysis methods is not straight forward. The Moving Window with PL1 method consists of both a numerical optimisation method, a Kalman filter implementation, the model structure, and the Profile Likelihood algorithm, all of which influence the computation time.

3.3 Method recommendations

Each of the methods presented in this paper has its advantages and disadvantages. Since they each compute and represent the uncertainty of estimated parameters in different ways, they can be used for different applications.

First, with regards to representation of uncertainty as profiles or regions, this is a question of usage. As an uncertainty estimate for comparison, regions or intervals may be preferable, since they can be quantitatively compared. Profiles are more descriptive, since they can represent how the uncertainty is *distributed* across an entire parameter domain. Hence, for applications where the parameters themselves are of interest, i.e., assumed to be determined by the physical properties of the system, representing parameters as distributions is perhaps preferable since they capture the most information about the underlying physical system.

Second, with respect to choosing what methods to use, the first question to consider is whether it is reasonable to assume that the parameters are well approximated by a Gaussian distribution, such that a quadratic approximation can be used to obtain ellipsoid regions for describing the uncertainty. In such cases, and when confidence regions rather than profiles are desirable representations, the Hessian method for computation of estimation covariance is preferable, due to its computational simplicity and speed. The Hessian method is based on analysing the curvature of the likelihood function $\ell(\theta)$ around an optimal estimate $\hat{\theta}$, which must first be obtained by calibration of all para-

meters and hence is subject to local minima problems. Hence, the Hessian method may only estimate the uncertainty of a pre-determined, *presumed* optimal, $\hat{\theta}$. Therefore, it should be ascertained, if possible, whether a particular $\hat{\theta}$ is a global or local optimum.

The *Profile Likelihood* (PL1) method Maiwald and Timmer [2008], Meeker and Escobar [1995], Murphy and Van der Vaart [2000], Raue et al. [2009], Venzon and Moolgavkar [1988] is an attractive choice if the practical identifiability of parameters is questionable. This method, unlike the Hessian based method, can represent non-symmetric confidence regions which can be used to diagnose identifiability Raue et al. [2009]. Further, the method allows representation as profiles, which may also be an advantage in some cases. The PL1 method can also be used for obtaining reasonable limitations on parameters in an exploratory analysis. Although it is known to give projections onto single parameters, which can be too wide if there are interdependent parameters, it is still a useful analysis tool.

The *Uninformed Random Picking* (URP) method Hoos and Stützle [2004] is a simple alternative to PL1, and provides approximately the same results if the number of randomly drawn parameters K is large enough. However, being a stochastic method, the distribution of randomly drawn parameters across parameter space can not be guaranteed. Hence, the optimal front in parameter space may not be detected unless a sufficiently large number of parameters is used. This is challenging for high dimension parameter spaces. The main advantage of URP is its simplicity, and that it does not require an optimisation algorithm.

The *two-dimensional Profile Likelihood* (PL2) method provides the most information about the parameter domain. In particular, it is the only method presented in this paper which can diagnose parameter inter-dependency and identifiability, as well as handle multimodal objective functions with local minima. Bootstrapping methods may show large dispersion in estimated parameters if parameters are nonidentifiable, but the exhaustive exploration of the entire parameter space Θ offered by the PL2 method still provides more detailed and clear diagnostic conclusions. Since the method obtains highly descriptive profiles of combinations of parameters, this method provides the most detailed information about the parameter space Θ. Hence, if methods like PL1 or URP indicate problems with identifiability, it may be useful to apply the PL2 method to obtain a better analysis of the parameter space. Finally, the PL2 method is guaranteed to find the global optimum in Θ, within the accuracy allowed by the discretisation for the brute force search.

Repeatedly optimising the parameters with random initial guesses can be used to test the parameter optimisation procedure for sensitivity to the initial conditions. Additionally, this method is a useful tool for identification of local minima in the objective function. If there are multiple locally optimal solutions, this method will likely find them faster than the PL2 method, provided that the distribution of randomised initial conditions is dense enough, i.e., it needs a large enough number of repeated randomised initial conditions with subsequent optimisation of parameters such that at least one of the randomly drawn initial guesses will be close enough to the local optima to find them.

Bootstrapping Politis [2003] is perhaps the most intuitive way to obtain confidence regions, since it resembles the basic idea of computing coverage probabilities for multiple experiments Neyman [1937]. However, as the results have shown, if the dynamic information content in the data varies in time, block based bootstrapping can create pseudo data-sets that are *uninformative* w.r.t. parameter estimation and hence provide poor parameter identifiability. Subsequently, there can be *outlier* parameter estimates among the M iterations which affect the computation of mean parameters and the covariance. When there are variations in dynamic information content in the calibration data, special care should be taken when selecting the block lengths for bootstrapping. Regardless, bootstrapping is much faster than the PL2 method, and is therefore a useful alternative or augmentation to the PL2 method, in particular where computational resources and/or time is a challenge. Arguably, bootstrapping may also provide a more *realistic* estimation of the uncertainty of the parameters, provided the consistency of dynamic information in the calibration data is acceptable, since the method approximates running repeated experiments in a way that is similar to the idea of *coverage probability* calculation for confidence intervals. Due to its simplicity of implementation, bootstrapping methods may be preferable as an initial estimate of the uncertainty of estimated parameters.

Finally, a moving window combined with the PL1, or the Hessian method, can be used to test for consistency in dynamic information w.r.t. a particular model. Since this method, especially based on the PL1 method, is somewhat time consuming, it is most useful as a diagnostic tool to test for sources of diverging results in other methods, such as block based bootstrapping.

4 Conclusion

In this paper, a number of different methods for parameter estimation and analysis has been presented. Two test cases, a simple first order model with simulated

data, and a thermal network building grey-box model with measurement data from a physical building, was used to demonstrate the application of these methods.

The main results from these two test cases are, firstly, demonstrating the usefulness of one- and two-dimensional *Profile Likelihood* Raue et al. [2009]. These methods obtain *descriptive profiles* for each parameter, which can both estimate the *uncertainty* of the parameter estimate, diagnose the *identifiability* of the parameters and test for presence of local minima. The two-dimensional Profile Likelihood was shown to be particularly useful for detecting *over-parametrisation* for the second test case. Further, the one dimensional profile likelihood method was used with a moving window to check the *consistency* of dynamic information, and subsequently the identifiability and estimation uncertainty of the parameters as a function of *time*, with respect to a specific model structure. The latter was shown to be useful in combination with *block based bootstrapping*, to test for segments of data that are *uninformative* w.r.t. parameter estimation.

For the first test case, six different simulated data-sets were used. Of these six sets, the simple input step and the Pseudo Random Binary Sequence with $0.5s$ bit length gave the lowest overall estimation uncertainty. However, since the step data-set contains significant segments of data in which the system is in steady state, and hence produce non-identifiable parameters, the use of block based bootstrapping method results introduce outliers in the parameter estimates which significantly inflate the covariance of the mean parameter estimate. Hence, the interesting conclusion for this test case is that the data-set which produces the *lowest* estimation uncertainty for the Profile Likelihood and Hessian based method gives the *highest* uncertainty for the block based bootstrap method. Hence, what methods to use is also affected by the *dynamic information* content in the calibration data, and consequently the experimental design used to obtain that data, in addition to the application requirements and desired representation of resulting parameters.

References

Akaike, H. Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike*, pages 199–213. Springer, 1998. doi:10.1007/978-1-4612-1694-0_15.

Bacher, P. and Madsen, H. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 2011. 43(7):1511 – 1522. doi:10.1016/j.enbuild.2011.02.005.

Bentley, J. P. *Principles of measurement systems*. Pearson education, 2005.

Berthou, T., Stabat, P., Salvazet, R., and Marchio, D. Development and validation of a gray box model to predict thermal behavior of occupied office buildings. *Energy and Buildings*, 2014. 74:91–100. doi:10.1016/j.enbuild.2014.01.038.

Bohlin, T. and Graebe, S. F. Issues in nonlinear stochastic grey box identification. *International journal of adaptive control and signal processing*, 1995. 9(6):465–490. doi:10.1002/acs.4480090603.

Brastein, O., Perera, D., Pfeiffer, C., and Skeie, N.-O. Parameter estimation for grey-box models of building thermal behaviour. *Energy and Buildings*, 2018. 169:58 – 68. doi:10.1016/j.enbuild.2018.03.057.

Brastein, O. M., Lie, B., Sharma, R., and Skeie, N.-O. Parameter estimation for externally simulated thermal network models. *Energy and Buildings*, 2019a. 191:200–210. doi:10.1016/j.enbuild.2019.03.018.

Brastein, O. M., Sharma, R., and Skeie, N.-O. Sensor placement and parameter identifiability in grey-box models of building thermal behavior. In *Proceedings of The 60th Conference on Simulation and Modelling (SIMS 60), 13-16 August 2019, Västerås, Sweeden*. Linköping University Electronic Press, 2019b.

Deconinck, A.-H. and Roels, S. Is stochastic grey-box modelling suited for physical properties estimation of building components from on-site measurements? *Journal of Building Physics*, 2017. 40(5):444–471. doi:10.1177/1744259116688384.

Efron, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 1979. 7(1):1–26. doi:10.1007/978-1-4612-4380-9_41.

Ergon, R. and Di Ruscio, D. Dynamic system calibration by system identification methods. In *European Control Conference (ECC), 1997*. IEEE, pages 1556–1561, 1997. doi:10.23919/ECC.1997.7082324.

Farrell, J. A. and Polycarpou, M. M. *Adaptive approximation based control: unifying neural, fuzzy and traditional adaptive approximation approaches*, volume 48. John Wiley & Sons, 2006.

Ferrero, C. S., Chai, Q., Dueñas Díez, M., Amrani, S. H., and Lie, B. Systematic analysis of parameter identifiability for improved fitting of a biological wastewater model to experimental data. *Modeling, Identification and Control*, 2006. 27(4):219. doi:10.4173/mic.2006.4.2.

Fux, S. F., Ashouri, A., Benz, M. J., and Guzzella, L. EKF based self-adaptive thermal model for a passive house. *Energy and Buildings*, 2014. 68:811–817. doi:10.1016/j.enbuild.2012.06.016.

Hoos, H. H. and Stützle, T. *Stochastic local search: Foundations and applications.* Elsevier, 2004.

Jazwinski, A. H. *Stochastic processes and filtering theory.* Dover Publications, Inc, 1970.

Johansson, R. *System Modeling and Identification.* Information and system sciences series. Prentice Hall, 1993.

Johnson, R. and Wichern, D. *Applied Multivariate Statistical Analysis.* Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007.

Juhl, R., Møller, J. K., Jørgensen, J. B., and Madsen, H. Modeling and prediction using stochastic differential equations. In *Prediction Methods for Blood Glucose Concentration*, pages 183–209. Springer, 2016a. doi:10.1007/978-3-319-25913-0_10.

Juhl, R., Møller, J. K., and Madsen, H. ctsmr-Continuous Time Stochastic Modeling in R. *arXiv preprint arXiv:1606.00242*, 2016b.

Killian, M. and Kozek, M. Ten questions concerning model predictive control for energy efficient buildings. *Building and Environment*, 2016. 105:403–412. doi:10.1016/j.buildenv.2016.05.034.

Kristensen, N. R. and Madsen, H. Continuous time stochastic modelling. *Mathematics Guide*, 2003. pages 1–32.

Kristensen, N. R., Madsen, H., and Jørgensen, S. B. Parameter estimation in stochastic grey-box models. *Automatica*, 2004. 40(2):225–237. doi:10.1016/j.automatica.2003.10.001.

Kullback, S. A Note on Neyman's Theory of Statistical Estimation. *The Annals of Mathematical Statistics*, 1939. 10(4):388–390. URL https://www.jstor.org/stable/2235617.

Kunsch, H. R. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 1989. pages 1217–1241. URL https://www.jstor.org/stable/2241719.

Lie, B. Model uncertainty and control consequences: a paper machine study. *Mathematical and Computer Modelling of Dynamical Systems*, 2009. 15(5):463–477. doi:10.1080/13873950903375452.

Ljung, L. *System Identification: Theory for the User.* Prentice Hall information and system sciences series. Prentice Hall PTR, 1999.

Lodhi, H. and Gilbert, D. Bootstrapping parameter estimation in dynamic systems. In *International Conference on Discovery Science*. Springer, pages 194–208, 2011. doi:10.1007/978-3-642-24477-3_17.

Madsen, H. *Time series analysis.* Chapman and Hall/CRC, 2007.

Madsen, H. and Holst, J. Estimation of continuous-time models for the heat dynamics of a building. *Energy and buildings*, 1995. 22(1):67–79. doi:10.1016/0378-7788(94)00904-X.

Maiwald, T. and Timmer, J. Dynamical modeling and multi-experiment fitting with PottersWheel. *Bioinformatics*, 2008. 24(18):2037–2043. doi:10.1093/bioinformatics/btn350.

Meeker, W. Q. and Escobar, L. A. Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician*, 1995. 49(1):48–53. doi:10.1080/00031305.1995.10476112.

Murphy, S. A. and Van der Vaart, A. W. On profile likelihood. *Journal of the American Statistical Association*, 2000. 95(450):449–465. doi:10.1080/01621459.2000.10474219.

Neyman, J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1937. 236(767):333–380. doi:10.1098/rsta.1937.0005.

Nocedal, J. and Wright, S. *Numerical optimization.* Springer Science & Business Media, 2006.

Pohjanpalo, H. System identifiability based on the power series expansion of the solution. *Mathematical Biosciences*, 1978. 41(1):21–33. doi:10.1016/0025-5564(78)90063-9.

Politis, D. N. The impact of bootstrap methods on time series analysis. *Statistical Science*, 2003. pages 219–230. URL https://www.jstor.org/stable/3182852.

Politis, D. N. and Romano, J. P. The stationary bootstrap. *Journal of the American Statistical association*, 1994. 89(428):1303–1313. doi:10.1080/01621459.1994.10476870.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical recipes in C++*, volume 3. Cambridge University Press, 2007.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 2009. 25(15):1923–1929. doi:10.1093/bioinformatics/btp358.

Rosen, R., Wichert, G. v., Lo, G., and Bettenhausen, K. D. About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine*, 2015. 48(3):567 – 572. doi:10.1016/j.ifacol.2015.06.141. 15th IFAC Symposium onInformation Control Problems inManufacturing.

Rossi, R. J. *Mathematical Statistics: An Introduction to Likelihood Based Inference.* John Wiley & Sons, 2018.

Runge, C. Ueber die numerische Auflösung von Differentialgleichungen. *Mathematische Annalen*, 1895. 46(2):167–178. doi:10.1007/BF01446807.

Simon, D. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches.* John Wiley & Sons, 2006.

Venzon, D. and Moolgavkar, S. A method for computing profile-likelihood-based confidence intervals. *Applied statistics*, 1988. pages 87–94. doi:10.2307/2347496.

Wang, L. *Model predictive control system design and implementation using MATLAB®.* Springer Science & Business Media, 2009.

Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 1938. 9(1):60–62. URL https://www.jstor.org/stable/2957648.

Article E

Analysing uncertainty in parameter estimation and prediction for grey-box building thermal behaviour models

Authors O. M. Brastein, A. Ghaderi, C. F. Pfeiffer and N.-O. Skeie

Analysing uncertainty in parameter estimation and prediction for grey-box building thermal behaviour models☆

O. M. Brastein[a,*], A. Ghaderi[b], C. F. Pfeiffer[a], N.-O. Skeie[a]

[a]*Department of Electrical Engineering, Information Technology and Cybernetics, University of South-Eastern Norway, N-3918 Porsgrunn*
[b]*Department of Mathematics and Science Education, University of South-Eastern Norway, N-3918 Porsgrunn*

Abstract

The potential reduction in energy consumption for space heating in buildings realised by the use of predictive control systems directly depends on the prediction accuracy of the building thermal behaviour model. Hence, model *calibration* methods that allow improved prediction accuracy for *specific* buildings have received significant scientific interest. An extension of this work is the potential use of calibrated models to estimate the *thermal properties* of an existing building, using measurements collected from the actual building, rather than relying on building specifications.

Simplified thermal network models, often expressed as *grey-box* Resistor-Capacitor circuit analogue models, have been successfully applied in the prediction setting. However, the use of such models as *soft sensors* for the thermal properties of a building requires an assumption of *physical interpretation* of the estimated parameters. The parameters of these models are estimated under the effects of both *epistemic* and *aleatoric* uncertainty, in the model structure and the calibration data. This uncertainty is propagated to the estimated parameters. Depending on the model structure and the dynamic information content in the data, the parameters may not be *identifiable*, thus resulting in *ambiguous* point estimates.

In this paper, the Profile Likelihood method, typical of a *frequentist* interpretation of parameter estimation, is used to diagnose parameter *identifiability* by projecting the likelihood function onto each parameter. If a Bayesian framework is used, treating the parameters as random variables with a probability distribution in the parameter space, *projections* of the posterior distribution can be studied by using the Profile Posterior method. The latter results in projections that are *similar* to the *marginal distributions* obtained by the popular Markov Chain Monte Carlo method. The different approaches are applied and compared for five experimental cases based on *observed* data. Ambiguity of the estimated parameters is resolved by the application of a *prior distribution* derived from a priori knowledge, or by appropriate modification of the model structure. The posterior predictive distribution of the *model output predictions* is shown to be mostly *unaffected* by the parameter non-identifiability.

Keywords: thermal network models; grey-box models; Profile Likelihood; Bayesian parameter estimation; Markov Chain Monte Carlo; parameter distribution; posterior predictive distribution; parameter identifiability

1. Introduction

1.1. Background

The reduction of anthropogenic CO_2 emissions is perhaps the most important task in modern science. The energy consumed by space heating in buildings is considerable [1]. According to the Energy Performance of Buildings Directive (EPBD) [2], the energy consumed by buildings accounts for 40% of the total energy consumption within the European Union (EU). Hence, the development of model predictive control strategies that can effectuate energy reductions by improved thermal control has received significant scientific interest [3, 4]. For control systems, development of accurate *prediction* models is essential.

Another application of interest for building thermal modelling is the *classification* of building properties related to space heating, for improved evaluation of the energy *performance* of existing buildings [5]. By classifying actual energy performance, development of taxation schemes could be utilised to motivate investments in energy reduction technology. Given that there is often discrepancies between physical buildings and their blueprints, typically due to continuous modifications or workmanship issues, energy classification schemes could with benefit be based on energy and temperature data recorded from the building to be classified. A popular method for modelling building thermal behaviour is the use of simplified thermal network models expressed as a Resistor-Capacitor analogue [3, 4, 6–9]. Regardless of their proven efficiency in the prediction setting, the parameters of a thermal network model may not be suitable as *soft sensors* for monitoring building properties, since this assumes a physical interpretation of the parameters as constants of the physical building [5, 10]. For such an assumption to be justified, the verification of *parameter identifiability* is essential, in order to ensure unambiguous parameter estimation.

1.2. Previous work

Thermal behaviour models of buildings

For widespread use of model predictive control and/or classification systems in buildings, a simple modelling method that can both produce *physically interpretable* parameters *and* make accurate *predictions* of future thermal behaviour is needed. Models of building thermal behaviour based on exact physical specifications of a building often become intractable due to the complexity of building structures, and may require specialised software to simulate [1]. Additionally, existing buildings may deviate from blueprints and specifications of the building, which further exacerbates the challenge of developing a physics-based *white-box* model [5]. In contrast, data-driven models typically use simple model structures, with parameters that are calibrated from data acquired from existing buildings. Such data-driven *black-box* models, e.g., from system identification methods, typically have improved prediction performance due to being calibrated for specific buildings, but in general lack physical interpretability [11–15].

A reasonable compromise between the physics-based white-box and the data-driven black-box models is the use of *grey-box* thermal network models [3–5, 7, 9]. Thermal network (TN) models, typically expressed as Resistor-Capacitor (RC) electric analogue models, are based on a *naive* physical, *cognitive* understanding of the building thermodynamics, with relatively few *lumped* parameters that are calibrated from observational data. These models contain significant *epistemic* uncertainty in their formulation, resulting from model approximations and unmodelled or unrecognised disturbances [8], in addition to the *aleatoric* uncertainty introduced by random measurement noise. Hence, they can with advantage be formulated using stochastic differential equations (SDE) [8, 16–19]. Since the structure of a grey-box TN model is developed based on a physical description of the building, the parameters are often assumed to have a physical interpretation. However, due to the inherent uncertainties involved in the formulation of such models, a through analysis of parameter identifiability, which may lead to ambiguous parameter estimates, is needed prior to such interpretation.

Another point of interest regarding interpretation of TN models is the model *states*. The temperature *state nodes* in the RC circuit are typically chosen to representation a specific part of the building, e.g., the room interior or the building envelope internal surface, hence a physical interpretation of the states are assumed from the model structure. However, since the parameters that determine the *relationship* between these nodes are calibrated from measured data, the model is trained to predict the temperature at the specific sensor locations [20]. If the states are directly measurable, each state corresponds to a specific sensor location. Hence, the physical interpretation of TN model states is determined by both the model structure and the sensor location. Compared to the black-box system identification (SSID) paradigm, where the model structure to be calibrated is some *general* state mapping, the grey-box TN structure *constrains* the state representation. in comparison for an SSID model, which also effectually learns to predict the system response at the sensor locations, a change of *basis* for the state space will result in equivalent descriptions of the system, with the same outputs given the same measured data, but with *different* state representations.

Parameter estimation

Estimation of parameters requires a well-defined objective function. Using a *statistically founded* objective function, such as the *likelihood* function or the *posterior distribution,* computed from Bayes' theorem by the inclusion of a *prior* distribution, of the parameters, allows the use of statistical tools for model validation and analysis [10, 16, 21]. The *evaluation* of the likelihood function and/or the posterior parameter distribution for SDE models has previously been presented in detail in the Continuous Time Stochastic Modelling (CTSM) framework [16, 22]. By utilising a Kalman Filter (KF) to compute the one-step ahead prediction *residuals,* which are subsequently assumed normally distributed, the likelihood can be efficiently evaluated for an SDE model [16]. The grey-box SDE approach has been claimed as a natural framework for modelling dynamic systems in general [23].

Parameter identifiability analysis and prediction accuracy

A common assumption for parametrised model structures is that there exist an *unambiguous* set of parameters, which is optimal in the sense that it produces the best *model fit* in some specified statistical sense. However, there are cases for which the objective function used for the estimation of parameters is in some way *non-informative* for a subset of the parameters, thus resulting in *ambiguous* solutions. This subset of parameters is denominated as *non-identifiabl*e. If the *non-identifiable* parameters are perturbed in some way, the objective function is either *unchanged,* or the change is *insufficient* to determine the bounds of the estimated parameter with a desired prescribed level of *confidence* [10]. A good diagnostic tool is found in the framework of the *Profile Likelihood* (PL) method [5, 10, 21, 24].

Since the objective function compares model predictions with measured data, non-identifiability may be caused by either the *model structure* or by a lack of *dynamic information* in the data. The former is the cause of *structural* non-identifiability, which presents as a flat equipotential *manifold,* bounded or unbounded depending on the model structure*,* in the parameter space [10]. Structural identifiability is well covered in the literature, and there exist several diagnostic methods based on a multitude of theoretical foundations [10, 13, 21, 25, 26].

If non-identifiability results from a lack of *dynamic information* in the calibration data, the affected parameters are diagnosed as *practically* non-identifiable. For a parameter to be identifiable according to the PL method [10], the *likelihood-based confidence interval* (CI), and subsequently also the likelihood profile, must be bounded in *both* directions. Hence a *practically* non-identifiable parameter may be diagnosed by inspecting the likelihood profile for the presence of a well-defined optimum that is *insufficiently pronounced* to produce a bounded CI [10].

The PL method, based on the likelihood function and computation of CIs, has a distinctly *frequentist* approach to parameter estimation. If a *Bayesian* framework is used, where parameters are treated as *random variables* that have a distribution in parameter space, the *Markov Chain Monte Carlo* (MCMC) method [21, 27–31] can be used to *infer* the posterior distributions from the measurement data, typically visualised by obtaining marginal posterior distributions for single parameters or pairs of parameters [27, 30, 31]. The Bayesian framework combines the likelihood function with a *prior* by use of Bayes' theorem, thus computing the posterior distribution of the parameters [27]. The use of the Bayesian framework and MCMC for calibration of TN models was also reported in [32]. Alternatively, a variation of the PL method, called the *Profile Posterior* (PP) method [21], may be used to visualise the posterior distribution by obtaining projections, rather than marginal distributions. Similar arguments w.r.t. the identifiability of parameters drawn from the PL method can be applied to the posterior distribution [21].

There are also several other methods that can be used to investigate parameter identifiability, some of which are reviewed in [33]. Some possibilities are the use of the Hessian matrix evaluated at the optimal estimate to compute confidence bounds, and the testing for convergence problems in the optimisation algorithm by repeated optimisations with randomized initial guess [34]. For simple linear models, structural identifiability can sometimes be evaluated analytically [35]. Another possibility is the use of graphing tools to analyse the interactions between parameters and model output [36].
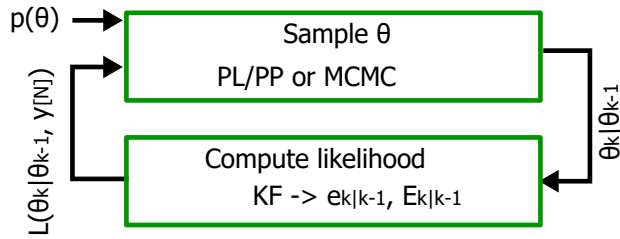
Figure 1: Both the PL/PP and the MCMC methods explore parameter space on the same likelihood/posterior hyper-surface.

Table 1: Method overview.

Name	Description
PL1D	$L\left(\theta;y\left[N\right]\right)$ *projected* to parameter θ_i
PL2D	$L\left(\theta;y\left[N\right]\right)$ *projected* to plane $\Theta_{i,j}$
PP1D	$p\left(\theta|y\left[N\right]\right)$ *projected* to parameter θ_i
PP2D	$p\left(\theta|y\left[N\right]\right)$ *projected* to plane $\Theta_{i,j}$
MCMC	$p\left(\theta|y\left[N\right]\right)$ *marginalised* to θ_i or $\Theta_{i,j}$

Since the parameter non-identifiability results from the objective function being *non-informative* for a sub-set of the parameters, adding *more information* to the estimation problem is a reasonable strategy towards resolving the non-identifiability. Experimental redesign may be used in order to collect more *informative* data, either by *improved* dynamic information content in existing measurements or by adding *new* measurements from the system $\mathcal{S}$ [10, 21]. The literature on system identification covers a range of experimental design considerations, including optimal experimental design for certain types of systems, see e.g. [11, 37]. A popular approach is the use of a Pseudo Random Binary Sequence (PRBS) applied to the actuator which may result in improved system excitation, thus improving practical identifiability of model parameters [6]. However, for occupied buildings, the choice of excitation for the active heating system may be limited due to occupant demands. If obtaining more data is not possible, redesigning the model structure $\mathcal{M}$, such that the model better represents the actual experimental data collected, may also resolve the non-identifiability [10, 21].

Finally, an important observation is that model structures with non-identifiable parameters can also provide reasonable predictions of the system outputs, but the non-identifiable parameters are arguably without a physical interpretation and can be considered *nuisance* parameters [5]. Indeed, ambiguous parameters without physical interpretation is the *norm* in traditional black-box calibration methods, such as system identification (SID) [11–15].

1.3. Overview of paper

In this work, the two projection-based methods, PL and PP, are compared to the MCMC method, on the basis of five experimental cases with differences in model structure, use of priors, identifiability of parameters and choice of training data. The theoretical foundation for the methods is presented in Section 2. The model, data and experimental setup of each case is presented in Section 3. The results are presented and discussed in Section 4, and the work concluded in Section 5.

2. Methods

2.1. Overview

In the sequel, the *Profile Likelihood* (PL) and the *Profile Posterior* (PP) methods [10, 21] are discussed and compared with the *Markov Chain Monte Carlo* (MCMC) method [27, 30, 31]. These methods are ideal for the study of parameter identifiability and allows detection of ambiguous parameter estimates. Despite fundamental differences in theoretical basis, i.e., the PL/PP methods are based on a *frequentist* interpretation of parameter estimation while the MCMC is typical of the

Bayesian statistics framework, the methods share certain similarities. As shown in Fig. 1, all these methods seek to obtain estimates of the likelihood function $L\left(\theta; y_{[N]}\right)$, or by inclusion of a *prior* $p\left(\theta\right)$, the posterior distribution $p\left(\theta|y_{[N]}\right)$. Each method *explores* the parameter space by taking samples θ_k and evaluating them on the same *likelihood/posterior hyper-surface*. However, there are some important differences; the use of *deterministic* vs. *stochastic* exploration of the parameter space, and the use of *projection* in the PL/PP methods vs. *marginalisation* in MCMC to obtain *partial* projections/distributions of selected parameters. An overview of relevant variations of the methods is given in Table 1, together with a short-hand name for each method for future reference. The PL1D/PL2D and PP1D/PP2D are collectively referred to as the PL and PP methods, respectively.

2.2. Parameter estimation and analysis

For simplified models, e.g., thermal network models, the uncertainty in the state transition can be large. Hence, it is convenient to express such models as a grey-box model using a continuous time *stochastic differential equation* (SDE) for the state transition Eq. (1); adopting the notation of [16]:

$$dx_t = f\left(x_t, u_t, t, \theta\right) dt + \sigma\left(u_t, t, \theta\right) d\omega_t \tag{1}$$

$$y_k = h\left(x_k, u_k, t_k, \theta\right) + v_k \tag{2}$$

where $t \in \mathbb{R}$ is the time variable and $x_t \in \mathbb{R}^{n_x}$ is the continuous time state vector. The first and second terms in Eq. (1) are commonly referred to as the *drift* and *diffusion* term, respectively [16, 38]. The *drift* term expresses the deterministic transition of the conditional mean state, while the *diffusion* term expresses the increments of the uncertainty linked to the conditional state covariance. The diffusion term, i.e. the *process noise*, is expressed as the function σ multiplied with the differential of a standard Wiener process ω_t [16, 38]. The *measurement equation*, given in Eq. (2) is formulated in discrete time where $v_k \sim \mathcal{N}\left(0, V\right)$ is the measurement noise. The continuous time input $u_t \in \mathbb{R}^{n_u}$ and output $y_t \in \mathbb{R}^{n_y}$ have the corresponding ordered sequences of discrete time measurements u_k and y_k taken from the system $\mathcal{S}$:

$$y_{[N]} = [y_0, y_1, \ldots, y_N] \tag{3}$$

$$u_{[N]} = [u_0, u_1, \ldots, u_N]$$

Here, the integer subscripts $k = 0, 1, \ldots, N$ denote the discrete time sampling instants, and the subscript enclosed in $[\cdot]$ is used to indicate an ordered sequence. The estimation of unknown parameters can be formulated as an optimisation problem, defining the *objective function* $g\left(\theta\right)$:

$$\hat{\theta} = \arg \underset{\theta}{\text{opt}}\, g\left(\theta; \mathcal{M}, \mathcal{K}, \mathcal{A}\right) \tag{4}$$

$$s.t.\ \theta \in \Theta$$

Here, $\mathcal{M}$ is a predetermined model structure parametrised by $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^{n_\theta}$ is a set of feasible values for the model parameters that form *inequality constraints* for the optimisation problem. Parameters in θ are sampled from the parameter space Θ by an algorithm $\mathcal{A}$. The experimental conditions $\mathcal{K}$ include the input and output measurements $u_{[N]}$ and $y_{[N]}$ as defined in Eq. (3). In the sequel, the dependency on $\mathcal{M}$, $\mathcal{K}$ and $\mathcal{A}$ is omitted for simplicity of notation.

A statistically well-founded choice of objective $g\left(\theta\right)$ is the *likelihood* function

$$L\left(\theta; y_{[N]}\right) = p\left(y_{[N]}|\theta\right) \tag{5}$$

which describes the *joint probability* of observing the measurement sequence $y_{[N]}$ given $\mathcal{M}\left(\theta\right)$. An elegant method for evaluating the likelihood function $L\left(\theta; y_{[N]}\right)$ for grey-box SDE models on the form of Eqs. (1) and (2) is presented in the framework named Continuous Time Stochastic Modelling (CTSM) [16]. The CTSM approach is summarised in Section 2.3. An alternative choice for $g\left(\theta\right)$ is the *posterior distribution* of the parameters $p\left(\theta|y_{[N]}\right)$, which combines the likelihood, by Bayes' theorem, with a *prior* distribution $p\left(\theta\right)$ on the parameters, and with the *evidence* $p\left(y_{[N]}\right)$, a scaling factor that is

137 independent of θ;

$$p\left(\theta|y_{[N]}\right) = \frac{p\left(y_{[N]}|\theta\right)p\left(\theta\right)}{p\left(y_{[N]}\right)} \tag{6}$$

138 Both the likelihood $L\left(\theta; y\left[N\right]\right)$ and the posterior $p\left(\theta|y\left[N\right]\right)$ are *statistical quantities* that relates different values of θ with the
139 *data* $y_{[N]}$, hence representing *density functions* [39] over the parameter space Θ. Observe that, unlike the posterior distribu-
140 tion, the likelihood is *not* a probability distribution over the parameters but takes its *random variable* as the measurements
141 $y_{[N]}$ in the *sample* space, *given* a known parameter θ.

142   It is interesting to note that the maximisation of the likelihood function is typically associated with a *frequentist* statistics
143 framework, whereas the use of a posterior *distribution* is typical of a *Bayesian* approach. In the frequentist framework, as for
144 the likelihood function, the model parameters are considered constants, while the *data* is the random variable. Hence, the
145 frequentist goal is to estimate some statistic of the *true* parameter θ^*, such as a *confidence interval (CI)* [40, 41]. Observe
146 that the confidence level of a CI is *not* a probability statement, as unequivocally stated in [40], since neither the CI nor the
147 true parameter θ^* are considered to be random variables.

148   In contrast, the Bayesian approach to statistics treats the *parameters* as random variables that are subject to probabilistic
149 treatment, i.e., described by a probability distribution rather than as constants. Typically, the posterior distribution cannot
150 be obtained analytically, and some variation of the *Markov Chain Monte Carlo (MCMC)* method is used instead to *estimate*
151 the posterior distribution of the parameters given the data.

152   Both the likelihood function and the posterior distribution can be directly optimised to obtain a parameter *point estimate*,
153 respectively denominated the *Maximum Likelihood Estimate (MLE)* $\hat{\theta}_{\mathrm{MLE}}$ and the *Maximum Aposteriori Estimate (MAP)*
154 $\hat{\theta}_{\mathrm{MAP}}$. However, for the purpose of *analysing* the results of the parameter estimation, it is useful to visualise the objective
155 function over the feasible region Θ; either the whole of Θ or some sub-region of particular interest. Since Θ is typically high
156 dimensional, it is necessary to create plots for single parameters, or combinations of two parameters. Since the posterior
157 $p\left(\theta|y\left[N\right]\right)$ is a probability density function (p.d.f.), the posterior for individual parameters or combinations of two parameters
158 can be found by *marginalisation*, i.e., integrating out the remaining parameters. The likelihood function $L\left(\theta; y\left[N\right]\right)$, however,
159 is *not* a p.d.f., and results for individual parameters are therefore obtained by *projections* onto individual parameters or
160 planes of two parameters. These projections can be computed and analysed in the framework of the *Profile Likelihood (PL)*
161 method, typically considered part of the *frequentist statistics* framework [21], in order to diagnose parameter identifiability
162 [5, 10, 21, 42].

163   If the prior $p\left(\theta\right)$ is chosen as *flat*, i.e., a *diffuse* prior is used, $p\left(\theta\right) = c$ for $\theta \in \Theta$ and $p\left(\theta\right) = 0$ for $\theta \notin \Theta$ where typically
164 $c = 1$, the posterior is *proportional* to the likelihood $p\left(\theta|y_{[N]]}\right) \propto p\left(y_{[N]}|\theta\right)$ over the *support* of the prior, i.e., where $p\left(\theta\right) \neq 0$
165 , since the *evidence* scaling constant $p\left(y_{[N]}\right)$ is independent of θ. If the prior is *flat* and *unbounded*, i.e., $p\left(\theta\right) = c$ for
166 $\theta \in \mathbb{R}^{n_\theta}$, the proportionality $p\left(\theta|y_{[N]]}\right) \propto p\left(y_{[N]}|\theta\right)$ holds for all θ. Hence, methods that operate on a *target distribution*
167 $\pi\left(\theta\right) \propto p\left(\theta|y_{[N]}\right)$, such as MCMC, can also be used with the likelihood $p\left(y_{[N]}|\theta\right)$ by assuming $p\left(\theta\right) = 1$ for $\theta \in \Theta$.

168   Observe that the use of a *feasible region* $\theta \in \Theta$ is equivalent to selecting a *uniform bounded* prior with a constant value
169 $c = 1$ in the defined space Θ and zero otherwise in $\mathbb{R}^{n_\theta}$. However, the introduction of such a feasible region does not exclude
170 the use of prior distribution $p\left(\theta\right)$, since one may well choose $\Theta = \mathbb{R}^{n_\theta}$. If a *non-uniform* prior is used *in addition* to a feasible
171 region Θ, this is equivalent to multiplying the non-uniform prior with a uniform bounded prior $p\left(\theta \in \Theta\right) = 1$.

172   Arguably, by effect of their *omission* in methods operating on the likelihood directly, the use of *flat unbounded* priors
173 is the *default* in the frequentist framework, but it is *non-typical* in Bayesian statistics [42, 43]. In practice, particularly
174 in engineering, there is often *some* prior information that could be made use of in the estimation in the form of a prior
175 distribution derived from physical system specifications.

176   For non-flat priors, many estimation methods based on the likelihood function can be modified to instead optimise on
177 the *posterior* by including the *prior* through Bayes' theorem in Eq. (6). An example of this is the modification of the PL

into the PP method presented in [21]. If numerical optimisation is used on the posterior distribution directly, i.e., a prior

is included with the likelihood function to form an objective function, the resulting parameter estimate is a MAP point

estimate. Indeed, this is supported in the CTSM framework as well [16, 18, 22].

2.3. Computing the likelihood and the posterior distribution for parameters of grey-box models

Both the MCMC and the PL/PP methods require evaluation of the likelihood function $L\left(\theta; y_{[N]}\right)$, either used directly

in PL, or for the evaluation of the posterior distribution $p\left(\theta | y_{[N]}\right)$ in PP and MCMC. The CTSM framework [8, 16, 17, 23]

presents a statistically well founded method for computing $L\left(\theta; y_{[N]}\right)$ for grey-box models on the SDE form of Eq 1.

The likelihood function is defined in Eq. (5). By application of the product rule $P\left(A \cap B\right) = P\left(A|B\right) P\left(B\right)$[38], Eq. (5)

can be expanded such that [16]:

$$L\left(\theta; y_{[N]}\right) = \left(\prod_{k=1}^{N} p\left(y_k | y_{[k-1]}, \theta\right)\right) p\left(y_0 | \theta\right) \tag{7}$$

In general, evaluating Eq (7) requires knowing the initial probability density function and successively solving the *Kolmogorov*

forward equation [16, 38]. However, by assuming a normal distribution for the one-step ahead prediction residuals, a simpler

alternative, the multivariate Gaussian distribution, can be used [16]:

$$L\left(\theta; y_{[N]}\right) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\epsilon_{k|k-1}^T \mathcal{E}_{k|k-1}^{-1} \epsilon_{k|k-1}\right)}{\sqrt{\det\left(\mathcal{E}_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^{n_y}}\right) p\left(y_0 | \theta\right) \tag{8}$$

By conditioning on knowing the *initial distribution* $p\left(y_0 | \theta\right)$, this expression can be iteratively evaluated in a Kalman Filter that estimates the quantities [16, 38]:

$$\hat{y}_{k|k-1} = \mathbb{E}\left[y_k | y_{[k-1]}, \theta\right] \tag{9}$$

$$\epsilon_{k|k-1} = y_k - \hat{y}_{k|k-1} \tag{10}$$

$$\mathcal{E}_{k|k-1} = \mathbb{E}\left[\epsilon_k \epsilon_k^T\right] \tag{11}$$

where $\hat{y}_{k|k-1}$ is the *predicted output* at time k given the measurements up to and including time $k-1$, i.e., the *one-step-ahead*

prediction. The choice of KF implementation depends on the type of state transition model; linear or non-linear, and in the

latter case, on the model being differentiable such that the model can be linearised for propagation of the covariance [44].

The assumption of normally distributed residuals can be verified by statistical testing [13, 16, 17, 22]. One possible

method is the use of a *cumulated periodogram (CP)*, which by use of plotting indicates if the resulting residuals are reasonably

approximated by a normal distribution [16, 17, 22]. Another, *numerical,* alternative is the use of the *Kolmogorov-Smirnov*

(KS) test criterion [13]. The KS criterion can also be used in combination with the CP diagram to compute confidence

bounds for the normality assumption on the CP diagram [17]. Other alternatives for normality testing include counting

zero-crossings, the *auto-correlation function (ACF)* [13], the *inverse* ACF or the *partial* ACF [17].

By taking the negative logarithm, and eliminating the factor $\frac{1}{2}$, the result $\ell_{\mathrm{L}}\left(\theta\right) = -2\ln L\left(\theta; y_{[N]}\right)$, where dependency

on $y_{[N]}$ is omitted in the sequel for notation simplicity, is obtained as

$$\ell_{\mathrm{L}}\left(\theta\right) = \sum_{k=1}^{N} \epsilon_{k1k-1}^T \mathcal{E}_{k|k-1}^{-1} \epsilon_{k|k-1} + \ln\left(\det\left(\mathcal{E}_{k|k-1}\right)\right) \tag{12}$$

If instead the *posterior* distribution $p\left(\theta | y_{[N]}\right) \propto L\left(\theta; y_{[N]}\right) p\left(\theta\right)$ is chosen, after eliminating the scaling by evidence $p\left(y_{[N]}\right)$

and applying the same transformation as above, $\ell_{\mathrm{P}}\left(\theta\right)$ is obtained as:

$$\ell_{\mathrm{P}}\left(\theta\right) = \ell_{\mathrm{L}}\left(\theta\right) - 2\ln p\left(\theta\right) \tag{13}$$

Hence, in log space, the application of a prior $p\left(\theta\right)$ is implemented by simply subtracting a value from $\ell_{\mathrm{L}}\left(\theta\right)$ that depends

only on the parameter θ. It is interesting to observe that the use of independent normal prior distributions $\mathcal{N}\left(\theta_{\mathrm{p,i}}, \sigma_{p,i}^2\right)$ for

each parameter in $\ell_{\mathrm{p}}\left(\theta\right)$ is similar to L^2-norm Tikhonov regularisation [45, 46], which indicates that application of non-flat

priors can be useful for improving *the generalisation* capability of a calibrated model.

2.4. The stochastic discrete time linear model

For a linear time invariant (LTI) model, which is the form typically used for thermal network models, Eqs. (1) and (2) can be written on discrete time form as [38]:

$$x_k = \tilde{A}x_{k-1} + \tilde{B}u_k + w_k \tag{14}$$

$$y_k = \tilde{C}x_k + v_k$$

where $w_k \sim \mathcal{N}(0, \mathcal{W})$ is the process noise (model error), $v_k \sim \mathcal{N}(0, \mathcal{V})$ is the measurement noise and the discrete time model matrices $\tilde{A} = \exp(\Delta t \cdot A)$ and $\tilde{B} = A^{-1}\left(\tilde{A} - I\right)B$ are computed from the standard linear continuous time model matrices A and B [45, 47]. Observe that the three model matrices $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$, and also the noise covariances $\mathcal{W}$ and $\mathcal{V}$ are typically functions of θ. For the noise covariances $\mathcal{W}$ and $\mathcal{V}$, the square root of the diagonal terms are included in θ, while the off-diagonal terms are assumed zero. This assumption is clearly reasonable for the measurement noise, but also commonly used for the process noise covariance [32]. A further extension on the presented work could be to include the off-diagonal terms of $\mathcal{W}$ in θ as well.

By using the SDE framework outlined in Section 2.3, the noise parameters in $\mathcal{W}$ and $\mathcal{V}$ influence $L\left(\theta; y_{[N]}\right)$ through the computed Kalman gain. In the limit case of zero measurement noise $\mathcal{V} \equiv 0$, the innovation covariance in the KF $\mathcal{E}_{k|k-1} = \tilde{C}X_{k|k-1}\tilde{C}^T$ and the standard equations for the linear Kalman Filter [48] give the Kalman gain

$$K_k = X_{k|k-1}\tilde{C}^T\left(\tilde{C}X_{k|k-1}\tilde{C}^T\right)^{-1} = \tilde{C}^{-1} \tag{15}$$

The aposteriori updated state is

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \tilde{C}^{-1}\left(y_k - \tilde{C}\hat{x}_{k|k-1}\right) = \tilde{C}^{-1}y_k \tag{16}$$

and the one-step ahead predicted output is

$$\hat{y}_{k|k-1} = \tilde{C}\left(\tilde{A}\tilde{C}^{-1}y_{k-1} + \tilde{B}u_k\right) + \epsilon_{k|k-1} \tag{17}$$

Hence, the model in the KF is *treated* as a first order autoregressive model in this limit case. However, the model structure and parametrisation are *still* the same grey-box TN structure and *not* the general black-box structure used in typical *Auto Regressive model with Exogenous input* (ARX) models. Since $X_{k|k} = \left(I - \tilde{C}^{-1}\tilde{C}\right)X_{k|k-1} = 0$, the state estimate covariance $X_{k|k-1} = \mathcal{W}$ and Eq. (8) with $\mathcal{E}_{k|k-1} = \tilde{C}\mathcal{W}\tilde{C}^T$ gives the *weighted least squares prediction error* parameter estimate.

In the limit case of $\mathcal{W} \equiv 0$, indicating a *deterministic* model with no diffusion term, it can be shown that the aposteriori state covariance $X_{k|k} \leq \tilde{A}^k X_0 \left(\tilde{A}^T\right)^k$ [38, 49] which will approach zero for a well-behaved stable system. If the initial state is also *deterministic*, $X_0 \equiv 0$, the state trajectory in the KF is *independent* of the measurements y_k, since $X_{k|k} = 0 \to X_{k|k-1} = 0$ and therefore

$$K_k = X_{k|k-1}\tilde{C}^T\mathcal{E}_{k|k-1}^{-1} = 0 \to \hat{x}_{k|k} = \hat{x}_{k|k-1} \tag{18}$$

Hence, Eq. (8) with $\mathcal{E}_{k|k-1} = \mathcal{V}$ gives the *weighted least squares* estimate for a *shooting/ballistic*, i.e., *deterministic*, state trajectory [38].

Both these limit cases are intuitively satisfactory and consistent with the common sense intuition of the KF. Given perfect measurements with $\mathcal{V} \equiv 0$, it is natural to rely exclusively on the measurement at the previous time-step at the expense of the previous estimates. In the case of $\mathcal{W} \equiv 0$, the perfect model *predictions* are trusted and the *data* ignored in the state propagation, with data only used to compute the error.

Since the SDE grey-box framework includes both of these limiting cases, it may arguably be considered a general framework, i.e., an *intermediate* between the purely autoregressive one-step-ahead prediction error and the deterministic output error, depending on the noise parameters. If both noise covariances $\mathcal{W}$ and $\mathcal{V}$ are non-zero, and correctly estimated or known apriori, the Kalman Filter gives the optimal estimate of the state.

Arguably, the limit case of $\mathcal{V} \equiv 0$ results in an LS parameter estimation that is similar to typical black-box methodology,

while the limit case $\mathcal{W} \equiv 0$ simulation error is more typical of a white-box modelling approach. Hence, the balance between these two limit cases through the Kalman gain can arguably be considered a mathematical expression for the intermediacy of grey-box models, between the white- and black-box approaches.

2.5. Identifiability of parameters

Since the model structure $\mathcal{M}$ is designed to be a *representation* of a system $\mathcal{S}$, it is often assumed that $\mathcal{S} \in \mathcal{M}(\Theta)$ and that consequently there exists a true parameter vector θ^* such that $\mathcal{M}(\theta^*) = \mathcal{S}$. However, this is rarely the case outside of simulation experiments, since the model structure $\mathcal{M}$ is only an *approximation* of $\mathcal{S}$. In the case of thermal network models based on a *naive* physical approximation of $\mathcal{S}$, the similarity of $\mathcal{M}$ to $\mathcal{S}$ is especially questionable. The estimate $\hat{\theta}$ depends on several factors, such as the amount of dynamic information in $\mathcal{K}$, the choice of objective function $g(\theta)$, and to some extent on the algorithm $\mathcal{A}$. Hence, the subject of *parameter identifiability* is of particular importance for simplified grey-box models, if the estimated parameters $\hat{\theta}$ are themselves of interest.

A model structure $\mathcal{M}$ may be *over-parameterised* such that a subset θ_s of the parameters has no effect on the model predictions $\hat{y}$, either because the model in Eqs. (1) and (2), and therefore also $g(\theta)$, is *free* of certain parameters, or the combined effect of several parameters cancels out. The parameters θ_s, denominated as *structurally* non-identifiable, result in *unbounded* confidence intervals (CI) [10]. Similarly, over-parametrisation may lead to the parameters in θ_s being *inter-dependant*, such that only some functional combination of the parameters are identifiable, resulting in equipotential, possibly bounded, manifolds in the parameter space. Additionally, if the dynamic information content in the data is *insufficient* for estimation of certain parameters, these parameters are *practically* non-identifiable [10]. Based on the definition given in [10, 21], parameters are practically non-identifiable when the likelihood is only *somewhat* affected by perturbations of the practically non-identifiable parameters, such that a well-defined optimum exists, but the likelihood is not sufficiently sensitive to produce a *bounded* CI at the desired level of confidence.

The use of CIs as diagnostic criteria for identifiability is a distinctly *frequentist* statistics approach [40]. A formal definition of non-identifiability, based on the Bayesian framework of computing probability distributions of parameters, is given in [25, 42]. The subset of *identifiable* parameters is defined such that $\theta = (\theta_i, \theta_s)$. Parameters θ_s are non-identifiable if [25]:

$$p\left(\theta_s | \theta_i, y_{[N]}\right) = p\left(\theta_s | \theta_i\right) \implies \theta_s \perp\!\!\!\perp y_{[N]} | \theta_i \tag{19}$$

That is, no *additional* information is obtained about θ_s from the data $y_{[N]}$ once the identifiable parameters θ_i are known [25]. Hence, the non-identifiable parameters are *conditionally independent* of the data, given the identifiable parameters [25]. Since $p\left(\theta_s | \theta_i, y_{[N]}\right) \propto L\left(\theta_i, \theta_s; y_{[N]}\right) p\left(\theta_s | \theta_i\right) p\left(\theta_i\right)$, Eq. (19) implies that the likelihood $L\left(\theta_i, \theta_s; y_{[N]}\right)$ is *free*, i.e., unaffected, by θ_s [25, 42], which is similar to the description of *structural* identifiability given in [10, 21].

As discussed in Section 2.4, the measurement and noise covariance matrices $\mathcal{W}$ and $\mathcal{V}$ are here considered functions of θ. Specifically, the noise covariance matrices are assumed diagonal, with the square root of the non-zero terms included in θ. Identifiability of these parameters is treated in the same way as for the thermal model parameters. For a more thorough analysis of noise model parameter identifiability, see e.g. [35].

Resolving non-identifiability by application of a prior

If both the likelihood and the priors are non-informative for a sub-set of the parameters, there is clearly a problem with the application of *any* parameter estimation method, since there is *no* information from which to estimate the non-identifiable parameters. The solution is to *introduce more information* into the parameter estimation problem, by either redesigning the experiment to obtain more *informative* data and/or new measurements, or by *revising* the model structure to better fit the available data. A third possibility is the addition of a *non-flat* prior distribution, based on prior physical information of the system. Experimental design is particularly challenging for the study of building thermal behaviour since buildings are

subject to weather conditions and occupancy demands that are usually beyond experimental control [4]. Hence, the use of priors to resolve non-identifiability is particularly interesting for building thermal modelling.

The local sensitivity of the log posterior distribution in Eq. (13) to perturbations of θ can be estimated by the Hessian:

$$H_P = \nabla^T \nabla \ell_{\mathrm{P}} \left(\theta; y_{[N]} \right) \big|_{\theta = \hat{\theta}} = H_{\mathrm{L}} - \nabla^T \nabla 2 \ln p \left(\theta \right) \big|_{\theta = \hat{\theta}} \tag{20}$$

where $H_{\mathrm{L}} = \nabla^T \nabla \ell_{\mathrm{L}} \left(\theta; y_{[N]} \right) \big|_{\theta = \hat{\theta}}$ is the Hessian of the likelihood function [10, 16, 50]. Hence, if the *likelihood* is insufficiently affected by perturbations of θ in certain directions, as indicated by H_{L}, the addition of a prior can be seen to introduce another source of sensitivity to perturbations of θ and therefore resolve the non-identifiability. Note that while a prior may resolve non-identifiability and therefore result in unambiguous parameter estimates, it does not necessarily guarantee a physical interpretability of the estimated parameters. Note also that the obtained H_{P} describes the sensitivity of $\ell_{\mathrm{P}} \left(\theta; y_{[N]} \right)$ which is data dependent [10, 16, 50].

2.6. Profile Likelihood and Profile Posterior

The PL method [5, 10] can be used to estimate uncertainty and diagnose identifiability of the parameters by *projecting* the likelihood function $L \left(\theta; y_{[N]} \right)$ onto each parameter θ_i. The *likelihood profile* $\ell_{\mathrm{PL1D}} \left(\theta_i \right)$ is defined as the *minimum negative log likelihood* $\ell_{\mathrm{L}} \left(\theta \right)$, computed for values of a single parameter θ_i, when the remaining parameters $\theta_{j \neq i}$ are *freely* optimised [10, 51]:

$$\ell_{\mathrm{PL1D}} \left(\theta_i \right) = \min_{\theta_{j \neq i}} \ell_{\mathrm{L}} \left(\theta_{j \neq i}; y_{[N]}, \theta_i \right) \tag{21}$$

Values of θ_i are chosen, either by a brute force discretisation of θ_i or using a gradient decent method, prior to optimising the remaining $\theta_{j \neq i}$ [10]. A likelihood-based CI can be obtained by applying a *threshold* to the likelihood function [10, 51]. Let

$$\left\{ \theta : \ell_{\mathrm{L}} \left(\theta \right) - \ell_{\mathrm{L}} \left(\hat{\theta} \right) < \Delta_\alpha \right\} \quad , \quad \Delta_\alpha = \chi^2 \left(\alpha, n_{\mathrm{df}} \right) \tag{22}$$

where $\hat{\theta}$ is a freely estimated, presumed optimal parameter vector, and the threshold Δ_α is the α percentile of the χ^2-distribution with n_{df} degrees of freedom [52]. By using Eq. (22) to set a threshold on the likelihood profile $\ell_{\mathrm{PL}} \left(\theta_i \right)$ of each parameter, it is possible to diagnose parameter identifiability. As discussed in Section 2.5, *structurally* non-identifiable parameters produce *unbounded* CIs, or equivalently, *flat* likelihood profiles [10]. A likelihood-based CI, unlike the Hessian based *asymptotic* CI, is not necessarily symmetric, and can therefore be unbounded in one direction. Hence, a *practically* non-identifiable parameter can be diagnosed if the at least half *unbounded* likelihood profile has a well-defined minimum [10]. Only parameters that produce bounded CIs, and consequently have sufficiently convex likelihood profiles, are identifiable by optimisation of $\ell_{\mathrm{L}} \left(\theta \right)$.

The PL method can be extended to project the posterior distribution, rather than the likelihood function, by inclusion of a prior $p \left(\theta \right)$ by Bayes' theorem [21]. The PP method is defined, similarly to Eq. (21), by obtaining the *posterior profile* $\ell_{PP1D} \left(\theta_i \right)$ as the *minimum negative log posterior, given in* Eq. (13), for a prescribed value of θ_i when the remaining parameters are freely estimated, i.e.:

$$\ell_{\mathrm{PP1D}} \left(\theta_i \right) = \min_{\theta_{j \neq i}} \ell_{\mathrm{P}} \left(\theta_{j \neq i}; y_{[N]}, \theta_i \right) \tag{23}$$

As for the PL method, the posterior profile is obtained for some selected values of θ_i, and subsequently plotting $\ell_{PP1D} \left(\theta_i \right)$. Observe that by replacing the log *likelihood* ℓ_{L} by the log *posterior* ℓ_{P}, the obtained profile is offset by the log of the *prior*, $-2 \ln p \left(\theta_i \right)$. Finally, observe that the PL method can be considered as a *special case* of the PP method, with the prior $p \left(\theta \right) = 1 \rightarrow -2 \ln p \left(\theta \right) = 0$ for all $\theta \in \mathbb{R}^{n_\theta}$.

Profiling in two parameter dimensions

The typical implementation of the PL/PP method [5, 10, 21] projects the likelihood/posterior of the n_θ dimensional space Θ onto the single parameter θ_i. These projections are known to *overestimate* the width of the obtained profiles if there are *inter-dependent* parameters. Hence it is of interest to project the likelihood/posterior in a way that visualises potential

10

parameter interactions. A possible modification of the PL method is then to hold out *two* parameters rather than one, hence the PL2D method obtains [44, 45];

$$\ell_{\text{PL2D}}\left(\theta_i, \theta_j\right) = \min_{\theta_{k \neq i,j}} \ell_{\text{L}}\left(\theta_{k \neq i,j}; y_{[N]}, \theta_i, \theta_j\right) \tag{24}$$

PL2D projects the log likelihood onto the plane $\Theta_{i,j} = (\theta_i, \theta_j)$ s.t. $\theta_i, \theta_j \in \Theta$. The resulting two-dimensional profiles can be analysed similarly to the one-dimensional profiles [10], using the definition in Eq. (22). The profiles are computed for all combinations of parameters, i.e., by projecting the objective function to all possible planes $\Theta_{i,j}$. Since $\ell_{\text{L}}(\theta)$ is typically similar for neighbouring θ, previous PL2D estimates can be used as a warm-start for new points in $\Theta_{i,j}$ to improve computational efficiency [20]. A *confidence region* in the $\Theta_{i,j}$ plane is obtained by applying the Δ_α threshold from Eq. (22). Observe that since the optimal estimate $\hat{\theta}$ has n_θ free parameters while the PL2D estimate has $n_\theta - 2$, this gives $n_{\text{df}} = 2$ for the computation of Δ_α from the χ^2-distribution in Eq. (22). Based on these two-dimensional profiles, and the computed confidence regions, parameters are considered identifiable if their corresponding confidence regions are bounded in all directions. If the region contains an unbounded equipotential *valley* in the log likelihood space, the parameter is considered structurally non-identifiable. If the profile has a well-defined minima, but is unbounded in one direction, i.e., the log likelihood is below the Δ_α threshold, this indicates a practically non-identifiable parameter [10]. Subsequently, the size and shape of a *bounded* region estimates the *accuracy* with which the parameters can be estimated.

The free estimate $\hat{\theta}$ may with advantage be chosen as the minimum $\ell_{\text{PL2D}}\left(\theta_i, \theta_j\right)$ obtained from *all* profiles, since such a search *approximates*, subject to the limitations imposed by discretisation in the brute force exploration, a *free* optimisation of *all* parameters, using the already computed ℓ_{PL2D} results. Since the PL2D profiles cover the entire parameter space Θ, this procedure is less affected by local minima than a direct numerical optimisation.

The PL2D method may also be modified to project the posterior rather than the likelihood, thus the PP2D method projects:

$$\ell_{\text{PP2D}}\left(\theta_i, \theta_j\right) = \min_{\theta_{k \neq i,j}} \ell_{\text{P}}\left(\theta_{k \neq i,j}; y_{[N]}, \theta_i, \theta_j\right) \tag{25}$$

This modification is analogous to the extension of the PL1D method into the PP1D method.

2.7. MCMC

The *projection* methods PL1D/PL2D, based on the interpretation of CIs, are typically considered part of a *frequentist* approach to parameter estimation [21]. In the *Bayesian* framework, the goal is to infer a *probability distribution* for the parameter θ, now considered a random variable. Given that the posterior distribution is often not analytically obtainable, the *Markov Chain Monte Carlo (MCMC)* method is instead used to compute an *estimate* of the posterior. Unlike regular Monte Carlo (MC) methods, MCMC draws samples of θ, such that each sample depends on the *previous* sample, by defining a *transition probability* $p\left(\theta_k | \theta_{k-1}\right)$. If the transition probability is chosen to fulfil the detailed balance equation

$$\pi\left(\theta_{k-1}\right) p\left(\theta_k | \theta_{k-1}\right) = \pi\left(\theta_k\right) p\left(\theta_{k-1} | \theta_k\right) \tag{26}$$

the generated samples will be drawn *proportional* to the *target distribution* $\pi(\theta) \propto p\left(\theta | y_{[N]}\right)$. Hence, the posterior and its parameters, e.g., mean and covariance, can be approximated by computing the *empirical* distribution as a histogram over the sequence of samples $\theta_{[K]}$. In this work, the MCMC method of choice is the basic *Metropolis* algorithm [30, 31] using a normal isotropic *proposal distribution* $\theta_k^c \sim q\left(\theta_k | \theta_{k-1}\right) = \mathcal{N}\left(\theta_{k-1}, \Sigma_q\right)$ where θ_k^c is a *candidate* for the next step θ_k in the Markov Chain, and Σ_q is the covariance of the proposal distribution, centred on the current step θ_{k-1} [27–29]. The work of Hastings [28, 31, 53], a generalisation of the work of Metropolis [28, 30], shows that if the proposal distribution $q\left(\theta_k | \theta_{k-1}\right)$ is chosen such that it ensures every possible value of θ will *eventually* be visited, and this is combined with an *acceptance probability* test of the generated proposal, the resulting *transition probability* $p\left(\theta_k | \theta_{k-1}\right)$, constituted of the combined proposal-acceptance scheme, fulfils the requirement of Eq. (26). The acceptance criterion using a normal proposal

11

distribution is defined from the probability ratio:

$$\alpha = \frac{\pi\left(\theta_k^c\right)}{\pi\left(\theta_{k-1}\right)} = \exp\left(0.5\left(\ell_{\mathrm{P}}\left(\theta_{k-1}\right) - \ell_{\mathrm{P}}\left(\theta_k^c\right)\right)\right) \tag{27}$$

The next step in the Markov Chain is then chosen as θ_k^c with probability $p_{\mathrm{a}} = \min\left(1, \alpha\right)$. Observe that α is greater than 1 if the proposal constitutes an *improvement*, in which case the proposal will be accepted with probability 1 [27, 28].

Posterior predictive distribution

An advantage of the Bayesian parameter estimation framework, and of the MCMC method, is that the representation of parameter uncertainty, expressed in MCMC as the empirical distribution of the samples $\theta_{[K]}$, enables better estimation of the model's *prediction* uncertainty. The *posterior predictive distribution* can be inferred from a set of simulated state/output trajectories obtained by Monte Carlo (MC) simulation of the model in Eq. (14).

Note that by using the covariance propagation equations of an LTI system [48], it is possible to compute the uncertainty of the predicted state and output trajectory for a *single* parameter estimate. However, the use of the MCMC sampled set $\theta_{[K]}$ allows accounting for uncertainty in the *parameters*. Additionally, the MC simulation method is not restricted to linear or time invariant systems. For a test dataset of length N, assume x_0 is known with covariance $X_0^{(i)}$ and let $\hat{x}_{0|0} \sim \mathcal{N}\left(x_0, X_0\right)$. Then, for each time $k \in [1, N]$ compute

$$\hat{x}_{k|0}^{(i)} = \tilde{A}\hat{x}_{k-1|0}^{(i)} + \tilde{B}u_k + w_k$$
$$\hat{y}_{k|0}^{(i)} = \tilde{C}\hat{x}_{k|0} + v_k \tag{28}$$

where $\hat{x}_{k|0}^{(i)}$ and $\hat{y}_{k|0}^{(i)}$ are the *estimated* future state and output at time k, given only measurement information at time 0, computed using the i-th accepted parameter proposal in $\theta_{[K]}$. The process noise $w_k \sim \mathcal{N}\left(0, \mathcal{W}\right)$ and the measurement noise $v_k \sim \mathcal{N}\left(0, \mathcal{V}\right)$ are *drawn* independently at each time-step, for each i-th trajectory, using a random number generator (RNG). The model matrices $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$, and the covariance matrices X_0, $\mathcal{W}$ and $\mathcal{V}$, are all potentially functions of the i-th sample in $\theta_{[K]}$, hence potentially different for each trajectory. Over these K trajectories, the *distribution* of the *predicted output* for the test set is computed, for each time-step k, as a histogram over the set of estimated outputs $\hat{y}_k^{(i)}$, $i \in 1, 2, \ldots, K$. A similar approach is used in [32].

2.8. Comparing MCMC and Profiling methods

Exploration by drawing samples

The projection based PL/PP methods explore Θ by selecting samples of θ *deterministically*. If a brute force method is used, where the parameter θ_i or the plane $\Theta_{i,j}$ is discretised with a prescribed resolution, the sampled values for each computed profile are completely determined apriori. If a hill-climbing method is used, the next sample is also determined deterministically by evaluating the gradient of the current sample. In contrast, the MCMC method explores the parameter space Θ *stochasticity*, using randomisation to select the next sample, such that each new sample is drawn proportionally to the target distribution $\pi\left(\theta\right)$ [27–29]. Hence, assuming proper mixing of the chains, the majority of the samples will be drawn from the regions of high posterior density. These are, naturally, the regions of most interest for inference about the parameters [27–29]. Subsequently, again assuming proper mixing of the chains, the majority of the computation time will be spent analysing the most interesting regions in Θ.

In contrast, the deterministic brute force sampling of the PL/PP methods explore the parameter space Θ *exhaustively* within the prescribed discretisation, which is significantly more time-consuming. The advantage of such exhaustive searches is that they are guaranteed to obtain the global optimum, within the precision allowed by the discretisation of Θ. Additionally, deterministic exploration is unaffected by the flat manifolds caused by non-identifiable parameters, whereas the stochastic exploration of MCMC in such conditions can result in convergence failure for chains of finite length [21]. Observe that the MCMC methods with appropriately selected proposal distributions are also theoretically guaranteed to obtain the global
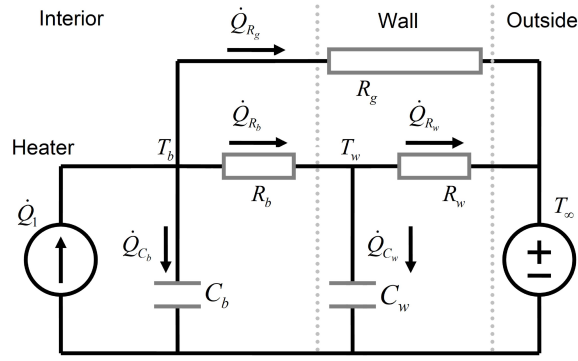
Figure 2: RC circuit model of the building.

optima for infinite chain lengths [27–29]. In practice however, MCMC samples Θ sufficiently for parameter inference even with reasonably short chain lengths.

Since the profiling methods explore the posterior by projections onto individual parameters, or planes of two parameters, the method must be executed repeatedly for each parameter or combination of parameters of interest. This further exacerbates the computational burden. For the one-dimensional projection methods PL1D and PP1D, computation time is linear in the number of parameters n_θ and usually comparable to MCMC. For the PL2D/PP2D methods, however, the computation time is exponential in n_θ, thus, even moderately large numbers of parameters may lead to infeasible computation times.

Projection and marginalisation

Since the MCMC method draws samples in proportion to the target distribution $\pi(\theta) \propto p(\theta|y_{[N]})$, the posterior distribution $p(\theta|y_{[N]})$, or its hyper-parameters, can be estimated directly on the set of samples $\theta_{[K]}$, e.g., by computing a histogram [27–29]. In order to plot the results, the posterior is often presented as *marginalised* distributions over one or two parameters. It is common practice to present *marginal distributions* for all possible combinations of parameters and present the results as *corner plots* [27].

In contrast, the PL/PP methods obtain the estimated profiles by *projecting* the likelihood/posterior onto individual parameters or planes of two parameters. The resulting profiles are *similar* to the *marginalised histograms* obtained by MCMC, but with one important difference. The projections are computed using *optimisation* over the remaining parameters, as illustrated in Eq. (23). This procedure returns the *optimal* density for the given θ_i, or given (θ_i, θ_j) pair for PL2D/PP2D. In contrast, the marginalisation used in MCMC computes the *integral* over the remaining parameters. For some distributions, such as the normal distribution, these two quantities are *proportional*. Hence, if the *scale* of the resulting profiles/distributions is ignored, these methods will, for some cases, result in *similar* profiles/distributions, particularly for the high posterior density regions where the stochastic exploration of MCMC gives the most accurate results.

3. Experimental setup

3.1. Model

Figure 2 shows a thermal network model structure, which was developed to approximate the thermal behaviour of an experimental building, located at Campus Porsgrunn of the University of South-Eastern Norway (USN). The model is partially based on the R4C2 model presented in [7]. The RC circuit consists of five components: the thermal resistance between room air and wall R_{b}, the building envelope R_{w}, and the thermal resistance of windows and doors R_{g}, and the two capacitances C_{b} and C_{w} representing the thermal capacitance of the building interior and envelope, respectively. The model has two outputs: the room temperature T_{b} and the wall surface temperature T_{w}, and two inputs: the consumed power by an electric heating element $\dot{Q}$ and the outside temperature T_∞. The model can be expressed on the form of Eqs. (14) with
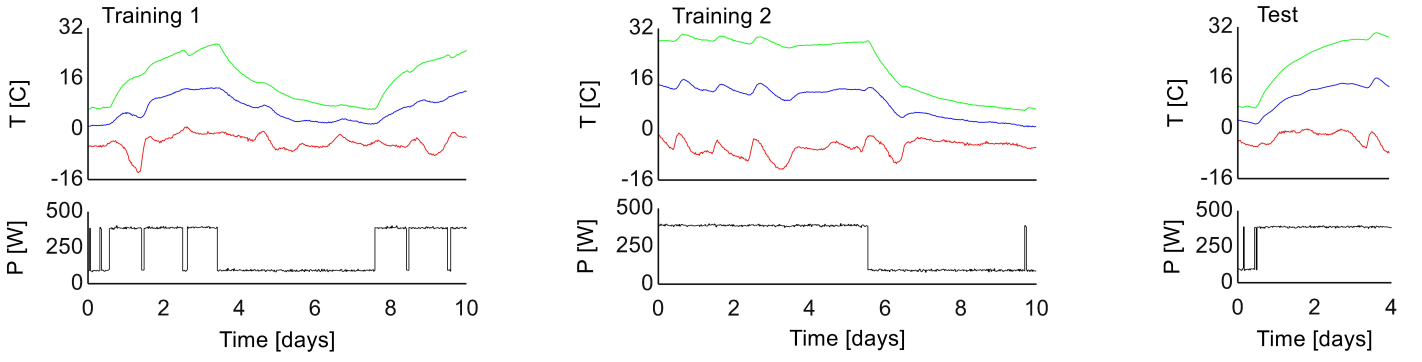
13

Figure 3: Training and test datasets, consists of three temperature measurements, T_b(green), T_w(blue) and T_∞(red), and one measurement of input electrical power, $\dot{Q}$.

state transition matrix A, input matrix B, state vector x and input vector u given as:

$$A = \begin{bmatrix} -\frac{1}{C_b R_b} - \frac{1}{C_b R_g} & \frac{1}{C_b R_b} \\ \frac{1}{C_w R_b} & -\frac{1}{C_w R_b} - \frac{1}{C_w R_w} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{1}{C_b} & \frac{1}{C_b R_g} \\ 0 & \frac{1}{C_w R_w} \end{bmatrix} \tag{29}$$

$$x_k = \begin{bmatrix} T_b \\ T_w \end{bmatrix}_{t=t_k}, \; u_k = \begin{bmatrix} \dot{Q} \\ T_\infty \end{bmatrix}_{t=t_k}$$

Since all states are observable, the measurement matrix $C = I \to \hat{y}_k = \hat{x}_k$. The model is LTI, hence a standard KF can be used. The noise covariance matrices $\mathcal{W} = \text{diag}\left(\text{w}_b^2, \text{w}_w^2\right)$ and $\mathcal{V} = \text{diag}\left(\text{v}_b^2, \text{v}_w^2\right)$ are also estimated from data, and are assumed diagonal. The parameter vector is then $\theta = \begin{bmatrix} R_g & R_b & R_w & C_b & C_w & w_b & w_w & v_b & v_w \end{bmatrix}$.

A variation of this R3C2 model, is the R2C2 model where the thermal resistance R_g is removed, equivalent to setting $R_g \equiv \infty$ in the R3C2 model.

3.2. Training and test datasets

Figure 3 shows three *independent* sets of data, collected from the experimental building in February 2018, which consist of three temperature measurements, T_b, T_w and T_∞, and one measurement of input electrical power, $\dot{Q}$, supplied to an electric heater. The data has been downsampled to a sampling interval of 30 minutes. This sample interval was determined experimentally by repeatedly increasing the downsampling ratio and using the PL1 method to test that the downsampled data produced similar results as the higher sample rate original data-set. Note that a sample time of 30 minutes is arguably reasonable for the main thermal behaviour of a building, but may be excessively long for the heater dynamics and solar gains. However, for this particular data-set, a sample rate of 30 minutes was found acceptable. The temperatures T_b and T_w are used as reference data for the model outputs, while T_∞ and $\dot{Q}$ are the model inputs. The two training datasets are used for parameter estimation and analysis, while the testset is used only for evaluation of the *posterior predictive distribution*, i.e., to evaluate how well the calibrated model predicts future system behaviour.

3.3. Experiment cases and setup

In the sequel, five different experiment configurations, as listed in Table 2, are analysed and compared.

Case 1 uses the full R3C2 model from Fig. 2 with the priors for all parameters $p(\theta) = 1$ for $\theta \in \mathbb{R}^{n_\theta}$. As the results in Section 4 show, Case 1 results in *non-identifiable* parameters. As discussed in Section 2.5, there are several ways to resolve parameter non-identifiability.

Case 2 uses the same model structure, but with the addition of a *prior* on the parameter R_g. The parameter R_g represents the thermal resistance of windows and the door, and can hence be computed by hand. The door in the building has a U-value

14

Table 2: Configuration for each experiment case.

#	Model	Description
1	R3C2	uniform priors $p(\theta) = 1$
2	R3C2	$p(R_{\mathrm{g}}) = \mathcal{N}\left(0.24,\, 0.01^2\right)$
3	R2C2	R_{g} removed, $p(\theta) = 1$, $\theta_0 = \hat{\theta}_{\mathrm{MAP}}$
4	R2C2	Same as 3 + added noise $\sigma = 0.1$
5	R2C2	Same as 3, but using Training 2 dataset

of $1.2\left[\frac{W}{m^2 K}\right]$ and an area of $1.76\left[m^2\right]$, while the two windows have U-values $1.3\left[\frac{W}{m^2 K}\right]$ and a total area of $1.57\left[m^2\right]$. The resulting total UA value is then $4.1\left[\frac{W}{K}\right]$, which gives an estimated thermal resistance $R_{\mathrm{g}} = 0.24$ [34]. The covariance of the prior, i.e., the uncertainty of the estimated mean value 0.24 is *chosen* as 0.01^2. With application of a prior distribution based on physical information of the building, the parameters are shown to be identifiable.

Case 3 instead resolves the non-identifiability by modifying the model structure into the R2C2 model, by removing the parameter R_{g} from the model and effectively lumping the thermal resistance of windows and the door together with the remaining R_{b} and R_{w}. All four parameters of the R2C2 model structure are identifiable, despite using uniform priors. Additionally, Case 3 starts the MCMC chains from the MAP estimate $\hat{\theta}_{\mathrm{MAP}}$, rather than drawing the initial sample uniformly from the feasible region Θ as is done in Cases 1 and 2, thus negating the need for a burn-in phase in MCMC.

Case 4 uses the same setup as Case 3, except that a random noise component $v'_k \sim \mathcal{N}\left(0,\, 0.1^2\right)$ is *added* to the data for T_b and T_w prior to analysing the estimated parameters. As the results will show, comparing Cases 3 and 4 reveal some interesting insight into the estimation of *noise covariance parameters* for this model. Case 5 also uses the same setup as Case 3, but now a different dataset, *Training 2*, is used. The other four cases all use *Training 1* for estimation and analysis. For Case 5, however, the Training 2 dataset has slightly more dynamic information content, which, as the results will show, is reflected in the parameter analysis.

For each case, the posterior distribution of the parameters $p\left(\theta | y_{[N]}\right)$ is estimated using the MCMC method. The results are presented as *marginal* distributions, both as one dimensional (1D) for each parameter, and as two dimensional (2D) distributions over two parameters. Additionally, each case is analysed using the profiling methods of Section 2.6 in one and two dimensions in order to obtain *projected profiles* of the log posterior $\ell_{\mathrm{P}}(\theta)$. Note that the experimental cases use different feasible regions Θ, as evident from the result plots in Section 4. Hence, the results for different cases must be compared by taking into account the differences in parameter limits. Note that, for simplicity, the projection method is refereed to in the sequel as PP, since the PL variation can be considered a special case of PP with uniform priors.

Post-processing of results

The marginal posterior distributions from MCMC are *log-transformed*, in the form of Eq. *(13),* to facilitate comparison with the PP1D/PP2D methods. Additionally, the results are *shifted* in log space such that the minimum of each profile/log distribution is zero, as discussed in Section *2.8.* Since the goal is to analyse the parameter estimation problem, it is the *shape* in log space and *distribution* over the parameters that are of interest, not the scale or the minimum log posterior value.

The PP1D/PP2D methods, being based in numerical optimisation, naturally respect the bound of the feasible region. In the Bayesian framework, the constraint $\theta \in \Theta$ is equivalent to a prior distribution $p(\theta \in \Theta) = 1$ and $p(\theta \notin \Theta) = 0$, hence in the MCMC implementation any proposed $\theta \notin \Theta$ is automatically rejected.

The results from MCMC and PP2D are presented as *corner plots*, with 2D profiles/marginal posterior distributions for each possible combination of parameters. Additionally, the marginal posterior for each parameter is plotted together with the PP1D profile, for comparison of results.
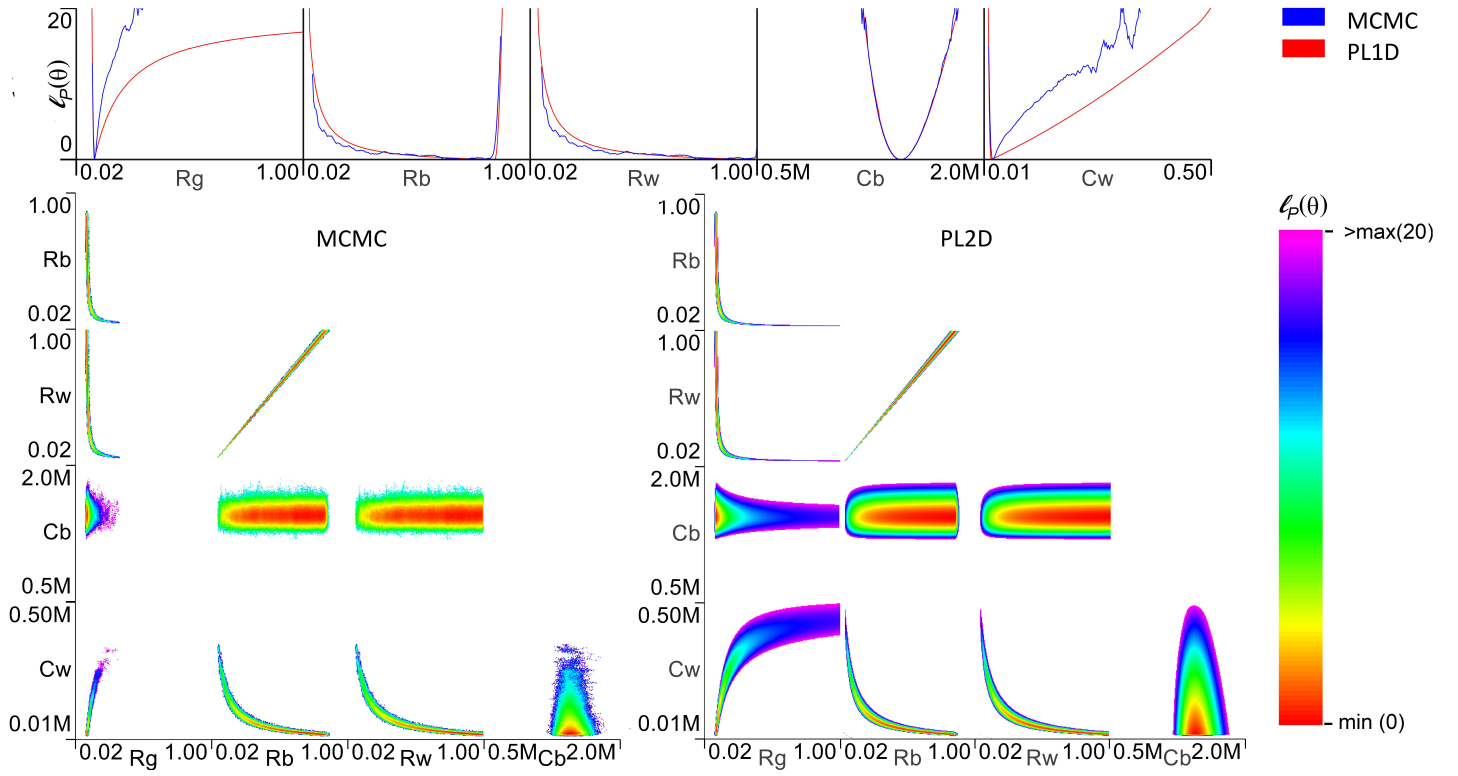
Figure 4: Comparison of projected profiles from PP1D and PP2D, with the marginalised posteriors obtained from MCMC, for Case 1. The top panel shows the 1D marginals/projections plotted together, where the PL1D projections are plotted in red and the marginalised posterior for each parameter is plotted in blue. The lower left and right panels show corner plots, a set of two-dimensional distributions/projections, one for each possible parameter combination, for the MCMC marginalised posterior and the PL2D projections, respectively.

Tuning

The PP2D method is executed with an experimentally obtained discretisation resolution of 200×200 grid, and a resolution of 400 for PP1D. The MCMC method is applied with 12 chains of length 10^5, except for in the non-identifiable Case 1 which uses chain length 2×10^6 and a thinning factor of 20. For Cases 1 and 2, a fixed burn-in of 10000 samples is used. Cases 3, 4 and 5 initialise the chains at $\hat{\theta}_{\mathrm{MAP}}$, hence, no burn-in phase is needed. The proposal distribution is chosen as normal isotropic: $q\left(\theta_k | \theta_{k-1}\right) \sim \mathcal{N}\left(\theta_{k-1}, \sigma_q^2\right)$ where $\sigma_q = l \cdot diag\left(\theta^0\right)$ and θ^0 is some nominal parameter vector close to the MAP estimate $\hat{\theta}_{\mathrm{MAP}}$. The step length $l = 0.01$ has been selected for all cases, such that for Cases 2 to 5 the proposal *acceptance rate* is around 50%. For Case 1, the acceptance rate is found to be around 25, due to the thin elongated valley in the posterior hyper-surface for the non-identifiable case resulting in an increase in rejected proposals.

4. Results and analysis

4.1. Marginal and projected posteriors

Figure 4 presents the marginal posterior plots from MCMC together with the PP1D and the PP2D projections. Observe that the resulting projections/profiles are *similar* for most parameters. Since the plots are shifted in log space, this similarity indicates *proportionality* in non-log space.

The presence of flat, equipotential regions in the posterior hyper-surface indicates that parameters R_b and R_w are *non-identifiable*. The corresponding PP2D profile in Fig, 4 shows a linear *inter-dependence* between the parameters R_b and R_w, which is indicative of a *structural* problem with the R3C2 model, resulting in non-identifiable parameters because of over-parameterisation.

Next, observe that the marginal posteriors for parameters R_b and R_w show considerable random fluctuation in regions that the PP1D and PP2D methods identify as *flat*. Since MCMC is based on a *stochastic* exploration of the parameter space
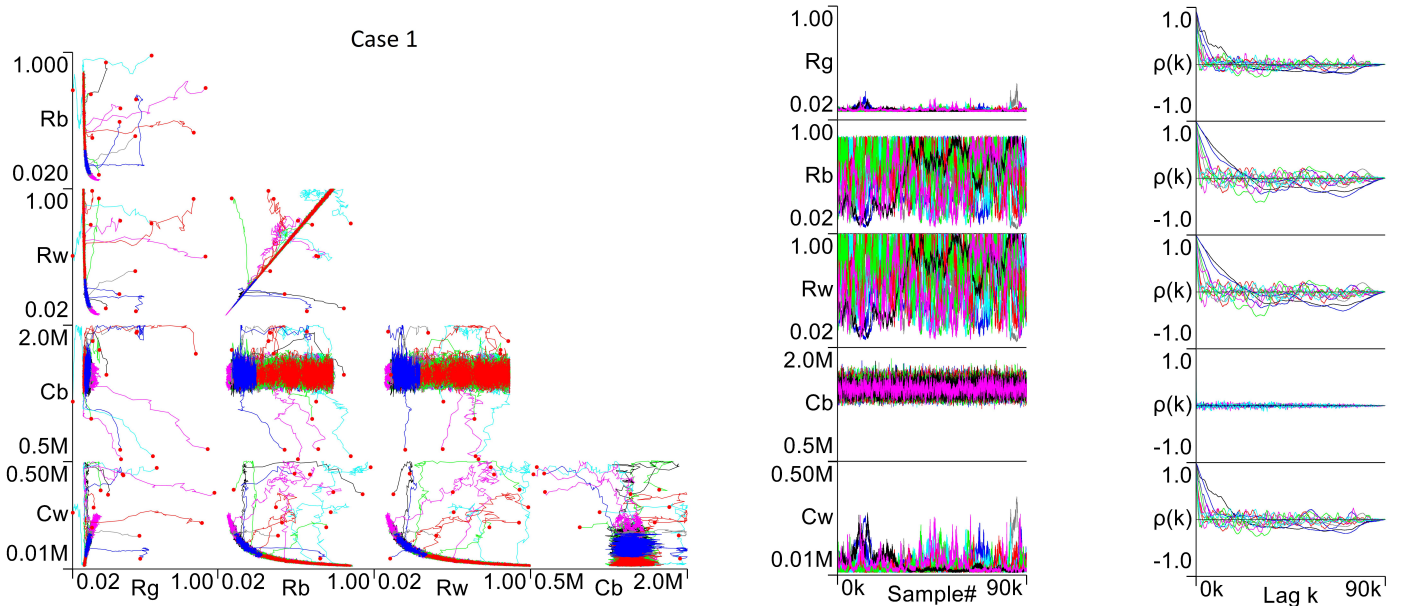
Figure 5: Diagnostic plots for Case 1: 2D and 1D trace plots are shown in the left and middle panels. The 2D trace plots contain the burn-in phase of 10000 samples, indicating that all chains reach the high-density regions. The right panel shows the ACF plot for each chain after removal of the burn-in phase, which indicates that the MCMC method does not properly converge for Case 1.

it is not surprising that the similarity of the results is only *approximate*. The MCMC method draws samples in *proportion* to their posterior density, focusing exploration of Θ on areas of high posterior density, hence the similarity of the results is stronger in these regions.

The differences between projected and marginal posteriors are most pronounced for the parameters R_g and C_w. The marginal posteriors are *not* proportional to the projected posteriors for R_g and C_w, since for these parameters, the optimum obtained by projections is *not* proportional to the integral over remaining parameters. In contrast, the marginal and projected posterior of the parameter C_b are nearly identical in shape.

Since the parameters are subject to the constraint $\theta \in \Theta$, the inter-dependence between R_b and R_w introduces *artefacts* in the profiles, such as the *sharp bend* that occurs in the profile for R_b at ~ 0.9. This phenomenon is caused by the dependant parameter R_w being *actively constrained* < 1.0, hence producing sub-optimal posterior projections for higher values of R_b. The same effect is observed on the MCMC marginal posterior plots, since the bends are caused by the bounds on the parameters and not the analysis method. However, in the Bayesian framework, the constraint $\theta \in \Theta$ can be interpreted as a *prior* on the parameters which *reshapes* the likelihood hyper-surface, consequently resulting in the observed bends in the marginal posteriors of R_b and C_w, as discussed in Section 2.2. The Bayesian interpretation of this phenomenon is arguably more satisfactory than that of *artefacts* induced by active constraints in optimisation.

Because of inter-dependant, and therefore non-identifiable, parameters, the *shape* and *extensiveness* of the posterior hyper-surface become difficult to traverse using the stochastic predict-accept/reject scheme of the Metropolis algorithm. These difficulties are evident by the diagnostic plots given in Fig. 5. For Case 1, the elongated, narrow and flat structure of the posterior hyper-surface for parameters R_b and R_w causes the Metropolis algorithm to sample the posterior somewhat ineffectively, resulting in an average proposal *acceptance rate* of $\sim 25\%$ for the defined proposal distribution $q\left(\theta_k | \theta_{k-1}\right)$, subsequently with high autocorrelation over the chains. Hence, significantly longer chain lengths were required for Case 1, where $K = 2 \times 10^6$ for all 12 chains, than for the other four cases.

Although the 2D trace plots in Fig. 5 show that all chains quickly reach the high posterior density region for all parameters, from their uniformly drawn starting points $\theta_0 \sim \mathcal{U}\left(\theta_{min}, \theta_{max}\right)$, the 1D trace plots show that the chains are not reaching equilibrium, except for in the parameter C_b. Since the chains do not converge, the resulting parameter samples $\theta_{[K]}$ are not properly representing the posterior distribution, hence producing less accurate estimates of $p\left(\theta | y_{[N]}\right)$. The
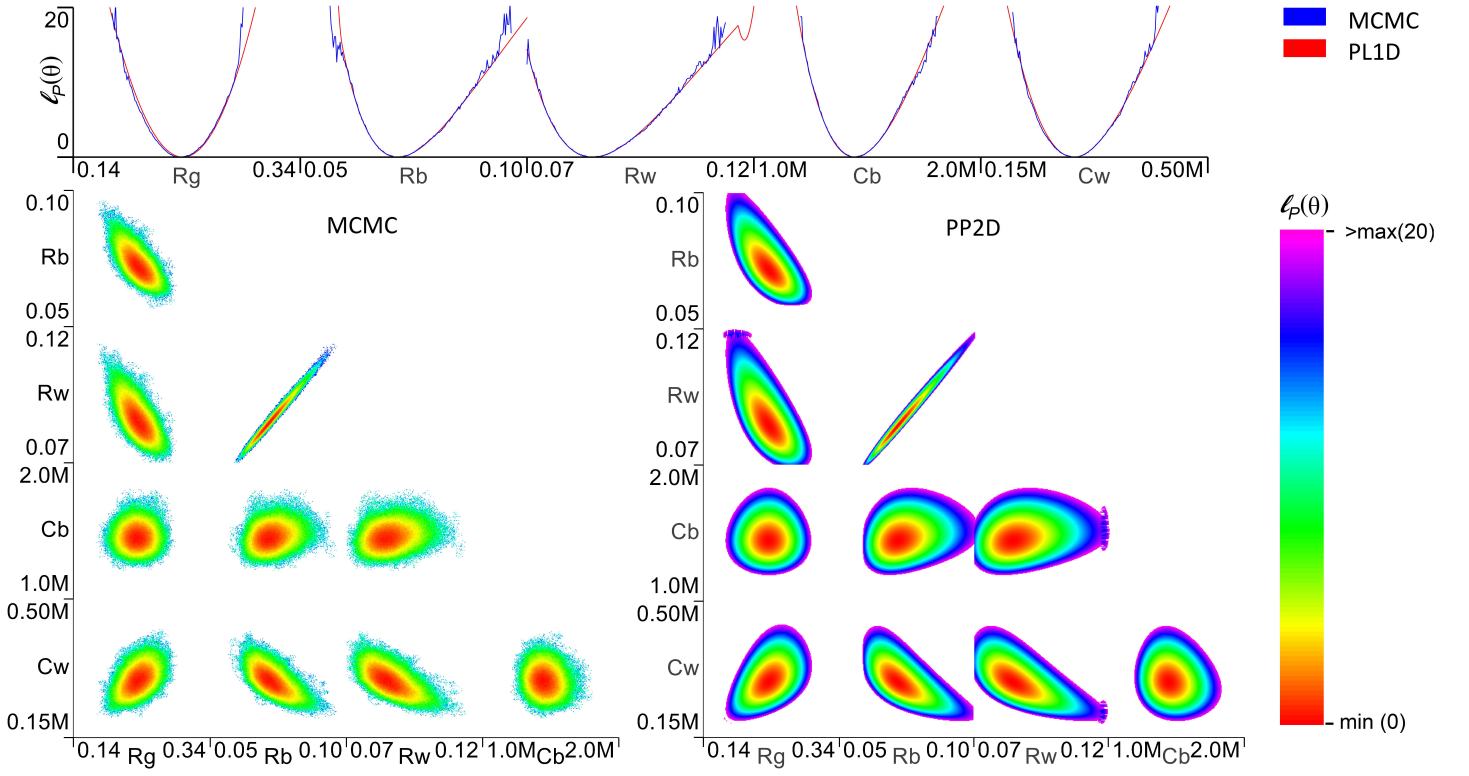
17

Figure 6: Marginal and projected posterior for Case 2, with left and right panels showing the MCMC and PP2D results, respectively. A comparison between the 1D marginal distributions (blue) and the PP1D projections (red) is shown in the top panel.

autocorrelation function (ACF) plots show significant correlation even at high lag values for all parameters *except* C_b. This is indicative of MCMC chains that are "clumpy" [27, 29], as a consequence of failure to converge. Despite the convergence failure, comparison with PP1D and PP2D projected posteriors suggests that the MCMC results are representative of the posterior, although with reduced accuracy.

The resulting marginal distributions and projections of the posterior hyper-surface of Case 2, *reshaped* by the addition of the prior of R_g, i.e., $p\left(R_\text{g}\right) = \mathcal{N}\left(0.24,\, 0.01^2\right)$, are presented in Fig. 6. All five parameters are now indicated as identifiable by *bounded* marginal and projected posterior distributions. Despite the different theoretical foundation of the methods discussed in Section 2.6, the marginal and projected posteriors are nearly identical once shifted in log space. The similarity is much stronger than for Case 1, since the challenging equipotential regions in the posterior hyper-surface have been eliminated. As evident from the marginal and projected 2D posterior of parameters R_b and R_w, there is still a strong correlation between them, but there is now a well-defined optimum.

The differences between the use of a *stochastic* rather than *deterministic* exploration of the parameter space is most pronounced in the low posterior density regions. Since the high-density regions are the primary area of interest for these analyses, the somewhat random exploration of the low posterior density regions of MCMC is of no practical consequence. Hence, both methods arguably provide the *same* insights of the parameter space of Case 2. Since the parameters now have well-defined optima, convergence of the MCMC chains occurs well within the predetermined burn-in phase, hence shorter chain lengths where required for Case 2.

The results for Case 3 are presented in Fig. 7. As shown by the marginal and projected posteriors, the posterior hyper-surface has been further reshaped by the removal of R_g. As for Case 2, all parameters have bounded profiles and are therefore identifiable. However, the *uncertainty*, i.e., the span of the log posterior projections and marginal distributions, is greatly reduced for Case 3. Hence a reduction in the region of interest Θ is required, compared to Case 2, as shown in the ranges of the plots in Fig. 7.

Next, observe that the similarity between the marginal and projected posteriors is stronger for Case 3 compared with
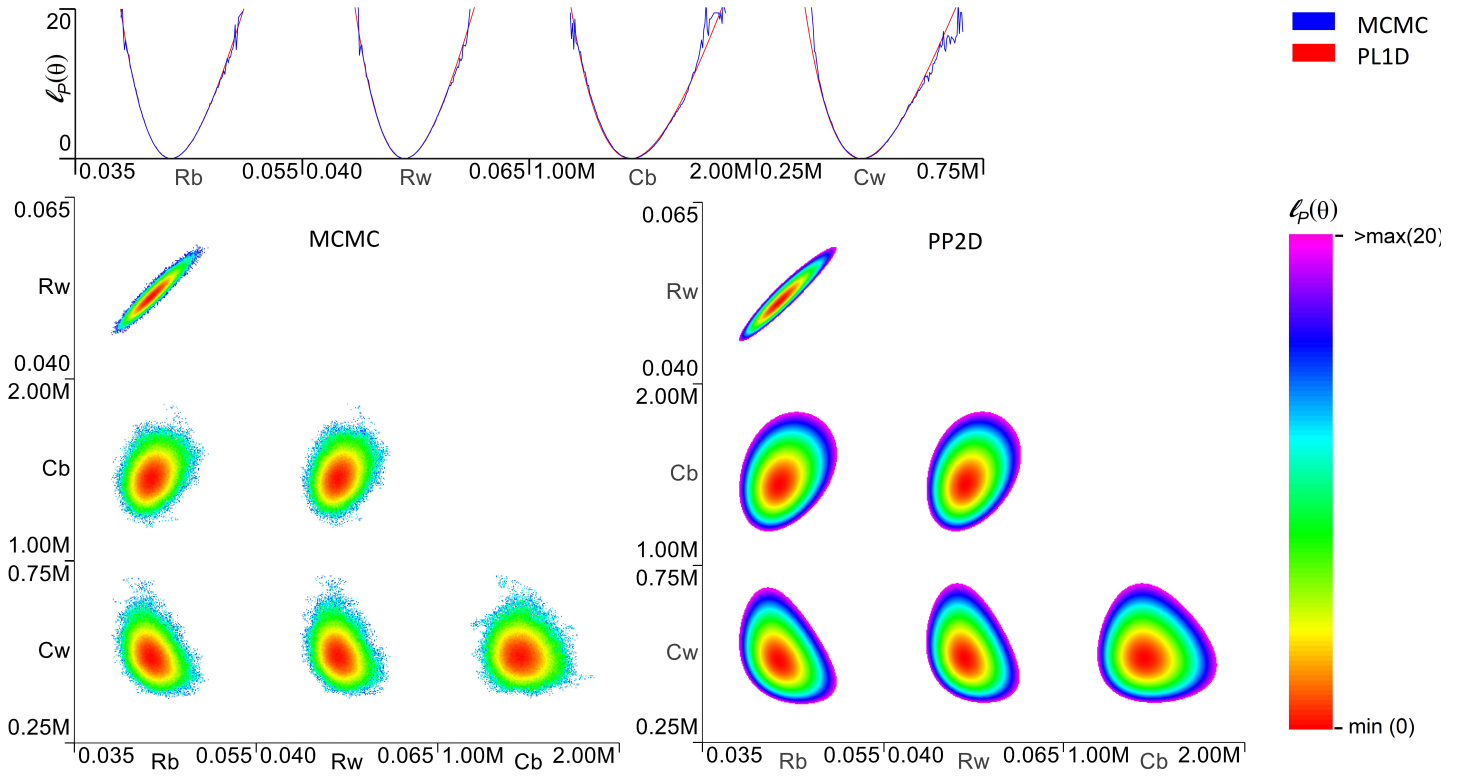
Figure 7: Marginal and projected posterior for Case 3, with left and right panels showing the MCMC and PP2D results, respectively. A comparison between the 1D marginal distributions (blue) and the PP1D projections (red) is shown in the top panel.
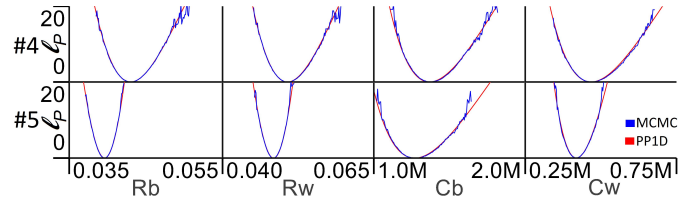

Figure 8: 1D marginal distributions (blue) and the PP1D projections (red) for Cases 4 (top) and 5 (bottom).

Case 2, although there are still some minor differences in the low posterior density regions due to the stochastic exploration of MCMC. These results further confirm that the two methods produce results that are *proportional* for the reshaped hyper-surface of Case 3. Hence, both methods provide the same insight into the parameter estimation problem.

The 1D marginal and projected posterior results for Cases 4 and 5 are shown in Fig. 8. Since the model structure and prior configuration are the same as for Case 3, the structural identifiability and parameter inter-dependency are also the same. The results for Case 4 are nearly identical to Case 3, but Case 5 obtains slightly different MAP estimates and uncertainties, since a different dataset is used for the parameter estimation.

Noise parameters

When calibrating grey-box thermal network models for the purpose of using the estimated parameters to classify building thermal behaviour, naturally, the thermal resistance and capacitance parameters are of primary interest. However, in this paper, the parameters of the noise covariance matrices $\mathcal{W}$ and $\mathcal{V}$ are also estimated from the data. Hence, it is interesting to study the identifiability of the noise parameters; the square root of the diagonal elements of each noise covariance matrix.

The PP1D projected profile and the MCMC marginal 1D posterior for noise parameters w_{b}, w_{w}, v_{b} and v_{w} for all five cases are presented in Fig. 9. First, observe that the noise parameters for Cases 1, 2 and 3 are nearly identical, despite some of the *thermal* parameters of Case 1 being non-identifiable. Observe also that the projections/marginal distributions are quite similar, indicating that similar information is obtained by both PL/PP and MCMC methods also for the noise parameters.
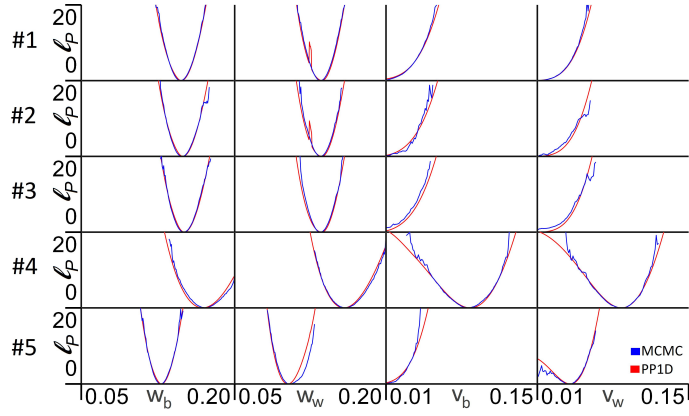
19

Figure 9: 1D marginal distributions (blue) and the PP1D projections (red) for the noise parameters w_b, w_w, v_b and v_w for all five cases, presented in increasing order.

Table 3: MAP parameters with normalised standard deviations computed with the Hessian method, together with normality test results from using Zero-Crossing (ZC) and Kolmogorov-Smirnov (KS) on residuals using the $\hat{\theta}_\mathrm{MAP}$ estimate.

#		R_g	R_b	R_w	C_b	C_w	w_b	w_w	v_b	v_w	Output	T_b	T_w
1	$\hat{\theta}_\mathrm{MAP}$	(0.101)	(0.515)	(0.607)	1.449	(0.041)	0.148	0.137	(0.000)	(0.004)	*ZC*	247	253
	$\frac{\sigma}{\hat{\theta}_\mathrm{MAP}}$	×	×	×	4.6%	×	3.2%	3.6%	×	×	*KS*	0.035	0.049
2	$\hat{\theta}_\mathrm{MAP}$	0.236	0.072	0.084	1.444	0.293	0.149	0.137	(0.003)	(0.002)	*ZC*	243	253
	$\frac{\sigma}{\hat{\theta}_\mathrm{MAP}}$	6.0%	5.7%	5.7%	4.7%	8.8%	3.2%	3.7%	×	×	*KS*	0.039	0.050
3	$\hat{\theta}_\mathrm{MAP}$	n/a	0.043	0.051	1.446	0.481	0.151	0.136	(0.010)	(0.010)	*ZC*	243	253
	$\frac{\sigma}{\hat{\theta}_\mathrm{MAP}}$	n/a	2.6%	2.7%	4.9%	7.1%	3.2%	4.9%	×	×	*KS*	0.038	0.050
4	$\hat{\theta}_\mathrm{MAP}$	n/a	0.043	0.051	1.369	0.486	0.169	0.158	0.088	0.088	*ZC*	246	259
	$\frac{\sigma}{\hat{\theta}_\mathrm{MAP}}$	n/a	2.7%	2.8%	5.3%	8.0%	5.7%	5.5%	11.5%	10.7%	*KS*	0.030	0.027
5	$\hat{\theta}_\mathrm{MAP}$	n/a	0.040	0.048	1.270	0.419	0.128	0.103	(0.002)	0.041	*ZC*	247	259
	$\frac{\sigma}{\hat{\theta}_\mathrm{MAP}}$	n/a	1.4%	1.5%	6.3%	4.5%	3.2%	5.2%	×	18.0%	*KS*	0.052	0.041

Next, observe that the profiles for the measurement noise parameters in $\mathcal{V}$ are *unbounded* towards the minimum of Θ. As discussed in Section 2.4, the noise parameters influence the likelihood *through* the computed *Kalman gain*. Since the Kalman filter estimated state trajectory is optimal when both $\mathcal{W}$ and $\mathcal{V}$ are correctly estimated [48], the values of all four noise parameters are structurally identifiable. However, if the optimal estimate of the measurement noise $\mathcal{V}$ is much smaller than the process noise $\mathcal{W}$, the Kalman gain approaches the inverse of the measurement matrix, i.e. $K \to \tilde{C}^{-1}$. Hence, the updated state depends almost exclusively on the measurement, such that $\hat{x}_{k|k} \approx \tilde{C}^{-1} y_k$. If the model uncertainty is indeed much larger than the measurement uncertainty, relying exclusively on measurements to update the state trajectory is arguably reasonable. However, this results in *practically* non-identifiable measurement noise parameters, since estimating lower values for the elements of $\mathcal{V}$ only drives K slightly closer to $\tilde{C}^{-1}$, and therefore only produces an upper bound on v_b and v_w. This effect can be observed in Fig. 9 for both measurements in Cases 1, 2 and 3, and for measurement T_b in Case 5. For Case 4, with the addition of artificial noise, the measurement noise parameters are both structurally and practically identifiable with well-defined minima and bounded CIs.

4.2. MAP point estimates with uncertainty

Since the posterior hyper-surface for most parameters and experimental cases is known to be asymptotically Gaussian [17], the uncertainty of the MAP estimate for identifiable parameters can be estimated by the Hessian from Eq. (20), i.e., the *covariance* of the MAP estimate is $\Sigma_{\theta_\mathrm{MAP}} = 2H_P^{-1}$. The $\hat{\theta}_\mathrm{MAP}$ estimate and estimated standard deviation $\sigma_i = \sqrt{\Sigma_{i,i}}$ are shown, together with normality test results of the $\hat{\theta}_\mathrm{MAP}$ estimate, in Table 3. The parameter values enclosed in $(\cdot)$ are

ambiguous, previously diagnosed as non-identifiable, and used only to test the residuals for normality. The corresponding uncertainty estimates are noted as ×. The elements denoted as n/a are not relevant due to use of the R2C2 model. The standard deviation σ is normalised over the $\hat{\theta}_{\mathrm{MAP}}$ estimate to facilitate comparison of different parameters. The residual for the $\hat{\theta}_{\mathrm{MAP}}$ estimate for all five cases passes both the Zero-Crossing (ZC) test with acceptance range $\langle 219, 262 \rangle$ and the Kolmogorov-Smirnov (KS) test with threshold (< 0.062), both at confidence $\alpha = 95\%$. The three resistance parameters are given in unit $[K/W]$, the two capacitances in unit $10^6 \, [J/K]$ and the four noise parameters in unit $[K]$.

There are several interesting observations to be made from Table 3. First, comparing Case 2 and 3, lumping R_{g} into the remaining resistance parameters R_{b} and R_{w} results in correspondingly *decreased* estimates of thermal resistance. Next, observe that the MAP estimate of C_{b}, and the corresponding uncertainty, is approximately the same for all five cases, although slightly lower for Case 5. Note also that the model prediction uncertainty parameters w_{b} and w_{w} are nearly identical for the first three cases. These observations indicate at least some correlation of the estimated parameters to the physical properties of the building, which is further discussed in Section 4.4.

Further, observe that although the inclusion of a prior on R_{g} in Case 2 produced unambiguous MAP estimates, the uncertainty of the remaining estimated parameters is significantly lower in Case 3, where $R_{\mathrm{g}} = \infty$.

Comparing Cases 3 and 4, the uncertainty in the four thermal parameters is not significantly affected by the addition of artificial measurement noise in Case 4. This comparison indicates that a slight increase in measurement noise does not adversely affect the parameter estimation uncertainty. However, the use of a different dataset in Case 5 significantly *reduces* the uncertainty of the thermal parameters. The important factor determining the uncertainty of the estimated parameters is the *dynamic information* content related to the system behaviour contained in the data, assuming a reasonable signal to noise ratio. Finally, observe that the MAP estimates of the four thermal parameters are in reasonable agreement for Cases 3, 4 and 5, which indicates that the estimated parameters are consistent irrespective of the data-set used for calibration, at least to some degree considering the datasets where recorded consecutively.

4.3. Posterior predictive distribution for the Test dataset

Figure 10 shows the posterior predictive distributions discussed in Section 2.7, for each experimental case, computed by *Monte Carlo (MC) simulation* of the model in Eq. (14) for a *thinned* subset of the parameter samples in $\theta_{[K]}$. To reduce computation time, a thinning factor of 100 is used for this computation. The plots are created by repeatedly simulating the test-set ballistically, with randomly generated measurement and process noise v_k and w_k, thus creating one simulated trajectory for each $\theta \in \theta_{[K]}$. The lower right plot shows the standard deviation σ_k of at each time step over the set of K ballistic simulations such that

$$\sigma_k^2 = \mathbb{V}\left[\hat{y}_{k|0}\right] = \frac{1}{K-1} \sum_{i=1}^{K} \left(\hat{y}_{k|0}^{(i)} - \bar{y}_{k|0}\right)^2$$

where $\bar{y}_{k|0} = \mathbb{E}\left(\hat{y}_{k|0}\right) = \frac{1}{K} \sum_{i=1}^{K} \hat{y}_{k|0}^{(i)}$. Note that the test-set *measurements* of future system inputs u_k are used to compute the posterior predictive distributions in order to separate the uncertainties of the model predictions with those introduced by using more realistic *predicted* system inputs.

First, observe from Fig. 10 that Cases 1, 2 and 3 produce *similar* prediction results, despite the differences in model structure and parameter posterior distributions, and from the lower right panel showing the standard deviation (SD) that the empirical SD is nearly identical for the first three cases. Comparing cases 4 and 5 to Case 3 shows that the SD of the predictions is *increased* for Case 4 but *decreased* for Case 5. Since Case 4 has artificially added measurement noise, it is expected that the output predictions will have increased uncertainty, due to larger values of the generated measurement noise parameters v_{b} and v_{w}. For Case 5, the variance of the output trajectories is reduced, since the estimated parameters in $\theta_{[K]}$ have less variation due to improved dynamic information in the *Training2* dataset. The similarity of the model predictions for each case is further demonstrated by the root mean square error (RMSE) of the MAP predictions shown in Fig. 10. The
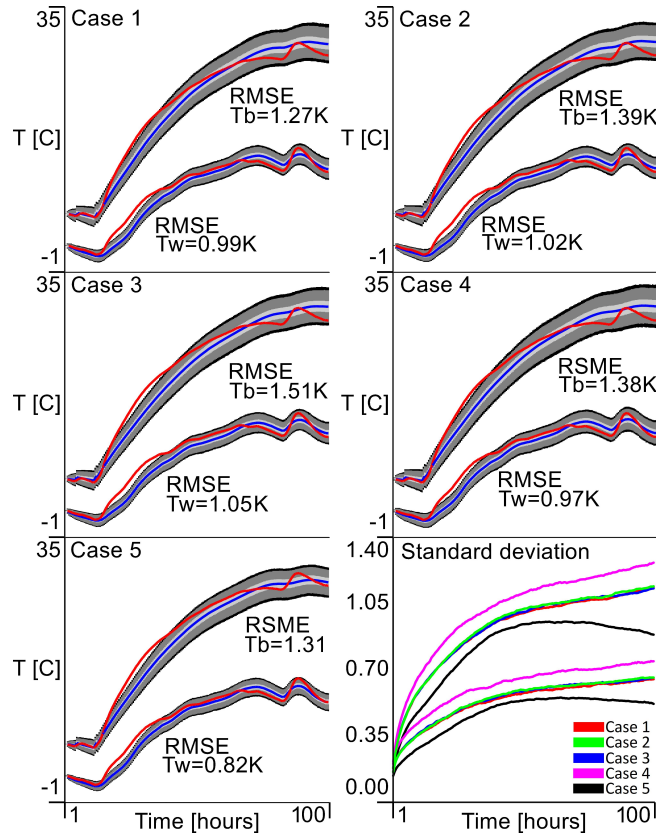
Figure 10: Prediction posterior from repeated ballistic simulations over $\theta_{[K]}$ of the Test dataset for all five cases. For each case, the temperatures T_b and T_w are plotted as upper/lower temperature, respectively. The plots show the MAP estimate (blue), credibility bands of 50%, 95% and 99% (light to dark) and the reference measurements (red). The lower right plot shows the standard deviation of the outputs, computed over all K trajectories, at each time step.

observed differences in predictions over the four-day horizon are likely to be of no practical significance for use in a predictive controller.

The similarity of the posterior predictive distributions of Cases 1, 2 and 3 shows that all three variations of the model are in fact able to *learn* the information in the training set necessary to *predict* the test set. The fact that Case 1 has some non-identifiable parameters with a significant equipotential region in their posterior distributions does *not* prevent the model from successfully predicting the output. Comparing Cases 4 and 5 to Case 3 further indicates that it is the dynamic information content in the training data that most significantly affects the posterior predictive distributions, as long as the model structure is sufficiently complex to learn the appropriate system behaviour.

These results show that the presented grey-box model may *adequately predict* the system behaviour, even if the parameters are not *unambiguously* identifiable, and that the prediction accuracy largely depends on the information content in the training data. For black-box models, there is usually no assumption of physical interpretability of the model coefficients, hence unambiguous optimal parameter estimates are of no consequence. Methods such as system identification [11] and Artificial Neural Networks (ANN) [54, 55] typically produce non-unique system description models whose ability to predict future data, assuming adequate model complexity, depends mostly on the information content in the training data.

By including the stochastic process and measurement noise terms in the model and learning their parameters from data, the model predictions can also reflect these important uncertainties. The computation of a posterior predictive distribution, rather than a single MAP or MLE trajectory, could *facilitate* use of *stochastic* MPC methods [56]. Calculating the *simulation* RMSE to a worst case error of $1.5K$, computed over a reasonably long prediction horizon of four days is likely sufficient for the purposes of model based control [3].
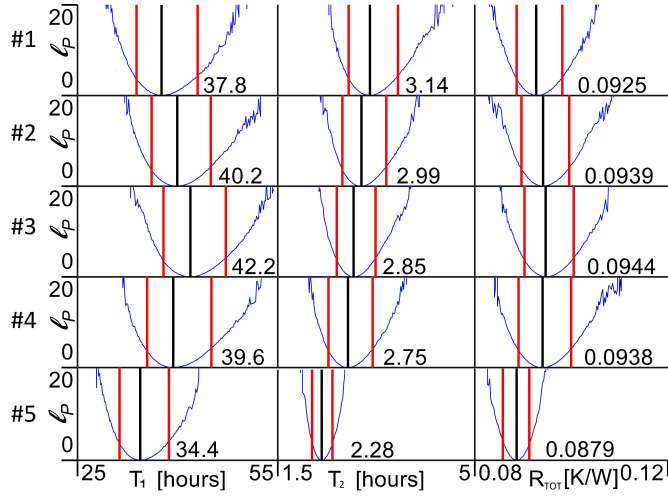
22

Figure 11: Log posterior marginal distributions with credibility 2.5 and 97.5 percentiles (red) MAP estimates (black) of the two time-constants, T_1 and T_2, and the total heat-loss resistance R_{TOT} of the building.

4.4. Physical interpretation of estimated parameters

Given that all five cases show similar predictive capabilities, despite Case 1 having some non-identifiable parameters and Table 3 showing large variation in the model parameters, it is interesting to consider if there are any similarities between the cases that are *not* expressed in the model parameters. Three properties of interest are the eigenvalues λ_1 and λ_2 of A, or rather their negative inverse, i.e., time-constants $T_1 = -\lambda_1^{-1}$ and $T_2 = -\lambda_2^{-1}$, and the total resistance to heat-loss $R_{\mathrm{TOT}} = R_{\mathrm{g}} || (R_{\mathrm{b}} + R_{\mathrm{w}})$ between the indoor temperature T_{b} and the outdoor temperature T_∞, where $||$ indicates a *parallel* connection of resistors, i.e., a *harmonic* sum. Note that for Cases 3, 4 and 5, $R_{\mathrm{g}} = \infty \rightarrow R_{\mathrm{TOT}} = R_{\mathrm{b}} + R_{\mathrm{w}}$.

Since the posterior distribution generated by MCMC is represented by a set of samples $\theta_{[K]}$, the quantities T_1, T_2 and R_{TOT} can be computed for each sample in $\theta_{[K]}$ and the marginal posterior distribution for each quantity computed by histogram. The marginal log posterior $\ell_{\mathrm{P}}(\theta) = -2 \ln p\left(\theta | y_{[N]}\right)$ for each quantity is given for all five cases in Fig. 11, together with the MAP estimate and $2.5/97.5$ credibility percentiles, computed from interpolation on the cumulative empirical distribution.

First, observe that these quantities have bounded profiles with a well-defined optima, also for Case 1. Even though the MCMC method's trace plots in Fig. 5 show a large variation in the model parameters, the time-constants and total resistance are well-defined. When MCMC proposes a new sample θ_k^c, if that sample gives time-constants T_1 and T_2, or a total thermal resistance R_{TOT} that differs *substantially* from the MAP estimates shown in Fig. 11, the resulting log posterior $\ell_{\mathrm{P}}(\theta)$ would produce a very low acceptance probability α.

Next, observe that all three quantities are in reasonable agreement for the first four cases, with the MAP estimate for each case falling within the credibility limits for all the other cases, except the MAP of T_1 for Case 1 falling just below the 2.5% credibility limit of Case 3. The similarity, despite using different model structures, priors and noise on training data, indicates that the time-constants and total resistance are *somewhat* invariant to the experimental setup. The consistency of $R_{\mathrm{TOT}} \sim 0.094$ for Cases 1 to 4 explains why there is a strong correlation between R_{b} and R_{w} as illustrated in Section 4.1. Given that R_g is determined by the prior, omitted, or has a well-defined optimum, the values of R_{b} and R_{w} must fulfil $R_{\mathrm{TOT}} \sim 0.094$. However, it is difficult to see any *physical* reason for this correlation. The interpretation of the individual R_{b} and R_{w} parameters as physical properties of the building is therefore questionable.

Finally, note from the marginal posterior distributions in Fig. 11 that the differences between Case 5 and Cases 1 through 4 are *significant*, e.g. a value of $R_{\mathrm{TOT}} = 0.094$ would give a very low posterior probability based on the distribution of Case 5. The Case 5 MAP estimate for each quantity is outside the 2.5/97.5 percentiles for Cases 1 to 4. While the results of Case 5 *cognitively appear* similar to Cases 1 to 4, the presented results are not similar *enough* to conclude with sufficient statistical

credibility that the parameters are consistent also when different calibration data is used.

4.5. Comparing MCMC and PP methods

As the results from the five experimental cases have shown, the MCMC and PL/PP methods indeed provide similar results. Even though the theoretical foundation of the methods differ significantly, in particular in the stochastic vs. deterministic exploration of Θ and the use of projection vs. marginalisation to present results, both methods produce estimates of the same posterior parameter distribution. Note, however, the projected and marginalised posterior is *not* always proportional as shown in Case 1. The advantage of the projection-based methods is mainly that they are not affected by flat regions in the likelihood function or posterior distribution. Additionally, the deterministic projections of each prescribed point in $\Theta_{i,j}$ allow an *exhaustive* exploration of the feasible region Θ.The method will therefore obtain both global and local minima, including any equipotential manifolds. The main advantages of MCMC are computational speed and the way that the target distribution is represented as a set of samples $\theta_{[K]}$ that can be used for further analysis, such as computing derived parameters, e.g. time constants or total thermal resistance.

Computation time

Deterministic brute force exploration is, naturally, quite time-consuming. The accuracy at which a global optimum can be found depends on the resolution of the parameter discretisation used in the brute force grid exploration. Hence, the key to successful use of the PL2D/PP2D methods is a reasonable compromise between computation time and resolution. The computational burden is further exacerbated by the need to project the log likelihood or log posterior to all parameter combinations $\Theta_{i,j}$ of interest. In contrast, the MCMC method is specifically designed to explore the most interesting areas, i.e., the areas of Θ with the highest posterior density. Additionally, since the resulting 2D distributions are computed from histograms by marginalising out the other parameters, there is no need to run the method multiple times. All computation times discussed here are given for the method configurations stated in Section 3.3. All the methods discussed in this paper can easily be parallelised, and can thus take advantage of modern multi-core CPU architectures.

The MCMC method computes in around $\sim 12min$ for Cases 2 through 5.Due to non-identifiable parameters, Case 1 required around 20 times long chains to produce a reasonable approximation of the posterior, thus taking a computation time of $\sim 4.5h$. In contrast, the PP2D method takes around $\sim 6.5h$ to compute all ten projections for Cases 1 and 2. The PP2D method is not affected by the shape of the posterior surface. Case 1 and 2 therefore takes approximately the same time to compute. The six projections in Cases 3, 4 and 5 are computed in around $\sim 2.5h$. The reduction in computation time is due to a lower number of parameters in the last three cases, resulting in fewer projection planes $\Theta_{i,j}$ and also fewer free parameters to optimise for each projected point. The PL1D/PP1D methods, requiring discretisation only of single parameters and only one projection per parameter, are significantly faster at around $\sim 3min$. Note that if only 1D posterior distributions were of interest, the MCMC method could likely have been configured with significantly shorter chains.

Predictive posterior and combined parameter distributions

A distinct advantage of MCMC is the ability to compute posterior predictive distributions for the model output. By using MC simulations over the set $\theta_{[K]}$, a set of K independent state and output trajectories can be computed. The uncertainty of the predictions, given both model uncertainty $w_k \sim \mathcal{N}(0, \mathcal{W})$, measurement uncertainty $v_k \sim \mathcal{N}(0, \mathcal{V})$ and the uncertainty in the parameter estimates as expressed in $\theta_{[K]}$, can be estimated for each time-step over the K trajectories, as discussed in Section 2.7.

Another use of the sampled set $\theta_{[K]}$ is the possibility to compute combined parameters, such as the eigenvalues of A or the total resistance to heat-loss R_{TOT} discussed in Section 4.4. Marginal distributions for these combined parameters can then be computed and analysed to provide a more flexible analysis and deeper insight into the model's behaviour.

24

5. Conclusion

In this paper, both *frequentist* and *Bayesian* frameworks for parameter estimation were used to obtain a detailed analysis of the parameter space of a grey-box thermal network model for a building [21, 43]. The Profile Likelihood (PL), the Profile Posterior (PP) and the Markov Chain Monte Carlo (MCMC) methods were used to estimate the *shape* of the posterior distribution for the parameters of a thermal network grey-box model expressed as a stochastic differential equation (SDE) [16].

Five experimental cases were investigated, one of which has non-identifiable parameters. This non-identifiability was shown to be resolved by application of either a *prior distribution* for the parameter R_g, or by the *removal* of R_g from the model, in Case 2 and 3, respectively. Cases 4 and 5 showed how, and under what conditions, the covariance of the process uncertainty w_k and the measurement uncertainty v_k can be estimated from data for the given model. By using the sampled set $\theta_{[K]}$ from MCMC, the eigenvalues, and subsequently the time-constants, of the state transition model A, and also the total thermal resistance R_{TOT}, were shown to have well-defined *bounded* distributions even for the non-identifiable Case 1. The estimates of the time-constants and total thermal resistance were found to be similar for the first four cases, but with significant differences in Case 5, which used a different training dataset.

A distinct advantage of the MCMC method is the ability to use the sampled set of parameters $\theta_{[K]}$ to propagate the uncertainty of the parameter estimates into the model output predictions, by computing the *posterior predictive distribution*. The resulting distributions for all five cases were found to be in reasonable agreement. Hence, all the models are found able to *learn* the necessary *knowledge* about the physical building from the training data necessary to predict the independent test set. This result indicates that while *parameter identifiability* is important for justifying a physical interpretation of the model parameters [5], the presented model's ability to predict system behaviour is not significantly affected by non-identifiable parameters. This result is well-know from the black-box modelling paradigm [11]. Since grey-box models explicitly applies prior physical knowledge of the system to create a model structure, the interpretation of parameters as physical constants of the system is often assumed [5]. The results presented here show that, even if the model correctly predicts the system behaviour, assumptions of physical interpretation of parameters should be supported by an identifiability analysis.

Finally, the use of both PP and MCMC methods to explore the posterior distribution shows that the *shapes* of the respectively resulting *projected* and *marginal* distributions are near identical in log space, i.e., proportional, and therefore convey the same diagnostic information about the parameter space for most of the presented cases [21]. The main advantage of the projection methods, due to the deterministic exploration of the parameter space, is that the equipotential manifolds in the log posterior space caused by non-identifiable parameters do not affect the method's ability to obtain projections of the posterior [21]. The MCMC method's main advantages are computational efficiency, achieved by focusing exploration of the parameter space on regions of high posterior density, and also the possibility of utilising the sampled set of parameters $\theta_{[K]}$ to compute the posterior predictive distribution and marginal distributions for other parameters derived from the sampled θ [27, 29]. Producing a stochastic forecast for the temperatures in the building could facilitate use of stochastic Model Predictive Control (MPC) [56, 57], which also accounts for uncertainty in the calibrated model parameters.

References

[1] D. Perera, C. F. Pfeiffer, N.-O. Skeie, Modelling the heat dynamics of a residential building unit: Application to Norwegian buildings, Modeling, Identification and Control 35 (1) (2014) 43–57. doi:10.4173/mic.2014.1.4.

[2] Recast, EPBD, Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), Official Journal of the European Union 18 (06) (2010) 2010.

[3] M. Killian, M. Kozek, Ten questions concerning model predictive control for energy efficient buildings, Building and Environment 105 (2016) 403–412.

[4] S. F. Fux, A. Ashouri, M. J. Benz, L. Guzzella, EKF based self-adaptive thermal model for a passive house, Energy and Buildings 68 (2014) 811–817.

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

[6] P. Bacher, H. Madsen, Identifying suitable models for the heat dynamics of buildings, Energy and Buildings 43 (7) (2011) 1511 – 1522. doi:https://doi.org/10.1016/j.enbuild.2011.02.005.

[7] T. Berthou, P. Stabat, R. Salvazet, D. Marchio, Development and validation of a gray box model to predict thermal behavior of occupied office buildings, Energy and Buildings 74 (2014) 91–100.

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

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

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

[11] L. Ljung, System Identification - Theory for the User, Prentice Hall information and system sciences series, Prentice Hall PTR, 1999.

[12] L. Ljung, Prediction error estimation methods, Circuits, Systems and Signal Processing 21 (1) (2002) 11–21. doi:10.1007/BF01211648.

[13] R. Johansson, System Modeling and Identification, Information and system sciences series, Prentice Hall, 1993.

[14] D. Di Ruscio, Combined Deterministic and Stochastic System Identification and Realization: DSR - A Subspace Approach Based on Observations, Modeling, Identification and Control 17 (3) (1996) 193–230. doi:10.4173/mic.1996.3.3.

[15] R. Ergon, D. Di Ruscio, Dynamic system calibration by system identification methods, in: European Control Conference (ECC), 1997, IEEE, 1997, pp. 1556–1561.

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

[17] H. Madsen, Time series analysis, Chapman and Hall/CRC, 2007.

[18] R. Juhl, J. K. Møller, H. Madsen, ctsmr-Continuous Time Stochastic Modeling in R, arXiv preprint arXiv:1606.00242.

[19] R. Juhl, J. K. Møller, J. B. Jørgensen, H. Madsen, Modeling and prediction using stochastic differential equations, in: Prediction Methods for Blood Glucose Concentration, Springer, 2016, pp. 183–209.

[20] O. M. Brastein, R. Sharma, N.-O. Skeie, Sensor placement and parameter identifiability in grey-box models of building thermal behavior, in: Proceedings of The 60th Conference on Simulation and Modelling (SIMS 60), 13-16 August 2019, Västerås, Sweeden, Linköping University Electronic Press, 2009, p. tbd.

[21] A. Raue, C. Kreutz, F. J. Theis, J. Timmer, Joining forces of bayesian and frequentist methodology: a study for inference in the presence of non-identifiability, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371 (1984) (2013) 20110544.

[22] N. R. Kristensen, H. Madsen, Continuous time stochastic modelling, Mathematics Guide (2003) 1–32.

[23] T. Bohlin, S. F. Graebe, Issues in nonlinear stochastic grey box identification, International journal of adaptive control and signal processing 9 (6) (1995) 465–490.

[24] S. A. Murphy, A. W. Van der Vaart, On profile likelihood, Journal of the American Statistical Association 95 (450) (2000) 449–465.

[25] A. P. Dawid, Conditional independence in statistical theory, Journal of the Royal Statistical Society: Series B (Methodological) 41 (1) (1979) 1–15.

[26] H. Pohjanpalo, System identifiability based on the power series expansion of the solution, Mathematical Biosciences 41 (1) (1978) 21–33. doi:https://doi.org/10.1016/0025-5564(78)90063-9.

[27] J. Kruschke, Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, Academic Press, 2014.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical recipes in C++, Vol. 3, Cambridge University Press, 2007.

[29] C. Bishop, Pattern Recognition and Machine Learning: All "just the Facts 101" Material, Information science and statistics, Springer, 2013.

[30] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, The journal of chemical physics 21 (6) (1953) 1087–1092.

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

[32] S. Rouchier, M. Rabouille, P. Oberlé, Calibration of simplified building energy models for parameter estimation and forecasting: Stochastic versus deterministic modelling, Building and Environment 134 (2018) 181–190.

[33] O. M. Brastein, B. Lie, C. F. Pfeiffer, N.-O. Skeie, Estimating uncertainty of model parameters obtained using numerical optimisation, Modeling, Identification and Control 40 (4) (2019) 213–243. doi:10.4173/mic.2019.4.3.

[34] O. Brastein, D. Perera, C. Pfeiffer, N.-O. Skeie, Parameter estimation for grey-box models of building thermal behaviour, Energy and Buildings 169 (2018) 58 – 68. doi:https://doi.org/10.1016/j.enbuild.2018.03.057.

[35] H. Madsen, J. Holst, E. Lindström, Modelling non-linear and non-stationary time series, Lecture Notes, Technical University of Denmark, Dpt. of Informatics and Mathematical Modeling, Kgs. Lyngby, Denmark.

[36] M. A. S. Perera, B. Lie, C. F. Pfeiffer, Structural Observability Analysis of Large Scale Systems Using Modelica and Python, Modeling, Identification and Control 36 (1) (2015) 53–65. doi:10.4173/mic.2015.1.4.

[37] G. C. Goodwin, R. L. Payne, Dynamic system identification. Experiment design and data analysis., 1977.

[38] A. H. Jazwinski, Stochastic processes and filtering theory, Dover Publications, Inc, 1970.

[39] I. Goodman, R. Mahler, H. Nguyen, Mathematics of Data Fusion, Theory and Decision Library B, Springer Netherlands, 2013.

[40] J. Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 236 (767) (1937) 333–380.

[41] S. Kullback, A Note on Neyman's Theory of Statistical Estimation, The Annals of Mathematical Statistics 10 (4) (1939) 388–390.

[42] A. E. Gelfand, S. K. Sahu, Identifiability, improper priors, and gibbs sampling for generalized linear models, Journal of the American Statistical Association 94 (445) (1999) 247–253.

[43] M. J. Bayarri, J. O. Berger, The interplay of bayesian and frequentist analysis, Statistical Science (2004) 58–80.

[44] O. M. Brastein, B. Lie, R. Sharma, N.-O. Skeie, Parameter estimation for externally simulated thermal network models, Energy and Buildingsdoi:https://doi.org/10.1016/j.enbuild.2019.03.018.

[45] S. Rouchier, Solving inverse problems in building physics: An overview of guidelines for a careful and optimal use of data, Energy and Buildings 166 (2018) 178–195.

[46] A. N. Tikhonov, A. Goncharsky, V. Stepanov, A. G. Yagola, Numerical methods for the solution of ill-posed problems, Vol. 328, Springer Science & Business Media, 2013.

[47] C. Van Loan, Computing integrals involving the matrix exponential, IEEE transactions on automatic control 23 (3) (1978) 395–404.

[48] D. Simon, Optimal state estimation: Kalman, H infinity, and nonlinear approaches, John Wiley & Sons, 2006.

[49] M. Bocquet, K. S. Gurumoorthy, A. Apte, A. Carrassi, C. Grudzien, C. K. Jones, Degenerate kalman filter error covariances and their convergence onto the unstable subspace, SIAM/ASA Journal on Uncertainty Quantification 5 (1) (2017) 304–333.

[50] T. J. Rothenberg, et al., Identification in parametric models, Econometrica 39 (3) (1971) 577–591.

[51] D. Venzon, S. Moolgavkar, A method for computing profile-likelihood-based confidence intervals, Applied statistics (1988) 87–94.

[52] S. S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, The Annals of Mathematical Statistics 9 (1) (1938) 60–62.

[53] N. Cressie, Statistics for spatial data, Vol. 4, Wiley Online Library, 1992.

[54] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, Vol. 1, MIT press Cambridge, 2016.

[55] M. Kuhn, K. Johnson, Applied predictive modeling, Vol. 26, Springer, 2013.

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

[57] T. A. N. Heirung, J. A. Paulson, J. OLeary, A. Mesbah, Stochastic model predictive controlhow does it work?, Computers & Chemical Engineering 114 (2018) 158–170.

—

usn.no