Models for the energy performance of low-energy houses

Andersen, Philip Hvidthøft Delff; Madsen, Henrik; Rode, Carsten

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):

DTU Library
Technical Information Center of Denmark

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Models for the energy performance of low-energy houses

Philip Delff Andersen

Kongens Lyngby 2013
PhD-2013-312
The aim of this thesis is data-driven modeling of heat dynamics of buildings. Traditionally, thermal modeling of buildings is done using simulation tools which take information about the construction, weather data, occupancy etc. as inputs and generate deterministic energy profiles of the buildings. This approach often fails in predicting the actual heat consumption of buildings once they are constructed. The approach taken in this work is deriving models from observations collected after the construction, aiming at describing the actual characteristics of the building.

Identification of heat dynamics of buildings is needed both in order to assess energy-efficiency and to operate modern buildings economically. Energy signatures are a central tool in both energy performance assessment and decision making related to refurbishment of buildings. Also for operation of modern buildings with installations such as mechanical ventilation, floor heating, and control of the lighting effect, the heat dynamics must be taken into account. Hence, this thesis provides methods for data-driven modeling of heat dynamics of modern buildings.

While most of the work in this thesis is related to characterization of heat dynamics of buildings, the first topic analyzed is the variation of presence of occupants. As buildings get more energy-efficient, internal loads and user-behavior increasingly influence the energy consumption. Most simulation tools use deterministic occupancy profiles to simulate internal loads. However, such occupancy patterns will largely depend on the specific use of the building, and hence the profiles must be empirically based. A probabilistic method for modeling time-dependence and dynamics of presence of occupants is developed and applied
by estimation and model validation on data from an office building. The approach to modeling occupants’ presence provides a flexible method where no assumptions in the application.

The rest of the thesis deals with statistical modeling of heat dynamics of buildings. First, discrete-time models are applied. Discrete-time models are computationally relatively simple and provide a flexible framework for dynamical modeling as a natural extension of the often-used static energy-balance models. The importance of applying dynamical models, even for deriving thermostatic or steady-state properties, is stressed, and methods for doing so are outlined.

Since heat transfer is fundamentally described by partial differential equations, modeling of heat dynamics using differential equations is an obvious approach. A quasi-Gaussian maximum likelihood estimation technique, where the likelihood function is evaluated using the extended Kalman filter on state-space models, is used. In this framework – referred to as “grey-box” modeling – one-step predictions can be generated and used for model validation by testing statistically whether the model describes all variation and dynamics observed in the data. The possibility of validating the model dynamics is a great advantage from the use of stochastic differential equations compared to ordinary differential equations.

The strengths of the discrete-time and the continuous-time approach are discussed. Besides the parametrization, which is directly physically interpretable, grey-box models intrinsically provide variable prediction uncertainty, which is important in relation to design of controllers and decision making for comfort requirements. In the framework of stochastic differential equations, there are normally more parameters related to noise processes than in discrete-time models which increases the complexity of the estimation. Here, the state space formulation is often used. Since there is normally infinitely many state space representations corresponding to a transfer function model, structural identifiability is important in relation to state space modeling.

A low-energy building in Sisimiut, Greenland is used as a test-building. The building is well-insulated and features large modern energy-efficient windows and floor heating. These features lead to increased non-linear responses to solar radiation and longer time constants. The building is equipped with advanced control and measuring equipment. Experiments are designed and performed in order to identify important dynamical properties of the building, and the collected data is used for modeling.

The thesis emphasizes the statistical model building and validation needed to identify dynamical systems. It distinguishes from earlier work by focusing on modern low-energy construction and going further into studying and character-
izing the dynamical properties of the fitted models.
Sammenfatning (Danish)

Formålet med denne afhandling er databaseret modellering af bygningers varmedynamik. Traditionelt udføres termisk modellering af bygninger ved brug af simuleringsværktøjer, som inddrager information om konstruktionen, vejrdatalære, bygningernes anvendelse mm. som input og genererer deterministiske energiprofiler. Denne tilgang mislykkes ofte med at forudsige det faktiske energiforbrug, som observeres efter bygningernes opførsel. Tilgangen i foreliggende arbejde er statistisk identifikation af modeller fra observationer indsamlet efter bygningernes opførsel med den hensigt at beskrive de faktiske egenskaber for bygningerne.

Identifikation af bygningers varmedynamik er nødvendig både for vurdering af energieffektivitet og for at drifte moderne bygninger økonomisk. Energisignaturer er et centralt værktøj i både måling af energimæssig ydelse og i forbindelse med beslutningstagen ved renovering af bygninger. Også i driften af moderne bygninger med installationer som f.eks. mekanisk ventilation, gulvarme og styring af lysindfald, må varmedynamikken tages i betragtning. Derfor giver denne afhandling metoder for databaseret modellering af moderne bygninger.

brugertilstedeværelse leverer en fleksibel metode, hvor ingen antagelser er nødvendige i anvendelsen.


Styrkerne ved tilgangene i diskret og kontinuert tid diskuteres. Udobere fordele ved parametriseringen, som er direkte fysisk fortolkelig, udtrykker grey-box modeller ibo ende usikkerhed af forudsigelser, hvilket er vigtigt i forhold til design af styring under komfortkrav. I stokastiske differentialligninger er der normalt flere parameter relateret til støjprocesser end i diskret-tids modeller, hvilket øger kompleksiteten af estimationen. Eftersom der normalt er uendelig mange repræsentationer på tilstandsform svarende til en overföringsfunktion, er strukturel identificerbarhed vigtigt i relation til tilstandsmodellering.


Afhandlingen lægger vægt på statistisk modelbygning og nødvendig validering ved identifikation af dynamiske systemer. Det afviger fra tidligere arbejder ved at fokusere på moderne lavenergi byggeri and ved at gå dybere ind i studiet og karakteriseringen af de dynamiske egenskaber ved de estimerede modeller.
This thesis was prepared at DTU Compute at the Technical University of Denmark as part of fulfillment of the requirements for acquiring a PhD in statistics and dynamical systems. The work behind the thesis was carried out 2010–2013 in Denmark, in Greenland, and in Spain. The thesis deals with modeling the heat dynamics of buildings based on measurements. It also features a contribution to modeling of presence of occupants. The thesis consists of an introduction to the topic including theoretical background, a general discussion and conclusions, followed by papers that were written during the PhD studies. The papers include three research articles submitted to peer-reviewed journals (one published at the time of writing), two technical reports, and one manuscript of a paper in progress.

I would like to thank my supervisors, Henrik Madsen and Carsten Rode, for making this work possible, for support and guidance. Especially, I owe thanks to Henrik for the faith you showed in me in the five very exciting and educational years we have been working together. It has been a pleasure to learn from your passion for spreading the application and communicating the potentials of the application of statistics. Uffe Thygesen, I want to thank you for all the patience and help over the years. Your teaching is a huge inspiration, and your passion for our field is contagious.

During my PhD studies I have had many great travel experiences. I’d like to thank all the nice people I met during my field work in Sisimiut, Greenland. I would never have been able to do all the experiments without the professional and careful help from Rasmus Kruse-Nielsen. Also Egil Borchersen was always ready to help. During my external stays in Málaga I was kindly hosted at the University of Málaga by Antonio Carrillo. Thanks for many great discussions.
I spent five months in total at the Plataforma Solar, Almería. A special thanks to Maríá José Jiménez. We managed to combine our expertise and get some great work out of it. I met so many wonderful people in Almería, I can’t list you all here, but Luís Castillo, you were extraordinarily helpful and welcoming. Whenever I can return some of your help, I hope you let me know.

I would like to thank so many dear colleagues at DTU IMM and now DTU Compute for lots of inspiration. I may forget many, but at least for capturing my personal group picture for these great years, I would like to send my warmest regards to (order randomized) Marco, Anne Kathrine, Rune Juhl, Tryggvi, Rune Haubo, Jacobo, Jan Kloppenborg, Niamh, Ewa, Juanmi, Merete, Lasse, Julija, Niels, Fannar, Jan Frydendall, Anna Helga, Per, Javier, and Emil. Special thanks to my office mates Piju and Peder – great friends and always ready for a passionate discussion and always supporting. Many thanks to Hanne for always being so helpful and cheering.

Family and friends: Thank you for all the support and patience – especially in the final period of completion of this work. And Ewa, thank you for everything we share.

Philip Delff Andersen
Kgs. Lyngby, Denmark
August 2013
Acknowledgments

The author would like to thank the Technical University of Denmark, and the Arctic Technology Centre for funding this work. Also Ciemat, Spain, namely Plataforma Solar in Almería for hosting parts of the work and for fruitful collaboration.
List of publications

Papers included in the thesis


List of publications


Other publications

Journal papers


Conference papers


Andersen, Philip Delff, Carsten Rode, and Henrik Madsen (2012c). “Modeling desktop occupancy with inhomogeneous Markov chains”. In: Evaluating and


Technical reports


Other works


Earlier publications by the same author

Andersen, Philip Delff, Jakob Skårhøj, Jens Wenzel Andreasen, and Frederik C Krebs (2009). “Investigation of optical spacer layers from solution based pre-
Contents

Summary i
Sammenfatning (Danish) v
Preface vii
Acknowledgments ix
List of publications xi

1 Introduction 1
  1.1 Reducing building energy consumption .................. 2
  1.2 A test house in the Arctic ............................... 5
  1.3 Existing research on the topic ........................... 6
  1.4 Outline of the thesis .................................... 8

2 Discrete time models 9
  2.1 Linear regression ........................................ 10
  2.2 Auto-Regressive models ................................. 11
  2.3 Auto-Regressive Moving-Average models .................... 12
  2.4 Box-Jenkins models ...................................... 13
  2.5 Remarks .................................................. 13

3 Stochastic differential equations in state space 15
  3.1 Ordinary differential equations .......................... 16
  3.1.1 Disadvantages of ODEs in data-fitting ................. 17
  3.2 Stochastic differential equations ........................ 18
  3.2.1 The estimation problem ............................... 19
  3.2.2 The extended Kalman filter ........................... 20
C Characterization of heat dynamics of an arctic low-energy house with floor heating 103
C.1 Introduction 104
C.2 Description of the building 107
C.3 Methods 109
C.3.1 Grey-box models and estimation 110
C.3.2 Forward selection, likelihood ratio tests 112
C.3.3 Scope of the model 113
C.3.4 Model validation 119
C.3.5 Notes on model building 120
C.4 Experimental design 120
C.5 Data 121
C.6 Results 122
C.6.1 Model building 122
C.6.2 Correlation of inputs 124
C.6.3 Parameter estimates 134
C.7 Interpretation and application 137
C.7.1 Estimation of physical properties 137
C.7.2 Further model validation 139
C.7.3 Discussion of the accuracy of the method 140
C.7.4 Simulations 141
C.8 Conclusions 141
References 145

D Low-energy house in Sisimiut – Measuring equipment 147
D.1 Introduction 148
D.2 Overview of the system 149
D.3 Consumption measures 151
D.3.1 Oil consumption 152
D.3.2 Electricity consumption 152
D.3.3 Water consumption 154
D.4 Room temperature measurements 157
D.4.1 Standard room temperature measurement 158
D.4.2 Principal component analysis of temperature distribution 159
D.5 Activity sensors 161
D.5.1 Passive infrared sensors ............................................. 161
D.5.2 Open/closed sensors .................................................. 164
D.5.3 CO₂ sensors ........................................................... 164
D.5.4 A short period of data ................................................. 164
D.6 Floor heating .............................................................. 165
D.7 Ventilation ............................................................... 172
D.7.1 Cooker hoods .......................................................... 174
D.8 Weather station ......................................................... 175
D.9 Solar thermal system .................................................... 177
D.10 Data logging ............................................................... 179
D.10.1 Network overview ................................................... 179
D.10.2 Data acquisition hardware ......................................... 179
D.10.3 Data acquisition software ......................................... 180
D.10.4 Backup system ....................................................... 181
D.11 Experimental planning ................................................ 182
D.11.1 Using Unity Pro M .................................................... 182
D.11.2 Using MS Excel ...................................................... 184
D.12 Data extraction ........................................................... 184
D.12.1 Data extraction using the PLC GUI ............................... 185
D.12.2 Data extraction from the MSSql server ......................... 186
D.12.3 Using the SQL server at DTU ...................................... 187
D.12.4 Using extracted data ................................................. 188
D.13 Conclusions .............................................................. 191
D.14 List of sensors ............................................................ 193
D.15 Network overview ....................................................... 200
References ........................................................................ 202

E Low-energy house in Sisimiut – Data overview .......................... 203
E.1 Introduction .................................................................. 204
E.2 Overview of the collected data ......................................... 205
E.2.1 Energy signatures ...................................................... 205
E.3 Conducted experiments .................................................. 208
E.3.1 Experiment 1 – A single-room experiment ...................... 209
E.3.2 Experiment 2 – repetition of Experiment 1 ....................... 213
E.3.3 Experiment 3 – A multi-room experiment ....................... 215
E.3.4 Experiment 4 – repetition of Experiment 1 ....................... 218
E.3.5 Experiment 5 – A multi-room experiment with electric radiators .................................................. 221
E.4 Conclusions .................................................................. 224
References ........................................................................ 225
F Parameter Identifiability in Grey-box Models of Heat Dynamics of Buildings 227

F.1 Introduction .................................................. 228
  F.1.1 The transfer function .................................. 229
  F.1.2 Obtaining the transfer function ....................... 230
  F.1.3 Design of input signals ................................ 231

F.2 Example: An insulated wall .................................. 231
  F.2.1 Simulating data .......................................... 233
  F.2.2 First solution: Two known parameters ............... 234
  F.2.3 Other solutions - briefly .............................. 236
  F.2.4 Summary .................................................. 237

F.3 Example: The envelope of a heated building ............. 237
  F.3.1 One state ................................................ 238
  F.3.2 Two states .............................................. 238
  F.3.3 Three states - modelling floor heating .............. 241

F.4 Conclusions .................................................. 242

References ...................................................... 243
The work behind this thesis was initiated because of a lack of reliable methods to characterize the thermal dynamics of existing buildings. Such methods are needed for two purposes. One is extracting knowledge such as energy signatures and deficiencies in the energy performance. The other is for design of controllers and intelligent systems which can ensure comfort and reduce energy consumption in buildings.

Another important possible application of knowledge about the thermodynamics of buildings relates to growing electricity consumption and the increasing power generation from renewable sources. The thermal capacity of buildings can potentially be used for demand side load-balancing in power grids.

A building in use is a complex dynamical thermal system under numerous influences including weather, heating, cooling, ventilation systems, user behavior and other internal loads. Identification of reliable models of such systems requires thorough application of statistics and time series analysis.
1.1 Reducing building energy consumption

Easy, cheap, and reliable access to energy has characterized the western part of the world since the beginning of industrialization. Economic growth has been sustained by increased energy consumption with the consequence of increasing pollution and CO₂ emissions.

In recent years, energy consumption has been receiving huge public and political attention due to a mix of increasing energy prices, the wish of independence from some energy supplying countries, and last but not least alarming reports about the impacts of CO₂ emissions on the global climate. This is while the global energy demand is still increasing – a development expected to continue for years to come due to rapidly growing economies, mainly in Asia (Asif and Muneer 2007).

In the European Union and the U.S., 20–40% of the energy consumption takes place in buildings – a share that is expected to increase (Perez-Lombard et al. 2008). Figure 1.1 shows the development of the energy consumption in buildings together with the total energy consumption in the European Union from 1990 to 2011 (European Commission 2013).

It is clear that the energy consumption in buildings needs to be reduced by further improvement of building design, improved control of heating, cooling, and ventilation systems if the political goals are to be met and in order to reduce modern society’s global footprint. The energy-efficiency of existing buildings must be improved. Providing the information needed for economic decision-making in refurbishment requires reliable energy signatures of buildings. An

![Figure 1.1 The evolution of the total energy consumption in the European Union. (European Commission 2013, online data code: ten00094).](image-url)
1.1 Reducing building energy consumption

energy signature of a building should be able to accurately describe its thermal performance. It must be able to characterize the effect of influences like indoor and outdoor temperature, solar radiation, wind, and occupant behavior.

Static and dynamical relationships are needed for many important purposes such as control of heating and ventilation with respect to indoor comfort and maybe fluctuating energy price signals. If reliable models of the heat dynamics of buildings can be obtained, the thermal mass in buildings provides an energy storage that may be used to shift some of the energy demand away from demand at peak hours. For some buildings, there may even be a potential of shifting some of the heating demand to days where more renewable energy is available.

Many building simulation tools exist: EnergyPlus (Crawley et al. 2000), Esp-R (Strachan et al. 2008), Trnsys (Fiksel et al. 1995), the Danish BSim (Wittchen et al. 2005), and others (overview of building energy simulation tools given by Hong et al. 2000)) can simulate heat loss of buildings based on detailed information about the construction, building materials etc. However, the energy demand in buildings often turns out very different than predicted by such simulations. Figure 1.2 shows the performance gap between actual and expected in a study of 33 simulations and measured heat losses in 21 buildings by Gorse et al. 2012). The average deviation from the simulated to the measured heat loss is about 50%. Furthermore, in almost all cases, the actual heat loss is larger than the simulated.

![Figure 1.2](image)

**Figure 1.2** Comparison of simulated and measured heat loss. A positive performance gap means that the observed heat loss is larger than predicted (Gorse et al. 2012).
Poor workmanship is one of the reasons for constructed buildings not living up to expected energy performance. Figure 1.3 shows infrared pictures of the outer leaf of two highly insulated walls. The walls were designed to be exactly similar and have a U-value of 0.2 W/(m²K). However, in-situ testing led to estimates of 0.21 and 0.86 W/(m²K). The main reason for the U-value of 0.86 W/(m²K) is air looping due to poor workmanship (Hens et al. 2007). The construction of a building is a complex process which cannot be neglected in estimation of thermal dynamics of buildings.

![Figure 1.3](image.png)

**Figure 1.3** Two walls that are designed to be exactly identical (Hens et al. 2007).

In Northern Europe, the requirement of increased energy efficiency leads to better insulated buildings with increased reception of solar radiation. This is both seen in improvement work on existing buildings and the increasing interest in building types such as low-energy buildings, zero-emission buildings, near-zero energy buildings, passive houses, etc.

In refurbishment, a tool for identifying causes for heat loss and what parts of the building to refurbish is needed. Once the refurbishment has been performed, the quality of the work should be validated using similar modeling techniques. Good and reliable models are crucial, both in order to refurbish the parts that will improve the energy performance the most for the lower costs, and for measuring the obtained effect.

In the construction of buildings of improved energy efficiency, validation of the energy performance should provide a solid reference for comparison of different building designs, new technologies etc. Simulation tools stay an important tool in the design-phase of a building, but evaluation of the performance must be based on data from the constructed building rather than simulations.
Modern energy-efficient buildings, however, comprise more complex modeling than traditional, simpler ones. As buildings get better insulated, they are characterized by larger time constants, adding to the need for dynamical models. At the same time, increasing window areas introduce strong nonlinearities in the response to solar radiation. In all brevity, modeling of heat dynamics of buildings is becoming more and more important, and more and more challenging.

The last aspect of building energy in this section is the increased importance of modeling of occupancy of buildings. As buildings get better insulated, an increasing share of the heating demand is being covered by internal loads. The heat input provided by the presence of people and the devices and systems they use is becoming an increasingly important heat input.

### 1.2 A test house in the Arctic

Although only a very small part of the global CO$_2$ emissions takes place in Arctic areas, the influence of the consequent climate changes is obvious in these regions. Diminishing icecaps and the threatened polar bear have become symbols of global climate change.

The symbolism is however not the only attraction of doing experiments in a cold climate. Long winters give flexibility on planning of experiments, and generally low temperatures give a large signal-to-noise ratio. The large difference in amount of sunlight from summer to winter provides excellent conditions for studies of the influence from solar radiation.

A low-energy building in Sisimiut, Greenland, owned by Technical University of Denmark has been used as a test facility in the present studies. The location is shown in Figure 1.4. Apart from the advantages of the location related to identification and estimation of thermal properties, the house is interesting because it is a modern low-energy building. It is well-insulated, has large window areas and features floor heating and mechanical ventilation. Moreover, it consists of two semi-detached symmetrical apartments, separated by common spaces. This provides the possibility of comparing the influence of different uses of the apartments on the energy consumption. The test building has been equipped with comprehensive measuring and control equipment, which has been used to collect the data analyzed in the studies.
1.3 Existing research on the topic

Approaches to data-driven estimation of heat dynamics of buildings make use of different time series methods. Most of them are in the time domain in which all models considered in this work are as well. This section briefly outlines some relevant research that this thesis builds upon. It uses terms from time series analysis that will be introduced in detail in the following chapters.

Kusuda et al. (1971) have already used experimental data to estimate heat dynamics of the thermal mass of a building. Their approach was based on discretization of ordinary differential equations into a difference equation using ordinary least squares to estimate parameters. This approach does not include any noise on the evolution of the system itself, only on the observations. The same idea was pursued by Sonderegger (1977). Letherman et al. (1982) used pseudo random binary signals (PRBS) to estimate impulse response from heating to indoor temperature.

In discrete time, ARMAX models were discussed by Crawford and Woods (1985) and applied to data from experimental data from a test cell by Norlén (1990) and to heat fluxes through a wall by Jiménez et al. (2008a). In continuous-time, modeling of thermal properties of building components using RC-networks was done by Hammarsten et al. (1988). An early overview of types of models,
1.3 Existing research on the topic

especially ARMAX and RC-networks, was presented by Rabl (1988).

The PASSYS project (1986-1992) was a large European project on developing methods for testing solar and thermal characteristics of building components. The project was initiated by the European Commission and included testing on test cells in several European countries. Cools and Gicquel (1989) provides an overview of the work in the first half of the project period. An overview of the PASSYS project was provided by Wouters et al. (1993).

In 1994, the Joint Research Centre under the European Commission launched a competition on identification of heat dynamics of building components. Bloem (1994) is a compilation of the selected contributions, demonstrating state-of-the-art methods in the field.

The continuous-time approach has been given increasing attention in recent years. A two-state linear stochastic differential equation formulated in a grey-box model showed promising abilities to describe the heat dynamics of parts of a very well-insulated building in Madsen and Holst (1995). Andersen et al. (2000) included a multi-room model based on the grey-box framework as well. Jiménez et al. (2008b) used non-linear grey-box models to describe heat exchange dynamics of photovoltaic module. Bacher and Madsen (2011) described a forward selection procedure to be used in grey-box modeling of heat dynamics of buildings. The method starts by the simplest feasible model and stepwise extends it with the most significant of a finite number of predefined states. The method is illustrated on a light and relatively simple building.

For parameter estimation in grey-box models, an implementation based on the extended Kalman filter called IdKit was presented by Bohlin (1994) and Bohlin and Graebe (1995). However, the implementation lacks the ability to describe input-dependent diffusion. Another implementation, which is also based on the extended Kalman filter but supports input-dependent diffusion was presented by Kristensen et al. (2004b). The latter implementation has other advantages to earlier implementations, using an algorithm better suited for estimation of nonlinear models. Kristensen et al. (2004a) presents useful methods for identification of the model structure in grey-box models. A recent overview of estimation methods for stochastic differential equations was given by Donnet and Samson (2013). The paper has a focus on pharmacokinetic/pharmacodynamic applications.
1.4 Outline of the thesis

The thesis contains a brief introduction to some important statistical methods that have been used in the papers. First, some important models in discrete time are included, and then follows a brief introduction to the continuous-discrete time state-space models. The last of the introductory chapters contains general discussions of the papers and a condensed presentation of some of the most important conclusions.

The main contributions are found in the included papers. They include three journal papers, two technical reports, and a manuscript. Paper A presents a framework for modeling presence of occupants in buildings. The models used in the paper are essentially Markov chains. They are introduced in the paper, and the reader will not need further introduction to be able to read it.

Paper B is on discrete-time modeling of the heat-dynamics of a test-building. The paper includes a description of the building and a description of how to extract important physical properties of a building using discrete-time models. The methods applied in this paper belong to the class treated in Chapter 2. However, the paper also includes methods for deriving estimates of properties such as time constants and UA-values from such models.

Paper C presents grey-box modeling of data from an experiment that was carried out in the same test-building as described in Paper B. The paper relies on the theory introduced in Chapter 3.

The technical reports in Papers D and E serve to document the experiment facilities used and the experiments performed. This includes the experiment from which the data is modeled in Paper C.

Finally, Paper F is a manuscript about identifiability of dynamical systems and experimental planning with the aim of characterization of heat dynamics of buildings.
Chapter 2

Discrete time models

Since data consists of a finite number of measurements on a physical system, it seems natural to formulate a model in discrete time. If the sample period is constant – possibly except from missing data points – discrete-time modeling certainly provides a powerful and flexible toolbox. The framework ranges from simple non-dynamical linear regression to models where different dynamics are related to different inputs. An advantage of these types of models is that most of them are implemented in standard modeling software such as R and MATLAB. In the simplest cases, closed-form solutions to the maximum likelihood problem exist.

Suppose that recordings of a univariate output, \{Y_k\}, are available at equidistant time instants \(k \in \{1, \ldots, N\}\). This could be the heat consumption in a building with approximately constant indoor temperature. The heat consumption will then obviously depend on weather variables and other inputs, \{u_k\} where for all \(k\): \(u_k \in \mathbb{R}^u\). This chapter gives an overview of discrete-time models describing such relations. We shall use the matrix notation, \(u_k\), to denote the \(k\)'th time instance of the inputs, \(u\). Notation will be introduced as it is being used.

The introduced models are all described in Madsen (2008) and Madsen and Thyregod (2011). For time series models, a classical reference is Box and Jenkins (1970). Since the aim is only to make the reader familiar with the different classes of models, we shall restrict this introduction to deal with single-output
systems. The mentioned literature treats multiple-output systems as well.

The chapter starts by non-dynamical linear regression models and then dynamical and more advanced linear models are introduced. Finally remarks are given about the model identification process.

2.1 Linear regression

The simplest model of the output given the inputs, is linear regression. For $Y \in \mathbb{R}^N$ (strictly speaking, $Y$ is a random vector which sample space is in $\mathbb{R}^N$), the linear model is:

$$Y = x\theta + e$$

where the elements of $e$ are independent and identically distributed (i.i.d.) with $e_k \sim N(0,\sigma^2)$, $x \in \mathbb{R}^N \times \mathbb{R}^p$ is the known design matrix, $\theta \in \mathbb{R}^p$. $\theta$ is a parameter vector, and $x$ is called the design matrix and can contain the input variables (elements in $u$), linear or nonlinear functions of these, functions of time etc. It is noticed that this general linear model can describe nonlinear relations between $u$ and $Y$. The strictly linear relation is between the design matrix, $x$, and the output, $Y$.

Advantages of the general linear model are the simplicity of the solution of the maximum likelihood problem, well-known procedures for model reduction, and model validation through residual analysis. Provided that $x$ has full rank, the ordinary least-squares estimate of the parameters in (2.1) is simply (Madsen and Thyregod 2011):

$$\hat{\theta} = (x^Tx)^{-1}x^TY$$

As long as the noise is Gaussian, the least squares estimate equals the maximum likelihood estimate. In $\mathbb{R}$, the function $\text{lm}$ can be used to perform maximum likelihood estimates of such models.

An assumption behind linear regression is, as mentioned, that the error terms are i.i.d. In practice this means that data points must be of a low enough sample frequency for correlation between them to be insignificant. But the time constants in heat dynamics of buildings are often a day or more (Madsen and Holst 1995; Andersen et al. 2000) – which is a result in Paper C as well. This static limitation makes this approach demand very long periods of data. Also, since they are non-dynamical models, they can of course only be used to estimate steady-state relations.
Steady-state relations like UA, gA values etc. can in principle be derived using (2.1). But the requirement of a sample time “beyond dynamics” of the system leads to very long testing periods. If a two-day sample time is used, two months of measurements will only lead to thirty data points. This is still little for statistical analysis.

The model 2.2 can be generalized to describe autocorrelated noise process, using weighted least squares estimation. In order to increase the sample frequency and describe the dynamical behavior of the system, we will instead turn towards models that directly express the dynamics of the system.

2.2 Auto-Regressive models

The first dynamical model to consider can be interpreted as a linear model as (2.1) where the design matrix consists of historical values of the output, \( x_k = (y_{k-1}, \ldots, y_{k-p}) \). This can also be formulated as

\[
Y_k = -\sum_{i=1}^{p} \phi_i Y_{k-i} + e_k
\]  

(2.3)

Notice that this is the solution to an ordinary differential equation with noise. We shall use the polynomial notation:

\[
\phi(B)Y_k = e_k, \quad \phi_0 = 1
\]  

(2.4)

where \( B \) is the back-shift operator such that

\[
B^j Y_k = Y_{k-j}, \quad j \in \mathbb{Z}
\]  

(2.5)

and \( \phi \) is a polynomial of order \( p \) with the restriction that \( \phi_0 = 1 \). This is the autoregressive (AR) model of order \( p \). The polynomial, \( \phi \), determines the dynamics of the system. For a stationary process, all solutions to

\[
\phi(z^{-1}) = 0, \quad z \in \mathbb{C}
\]  

(2.6)

with respect to \( z \) lie within the unit circle.

The AR model can be extended with inputs or rather exogenous processes. Without loss of generalization, we will consider only one exogenous process, \( x \):

\[
\phi(B)Y_k = \gamma(B)x_k + e_k
\]  

(2.7)
where $\gamma$ is another polynomial, this one without restrictions related to the coefficients. This is an autoregressive model with exogenous inputs, abbreviated an ARX model. The polynomial fraction

$$H(z) = \frac{\gamma(z)}{\phi(z)}$$

(2.8)

is the transfer function from the input to the output, where $z$ is the $z$-operator, i.e. the discrete-time equivalent to the Laplace operator.

In many applications on heat dynamics in buildings, ARX models provide a good framework for describing basic dynamical relations, such as time constants. This is done in Paper B.

AR models can be estimated using the “method of moments”, where the analytic expression of the autocorrelation (a function of the elements in $\phi$) are matched with the estimated autocorrelation. Another approach, which can also easily be used with exogenous processes (in ARX models), is formulating a model of the form (2.1) and using the estimate (2.2). The design matrix of an ARX model could have the structure

$$x = \begin{pmatrix} Y_{N-1} & \cdots & Y_{N-p} & u_N \\ \vdots & \vdots & \vdots & \vdots \\ Y_p & \cdots & Y_1 & u_{p+1} \end{pmatrix}$$

(2.9)

where a single input is used without delay.

## 2.3 Auto-Regressive Moving-Average models

The moving-average model is a regression model of the output against historical noise terms. The model is given as

$$Y_k = \theta(B)e_k, \quad \theta_0 = 1$$

(2.10)

This can be combined with the AR model to the autoregressive moving average (ARMA) model

$$\phi(B)Y_k = \theta(B)e_k$$

(2.11)

where $\phi_0 = \theta_0 = 1$, and $\theta$ is a polynomial of order $q$. Because of the transfer function related to the noise process, ARMA models cannot be fitted using least squares estimation, cf. (2.2). Instead, maximum likelihood estimation can be
2.4 Box-Jenkins models

used. Madsen (2008) presents both exact likelihood and conditional likelihood techniques. In R, the function arima can be used.

With exogenous process as well, the model becomes

\[
\phi(B)Y_k = \gamma(B)x_k + \theta(B)e_k
\]

which is the ARMAX model. The moving-average term can accommodate for uncertainty on inputs and on the output \((u \text{ and } y)\) in data deviating from their true physical values.

In R, the package “DSE”\(^1\) can be used for maximum likelihood estimation of ARMAX processes based on the Kalman filter. However, it only supports estimation of one common transfer function for all inputs.

2.4 Box-Jenkins models

The denominators in all transfer functions from all exogenous processes and from the noise term in (2.12) are the same, \(\phi(z)\). The Box-Jenkins model does not have this restriction and is the most general linear dynamical model, assigning full distinct transfer functions to each input and to the noise process:

\[
Y_k = \frac{\omega(B)}{\phi(B)} x_k + \frac{\theta(B)}{\phi(B)} e_k
\]

These types of models are referred to as Box-Jenkins models or transfer function models. They can be fitted with the R-package TSA\(^2\). Even though the ARMAX model is a sub-model of the transfer function model, TSA does not seem to be able to estimate ARMAX models.

2.5 Remarks

The model identification process is normally an iterative process where both knowledge about the physical system and data analysis is used. In the data analysis, the estimated correlation function is a central tool. The aim of the model building is to satisfy the assumption of white noise.

\(^1\)http://www.cran.r-project.org/web/packages/dse/
\(^2\)http://www.cran.r-project.org/web/packages/TSA/
Finally, it is mentioned that the model framework described here can be extended in different ways. An important example is that seasonality can be included in the ARMA processes such that periodic dependencies can be included. In building energy, seasonal (diurnal, weekly, annual) dependencies are likely to be significant. Also, it is mentioned that conditional parametric models are a very flexible class of models and have been applied in e.g. Bacher et al. (2013).
Since heat transfer relations are normally expressed in differential equations, it is intuitive to stay within this framework when analyzing data from thermodynamical systems. This makes it easier to apply physical knowledge about the system and relate to parameter estimates.

The framework of continuous-time models has other advantages than being intuitive. A restriction on the discrete-time models introduced in Chapter 2 is equidistance of observations. In the following, the very flexible class of stochastic differential equations (SDEs) will be introduced together with an estimation technique that allows for variable sample time and variable noise variance. The model is formulated in state space which separates system noise from observational noise. This is the type of models referred to as grey-box models.

The chapter starts by introducing ordinary differential equations and discussing why the reformulation into SDEs is crucial in parameter estimation, forecasting, and control. Then stochastic differential equations are introduced, and a method for state and parameter estimation for a class of SDEs is outlined. Finally, the subject of identifiability of state space models is mentioned.
3.1 Ordinary differential equations

Differential equations are one of the most important tools in modeling physical, biological, economic, and many other systems in continuous time (Gershenfeld 1999). Heat transfer as well is described by differential equations (Cengel 2006). It is therefore natural to turn to models in differential form for modeling heat dynamics. Consider the state vector, \( x_t \in \mathbb{R}^n \) at time \( t \), and let \( x \) denote the element-wise differential of \( x_t \) with respect to time. Then the first-order ordinary differential equation (ODE) of \( x_t \) can be established:

\[
\dot{x}_t = f(x_t, u_t, \theta_t, t) \tag{3.1}
\]

where \( f : (\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}) \rightarrow \mathbb{R}^n \). \( u_t \in \mathbb{R}^u \) is a vector of (normally time dependent) known inputs, \( \theta_t \in \mathbb{R}^p \) is a parameter vector. The time subscripts mean that all the components can depend on time. Since \( f \) expresses the evolution of \( x \) in time given the inputs, \( u \), the description is by nature dynamic. If a solution, \( x_t \) exists, it is a deterministic function which expresses the trajectory of the state, given an initial value, \( x_0 \). The class of first order ODE’s is indeed a very important class of models, since an ODE of any order can be reformulated into a first order ODE (Christensen 2006).

If the system (3.1) is an observed physical system, say temperatures at different locations in a wall, the system itself is not observed but information about the state, \( x \), must be inferred from observations, say readings, \( \{y_k\} = \mathbb{R}^l \), of thermometers. The observations are obtained at discrete time instants, \( k \in \{1, \ldots, N\} \), and as measurements of physical systems they are normally subject to noise. Introduce the observation equation

\[
y_k = g(x_{t_k}) + e_k \tag{3.2}
\]

where \( g : \mathbb{R}^n \rightarrow \mathbb{R}^l \), and \( \{e_k\}_{k=1}^N \) are independent random vectors. (3.1) and (3.2) compose an ODE in state space representation. The variable of special interest is the state, \( x \), while the realization of \( Y \) is observed. Typical problems in such a system are to estimate the function \( f \) and to re-construct, filter, or predict the state, \( x \). By assuming a structure of \( f \), \( g \), and a distribution of \( e \), a maximum likelihood approach may be used to infer on the parameters, \( \theta \). For this, the ODE in (3.1) must be solved either analytically or an approximation to the solution must be calculated using e.g. a Runge-Kutta method (Hull et al. 1972). Notice that the subscripts, \( k \in \{1, \ldots, N\} \), do not assume anything about the discretization scheme. The measurements do not need to be equidistant.
3.1 Ordinary differential equations

3.1.1 Disadvantages of ODEs in data-fitting

Once a model estimate is derived and the ODE is solved, the state can be predicted at all future using only an initial value. Denote these predictions, $y_k|0$. The series of prediction residuals, $\epsilon_k|0$, can be derived using the estimated model to predict the output of the system

$$\epsilon_k|0 = Y_k - \hat{Y}_k|0$$

(3.3)

As mentioned the errors, $\{e_k\}$, are independent over time. Hence, if the correct model structure has been chosen, are independent. However, this can only be the case if the functions, $f$ and $g$, are perfectly known and perfectly describe the observed system. In practice, this is most often an unrealistic assumption.

Imagine a disturbance enters the system at a given time. The disturbance results in a change in the state that is not described by the model, and a new initial value ought be used for solving the ODE in order to perform new predictions. If this is not considered in the analysis, the residuals will contain dynamics that the model does not account for. Obviously from this point, the predictions of the system are of lower quality. But also, statistical tests regarding inference on parameters may be invalid because fundamental assumptions about independence of residuals are violated (Madsen 2008).

Disturbances in the evolution of the state can be due to the true value of inputs deviating from recordings, or to deficiencies in the model formulation. Such changes are either out of the scope of the physical model description or they are due to effects that the modeler does not have information about. In continuous time, disturbance is not one or more isolated events but rather something that happens continuously with infinitesimal increments (Øksendal 2007).

A good physical description of a system may allow the formulation of a good model of its trajectory, but in reality an input is slightly different than what is used in the model run, or something happens that is not even included in the model. And once the trajectory has changed, the system is unlikely to come back to the one predicted. The ODE solution provides a perfect trajectory prediction of e.g. the temperature at a certain position in a medium from the moment it is solved for all future time points. Unfortunately the next moment the solution is wrong and will never be right again. So while ODE’s have the strength of expressing a physical system, they lack the ability describe “random” variation, i.e. to allow variation that is not perfectly predictable.
3.2 Stochastic differential equations

Random variation needs to be incorporated in the differential equation. This will facilitate obtaining uncorrelated residuals, which is necessary for reliable inference on parameters. It seems natural to include a noise term on the differential equations like

$$\dot{x} = f(x_t, u_t, \theta_t, t) + \text{“noise”}$$ (3.4)

This noise process must be defined in continuous time and it should share some properties with Gaussian noise in discrete time. These properties are formulated on increments of the process. Denote the univariate noise process $\omega$.

Let the process start in 0:

$$\mathbb{P}(\omega \neq 0 = 0)$$ (3.5)

Second, let increments be zero-mean Gaussian with variance scaling with time:

$$(\omega_{t+s} - \omega_t) \sim N(0, s)$$ (3.6)

Finally, let the increments be independent:

$$(\omega_{t+s} - \omega_t) \perp \perp (\omega_{t+s+u} - \omega_{t+s}), \quad t \in \mathbb{R}, \quad s, u > 0$$ (3.7)

Such a process is called a Brownian motion or a Wiener process\(^1\). When – as here – the incremental variance is the time increment (and not scaled), it is a standard Wiener process. Notice that even though it is natural to think of $s > 0$, it is not a necessity. This is because the Wiener process is time-reversible. The process obeys the martingale property:

$$\mathbb{E}(\omega_{t+s} - \omega_t) = 0, \quad s, t \in \mathbb{R}$$ (3.8)

which follows directly from the independence and central Gaussian distribution of increments. Notice also that the variance of $\omega_{t+s} - \omega_t$ equals $s$. From this follows that the unit of the Wiener process is the square root of the unit used for time. For thorough treatment of stochastic differential equations, refer to Øksendal (2007), Kloeden and Platen (1992), and Steele (2001).

A property of the Wiener process resulting from (3.6) is that it is nowhere differentiable. Even though the variance of the increments go to zero as the time increment goes to zero, the independence of increments leads to this theoretically complicated property. In order to be able to integrate with respect to such

---

\(^1\)Robert Brown was the first to observe this physical motion which Albert Einstein later explained with theoretical physics. The stochastic process is named after Norbert Wiener.
3.2 Stochastic differential equations

A process, a definition of the integral is needed. Itô and Stratonovich have suggested definitions of integrals to use, and each of them have their advantages. While the Itô integral has advantages in parameter estimation, the Stratonovich integral facilitates use of the chain rule. Conversion between the two integrals can be performed using a relatively simple relation (Øksendal 2007). Here, only the Itô integral will be used.

Equipped with the Itô integral, a stochastic differential equation (SDE) can be established:

\[ \text{d}X = f(X_t, u_t, \theta_t, t)\text{d}t + \sigma(X_t, u_t, \theta_t, t)\text{d}\omega_t \quad (3.9) \]

where the first (and deterministic) term is called the drift term. The second (and stochastic) term is called the diffusion term. \(f\) and \(\sigma\) are differentiable mappings into \(\mathbb{R}^n\) and \(\mathbb{R}^n \times \mathbb{R}^n\) respectively, and \(u_t \in \mathbb{R}^u\) and \(\theta_t \in \mathbb{R}^m\) are defined as in Section 3.1. \(\omega\) are standard Wiener processes. The solution to a stochastic differential – which can only be found analytically in simple cases – is a stochastic process.

Again, the system is observed through an observation equation:

\[ Y_k = h(X_{t_k}, u_k, t_k, \theta) + e_k \quad (3.10) \]

where \(h\) maps into \(\mathbb{R}^l\). (3.9) and (3.10) form a system of coupled stochastic differential equations in state-space representation. \(e\) is the observation noise which is a white noise process. Here, only Gaussian observation noise will be treated, so let \(e_k \sim N(0, \Sigma)\) where \(\Sigma\) can be a function of \(u, \theta,\) and \(t\). The system consists of a stochastic differential equation (3.9) in continuous time and an observation equation (3.10) in discrete time. We shall refer to an SDE in state space representation as a grey-box model.

3.2.1 The estimation problem

Given a series of measurements \(y\) of the stochastic process, \(Y\), a typical statistical problem is to reconstruct, filter, or predict the state of the system, to estimate parameters in the system, or maybe all of this. The software CTSM-R provides an implementation of the extended Kalman filter and optimization of the parameters for a sub class of this type of models. The methods used will be briefly explained. For more details, refer to Kristensen et al. (2004b) and Kristensen and Madsen (2003).

The subclass of (3.9) and (3.10) that can be estimated by CTSM-R has the limitation that the noise processes cannot depend on the state. The reduced
type of models can be written as

$$dX = f(X_t, u_t, \theta_t, t)dt + \sigma(u_t, \theta, t)d\omega_t$$

(3.11)

$$Y_k = h(X_{t_k}, u_{t_k}, t_k, \theta) + e_k,$$

(3.12)

where $\sigma$ is a mapping into $\mathbb{R}^n \times \mathbb{R}^n$ which determines how the used Wiener processes affect each state as in (3.9). $e_k \sim N(0, S^2)$ is the observation noise.

Let $\mathcal{Y}_N$ denote all observations available up to time $t_N$:

$$\mathcal{Y}_N = (Y_0, \ldots, Y_N)^T$$

(3.13)

Recall the definition of conditional probability (Pitman 1993). The joint probability of the events $A$ and $B$ is the product of the probability of $A$ given $B$ times the probability of $B$:

$$P(A \cap B) = P(A|B)P(B), \quad P(B) > 0$$

(3.14)

Then the joint likelihood function of the parameters in the system given the available observations follows as (Kristensen et al. 2004b):

$$L(\theta; \mathcal{Y}_N) = \prod_{k=1}^{N} P(Y_k = y_k|\mathcal{Y}_{k-1}, \theta) P(Y_0|\theta)$$

(3.15)

The joint likelihood function has been partitioned into a sequential product of probability densities of one-step-ahead predictions.

The Fokker-Plank equation or the Kolmogorov forward equation describes the future evolution of the density function for a diffusion process by a partial differential equation in space and time (Øksendal 2007). This could be solved at each time step to calculate the conditional distributions of the output for each prediction in (3.15). However, by assuming that the conditional densities are Gaussian, the extended Kalman filter can be applied. This assumption can be checked after running the algorithm by comparing the residuals to the normal distribution. The residuals must also be checked for correlation.

### 3.2.2 The extended Kalman filter

The Kalman filter is used to re-construct and predict the state of the system given a set of measurements, a model and known inputs. The original Kalman filter as presented by Bucy and Kalman (Kalman et al. 1960) is a filter for linear systems. It can be shown to be the optimal state estimate under the assumption of Gaussian noise (Madsen 2008). The filter consists of two steps that are being
carried out repeatedly for each data point: updating and prediction. It provides a recursive filter with the advantage that only the last state of the filter has to be used to generate the next prediction. It is therefore ideal for online use, and has many applications such as position tracking, control of physical systems, noise reduction, and many others. The filter used here is a so-called extended Kalman filter (EKF) using a linearization around the prediction value for forecasting the system noise.

Consider a system on the form (3.11)–(3.12). Introduce the following one-step ahead prediction of the output:

\[
\hat{Y}_{k|k-1} = \mathbb{E}(Y_k|\mathcal{Y}_{k-1}, \theta)
\]

\[
= h(\hat{X}_{k|k-1}, u_k, t_k, \theta)
\]

which is seen directly from (3.12). Then by using the linearizations

\[
\hat{A} = \left. \frac{\partial f}{\partial X_t} \right| \hat{X}_{k|k-1}, u_k, t_k
\]

\[
\hat{C} = \left. \frac{\partial h}{\partial X_t} \right| \hat{X}_{k|k-1}, u_k, t_k
\]

The variance-covariance matrix of \(\hat{Y}_{k|k-1}\) is given by

\[
R_{k|k-1} = \mathbb{V}(Y_k) - \hat{Y}_{k|k-1}
\]

\[
= \hat{C}P_{k|k-1}\hat{C}^T + S
\]

where \(P\) is the state covariance matrix. The summation of the two variance contributions follows the fact that the system noise and the observation noise are independent (Shumway and Stoffer 2006). The one-step-ahead prediction errors, also referred to as the innovations, are given by

\[
\epsilon_k = Y_k - \hat{Y}_{k|k-1}
\]

where \(\epsilon_k\) is short-hand for one-step-ahead prediction error, i.e. \(\epsilon_k = \epsilon_{k|k-1}\) The last thing to introduce before the two steps of the EKF, is the very important Kalman gain:

\[
K_k = P_{k|k-1}\hat{C}^T R_{k|k-1}^{-1}
\]

The Kalman gain expresses the ratio of the total variation of the innovation that is due to the diffusion process. So if the Kalman gain is zero (no system noise), the new observation is disregarded. If the Kalman gain is the identity matrix, no noise is on the observations, and the observation is given full weight in the update of the state estimate. Normally, the Kalman gain is a weight in between these two extremes.
The updating step updates the prediction to estimate the state of the system using the newly obtained information in the recent observation:

\[ \hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k \epsilon_k \] (3.24)

and the covariance is updated accordingly:

\[ P_{k|k} = P_{k|k-1} - K_k R_{k|k-1} K_k^T \] (3.25)

The prediction step extrapolates the state by using the estimate and the model structure. The prediction is given by

\[ \frac{d\hat{X}_{t|k}}{dt} = f(\hat{X}_{t|k}, u_t, t, \theta) \] (3.26)

and

\[ \frac{d\hat{P}_{t|k}}{dt} = \hat{A} P_{t|k} + P_{t|k} \hat{A}^T + \sigma \sigma^T \] (3.27)

which have to be solved at each time step. The integrals of (3.26) and (3.27) to extrapolate the state and the system noise are performed with a numerical integrator using subsampling in the non-linear case. The default in CTSM-R is to use the Adams algorithm which places more subsamples where the state is more non-linear (Hull et al. 1972). Notice that if an observation is missing, and the prediction error cannot be calculated, the update (3.26) can be evaluated disregarding the second term, corresponding to using zero Kalman gain. However, in order to perform the prediction in (3.26), the inputs, \( u_t \) must be known. Hence, observations can be missing, but input values cannot.

The Kalman filter estimates the state \( X \) for each time step, given that the structure and all parameters in the equations (3.11) and (3.12) are perfectly known. The next section deals with estimating the parameters given an assumed model structure.

### 3.2.3 Parameter estimation

Using the assumption of Gaussian innovation, the likelihood of a set of parameters, \( \theta \), as given in (3.15) can is calculated based on the set of observations \( \mathcal{Y}_N \) using the multivariate normal distribution (Kristensen et al. 2004b):

\[
\mathcal{L}(\theta; \mathcal{Y}_N) = \prod_{k=1}^{N} \frac{\exp \left( -\frac{1}{2} \epsilon_k^T R_{k|k-1}^{-1} \epsilon_k \right)}{\sqrt{\det(R_{k|k-1})} (\sqrt{2\pi})^l} P(y_0|\theta) \tag{3.28}
\]
This function can be used to formulate an objective function – the negative log likelihood function – and the minimum of that is the maximum likelihood estimate of the problem:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \{- \ln(L(\theta; \mathcal{Y}_N))\} \quad (3.29)$$

where $\Theta$ is some predefined parameter space. This is what is implemented in CTSM-R. Initial values of the states and the discretized system noise are needed to initiate the algorithm. In CTSM-R, the following are used:

$$\hat{X}_{t|t_0} = X_0$$

$$P_{t|t_0} = P_s \int_{t_0}^{t} e^{\hat{A}_s} \sigma \sigma^T (e^{\hat{A}_s})^T ds \quad (3.31)$$

where $X_0$ can be estimated.

In the state prediction (3.26), the inputs are assumed known in continuous time. However, in practice the inputs will often be measurements in discrete time. Therefore, an assumption must be used on the evolution of the inputs between samples in order to perform the integration. CTSM-R provides the options zero order and first order hold. The choice should depend on the nature of the inputs.

### 3.2.4 The linear case

The linear class of coupled stochastic differential equations is a very important subclass of the class of systems covered in this chapter. The Kalman filter steps simplify and become much faster to compute, which is an obvious advantage. Also, the theory in the linear framework is simpler, and often one will linearize a non-linear system in order to derive important properties such as time constants or to study observability and identifiability of the system.

$$dX = (AX + Bu_t)dt + d\omega$$

$$Y_k = CX_{t_k} + Du_k + e_k \quad (3.33)$$

The system matrix, $A$, contains the dynamics of the system. Remember that for the homogeneous ordinary differential equation

$$\dot{x} = Ax \quad (3.34)$$

Recall that the eigenvalues of $A$ are roots in $\det(A - \lambda I)$ with the eigenvectors $v$ that satisfy $(A - \lambda I)v = 0$. In this work, all eigenvalues are negative but
theoretically they could be any complex numbers. Assume that $A$ has $p$ distinct negative eigenvalues $\lambda_i < 0, i \in \{1, \ldots, p\}$. Then the solutions to (3.34) are entirely given by

$$\textbf{x}(t) = \sum_{i=1}^{p} c_i e^{\lambda_i t} \textbf{v}_i, \quad i \in \{1, \ldots, p\}$$

The time constants of the system are given by

$$\tau_i = -\lambda_i^{-1}, \quad i \in \{1, \ldots, p\}$$

A time constant, $\tau_i$, can be interpreted as the time it takes for the system vector to multiply by $e^{-1}$ when it starts in the corresponding eigenvector, $\textbf{v}_i$.

### 3.2.5 Observability and identifiability

The state-space formulation allows for a partitioning of a model into a system equation describing the physical system and an observation equation describing the measurements. However, it may be that, given the inputs and the measurements, one is not able to filter and estimate the assumed model reliably. These properties are called observability and identifiability. If interpretation of the model parameters is of interest and if the model estimate will be used on simulations or forecasts using new input data, these properties are crucial to ensure.

Observability means that for a given set of parameters and known inputs and outputs, it is possible to estimate the state of the system over time. The system (3.32)–(3.33) is observable if

$$\begin{pmatrix} C \\ \vdots \\ CA^{n-1} \end{pmatrix}$$

has full rank.

Identifiability is related to the inference on the parameters of the system. For a model to be identifiable, different values of the parameters must lead to different probability distributions of the output of the system. In other words, the joint likelihood function of all the parameters must have a unique global maximum given the output of the system. If this is theoretically the case for the system given infinite amounts of data from an ideal experiment, the system is said to be structurally identifiable.
To analyze structural identifiability for a state-space model, the structure must be compared with the related transfer functions. The transfer functions from the inputs, $U$ to the output $Y$ are given by

$$H(s) = C(sI - A)^{-1}B + D, \quad s \in \mathbb{C}, \quad sI \neq A$$

(3.38)

If the parameters in $A, B, C,$ and $D$ are uniquely defined given $H(s)$, the system is structurally identifiable with respect to the physical parameters. Paper F is a short analysis of the structural identifiability of the physical parameters of the models fitted in Paper C. For structural identifiability of the parameters related to the noise processes, see e.g. Madsen and Holst (1995).

While structural identifiability has to do with identification of the model parameters from the transfer function, practical identifiability is related to whether the transfer function can be estimated from data. This is related to the dynamics of the inputs and the amount of data. Input signals must be mutually uncorrelated and persistently exciting, which PRBS (Godfrey 1980) signals can be shown to be (Ljung and Söderström 1983). This is why the experiment used in Paper C is based on the floor heating following PRBS signals.
Stochastic differential equations in state space
The main contributions of this thesis fall in three categories, which can be further divided. The main categories are:

- Modeling of presence of occupants in buildings
- Energy performance characterization of buildings
- Identification of heat dynamics of buildings

This chapter presents an overview of the main findings in the three areas and discusses applications and further challenges. It is organized by subject, and references will be given to the individual papers. Finally, the main conclusions are briefly given.

### 4.1 Modeling of presence of occupants in buildings

The first paper (A) is related to modeling of presence of occupants in buildings. The presented approach provides two major abilities. The one is to describe the
variability of occupant presence as a function of time of day. The other is to describe the per-occupant dynamics of presence. The result is a method that can be implemented in building simulation tools for simulating reliable scenarios of internal loads. The method was used in (Iversen et al. 2012).

The analysis is performed on data from an office building with 57 work spaces. The developed model is an inhomogeneous Markov chain (Grimmett and Stirzaker 2005). Generalized linear models (Madsen and Thyregod 2011) provide a very general framework for modeling transition probabilities. The diurnal variability is obtained using logistic regression with periodic splines as inputs. It certainly shows the ability to reproduce the average occupancy pattern along the day. Generally, the variance of the total number of occupants is well matched with data as well. However, the variance in data is somehow larger than provided by the model in the afternoon. This seems to be due to a single Friday where the occupants have left earlier than the rest of the days. Simulations using the developed model and mean and confidence levels for the developed and simpler models are compared in Figure 4.1.

![Figure 4.1](image)

**Figure 4.1** A comparison of model simulations and data studied in Paper A. The upper panel shows 16 simulations of total number of present employees as a function of time of day using the chosen model. The lower panel shows mean values and non-parametric confidence intervals for data and different models. The inhomogeneous Markov chains fit the daily variation well. Figure A.12.

Dependence on time of day is modeled using periodic splines as inputs. This method results in quite a large number of parameters. This could be reduced
4.2 Characterization of energy performance of buildings

by using for instance kernel smoothing or conditional parametric models for the inputs.

Also the per-employee dynamics are well matched by the developed method. This is achieved by using exponential smoothing – another Markov chain (see e.g. Madsen 2008) – as input in the generalized linear model. Again this method has not been compared with other methods and others which may just as well account for this over-dispersion. Exponential smoothing is however a simple method and leads only to one extra parameter. The way the model is formulated using exponential smoothing can easily be taken out or replaced by another method. It is noticed that for larger numbers of occupants, if the aim is only to simulate the total load or total number of present occupants, the exponential smoothing is unnecessary.

A weakness of the analysis is that it is based on a quite limited amount of data. The method needs to be further tested on similar and other types of buildings and on larger data sets.

4.2 Characterization of energy performance of buildings

Two different approaches have been studied relating weather data to energy consumption in buildings. Paper B derives energy signatures based on discrete-time input-output models, and Paper C derives similar characteristics using continuous-discrete time state-space models.

Paper B suggests a dynamical model inspired by the literature (namely Mortensen and Nielsen 2010) and estimates the parameters based on data from the same house from two distinct periods. The two periods differ in several ways. Most notably in the weather variables, the amount of solar radiation is considerably larger in one of the periods. The domestic hot water consumption is more than three times larger in one period than the other, suggesting considerable difference in occupancy over the two periods. The estimated UA and gA values are shown in Table 4.1.

Improvements on the building envelope have been done in between the two considered periods because of substandard workmanship in the original construction. Hence it was expected to see differences in the energy performance. However, such improvements are not found significant in the analysis. This can of course be due to a lack of real improvement of the energy performance. But the significantly different amounts of solar radiation are seen from the experience
Table 4.1  Estimates of physical properties of the building in two distinct periods in Paper B. Results from Table B.4.

<table>
<thead>
<tr>
<th>Property</th>
<th>Period 1</th>
<th></th>
<th>Period 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>Std. E.</td>
<td>Est</td>
<td>Std. E.</td>
</tr>
<tr>
<td>$\hat{U}A_{\text{mean}}$ (W/K)</td>
<td>119.8</td>
<td>23.6</td>
<td>121.0</td>
<td>12.3</td>
</tr>
<tr>
<td>$gA$ (m$^2$)</td>
<td>5.7</td>
<td>5.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

in Paper C to be very important. In the latter paper, the solar radiation has been given a lot of attention leading to a nonlinear model, reacting differently to solar radiation depending on azimuth angle and solar elevation. Adding the indications of differences in occupancy of the building, such a comparison of two distinct periods seems difficult using only daily values. There were issues with low-resolution of data, which was ultimately the reason for using daily values in the analysis.

Paper C takes the approach of continuous-discrete time grey-box models using 15-minute resolution data. The resulting estimate of the UA-value has very little uncertainty and is being validated by splitting the data set in two halves, and re-estimating on each of these. Also the window areas are consistently estimated over the two half data sets. However, for the latter, the interpretation is harder because a projection of solar radiation is related to one of them. The estimated “window areas” and the estimated UA value and their uncertainties are shown in Table 4.2.

Table 4.2  Main thermostatic properties estimated on an apartment in Paper C. Results from Table C.6.

<table>
<thead>
<tr>
<th></th>
<th>est</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Areas, m$^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_s$</td>
<td>6.1</td>
<td>0.5</td>
</tr>
<tr>
<td>$A_w$</td>
<td>1.8</td>
<td>0.5</td>
</tr>
<tr>
<td>UA-value, W K$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UA</td>
<td>82.8</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Both the window areas and the UA-values found using grey-box models in Paper C on only one apartment of the building seem relatively large compared to the values found for the whole building using discrete-time models in Paper B. However, it must be noticed that the analyses in the two papers are not di-
rectly comparable. In the analysis of daily values in Paper B, the average of a temperature measurement in each apartment is used as the indoor temperature. In the grey-box time model in Paper C, the average of six central temperature measurements in the one apartment is used. The heat supplies used in the models are different corresponding to estimates for one apartment and the whole building, respectively. Furthermore, the discrete-time model is estimated on periods where the building was in use, whereas the continuous-discrete time model is estimated on a relatively short planned experiment where one of the two apartments was not used. The floor heating signal was designed to excite the dynamics, and due to extremely low outdoor temperatures (-30°C), the indoor temperature went down to about 14°C in this period. The last and maybe the most important difference is the use of new measuring equipment for the data collection in Paper C. Obtaining a reliable estimate of the floor heating input was an issue which is discussed in Paper C. Discrete-time models were also applied on the data analysed in Paper C, and the obtained results were consistent with the results obtained using grey-box models. This is, however, not documented in the papers.

It is underlined that dynamical modeling has been applied even though measures of energy performances are typically generalizations of steady-state characteristics. Due to constant changes of the inputs, a building is never in steady-state. Hence, a dynamical approach must be taken. Consequently, steady-state characteristics must be derived from estimated dynamical models. Characterization of steady-state thermal properties of a building becomes a sub task of identification of heat dynamics of the building. Alternatively, single data points must be average values of periods longer than the largest time constant of the system. This will lead to a need for very long testing periods.

4.3 Identification of heat dynamics of buildings

While both Paper B and C present application of dynamical models, the used sample times – one day and 15 minutes, respectively – make them hard to compare. The aim of Paper B is rather derivation of steady-state properties, and daily values provide a low resolution for dynamical analysis. Mortensen and Nielsen (2010) recommend that four-hour or finer resolution is used for dynamical modeling of heat dynamics of the considered buildings in order to describe the response to outdoor temperature and solar radiation. This will of course heavily depend on the properties of the building and on which part of the dynamics are of interest.

In Paper C a dynamical model of a central temperature in one of the two apart-
Table 4.3  Estimates and their standard deviations of dynamical properties of an apartment studied in Paper C. Results from Table C.6.

<table>
<thead>
<tr>
<th></th>
<th>est</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total thermal capacity, kW h K$^{-1}$</td>
<td>6.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Time constants, h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>$\tau_3$</td>
<td>53.6</td>
<td></td>
</tr>
</tbody>
</table>

ments is developed. The model matches the dynamics of the output quite well, only a slight diurnal autocorrelation is found in the residuals. Solar radiation is modeled using both global horizontal radiation and a projection of this onto the main window facade of the building. Also, the floor heating system has been modeled by a separate state. The estimated total heat capacity of the apartment and the time constants of the system are shown in Table 4.3. The largest time constant of the system is about two days, reflecting the very slow response to floor heating.

Paper C also demonstrates how simulations of step and impulse responses can be used to study the dynamics of the system. These responses are shown in Figure 4.2. The figure clearly shows the distinct responses to ventilation and floor heating. For control of the indoor temperature with horizons up to a couple of hours, ventilation must be used. Such knowledge about the heat dynamics is central in design of model predictive control (MPC) of heating inputs.

4.3.1 Experimental design

An experimental setup was designed and installed in a low-energy house in Sisimiut, Greenland, as documented in Paper D. Experiments were designed and performed on this system yielding the data presented and modeled in Papers C and E.

Early in the process of data analysis, challenges appeared related to determining the power dispatched in the system. Briefly described, parameters were to be estimated for all heat inputs. But since thermal capacity is a key property in the heat dynamics of the system, this led to over-parametrization of models. The floor heating input was measured through volume flow and temperature
measurements. It became important to estimate density and heat capacity of the heating liquid before starting modeling. This problem can be avoided by always using electric heating. The recommendation is always to have electric heaters at least as a part of the heating. This ensures that a part of the heating is easily measured and also that high frequency variations of the input can be used in the experimental design.

Floor heating itself leads to a related problem, even if it is electric and the power can easily be measured. Since the floor provides a boundary between the indoor temperature and the ground or outdoors (if the building is suspended), the floor can in practice exchange heat both with indoors and the ground (or outdoors). Again, if an electric heater is added inside the building, the floor heating contribution can be estimated by formulating the model such that a certain amount of energy reaching a certain state will change the temperature of this state equally, no matter from where this energy comes. The contribution from electric heaters must be quantitatively comparable to that of floor heating in order to obtain better estimates.

This recommendation was followed in a new experiment carried out in the spring 2013. This is documented as “Experiment 5” in Paper E. The experiment was designed such that different inputs were given to different rooms. This makes it a good experiment for studying multi-room models of the apartment as well. Initial modeling was done but has not been included in this thesis due to the late availability of the data.
These problems with over-parametrization lead to a more general study of structural identifiability. Paper F is a manuscript describing how to handle these issues. It contains calculations showing that even some simple and intuitive systems turn out not to be identifiable. The plan is to make it into a paper about identifiability for studies of heat dynamics in buildings. So far, it is at least advised always to check the structural identifiability of a model before fitting it to data.

As mentioned, a little diurnal correlation is seen in the residuals of the proposed model in Paper C. It is demonstrated how it is related to solar radiation. It is recommended that for measurements on buildings with large window areas, pyranometers are installed directly in the direction of those window areas. E.g. for a building with a large window facade towards south, both a pyranometer measuring global horizontal radiation and one preferably mounted directly on the wall with the window facade should be used. Another interesting approach would be to let blinds open and close following a PRBS signal.

As already discussed, the experience in this work is clearly in favor of integrating electric heating in experiments because of the simplicity with which their contributions are measured. Moreover it gives a fast response of the heating system, so the dynamical characteristics of the heating system are easy to separate from the dynamics of the whole building. It can even be beneficial to include multiple electric heating systems. If a liquid-borne floor heating is used, it would be very interesting to also have electric heating placed together with the liquid-borne floor heating. Then the response of temperatures in the system to these two heat sources can be compared, and one can obtain good understanding of heat capacity and possibly time delays in the liquid-borne system. If also compared with electric heating dispatched inside the building, one can obtain estimates of the thermal resistance of the floor.

Another issue related to experimental setup is independence of inputs. In the experiments described in Paper E and in the one analysed in Paper C, the modeled area of a building is surrounded partly by the building envelope and partly by walls to adjacent room. An estimate of the thermal resistance of the envelope is of particular interest. In one of the adjacent rooms, the heating was not on during the experiment. This leads to a sub-optimal experimental design where one input (the temperature in the adjacent room) is a low pass filter of other inputs (the outdoor temperature, solar radiation etc.) and the output (the temperature measurement in the modeled area). In order to be able to distinguish the responses from different inputs, the inputs must not be too correlated. The experience in this work is that adjacent rooms must be heated following mutually uncorrelated control signals. A good design can be obtained by using distinct pseudo random binary signals (PRBS) in each adjacent rooms – of course all different from control signals for other input signals.
4.3 Identification of heat dynamics of buildings

4.3.2 Variable noise variance

The recursive filter used (EKF as explained in Section 3.2.2) makes it easy to use input- or time-dependent noise processes. The model suggested in Paper C features variable observational noise. This leads to a weighting of the observations depending on the input. In this case, it improves on the model fit significantly measured on both autocorrelation of residuals and the likelihood value.

If the model deficiency is due to periodic uncertainty or bias on inputs, it seems however more reasonable to use variable system noise than observational noise. This approach was also attempted in the modeling work behind Paper C. Promising tendencies were seen. The likelihood value was larger and reduced autocorrelation of residuals was obtained. However, the estimates of thermal capacities were heavily affected. Therefore, a conservative choice was made, and the variable system noise variance was omitted. There may be a good potential in using input-dependent system noise but an improved approach for model validation could then be considered. As discussed in the paper, the distribution of the residuals is obviously non-Gaussian, making it a quasi-maximum-likelihood method. If possible, it would be preferable to collect enough data to partition the data set into “training”, “validation”, and “testing” periods (see i.e. Hastie et al. 2009; Izenman 2008).

4.3.3 Discrete-time or continuous-time models

The choice of one framework or the other makes important differences in the modeling work and in the interpretation of the obtained models. As explained in Section 3.1, unless the diffusion processes in the system are completely dominated by the drift term, ODE’s are unfeasible for simulation, prediction, and control. Hence, by continuous-time modeling we shall refer to the framework of SDE’s on state-space form.

The continuous-time models have the advantage of containing directly interpretable physical parameters. This makes it easy to perform “sanity checks” of model-fits, and to learn about the physical system from these. In that sense, grey-box modeling bridges the gap between physical (deductive) and statistical (inductive) modeling. In general, continuous-time models of physical systems are intuitive to work with because of the close relation to physics which also brings in some robustness in the modeling procedure.

In structural identification of grey-box models, a parameter can be re-defined as a hidden state evolving as a random walk (Kristensen et al. 2004a; Møller et al.
\[ d\theta_t = \sigma_\theta d\omega_\theta \]  

(4.1)

The idea is using \( \theta \) as a parameter in another state equation. If the estimated trajectory of \( \hat{\theta} \) shows systematic dependence on time, states, or inputs, this dependence can reveal model deficiencies.

The partitioning of the model into a state equation in continuous time and observations in discrete time allows for non-constant measurement sampling. In discrete-time formulation one often has to interpolate inputs and/or outputs to obtain this, applying a low-pass filter to data and hence losing dynamics in the signals. Furthermore, this means that missing observations is not a problem in the continuous-time framework. However, inputs must be known in order to be able to predict the system.

Another advantage of the continuous-time state-space formulation is the freedom and ease with which the noise processes are formulated. Both the system and observation noise can be modeled as functions of inputs, time, and states. CTSM-R does not support state-dependent noise terms but in simple cases, a relevant transformation of the SDE such as the Lamperti transform can be used to overcome this (Møller et al. 2011b).

The theoretical complexity and the heavy computations in numerical integration are drawbacks of continuous-time stochastic processes in many engineering applications. However, with the introduction of the multithreading-enabled implementation in a user-friendly \( \text{R} \)-package, CTSM-R\(^1\), these challenges have been reduced.

It must be remembered that what can be estimated directly from data is transfer functions from inputs and noise processes to the output. It is important to notice that given a transfer function model there exists infinitely many state-space representations. This means that the user must have structural identifiability in mind when formulating models with the aim of interpreting on the physical system, predicting or controlling. For a transfer function model, this will normally not be an issue. On the other hand, the lack of direct physical interpretation can lead to formulation of unfeasible transfer function models. A way to avoid issues with structural identifiability of state-space models is to consider canonical forms which are ensured to be identifiable. Then, on the other hand, some of the physical interpretation may be changed. This difference between transfer function models and state-space models only concerns structural identifiability. The issues related to practical identifiability such as persistently exciting and uncorrelated inputs are shared between the two modeling approaches.

\(^1\)http://ctsm.info/
For derivation of steady-state relations such as estimates of UA-value, gA-value etc., simple transfer function models of rather low-resolution averaged data are likely to be a good approach if the relations can be assumed to be linear. If the aim is higher data efficiency and/or a more detailed dynamical description, the experiences in this work are in favor of a grey-box approach.

4.4 Main conclusions

Data-driven models have been applied to modeling of heat dynamics of buildings. The contributions include modeling of presence of occupants and different approaches to characterization of the energy-performance of building envelopes and the thermal dynamics of buildings.

A method has been developed to estimate patterns of occupant presence from data. The method can be used to simulate reliable scenarios of occupant presence. Both diurnal variability and per-employee dynamics are described by the model. The method was developed using data from an office building. However, the framework is general and can be used on data from different types of buildings, i.e. dwellings, office buildings, hospitals, schools, etc.

Data-driven approaches to modeling of heat dynamics of buildings and characterization of their energy performance have been investigated. Since buildings are always under dynamic conditions, static characterizations based on ordinary least squares methods can only be used if data is averaged to sample periods longer than the largest time constant of the system. This leads to very long testing periods and inefficient use of data. Instead, dynamical models must be applied. From these, steady-state characteristics can be derived.

Two different approaches have been outlined and reported for statistical modeling of heat dynamics. The one is discrete-time autoregressive moving average models with exogenous inputs (ARMAX), closely related to transfer function modeling. This is sometimes referred to as black-box modeling. It was seen to be a strong tool for characterization of energy performance of buildings. A framework of linear dynamic models (ARMAX) was described together with methods to extract important information about heat dynamics from model estimates. The method was successfully applied, and the model and the estimation results were validated and reported.

The other approach is continuous-discrete time state-space modeling. Differential equations based on physical knowledge are used to describe the physical system. This is combined with stochastic diffusion processes and observational
noise into a continuous-discrete time state-space model often referred to as a grey-box model. A procedure has been outlined to build and validate stochastic grey-box models of the heat dynamics of a building with multiple heat inputs, including floor heating and solar radiation through a large window area. Solar radiation was given special attention due to its complex influence on the indoor temperature. Only global horizontal radiation was measured. A projection of this onto the large window area proved to be particularly important in the modeling of the dynamics. The variance of the observational noise was described as a function of this projected solar radiation as well.

The presented grey-box approach is shown to be highly data efficient; based on a 16-day experiment, key thermal parameters are estimated with relatively little uncertainty. A UA-estimate was given with uncertainty corresponding to 1% and total thermal capacities with an uncertainty of 3%. These uncertainties are conditioned on the model being correct which includes several assumptions. These assumptions are probably sources of larger uncertainty than the statistical uncertainty in the parameter estimation.

Recommendations were given for experiments with the aim of identification of heat dynamics of buildings. In the present experimental setup there are adjacent rooms to the area of a building to be modeled. It is advised that the heat input to such adjacent rooms be designed with independent and persistently exciting signals in order to avoid problems with correlation with other inputs. Challenges estimating the heating power provided by the floor heating system lead to the recommendation of including electric heating giving an easily measurable heat input. This can then be used to estimate heat inputs and heat capacities in the building. Also, it was recommended to use pyranometers installed directly on walls in the direction of large window areas in order to model the fast dynamics related to solar radiation.

Grey-box models have the advantage of having physically interpretable parameters, and estimates of properties like UA-value, thermal capacities can be derived. Dynamics of the system can be understood from the obtained model fit as well. In the linear case, time constants for the building can be estimated. Simulations demonstrated different dynamic responses to ventilation heating, floor heating, and outdoor temperature. For example, impulse response from floor heating to indoor air temperature showed that this system is suitable for shifting the heating demand by several hours. This is not only an important result for control of heating, cooling, and ventilation systems in buildings, but also for design of intelligent energy grids and controllers for use in intelligent energy grids.


namics of building integrated photovoltaic modules”. In: Energy and Build-
component UA and gA from outdoor tests in warm and moderate weather
conditions”. In: Solar Energy 82.7, pp. 573–587.
Differential Equations”. Supervised by Professor Henrik Madsen, hm@imm.dtu.dk,
of Denmark, DTU Informatics.
Kalman, Rudolph Emil et al. (1960). “A new approach to linear filtering and
prediction problems”. In: Journal of basic Engineering 82.1, pp. 35–45.
Kristensen, Niels Rode and Henrik Madsen (2003). Continuous Time Stochas-
tic Modelling – Mathematics Guide. CTSM 2.3. Technical University of
Kristensen, Niels Rode, Henrik Madsen, and Sten Bay Jørgensen (2004a). “A
method for systematic improvement of stochastic grey-box models”. In:
— (2004b). “Parameter estimation in stochastic grey-box models”. In: Auto-
Kusuda, T, T Tsuchiya, and FJ Powell (1971). “Prediction of indoor temperature
by using equivalent thermal mass response factors”. In: Proceedings of the
5th Symposium on Temperature. Washington DC.
trical and computer engineering. Oxford University Press. isbn: 9780195219173.
dynamic thermal response in rooms using pseudo-random binary sequences”.
1016/0360-1323(82)90004-X. URL: http://www.sciencedirect.com/
science/article/pii/036013238290004X.
information and system sciences series. Prentice-Hall. isbn: 9780138816407.
Ljung, Lennart and Torsten Söderström (1983). Theory and Practice of Re-
cursive Identification. The Massachusetts Institute of Technology. isbn:
0-262-12095-X.
isbn: 978-1-4200-5967-0.
of Thermal Characteristics and Building Energy Performance Parameters”.
In: Whole Building Testing, Evaluation and Modelling for Energy Assess-
ment. Dynastee. Kgs Lyngby, Denmark.


Dynamic Modelling of Presence of Occupants using Inhomogeneous Markov Chains

Authors:
Philip Delff Andersen\textsuperscript{1} Anne Iversen\textsuperscript{2}, Henrik Madsen\textsuperscript{1}, Carsten Rode\textsuperscript{3}.

Accepted for publication in:
Energy and Buildings (Elsevier).

\textsuperscript{1}DTU Compute, Matematiktorvet, Building 322, DK-2800 Kgs. Lyngby, Denmark
\textsuperscript{2}Danish Building Research Institute, Dr. Neergaards Vej 15, 2970 Hørsholm
\textsuperscript{3}DTU Civil Engineering, Brovej, Building 118, DK-2800 Kgs. Lyngby, Denmark
Abstract

Occupancy modeling is a necessary step towards reliable simulation of energy consumption in buildings. This paper outlines a method for fitting recordings of presence of occupants and simulation of single-person to multiple-persons office environments. The method includes modeling of dependence on time of day, and by use of a filter of the observations it is able to capture per-employee sequence dynamics. Simulations using this method are compared with simulations using homogeneous Markov chains and show far better ability to reproduce key properties of the data.

The method is based on inhomogeneous Markov chains with where the transition probabilities are estimated using generalized linear models with polynomials, B-splines, and a filter of passed observations as inputs. For treating the dispersion of the data series, a hierarchical model structure is used where one model is for low presence rate, and another is for high presence rate.

A.1 Introduction

Occupants interact with the indoor environment through heat and carbon dioxide emission, switching lights on/off, opening windows etc. Occupancy profiles are therefore a necessary input to building simulation models that include indoor environment variables, ventilation loads, electric power consumption etc. The most common way of considering occupancy in simulation tools is by using, and if necessary repeating, one static occupancy profile (Haldi 2010; Hoes et al. 2009). Typically, the used profile is constant for weekdays and weekends, respectively. However, occupants do not arrive in buildings or leave buildings at fixed times. A study by (Manicca et al. 1999) reported that on average the offices were occupied 46% of the time, which is supported by the study by (Page et al. 2008) where it was found that only half of the work day was spent at the work station. Therefore building systems controlled by occupant presence have shown great energy saving potential in office buildings.

Recently, occupants presence models have been developed by (Tabak and Vries 2010), (Page et al. 2008), (Wang et al. 2005), (Bourgeois 2005) and (Reinhart 2004). These models include behavior of occupants based on empirical data. (Wang et al. 2005) and (Richardson et al. 2008) both developed occupant presence models as a first order Markov chain. Wang’s data fits well with the exponential distribution when observing individual offices and vacant intervals. However the exponential model was not validated for occupied intervals.
A.1 Introduction

(Page et al. 2008) considered occupant presence as an inhomogeneous Markov chain interrupted by occasional periods of long absence. By using a profile of probability of presence as input to a Markov chain they were able to reproduce intermediate periods of presence and absence distributed exponentially with a time-dependent coefficient as well as fluctuations of arrivals, departures and typical breaks. They defined a parameter called the “parameter of mobility”. This parameter indicates how much people move in and out of the zone, by correlating the tendency of coming to work with the tendency of leaving.

(Tabak and Vries 2010) looked at occupancy based on more detailed prior knowledge about time consumption on different tasks in a working day. As input to their model they included information on the intermediate activities of the occupants such as ‘receiving unexpected visitor’, ‘walking to printer’, ‘having lunch’, etc. They were able to simulate occupancy patterns using a probabilistic method for different intermediate activities. The model by (Tabak and Vries 2010) is a step towards a more behavioral approach to simulating occupancy.

The focus of the study presented in this paper is to develop a model for presence of occupants for simulation of single person presence sequences in an office environment. The study seeks answers to the following questions:

- How can the dependence of the tendency of being present on the time of day be modeled?
- Can model dependence based on past behavior improve predictions?

The aim is to present a framework for modeling occupancy in an office environment and then apply it to fitting data that is believed to be representative. The focus will be on modeling what can be considered “typical” occupant presence sequences, where “typical” is to be judged from data. Sequences of very little presence are expected to be frequent because of vacation, sickness, etc. Sequences of significantly more presence than “typical” will be omitted. The focus is on single-person simulation, and correlation structures in data will not be modeled.

The outcome is techniques for an occupancy simulation model that can be used in building simulation programs when simulating demand responsive systems such as lighting or ventilation systems.
A.2 Method

In this section, the data collection method and the mathematical framework to be used in the analysis will be described.

A.2.1 Data collection

Occupancy patterns have been measured in an office building in San Francisco, California. Data comes from ballast status records in the control system and have been registered every 2 minutes. If an occupant is present at the workspace, the lamp is switched on, and the ballast status is on. Once the workspace is unoccupied the lights drop to preliminary power and are turned off after a delay of 20 minutes. The occupants cannot override anything manually. The data collected have been corrected for the delay by setting the last 20 minutes of intervals of “presence” to “absence”. However absences shorter than 20 minutes have not been encountered because of the delay in the equipment.

Data from 86 workspaces were collected, out of which 29 were unoccupied or occupied by interns. Only data from the 57 workspaces that have been occupied by full-time staff for the entire measurement period is used.

The model fitting is based on full days in September and December 2009 and January 2010; 16 days in total. No data points are missing.

A.2.1.1 Description of models

All models in the present work are in discrete time. Let $t \in [0, T]$ be a continuous time scale. Choose a natural number, $N$, and let $\tau := \frac{T}{N}$. Then $t_n = n\tau, n \in \{0, 1, \ldots, \frac{T}{\tau}\}$ is a discretization of $t$ with sample period $\tau$. The sample period is equal to the measuring period, 2 minutes in this work.

The notation $X_n$ is introduced as shorthand for the state of the discrete-time random process $\{X_t\}$ at time $t_n$. In other words, $X_n$ refers to $\{X_t\}$ at time $t = n\tau$, in this case $n \cdot 2$ minutes.

Markov Chains A Markov Chain is a time series that meets the Markov condition stating that conditioned on the present state, the future is independent
A.2 Method

Table A.1  Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike information criterion.</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayes information criterion.</td>
</tr>
<tr>
<td>HOR</td>
<td>High occupancy rate.</td>
</tr>
<tr>
<td>LOR</td>
<td>Low occupancy rate.</td>
</tr>
<tr>
<td>rmse</td>
<td>Root mean square error.</td>
</tr>
<tr>
<td>$\tau$</td>
<td>A time stamp in continuous time.</td>
</tr>
<tr>
<td>$t$</td>
<td>A time stamp in continuous time.</td>
</tr>
<tr>
<td>$T$</td>
<td>Maximum of $t$, i.e. $t \in [0, T]$</td>
</tr>
<tr>
<td>$n$</td>
<td>A time stamp in discrete time.</td>
</tr>
<tr>
<td>$N$</td>
<td>Maximum of $n$, i.e. $n \in {0, 1, \ldots, N}$</td>
</tr>
<tr>
<td>${X_n}$</td>
<td>A random process in discrete time.</td>
</tr>
<tr>
<td>$X_n$</td>
<td>The state of the random process ${X_n}$ at time $n$.</td>
</tr>
<tr>
<td>$x_n$</td>
<td>The observation of the random process ${X_n}$ at time $n$.</td>
</tr>
<tr>
<td>$\mathbf{X}^{(i)}$</td>
<td>The $i$’th sequence of observations.</td>
</tr>
<tr>
<td>$p_n \in [0, 1]$</td>
<td>The unconditioned probability of $X_n = 1$.</td>
</tr>
<tr>
<td>$\mathbf{A}$</td>
<td>A matrix.</td>
</tr>
<tr>
<td>$\mathbf{A}^T$</td>
<td>$\mathbf{A}$ transposed.</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of states in a Markov Chain</td>
</tr>
<tr>
<td>$I$</td>
<td>The characteristic function.</td>
</tr>
<tr>
<td>$Q$</td>
<td>The number of sequences of observations.</td>
</tr>
<tr>
<td>$\log : \mathbb{R} \rightarrow \mathbb{R}$</td>
<td>The natural logarithm.</td>
</tr>
<tr>
<td>$\mathbb{N}$</td>
<td>The set of natural numbers, ${1, 2, \ldots}$</td>
</tr>
<tr>
<td>$i, j, k \in \mathbb{Z}$</td>
<td>Integers.</td>
</tr>
<tr>
<td>$\mu_i$</td>
<td>The mean of the $i$’th sequence of observations.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>The intercept in the linear domain of the generalized linear model.</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>The weighting of the $i$’th basis spline in the linear domain of the generalized linear model.</td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>The weighting of the $i$’th power of time of day in the linear domain of the generalized linear model.</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>The weighting of the exponential smoothing in the linear domain of the generalized linear model.</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The parameter for the exponential smoothing filter.</td>
</tr>
<tr>
<td>$\Lambda_n$</td>
<td>The value of the exponential smoothing filter at time step, $n$.</td>
</tr>
<tr>
<td>$\theta$</td>
<td>A parameter vector.</td>
</tr>
</tbody>
</table>

of the past (Grimmett and Stirzaker 2005). Let $\Omega$ represent the set of possible states of $X$. Then, in discrete time, $\{X_n\}$ is a Markov chain if
∀k ∈ N : n + k < N, ∀s ∈ Ω : 
\Pr(X_{n+k} = s | X_0, X_1, \ldots, X_n) = \Pr(X_{n+k} = s | X_n) \quad (A.1)

This is illustrated in Figure A.1.

A Markov chain with $M$ states is completely characterized at time $n$ by the probabilities of transitions to all states:

$$
\Pr(X_{n+1} = j | X_n = i), \quad i, j \in \{1, \ldots, M\} \quad (A.2)
$$

This means that the transition probabilities contain the distributions of the transitions from the states in the Markov chain. Hence, for each state they sum to one:

$$
\forall i \in \{1, \ldots, M\}, \forall n \in \{1, \ldots N\} : 
\sum_{j=1}^{M} \Pr(X_{n+1} = j | X_n = i) = 1 \quad (A.3)
$$

Often the transition probabilities are represented in a probability transition matrix.

**Figure A.1** Illustration of dependence in a Markov chain. The Markov condition says that the distribution of future states (here $X_{n+1}$ and on) conditioned on the present state ($X_n$) and all past states (up to $X_{n-1}$) is the same as the distribution of future states that are only conditioned on the present state. Therefore, in the graph, $X_{n-1}$ and $X_{n+1}$ are only connected through $X_n$.

Because of the constraint in Equation (A.3), at each time step the transition probabilities have $M - 1$ degrees of freedom for each state, corresponding to $(M - 1)M$ in total for each time step. When applied to binary data, $M = 2$, and hence the model has two degrees of freedom at each time step. If the transition probability matrix is constant, i.e. $\Gamma(n) = \Gamma, \Gamma \in \mathbb{R}^M \times \mathbb{R}^M$ the Markov chain is said to be homogeneous. A homogeneous Markov chain has $(M - 1)M$ degrees of freedom.

**Two-state Markov chains with covariates** Covariates in Markov chains with only the two states, 0 and 1, can be modeled as

$$
\text{logit}(\Pr(X_{n+1} = 0 | X_n = 0)) = Z_{1,n} \theta_1, \theta_1, Z_{1,n} \in \mathbb{R}^p \quad (A.4a)
$$
\[ \text{logit}(P(X_{n+1} = 1 \mid X_n = 1)) = Z_{2,n} \theta_2, \theta_2, Z_{2,n} \in \mathbb{R}^q \quad \text{(A.4b)} \]

where the logistic function denoted logit is defined as
\[ \text{logit} : ]0, 1[ \rightarrow \mathbb{R}, \quad \text{logit}(x) = \log \left( \frac{x}{1-x} \right) \quad \text{(A.5)} \]

and \( \log \) is the natural logarithm. \( \theta_1 \) and \( \theta_2 \) are parameter vectors while \( Z_1 \) and \( Z_2 \) are design vectors. \( P(X_{n+1} = 1 \mid X_n = 0) \) and \( P(X_{n+1} = 0 \mid X_n = 1) \) are calculated by application of Equation (A.3). This formulation has the advantages that the parameters are unconstrained while the resulting probabilities span and never exceed \( ]0, 1[ \). This is a generalized linear model (Madsen and Thyregod 2011) for binomial data, and logit is the canonical link function which maps from the full range of the real numbers into \( ]0, 1[ \). This model has \( p + q \) free parameters.

\( Z \) is a design matrix that can contain any observable real input. Here, functions of time will be used. One design matrix could be
\[ Z^T = (1, n, n^2) \quad \text{(A.6)} \]

where \( Z^T \) denotes \( Z \) transposed. This would result in a second order polynomial of time to be passed through the logistic function. In (A.6), 1 means that an offset is included in the model (for \( n = 0 \)), and the parameter representing this offset is denoted \( \alpha \).

The dependence of \( \{X_n\} \) on past values and on the exogenous process is illustrated in Figure A.2. Since Equations (A.4) describe transition probabilities which vary with some exogenous process, this Markov chain is inhomogeneous. When dependence on time of day is used, the parameter in the linear domain of the generalized linear models will be denoted \( \rho_i \) where \( i \) is the power of the time of day.

![Figure A.2](image)

**Figure A.2**  Illustration of dependence in a Markov chain, \( \{X_n\} \) with a covariate, \( \{Z_n\} \). The input process is a deterministic process which is assumed to be known.

Generalized linear models are implemented in R and can be fitted using the `glm` function.
Natural splines  Splines are piecewise polynomial functions. In this work, B-splines with natural boundary conditions are used. These are piecewise third order polynomials with the boundary condition that the second derivatives are zero at the end-points (Eldén et al. 2004). The polynomials are between knots for which number and positions have to be chosen. In this work, knots are always equidistantly spread.

In R, the basis functions of natural splines can be calculated using the splines package. By using the basis functions in the design matrix, splines are fitted as input to the generalized linear model.

Where natural splines are used in the general linear models, the parameters in the linear domain are denoted $\beta_i$ where $i$ means that it relates to the $i$'th basis function.

Exponential smoothing  Exponential smoothing is a lowpass filter. It is a weighted average, with the weights decaying exponentially with time difference. The speed of the decay is contained in the only parameter, $\lambda \in [0, 1]$:

$$\Lambda_n = \lambda X_n + (1 - \lambda)\Lambda_{n-1} \quad (A.7)$$

Since $\{\Lambda_n\}$ is a weighted average of $\{X_n\}$, it has the same range as $\{X_n\}$.

In the framework of Equations (A.4), the design matrix for a model using exponential smoothing and no covariates is

$$Z_n = (1, \Lambda_{n-1})^T \quad (A.8)$$

Figure A.3 is a graph of the information flow using exponential smoothing and no covariates. As seen from Figure A.3, the Markov condition is still respected when using the exponential smoothing as input as long as the most recent, and only the past states of $\{X_n\}$ are used in the design matrix as in Equation (A.8). Notice that the exponential smoothing adds two parameters to the model, one is the exponent, $\lambda$, the other is the parameter in the linear domain of the generalized linear model. The latter is denoted $\gamma$.

Finally, both filtered states and exogenous processes can be used in the design matrix. A graph of this model is shown in Figure A.4. The design matrix can now include both a column with ones, polynomial functions of time of day, basis splines of time of day, and exponential smoothing of the observations.
A.2 Method

A.2.1.2 Model performance assessment

The model estimation is based on the maximum likelihood principle. Let \( X_n \) be 1 for occupant presence, 0 for occupant absence at time \( n \). Assume that it follows the Bernoulli distribution with parameter, \( p \), the probability of \( X_n = 1 \). Then the likelihood function of \( p \) given the observation, \( x_n \), is:

\[
L(p; x_n) = \mathbb{P}(X_n = x_n) = \begin{cases} 
1 - p, & x_n = 0 \\
p, & x_n = 1 
\end{cases} \quad (A.9)
\]

The joint likelihood of observations \( x_1, x_2, \ldots, x_N \) is the product of the individual likelihood values:

\[
L(p; x^{(N)}) = \prod_{n=1}^{N} L(p; x_n) \quad (A.10)
\]

The maximum likelihood estimate of \( p \) refers to the value of the parameter that maximizes the likelihood function.

\[
\hat{p}(x^{(N)}) = \arg \max_p L(p; x^{(N)}) \quad (A.11)
\]
Here, $p$ can also be a function of other parameters, $\theta$. Then the maximum likelihood estimate of $\theta$ is parameters that maximizes the likelihood function.

$$\hat{\theta}(x^{(N)}) = \arg \max_{\theta} \mathcal{L}(p(\theta); x^{(N)})$$  \hspace{1cm} (A.12)

Instead of the likelihood function itself, the logarithm of the likelihood function, simply called the log-likelihood and denoted $\ell$, is often used. This has the advantage that the joint log-likelihood function is a sum instead of a product:

$$\ell(p(\theta); x^{(N)}) = \log \left( \prod_{n=1}^{N} \mathcal{L}(p(\theta); x_n) \right)$$

$$= \sum_{n=1}^{N} \log(\mathcal{L}(x_n; p(\theta)))$$  \hspace{1cm} (A.13)

Since the natural logarithm is an increasing function of all positive numbers, the log-likelihood can be maximized just as well as the likelihood itself.

In a homogeneous Markov chain, the transition probabilities can be estimated in the same way. The transition from $i$ to $j$ is a Bernoulli experiment that happens with probability $\mathbb{P}(X_{n+1} = j \mid X_n = i)$. Notice that the likelihood function of the conditional probability in (A.10) should only be based on the data where $X_n = i$.

For an inhomogeneous Markov chain, a parametric relation over time can be determined using parametric expressions of the transition probabilities as in (A.12).

**The estimation routine**  For a given smoothing parameter, $\lambda$, for the exponential smoothing (A.7), the following steps are carried out.

1. Exponential smoothing is calculated for the whole data sequence using (A.7).
2. Parametric expressions (basis splines or other polynomial expressions) of time of day are calculated.
3. The design matrix is formed by exponential smoothing (one column) and polynomial relations (several columns, for splines, one less than the number of knots).
4. The parameters in the general linear model are fitted using `glm` in R.

The log-likelihood value of this total model is used as the objective function in an optimization algorithm. For the optimization, the implementation of the Brent algorithm in `optimize` in R is used (Brent 2002).
Information criteria  For testing models against each other, likelihood-ratio tests can be used if the models are nested (one model can be obtained by equaling parameters in the other to zero). Since change of positions of the spline knots leads to models that are not sub-models of each other (not nested), an information criterion is needed to compare the performance of different models.

The Akaike Information Criterion (AIC) is a popular choice of information criterion (Wasserman 2003). For the model, $S$, it is given by

$$AIC(S) = -2 \cdot \ell_S + 2 \cdot k$$ (A.14)

where $\ell_S$ is the log-likelihood value of the parameters of $S$ at the maximum likelihood estimate. $k$ is the number of parameters in the model. However, it may be an advantage to use the Bayesian Information Criterion (BIC) which takes the amount of data into account.

$$BIC(S) = -2 \cdot \ell_S + \log(N) \cdot k$$ (A.15)

where $N$ is the number of data points. In this work, BIC is used for model choice.

A.3 Results

A.3.1 Data overview and preparation

Data recordings for every two minutes from 57 sensors over 16 full days were considered. The first records were from August 2009, the last from January 2010. The time stamps in the data files were in PST/PDT (Pacific Standard Time/Pacific Daylight Time). Working hours were assumed to follow local time. Therefore “time of day” is used for modeling referring to the local time zone, i.e. PST/PDT.

A.3.1.1 Choosing periods to model

It was investigated if some times of the day, some sensors, or even whole days should be skipped. The total number of activated sensors was inspected throughout each of the available days to ensure that none of them were holidays. The total number of occupants was plotted for all of the 16 considered days in the upper region of Figure A.5. Two days look a bit different than the rest with lower occupancy in the afternoon, but none of the days were so different that
they could be considered non-working days. They are a Tuesday and a Friday and hence not one day of the week that could be different from the others. Apart from these two days where there is slightly lower afternoon occupancy, the days are quite similar. All days were kept for the analysis.

Narrow spikes of high occupancy, even after 8 p.m., are seen in many – if not all – of the sequences of total occupancy. This means that the status of the sensors are correlated. The spikes are unlikely to be caused by employees coming to and leaving their desk but rather by one or more persons activating several sensors. It is known that a guard walks through the building every night and this could be the cause of some of these spikes. Since these spikes are likely not to caused by usage of the workspaces, they are not of considered particularly interesting in this work.

The lower region of Figure A.5 is a boxplot of total occupant presence in the building grouped on hour of the day. It is seen that until 6 a.m., the activity is very close to zero, except for between 5 a.m. and 6 a.m. where there is a slight activity on some of the days. From between 6 a.m. and 7 a.m. to between 10 and 11 a.m. the activity increases to around 30 simultaneous positive measurements. From between 10 to 11 a.m. to between noon and 1 p.m. the total occupation decreases to slightly more than 20 as median. This drop could be explained by a lunch break. The activity increases until between 2 and 3 p.m. after which it starts dropping. After between 3 and 4 p.m. the activity drops quickly until
between 6 and 7 p.m. where the median is below 5 sensors again. Also, from this plot it is clearly seen that the many narrow peaks in occupancy after 7:30 p.m. are caused by relatively few outliers from the generally low occupancy. It is seen that the variance of the occupancy is larger in the afternoon than in the morning. Time intervals where the occupancy is small were left out, and based on Figure A.5, only model occupant presence from 6 a.m. to 7 p.m. were included in the model. Only this part of data is considered from this point.

A.3.1.2 Identifying outlier employees

It was then checked if data from some sensors was significantly different from the rest and should be considered outliers. It was expected that single sensors would be inactive almost throughout whole days because of employees being away. A boxplot of the mean activity over each day for each sensor is shown in Figure A.6. The distribution of the daily means of the different sensors are quite different, both in medians and in variance. Many days of low occupant presence are seen, and also workspaces with generally very low occupant presence. This seems to be too many to simply disregard them as outliers and will be further investigated below. However, a few sensors have very high occupancy (6, 20, 26, 56, 57) and some of these (especially 6, 20, and 56) have low variance in occupancy. These could be located in areas that are passed by other employees throughout the day. They are considered significantly different from the rest. The vertical lines at the upper edge of the plot show the sensors that were left out.

In the data modeling description, the data considered is stripped from the outliers described here.

A.3.2 A hierarchic model

To determine a threshold of when to consider a sequence of measurements from one day as a working day or not, the distribution of the mean occupancy throughout a whole day of all sensors is considered. A histogram of this is seen in Figure A.7. There is a high density close to zero, and then the density is generally decreasing until mean occupant presence of a bit less than 0.2. It could be a mixture of one distribution with mode close to zero (not at work) and another with mode close to 0.6 (a work day). Based on this it is decided to make a threshold at a mean of 0.2 activity for a day-sequence. This corresponds to 2.6 hours of activity. Sequences with less occupancy than 20% (from 6 a.m.–7 p.m.) will be used to fit a low presence rate model, sequences with more than 20%
presence are used to fit a high presence rate model. This is done after removal of the outliers detected in Section A.3.1. The densities in Figure A.7 are colored according to this division of data. The blue color represents the sequences that fall into the “low occupancy” category, the green ones into the “high occupancy” category whereas the red ones are the outliers which are not used in the model fitting.

The model of the presence of one employee becomes a hierarchic model, see Figure A.8. With a certain probability, \( P_{HPR} \), the employee is modeled with a model describing occupant presence patterns with a mean presence higher than 0.2, whereas another model with mean presence lower than 0.2 will be used with probability \( 1 - P_{HPR} \). The model of particular interest in the present paper is the model describing presence – the “High presence rate model”. Some key properties of the partitions of the data are shown in Table A.2. The procedure of estimating this model is outlined. For the low-presence sequences, the same procedure has been carried out and the results will be given.

The probability, \( P_{HPR} \) was estimated as

\[
\hat{P}_{HPR} = \frac{1}{N_s} \sum_{s=1}^{N_s} I(\mu_s > 0.2) \approx 0.686
\]  
(A.16)

where \( \mu_s \) is the mean presence in the sequence, \( s \).

\[
\mu_s = \frac{1}{N} \sum_{n=1}^{N} x_{t(s)}
\]  
(A.17)
A.3 Results

Figure A.7 Distribution of occupant presence per day for all sensors. The colors indicate what groups the different series fall into.

Figure A.8 The hierarchic structure of the model. With probability $\hat{P}_{\text{HPR}}$ an occupant presence sequence is generated with the High presence rate model.

A.3.3 Initial state

Since inhomogeneous processes do not have steady state properties, it was decided to base the initial conditions on the expected occupancy presence at the start of the simulations (6 a.m.). The expected presence was estimated for the HPR and the LPR independently as the mean presence of the occupants in the data sequences for the HPR and the LPR group respectively. A Bernoulli ex-
Table A.2  *Overview of the partitioning of the occupant presence sequences.*

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of sequences</th>
<th>Mean</th>
<th>Variance of mean of sequences</th>
<th>Min. mean of sequences</th>
<th>Max. mean of sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPR</td>
<td>571</td>
<td>0.521</td>
<td>0.018</td>
<td>2.05 \cdot 10^{-1}</td>
<td>0.836</td>
</tr>
<tr>
<td>LPR</td>
<td>260</td>
<td>0.031</td>
<td>0.002</td>
<td>0</td>
<td>0.197</td>
</tr>
<tr>
<td>Outliers</td>
<td>80</td>
<td>0.672</td>
<td>0.024</td>
<td>3 \cdot 10^{-3}</td>
<td>0.964</td>
</tr>
<tr>
<td>Total</td>
<td>911</td>
<td>0.394</td>
<td>0.068</td>
<td>0</td>
<td>0.964</td>
</tr>
</tbody>
</table>

The experiment was then carried out to start each simulation in either “absence” or “presence” for each simulation. The mean value of this Bernoulli experiment was either $\hat{p}_{0,\text{LPR}} = 0$ or $\hat{p}_{0,\text{HPR}} \approx 0.045$.

**A.3.4 High occupancy rate model**

Two different events must be described, namely the transition from absent (0) to present (1) and from present to absent. Different models will be applied, their performances assessed, and the best one will be picked.

For every two-minute interval, the conditional probability of a transition to 1, given that 0 is observed was estimated. This is an estimate for a time of day, $n$. These local estimates are shown as points in Figure A.9. Also fits of generalized linear models with splines of 11 knots (10 basis splines) and different exponential smoothing levels are shown. The range of the exponential smoothing is the range that the model can take for this data. Because of the ten observations of absence which will always follow a sequence of presence, this interval is $[0, 0.102]$. The tendency to start working is small at 6 a.m. and it only slowly increases the first hour. Then, from 7 a.m. to 9 p.m., this tendency grows rapidly. The growth is then slower but persists until around 10.30 a.m. where it starts decaying from about 7%. A “valley” is then seen over lunch time at around twelve. The global maximum is seen just before 2 p.m. after which it drops for a small valley before a local peak at 4 p.m. From there, it drops again and approaches zero at 7 p.m.

Similar estimates of local conditioned probabilities have been made for transitions from presence to presence. These are shown in Figure A.10 together with generalized linear models based on 8 knots (7 basis splines) and exponential smoothing at different levels. The smoothing levels here range from 0.186 corresponding to the lowest possible level given that the process is in “presence” to 1 – corresponding to having been in present in all history. The main tendencies
are that given the smoothing level, the tendency to remain at one’s work desk is quite constant except for during lunchtime and after around 3 p.m. where it drops quickly.

The decision on a model structure is based on BIC. BIC values for the different models applied are plotted in Figure A.11. A large gain is seen in going from using homogeneity or a 1st order polynomial to at least a third order polynomial or splines. The increase in BIC between these models could be because of a suboptimal positioning of the knots. The exponential smoothing improves all the models implemented measured on BIC. The best model is found to be based on a spline with 11 knots and the exponential smoothing. This gives 13 parameters in total. Table A.3 shows the parameter estimates in the chosen generalized linear model of the probability of occupancy at time $n + 1$ conditioned that an employee is idle at time $n$.

Using likelihood-ratio tests, it was checked that all parameters in this model are significant (p-values shown in Table A.3). The exponential smoothing parameter is 0.205. The glm parameter estimate related to the exponential smoothing is 8.4. Since there will never be a switch back from 0 to 1 after less than 10 zeros, the exponential smoothing level cannot exceed $1 \cdot (1 - 0.205)^{10} \approx 0.1$.

The same analysis has been carried out for modeling the probability of occupancy at time $n + 1$ given that the employee is occupant at time $n$. The improved model found here is a generalized linear model with an intercept and
Figure A.10 Two-minute estimates of the probability of presence for an employee at the next time step given that he or she is present. The smoothing level ranges from 0.186 to one, with increasing probability of staying work.

Table A.3 Parameter estimates in the HOR model of transitions from occupant absence to presence, their confidence intervals, and p-values for the test of the hypothesis that the individual parameters are zero.

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>2.5 %</th>
<th>97.5 %</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-4.56</td>
<td>-4.83</td>
<td>-4.31</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.88</td>
<td>1.58</td>
<td>2.19</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.50</td>
<td>1.08</td>
<td>1.92</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>2.03</td>
<td>1.65</td>
<td>2.41</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.71</td>
<td>0.34</td>
<td>1.09</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>2.73</td>
<td>2.35</td>
<td>3.11</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.55</td>
<td>0.13</td>
<td>0.97</td>
<td>0.010</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>2.24</td>
<td>1.84</td>
<td>2.64</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_8$</td>
<td>-0.78</td>
<td>-1.19</td>
<td>-0.38</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_9$</td>
<td>-0.77</td>
<td>-1.44</td>
<td>-0.09</td>
<td>0.027</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>-1.26</td>
<td>-1.72</td>
<td>-0.83</td>
<td>0.000</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>7.67</td>
<td>6.40</td>
<td>8.93</td>
<td>0.000</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A.3 Results

Figure A.11 BIC for different models applied to the transitions in the high presence rate part of data. 0 represents absence, 1 presence in legends.

Table A.4 Parameter estimates in the HOR model of transitions from occupant presence to occupant presence, their confidence intervals, and p-values for the test of the hypothesis that the individual parameters are zero.

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>2.5 %</th>
<th>97.5 %</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>1.76</td>
<td>1.38</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.66</td>
<td>0.25</td>
<td>1.04</td>
<td>0.002</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-1.19</td>
<td>-1.68</td>
<td>-0.72</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.32</td>
<td>-0.12</td>
<td>0.74</td>
<td>0.157</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-0.08</td>
<td>-0.55</td>
<td>0.37</td>
<td>0.729</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-1.46</td>
<td>-1.78</td>
<td>-1.14</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>-2.22</td>
<td>-3.14</td>
<td>-1.34</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>-2.02</td>
<td>-2.34</td>
<td>-1.69</td>
<td>0.000</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.82</td>
<td>2.68</td>
<td>2.97</td>
<td>0.000</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5 basis spline functions. Exponential smoothing did not improve this model significantly. The resulting parameter estimates in the generalized linear model are shown in Table A.4. It is seen that some p-values (for the likelihood-ratio tests in which the parameters are zero) are large here, meaning that at least one spline basis function is insignificant. This can occur because the knot placements are not optimized but determined to be equidistant, and the number of knots is decided from BIC.
Table A.5  

Performance measures for the best of each type of the models applied on the high occupancy rate part of data. The inhomogeneous Markov chain using exponential smoothing has both the better rmse and bias.

<table>
<thead>
<tr>
<th>HOR Model</th>
<th>k</th>
<th>rmse</th>
<th>bias</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hom. MCs</td>
<td>4</td>
<td>0.146</td>
<td>3.96⋅10^{−11}</td>
<td>−22875</td>
</tr>
<tr>
<td>Inh. MCs</td>
<td>18</td>
<td>0.145</td>
<td>−1.73⋅10^{−14}</td>
<td>−21711</td>
</tr>
<tr>
<td>Inh. MCs, e.s.</td>
<td>23</td>
<td>0.144</td>
<td>9.92⋅10^{−15}</td>
<td>−21087</td>
</tr>
</tbody>
</table>

Table A.6  

Parameter estimates in the LOR model of the probability of presence at \( n+1 \) given absence at \( n \).

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>2.5 %</th>
<th>97.5 %</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>−6.69</td>
<td>−7.31</td>
<td>−6.13</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>2.03</td>
<td>1.47</td>
<td>2.63</td>
<td>0.000</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1.33</td>
<td>0.61</td>
<td>2.11</td>
<td>0.000</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>3.18</td>
<td>2.69</td>
<td>3.68</td>
<td>0.000</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>2.58</td>
<td>1.25</td>
<td>4.02</td>
<td>0.000</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>−1.93</td>
<td>−2.59</td>
<td>−1.32</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table A.5 shows an overview of the aggregated performance (all transitions – from absence as well as presence) of the best (measured on BIC) of the different types of models that were applied on the high occupancy rate data. It is seen that the inhomogeneous models outperform the homogeneous ones measured on bias and rmse, and that the exponential smoothing further increases the performance.

A.3.5 Low occupancy rate model

The same procedure as for the High occupancy rate model has been carried through to find a Low occupancy rate model. In this case, the exponential smoothing was not found significant to include in the generalized linear model. For the model of the probability of presence at time \( n+1 \) given that the occupant is idle at time \( n \), a generalized linear model based on a spline and a total of six parameters was found to perform best. The parameter estimates are listed in Table A.6.
Table A.7  Parameter estimates in the LOR model of the probability of presence at \( n + 1 \) given presence at \( n \).

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>2.5 %</th>
<th>97.5 %</th>
<th>( \text{Pr}(&gt;\text{Chi}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>-2.88</td>
<td>-4.49</td>
<td>-1.26</td>
<td></td>
</tr>
<tr>
<td>( \rho_1 )</td>
<td>0.66</td>
<td>0.39</td>
<td>0.93</td>
<td>0.000</td>
</tr>
<tr>
<td>( \rho_2 )</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.000</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.32</td>
<td>0.13</td>
<td>0.51</td>
<td>0.001</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.8  Performance measures for the applied models on the low occupancy rate part of data. The inhomogeneous Markov chain without exponential smoothing has both the smallest rmse and the smallest bias.

<table>
<thead>
<tr>
<th>LOR Model</th>
<th>( k )</th>
<th>rmse</th>
<th>bias</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hom. MCs</td>
<td>2</td>
<td>0.112</td>
<td>(-4.10 \cdot 10^{-11})</td>
<td>(-5875)</td>
</tr>
<tr>
<td>Inh. MCs</td>
<td>9</td>
<td>0.111</td>
<td>(-1.70 \cdot 10^{-5})</td>
<td>(-5734)</td>
</tr>
<tr>
<td>Inh. MCs, e.s.</td>
<td>11</td>
<td>0.111</td>
<td>(-3.43 \cdot 10^{-4})</td>
<td>(-5710)</td>
</tr>
</tbody>
</table>

For the modeling of the probabilities of occupancy at time \( n + 1 \) given occupancy at time \( n \), the chosen generalized linear model is based on a second order polynomial and no exponential smoothing. The parameter estimates are listed in Table A.7.

Table A.8 lists aggregated performance measures of models on the low occupancy rate part of data. Again the inhomogeneous Markov chains perform better when measured on bias and rmse and the performance is further improved by adding exponential smoothing. However, the latter has little effect on the Low occupancy rate data. In this model, exponential smoothing is only used on the transitions from presence, see Tables A.6 and A.7.

A.4  Simulations

The estimation was based on data from a 16-day period. The estimated models were then used to simulate a new 16-day period. These are simulations of the full system as sketched in Figure A.8 for as many occupant day sequences as available in data after omitting outliers. This corresponds to monitoring 52 employees.
for 16 days, resulting in 832 sequences in total. As in Figure A.8 each sequence is simulated with the High occupancy rate model with probability $\hat{P}_{HPR}$ (see Equation A.16), and with the Low occupancy rate model with probability $1 - \hat{P}_{HPR}$. This gave 565 sequences simulated with the High occupancy rate model and 267 simulated with the Low occupancy rate model. Once the choice between LOR and HOR has been made, the initial value of the sequence is determined by a Bernoulli experiment. In the LOR model, the mean value of the Bernoulli experiment is the average of the LOR group at 6 a.m., and for the HOR model the mean value of the Bernoulli experiment is the average of the HOR group at 6 a.m..

The upper plot in Figure A.12 shows the sequences of total occupancy versus time of day for the simulated data using the model chosen in Section A.3. This is to be compared with the plots in Figure A.5. The simulations all start with low occupancy (due to initial conditions), they have a peak before lunch, and one after. At 7 p.m. the occupant presence has dropped close to zero. This general tendency captures the tendency seen in the data very well. However, the data seems to vary slightly more, especially after the lunch break, mainly because of the two days discussed in Section A.3.1.
The lower plot in Figure A.12 shows the mean of total occupant presence over the day and an estimated confidence interval for the total simulated occupant presence. The statistics are shown for the data series, the homogeneous Markov Chain simulations (both for LOR and HOR), and the inhomogeneous MCs with and without exponential smoothing. Whereas the Markov chain due to the homogeneity does not capture the dependence of time, the two inhomogeneous models both have this ability. It is seen that the exponential smoothing does not have a big influence on the mean occupancy over the day. This is expected as exponential smoothing is a filter that influences the dynamics at per-employee level. Hence exponential smoothing is not important for mean value considerations for large systems. From the confidence intervals, it is again seen that in the afternoon, the variance in total occupancy is larger for the data than for any of the models.

The distribution of the simulated occupancy for employees throughout single days is shown in Figure A.13. This should be compared with Figure A.7. It is seen that the fitted LOR model tends to give fewer days of almost no occupancy and fewer days with occupancy over 0.1. The HOR model seems to fit the distribution in the data nicely. However, the tails of the distribution are slightly longer than what is seen in the data.
A.5 Discussion

A central assumption in this work is that the ballast status records are representative for each individual present. The validity of this assumption will depend on the office environment. But for the application of evaluating consumption which is controlled using passive infrared sensors this assumption is less important, since the data reflect activation of such sensors (apart from the delay on the turning off).

The outlined method clearly demonstrates its ability to model the variation in occupants’ transitions between present and absent. However, splines and polynomials are only examples of how this can be done. Kernel smoothing provides other methods which could also be used.

Describing the variation of the transitions over the day is one problem, describing the per sequence dynamics is another. Using exponential smoothing of the occupant presence sequences significantly improved the prediction ability of the model, especially for the High occupancy rate data. Only this one method for describing the variation was tried, and others may do just as well or better. This result however shows that there is a need for modeling the per sequence dynamics if reliable single-occupant sequences are wanted.

The idea of using exponential smoothing comes from reading the work of (Zucchini et al. 2008) where exponential smoothing of observations is used as input in the model of the transition probabilities in a hidden Markov chain. Such a model was also tried on this data, but the state dependent distributions turned out to be Bernoulli distributions with practically certain success. This means that the Markov chain is practically not hidden. Hence the idea of directly observing the Markov chain. However, this limits the model framework to only two states. It is possible that for other data sets, more (hidden) states would give a better fit. Such states could be interpreted as “meeting”, “short break”, “gone home”, etc. These would then tend to lead to absence of different lengths.

This method could be directly used to simulate occupant presence profiles in a building simulation program. However, more data must be analyzed in order to provide good standard values for different kind of office environments and other uses.

In general the choice of model depends on the data set. Before more general conclusions can be drawn on the subject, different data sets from different sources must be analyzed.

Only independent single-occupant profiles were fitted and simulated. Correla-
A.6 Conclusions

Occupant presence patterns for employees in an office environment have been modeled based on data collected from electrical ballasts triggered by passive infrared sensors. After compensation for a delay in switching off the ballasts and removal of outliers, data was divided into “low occupancy rate” and “high occupancy rate” patterns which were fitted independently and the probability of activation of the two resulting models was estimated.

By use of generalized linear models based on natural splines and exponential smoothing of observations, the daily patterns were fitted. By use of the fitted models, new occupant presence patterns were simulated, and they demonstrated similar mean occupancy over the day, and the distribution of the occupancy per day had the same two-peak property as the data. The mean occupancy per versus time-of-day fit using homogeneous Markov chains did not capture the two-peaks tendency with a drop around lunch time and the drop in the afternoon.

While using exponential of the observations as a covariate in the Markov chains did not seem to have any large effect on the dependency of the time of day, it significantly improved the one-step predictions. This is thought to reflect an improved model of the dynamics of the sequences.

The outlined method can be used for generating reliable occupant presence sequences and can be included in building simulation tools. Some objectives for further studies of the subject were given, and they include modeling of modeling of data from different environments, and modeling of correlation structures between occupants and/or between days.

Acknowledgments

The authors would like to thank Francis Rubinstein and his research group at Lawrence Berkeley National Laboratory for giving us access to the measured data.
References


An Arctic Low-Energy House as Experimental Setup for Studies of Heat Dynamics of Buildings

Authors:
Philip Delff Andersen¹, Carsten Rode², Henrik Madsen¹.

In press at:

¹DTU Compute, Matematiktorvet, Building 322, DK-2800 Kgs. Lyngby, Denmark
²DTU Civil Engineering, Brovej, Building 118, DK-2800 Kgs. Lyngby, Denmark
Abstract

This paper addresses the difficulties in pinpointing reasons for unexpectedly high energy consumption in construction, and in low-energy houses especially. Statistical methods are applied to improve the insight into the energy performance and heat dynamics of a building based on consumption records and weather data. Dynamical methods separate influences from outdoor temperature, solar radiation, and wind on the energy consumption in the building. The studied building is a low-energy house in Sisimiut, Greenland. Weather conditions like large temperature differences between indoors and outdoors throughout long winters, strong winds, and very different circumstances regarding solar radiation compared to areas where low-energy houses are usually built, make the location very interesting for modeling and testing purposes. In 2011 new measurement equipment was installed in the house, which will be used to develop more detailed models of the heat dynamics and energy performance in relation to different meteorological variables, heating systems, and user behavior. This type of models is known as a grey-box model and is been introduced in this paper.

Keywords
Low-energy houses, Heat dynamics, Arctic climate, Statistical modeling, Gray-box modeling

B.1 Introduction

Increasing consciousness of the impact of human activities on the global environment – namely the consequences of emissions of greenhouse gases – has lately led to political goals of lowering energy consumption and switching to more sustainable energy supply systems. In buildings, this calls for improved methods for assessment of energy performance and characterization of heat dynamics.

In order to lower costs of collecting digital consumption data and for the consumer to be able to monitor his or her consumption pattern and maybe even adapt to price fluctuations, online data collection devices such as “Smart Meters” are getting increasingly common in dwellings. They typically monitor and log at least one consumption variable and possibly indoor climate variables. Already in (Westergren et al. 1999) a framework is developed to estimate physical parameters of buildings based on weather and consumption data, and the energy consumption is modelled for a sample of buildings. In (Mortensen and Nielsen 2010) rather simple methods are presented on how to estimate UA-values, gA-
values and sensitivity to wind speed of buildings using only consumption and weather data. In (Bacher et al. 2013) simple lowpass filters are applied to inputs and outputs in order to obtain reliable predictions of heat consumption in buildings on time intervals down to one hour.

Discrete-time dynamical models have a large potential for use on automated and standardized measurements. ARMAX (Autoregressive Moving Average with eXogeneous inputs) models are a wide class of dynamical linear models. (Norlén 1990) implements a recursive algorithm to estimate the UA-value of a test cell with ARMAX models, and in (Jiménez et al. 2008a) ARMAX models are used on data from a test wall.

More detailed information about the heat dynamics of a building can be achieved by applying continuous-time models such as grey-box models to data of higher resolution. This has been done for a part of a highly insulated building in (Madsen and Holst 1995). In (Andersen et al. 2000) it was applied on a multi-room building, and (Bacher and Madsen 2011) presents a method for a consistent model selection procedure.

In 2005, a low-energy house was inaugurated in Sisimiut, Greenland. The objective was to build a house with very low energy consumption for heating, which should inspire the development of energy-efficient housing in Greenland and demonstrate the potentials for energy efficiency in a house which should also be a leading example of good indoor thermal environment. The house and its objective was also to be presented in (Norling et al. 2006). Therefore the current paper will only briefly introduce the building and then focus on how well the house has lived up to its performance targets, and which challenges it has incurred. Some preliminary performance results were presented also in (Rode et al. 2009), but significant improvements have occurred since then. The statistical analysis will be performed on data from before and after the work was conducted on the building and the results will be compared.

Apart from improvements on the energy performance of the envelope, the building has been equipped with numerous sensors and control equipment for conducting experiments in the building. Long winters of low outdoor temperatures give a high signal/noise ratio and ease planning of experiments. The modern design of the building with a high level of insulation, large window areas, and floor heating, makes it interesting for studies of heat dynamics of modern low-energy construction. Moreover, it consists of two symmetrical apartments which enables studies of the influence of occupancy.

This paper presents results of statistical modeling of historical consumption data from the house in order to quantify the alleged improvement of the building envelope. It also describes the new measurement setup, and finally presents
suggestions for obtaining more detailed heat dynamics models. The paper is structured in the following way: Section B.2 gives a brief presentation of the low-energy house, Section B.3 describes a statistical methods for analysis of data before and after the repair work on the building. Section B.4 lines out plans for future experiments and analysis, and finally conclusions are given in Section B.5.

B.2 Description of the house

A target for the house was that the energy consumption for heating and ventilation should be only half of that permitted by the 2006 version of the Greenlandic Building Regulations: 230 kWh/m$^2$/yr (Government of Greenland 2006). Furthermore, considering that the house was planned to have a ventilation system with heat recovery – something that was not assumed for residential dwellings in the building regulations – the target value 80 kWh/m$^2$/yr was chosen. Building energy simulations were executed to substantiate that this level of annual energy consumption was possible. The means to reduce the energy consumption in comparison with common Greenlandic houses have been to use extra insulation in floors, exterior walls and the roof. Advanced windows have been used with low energy glazing using normally 3 layers of glass. A solar collector has been installed on the roof for domestic hot water heating. The house has been orientated to exploit the light and its geometry optimizes the daylight absorption. The ventilation system is supplied with a counter-flow heat exchanger that uses the warm exhaust air to preheat the cold inlet air. Sisimiut is the second largest city of Greenland (5500 inhabitants) located on the west coast just 42 km north of the Polar Circle. The mean average temperature is around 6 °C in summer and around −13°C in the winter months. The number of heating degree days is around 8000 K-days (base 19°C). The house is approximately 200 m$^2$ and is made as a semi-detached house, where the two living areas are built on each side of the boiler room and an entrance hall. Figure B.1 shows a picture of the house, and Figure B.2 shows the cross section and floor plan. One of the two dwellings serves as home for a family, while the other is used as a guest house for visitors and for research experiments.

B.2.1 The building envelope

The building is generally made as a wood frame construction. The inhabited part is all on one floor, which is distributed over two slightly displaced levels, and there is a cold attic above the whole building, and an open crawl space
B.2 Description of the house

The heat loss due to thermal transmittance of the building envelope constructions is kept at a minimum by using large insulation thicknesses and wooden posts and girders in two separate layers that do not touch each other, so thermal bridges are practically eliminated, see Table B.1 and Figure B.3. As it can be seen from Table B.1 all the constructions have U-values below the demands.

Figure B.1  Photo of the low-energy house in Sisimiut as seen from the west.

Figure B.2  Cross section and floor plan of the low-energy house. The house is built as a double house with common scullery/boiler room and entrance hall.
Table B.1  Calculated U-values of the different constructions compared with the demands from the Greenlandic Building Regulations (GBR). The values include thermal bridge effects.

<table>
<thead>
<tr>
<th>Construction</th>
<th>Floor</th>
<th>Walls</th>
<th>Roof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulation thickness [mm]</td>
<td>350</td>
<td>300</td>
<td>350</td>
</tr>
<tr>
<td>U-value calculated [W/m²/K]</td>
<td>0.14</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>U-value GBR 2006 [W m⁻² K⁻¹]</td>
<td>0.15</td>
<td>0.20</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Figure B.3  Wood based structural members near a corner of the building are configured such that thermal bridges are avoided. To the right: A plot of the calculated temperature distribution around the corner. The calculated linear thermal transmission coefficient is $\psi = 0.015$W/(m K).

B.2.2 Windows

Three types of glazing units are used in the low-energy house:

**Type 1** 1+2 solution: Made of one single glass layer with a hard low-emission coating and a sealed unit with two glass layers.

**Type 2** Combined double energy glazing and a vacuum glazing unit.

**Type 3** 2+1 solution: Made of a sealed unit with two layers of glass and a separate single layer of glass with a hard low emission coating.

The three types of glazing units are shown in Figure B.4. Data for the glazing units are shown in Table B.2. The net energy gain is calculated as a mean value
Table B.2  *Heat transmission coefficient* (\(U_g\), \(U_w\)), *solar energy transmission* (\(g_g\), \(g_w\)) and *net annual energy gain* (\(Q_g\), \(Q_w\)). Index \(g\) for glazing and \(w\) for window.

<table>
<thead>
<tr>
<th>Index</th>
<th>1: 1+2</th>
<th>2: 2+Vac.glaz</th>
<th>3: 2+1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(U_g) [W/(m(^2)K)]</td>
<td>0.7</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>(g_g) [-]</td>
<td>0.45</td>
<td>0.40</td>
<td>0.56</td>
</tr>
<tr>
<td>(Q_g) [kWh/m(^2)]</td>
<td>172</td>
<td>136</td>
<td>228</td>
</tr>
<tr>
<td>(U_w) [W/(m(^2)K)]</td>
<td>1.0</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>(g_w) [-]</td>
<td>0.30</td>
<td>0.27</td>
<td>0.47</td>
</tr>
<tr>
<td>(Q_w) [kWh/m(^2)]</td>
<td>−17.3</td>
<td>−59.3</td>
<td>67.1</td>
</tr>
</tbody>
</table>

of windows oriented north, east, west and south for a reference house.

![Figure B.4](image)

*Figure B.4*  *The three glazing units used in the low-energy house.*

### B.2.3 Heating system

The low-energy house is constructed with a hydronic floor heating system based on PEX-tubes installed in aluminum plates just below the wooden floor boards. The floor heating system in the bathrooms is based on PEX-tubes cast in the concrete. The ventilation system is equipped with a heating coil which is positioned in the supply air duct after the heat exchanger. The heating coil is meant to ensure that the air supply is at a minimal temperature of 18°C. The ventilation system’s heating coil is based on the same hydronic system as the floor heating.

Hot water for the floor heating and heating coil is supplied from an oil furnace,
which is located in the boiler room of the house. Heat for the domestic hot water comes from a solar collector. The oil furnace supplies back up heat in periods when the solar heating is insufficient. Finally, a radiator in the entrance hall is meant to be heated with excess heat from the solar collector system when available.

B.2.4 The ventilation system

Mechanical ventilation with heat recovery in cold climates can present problems with ice formation in the heat exchanger. When warm humid room air is brought in contact with the cold surfaces of the exchanger (cooled by the outside air), the moisture in the exhaust air condenses in the heat exchanger. If the outside air is below freezing, the water vapor will freeze, resulting in a larger air flow resistance on the exhaust side of the exchanger, which in turn decreases the air flow. The decrease in the amount of warm air through the exchanger will result in the exchanger being cooled further, and eventually the system will become fully blocked with ice and stop. This problem can be prevented by preheating the inlet air before it reaches the exchanger. This will however result in extra energy consumption and higher installation costs, and is therefore not an optimal solution.

A new design of a heat recovery unit was developed for the low-energy house in Sisimiut in cooperation between EXHAUSTO A/S and the Technical University of Denmark. The dimensions of the unit are: Length 1,760 mm, width 930 mm and height 660 mm. The unit consists of two highly efficient aluminum counter flow heat exchangers coupled in a serial connection. A damper is able to switch the air flow direction through the units. When ice formation starts to reduce the air flow in the coldest exchanger, the air flow direction is switched. The exchangers, damper and filters are mounted in a cabinet with 50 mm insulation, although the unit is recommended to be placed in a heated place to minimize risks of frost damage from the condensing water. A diagram of the system is shown in Fig. 5. The theoretical temperature efficiency of the heat recovery unit is approximately 90%.

B.2.5 The solar collector

Solar hot water panels installed on the low-energy house constitute a flat plate collector. It has a total surface area of 8.1 m² and the system is able to collect 1700 kWh/yr. This covers approximately 57% of the hot water consumption of the house. The house and its inhabitants use around 150 L of hot water per day.
The solar collector faces south-east and is tilted 70° from horizontal to have the optimal position in relation to the sun.

### B.3 Linear modeling of existing data

Since the completion of the house, consumption and some indoor climate variables have been measured. The consumption recordings consist of common oil consumption and electricity consumption recordings for each apartment and for common areas. For the ventilation and heating systems, all inlet and outlet flows and temperatures have been measured. Moreover, measurements have been taken of temperatures and relative humidity both indoors and in some construction parts. Consumption recordings are cumulative, temperature and flow measurements are instant, and all data was logged every hour. Unfortunately, the measurement recordings have been interrupted, which limits the periods which can be used for modeling.

Two periods of approximately 2.5 months each have been chosen for analysis. The first period starts on September 1st 2009 while the second starts on February 1st 2010. Mending had been carried out between these two periods so the house was expected to perform better – namely be tighter and have a better control of the heating – in the second period.

A first comparison of the energy consumption for the two periods is seen in Figure B.6. The largest power consumption is in floor heating which has been reduced by 545 W on average or more than 15%. The ventilation heating is in general only 5 to 10% of the contribution from floor heating but it has increased by around 35%. Energy consumption for water has dropped significantly by 69%. All three electricity consumptions have increased. While for a household
all power consumption is equally interesting, in the modeling of the performance of the building envelope, hot water consumption will be left out. This is because the hot water consumed is largely assumed to be drained while still warm. Omitting heating of domestic water, the average (heating and electricity) power consumption is 158 W lower in period 2 compared to period 1, which corresponds to a reduction of 3.5%.

### B.3.1 Statistical model framework

The energy performance is hard to compare between periods because of the different weather conditions, possible differences in the use of the house etc. For this reason, a statistical model is needed to describe the influences of different variables. The variables that will be used here are indoor and outdoor temperatures, solar radiation, and wind speed. The use of the building is assumed to have been similar in the two periods. The increased energy consumption for heating of water could suggest more occupancy in the second period, though.

An average UA-value for the building is of special interest. The average UA-value is interpreted as the steady-state heat conductivity of the envelope. Using only indoor and outdoor temperature and solar radiation, this interpretation yields the following steady-state expression for the heat loss, $P_h$, through the

![Figure B.6](image-url) Distribution of the power consumption in the building in the two considered periods.
envelope:

\[ P_h = UA(T_i - T_a) - gA \cdot P_s \]  \hspace{1cm} (B.1)

However, the system is never in steady state and hence must be modeled with dynamic techniques. The UA-estimate is then derived as the steady state response of the temperature difference to the heat consumption. A dynamical relation between the heat consumption and indoor temperature, outdoor temperature, and solar radiation can be modeled with an ARMAX model of the heat transfer at time \( n \):

\[ \phi(B)P_h(n) = \gamma_1(B)T_i(n) + \gamma_2(B)T_a(n) + \gamma_3(B)P_s(n) + \theta(B)e(n) \]  \hspace{1cm} (B.2)

where \( \phi, \gamma_1, \gamma_2, \gamma_3, \) and \( \theta \) are polynomials, and \( B \) is the backward shift operator given by

\[ B \cdot x(n) = x(n - 1) \]  \hspace{1cm} (B.3)

and \{ \( e(n) \) \} is gaussian white noise. Notice that \{ \( n \) \} is discrete-time-normalized so that the sample period equals 1. The parameters in the \( \gamma_1, \gamma_2, \gamma_3 \) polynomials are any real numbers, while \( \phi_1 = \theta_1 = 1 \). For the system to be stable, the roots of \( \phi(B) \) must lie within the unit circle. See Table B.6 for nomenclature. In-depth treatment of ARMAX processes can be found in (Madsen 2008).

Since the temperature difference is included in (B.2) through both indoor and outdoor temperature, the UA-value can be estimated both as the estimated stationary response from \( T_i \) to \( P_h \) and as the negative stationary response from \( T_a \) to \( P_h \). Compare with (B.1) for the sign convention.

The discrete-time transfer function, \( H(z) \), is given in the \( z \)-domain as

\[ Y(z) = H(z)X(z), \quad z \in \mathbb{C} \]  \hspace{1cm} (B.4)

where \( Y \) and \( X \) are Z-transforms of the discrete-time processes \{ \( x_n \) \} and \{ \( y_n \) \}. For general treatment of signal processing and the \( z \)-domain, see e.g. (Oppenheim et al. 1983) and (Madsen 2008). In the (discrete) time domain, the impulse response \{ \( h_k \) \} is given as

\[ y_n = \sum_{k=-\infty}^{\infty} h_k x_{n-k} \]  \hspace{1cm} (B.5)

The steady state response is the step response for time going to infinity. Since for the step response, \( x_k = 0 \) for \( k < 0 \), and \( x_k = x \) for \( k \geq 0 \), the steady state
value of $y$ becomes

$$y_\infty = x \sum_{k=0}^{\infty} h_k$$  \hspace{1cm} (B.6)

where causality of the system is assumed ($h_k = 0$ for $k < 0$).

Since the transfer function in a causal system can be calculated from the impulse response function as

$$H(z) = \sum_{k=0}^{\infty} z^{-k} h_k$$  \hspace{1cm} (B.7)

it follows that

$$H(1) = \sum_{k=0}^{\infty} h_k$$  \hspace{1cm} (B.8)

and so the step response simplifies to

$$y_\infty = x \cdot H(1)$$  \hspace{1cm} (B.9)

Since the UA-value is the steady state extra heat transfer when the temperature difference increases by 1°C, this further simplifies to

$$y_\infty = H(1)$$  \hspace{1cm} (B.10)

in this case.

Hence – given that the considered process is stationary – the steady state value of the step response from $T_a$ to $P_h$ is given by the transfer function from $T_a$ to $P_h$ for $z = 1$:

$$\hat{U}A_{T_i} = \hat{\gamma}_2(1)/\hat{\phi}(1)$$  \hspace{1cm} (B.11)

The steady state value of the step response from $T_i$ to $P_h$ is given by

$$\hat{U}A_{T_a} = \hat{\gamma}_2(1)/\hat{\phi}(1)$$  \hspace{1cm} (B.12)

B.3.1.1 An optimal UA estimate based on both indoor and outdoor temperature

To obtain an estimate of the UA-value as the steady state response of the heat consumption to a step in difference between indoor and outdoor temperature,
the two estimates from Equations (B.11) and (B.12) must be combined. The linear combination of the two yielding the lower variance is used. This is the estimate

\[ \hat{U}A = \lambda^* \cdot \hat{U}AT_i + (1 - \lambda^*)\hat{U}AT_a \]  

(B.13)

where

\[ \lambda^* = \arg \min_{\lambda \in \mathbb{R}} \mathbb{V} \left( \lambda \hat{U}AT_i + (1 - \lambda)\hat{U}AT_a \right) \]  

(B.14)

The variance of the linear combination of the two estimates of UA is calculated and minimized. The variance is

\[ \mathbb{V} \left( \hat{U}A \right) = \lambda^2 \mathbb{V} \left( \hat{U}AT_i \right) + (1 - \lambda)^2 \mathbb{V} \left( \hat{U}AT_a \right) + 2\lambda(1 - \lambda)\text{Cov} \left( \hat{U}AT_i, \hat{U}AT_a \right) \]  

(B.15)

And the minimization yields

\[ \lambda^* = \frac{\mathbb{V} \left( \hat{U}AT_a \right) - \text{Cov} \left( \hat{U}AT_i, \hat{U}AT_a \right)}{\mathbb{V} \left( \hat{U}AT_i \right) + \mathbb{V} \left( \hat{U}AT_a \right) - 2\text{Cov} \left( \hat{U}AT_i, \hat{U}AT_a \right)} \]  

(B.16)

Notice that \( \lambda \) is unconstrained on \( \mathbb{R} \). The estimate cannot directly be interpreted as a weighted average of \( \hat{U}AT_i \) and \( \hat{U}AT_a \), since \( \lambda \) can exceed \([0, 1]\). This will happen when \( \text{Cov} \left( \hat{U}AT_i, \hat{U}AT_a \right) > \mathbb{V} \left( \hat{U}AT_a \right) \).

The variances of \( \hat{U}AT_i \) and \( \hat{U}AT_a \) and their covariance can be estimated by linearization (Westergren et al. 1999). Let \( \hat{x} \) be the estimated parameters of an ARMAX model, and \( \mathbb{V} \left( \hat{x} \right) = \hat{P} \). Then the vector of the estimates, \( \hat{U}AT_i, \hat{U}AT_a \) is given by a possibly non-linear but differentiable function, \( f \), of \( x \):

\[ f(x) = \begin{pmatrix} \hat{U}AT_i \\ \hat{U}AT_a \end{pmatrix} \]  

(B.17)

The variance-covariance matrix of the vector of estimates can be approximated by the first-order-covariance Taylor expansion:

\[ \mathbb{V} \left( f(\hat{x}) \right) \approx \left( \frac{\partial f}{\partial x} \right) \hat{P} \left( \frac{\partial f}{\partial x} \right)^T \]  

(B.18)

where \( \left( \frac{\partial f}{\partial x} \right) \) is the Jacobian matrix. While \( \hat{P} \) can be estimated from the estimation of \( x \), the Jacobian can be derived from \( f \), in this case by differentiating the expressions, (B.11) and (B.12). It is possible to estimate the variances of (B.11) and (B.12) without linearization, see e.g. Tellinghuisen 2001. The advantage of using (B.18) is that it yields the covariance between the estimates directly.
B.3.2 Deriving time constants from a linear model

Time constants summarize valuable information about the dynamics of a system. Consider the linear first order system

$$\frac{dy(t)}{dt} = ay(t), \quad a \in \mathbb{R} \quad (B.19)$$

The discrete time solution to this system is

$$y_{t+\Delta t} = e^{a\Delta t} y_t = e^{-\Delta t/\tau} y_t \quad (B.20)$$

where $\Delta t$ is the sampling period, and $\tau$ is the time constant.

For each pole, $\phi_i$, of the transfer function of an ARMAX model, the corresponding time constant can be identified by solving

$$\phi_i = e^{-\Delta t/\tau_i}, \quad \tau_i, \phi_i > 0 \quad (B.21)$$

for $\tau_i$.

B.3.3 The applied model

In (Mortensen and Nielsen 2010) the following model is used to estimate both UA and gA-values based on daily average data. Let $T_d(n) = T_i(n) - T_a(n)$:

$$P_{h}(n) = UA_1 \cdot T_d(n) + UA_2 \cdot T_d(n-1) - gA \cdot P_s(n) + c_w \cdot W_a(n) \cdot T_d(n) + e(n) \quad (B.22)$$

$n$ is the day number, and $\{e\}$ is white noise. This is clearly a sub-model of (B.2). The work in Mortensen and Nielsen 2010 is based on measurements where only the outdoor temperature was measured (and the indoor temperature was estimated). In the data from Sisimiut, indoor temperature measurements are available, however, and instead of using only the outdoor temperature, the difference between the indoor and outdoor temperature can be used.

Now let

$$WT(n) = W_s(n) \cdot (T_i(n) - T_a(n)) \quad (B.23)$$

and

$$B \cdot WT(n) = W_s(n-1) \cdot (T_i(n-1) - T_a(n-1)) \quad (B.24)$$
B.3 Linear modeling of existing data

Including an autoregressive term, this leads to the model

\[
P_h(n) + \phi_1 P_h(n-1) = \gamma_{1,0} T_i(n) + \gamma_{1,1} T_i(n-1) + \gamma_{2,0} T_a(n) + \gamma_{2,1} T_a(n-1) \\
+ \gamma_{3,0} P_s(n) + \gamma_{4,0} \cdot WT(n) + e(n)
\]  \hspace{1cm} \text{(B.25)}

where \{e(n)\} are independent and for all \(n\): \(e(n) \sim N(0, \sigma^2)\). This model will be fitted to the two periods of data. The estimates of the UA-values are calculated as in (B.13)–(B.14).

B.3.4 Data

The indoor temperature is the average of a temperature measurement placed centrally in each of the two apartments of the building. The heat input is based on an energy meter in the building measuring the energy dissipated in both floor heating and ventilation heating. The weather data is measurements from the local weather station in Sisimiut run by Asiaq (part of the Greenlandic Ministry of Housing).

The raw data has been averaged to daily values. Residuals of model (B.25) are plotted together with the averaged data for the two periods in Figure B.7. In the first period, the heat consumption is generally increasing as the outdoor temperature and solar radiation drop, whereas the opposite is the case in the spring (period 2). This is as expected. Three points from each data set were not well fitted and had too much influence on the parameter estimates (i.e. large Cook’s distances (Ersbøll and Conradsen 2005)). Hence they were considered outliers and removed before the reported estimates were calculated. These points are indicated with red circles in Figure B.7.

B.3.5 Results and discussion

Model (B.25) was fitted to the data from the two periods omitting the classified outliers, and the parameter estimates are listed in Table B.3. Supporting the model parameters themselves, estimates of UA and gA-values are listed in Table B.4. The estimates of the UA-values are calculated as in Equations (B.11)–(B.16), \(\dot{g}A = -\dot{\gamma}_{3,0}\), and a parameter is related to the sensitivity to wind speed \(\dot{c}_W = \dot{\gamma}_{4,0}\). Also, properties called called \(\dot{UA}_{\text{mean}}\) and \(\dot{UA}_{\text{max}}\) are reported. The term \(WT(n)\) in Equation (B.25) includes both \(W_s\), \(T_i\) and \(T_a\). As it contains a product of inputs, it is a nonlinear term in the model. This makes the estimate of the UA-value depend on wind speed. Hence, what is reported as UA is for \(W_s = 0\), and \(\dot{UA}_{\text{max}}\) is calculated for \(W_s = 8\) m/s. Comparing UA and
Figure B.7  Time series plots of average daily heating and explanatory variables for the first period to the left and for the second period to the right. The red points indicate outliers that have not been used for the model fits.
B.3 Linear modeling of existing data

| Parameter | Estimate | Std. Error | t value | Pr(|t|) |
|-----------|----------|------------|---------|--------|
| Period 1  |          |            |         |        |
| $\phi_1$  | -0.40    | 0.11       | 3.80    | 0.000  |
| $\gamma_{1,0}$ (W/K) | -343.43  | 170.62     | -2.01   | 0.048  |
| $\gamma_{1,1}$ (W/K) | 420.36   | 160.51     | 2.62    | 0.011  |
| $\gamma_{2,0}$ (W/K) | -57.09   | 47.00      | -1.21   | 0.229  |
| $\gamma_{2,1}$ (W/K) | -53.76   | 50.13      | -1.07   | 0.287  |
| $\gamma_{3,0}$ (m$^2$) | 0.06     | 3.23       | 0.02    | 0.985  |
| $\gamma_{4,0}$ (J/mK) | 9.07     | 2.33       | 3.89    | 0.000  |
| $\sigma$ (W) | 735.41   |            |         |        |
| Period 2  |          |            |         |        |
| $\phi_1$  | -0.29    | 0.08       | 3.58    | 0.001  |
| $\gamma_{1,0}$ (W/K) | 203.87   | 116.69     | 1.75    | 0.085  |
| $\gamma_{1,1}$ (W/K) | -105.48  | 108.30     | -0.97   | 0.333  |
| $\gamma_{2,0}$ (W/K) | -30.21   | 33.42      | -0.90   | 0.369  |
| $\gamma_{2,1}$ (W/K) | -83.81   | 35.17      | -2.38   | 0.020  |
| $\gamma_{3,0}$ (m$^2$) | -6.28    | 1.67       | -3.76   | 0.000  |
| $\gamma_{4,0}$ (J/mK) | 10.43    | 2.42       | 4.31    | 0.000  |
| $\sigma$ (W) | 658.62   |            |         |        |

$\hat{U}_{A_{\text{max}}}$ gives information about the sensitivity to wind speed, i.e. the tightness of the building. 8 m/s is chosen because it is a relatively high wind speed for both considered periods. For period 1 it corresponds to about the 0.97 quantile, for period 2, the 0.98 quantile. $\hat{U}_{A_{\text{mean}}}$ is included to give an estimate of the average heat transfer coefficient for average wind speed, 3.1 m/s.

First, the full model in Equation (B.25) is fitted to the two periods, one period at a time. The estimated physical properties for the two periods are shown in the upper half of Table B.4. The UA estimates increase from period 1 to period 2. However, notice that the $g_A$-value is $-6 \cdot 10^{-2}$ m$^2$ (not significantly different from zero) in the first period, while it is estimated to be around 6.3 m$^2$ in the second period. This could be part of the reason for the large difference in the UA estimates. In the lower part of Table B.4 a common $g_A$-value has been estimated. This is reasonable because the mending of the building did not include any changes related to the glass facades. On the other hand the surroundings may have changed, especially the reflectivity of the surroundings can have changed due to change in snow cover and vegetation. Using the common $g_A$ estimate, there is a drop from 95 to 88 for no wind, but the estimate of UA increases
Table B.4  Estimates of physical properties of the building in the two periods.

<table>
<thead>
<tr>
<th>Property</th>
<th>Period 1</th>
<th>Period 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>Std. E.</td>
<td>Est.</td>
<td>Std. E.</td>
</tr>
<tr>
<td>Individual gA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{U}A$ (W/K)</td>
<td>88.9</td>
<td>23.5</td>
<td>103.3</td>
<td>14.9</td>
</tr>
<tr>
<td>$\hat{U}A_{\text{mean}}$ (W/K)</td>
<td>116.7</td>
<td>23.5</td>
<td>135.2</td>
<td>14.9</td>
</tr>
<tr>
<td>$\hat{U}A_{\text{max}}$ (W/K)</td>
<td>161.5</td>
<td>23.5</td>
<td>186.8</td>
<td>14.9</td>
</tr>
<tr>
<td>gA (m²)</td>
<td>-0.1</td>
<td>3.2</td>
<td>6.3</td>
<td>1.7</td>
</tr>
<tr>
<td>$c_W$ (J m⁻¹ K)</td>
<td>9.1</td>
<td>2.3</td>
<td>10.4</td>
<td>2.4</td>
</tr>
<tr>
<td>$\tau$ (days)</td>
<td>1.0</td>
<td></td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Common gA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{U}A$ (W/K)</td>
<td>94.6</td>
<td>23.6</td>
<td>87.8</td>
<td>12.3</td>
</tr>
<tr>
<td>$\hat{U}A_{\text{mean}}$ (W/K)</td>
<td>119.8</td>
<td>23.6</td>
<td>121.0</td>
<td>12.3</td>
</tr>
<tr>
<td>$\hat{U}A_{\text{max}}$ (W/K)</td>
<td>160.5</td>
<td>23.6</td>
<td>174.8</td>
<td>12.3</td>
</tr>
<tr>
<td>gA (m²)</td>
<td>5.7</td>
<td></td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>$c_W$ (J m⁻¹ K)</td>
<td>8.2</td>
<td>2.2</td>
<td>10.9</td>
<td>2.4</td>
</tr>
<tr>
<td>$\tau$ (days)</td>
<td>1.0</td>
<td></td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

from 160 to 175 for wind speed $8 \text{ m s}^{-1}$. For mean wind, the increase is from 120 to 121. However, these changes must be compared with the uncertainties of the estimates. For the models using a common gA-value, neither estimates of $U_A$, nor of $c_W$ are significantly different between the two periods using t-tests. The estimated time constants are a little larger in the first period than in the second, dropping from 1 to 0.8 days.

It was tested whether the gA-value can be assumed to be the same for the two periods. The model fit is based on the assumption of gaussian noise, and hence, the test becomes an F-test. The p-value is 8.1%, which means that the difference in gA-values over the two periods is statistically insignificant. Hence, the models with common gA-values could be used.

A fundamental assumption of ARMAX models is that the residuals are independent. Therefore, the autocorrelation of the residuals must be checked. Sample autocorrelation functions of the residuals of the model with common gA-values applied on the two data sets are plotted in Figure B.8 together with confidence bands for white noise. The estimated autocorrelations are relatively small for both periods, and hence the dynamics of both data sets are well described by the fitted model.
Using this simple ARX model, none of the properties of the building are significantly different between the two periods. It should be noticed that it is difficult to compare model fits on two different data sets. However, it demonstrates consistency in the results using this model. A clear weakness of the model is that averaging to daily values reduces the number of data points dramatically.

The averaging has another side-effect that may cause problems. Figure B.9
shows average solar radiation versus average indoor temperature for the two periods. For period 2, there is a strong correlation (more than 0.9), which means that the impacts of the two are hard to distinguish. An approach to overcome such issues is to excite the system better by varying the indoor temperature independently of other inputs. Section B.4.3 introduces methods to avoid this kind of issue in future work.

B.4 Future work

B.4.1 Improved measurement and control equipment

In the spring 2011, new measurement equipment and a programmable logic controller (PLC) system was installed in the house. These facilitate online and centralized scheduling and surveillance of experiments. Air temperatures in all rooms, heating and ventilation inlet and outlet temperatures and flows are measured. Moreover open/closed sensors are installed on all exterior doors and windows, and CO₂ concentration is measured in the apartment used for experiments.

A weather station taking meteorological measurements is installed on-site. Ambient temperature and horizontal solar radiation as well as wind speed and direction are measured. An overview of the measurements most relevant to modeling the heat dynamics is provided in Table B.5. Meteorological data is also available from the governmental weather station nearby used for analysis in this paper.

With the new control system, heating and ventilation systems can be controlled based on all measurements, functions hereof, or even exogenous inputs. An overview of the state of the system is available on-line, and a screen dump of this is seen in Figure B.10. The overview intuitively shows how the different systems are connected and interact. There are two circulation systems, illustrated by different colors of the pipes. Follow the one leading from the boiler; it goes to the domestic hot water tank (if the return valve is open), to ventilation after heating, and/or to floor heating. Before it comes back to the furnace, it passes through a heat exchanger. The other pipe system goes from the solar panel. It goes to either heating the domestic water tank or to the radiator and buffer tank when a surplus of heat from the solar panel is present. The storage tank is both loaded and unloaded from the top so that a vertical temperature gradient can be maintained in the tank.
### Table B.5

The most important measurements in the house for modeling the heat performance of the envelope.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Common areas</th>
<th>Rental apartment</th>
<th>Experimental apartment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indoor temperatures</strong></td>
<td>All rooms</td>
<td>All rooms</td>
<td>All rooms</td>
</tr>
<tr>
<td>Full standard indoor temperature measurement in living room</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heating</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor heating</td>
<td>2/2</td>
<td>5/5</td>
<td>5/5</td>
</tr>
<tr>
<td>Ventilation afterheating</td>
<td>All measured together</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ventilation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central ventilation</td>
<td>All measured together</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer doors open/closed</td>
<td>2/2</td>
<td>3/3</td>
<td>3/3</td>
</tr>
<tr>
<td>Windows open/closed</td>
<td>No windows</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>Cooker hood</td>
<td>0/0</td>
<td>0/1</td>
<td>1/1</td>
</tr>
<tr>
<td><strong>Occupancy indicators</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIR sensors</td>
<td>0</td>
<td>0</td>
<td>Living room/kitchen, corridor</td>
</tr>
<tr>
<td>CO₂ concentration</td>
<td>0</td>
<td>Deactivated</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Consumption</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oil</td>
<td>All measured together</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electricity</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td><strong>PV system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total heat collection</td>
<td></td>
<td></td>
<td>Only one system</td>
</tr>
<tr>
<td>Domestic water heating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buffer water heating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heating system</td>
<td></td>
<td></td>
<td>Contributions to common system</td>
</tr>
<tr>
<td><strong>Meteorological variables</strong></td>
<td>All at one common weather station</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient temperature</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solar radiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind direction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
B.4.2 Grey-box modeling

Formulation and selection of a statistical model is an iterative process, and when modelling systems of high complexity it is often fruitful to start from a simple description and then step-by-step include new terms if they significantly improve the description of the system. An initial description of the heat loss was given in the preceding section. In order to describe the heat dynamics in more detail, grey-box models (Madsen and Holst 1995; Andersen et al. 2000; Bacher and Madsen 2011) can be applied. Grey-box modelling combines the advantages of using physical knowledge about the system with statistical methods to obtain precise descriptions of the dynamics behind measurements of a physical system. The stochastic differential equations used are based on the well-known differential equations of heat dynamics which are naturally dynamic. A very simple linear dynamical model of the indoor temperature, \( T_i \), is formulated in (Bacher and Madsen 2011):

\[
dT_i = \frac{1}{C_h} \left( \frac{T_a - T_i}{R} + A_w \cdot P_s + P_h \right) dt + \sigma_i d\omega_i(t) \tag{B.26}
\]

where \( T_i \) is the indoor temperature, \( R \) is the thermal resistance of the building envelope, \( C_h \) is the heat capacity of the building, \( A_w \) is the effective area of the windows, and \( P_h \) is the heat supply from the heating system. \( \omega_i \) is a standard Wiener process (a white noise process in continuous time), and \( \sigma_i \) is a constant. (B.26) describes the indoor temperature evolution in continuous time.

Let \( Y(n) \) be the measurement of the indoor temperature at discrete time \( n \):

\[
Y(n) = T_i(n) + \epsilon(n), \quad \epsilon(n) \sim N(0, \sigma_o^2) \tag{B.27}
\]

i.e. the measurements are encumbered with white noise, \( \epsilon(n) \). For in-depth treatment of stochastic differential equations, see (Øksendal 2007). For filtering, i.e. re-construction and prediction of the temperature based on measurements, see (Jazwinski 2007).

The test facilities available for the project are expected to enable more detailed observation of the heat dynamics. Hence a more general dynamic heat balance in the house is considered. Let \( P \) in general denote a heat flux, and the subscripts \( h, v, c, s, \) and \( i \) denote the heating system, ventilation, conduction, solar radiation, and infiltration. Then

\[
dT_i = \frac{1}{C_h} (P_h + P_v + P_c + P_s + P_i) dt + \sigma_2 d\omega_2 \tag{B.28}
\]

expresses the interior temperature development. \( \sigma_2 \) is a constant and \( \omega_2 \) is a standard Wiener process.
The conduction $P_c$ through walls, roof, doors and windows is expected to be of major importance. Let this be an example of how the model can be extended to contain more states, i.e. consist of coupled stochastic differential equations. Let it be given by a conduction from the outside surface temperature of the building, $T_o$, and the indoor temperature:

$$P_a = \frac{1}{R_{oi}C_e} (T_o - T_i)$$

The outer envelope surface could be cooled (or heated) as modeled by convection. Then that state could be written like this:

$$dT_o = \left( \frac{1}{R_{oi}C_e} (T_i - T_o) + f_W(W_s, W_d)(T_a - T_o) \right) dt + \sigma^2 d\omega$$  \hspace{1cm} (B.29)

where $f_W$ could be a non-linear function of wind speed and wind direction. In (Jiménez et al. 2008b) with a model similar to this one, $f_W$ is modeled for a PV-module with an allometric function. Further extension in the present case could include dependence on the wind direction.

This non-linear extension of the linear dynamic model is only one of many possible extensions. It has been justified from physical considerations but the
main criterion is the ability to describe the behavior of the system, i.e. what is reflected in the data. Hence, standardized model search procedures as in (Bacher and Madsen 2011) are needed.

### B.4.3 Experiments in early 2012

Estimation of parameters in grey-box modelling requires - depending on the complexity of the models - not only good models, but also good experiments. Parameter estimates can be correlated if parametrization or the experimental plan are sub-optimal. Take Equation (B.26) as an example. If $T_i$ is constant, $R$ and $C_h$ turn out not to be identifiable. Therefore, the indoor temperature has to be varied. For being able to distinguish influences from each other, the input signal (here, the floor heating) must be varied on all frequencies, which can be done using PRBS (Pseudo-Random binary signals) or ROLBS (Randomly Ordered Logarithmically distributed Binary Sequence) signals. The system is said to be excited. This was the focus of experiments that were carried out in 2012. Such input signals are said to be persistently exciting.

### B.5 Conclusions

Statistical modeling of heat dynamics is a strong tool for characterization of and improving energy performance of buildings. A framework of linear dynamic models (ARMAX) was described together with methods to extract important information about heat dynamics from model estimates.

Promising results have already been obtained by applying linear dynamic models on heat dynamics in buildings. An example study was given where 2 periods of 2.5 months each were compared for a described low energy house in Greenland. Repair work was carried out between the two periods, and improved tightness of the envelope was expected in the second period. Heat consumption was modeled using indoor temperature and weather variables, and properties of the building were estimated and compared over the two periods. From the available measurements, the expectation that repair works had made the heating consumption in the building less sensitive to wind could not be demonstrated.

However, long testing time was needed, and separating impacts of different inputs seemed to lead to problems. Test facilities in the Arctic area have been described and the advantages of these in relation to more detailed modeling have been discussed. Finally, some examples on modeling, non-dynamic and
dynamic, linear and non-linear, have been given. Experiments will be carried out to apply models of this framework on experiments carried out in 2012.
## Table B.6 Nomenclature.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi, \gamma_i, \theta$</td>
<td>Polynomials</td>
</tr>
<tr>
<td>$C_h$</td>
<td>Thermal capacity</td>
</tr>
<tr>
<td>$P_h$</td>
<td>Heat and electricity input</td>
</tr>
<tr>
<td>$P_s$</td>
<td>Solar radiation</td>
</tr>
<tr>
<td>$R$</td>
<td>Thermal resistance</td>
</tr>
<tr>
<td>$T_a$</td>
<td>Outdoor temperature</td>
</tr>
<tr>
<td>$T_d$</td>
<td>$T_i - T_a$</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Indoor temperature</td>
</tr>
<tr>
<td>UA</td>
<td>Common UA-value for the building envelope</td>
</tr>
<tr>
<td>$\hat{UA}$</td>
<td>Estimate of UA value at $W_s = 0$</td>
</tr>
<tr>
<td>$\hat{UA}_{\text{max}}$</td>
<td>Estimate of UA value at $W_s = 8 \text{ m/s}$</td>
</tr>
<tr>
<td>$Y$</td>
<td>Observation</td>
</tr>
<tr>
<td>$W_s$</td>
<td>Wind speed</td>
</tr>
<tr>
<td>$c_w$</td>
<td>A constant related to the effect of wind speed</td>
</tr>
<tr>
<td>$n$</td>
<td>Day number</td>
</tr>
<tr>
<td>$B$</td>
<td>Backward shift operator</td>
</tr>
</tbody>
</table>
References


tions - a case study” In: Research in Building Physics and Building Engineering, pp. 587–594.


Characterization of heat dynamics of an arctic low-energy house with floor heating

Authors:
Philip Delff Andersen¹, María José Jiménez² Henrik Madsen¹, Carsten Rode³.

Accepted for publication in:
Building Simulation (Springer).

¹DTU Compute, Matematiktorvet, Building 322, DK-2800 Kgs. Lyngby, Denmark
²Department of Energy, Energy Efficiency in Buildings Unit, CIEMAT, Av. Complutense 22, E-28040, Madrid, Spain
³DTU Civil Engineering, Brovej, Building 118, DK-2800 Kgs. Lyngby, Denmark
Abstract

This paper presents grey-box modeling of the heat dynamics of an apartment in a highly insulated test building located in the Arctic. Data from a 16-day-long experiment is analyzed and used to fit lumped parameter models formulated as coupled stochastic differential equations. The output of the models is the measured indoor air temperature, and the models are fitted using maximum likelihood techniques with the software CTSM-R. Models are compared using likelihood-ratio tests and validated considering autocorrelation and periodograms of residuals. The fitted models facilitate description of both the fast responses to mechanical ventilation and solar radiation through a large window facade, and the slow responses to floor heating and outdoor temperature. To successfully describe the dynamics of the system, solar radiation is given special attention in modeling of both the physical system and the observational noise. The estimated physical parameters which include UA-value, total heat capacity, and time constants for the apartment are discussed. Simulations are performed to illustrate step and impulse responses of inputs.

C.1 Introduction

In recent years, there has been increasing interest in the characterization of the heat dynamics of buildings using data. Due to generally increasing energy prices and in order to lower CO$_2$ emissions, the overall evaluation of the energy performance of buildings is of growing interest. As more detailed data from normal operation of buildings becomes available, more detailed information about the building can be derived from data-driven models. Single-number characteristics of a building’s energy performance such as an overall UA-value (the sensitivity of the energy demand to the difference between indoor and outdoor air temperature) are crucial for comparison of energy performance of different buildings or for tracking energy performance of a building over time.

Characterization of the energy performance of a building involves establishing an energy balance which is never in steady-state. This can include contributions that are measured absolutely such as a known heat input signal and some that are measured relatively, such as solar radiation where only the intensity is known but not how much the building absorbs or for estimation on parts of a building where heat transfer with other parts are significant but only temperatures are measured. In order to estimate the total heat input, one must model the system at a resolution where periods with different levels of the heat inputs are observed. E.g. for estimating the influence of solar radiation, daily resolution
will normally be insufficient. When increasing the resolution, the dynamics of
the system become more important. Hence dynamical modeling of the building
will normally be necessary in order to estimate steady-state properties such as
an average UA-value. A simple average consideration of consumed power per
temperature difference between inside and outside can be heavily biased because
of the ignored solar radiation or other important contributions.

The information obtained from dynamical modeling can be used for much more
than a characterization of the energy performance of the envelope. In general,
it can be used for control of heating systems. And also in smart grids it is
potentially very valuable. As more and more renewable energy is being imple-
mented on the supply side of the energy grid, flexibility on the demand side is
going more important to compensate for the lack of storage technologies. On
the demand side, the thermal capacity of buildings may be used for dispatching
energy when it is relatively cheap and then save on heating when the price is
higher (Corradi et al. 2013).

Several methods and models for describing the thermal characteristics of build-
ing components exist. Many of them have been developed and tested during a
number of European research projects focusing on outdoor testing under real
weather conditions, the first being the PASSYS project (Cools and Gicquel
1989). The European PASLINK Group (1990-2005) carried out a number of
EU projects on empirical test procedures for different types of building compo-
nents under real weather conditions applying dynamic analyses based on time
series analysis. A special issue of Building and Environment published in 2008
(vol. 43, issue 2) summarized the activity of this group during the previous
10 years. Some of the papers in this issue are fully focused on data analysis
(Jiménez and Madsen 2008; Jiménez et al. 2008a; Gutschker 2008). Some of
these experiences pointed to the relevance of a proper modeling of some the ef-
teffects caused by solar radiation as one of the key factors to the analysis (Jiménez
et al. 2008c). This reference also reports a systematic procedure based on phys-
ical considerations, to build candidate models and select the most appropriate
model to estimate the required parameters.

Estimation of a UA-value of whole buildings has lately been performed based
on different data-driven models in works such as (Mortensen and Nielsen 2010)
and (Bacher and Madsen 2011). In (Mortensen and Nielsen 2010), UA and gA
values are estimated for 57 households based on smart-meter readings of en-
ergy consumption, weather variables, and basic knowledge about the individual
houses. The results are obtained with daily averages using first-order autore-
gressive models. Using higher sample frequency, dynamics are estimated by
estimating simple transfer function models.

The use of grey-box models for modeling of heat dynamics of buildings was
explored and successfully reported already in (Madsen and Holst 1995) and (Andersen et al. 2000). (Bacher and Madsen 2011) outlines a method to choose a model structure using grey-box models or stochastic differential equations in a hidden Markov model framework. Through a forward selection method, a given model structure is justified, facilitating detailed description of the dynamics of a building. The method is demonstrated on a relatively light and poorly insulated building in Denmark. This type of models provides a dynamical description of the dynamics that is closely related to physics yielding easily interpretable parametrization. At the same time, statistical methods are used in the model development to reach a data-driven and simple model.

In the present work, an apartment in a modern low-energy building with floor heating and mechanical ventilation is modeled, see Figure C.1. The building features relatively slow heating from floor heating in a wooden floor and fast heating through ventilation. Furthermore, the building has a large south-east-facing glass facade which makes the effect of solar heating heavily dependent on the angle of incidence. The modeled apartment is adjacent to rooms with which heat transfers have to be taken into consideration. It turns out that models with multiple time constants are required to model the system.

To account for the fast response to solar radiation due to the window facade, a projection of the solar radiation is used. Including ventilation heating, floor heating, heat exchanges with adjacent rooms, and effects of wind speed, the number of possible model formulations is huge. Therefore, a rigorous model-building algorithm is outlined and followed.

The resulting model has three time constants. A systematic prediction error related to the fast response to solar radiation when the radiation hits directly onto the glass facade of the building is investigated. Scaling of the observational noise with a projection of solar radiation, the autocorrelation of standardized residuals is reduced significantly, clearly indicating an improvement. This avoid the deficiencies in the model of solar radiation having too much influence on the estimate of parameters of special interest, such as thermal resistances.

The main contributions of this paper lie in the demonstration of a method to construct and validate a model of the heat dynamics of a building or part of a building leading to estimates of physical parameters with relatively low uncertainty. It is demonstrated on a modern low-energy building with complex response to solar radiation and floor heating. The use of the variable observational noise variance demonstrates how modeling of the physical system can be combined with the strengths of data analysis. Moreover, the dynamics of the system as fitted from data are demonstrated through simulations.

The paper is structured in the following way. Section C.2 briefly introduces
the reader to the building used for experiments, Section C.3 features a short introduction to the framework of grey-box models used, an overview of the inputs used in the modeling, and the method used to build the models. The design of the experiment is outlined in Section C.4, and Section C.5 provides an overview of the collected data. The model building and discussion of the accuracy of the models obtained is in Section C.6. This is followed by a general discussion of the validity of the models in Section C.7 and conclusions in Section C.8.

C.2 Description of the building

The building used for the experiments is a semi-detached one-story low-energy building in Sisimiut, Greenland, see Figures C.1 and C.2. The two apartments are symmetrically aligned around common spaces of an entrance hall, a scullery, and in the center of the building a small room where the boiler and technical equipment are placed. In addition, the building has a cold attic. See Figure C.3 for an overview of the area of the building most relevant to this study. The gross floor area of the building is 197 m², and the apartments have a floor area of approximately 85 m² each. One of these apartments has been used for the experiment.

![Image](image)

**Figure C.1** The building used for experiments seen from the south. Notice the relatively large window area towards the valley (south-east) compared to the small windows on the side of the house. On the roof, the weather station is seen.

The construction of the house is mainly a wood frame which is partly suspended. The building is well insulated with 300 mm mineral wool in walls and 350 mm
mineral wool in roof and floor, and the wood frame is designed to avoid thermal bridges. U-values for walls, roof, and floor are calculated to lie in the range of $0.13 - 0.15 \text{W/m}^2\text{K}$. The air tightness of the apartment used for the experiment presented in this work was estimated using a blower-door test in 2010. The estimate was $2.59 \text{l/(s m}^2\text{)}$.

Windows are energy-efficient with stated U-values in the range $1.0 - 1.1$ and g-values in the range $0.27 - 0.47$. A large floor-to-ceiling window area is oriented towards the south-east whereas the rest of the window areas in the apartments are relatively small. The orientation of the main window facade is $124^\circ$ clockwise from north.

The building is equipped with water/glucola-borne floor heating and mechanical ventilation. The floor heating system has 12 strings in parallel, 5 for each apartment, and 2 for the common areas. Flows through and return temperatures from each of these strings are measured separately. The ventilation flows and temperatures are measured for the whole building. The building and its components are described in more detail in (Norling et al. 2006; Andersen et al. 2013c).

Heating and ventilation are controlled by a central programmable logic controller.
(PLC) which also collects all measurements. Apart from numerous measurements inside the house, the PLC is connected to a weather station on the roof, measuring outdoor temperature, global horizontal solar radiation, wind speed, and wind direction. The PLC facilitates programming of all controlled valves etc., making the building very well-suited for experiments. The data acquisition system is documented in (Andersen et al. 2013e).

![Figure C.3](image.png)

**Figure C.3** Sketch of the floor plan of the experimental apartment. The turquoise area is the area modeled. The arrow at the entrance indicates north. The red dots indicate positions of thermometers. The large red dot in the center of the experimental apartment indicates the position of a column of six high-resolution thermometers. The thick black lines represent insulated walls.

### C.3 Methods

The aim is a model of the heat dynamics of parts of the building driven by data from experiments with an unoccupied apartment. In the formulated models weather measurements, heat supplies and temperatures of adjacent rooms are known inputs, and temperature measurements in the building provide a measured output. Not measured temperatures in other parts of the building or the construction will be modeled as unknown states which can be estimated. This
section describes the type of models considered, the iterative model-building process, and the inputs that are considered.

An important part of the heat input to modern buildings is internal gains due to occupancy. The focus in this work is rather the heat dynamics of the building when known inputs are known, and internal gains are (almost) avoided. Modeling of presence of occupants is studied in (Andersen et al. 2014a).

C.3.1 Grey-box models and estimation

The models that will be used are lumped parameter models formulated as a continuous-discrete time stochastic state space model consisting of a set of coupled stochastic differential equations describing the physical system and an observation equation describing the measurements. For an in-depth description of stochastic differential equations, see (Øksendal 2007). For continuous-discrete time state space models, see (Kristensen et al. 2004b).

First, consider a single-state model, and let $T_i(t)$ denote the state at time $t$. $T_i$ represents a common temperature for a whole building mass and indoor air considered. In this work, states are always temperatures. $U(t)$ is the input vector, $\theta$ a parameter vector, and $f_i, \sigma_i, h$ functions on these. Then the considered continuous-discrete time stochastic state space model is given by

$$dT_i = f_i(T_i(t), U(t), t, \theta)dt + \sigma_i(t, \theta, U(t))d\omega_i$$  \hspace{1cm} (C.1a)

$$Y_k = h(T_i(t_k), U(t_k), t_k, \theta) + e_k$$  \hspace{1cm} (C.1b)

$$e_k \sim N(0, \sigma^2_o(U(t_k), t_k, \theta))$$  \hspace{1cm} (C.1c)

where $\omega_i$ is a standard Wiener process, $k$ is a discrete sampling time, $Y_k$ is shorthand for the observation at time $t_k$, and $e_k$ is the observation noise at time $t_k$. $f_i(...)$ is called the drift term for the state $i$, while $\sigma_i(...)$ is called the diffusion term. $\sigma^2_o(...)$ is the observational variance. This is a very flexible model formulation build upon the both classical dynamical ODE-modeling and stochastic modeling. The two noise processes can depend on time and inputs. However, the noise processes are limited not to depend on the state. This is due to a limitation in the software that will be used for the parameter estimation. The observational noise is uncorrelated, i.e. expresses measurement errors. The system noise goes through the system leading to a dynamical response. This could be inputs that are not included in the model or noise on measurements on the inputs or other effects on the states that are not included in the drift term. Hence this model formulation facilitates estimation of the variance on both types of noise.
This simple model is illustrated in Figure C.4, where expressions are given for different inputs to the system. These expressions will be explained in Section C.3.3. The model can be extended by partitioning the building mass into distinct parts, such as indoor air, building envelope, floor, etc. How the different inputs affect the different states, and how the states interact is defined by introducing additional (coupled) stochastic differential equations in the system equation (C.1a). The number of observation does not have to increase with the number of states, typically only a linear combination or just one state is observed. In this work, only one state will be observed. This is expanded on in Section C.3.3.1.

**Figure C.4** Illustration of the heat flows in a simple model of the indoor air temperature. The single state, $T_i$, is marked in blue, inputs in black, and parameters to be estimated in red. The building is partly suspended as in the illustration. The part of the building used for experiments was unoccupied throughout the experiment.

Maximum likelihood estimates of such grey-box models can be calculated based on the extended Kalman filter with the software CTSM. The mathematical routine is described in (Kristensen et al. 2004b; Kristensen and Madsen 2003). An R interface has been developed (Juhl 2011) making scripting possible and with parallelization increasing the calculation speed significantly. This implementation is called CTSM-R$^4$.

$^4$[http://ctsm.info](http://ctsm.info)
Each system equation represents the dynamic heat balance equation for a state. Lumped parameter models assume by nature homogeneous temperature in each state. The aim is to get a physically feasible and statistically well-performing model description with parameters containing valuable information about the heat dynamics of the building. The criterion for the model to perform statistically well is that the prediction errors are white noise, especially that they are uncorrelated.

C.3.2 Forward selection, likelihood ratio tests

A forward selection procedure is applied in order to build models to efficiently describe the variation of the data given the heat flux terms considered. In (Bacher and Madsen 2011), the simplest feasible single-state model was formulated and iteratively extended with the more significant of a predefined set until a saturated model was obtained. The level of significance of a given term is determined by the \( p \)-value of the likelihood-ratio test of including it (Madsen and Thyregod 2011). The smaller the \( p \)-value, the greater the significance. In this work, a confidence level of 5\% is used, meaning that an extension is considered significant if the \( p \)-value is less than or equals 5\%. If all significant terms are included, the model is said to be saturated.

As we shall see in the following, there are different heat fluxes about which different hypotheses can be formulated, each leading to different feasible approximations using the available inputs. Hence many different combinations of the inputs in different states are possible. Therefore, the forward selection is performed from models restricted to having a certain number of states where heat flux terms are added sequentially until the model is saturated. This is repeated for models with increasing number states until the model performance is considered satisfactory. The procedure can be summarized as follows:

- Define a finite number of heat flux terms to consider.
- Keep simple assumptions about system and observation noise, i.e. keep variances constant.
- For a given number of states, the full model is given by linear combinations of all the defined heat flux terms in all states.
- Fit a simplest feasible single-state model. Add the most significant terms sequentially until no term is significant.
- Until the dynamics in data are sufficiently well described:
– Add a state to former starting point
– Add the most significant terms sequentially until no term is significant. All terms should be tried in all states.

By looking at the autocorrelation and the cumulative periodogram of the residuals it can be determined if the dynamics are sufficiently well described. At some point other steps may be better than adding another state. In each step one must perform residual analysis and consider the autocorrelation of residuals. In this process, it may also show that transformations of inputs or extra inputs are needed to get a better fit. For forward selection and likelihood ratio tests, see (Bacher and Madsen 2011) and (Pawitan 2001).

The drift term \( f_i \) in Equation (C.1a) contains the model of the physical system. This part of the model is constructed first under simple assumptions about the diffusion or system noise \( \sigma_i \) in (C.1a) and the observations (i.e. Equations (C.1b) and (C.1c)). Initially, it is assumed that the variances of the increments of the system noise processes are constant, and that the state \( T_i \) is measured without bias \( h(\ldots) = T_i(t_k) \) and with gaussian white noise of constant variance.

The drift term will be iteratively constructed as a linear combinations of feasible heat flux terms. Each term tried is predefined and chosen from physical considerations which are argued in the following. The following heat flux terms provide the model candidates.

### C.3.3 Scope of the model

Several inputs have been considered in the search for a satisfactory description of the heat dynamics of the considered part of the building. Both the output and the many considered inputs are briefly described in this section. The inputs terms are listed in Table C.1. Recall that Figure C.4 illustrates a heat balance with these inputs.

#### C.3.3.1 The output

The indoor air temperature measurement, \( y \), is the mean value of 6 high-resolution temperature sensors placed in a central column in the apartment (see Figure C.3).
### Table C.1 Inputs. Description and notation.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Measurements</th>
<th>Parameters in heat flux terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat losses to outdoors through the building envelope</td>
<td>$T_a$</td>
<td>$R_{ai}, R_{fi}$</td>
</tr>
<tr>
<td>Heat losses due to air leakage or wind dependent external surface coefficient</td>
<td>$T_a, W_s$</td>
<td>$c_{W_s}$</td>
</tr>
<tr>
<td>Fraction of the global horizontal solar radiation</td>
<td>$P_s$</td>
<td>$A_s$</td>
</tr>
<tr>
<td>Fraction of the global horizontal solar radiation projected on the main window surface</td>
<td>$P_{s,\text{win}}$</td>
<td>$A_w$</td>
</tr>
<tr>
<td>Floor heating</td>
<td>Floor heating flows, forward temperature and return temperatures.</td>
<td></td>
</tr>
<tr>
<td>Heat supplied by the ventilation system</td>
<td>Ventilation in/out flows and temperatures</td>
<td>$c_v$</td>
</tr>
<tr>
<td>Heat exchange with the adjacent rooms: Entrance hall, scullery</td>
<td>$T_h, T_s$</td>
<td>$R_{hi}, R_{si}$</td>
</tr>
</tbody>
</table>

An alternative approach would be to consider the principal component of the measurements, yielding the linear combination with the largest variation. A principal component analysis was performed of the temperature measurements, and it was seen that the first principal component was very close to the mean value of the six measurements. Because of the little difference between the principal component and the mean value, the mean value is chosen as the output for the ease of interpretation of the estimated parameters.

#### C.3.3.2 Heat losses through the building envelope

Since all outer windows and doors have been kept closed in the experiment, it is natural to include a heat flux term mainly based on conduction through the
envelope. In the single-state model, this heat flux becomes

\[ \frac{1}{R_{ai}} (T_a - T_i) \]  

(C.2)

Other states and inputs are introduced later on. In general, let the resistance between the states or inputs \( T_i \) and \( T_j \) be denoted \( R_{ij} \).

In the linear cases, the models are essentially RC-networks. The model formulations will be kept within the RC-network terminology as which they have been estimated as well. The overall average heat conductivity of the envelope is the inverse of the resistance between indoors and outdoors. This means however that the distribution of \( \hat{U}A = \frac{1}{\hat{R}_{ai}} \) may deviate significantly from the Gaussian distribution. This may as well lead to optimistic uncertainty estimates on \( \hat{U}A \). If estimation of the \( \hat{U}A \) is the central aim of the analysis, it is better to include \( UA \) and estimate as a parameter in the model directly.

### C.3.3.3 Solar gains

Global horizontal solar radiation is measured by a pyranometer at the weather station on the building. This signal is denoted \( P_s \). The building has most of its window area towards the south-east. A projection onto this side of the building has been calculated and denoted \( P_{s,\text{win}} \). It is based on calculations of azimuth and elevation angles. Projections have been applied to the available global horizontal solar radiation, assuming that for this measurement the contribution due to the direct radiation is predominant with regard to the rest of contributions. This approximation will be best under clear sky conditions.

Both solar radiation signals have unit W m\(^{-2}\) and scaling constants must be estimated. These constants have the unit m\(^2\) and can be interpreted as effective areas of the building receiving the radiation. The two heat flux contributions become

\[ A_s P_s \]  

(C.3)

and

\[ A_{s,\text{win}} P_{s,\text{win}} \]  

(C.4)

Models including only \( P_{s,\text{win}} \) assume that only solar radiation incident in the main window has significant contributions to the energy balance equations. Models including \( P_s \) assume that also solar radiation incident in other parts of the building envelope significantly affects the energy balance equations.

The pyranometer used has the limitation that it cannot measure direct solar radiation with an angle of incidence larger than 82 degrees. Extrapolation of the solar
radiation in this interval was attempted but the resulting signals did not show any significant improvement of the models.

C.3.3.4 Heat losses dependent on wind speed

A term depending on the wind speed in the state representing indoor air temperature is included because of possible air leakages of the building. Including it in the state representing the envelope can be interpreted as the surface properties being dependent on the wind speed.

Heat losses due to air leakage or lack of tightness of the building are thought to be dependent on wind speed and the temperature difference between outdoors and where the leakage is modeled. The heat flux is modeled simply as a constant times wind speed times the temperature difference between outside and the temperature of the state that it is affecting, i.e. if used in state, $T_i$:

$$c_{Ws} W_s (T_a - T_i) \tag{C.5}$$

Notice that this term is nonlinear, since it includes a product of two inputs and a product of an input and a state. It could be modeled more generally with a power function like $c_{Ws} W^\gamma (T_a - T_i)$ where $\gamma$ is a parameter to be estimated. This approach is used in (Friling et al. 2009). Here, the simpler approach in (C.5) is chosen due to its simplicity.

C.3.3.5 Heat supplied by the ventilation system

The ventilation system was controlled based on humidity measurements in the apartment and is considered an input to the system. Air flow rate and temperature have been measured for both inlet and outlet in the ventilation system. Based on these, the net heat contribution to the building have been estimated using

$$c_v P_v = c_v c_{air} \cdot \rho_{air} \cdot (\dot{V}_{v,in} \cdot T_{v,in} - \dot{V}_{v,out} \cdot T_{v,out})/2 \tag{C.6}$$

Inflow and outflow rates and temperatures are measured for the whole building. The division by two reflects the fact that ventilation heating only goes to the two apartments and not to the common spaces. The heat capacity, $c_{air}$ and density, $\rho_{air}$, of the air are assumed to be constant and they are taken for air at 25°C. The constant scaling, $c_v$, is estimated. Notice that $P_v$ is the ventilation heating contribution to the whole building divided by two.

The difference in ventilation inlet and outlet must be balanced by infiltration.
C.3 Methods

Introduce the combined ventilation heating and infiltration from outdoor temperature, $T_a$:

\[ P_{v,a} = c_v P_v + c_v c_{\text{air}} \rho_{\text{air}} (\dot{V}_{v,out} T_a - \dot{V}_{v,in} \cdot T_i) / 2 \]  \hspace{1cm} (C.7)

\[ = c_v c_{\text{air}} \rho_{\text{air}} \left( \dot{V}_{v,in} (T_{v,in} - T_i) + \dot{V}_{v,out} (T_a - T_{v,out}) \right) / 2 \]  \hspace{1cm} (C.8)

Since infiltration can come from both adjacent rooms and outdoors, the subscript will vary depending on the infiltration term. Hence, $P_{v,a}$ means that $T_a$ is replaced by $T_s$ in Equation (C.7).

C.3.3.6 Floor heating

Floor heating is the main heat input to the building, and dynamics due to floor heating is the main purpose of the paper. Hence, it is the input signal used for the experimental design. The heat flows through the floor heating strings have been estimated based on measurements of individual forward flows, common forward temperature, and individual return temperatures for each string. The power supply to each string is then calculated using the relation

\[ P_{h,i} = c_h \cdot \rho_h \cdot \dot{V}_{i,h} \cdot (T_{h,in} - T_{i,h,out}) \]  \hspace{1cm} (C.9)

Notice that since the floor heating is liquid borne, inflow equals outflow. All inflows come from the same pipe, so all inflow temperatures are identical.

The liquid in the floor heating system was an unknown mixture of water and glucola so the heat capacity, $c_h$, and density, $\rho_h$, are uncertain. A Nuclear Magnetic Resonance Spectroscopy (NMR) analysis of a sample of the liquid was performed at Department of Energy Conversion and Storage at Risø, DTU. The estimate should have an uncertainty of around 5%. This ratio was then used to estimate the heat capacity and the density of the mixture.

The calculated power flows in the twelve loops exceeded the energy provided by the boiler and the solar thermal system. Flow and temperature measurements on the total floor heating system (before and after the flow is split into twelve), were also available, and these matched the power from boiler and solar thermal system very well. When comparing with the individual strings, a linear relation was seen between the two estimated powers. The scaling constant was estimated to be 0.805. Hence, the total floor heating contribution to the apartment is estimated as
\[ P_h = 0.805 \sum_{i=1}^{5} P_{h,i} \]  

Due to lack of parameter identifiability the direct heat loss from the heating system to outdoors cannot be addressed with the models considered in this work. Measuring the heat flow from the floor to the ground is rather complex. A solution would be to include convective heaters in the apartment. The contribution from these heaters would then affect the indoor air temperature directly, facilitating estimation of the heat capacity of the inner parts of the building. Then it would be possible to estimate how much of the floor heating affects indoors and how much is being lost through the floor.

### C.3.3.7 Heat exchange with adjacent rooms

In all rooms with floor heating, the air temperature is measured using thermometers placed on the walls. These have a resolution of 0.5\(^\circ\)C. The adjacent rooms are used as boundary conditions with resistances of walls to be estimated. Both the temperature of the scullery and the entrance hall are used but also the average temperature of these two is used. When including this term for the scullery in state, \( T_i \), the heat flux term becomes

\[ \frac{1}{R_{si}} (T_s - T_i) \]  

The heating in the two adjacent rooms is controlled based on temperature measurements in the rooms in order to maintain a normal comfort level.

The layout of the apartment and the adjacent rooms were seen in Figure C.3. The temperature has not been measured in the boiler room but only in the scullery and entrance hall. In the model building, the scullery temperature is used for the scullery including the boiler room.

### C.3.3.8 Electricity consumption and other internal loads

The electricity consumption in the apartment was measured yielding a pulse for every kWh consumed. Based on this, the consumption was very constantly at approximately 53 W. This includes a refrigerator that was running in the apartment during the experiment and standby-consumption of consumer electronics.
This input will be included in all models and will be denoted $P_{el}$. It will be included in the description of the measured state. Because of the low power, this is considered of minor importance.

Apart from the measured electricity consumption, a small unmeasured contribution can be expected from the domestic hot water system in the apartment. However there is no hot water consumption in the apartment during the experiment, and the distance that the circulation pipes traverse the guest apartment is only a couple of meters, and it is located under the somewhat insulated bathroom floor or behind the somewhat insulated wall behind the kitchen cabinets. Mainly due to the lack of hot water consumption, this contribution can be expected to be very small and is ignored.

### C.3.4 Model validation

The iterative process of model building is closely related to residual analysis. Let the residuals be defined as one-step-ahead prediction errors, $\{\epsilon_k\}$:

$$ Y_k = \hat{Y}_{k|k-1} + \epsilon_k $$

(C.12)

The aim is a model, $\hat{Y}$, that describes the dynamics of the signal, $Y$. This means that there are no dynamics left in the model residuals, $\epsilon$. However, the size of the residual itself is not as important as the standardized residual. The standardized residual, $\tilde{\epsilon}_k$, is calculated as

$$ \tilde{\epsilon}_k = \frac{\epsilon_k}{r_k} $$

(C.13)

where $r_k$ is the prediction uncertainty for the one-step-ahead prediction of $Y_k$. If the one-step-ahead prediction errors are normally distributed with variable variance, then the standardized residuals are standard normally distributed (fixed variance $(1^\circ C)^2$).

An important measure of dynamics of a signal in the time domain is the sample autocorrelation function (SACF). If there are no dynamics left in the residuals, i.e. they are white noise, the SACF will be insignificantly different from zero for all lags. Another way to see the dynamics is by distributing the power of the signal on frequencies. If the power is uniformly distributed on all frequencies, the spectrum is constant over frequencies, and hence the cumulative periodogram is converging to a straight line. SACF and periodograms of standardized residuals can be used in this way for model validation (Madsen 2008).
C.3.5 Notes on model building

For establishing a heat balance, at least one positive heat input, and one heat loss must be included. Only two heat losses are considered, the one only depending on the difference between state and outdoor temperature, and the other loss depending on wind speed.

As seen in this section, most of the heat fluxes include parameters to be estimated. But in order to estimate the thermal capacity of the system, at least one heating signal must be known. In this system, ventilation heating is shared between two apartments, and the amount of solar radiation entering the building is unknown. Hence, the measurements of the floor heating input are important. It would be easier to measure accurately if using electric heaters. This issue is about structural identifiability which is treated for dynamic systems in (Bellman and Åström 1970). All models fitted in this work have been ensured to be structurally identifiable.

C.4 Experimental design

The modeling focuses on description of the heat dynamics of the north-east apartment in an unoccupied state. The apartment is considered as one homogeneous area, and hence all doors inside the apartment were kept wide open, all exterior doors and windows were closed, and curtains were open. The experiment was run in 2012 from February 21 at 10:45 a.m. until March 9 at midnight. This gives around 16.5 days. The limitation on the experimental period is due to the need for the apartment to be unoccupied throughout the experiment. However it is known from other studies (Madsen and Holst 1995; Andersen et al. 2000; Bacher and Madsen 2011) that about one week should be sufficient for reliable characterization.

The experiment aims at exciting the dynamics of the system (Ljung 1987). The PLC system facilitates controlling the floor heating (as well as any other controlled input) by a given signal. In the implementation used, the floor heating valves are controlled to switch between fully open and closed in intervals of 30 minutes.

A pseudo-random binary signal (PRBS) is persistently exciting (Ljung and Söderström 1983; Godfrey 1980). Hence, to excite the dynamics, floor heating is controlled with PRBS signals, switching the floor heating between zero and full power. Since the apartment is considered as a whole, all the five floor
heating loops in the apartment are controlled by the same signal, and the considered floor heating input in the modeling is the total floor heating supplied to the apartment.

Three different PRBS signals were used subsequently, first one focusing on relatively high frequencies, then two of lower frequencies. A PRBS signal has an expected value of $\frac{1}{2}$ which means that following a PRBS signal, heating will be on about half of the time on average. However, the indoor air temperature must be held more or less within the bounds of normal operation of the system to be as close to normal use as possible. In this case, the maximum power of the heating system was expected to be too small to only run half of the time and still maintain normal indoor air temperatures. Therefore, the length of the heating periods was scaled by two, giving the resulting signal an expected value of $\frac{2}{3}$. The resulting control signal will be shown together with the collected data in Section C.5.

### C.5 Data

The PLC was configured to collect all measurements every 30 seconds. Afterwards, data were averaged to 15-minute resolution. The resampling algorithm first interpolates linearly to fixed time stamps of 30-second intervals, then it takes the mean value of 15-minute intervals and assigns the value to the time in the middle of this period. Averaging has several advantages. Observational noise is smoothed out, all data series get synchronized, and since around 30 measurements are taken for every averaged data point, and the data point can still be calculated even if some points are missing. The 15-minute period is chosen because the dynamics are expected to be much slower. However, i.e. a 5 or 10 minute period could also be used. The resampled data used for modeling are shown in Figure C.5. Since the output is based on 6 temperature measurements taken every 30 seconds, each point in the output is an average of about 180 measurements. The observational variance is therefore expected to be very small.

From the passive infrared and the open/closed sensors on doors it was seen that a door had been opened and there had been activity in the apartment on one occasion. The interruption of the experiment lasted around one minute. The indoor air temperature measurements were studied around this, and conservatively 4 hours of data have been omitted starting from immediately before the interruption. This period is indicated with dashed vertical lines in Figure C.5. One can see the interruption if looking closely at the indoor air temperature measurement between the two lines. The data between the dashed lines is not
The indoor air temperature measurements range between approximately 13°C and 24°C with large fluctuations as intended with the experimental design. The scullery \(T_s\) is warm with quite stable temperatures between 22°C and 25°C. The temperature in the entrance hall \(T_h\) is characterized by diurnal variations of 2 to 4°C. The few short colder periods could be caused by the door to the outdoors being opened, and the reason why it is only seen a couple of times is likely to be caused by the 15-minute averaging.

Under the indoor temperature the resulting floor heating signal, \(P_h\), and the generating experimental design (PRBS signals) are plotted. The floor heating power follows the PRBS signal well, switching between off and maximum power which will depend on the demand from the rest of the building. When on, it delivers 3–6 kW to the apartment. The PRBS signal is shown in three different colors indicating the three different signals used as described in Section C.4.

The ventilation heating power \((P_v)\) is varying fast and is quite small compared to the floor heating with values in the range −400 to 400 W. The ventilation heating signal seems negatively correlated with the indoor temperature. As the inlet temperature is kept constant, this follows from Equation C.6.

The reason for the very low air temperatures indoors is seen in the outdoor temperature, \(T_a\) which goes as low as −30°C. The solar radiation goes up to around 300 W m\(^{-2}\). Many of the days have quite similar smooth solar radiation curves. A tendency of increasing solar radiation throughout the period is seen due to the time of year. The wind speed stays lower than 6 m s\(^{-1}\) except for March 4th where it reaches 13 m s\(^{-1}\).

### C.6 Results

#### C.6.1 Model building

In the following, one model of each number of states considered will be selected, and cumulative periodograms and SACF are shown together in Figure C.6 for easy comparison.
Figure C.5 Indoor air temperature followed by heating and weather measurements. The dashed vertical lines indicate a period disregarded because of brief activity in the apartment.
C.6.2 Correlation of inputs

The temperature, $T_s$, in the scullery turned out to be a highly significant input in the model building, and small estimates of $R_{si}$ were obtained. This is believed to be due to lack of excitation of this input. In fact, because of the heat from the boiler, floor heating was never activated in this room, and as seen in Figure C.5, $T_s$ varies little. This is an issue with practical identifiability. To overcome the problem and still be able to estimate the resistance of the envelope, the resistance to the adjacent rooms is calculated from design values which leads to the value $R_{si} = R_{hi} = 0.287\text{K}/\text{W}$. This boundary condition is used in all models.

C.6.2.1 Single-state models

To construct the simplest feasible model, a heat gain and a heat loss is needed. Only two terms represent heat losses, namely linear heat transfer to $T_a$ and the product $W_s(T_a - T_i)$, where $T_i$ is the only state in the model and $P_h$ is a directly measured heat input. The most significant heat loss was $T_a$ (log likelihood equaling 1389).

Table C.2 shows the sequence of model building for using 1, 2, and 3 states. For single-state models, the sequence of the terms included (after $T_a$ and $P_{s,\text{win}}$ were already included) are $P_h$, $T_s$, and $P_{v,a}$. Table C.2 also shows the number of parameters, log likelihood value, and a $p$-value for each model. The $p$-value is of the likelihood ratio test of the respective model against the null hypothesis of the preceding model. It is seen that the $p$-value exceeds the chosen 5% limit and is practically 1 when testing the best next extension, $P_s$. Hence, this term is not included. The “null” model is included in Table C.2. This is the simplest possible model within the framework of (C.1) where the single state is modeled by a straight line plus diffusion. It has no inputs.

The resulting single-state model is

$$
\begin{align*}
    dT_i &= \frac{1}{C_i} \left( P_h + P_{el} + \frac{T_h - T_i}{R_{hi}} + \frac{T_s - T_i}{R_{si}} + \frac{T_a - T_i}{R_{ai}} + A_w \cdot P_{s,\text{win}} + c_v \cdot P_{v,a} \right) dt \\
    &\quad + \sigma_i d\omega_i(t) \\
    Y_k &= T_i(t_k) + e_k, \quad e_k \sim N(0, \sigma^2_o) 
\end{align*}
$$

(C.14a)  \hspace{1cm} (C.14b)

$T_i$ is interpreted as a common temperature for the whole building, $C_i$ as the total heat capacity. $R_{hi} = R_{si}$ are the resistances to the hall and the scullery,
Table C.2  The forward selection procedures. The three first columns are the inputs in the three states representing indoor ($T_i$), floor heating system ($T_f$), and envelope ($T_e$) temperature, the total number of parameters ($n$), the log-likelihood value ($\ell$), and the p-value of the test against the preceding model. The last step for each number of states is insignificant since the p-values exceed 5%. Electricity consumption and temperatures of adjacent rooms are not included in the table even though they are used in all models.

<table>
<thead>
<tr>
<th>States</th>
<th>$T_i$</th>
<th>$T_f$</th>
<th>$T_e$</th>
<th>$n$</th>
<th>$\ell$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 state</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(None)</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1287</td>
<td></td>
</tr>
<tr>
<td>$P_h, P_{el}, T_h, T_s, T_a$</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1389</td>
<td>0</td>
</tr>
<tr>
<td>$P_h, P_{el}, T_h, T_s, T_a, P_s,\text{win}$</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>1973</td>
<td>0</td>
</tr>
<tr>
<td>$P_h, P_{el}, T_h, T_s, T_a, P_s,\text{win},P_{v,a}$</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>1993</td>
<td>$2.6 \cdot 10^{-10}$</td>
</tr>
<tr>
<td>$P_h, P_{el}, T_h, T_s, T_a, P_s,\text{win},P_{v,a}, P_s$</td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>1993</td>
<td>1</td>
</tr>
<tr>
<td>2 states</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}, T_a$</td>
<td>$P_h$</td>
<td></td>
<td></td>
<td>10</td>
<td>2829</td>
<td></td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}, T_a, P_s$</td>
<td>$P_h$</td>
<td></td>
<td></td>
<td>11</td>
<td>2870</td>
<td>0</td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}, T_a, P_s, P_{v,s}$</td>
<td>$P_h$</td>
<td></td>
<td></td>
<td>12</td>
<td>2875</td>
<td>$1.3 \cdot 10^{-3}$</td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}, T_a, P_s, P_{v,s}, W_s$</td>
<td>$P_h$</td>
<td></td>
<td></td>
<td>13</td>
<td>2875</td>
<td>1</td>
</tr>
<tr>
<td>3 states</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}$</td>
<td>$P_h$</td>
<td>$T_a$</td>
<td></td>
<td>13</td>
<td>2834</td>
<td></td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}, P_s$</td>
<td>$P_h$</td>
<td>$T_a$</td>
<td></td>
<td>14</td>
<td>2921</td>
<td>0</td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}, P_s, P_{v,s}$</td>
<td>$P_h$</td>
<td>$T_a$</td>
<td></td>
<td>15</td>
<td>2927</td>
<td>$7.1 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>$T_h, T_s, P_{el}, P_s,\text{win}, P_s, P_{v,s}, W_s$</td>
<td>$P_h$</td>
<td>$T_a$</td>
<td></td>
<td>16</td>
<td>2927</td>
<td>$9.6 \cdot 10^{-1}$</td>
</tr>
</tbody>
</table>
which are not estimated. $R_{ai}$ is the resistance to outdoors, and the estimate of the UA-value is the inverse of this. The cumulative periodograms and SACF of the residuals of this model are shown in the upper plots in Figure C.6. The model clearly fails to describe the dynamics. An exponential decay is seen in the SACF which suggests adding a state, and around 24 hours there is a clear autocorrelation. From the cumulative periodogram, it is also very clear that the power is far from being uniformly distributed on the frequencies.

### C.6.2.2 Two-state models

To decide on a starting model for the search of an optimal model with two states, three approaches were tried. One approach tried to introduce a state representing the sensor and leave $P_{s,win}, T_{a}, P_{h}$ in a state together. Another approach tried to introduce a state representing the envelope exchanging heat with $T_{i}$ and $T_{a}$. A third approach it was tried to let a state represent the floor heating system which exchanges heat with the rest of the building. Based on the likelihood values, the third approach was preferred. As for single-state models, terms are added subsequently (see Table C.2) but now they are tried on both states, and the best-performing is chosen. The chosen two-state model is

$$
\begin{align*}
\frac{dT_{i}}{dt} &= \frac{1}{C_{i}} \left( \frac{T_{h} - T_{i}}{R_{hi}} + \frac{T_{s} - T_{i}}{R_{si}} + P_{el} + A_{w} \cdot P_{s,win} + \frac{T_{a} - T_{i}}{R_{ai}} + A_{s} \cdot P_{s} \\
& \quad + c_{v} \cdot P_{v,s} + \frac{T_{i} - T_{i}}{R_{fi}} \right) dt + \sigma_{i} d\omega_{i}(t) \\
\frac{dT_{f}}{dt} &= \frac{1}{C_{f}} \left( P_{h} + \frac{T_{i} - T_{i}}{R_{fi}} \right) dt + \sigma_{f} d\omega_{f}(t) \\
Y_{k} &= T_{i}(t_{k}) + e_{k}, \quad e_{k} \sim N(0, \sigma_{o}^{2})
\end{align*}
$$

where the state representing the floor heating system is denoted $T_{f}$.

The wish for structural identifiability gave certain restrictions on the model formulations. Briefly, it is important to make assumptions about the coupling of the floor heating with the states. If for instance floor heating can exchange energy with both indoors and outdoors, it is impossible to estimate the the heat capacity of the building and the heat loss to the outdoors at the same time.

The SACF and the cumulative periodogram (Figure C.6 second row) look much better than for the single-state model. However, there is significant autocorre-
Figure C.6  Cumulative periodograms and SACF of standardized residuals for the better (by row) single-state model, two-state model, three-state model, and three-state model with variable observational noise. The unit on the abscissa of the cumulative periodograms is sample frequency, i.e. frequency 1 corresponds to $4/h$. The ordinates are normalized so that 1 is the full power of the signal. In both plot types, the blue dashed lines are 95% confidence intervals for white noise.
lation left on lag 30 minutes and 45 minutes, the 24-hour autocorrelation is still there, and the cumulative periodogram still exceeds the 95% confidence band.

The autocorrelation on 30 minutes and not on 15 minutes points to problems modeling the floor heating input. As explained in Section C.4, the floor heating signal is controlled on a 30-minute resolution, which could explain why there is no significant autocorrelation on 15 minutes.

C.6.2.3 Three-state models

The model selection procedure is repeated with three states. To find a starting point for the forward selection procedure, an attempt was made to extend the starting point for the two-state models with a state representing the sensors and with a state representing the envelope, i.e. between indoors and outdoors. The model with a state representing the envelope was chosen because it has the lower p-value tested against the null-model. The added state representing the envelope is modeled as two symmetric resistances around the capacity. By this restriction (and parameter reduction) structural identifiability is obtained.

Again terms were added sequentially but tried on all three states. The resulting three-states model is

\[
\begin{align*}
\frac{dT_i}{C_i} &= \frac{1}{C_i} \left( \frac{T_h - T_i}{R_{hi}} + \frac{T_s - T_i}{R_{si}} + P_{el} + A_w \cdot P_{s,win} + A_s \cdot P_s \right) \\
&\quad + c_v \cdot P_v \cdot s + \frac{2(T_e - T_i)}{R_{ai}} + \frac{T_i - T_i}{R_{fi}} \right) dt + \sigma_i \omega(t) \\
\frac{dT_e}{C_e} &= \frac{1}{C_e} \left( \frac{2(T_h - T_e)}{R_{ai}} + \frac{2(T_i - T_e)}{R_{ei}} \right) dt + \sigma_e \omega_e(t) \quad \text{(C.16a)} \\
\frac{dT_f}{C_f} &= \frac{1}{C_f} \left( \frac{P_h}{R_{hi}} + \frac{T_i - T_f}{R_{fi}} \right) dt + \sigma_f \omega_f(t) \\
Y_k &= T_i(t_k) + e_k, \quad e_k \sim N(0, \sigma_e^2) \quad \text{(C.16d)}
\end{align*}
\]

The SACF of the residuals (Figure C.6, third row) of the three-state model is insignificantly different from zero for all lags except for around 24 hours, i.e. the significant spikes at 30 and 45 minutes from the two-state model are gone. The third state added facilitates distinguishing the dynamics of the house related to outdoor temperature and floor heating. The cumulative periodogram of the residuals shows no significant deviance from uniformly distributed power
C.6 Results

on frequencies. The following investigates which input causes the diurnal autocorrelation. Since the floor heating signal was designed to be uncorrelated with time, it was expected to be a weather variable.

C.6.2.4 Analysis of diurnal error

To investigate the reason for the diurnal autocorrelation of the residuals, the residuals have been plotted over weather inputs. Figure C.7 shows output, residuals, standardized residuals, and inputs for the three-state model. Looking at the residuals, it is clear that around 20 points are extreme residuals. One residual is highlighted with blue to illustrate the importance of considering standardized residuals rather than residuals themselves. This point comes after the period that was removed because a door had been opened. The extended Kalman filter has not been updated with data in this period and the prediction is highly uncertain. Therefore, a large prediction error is expected. In the standardized residuals this point is no longer sticking out. This is important because the parameter estimate is based on a minimization of the standardized residuals.

The number of residuals is 1574, and the probability of having one or more outcomes outside the range $[-5, 5]$ of 1574 standard normal distributed variables is less than $10^{-4}$. Here (plot 3 from the top) there are 14. This is a clear sign of over-dispersion. In general, a qq-plot is a much better graphical comparison of a data set to a normal distribution, but here the time series plot reveals important dependency on inputs.

Eight large standardized residuals have been highlighted with red. They are all positive (i.e. the model under-predicts) and they occur when the indoor air temperature is increasing and solar radiation comes on. They all occur when the solar radiation projected onto the glass facade of the building is peaking.

Figure C.8 further illustrates this periodic deficiency. The figure shows the sum of squared standardized residuals of the three-state model versus the azimuth angle in polar coordinates. The standardized residuals have been binned in intervals of $\pi/24$ rad of the azimuth angle. From this plot, it is very easy to see that the single big problem with the model occurs for a narrow interval of azimuth angles. Also, the orientation of the building is indicated in the plot. The problem occurs when the azimuth angle is close to perpendicular to the glass facade, i.e. when the effect of the solar radiation is expected to be largest. This illustrates the 24-hour lag autocorrelation.

The model under-predicts the output when the projected solar radiation peaks,
Figure C.7  Residuals, output and inputs for the three-states model with stationary noise processes.
Figure C.8  Sum of squared standardized residuals in bins of $\pi/24$ rad of the azimuth angle for the three-states model with stationary noise processes. The rectangle illustrates the orientation of the building, and the blue side is the window facade. Clearly, the bigger problem is seen when the sun is around perpendicular to the window facade.
i.e. the effect of the solar radiation directly onto this side of the building is larger than the model predicts. The projected solar radiation drops a bit before peaking. This is because the projection coefficient drops as the elevation angle increases and no reflection of the solar radiation is taken into account. It may be that the extra effect of the radiation reflection of the snow-covered ground would further amplify the radiation here and improve the model performance. It would be a good improvement of the data to add pyranometers on the main glass facades, in this case only one.

From Figure C.7 it also seems that large wind speeds influence on the residual variance. Some numerically large (both positive and negative) standardized residuals are marked with green, and they coincide with large wind speeds. The red-marked points account for 10 (including points just after that are caused by the same phenomenon) out of 14 standardized residuals numerically larger than 5. Therefore, this is considered by far the biggest single problem.

It is possible that the problem occurs due to solar radiation hitting the shielding of the sensors directly. This is not an important effect in relation to the heat dynamics of the building but makes the modeling more challenging. It was decided not to go further into the physical modeling of the system, but instead apply some modeling of the noise processes to overcome the problem.

C.6.2.5 Input dependent noise

Various ways of modeling the periodic over-dispersion (Madsen and Thyregod 2011) were tried. An extra state representing the sensor directly hit by solar radiation was tried. Also, letting the observations depend on \( P_{s,\text{win}} \) was tried. Finally, letting observational noise variance depend on \( P_{s,\text{win}} \) was tried. The latter gave good results, both in terms of likelihood value and SACF of standardized residuals. The log likelihood value increased from 2927 to 3000.

Taking another look at Figure C.7, all the large residuals that are not marked red (and not right next to one marked red) are in periods of relatively large wind speeds. This includes the extreme residual early on March 4. However, it was decided to stop the model extension here. The number of states has been extended until the dynamical description of the data was satisfactory, the drift terms have been extended until no more terms were significant, and the observational noise has been modeled to take solar radiation into account which was justified from earlier model residuals.
Table C.3 Backward selection for some parameters of the final model. All the parameters are significantly different from zero. $1/R_o$ is included as the parameter tested to be zero. In practice, the whole term is omitted. The full final model has 16 parameters and log likelihood 3000.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k$</th>
<th>$\ell$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{\text{noise}}$</td>
<td>15</td>
<td>2927</td>
<td>0</td>
</tr>
<tr>
<td>$c_v$</td>
<td>15</td>
<td>2921</td>
<td>0</td>
</tr>
<tr>
<td>$A_s$</td>
<td>15</td>
<td>2926</td>
<td>0</td>
</tr>
<tr>
<td>$A_w$</td>
<td>15</td>
<td>2911</td>
<td>0</td>
</tr>
</tbody>
</table>

C.6.2.6 Backward selection

The model building process has been based on forward selection meaning that terms have been sequentially added to the model. This may lead to a model including insignificant terms. Therefore, it is tested whether one or more terms that can be removed from the model without changing the model structure are insignificant. The null hypotheses sequentially tested are that each of these parameters are zero. The $p$-value for these tests are listed in Table C.3. All parameters are significantly different from zero, so all parameters are kept in the model.

The cumulative periodogram and SACF of standardized residuals are shown in the bottom row of Figure C.6. The cumulative periodogram shows no significant deviation from uniformly distribution of the power on frequencies. The SACF with lags around 24 hours has now been reduced from 0.25 to 0.17. At earlier lags, slight autocorrelation has been introduced. The overall picture of the autocorrelation is far better than for any other considered model.

Figure C.9 shows a time series plot of the standardized model residuals using a solar radiation dependent system noise process. The same points are marked as in Figure C.7, third plot. The large standardized residuals caused by the solar radiation are significantly reduced. There are still some numerically large residuals left.

This model with three states and observational noise variance depending on horizontal solar radiation projected onto the main window side of the building is considered sufficiently well performing. It is entirely given as
Figure C.9 Standardized residuals versus time for the three-state model with input-dependent diffusion. It should be compared with the third plot in Figure C.7. Most of the very large standardized residuals have been significantly reduced, including the ones where solar radiation is obviously related (red). The ones where wind speed is large (green) are unchanged.

\[
\begin{align*}
dT_i &= \frac{1}{C_i} \left( \frac{T_h - T_i}{R_{hi}} + \frac{T_s - T_i}{R_{si}} + P_{el} + A_w \cdot P_{s,\text{win}} + A_s \cdot P_s + c_v \cdot P_{v,\text{a}} + \frac{2(T_e - T_i)}{R_{ai}} + \frac{T_f - T_i}{R_{fi}} \right) dt + \sigma_i d\omega_i(t) \quad (C.17a) \\
dT_e &= \frac{1}{C_e} \left( \frac{2(T_a - T_e)}{R_{ai}} + \frac{2(T_i - T_e)}{R_{ai}} \right) dt + \sigma_e d\omega_e(t) \quad (C.17b) \\
dT_f &= \frac{1}{C_f} \left( P_h + \frac{T_i - T_f}{R_{fi}} \right) dt + \sigma_f d\omega_f(t) \quad (C.17c) \\
Y_k &= T_i(t_k) + e_k, \quad e_k \sim N(0, \sigma_o^2 + (A_{\text{noise}} \cdot P_{s,\text{win}})^2) \quad (C.17d)
\end{align*}
\]

i.e. as in Equations C.16 except that the observational noise variance depends on \(P_{s,\text{win}}\).

C.6.3 Parameter estimates

The estimated model parameters for the better one-state, two-states, three-states and three-states with variable observational noise variance are all listed
in Table C.4. Also standard errors on the estimates (except for on the noise terms) are given. Be aware that the interpretation of the parameters depends on the model formulation. Especially between models with different numbers of states, parameters with similar names can vary in meaning. For instance, $C_i$, is the thermal capacity of the whole building in the single-state model but for the other models, the capacity is distributed on two or three multiple states. The states must be interpreted differently and so must the parameters.

Only the estimated parameters are listed in Table C.4. The assumed resistance to the adjacent rooms of 0.287 K/W for each wall is naturally without uncertainty.
Table C.4  Parameter estimates for the better models with one state, two states, three states and three states with variable system noise processes. Notice that depending on model formulation, the parameters have different interpretations.

<table>
<thead>
<tr>
<th></th>
<th>1 state</th>
<th>2 states</th>
<th>3 states</th>
<th>3 states, noise model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>std</td>
<td>est</td>
<td>std</td>
</tr>
<tr>
<td><strong>Thermal resistances, K W</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{t1}$</td>
<td>$1.47 \cdot 10^{-2}$</td>
<td>$7.09 \cdot 10^{-4}$</td>
<td>$1.11 \cdot 10^{-2}$</td>
<td>$3.92 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>$R_{t2}$</td>
<td>$6.73 \cdot 10^{-4}$</td>
<td>$7.49 \cdot 10^{-5}$</td>
<td>$1.33 \cdot 10^{-3}$</td>
<td>$1.29 \cdot 10^{-4}$</td>
</tr>
<tr>
<td><strong>Thermal capacities, J K</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_i$</td>
<td>$4.60 \cdot 10^7$</td>
<td>$1.90 \cdot 10^6$</td>
<td>$7.22 \cdot 10^6$</td>
<td>$7.34 \cdot 10^5$</td>
</tr>
<tr>
<td>$C_e$</td>
<td>$1.75 \cdot 10^7$</td>
<td>$1.26 \cdot 10^6$</td>
<td>$1.56 \cdot 10^7$</td>
<td>$6.90 \cdot 10^5$</td>
</tr>
<tr>
<td><strong>Areas, m$^2$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_v$</td>
<td>$1.57 \cdot 10^1$</td>
<td>$7.59 \cdot 10^{-1}$</td>
<td>$8.02$</td>
<td>$8.61 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>$A_w$</td>
<td>$2.84$</td>
<td>$4.02 \cdot 10^{-1}$</td>
<td>$2.84$</td>
<td>$4.02 \cdot 10^{-1}$</td>
</tr>
<tr>
<td><strong>Other solar constant, K m$^2$s$^{1/2}$/J</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_{\text{noise}}$</td>
<td></td>
<td></td>
<td></td>
<td>$3.61 \cdot 10^{-5}$</td>
</tr>
<tr>
<td><strong>Ventilation constant, $-c_v$</strong></td>
<td></td>
<td></td>
<td></td>
<td>$3.61 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>$r_v$</td>
<td>$2.42$</td>
<td>$8.97 \cdot 10^{-2}$</td>
<td>$3.63 \cdot 10^{-1}$</td>
<td>$1.18 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>$2.40 \cdot 10^{-1}$</td>
<td>$6.93 \cdot 10^{-2}$</td>
<td>$2.20 \cdot 10^{-1}$</td>
<td>$6.62 \cdot 10^{-2}$</td>
<td></td>
</tr>
<tr>
<td><strong>Wind constant, Ks$^{1/2}$/m</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_{\text{wind}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2.19 \cdot 10^{-3}$</td>
<td>$7.39 \cdot 10^{-4}$</td>
<td>$6.67 \cdot 10^{-4}$</td>
<td>$1.96 \cdot 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>$6.00 \cdot 10^{-3}$</td>
<td>$6.67 \cdot 10^{-4}$</td>
<td>$3.78 \cdot 10^{-2}$</td>
<td>$3.77 \cdot 10^{-2}$</td>
<td></td>
</tr>
<tr>
<td><strong>System noise scalings, K/√s</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>$3.90 \cdot 10^{-5}$</td>
<td>$2.89 \cdot 10^{-5}$</td>
<td>$7.50 \cdot 10^{-3}$</td>
<td></td>
</tr>
<tr>
<td>$\sigma_f$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>$5.13 \cdot 10^{-6}$</td>
<td>$3.90 \cdot 10^{-5}$</td>
<td>$2.89 \cdot 10^{-5}$</td>
<td>$7.50 \cdot 10^{-3}$</td>
</tr>
</tbody>
</table>
C.7 Interpretation and application

C.7.1 Estimation of physical properties

In stead of the model parameters themselves, derived properties of the building, or as here the apartment, may be of interest. These properties are namely UA-value, heat capacities, and time constants. They are reported in Table C.5. The overall UA-value for the envelope is calculated as the heat input needed to maintain 1° C temperature difference between inside and outside under steady-state conditions and provided that there is no heat losses to adjacent rooms. The UA-value is simply the inverse of the resistance between indoor and outdoor air temperature, $UA = 1/R_{ai}$, hence the insulation between the apartment and adjacent rooms has no effect on it. The reported total heat capacities are the sum of the heat capacities ($C_i, C_e, C_f$) in the models. The uncertainties are calculated by the multivariate delta method (Wasserman 2003).

To estimate dynamical properties of a linear system, consider the following parameterization of the system on matrix form (omitting the noise processes):

$$dT = (AT + BU)dt$$  \hspace{1cm} (C.18)

where $T$ is the state vector, $U$ is the input vector. $A$ is called the system matrix and determines the dynamics of the system, whereas $B$ determines how the inputs enter the system. The time constants of the system are found as the negative to the inverse of the eigenvalues of the system matrix. The ventilation terms in the obtained models are non-linear. Hence, linearization must be performed in order to obtain estimates of the time constants. This has been done around average values of temperatures and ventilation flows.

The estimates of the UA-value and the heat capacity are relatively consistent for all but the single-state model. This is expected because of the large improvement

<table>
<thead>
<tr>
<th></th>
<th>1 state est</th>
<th>1 state std</th>
<th>2 states est</th>
<th>2 states std</th>
<th>3 states est</th>
<th>3 states std</th>
<th>3 states, noise model est</th>
<th>3 states, noise model std</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA-value, W K$^{-1}$</td>
<td>68.2 2.7</td>
<td>89.9 2.5</td>
<td>83.1 0.7</td>
<td>82.8 0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total thermal capacity, kW h K$^{-1}$</td>
<td>12.8 0.0</td>
<td>6.9 0.3</td>
<td>5.9 0.2</td>
<td>6.0 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time constants, h</td>
<td>33.9 0.9</td>
<td>50.7 1.3</td>
<td>53.9 1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
in the ability to describe the dynamics when going from one to two states. The single-state model clearly failed the model validation. When that is said, both the UA and the heat capacity estimates drop significantly when including the third state. This is when the cumulative periodogram gets to lie inside the confidence interval for white noise. Between the three-state models, the estimates are quite similar. The inclusion of the variable observational variance reduces the uncertainty on the UA estimate a little.

The estimate of the UA-value of the apartment for the final three-state model with variable observational variance is $82.8 \text{ W K}^{-1}$, and the thermal capacity is estimated to $6 \pm 0.22 \text{ kW h K}^{-1}$ (estimate $\pm$ standard error). The uncertainty on the estimate of the UA-value corresponds to a 95% confidence interval of $(81.5, 84.0)\text{ WK}^{-1}$, i.e. approximately plus/minus 2%. On the heat capacity, the relative uncertainty is 100%. The UA-value per gross floor area is $0.97 \text{ W K}^{-1} \text{ m}^{-2}$.

The very small uncertainty on the UA-value is considered optimistic. It is the uncertainty of the property under a number of assumptions, namely that the model is correct. Under these assumptions, the uncertainty is very small, but simplifications throughout the model building may lead to larger uncertainties than this.

In (Mortensen and Nielsen 2010), the UA-value is estimated for 57 houses. All except for one value range from $1.0$ to $1.9 \text{ W K}^{-1} \text{ m}^{-2}$. Hence, it seems that the UA-value estimated is a relatively low value, and the apartment is relatively well insulated. However, these values should be compared with care, especially because the analysis in this paper concerns an apartment in a semi-detached house, and whole houses are analyzed in (Mortensen and Nielsen 2010).

For the heat capacity obtained, it corresponds to about $71 \text{ W h m}^{-2} \text{ K}^{-1}$. What is obtained in (Bacher and Madsen 2011) for a light wooden building with little inventory corresponds to approximately $40 \text{ W h m}^{-2} \text{ K}^{-1}$. Hence, the heat capacity obtained seems reasonable.

In (Madsen and Holst 1995) and (Bacher and Madsen 2011) an “effective window area” is estimated. Due to the decomposition of the solar radiation into the two signals $P_s$ and $P_{s,\text{win}}$ would have to be derived as a function of elevation and azimuth angles or for a certain day of the year, as function of time of day or just azimuth angle. However, the scaling of the observational noise variance with solar radiation and wind speed has the consequence that the parameter estimates are less dependent on the observations where solar radiation is projected onto the main window side of the building. This also means that the estimation of the parameters directly related to these inputs is more uncertain. For this reason, such a window area is not given here. On the other hand, the
input dependent observational noise processes strengthens the estimation of the parameters related to outdoor temperature and heating.

### C.7.2 Further model validation

The reported model has been validated by considering SACF and the cumulative periodogram. However, for the estimation of physical properties, the consistency of the estimates is important and will be directly addressed. Especially in this case, the residuals include outliers which challenge some of the assumptions behind the residual analysis. For further validation of the model fit reported, the data set was divided in two equally large sets, and the estimation of the chosen three-state model with variable observational variance was fitted on the two subsets. The data were divided after exactly half of the time of the experiment, which was on February 29 at 3.15 pm. The two halves of the dataset are different in a number of ways. The average indoor and outdoor temperatures are considerably higher in the first half, the floor heating control signal has shorter average shifting time in the first half, there is on average more solar radiation in the second half, and the wind speed is very large for about half a day in the second half compared to the rest of the data set.

The three-state model with variable observational noise variance (C.17) has been re-estimated on the two halves of the data, and the derived estimates of physical properties are given in Table C.6. The estimated areas are consistent in the two half data sets. The significantly higher uncertainty on $A_w$ is due to the less data points where there is direct solar radiation onto the main window side of the building. The estimates of the UA-value are (82 and 81 W K$^{-1}$) using only half of the data are very close to the estimate found on the full data set (83 W K$^{-1}$), and the uncertainties of the estimates have increased as expected (to 0.91 and 1.6 W K$^{-1}$ respectively). The estimate using the full data set is less than one standard deviation from both of the estimates using subsets of data. Also the estimates of the total thermal capacity are very close and within one standard deviation of the common estimate for the subsets of the data.

The model succeeds well in obtaining consistent estimates of the physical properties of the system using only these half data sets. In this case, we can see that – despite the inhomogeneity of the data – we are able to reach the same estimates using only 8 days. However, the uncertainty of the estimates increases as smaller data sets are used.

The final model does not include any dependence on wind speed or wind direction. However, this does not mean that another formulation of the wind dependence than the one chosen in Section C.3.3.4 could be significant.
Model validation by re-estimation of physical properties only considering half of the data. All parameter estimates are consistent on the two half data sets, and notice how the uncertainty increases as well.

<table>
<thead>
<tr>
<th></th>
<th>1st half</th>
<th>2nd half</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>std</td>
<td>est</td>
</tr>
<tr>
<td>Areas, m²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_s</td>
<td>6.5</td>
<td>0.8</td>
<td>5.5</td>
</tr>
<tr>
<td>A_w</td>
<td>1.6</td>
<td>0.8</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.1</td>
</tr>
<tr>
<td>UA-value, W K⁻¹</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>82.5</td>
<td>0.9</td>
<td>81.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>82.8</td>
</tr>
<tr>
<td>Total thermal capacity, kW h K⁻¹</td>
<td>6.1</td>
<td>0.3</td>
<td>5.8</td>
</tr>
<tr>
<td>Time constants, h</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r_1</td>
<td>1.0</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>r_2</td>
<td>1.8</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td>r_3</td>
<td>46.3</td>
<td>64.2</td>
<td>53.6</td>
</tr>
</tbody>
</table>

C.7.3 Discussion of the accuracy of the method

The models are based on the assumption that the temperature measurements in the column of the center of the apartment are representative for the whole apartment. While the measurements are believed to be of high quality, the homogeneity assumption may not be that good. To investigate this assumption, a spatial multi-state model could be formulated using measurements from more locations in the apartment.

It is important that the estimation of these physical properties and the assessment of uncertainties of the estimates are based on the assumption that the models are correct. The reported uncertainty on the estimate of the UA-value is therefore considered very optimistic if interpreted as the uncertainty about the true overall UA-value of the envelope. It is not possible to conclude that this estimator follows a distribution with this standard deviation from the true value. The model validation serves to give a picture of how well these results can be believed.

Another concern about the reported estimates is that they are obtained using a short period of time. The 16.5 days are believed to provide a long enough period for a good excitation of the system and reliable model estimation. However, the estimates are based only on data from a cold period with relatively little sunlight. These properties of the inputs will give a good signal-to-noise ratio for estimation of thermal resistance but less of accuracy on parameters related to solar radiation. Moreover, parameter estimates rely on linearizations which may depend on the inputs. However, if the system can be assumed to be (almost) linear, this should not be a problem.
C.7.4 Simulations

To better understand a dynamical system, one can consider impulse responses and step responses to different inputs. Simulations of the deterministic part of (C.17) will be performed. Since the indoor air temperature is of special interest concerning comfort for users of the building, that state is considered in the simulations. For an impulse response, all states are initialized at zero. All except for one input are zero throughout the simulation. The input to which the impulse response is simulated is then instantly changed to release a given amount of energy.

In the upper plot of Figure C.10, 1 kWh is released in floor heating and ventilation heating respectively at time 0. For the ventilation heating, the temperature gain is obtained almost instantly, and in two hours the indoor air temperature drops about two thirds of the maximum observed gain. The maximum temperature gain is less in the impulse response to floor heating. But here, the energy dissipates out of the system much more slowly. After one day, the indoor air temperature is still at around % of the maximum indoor air temperature observed.

The curves in Figure C.10 show important features that must be used when controlling the indoor air temperature with floor and ventilation heating. It is seen that one must heat a few hours before the maximum effect is wanted. For use in an energy grid with variable energy prices, there are good possibilities of time shifting the heating demand in the considered system in order to consume e.g. electricity when the price is low. For obtaining good indoor comfort efficiently, both ventilation heating and floor heating must be controlled.

The lower plot of Figure C.10 shows a step response to outdoor temperature (i.e. the outdoor temperature is changed from 0 to 1°C at time 0, and all other inputs are kept at zero). The asymptotic indoor air temperature is approximately 0.92°C. This is because the temperatures, $T_h$ and $T_s$ of the entrance hall and scullery are kept at 0°C which leads to a heat transfer with the apartment. The response is slow which is expected because of the time constant of 54 hours.

C.8 Conclusions

A procedure has been outlined to build and validate stochastic grey-box models of the heat dynamics of a building with multiple heat inputs, including floor heating and solar radiation through a large window area. For a given number of
states, likelihood-ratio tests were conducted to determine which terms to include, and the number of states was increased until the dynamics of the system were satisfyingly well matched by the model.

The procedure was carried out on data collected in an experiment in a modern low-energy house with floor heating and large windows in the Arctic. The resulting model featured three states, and the significant terms were global horizontal solar radiation, a projection of the same horizontal solar radiation, ventilation heating, outdoor temperature, and floor heating. However, due to 24-hours autocorrelation of residuals, the observational noise variance was modeled by linear combinations of inputs. In this case, horizontal solar radiation projected onto the main window facade of the building proved significant for this. A potential for modeling the system noise processes in a similar way was observed.

The presented method is highly data efficient; based on a 16-day experiment, key thermal parameters were estimated with relatively little uncertainty. The UA-value of the apartment was estimated to $83 \pm 0.6 \text{ W K}^{-1}$ (of which the uncertainty was however argued to be optimistic). The obtained model was further validated by splitting the data set in two and re-estimating the model on the two halves. The physical properties found were consistent with the ones found on the whole data set. This was the case for both UA-value and total
thermal capacity. Hence, if estimation of these physical properties is the aim of the experiment, 8 days of data would in this case be enough with this type of model. In the analysis, assumptions about the heat exchange with adjacent rooms had to be made.

It was recommended that for similar experiments, the heat input to adjacent rooms should be designed with independent and persistently exciting signals in order to avoid problems with correlation with other inputs. Challenges estimating the heating power provided by the floor heating system led to the recommendation of including electric heating giving an easily measurable heat input which can be used to estimate heat inputs and heat capacities in the building. Also, it was recommended to use pyranometers installed directly on walls in the direction of large window areas in order to model the fast dynamics related to solar radiation.

The obtained models have the advantage of having physically interpretable parameters, and estimates of properties such as UA-value, thermal capacities, and time constants for the building can be derived. Dynamics of the system can be understood from the obtained model fit as well. Simulations demonstrated different dynamic responses to ventilation heating, floor heating, and outdoor temperature. For example, impulse response from floor heating to indoor air temperature showed that this system is suitable for shifting the heating demand several hours. This can contribute to the design of controllers for use in intelligent energy grids.
Table C.7 Nomenclature. Notice that interpretation of states depends depending on the model formulation.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>System matrix in a linear system</td>
<td></td>
</tr>
<tr>
<td>$A_s$</td>
<td>Area with which the global horizontal solar radiation is scaled</td>
<td>$[\text{W/m}^2]$</td>
</tr>
<tr>
<td>$A_w$</td>
<td>Area with which the projected solar radiation ($P_{s,\text{win}}$) is scaled</td>
<td>$[\text{W/m}^2]$</td>
</tr>
<tr>
<td>$A_{\text{noise}}$</td>
<td>Scaling of projected solar radiation in a system noise process</td>
<td>$[\text{Km}^2\text{s}^{1/2}/\text{J}]$</td>
</tr>
<tr>
<td>$B$</td>
<td>Matrix describing how inputs enter in a linear system</td>
<td></td>
</tr>
<tr>
<td>$C_i$</td>
<td>Thermal capacity of state i (or e,f)</td>
<td>$[\text{J/K}]$</td>
</tr>
<tr>
<td>$P_{\text{el}}$</td>
<td>Electric power</td>
<td>$[\text{W}]$</td>
</tr>
<tr>
<td>$P_h$</td>
<td>Floor heating power</td>
<td>$[\text{W}]$</td>
</tr>
<tr>
<td>$P_s$</td>
<td>Global horizontal solar radiation</td>
<td>$[\text{W/m}^2]$</td>
</tr>
<tr>
<td>$P_{s,\text{win}}$</td>
<td>Global horizontal solar radiation projected onto the window side of the building</td>
<td>$[\text{W/m}^2]$</td>
</tr>
<tr>
<td>$P_v$</td>
<td>Estimated ventilation heating to one apartment</td>
<td>$[\text{W}]$</td>
</tr>
<tr>
<td>$R_{ij}$</td>
<td>Thermal resistance between states or inputs i and j</td>
<td>$[\text{W/K}]$</td>
</tr>
<tr>
<td>SACF</td>
<td>Sample autocorrelation function</td>
<td>$[-]$</td>
</tr>
<tr>
<td>$T_a$</td>
<td>Outdoor temperature</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>$T_e$</td>
<td>Temperature of the building envelope</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Indoor air and surface temperature (in three-states model)</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>$T_f$</td>
<td>Floor heating system temperature</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>UA</td>
<td>Common UA-value for the building envelope</td>
<td>$[\text{W/K}]$</td>
</tr>
<tr>
<td>$W_s$</td>
<td>Wind speed</td>
<td>$[\text{m/s}]$</td>
</tr>
<tr>
<td>$c_v$</td>
<td>Scaling of ventilation heating signal</td>
<td>$[-]$</td>
</tr>
<tr>
<td>$c_w$</td>
<td>A constant related to the effect of wind speed</td>
<td>$[\text{Ks}^{1/2}/\text{m}]$</td>
</tr>
<tr>
<td>$\epsilon_k$</td>
<td>Observational noise at discrete time, $k$</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>$k$</td>
<td>Discrete time</td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>Number of parameters in a model</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>P-value of a hypothesis test</td>
<td></td>
</tr>
<tr>
<td>$r_k$</td>
<td>One-step-ahead prediction uncertainty of the indoor air temperature measurement at time $k$</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>$t$</td>
<td>Continuous time</td>
<td></td>
</tr>
<tr>
<td>$Y_k$</td>
<td>Indoor temperature measurement at time $t_k$</td>
<td>$[^\circ\text{C}]$. When referring to particular observations, $y_k$ is used.</td>
</tr>
<tr>
<td>$\epsilon_k$</td>
<td>Residual, i.e. one-step-ahead prediction error of the indoor temperature at time $k$.</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>$\hat{\epsilon}_k$</td>
<td>Standardized residual at time $k$, $\hat{\epsilon}_k = \epsilon_k/r_k$</td>
<td>$[-]$</td>
</tr>
<tr>
<td>$\omega_i$</td>
<td>Wiener processes in state i (or e,f)</td>
<td>$[\text{s}^{1/2}]$</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>Scaling of Wiener process in state i (or e,f)</td>
<td>$[\text{K s}^{-1/2}]$</td>
</tr>
<tr>
<td>$\sigma_o$</td>
<td>Standard deviation of observational noise</td>
<td>$[^\circ\text{C}]$</td>
</tr>
<tr>
<td>$\ell$</td>
<td>Log-likelihood value</td>
<td>$[-]$</td>
</tr>
</tbody>
</table>
References


Low-energy house in Sisimiut – Measuring equipment

Authors:
Philip Delff Andersen\textsuperscript{1}, Henrik Madsen\textsuperscript{1}, Carsten Rode\textsuperscript{2}.

Published by:
DTU Compute
(Technical report)

\textsuperscript{1}DTU Compute, Matematiktorvet, Building 322, DK-2800 Kgs. Lyngby, Denmark
\textsuperscript{2}DTU Civil Engineering, Brovej, Building 118, DK-2800 Kgs. Lyngby, Denmark
Abstract

This paper documents the measurement equipment in a low-energy house in Sisimiut, Greenland. Detailed measurements are being taken on energy consumption, indoor temperatures, floor heating, ventilation, open/closed state of doors and windows, and indoors climate. Equipped with a central control unit, experiments can be designed in order to study heat dynamics of the building. It is described how to plan and execute such experiments in one apartment in the building. The building also features both a solar thermal system and extra buffer tank facilitating testing of storage strategies on the power generated by the solar thermal system. A weather station equipped with thermometer, pyranometer and anemometer is installed on the building as well. Finally, it is described how to retrieve data from an SQL server which is configured to take monthly backups. R functions have been implemented to fetch and prepare the data for time series analysis. Examples are given on the use of these.

D.1 Introduction

Technical University of Denmark (DTU) have constructed a modern low-energy house in Sisimiut, Greenland, in order to gain experience with low-energy construction even under very demanding weather conditions and creating a full-scale research facility that enables experiment with heating and ventilation systems. Furthermore, the house is equipped with solar thermal panels to provide domestic hot water. A picture of the building is shown in Figure D.1. A thorough description of the building can be found in (Norling et al. 2006).

After the first five years of operation, repair work was carried out in order to improve on the energy accounts of the house which did not live up to the original expectations. Moreover, it was decided to invest in an improved measurement and control system. The new system was installed in 2011 and provides more than 200 signals online to use for supervision and modeling of the building and the energy consumption. It can be used to execute experimental plans, and that even from remote.

This report is a description of the installed system and provides example plots of many important measurements. The report has been carried out partly for documentation of work done during a PhD study on heat dynamics in buildings, and hence the focus tend to be more on the part of the measurements that concern heating, temperatures, and occupancy. The house consists of two apartments of which one is rented out, and the other is used as a guest house.
and for research. Many of the examples from inside the building particularly uses data from the apartment that is not rented out.

Examples on experiments carried out in the building can be seen in (Andersen et al. 2013d). One of the experiments, from which many examples in this paper have been taken is thoroughly analyzed in (Andersen et al. 2014b). Finally, an analysis of some of the data that was collected before the installation of the new data acquisition system can be found in (Andersen et al. 2013c).

D.2 Overview of the system

All measurements in the building are collected by a central unit called a programmable logic controller (PLC). Measurements are taken on indoor climate, heating systems, the solar thermal system, outdoor climate, and presence of occupants. Table D.1 shows an overview of the most important measurements for modeling of the heat dynamics of the building.

The PLC also provides a graphical user interface (GUI) for Microsoft Windows which is running on a computer in the house. This can be accessed from remote by using an application called Teamviewer\(^3\). From this, the most recent of many of the measurements can be seen. When opening the GUI, the user is being presented with an overview (See Figure D.2) of the heating system, and the solar thermal system. Leaving the boiler (“Oliefyr”) the heated liquid is first

\(^3\)http://www.teamviewer.com
Table D.1  Overview over measurements in the house of special interest for modeling of heat dynamics of the building.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Common areas</th>
<th>Rental apartment</th>
<th>Experimental apartment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temperature</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hall</td>
<td></td>
<td>All rooms</td>
<td>All rooms</td>
</tr>
<tr>
<td>Broiler room</td>
<td></td>
<td></td>
<td>Full standard indoor temperature measurement in living room</td>
</tr>
<tr>
<td><strong>Heating</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor heating</td>
<td>2/2</td>
<td>5/5</td>
<td>5/5</td>
</tr>
<tr>
<td>Ventilation afterheating</td>
<td></td>
<td>All measured together</td>
<td></td>
</tr>
<tr>
<td><strong>Ventilation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central ventilation</td>
<td></td>
<td>All measured together</td>
<td></td>
</tr>
<tr>
<td>Outer doors open/closed</td>
<td>2/2</td>
<td>3/3</td>
<td>3/3</td>
</tr>
<tr>
<td>Windows open/closed</td>
<td>No windows</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>Cooker hood</td>
<td>0/0</td>
<td>0/1</td>
<td>1/1</td>
</tr>
<tr>
<td><strong>Occupancy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIR sensors</td>
<td>0</td>
<td>0</td>
<td>Living room/kitchen, corridor</td>
</tr>
<tr>
<td>CO₂ concentration</td>
<td>0</td>
<td>Deactivated</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Consumption</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oil</td>
<td></td>
<td>All measured together</td>
<td></td>
</tr>
<tr>
<td>Electricity</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td><strong>Solar thermal system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total heat collection</td>
<td></td>
<td></td>
<td>Only one system</td>
</tr>
<tr>
<td>Domestic water heating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buffer water heating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heating system</td>
<td></td>
<td></td>
<td>Contributes to common system</td>
</tr>
<tr>
<td><strong>Meteorological variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient temperature</td>
<td></td>
<td></td>
<td>All at one common weather station</td>
</tr>
<tr>
<td>Solar radiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind direction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

connected to the domestic hot water tank (where the liquid can run through a spiral without mixing with the water in the tank), then to the floor heating loops, and finally to the ventilation after heating. Only 6 floor heating loops are shown in the overview, but actually there are 12 of them. The return temperatures are reported individually while the forward temperature is common for all of them. The water from the boiler is connected to the ventilation intake air through a heat exchanger. The ventilation system is equipped with heat recovery so the boiler is only used to heat after that.

The solar thermal system is connected to the domestic hot water tank. If all the water in the tank is already hot, the liquid can be passed to a buffer tank
under the building or it can be run through a heat exchanger to after heat the liquid coming back to the boiler. If also this tank is hot already, and there is no use of the energy in the heating system, energy from the solar system can be dispatched in a radiator located in the entrance hall.

All data is being collected using the PLC system in UTC time. Normally, data is being measured every 30 seconds or with pulse signals. Table D.4 in Section D.14 is an overview of all the sensors, and basic information about them. An overview of the computer network in the building is given in Section D.15.

D.3 Consumption measures

The consumption in the building is measured in terms of oil consumption, electricity consumption, and water consumption. Electricity consumption is measured for the two apartments and for common spaces separately. The water consumption is measured for the whole building. The PLC GUI provides the overview of online consumption readings shown Figure D.3.
D.3.1 Oil consumption

Three readings, A02_FM01, A02_FM01_SUM, A02_FM01_TRIP, are measurements of the oil consumption. A02_FM01 and A02_FM01_SUM are plotted for the year 2012 in Figure D.4. A02_FM01 seems to always be 0 and hence contains no information. The others are cumulative counts of pulses for each liter of oil that has been consumed, so the resolution is quiet limited. The “trip” version can be reset in the user interface but is apart from that exactly the same as the “sum”. However, it seems to be missing in the database. From the data, it is calculated that about 2930 liters of oil were consumed in 2012.

D.3.2 Electricity consumption

The readings from the three electricity meters in 2012 are shown in Figure D.5. The meters get a pulse for each kWh consumed. As, expected, the rental apartment has the larger electricity consumption. It is quite stable, but larger in the winter than in summer. The guest apartment electricity consumption comes more in jumps which is normal because of the occasional use. The common electricity consumption seems to be quite constant over the year. Again, that is as expected since this is mostly for pumps, data acquisition, a computer etc.
Figure D.4 The two available series related to the oil consumption throughout 2012. The instant flow measurement is obviously not working, and the “Trip” series is missing in the database.

Figure D.5 The three electricity meters throughout 2012.
It is known that at least a freezer belonging to the rental apartment has been connected to a power plug socket of the common areas. Thus, the measures may not exactly correspond to what is expected.

D.3.3 Water consumption

A flow measurement from boiler to spiral in domestic hot water tank is missing. Hence, the power consumption to hot water cannot be calculated. The description of the water consumption is therefore incomplete.

The water consumption is being measured in terms of water flows and energy consumed for water heating. Also the temperature of the water near the bottom and near the top of the hot water tank is being measured. Some related data is plotted for 2012 in Figure D.6. The first plot is the total energy to domestic water heating. The two next plots are the average daily consumption in m$^3$/h and the cumulative consumption in m$^3$. Figure D.7 compares these two signals by cumulating the instant flow (A04_X02) to the cumulative flow (A04_X03) subtracted by its own minimum. They obviously match except for periods of missing flow measurements. A04_X02 is called an instant power measure in the database. This must be a mistake. It must be a flow measure.

The fourth plot in Figure D.6 is apparently an instant volume flow (A04_X04). The mean value is around 230. What it represents is unknown, and so is the unit. According to Bo Holdt-Simonsen, it is likely to be a power measure. Then follows a water inflow temperature. The temperatures are too low to be from the boiler and seemingly too high to be from the solar panel. A04_X10, A04_FM01, and A04_FR01 are missing or zero and do not contain any information. A04_X11 is cumulative extra energy to hot water, i.e. the energy to hot water consumption not covered by the solar thermal panels. In 2012 the increase is larger than for A04_X01 which further questions what the latter represents.

The cumulative hot water flow from the tank – A04_FM01_SUM – has a few outliers. Even though it seems to be in liters, it seems to have similar shape to A04_X03 which is the cold water inlet to the tank. They are compared in Figure D.8 and and match very well.
Figure D.6 Measures related to water consumption.
Figure D.7  Comparison of $A_{04,X02}$ and $A_{04,X03}$ where the first has been cumulated, the latter has been substracted its minimum.

Figure D.8  Comparison of the cold water inlet and the hot water outlet from the domestic hot water tank. The signals have been scaled to match.
D.4 Room temperature measurements

The temperature is measured in each room by Thermokan SR04 wireless sensors. The sensors have two parameters to control the frequency with which they transmit measurements. The one is $T_{\text{wakeup}}$, the number of seconds between each measurement. If a measurement is more than 2% different from what previously recorded, a signal is transmitted. This condition is repeated $T_{\text{interval}}$ times after which a signal is transmitted no matter the measurement. $T_{\text{wakeup}}$ and $T_{\text{interval}}$ are both set to 10 seconds. That means that a signal is sent at least every 100 seconds. Adjustment of these variables is done by physically moving jumpers under the plastic covers on the units.

(a) The positions of the thermal sensors used to control the floor heating.

(b) An installed thermal sensor. (c) A thermal sensor with cover removed. The jumpers are seen in the bottom-left corner.

Figure D.9 Positions and pictures of the thermal sensors that are installed in all rooms except for the boiler room. Sketch and photots by Konstatinos Tsapralidis
The sensors are installed on the walls around 150 cm from the floor. A sketch of the positions and pictures of one of the sensors are shown in Figure D.9.

### D.4.1 Standard room temperature measurement

Placed centrally in the guest apartment, a column of six temperature measurements is installed. They are thermocouples type T sensors with Seneca K121 transmitters. See pictures and a sketch of the position heights in Figure D.10.

![The whole column of thermometers.](image1)

![One sensor in the column.](image2)

![Sketch of the heights of the sensors.](image3)

**Figure D.10** The column of temperature measurements in the presentation apartment. On (a) the cooker hood, the CO$_2$ sensor (near the ceiling to the left of the column of thermometers), and a PIR sensor (in the ceiling in the back) are also seen. D.10c shows the heights of the six sensors (Konstatinos Tsapralidis).
Figure D.11 shows different temperatures measured in the building. The upper plot shows temperature measurements in the different rooms of the apartment, measured by on-wall thermometers. The temperature difference between some rooms are up to about 5 °C. Another interesting feature is that when the temperatures increase rapidly, the living room temperature seems faster than the others. There are however also times where this is not the case. This could indicate that different heating inputs (e.g. floor heating and solar radiation) work differently on the different rooms.

The second plot in Figure D.11 shows temperature measurements from the column of standardised temperature measurements located centrally in the apartment. The temperature difference from top to bottom stays within a couple of °C, and the temperature gradient switches direction when the sensors are being heated or cooled respectively. When they are cold, the bottom measurement is the lower, when they are warmer, the top measurement is the warmer.

D.4.2 Principal component analysis of temperature distribution

Principal component analysis (Izenman 2008) is a method in multivariate statistics where the variation in multiple directions (here multiple sensors) is projected onto new orthogonal dimensions. These new dimensions are ordered so that the first (principal) component captures the most of the variation in data, the second component captures second most under the constraint to be orthogonal to
Figure D.12  Loadings of the different temperature measurements in principal components. For the column, the number on the abscissa means the sensor number from the top. For the rooms, the sequence is large room, small room, corridor, bathroom, living room.

The loadings are shown in Figure D.12 and express how the original measurements are weighted in the construction of the principal component. This means that the principal components are linear combinations of the original temperature measurements.

Figure D.13 shows principal components of both the room temperatures and the temperatures measured in the column of sensors. It is seen that the first principal component in both cases represent something close to the mean values of the groups of temperatures (with negative sign). The second principal component expresses the difference from top to bottom. The warmer the bottom is compared to the top, the larger this principal component. Furthermore, the measurements have the larger weight, the further they are from the center. For the rooms, the second principal component expresses the temperature difference between the living room on the one side and the two rooms and the bathroom on the other. Notice that the corridor is given approximately zero weight.

So the second component of the column of thermal sensors expresses the vertical temperature gradient. This dynamics is important because it is excited by floor heating. The second component of the room temperatures expresses the difference between living room and small rooms plus bathroom. This is likely to be because of the solar radiation coming into the living room through the large windows.
D.5 Activity sensors

The house is equipped with different sensors with the purpose of facilitating analysis of the influence of user behavior in models of heat dynamics. Open/closed sensors are on all external doors and windows of the two apartments and the common exterior doors. Moreover, the guest apartment has two passive infrared (PIR) sensors and a CO$_2$ sensor to measure presence of occupants. This section briefly lists the available measurements and gives an example of recorded data from a short period. Through the GUI of the PLC, an overview of each apartment can be seen. A screen dump of this is shown for the guest apartment in Figure D.14. In the overview, the user can see temperature measurements, status of floor heating (green disc for on, white for off), set temperature for floor heating, and PIR sensor status (see Section D.5.1). The six temperatures to the left are from the column of thermometers in the center of the apartment followed by temperature and humidity measurements in paper insulation under the house. Open/closed state of windows and doors is also seen in the overview (Section D.5.2).

D.5.1 Passive infrared sensors

In the rental apartment two passive infrared (PIR) sensors are installed in the ceiling. One is in the corridor (see Figure D.10a), one is in the living room.
Figure D.14 From the PLC GUI an overview of the activity in the two apartments can be seen. In the guest apartment, as shown here, CO₂ concentration is measured in the living room, and two PIR sensors register occupant presence.

They have a delay of less than one minute, meaning that they will keep showing activity for this delay after activity has been measured.
Two passive infrared sensors are installed in the guest apartment. The one in the living room is seen in the ceiling in the picture to the right. See the position of the one in the corridor in Figure D.10a. Photos by Konstatinos Tsapralidis.
Figure D.16 The occupancy sensors in the guest apartment during experiment 5.

D.5.2 Open/closed sensors

The outer doors of the building (entrance door, backdoor), and outer doors and windows of the apartments (doors to hall and scullery, windows and terrace door) have open/closed sensors.

D.5.3 CO₂ sensors

There is a CO₂ sensor in each apartment (see Figure D.10a), but only the one in the rental apartment is active. This and the missing PIR sensors in the rental apartment is due to respect for privacy of the tenants.

D.5.4 A short period of data

Figure D.16 shows data from the PIR sensors, the open/closed sensors and the CO₂ sensor in the guest apartment from March 12th to March 27th 2013. Green is a closed door or no activity while red means open door or activity. The data logger stops logging when there is no change in the signal, which is the reason for the white spaces. It has been checked that they correspond to no activity or closed as well as green.

It is noticed that the PIR sensors react even though no doors have been opened. It seems that something else than occupant presence is able to activate them or
that the open/closed sensors do not work. The CO$_2$ measure does not seem to be a good indicator of activity. It is periodically all large but this does not seem to be correlated with the other signals.

D.6 Floor heating

The floor heating system is liquid born and has twelve strings, including five for each apartment, one for the entrance hall, and one for the scullery. One common forward temperature is measured. For the return temperatures, both a common return temperature, and the twelve individual return temperatures are measured.

This means that the floor heating supply to the $i$’th string can be calculated as

$$P_{h,i} = c_h \cdot \rho_h \cdot \dot{V}_{i,h} \cdot (T_{h,in} - T_{i,h,out}) \quad (D.1)$$

$P_{h,i}$ is the power dispatched in the $i$’th string, $c_h$ is specific heat capacity of the liquid, $\rho_h$ is the density, $\dot{V}_{i,h}$ is the volume flow through the $i$’th string, $T_{h,in}$ is the common forward flow temperature, and $T_{i,h,out}$ is the $i$’th return temperature. However, some resampling may be needed to obtain common time stamps for the involved signals. How to obtain this will be described in Section D.12.4. The flow and temperature sensors on the floor heating system are shown in Figure D.18. Status of valves in the twelve strings and the inflow pump are logged as well.
The constants $c_h$ and $\rho_h$ are specific heat capacity and density for the fluid in the floor heating system which is a mixture of water and glucola. Liquid has been added multiple times resulting in an unknown ratio of the two components. A Nuclear Magnetic Resonance Spectroscopy (NMR) analysis of a sample of the liquid was performed at Department of Energy Conversion and Storage at Risø, DTU. The mixture was estimated to be 1:8.85 of glucola/water by molecular concentration. The estimate should have an uncertainty of around 5%. This ratio can be used to estimate the heat capacity and the density of the mixture. A linear interpolation using this concentration yields $c_h \approx 4.01 \cdot 10^3 \text{ J kg}^{-1} \text{K}^{-1}$ and $\rho_h \approx 1.01 \cdot 10^3 \text{ kg/m}^3$.

As an example, floor heating data from an experiment in the guest apartment is shown. Since there are five floor heating loops with individual flow and return temperature measurements (common forward temperature measurement), the example involves 11 signals. The experiment was designed so that the five floor heating signals followed a common Pseudo Random Binary Signal (PRBS) which is a deterministic powerful signal for examples to identify and infer on dynamic systems (Godfrey 1980). The flow measurements are shown in Figure D.19. It shows both the total flow and the flows for the individual rooms. The total flow is only shown interpolated, the individual ones are shown both as raw measurements, and interpolated. This is due to measurements not always being taken at the exact same time, so the sum of the individual measurements is not well defined. In the bottom, the PRBS used for the experiment is shown.
Figure D.19  Flow measurements and interpolations of these in all floor heating loops in the apartment. The black points are raw data, the green are interpolated.
Figure D.20  Return temperatures and interpolations of these in all floor heating loops in the apartment. The black points are raw data, the green are interpolated.
Figure D.20 shows the return flow temperatures in the five floor heating loops. Also it shows their mean at each interpolated time step. The total floor heating power to the apartment is now estimated by

\[ P_h = \sum_{i=1}^{5} P_{h,i} \]  

(D.2)

This is plotted together with the estimated power in the individual loops in Figure D.21.

Figure D.22 shows the distribution of the energy supplied to the floor heating system in the guest apartment during the experiment. In total, 1291 kWh is supplied.
Figure D.21  The estimated floor heating power supplied to the guest apartment.
Figure D.22 The distribution of the floor heating supplied to the guest apartment during the experiment.
D.7 Ventilation

For a description of the ventilation system, see e.g. (Vladyková et al. 2011). Here, the focus is on the describing the location of the sensors. Both flows and temperatures are measured different places in the ventilation system. However the contribution to the whole building is measured, not to the apartments individually.

A sketch of the ventilation system and the positions of the different sensors is shown in Figure D.23.

![Sketch of the ventilation system and the positions of the different sensors. By Bo Holdt-Simonsen and Philip Delff.](image)

The flow measurements are being calculated in the PLC from a pressure drop measurement using the relation

\[ Q = 27.8 \cdot \sqrt{\Delta p} \cdot 3.6 m^3/h/Pa^{1/2} \]  \hspace{1cm} (D.3)

However, until March 30 2012 the following relation was erroneously used:

\[ Q = 0.5 m^3/h/Pa \cdot \Delta p \]  \hspace{1cm} (D.4)

For data before March 30 2012, the user has to correct the data.
D.7 Ventilation

Figure D.24 shows ventilation in and out flows for a period in February-March 2012 (the data has been corrected to follow Equation (D.3)). The in flow stays in the interval 4 to 11 m$^3$/h while the out flow stays within 6 to 10 m$^3$/h. The cumulative flows show that the out flow is considerably larger than the in flow on average.

![Figure D.24](image)

**Figure D.24** In flow is red, out flow is blue.

Flow temperatures in the ventilation system are plotted in Figure D.25. The inlet temperature to the building varies little around 22°C. The inlet before any heating follows the outdoor temperature, while the outlet temperature follows the indoor temperature.

The energy supplied to the building by the ventilation system is estimated and plotted in Figure D.26. It is estimated based on inlet temperature after the after heating, outlet temperature before heat exchanger, in and out flows, and physical properties of air at 25°C. The relation is

$$ P_v = c_{\text{air}} \cdot \rho_{\text{air}} \cdot (\dot{V}_{v,\text{in}} \cdot T_{v,\text{in}} - \dot{V}_{v,\text{out}} \cdot T_{v,\text{out}}) $$  \hspace{1cm} (D.5)

The power supplied will be negatively correlated with indoor temperature according to Equation (D.5).
Figure D.25  Temperatures in the ventilation system. C01_IN03 is inlet before any heating, C01_IN01 is inlet after heat exchanger, before after heating, and C01_IN02 is inlet after after heating. C01_UD01 is outlet temperature from the apartment, before heat exchanger. See also Figure D.23.

Figure D.26  Estimated ventilation energy to the building. Based on C01_IN02, C01_FM01, C01_UD01, C01_FM02, and physical properties of air at 25°C.

D.7.1 Cooker hoods

The use of the cooker hoods are measured in both apartments. The data can be found under the tags C01_TT01 and C01_TT02 respectively.
D.8 Weather station

The building is equipped with a weather station connected to the data acquisition. The weather station includes a thermometer which is shielded from radiation, a pyranometer measuring global horizontal radiation, and an anemometer measuring both wind speed and direction. On the picture in Figure D.1 the weather station is seen on the roof. Figure D.27 shows the overview provided by the PLC GUI. Apart from online measures, it provides dynamic graphs where the user can specify time interval and which measures to plot.

Data from the weather station in the period January 16th to February 1st 2013 is shown in Figure D.28. The data shown has been averaged to 15 minute resolution. The outdoor temperature increases by more than 10 °C within about half an hour. That happens as the wind speed picks up from about 0 to 12 m/s with wind speeds up to 20 m/s to follow. The wind direction is remarkably steady in this period. Notice about the spikes in the wind direction that most of them correspond to a very small angle change around North. Moreover, when interpolating directions the angles have been split in sine and cosine, interpolated, and then the angles have been re-calculated. Interpolating 0 and $2\pi$ to $\pi$ would obviously be wrong.

The pyranometer has the limitation of only being able to measure radiation of angles of incidence lower than 82°. Also, notice that the readings from the
Figure D.28 Data from the weather station from January 2013. The period features an extreme temperature change on January 21st followed by a storm. Notice that the pyranometer seems badly calibrated.
D.9 Solar thermal system

The building features a solar thermal system providing energy for hot domestic water usage. The system is treated in (Dragsted 2011).

The system consists of six panels totalling an area of 8.31 m² (See Figure D.29). They are tilted 70° from level, and the orientation is 124° clockwise from North.

Temperatures and flows between the solar panels, the hot water tank, the buffer tank, and the radiator are measured. Also, an entry in the database contains...
the power produced by the system. The daily mean of this power has been calculated and plotted for 2012 in Figure D.30. Apparently, either the system or the data logging was not working the first 5 months of the year and maybe from November again. Moreover, the power is sometimes negative. It should be cleared what exactly this measures before using it.

As sketched in Figure D.2 the pipes from solar thermal panels go to the a heat exchanger in the domestic hot water tank. In case that the water in the hot water tank is already sufficiently warm, the water can be lead to a buffer tank or a radiator. At the time when (Dragsted 2011) was written, the buffer tank was not installed, and the surplus of hot water would go to the radiator. The radiator has been installed inside the house and then moved out because it was heating too much during summer. The buffer tank which is well insulated and located under the house is supposed to facilitate keeping the hot water longer so the radiator is no longer needed. It was installed in April 2011 and has a capacity of 800 l. It is designed so that hot and cold water mixes minimally.
The temperature is measured at four heights in the tank. The four temperature measurements are plotted for 2012 in Figure D.31. As expected the temperatures are higher in summer, and the temperature is generally increasing from bottom to top. However, the temperature curves do not go lower than -3°C and -1°C. This could be because of settings of the sensors or the data acquisition. But it should be checked whether these temperatures are actually representative. Also a maximum temperature of 10°C seem low.

According to Bo Holdt-Simonsen, the buffer tank has probably never been in use in the system.

D.10 Data logging

This section describes the overall Building Management System (BMS) installed in the building.

D.10.1 Network overview

The BMS system consists of a PC, on which SCADA software to monitor and log data is installed. The PC is connected to a local TCP/IP network on which the PLC and the router also are connected. The router is also connected to the internet for remote access. An network overview is found in Section D.15

D.10.2 Data acquisition hardware

The data acquisition hardware consists of the following:

- Schneider TSX P57 2634M Premium PLC [Programmable logic controller]
  - IO modules [Physical In and Out put modules for direct connected sensors and actuators ]
- Thermokon SRC65 ModBus to Wireless gateway [Gateway between PLC and wireless sensors]
  - Thermokon wireless sensors
- RESI ModBus to MBus gateway [Gateway between PLC and M-Bus meters]
– Kamstrup and Brunata energy meters

The PLC, with its connected IO modules, is programmed to monitor, control and regulate the heating, ventilation and other systems, as described earlier, either through the direct connected sensors and actuators or via gateways to wireless sensors and energy meters.

D.10.3 Data acquisition software

The data acquisition software consists of the following:

- Schneider Vijeo Citect: V7.20 [Supervisory Control and Data Acquisition (SCADA) – GUI software]
- Schneider OFS server: V3.34 [OPC data server – communication driver between PLC and SCADA software]
- Schneider Vijeo Historian: V4.30 [Data logger to Microsoft SQL server]
  - Microsoft SQL server [SQL database for data logged Vijeo Citect Tags]
- Schneider Unity Pro M: V5.0 [PLC programming software]
- Team Viewer [Remote access software for access through the internet]

Data to/from the PLC and SCADA software (Vijeo Citect V7.20) moves through the OPC server (OFS server V3.34) which acts as a communications driver between the PLC and the SCADA software. See Figure D.32.

Data that has to be logged and saved for historical analysis are chosen in Vijeo Citect and configured in Vijeo Historian which acts as a SQL Server client and selected meter-, sensor- and alarm- data are stored for extended periods, independently of a SCADA system, in an “Historian” SQL database for later access and historical analysis. See Figure D.33.

Data logged to the Historian can be accessed directly from the host SQL server database or from the Excel client.
D.10 Data logging

Figure D.32 The OFS interface between PLC and SCADA client.

Figure D.33 Vijeo Historian server-based architecture to collect and redistribute data.

D.10.4 Backup system

The data logging system in Sisimiut have several weak points, including easy write access from within the house, dependency on a consumer range laptop, a non-mirrored hard drive with all the risks that this implies (disk failures, theft, fire), and others. Moreover, the temperature in the room with the boiler used for the PLC system and the data logging, is high and certainly adds to the risk
of hardware failures (and especially disk failures). High risk or not, backup of data is essential, and a complete backup of the database is being taken every month. Throughout most of 2012, an incremental mirroring was running but due to occasional Internet connection failures from the house, the mirroring would stall and have to be re-initiated manually. Now a complete backup is transferred every month. This has been configured by Ole Brandt.

D.11 Experimental planning

For analyzing the heat dynamics of a building, experiments with the heat input is central. A module to “Unity Pro M” has been written by Jakob Nørby to schedule open/close signals to floor heating valves in the guest apartment (5 loops) plus the scullery (1 loop). The 6 floor heating strings are controlled individually and for each 30 minutes. Values for 672 time steps can be set, corresponding to an experimental plan of 14 days. The module can be controlled both directly through the “Unity Pro M” interface and using an Excel macro also written by Jakob Nørby.

D.11.1 Using Unity Pro M

To use the Unity Pro M interface do the following:

- Open Unity Pro M from the “Start” menu choosing sisimiut_lavenergihus.stu.
- In Unity Pro M, click PLC -> Connect.
- In Project Browser, double click Project -> Animation Tables -> Recept.

Figure D.34 shows a screendump of the interface. The value StartRecept controls if the module is active or not. Recept_sp contains the schedule to be used. The signal is binary and given as a sum of decimal values in Table D.2. To have no floor heating in the listed rooms, write 0. To have heating in the living room and the small room, write 5. To have heating in the whole guest apartment and in the scullery, write 63. In Figure D.34 the plan implemented means no heating for 3.5 hours, then heating in all the rooms in the guest apartment but not in the scullery. sp[0] is not used. If the pointer is “0” when the module is started, it will immediately switch to 1. When StartRecept is set to 1, any
Figure D.34  The Unity Pro M interface to the module implementing planned floor heating in the guest apartment.

Table D.2  The elements to control with the experimental planning module.

<table>
<thead>
<tr>
<th>Bit</th>
<th>Valve</th>
<th>Room</th>
<th>Decimal value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>A01_MG07</td>
<td>Living room</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>A01_MG08</td>
<td>Corridor</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>A01_MG09</td>
<td>Small room</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>A01_MG10</td>
<td>Large room</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>A01_MG11</td>
<td>Bathroom</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>A01_MG12</td>
<td>Scullery</td>
<td>32</td>
</tr>
</tbody>
</table>
other control of floor heating in the rooms than listed in Table D.2 is overruled. To edit the fields, make sure to have the Modifications button activated.

The binary series can be seen and edited directly by pressing Shift-F3. Press F3 to edit in decimal values again.

### D.11.2 Using MS Excel

In stead of typing in the experimental plan manually, it is possible to read a column of entries from Excel. This enables the user to generate the experimental plan with other software and feed it to the controller.

The spread sheet to use is called “Recept parametre.xls” and is located in “C:\Sisimiut\OFS”. A copy should be edited in stead of the original. A screen dump of the interface is seen in Figure D.35.

Do the following to use the Excel interface to implement an experiment plan:

- Write the values to submit to the PLC in the “Write” column.
- Press “Start”.
- Press “Write”. The values in the Write column should now be copied to the Read column. If not, something went wrong in the communication with the PLC.
- Go to the Unity Pro M interface and start the experiment by setting StartRecept to 1 whenever you want.

The user does not need to use the “Stop” button which may make Excel crash. If that happens, it has no consequences for the PLC. Also when copying hundreds of elements, Excel may crash. One may need to copy only about 200 at a time.

### D.12 Data extraction

Several methods are available for extraction of data from the system, and they will be described in this section. Most of the methods access the server in Sisimiut directly. However, this is not advised. Using the backup server at DTU to access the data is faster unless the user is physically in the low energy
D.12 Data extraction

When choosing to export to a text file, one will obtain a table with commas as
decimal points and white spaces as field separator. The encoding of this file will be utf-16. At least in Linux/Unix systems it may be necessary to convert this into utf-8. `iconv` can be used for this

```
$ iconv -f UTF-16 -t UTF-8 oldfile.txt > newfile.txt
```

In **R** the file can now be read with

```
R Example D.12.1

flh <- read.table("newfile.txt", header = TRUE, dec = ",")
```

The time stamp written to the file is in decimal days after Jan 1st 1900 00:00 UTC.

Only a very limited selection of the total data can be extracted in this way. The data obtained in the different sections is listed in Table D.3.

<table>
<thead>
<tr>
<th>Table D.3</th>
<th>The data series that can be extracted from the gui.</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Gulvvarme” (Floor heating)</td>
<td>A01_FM01-A01_FM06, A01_FR01, A01_RT07, A01_MG07 - A01_MG12</td>
</tr>
<tr>
<td>“Gæstebolig” (Guest apartment)</td>
<td>A01_TS01 - A01_TS12, A01_PI01, A01_PI02, C01_TT01, C01_VU01_drift</td>
</tr>
<tr>
<td>“Gæstebolig temp” (Guest apartment temperature)</td>
<td>A01_RU01-A01_RU05, A01_RU01_sp - A01_RU05_sp, A01_RU11, A01_RU12</td>
</tr>
<tr>
<td>“Vejr” (Weather)</td>
<td>A01_UE01, A01_VH01, A01_VR01, A01_LU01</td>
</tr>
</tbody>
</table>

**D.12.2 Data extraction from the MSSql server**

Data can be fetched directly from the MSSql server in Sisimiut by – on the local system – starting the tool called “Import and export data (32 bit)”.

1. Choose a Data Source
   - Data source: SQL Server Native Client 10.0
   - Server name: LP-14444\VIJEOHISTORIAN
   - Authentication Use SAL Server Authentication
   - Database: SisimiutData
2. Choose a Destination Destination: Flat File Destination File name: <you chose> Locale: English (United States), unicode
   Format: Delimited Text qualifier: <none>

3. Configure Flat File Destination Row delimiter: CRLF Column delimiter: Comma , Per default, data will be appended to the destination file. Go to “Edit Mappings...” to edit this and other settings.

4. Specify Table Copy or Query Copy data from one or more tables or views.

5. Save and Run “Package Run” immediately.

when exporting tags, “^A” and “^B” must be replaced. In Emacs, C-q C-a, C-q C-b are keystrokes representing those characters. This is again an encoding issue.

D.12.3 Using the SQL server at DTU

The backup facility (Section D.10.4) provides a fast SQL server located at DTU with the data. From within the DTU network, this server can be reached with any SQL client, given that one has an account on the server with read access. This is the recommended way to access the data. In R, functions have been written based on the RODBC library. Say, one wants to fetch the data from the room temperature in the bathroom and the registrations of presence in the living room (both in the guest apartment) in the period from February 20 to March 13 2012. This is done with:

```r
R Example D.12.2

names <- c("A01_PI01", "A01_RU04")
data.raw <- fetch.sensor(name = names, from = "2012-02-20", to = "2012-02-21")
```

The database has two tables, one for decimal measurements called NumericSamples and one with binary measurements called DigitalSamples. `fetch.sensor` looks up where to find the given sequences, creates the SQL queries and fetches the data. The SQL commands executed by `fetch.sensor` in this case are
fetch.sensor returns a list of dataframes. The names of the dataframes are the names of the measurements, in this case A01_PI01 and A01_RU04. These dataframes contain two columns each, one called time with time stamps in POSIXct format, and one called values containing the actual measurements. Each series of measurements have their own time column because they are not necessarily identical. If one wants to obtain shared time stamps, a filter must be applied. Such filters have also been implemented in R.

D.12.4 Using extracted data

When the raw data comes from the database, it needs some processing before it can be used for many time series applications. This is due to asynchronous sampling of signals and uneven sampling periods. Often the user will want to interpolate or average to some fixed timestamps before modeling the data.

Normally all measurements are logged every 30 seconds. The system uses a so-called “deadband” to avoid redundant sampling. That is if a signal has the same value plus/minus the deadband at the succeeding measurement, the measurement will not be logged. This has to be taken into account for interpolation or averages to be correct. The deadband is set to zero for all signals but in case the signal is exactly constant, it will still influence. This is often the case for e.g. flow measures (when there is no flow) and of course the binary signals. A function called revive.band has been written to re-construct the signal with 30
second resolution.

R Example D.12.3

```r
## revive one signal
A01PI01.revived <- revive.band(data.raw$A01_PI01)
## revive multiple signals
data.revived <- lapply(data.raw, revive.band)
## revive multiple signals using multithreading
data.revived <- mclapply(data.raw, revive.band, mc.cores = 2)
```

After "reviving" the signals, the data is ready to be interpolated, averaged, or processed in some other way in order to synchronize sampling and obtaining the sampling rate that the user wants. How to carry out this step depends completely on the processing wanted. A few different schemes have been implemented in an R function called `mergedata`.

- Interpolation between nearest points is available with the `approx` function in R. This method has the serious backdraw that it potentially disregards most of the data. For instance, if measurements are taken every 30 seconds, and one wants a 30 minute resampling, only 2 out of 60 measurements are used.

- Averaging the raw data is the second option. Then all of the 60 data points in the example above is used. However, in case of uneven distance between sampling times in raw data, they are not weighted correctly. This can be a serious problem in situations where sampling rates are erratic and/or signals are non-linear.

- The last option implemented interpolates to a specified sampling rate first, and then it averages. If interpolating to 30 seconds before averaging, all data points should be used, and the averaging can correctly use even weighting of the interpolated points.

Here follows examples on how to obtain sub-sampling taking into account these issues:

R Example D.12.4
```r
## a series of time stamps to average to
synctime <- seq(from = as.POSIXct("2012-02-20", tz = "UTC"),
              to = as.POSIXct("2012-01-21", tz = "utz"), by = 15 * 60)
## interpolation using nearest two points.
data.sync.15min <- mergedata(data.revived, mergeby = "time",
                          time = synctime, approxfun = approx, parallel = TRUE)
## averaging over +/-7.5 minutes
data.sync.15min <- mergedata(data.revived, mergeby = "time",
                          time = synctime, approxfun = resample, parallel = TRUE,
                          h = 15 * 60 - 1, kernel = "mean", pastonly = FALSE,
                          na.rm = TRUE)
## averaging over +/-7.5 minutes insuring correct weighting.
data.sync.15min <- mergedata(data.revived, mergeby = "time",
                          time = synctime, approxfun = resample, parallel = TRUE,
                          h = 15 * 60 - 1, kernel = "mean.int", int.step = 30,
                          pastonly = FALSE, na.rm = TRUE)
```

The `h` argument is the width of the interval considered. The `pastonly` argument decides whether the algorithm should only look back in time to evaluate the filtered value. This should be set to `TRUE` if one wants to simulate real-time modeling. `parallel` controls whether multithreading should be enabled, and `na.rm` controls if missing data points should simply be discarded or result in a missing point. The latter gets very important as the `h` increases since there will often be at least one missing data point in say one day.

Figure D.36 plots the obtained data using these steps resulting in interpolated signals of 15 minute resolution. From the raw (black) data to the revived (red) the sampling frequency is ensured to be constant by repeating preceding values. The green lines are interpolations to sample times predefined by the user. The green signals can be compared because they share sample times.

Table D.4 in Section can be read into R, and used to select sensors to retrieve. The following example fetches all data related to the floor heating system in the guest apartment:

```r
R Example D.12.5

sensor.names <- sensors$Tag[sensors$Group == "Floor heating" &
sensors$Apartment == "Guest"]
```
which can then be retrieved using `fetch.sensor`. Actually, `fetch.sensor` uses
the column "db\_table" to determine which table in the database to retrieve the
signals from.

## D.13 Conclusions

A low-energy house in Sisimiut, Greenland has been equipped with measurement
and control equipment enabling detailed surveillance and experiments related
to heat dynamics of the building, influence of occupant behavior, and storage
strategies of power generated solar radiation.

Different systems in the building have been described, and examples on mea-
urements from the data acquisition system have been given. This includes
consumption variables, the heating systems, the solar thermal system, occup-
ancy indicators, and the weather station. In the description, several issues
about the data series were addressed. In the treatment of the temperature mea-
surements from the building, an example on principal component analysis was
given indicating dynamical properties in one apartment of the house.

The data acquisition and control unit can be used to execute planned experi-
ments as well. It was described how to do this for experiments with the floor
heating system in an apartment available for research.
Several methods of data extraction were described, and R functions have been developed to handle and process the data for statistical modeling.
Table D.4 provides an overview of the sensors installed in the house. The table is not complete. For each sensor it contains the following fields:

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag</td>
<td>Tag name in the database.</td>
</tr>
<tr>
<td>Description</td>
<td>a few keywords describing the sensor.</td>
</tr>
<tr>
<td>Unit</td>
<td>The unit of the data.</td>
</tr>
<tr>
<td>Apartment</td>
<td>The apartment it relates. “Rent”, “Guest”, or “Common”.</td>
</tr>
<tr>
<td>db.table</td>
<td>The table in which it is found in the database.</td>
</tr>
<tr>
<td>Group</td>
<td>A grouping of the sensors that can be used when retrieving data.</td>
</tr>
<tr>
<td>Sensor</td>
<td>The type of sensor providing the measurements.</td>
</tr>
<tr>
<td>Tag</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>A02_CP01</td>
<td>Circulation pump from boiler start/stop</td>
</tr>
<tr>
<td>A02_Drift</td>
<td>Circulation pump, error</td>
</tr>
<tr>
<td>A02_FM01</td>
<td>Flow, oil to boiler</td>
</tr>
<tr>
<td>A02_FM01_TRIP</td>
<td>Trip cumulative flow, oil to boiler</td>
</tr>
<tr>
<td>A02_FM01_RUN</td>
<td>Cumulative flow, oil to boiler</td>
</tr>
<tr>
<td>A02_FM01</td>
<td>Temperature boiler, forward</td>
</tr>
<tr>
<td>A02_RT01</td>
<td>Temperature boiler water, before heat exchanger</td>
</tr>
<tr>
<td>A02_RT02</td>
<td>Temperature boiler water, after heat exchanger</td>
</tr>
<tr>
<td>A02_STOP</td>
<td>Boiler stop</td>
</tr>
<tr>
<td>A04_FM01</td>
<td>Flow, hot water from tank</td>
</tr>
<tr>
<td>A04_FM01_SUM</td>
<td>Cumulative hot water consumption</td>
</tr>
<tr>
<td>A04_FR01</td>
<td>Temperature, from boiler to spiral, forward</td>
</tr>
<tr>
<td>A04_FR01</td>
<td>Temperature, spiral to boiler, return</td>
</tr>
<tr>
<td>A04_FR01</td>
<td>Temperature, tank, top</td>
</tr>
<tr>
<td>A04_FR01</td>
<td>Temperature, tank, bottom</td>
</tr>
<tr>
<td>A01_MG01</td>
<td>Hall On/off</td>
</tr>
<tr>
<td>A01_MG01</td>
<td>Bathroom</td>
</tr>
<tr>
<td>A01_MG01</td>
<td>Large room On/off</td>
</tr>
<tr>
<td>A01_MG04</td>
<td>Small room On/off</td>
</tr>
<tr>
<td>A01_MG05</td>
<td>Corridor</td>
</tr>
<tr>
<td>A01_MG06</td>
<td>Living room On/off</td>
</tr>
<tr>
<td>A01_MG07</td>
<td>Living room On/off</td>
</tr>
<tr>
<td>A01_MG08</td>
<td>Corridor</td>
</tr>
<tr>
<td>A01_MG09</td>
<td>Large room On/off</td>
</tr>
<tr>
<td>A01_MG10</td>
<td>Small room On/off</td>
</tr>
<tr>
<td>A01_MG10</td>
<td>Bathroom</td>
</tr>
<tr>
<td>A01_MG11</td>
<td>Scullery, flow</td>
</tr>
<tr>
<td>A01_MG12</td>
<td>Scullery, flow</td>
</tr>
</tbody>
</table>

Table D.4 List of the sensors in the building and their names in the database.
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Unit</th>
<th>Apartment</th>
<th>db.table</th>
<th>Group</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A01_M02</td>
<td>Bathroom, flow</td>
<td>l/h</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M03</td>
<td>Large room, flow</td>
<td>l/h</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M04</td>
<td>Small room, flow</td>
<td>l/h</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M05</td>
<td>Corridor, flow</td>
<td>l/h</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M06</td>
<td>Living room, flow</td>
<td>l/h</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M07</td>
<td>Hall, flow</td>
<td>l/h</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M08</td>
<td>Bathroom, flow</td>
<td>l/h</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M09</td>
<td>Large room, flow</td>
<td>l/h</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M10</td>
<td>Small room, flow</td>
<td>l/h</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M11</td>
<td>Corridor, flow</td>
<td>l/h</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M12</td>
<td>Living room, flow</td>
<td>l/h</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Brunata HGQ1qp 1,2 m3/h</td>
</tr>
<tr>
<td>A01_M13</td>
<td>Bathroom return temp</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M14</td>
<td>Bathroom return temp</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M15</td>
<td>Large room return temp</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M16</td>
<td>Small room return temp</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M17</td>
<td>Combined return temp</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M18</td>
<td>Living room return temp</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M19</td>
<td>Hall return temp</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M20</td>
<td>Bathroom return temp</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M21</td>
<td>Large room return temp</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M22</td>
<td>Small room return temp</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M23</td>
<td>Corridor return temp</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M24</td>
<td>Living room return temp</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Floor heating</td>
<td>Sensonic LTT4205 (0-100°C)</td>
</tr>
<tr>
<td>A01_M25</td>
<td>Large room</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_M26</td>
<td>Small room</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_M27</td>
<td>Corridor</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_M28</td>
<td>Bathroom</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_M29</td>
<td>Living room</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
</tbody>
</table>

Continues on next page...
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Unit</th>
<th>Apartment</th>
<th>db.table</th>
<th>Group</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A01_RU06</td>
<td>Large room</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_RU07</td>
<td>Small room</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_RU08</td>
<td>Corridor</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_RU09</td>
<td>Bathroom</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_RU10</td>
<td>Living room</td>
<td>°C</td>
<td>Rent</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_RU11</td>
<td>Scullery</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>A01_RU12</td>
<td>Hall</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Room temperatures</td>
<td>Thermokon SR 04</td>
</tr>
<tr>
<td>C01_RU01</td>
<td>Living room, 1 from top</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Std temp meas</td>
<td>Thermocouple, type T, transmitter: Seneca K121</td>
</tr>
<tr>
<td>C01_RU02</td>
<td>Living room, 2 from top</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Std temp meas</td>
<td>Thermocouple, type T, transmitter: Seneca K121</td>
</tr>
<tr>
<td>C01_RU03</td>
<td>Living room, 3 from top</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Std temp meas</td>
<td>Thermocouple, type T, transmitter: Seneca K121</td>
</tr>
<tr>
<td>C01_RU04</td>
<td>Living room, 4 from top</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Std temp meas</td>
<td>Thermocouple, type T, transmitter: Seneca K121</td>
</tr>
<tr>
<td>C01_RU05</td>
<td>Living room, 5 from top</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Std temp meas</td>
<td>Thermocouple, type T, transmitter: Seneca K121</td>
</tr>
<tr>
<td>C01_RU06</td>
<td>Living room, 6 from top</td>
<td>°C</td>
<td>Guest</td>
<td>numerical</td>
<td>Std temp meas</td>
<td>Thermocouple, type T, transmitter: Seneca K121</td>
</tr>
<tr>
<td>A01_PI01</td>
<td>PIR, Living room</td>
<td>1=activity</td>
<td>Guest</td>
<td>digital</td>
<td>Occupancy</td>
<td>PEHA 482 FU-BM DE</td>
</tr>
<tr>
<td>A01_PI02</td>
<td>PIR, Corridor</td>
<td>1=activity</td>
<td>Guest</td>
<td>digital</td>
<td>Occupancy</td>
<td>PEHA 482 FU-BM DE</td>
</tr>
<tr>
<td>A01_TS01</td>
<td>Door, scullery</td>
<td>1=closed</td>
<td>Guest</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS02</td>
<td>Door, terrace</td>
<td>1=closed</td>
<td>Guest</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS03</td>
<td>Door, hall</td>
<td>1=closed</td>
<td>Guest</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS04</td>
<td>Door, scullery</td>
<td>1=closed</td>
<td>Rent</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS05</td>
<td>Door, terrace</td>
<td>1=closed</td>
<td>Rent</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS06</td>
<td>Door, hall</td>
<td>1=closed</td>
<td>Rent</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS07</td>
<td>Door, back</td>
<td>1=closed</td>
<td>Common</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS08</td>
<td>Door, front</td>
<td>1=closed</td>
<td>Common</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TS09</td>
<td>Window, large room</td>
<td>1=closed</td>
<td>Guest</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>A01_TU10</td>
<td>Window, small room</td>
<td>1=closed</td>
<td>Guest</td>
<td>digital</td>
<td>Occupancy</td>
<td>Enocean STM 250</td>
</tr>
<tr>
<td>C01_CO01</td>
<td>CO₂ out</td>
<td>ppm</td>
<td>Guest</td>
<td>numerical</td>
<td>Occupancy</td>
<td></td>
</tr>
<tr>
<td>C01_FM01</td>
<td>Flow in</td>
<td>m³/h</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_FM02</td>
<td>Flow out</td>
<td>m³/h</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_TT03</td>
<td>Pressure drop heat exchanger</td>
<td>??</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td>Carlo Gavazzi E83-2050</td>
</tr>
<tr>
<td>C01_TT04</td>
<td>Cooker hood, el current</td>
<td>A</td>
<td>Guest</td>
<td>numerical</td>
<td>Ventilation</td>
<td>Carlo Gavazzi E83-2050</td>
</tr>
<tr>
<td>C01_TT05</td>
<td>Temperature, outtake, after heat exchanger</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_VU01</td>
<td>Intake</td>
<td>1=On,0=off</td>
<td>Common</td>
<td>digital</td>
<td>Ventilation</td>
<td></td>
</tr>
</tbody>
</table>

Continues on next page...
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Unit</th>
<th>Apartment</th>
<th>db table</th>
<th>Group</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>C01_HY01</td>
<td>Rel. humidity, Outtake</td>
<td>%</td>
<td>Guest</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY02</td>
<td>Rel. humidity, Outtake</td>
<td>%</td>
<td>Rent</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY03</td>
<td>Rel. humidity, Outtake, before heat exchange</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY04</td>
<td>Rel. humidity, Outtake, after heat exchange</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY05</td>
<td>Rel. humidity, Intake, before heat exchange</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY06</td>
<td>Rel. humidity, Intake, after heat exchange</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY07</td>
<td>Rel. humidity, Basement pos. 1</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Basement</td>
<td></td>
</tr>
<tr>
<td>C01_HY08</td>
<td>Rel. humidity, Basement pos. 2</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Basement</td>
<td></td>
</tr>
<tr>
<td>C01_HY09</td>
<td>Rel. humidity, Basement pos. 3</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Basement</td>
<td></td>
</tr>
<tr>
<td>C01_HY10</td>
<td>Temp, intake, after heat exchange, before after heating</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY11</td>
<td>Temp, intake, after heat exchange, before after heating</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY12</td>
<td>Temp, intake, after heat exchange and after heating</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
</tr>
<tr>
<td>C01_HY13</td>
<td>Control valve, heat surface</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C01_HY14</td>
<td>Return temp, heat surface</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Ventilation</td>
<td>Sensonic LTT420S (0-100)</td>
</tr>
<tr>
<td>C01_HY15</td>
<td>Temp, Basement pos. 1</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Basement</td>
<td></td>
</tr>
<tr>
<td>C01_HY16</td>
<td>Temp, Basement pos. 2</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Basement</td>
<td></td>
</tr>
<tr>
<td>C01_HY17</td>
<td>Temp, Basement pos. 3</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Basement</td>
<td></td>
</tr>
<tr>
<td>C01_HY18</td>
<td>Temp, Basement pos. 4</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Basement</td>
<td></td>
</tr>
<tr>
<td>A01_X01</td>
<td>Floor heating, sum</td>
<td>kWh</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td></td>
</tr>
<tr>
<td>A01_X02</td>
<td>Floor heating, instant</td>
<td>kW</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td></td>
</tr>
<tr>
<td>A01_X03</td>
<td>Floor heating, flow, sum</td>
<td>m³</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td></td>
</tr>
<tr>
<td>A01_X04</td>
<td>Floor heating, flow, instant</td>
<td>l/h</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td></td>
</tr>
<tr>
<td>A01_X05</td>
<td>Floor heating, Temp, before</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td></td>
</tr>
<tr>
<td>A01_X06</td>
<td>Floor heating, Temp, return</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Floor heating</td>
<td></td>
</tr>
<tr>
<td>A01_X07</td>
<td>Electricity, cumulative</td>
<td>kWh</td>
<td>Rent</td>
<td>numerical</td>
<td>Consumption</td>
<td></td>
</tr>
<tr>
<td>A01_X08</td>
<td>Electricity, cumulative</td>
<td>kWh</td>
<td>Guest</td>
<td>numerical</td>
<td>Consumption</td>
<td></td>
</tr>
<tr>
<td>A01_X09</td>
<td>Solar circulation pump</td>
<td>Start/stop</td>
<td>Common</td>
<td>digital</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A01_X10</td>
<td>Buffer Circulation pump</td>
<td>Start/stop</td>
<td>Common</td>
<td>digital</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A01_X11</td>
<td>Flow, solar panel</td>
<td>kWh</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A02_MV01</td>
<td>Cumulative flow, solar panel</td>
<td>Common</td>
<td>digital</td>
<td>Solar thermal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continues on next page...
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Unit</th>
<th>Apartment</th>
<th>db_table</th>
<th>Group</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A03_R02</td>
<td>Temperature, glucola to buffer</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td>Sensonic LTT420S (0-100)</td>
</tr>
<tr>
<td>A03_R03</td>
<td>Temperature, after radiator</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td>Sensonic LTT420S (0-100)</td>
</tr>
<tr>
<td>A03_R01</td>
<td>Control valve, solar heat</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R02</td>
<td>Control valve, buffer</td>
<td></td>
<td>Common</td>
<td>digital</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R03</td>
<td>Control valve, radiator bypass</td>
<td></td>
<td>Common</td>
<td>digital</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R04</td>
<td>Control valve, circulation</td>
<td></td>
<td>Common</td>
<td>digital</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R01</td>
<td>Temperature, glucola after solar</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td>Sensonic LTT420S (0-100)</td>
</tr>
<tr>
<td>A03_RT02</td>
<td>Return temp, after buffer</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_RT01</td>
<td>Temperature, buffer, top</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Buffer tank</td>
<td>Sensonic LTT420S (0-100)</td>
</tr>
<tr>
<td>A03_RT02</td>
<td>Temperature, buffer, 2nd from top</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Buffer tank</td>
<td>Sensonic LTT420S (0-100)</td>
</tr>
<tr>
<td>A03_RT03</td>
<td>Temperature, buffer, 3rd from top</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Buffer tank</td>
<td></td>
</tr>
<tr>
<td>A03_RT04</td>
<td>Temperature, buffer, bottom</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Buffer tank</td>
<td></td>
</tr>
<tr>
<td>Solvarme_kW</td>
<td>Energy, sum</td>
<td>kWh</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R01</td>
<td>Power</td>
<td>kW</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R03</td>
<td>Flow, sum</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R04</td>
<td>Flow, instant</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_RT01</td>
<td>Temp, before</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_RT02</td>
<td>Temp, return</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R11</td>
<td>Excess energy, sum</td>
<td>kWh</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R12</td>
<td>Excess energy, Instant</td>
<td>kW</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R13</td>
<td>Excess flow, sum</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R14</td>
<td>Excess flow, Instant</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A03_R15</td>
<td>Excess temp, before</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Solar thermal</td>
<td></td>
</tr>
<tr>
<td>A04_R01</td>
<td>Water in, sum</td>
<td>kWh</td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>A04_R02</td>
<td>Water flow in, instant</td>
<td>m³/h</td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>A04_R03</td>
<td>Water flow in, sum</td>
<td>m³</td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>A04_R04</td>
<td>Water flow in, instant</td>
<td>?</td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>A04_R05</td>
<td>Water temp in, in</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>A04_R06</td>
<td>Water temp out</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>A04_R10</td>
<td>Hot water meter</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>A04_R11</td>
<td>Extra energy to hot water</td>
<td>kWh</td>
<td>Common</td>
<td>numerical</td>
<td>Water cons</td>
<td></td>
</tr>
<tr>
<td>CO1_R01</td>
<td>Ventilation heating, sum</td>
<td>kWh</td>
<td>Common</td>
<td>numerical</td>
<td>Vent heat circuit</td>
<td></td>
</tr>
<tr>
<td>CO1_R02</td>
<td>Ventilation heating, Instant</td>
<td>kW</td>
<td>Common</td>
<td>numerical</td>
<td>Vent heat circuit</td>
<td></td>
</tr>
<tr>
<td>CO1_R03</td>
<td>Ventilation heating flow, sum</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Vent heat circuit</td>
<td></td>
</tr>
<tr>
<td>CO1_R04</td>
<td>Ventilation heating flow, instant</td>
<td></td>
<td>Common</td>
<td>numerical</td>
<td>Vent heat circuit</td>
<td></td>
</tr>
<tr>
<td>CO1_R05</td>
<td>Ventilation heating temp, before</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Vent heat circuit</td>
<td></td>
</tr>
<tr>
<td>CO1_R06</td>
<td>Ventilation heating temp, after</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Vent heat circuit</td>
<td></td>
</tr>
<tr>
<td>A01_R01_middel</td>
<td>Wind speed, ten minutes mean</td>
<td>m/s</td>
<td>Common</td>
<td>numerical</td>
<td>Weather</td>
<td></td>
</tr>
</tbody>
</table>

Continues on next page...
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Unit</th>
<th>Apartment</th>
<th>db.table</th>
<th>Group</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A01_VH01</td>
<td>Wind speed, instant</td>
<td>m/s</td>
<td>Common</td>
<td>numerical</td>
<td>Weather</td>
<td>Young Yo05103L-45</td>
</tr>
<tr>
<td>A01_UE01</td>
<td>Outdoor temperature</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Weather</td>
<td>MB teknik 2.1280.00.141</td>
</tr>
<tr>
<td>A01_VR01</td>
<td>Wind direction (from N), ten minutes mean</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Weather</td>
<td>Young Yo05103L-45</td>
</tr>
<tr>
<td>A01_U01</td>
<td>Global solar radiation</td>
<td>W/m²</td>
<td>Common</td>
<td>numerical</td>
<td>Weather</td>
<td>MB teknik 7.1415.09.041</td>
</tr>
<tr>
<td>TWC</td>
<td>Temperature, corrected for chill factor</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Weather</td>
<td></td>
</tr>
<tr>
<td>C01_TF01</td>
<td>Temp, paper insulation 1</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
<tr>
<td>C01_TF02</td>
<td>Temp, paper insulation 2</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
<tr>
<td>C01_TF03</td>
<td>Temp, paper insulation 3</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
<tr>
<td>C01_TF04</td>
<td>Temp, paper insulation 4</td>
<td>°C</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
<tr>
<td>C01_HY07</td>
<td>Humid, paper insulation 1</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
<tr>
<td>C01_HY08</td>
<td>Humid, paper insulation 2</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
<tr>
<td>C01_HY09</td>
<td>Humid, paper insulation 3</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
<tr>
<td>C01_HY10</td>
<td>Humid, paper insulation 4</td>
<td>%</td>
<td>Common</td>
<td>numerical</td>
<td>Paper insulation</td>
<td>Thermokon LCN-FTK</td>
</tr>
</tbody>
</table>
D.15 Network overview

The following page features an overview of the network infrastructure in the building.
Sisimiut lavenerghus. PLC og Scada netværk

Software versioner:
Windows 7 Enterprise
Unity Pro M : V5.0
Vijeo Citect: V7.20
OPS: V3.34
Vijeo Historian: V4.30
TeamViewer. v6

Licens info:
Vijeo Citect: 500 tags
Vijeo Historian: 500 tags
Full licenses : 1
Web client: 2
SN: 0479 – 97739
Unity SN: 21105101086

Team Viewer Login:
ID: 564 027 677
PW: Sisimiut8607

PC Login:
PC navn: LP-14444
User: PC Service Technician
PW: sisimiut

Historian client Login:
Server: LP-14444
Database: Historian
User name: user
PW: user

WLAN:
SSID: ArtekNet
Key: Anet4all@low!

IPad:
Mocha RDP Login.
DTU lavenerghus
IP: 192.168.254.250
User: PC Service Technician
PW: sisimiut

Adresse:
Bolefhep Aqq. 36
3911 Sisimiut
Lejer: Laarseraq Skifte
Tlf: +299585901

Node 2
Gateway for trådløse sensorer
SRC/STC RS485
MODBUS

Node 10
Mbus gateway
REU
Mbus-MODBUS-RS485

Node 244
Varmtvand
Kamstrup
multical 66
Mbus
SN 4619244

Node 70
Solvarme
overskud
Brunata HQQ1
Mbus
SN 30424870

Node 71
Solvarme panel
Brunata HQG1
Mbus
SN 30424871

Node 72
Ventilation
værmeflade
Brunata HQQ1
Mbus
SN 40630272

Node 73
Gulvvarme
Brunata HQG1
Mbus
SN 40630273

12 Temperatur følere
12 Dør/vindue kontakter
2 PIR følere
References


Low-energy house in Sisimiut – Data overview

Authors:
Philip Delff Andersen¹, Henrik Madsen¹, Carsten Rode².

Published by:
DTU Compute
(Technical report)

¹DTU Compute, Matematiktorvet, Building 322, DK-2800 Kgs. Lyngby, Denmark
²DTU Civil Engineering, Brovej, Building 118, DK-2800 Kgs. Lyngby, Denmark
Abstract

Experiments with persistently exciting heat inputs are a fundamental tool in identification of heat dynamics in buildings. The Low-energy house in Sisimiut, Greenland, provides an advanced experimental setup with frequent measurements of temperatures, heat inputs, and much more. This paper presents an overview of data collected since the installation of the new measurement and control system. Focus is on heat dynamics so only data related to that will be shown. 5 experiments have been conducted. They are described, and resulting data is shown.

E.1 Introduction

A new data acquisition system was installed in a low-energy house in Sisimiut, Greenland, in April 2011. This is documented in Andersen et al. 2013e. Since then, two heating seasons have past. Here, a quick overview will be given of the data collected. Floor heating power is the main heat input in the building, and it is measured by different sensors.

For identification of heat dynamics of buildings, experiments that excite the dynamics are crucial. This can be achieved by designing the heat inputs with pseudo random binary signals (PRBS). This paper presents 5 experiments designed using PRBS signals. Three of the experiments are designed with the aim to model the apartment as one large room, using common heating signals in all rooms and open doors between them. Two experiments are designed with multi-room models in mind. They were both run with closed doors between the rooms and different floor heating inputs for each room. Moreover, in the last experiment a combination of floor heating and electric radiator heating is used. This enables separate models for the two heat sources which can provide valuable information about the floor heating system.

Data especially relevant for modeling heat dynamics of the building or of the guest apartment will be shown. Much more data is accessible as documented in Andersen et al. 2013e. Some of the data acquired before the installation of this new system was analyzed in Andersen et al. 2013c.
E.2 Overview of the collected data

This section presents a brief overview of the data that has been collected since the installation of the new measurement system. Here, a single measurement (the signal called A01_X02) for all the floor heating in the building will be plotted. For a common indoor temperature, the room temperature measurements have been weighted by the surface areas of the rooms. Also outdoor temperature, solar radiation, and wind speed are plotted. All signals have been averaged to daily values. The data for the two periods is seen in Figures E.1 and E.2. Refer also to the nomenclature in Table E.1.

| \(P_s\) (W/m²) | Global horizontal solar radiation measured on weather station. |
| \(P_v\) (W) | Estimate of ventilation heating to one apartment. |
| \(T_a\) (°C) | Outdoor temperature measured on weather station. |
| \(T_i\) (°C) | Area weighted indoor temperature. In experiments, only for guest apartment. |
| \(P_h\) (W) | Floor heating. In experiments, it’s the sum of the contributions to the rooms in the guest apartment. In the data overview, it’s the total floor heating supply which is measured by another sensor. |
| \(W_s\) (m/s) | Wind speed measured on weather station. |
| \(y\) (°C) | Mean of temperature measurements in central column in Guest apartment. |

The floor heating signal misses many data points. It is possible that the measurements on the floor heating loops in the individual rooms will give more information. However, notice that the sum of these normally exceeds the one plotted. They have to be scaled as argued in Andersen et al. 2014b.

E.2.1 Energy signatures

To get a quick idea of the energy signature of the building in the two periods, the following ARX model (Madsen 2008) has been fitted to average daily values:

\[
P_h(n) = P_h(n-1) + T_i(n) + T_i(n-1) + T_a(n) + T_a(n-1) + P_s(n) + e(n) \tag{E.1}
\]

where \(P_h\) is floor heating, \(T_i\) is indoor room temperatures of the whole building weighted by area, \(T_a\) is outdoor temperature, \(P_s\) is global horizontal solar radia-
Figure E.1 Data from the first heating season. The indoor temperature is an area-weighted average for the whole building.
Figure E.2 Data from the second heating season. The indoor temperature is an area-weighted average for the whole building.
ation, and \( n \) is the day number. \( \{e(n)\} \) are i.i.d. following \( N(0, \sigma^2) \). Table E.2 contains derived physical parameter estimates using the same methods as in Andersen et al. 2013c.

### Table E.2 Physical properties derived from Model E.1.

<table>
<thead>
<tr>
<th></th>
<th>Period 1</th>
<th></th>
<th>Period 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>std</td>
<td>est</td>
<td>std</td>
</tr>
<tr>
<td>UA, W/K</td>
<td>145.79</td>
<td>16.54</td>
<td>141.78</td>
<td>6.36</td>
</tr>
<tr>
<td>gA, –</td>
<td>9.27</td>
<td>6.74</td>
<td>12.07</td>
<td>2.91</td>
</tr>
</tbody>
</table>

#### E.3 Conducted experiments

This section briefly describes and provides overviews of data obtained from the experiments carried out in the guest apartment in the period February 2012 – March 2013. Table E.3 gives an overview of the 5 conducted experiments, their periods of execution, duration, if internal doors have been open or closed, which PRBS signal has been used for floor heating, and if electric radiators have been used. The ventilation system has been controlled from measurements humidity or it has been shut off by the tenants of then rental apartment.

### Table E.3 Overview of the five conducted experiments. The two first are intended for estimation of properties on the whole apartment, the 3 others on each room separately. Moreover, the last experiment can be used to estimate heat capacity of the floor heating system.

<table>
<thead>
<tr>
<th>Exp</th>
<th>start</th>
<th>end</th>
<th>#days</th>
<th>doors</th>
<th>PRBS</th>
<th>radiators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Feb 21 2012</td>
<td>Mar 09 2012</td>
<td>16.6</td>
<td>open</td>
<td>A</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>Jan 09 2013</td>
<td>Jan 16 2013</td>
<td>7</td>
<td>open</td>
<td>A</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>Jan 16 2013</td>
<td>Feb 04 2013</td>
<td>18.2</td>
<td>closed</td>
<td>B</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>Feb 22 2013</td>
<td>Mar 10 2013</td>
<td>16.5</td>
<td>closed</td>
<td>B</td>
<td>no</td>
</tr>
<tr>
<td>5</td>
<td>Mar 12 2013</td>
<td>Mar 24 2013</td>
<td>12.6</td>
<td>closed</td>
<td>C</td>
<td>yes</td>
</tr>
</tbody>
</table>
E.3 Conducted experiments

E.3.1 Experiment 1 – A single-room experiment

The aim of the experiment is estimation of thermal properties of the experimental apartment as a whole. For obtaining this, the floor heating inputs to all rooms are being varied together, and the dynamic thermal response of the house is being studied. All outer doors and windows of the apartment were kept closed, and all inner doors were kept wide open. This is to be able to model the envelope as well as possible without internal thermal resistances between rooms.

The experiment started on February 21 and finished on March 9, 2012. Plots of the obtained data are shown in Figure E.3. (Andersen et al. 2014b) presents graybox modeling of this data set. Also, many of the examples in (Andersen et al. 2013e) are from this experiment.

To vary the floor heating with as wide a range of frequencies as possible and independently of other inputs, pseudo random binary signals (PRBS) were used (Godfrey 1980). To get as large an excitation of the system as possible, the floor heating was either completely off or at maximum power in accordance with the designed signal. The expected mean value of a PRBS is 0.5. this means that if such a signal is followed, the heating system will in average deliver half of its maximum power.

From simple experiments and rough calculations, it was seen that following a PRBS would give too little a heating input to keep the indoor temperature within normal operation temperatures (say 17°C-26°C) because the floor heating system was relatively small compared to the low outdoor temperatures. Therefore, after the PRBS were generated, the length of the heating periods were doubled.

The heating system has an experimental module where floor heating schedules can be implemented (documented in Andersen et al. 2013e). It works with half hour steps so the PRBS are interpreted as one step per half hour. Also from rough calculations from initial experiments, it was estimated that shifts after less than one hour with or without heating was unnecessary.

Three PRBS were generated. One of fast shifting regimes, and two of slower shifts. To generate a PRBS one has to choose the length of the register \((n)\), the initial value of the register \((R_0)\), and the shortest allowed length of a subsequence of zeros or ones \((\lambda)\). The function used for the generation of the signals was written by Peder Bacher. After the generation of the PRBS signal and the scaling of the heating periods, the length of the signals did not match the wanted experiment duration. Hence, the were cut off to match the available time. Also the first signal was started at the second value instead of the first.
The starting indexes and the length of the used sequences are listed together with the properties of the three signals in Table E.4. Figure E.4 shows the signals used, the three signals colored differently. There are no gaps between the signals. The same signal was shown in Figure E.3.

Table E.4  Options used for generation of the PRBS signals used in the design of Experiment 1.

<table>
<thead>
<tr>
<th>Signal</th>
<th>n</th>
<th>$R_0$</th>
<th>$\lambda$</th>
<th>start</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>402</td>
<td>2</td>
<td>2</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>982</td>
<td>7</td>
<td>1</td>
<td>579</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>102</td>
<td>7</td>
<td>1</td>
<td>124</td>
</tr>
</tbody>
</table>
Figure E.3 The indoor air temperature together with temperatures in the scullery and entrance hall followed by weather data, heating signals, and a view of the incidences of disturbances of the experiment.
Figure E.4  The PRBS signals used for Experiment 1.
E.3 Conducted experiments

E.3.2 Experiment 2 – repetition of Experiment 1

The second experiment repeats the floor heating signal from Experiment 1 starting on January 9 at 4:35:00 PM UTC and ending on January 16 at 6:08:00 AM UTC. The doors inside the apartment were kept open like in Experiment 1. Data is plotted in Figure E.5. Before this experiment, a HOBO data logger was placed in the boiler room to measure the temperature every minute. This was in use during Experiment 2–5. This temperature measurement will be denoted $T_b$.

While there is considerably less solar radiation at this time of year, the outdoor temperature is higher during Experiment 2 than it was in Experiment 1. The indoor temperature stays within 15 and 21°C. The experiment seems to have been disturbed a couple of times.
Figure E.5 Mean temperature measured in the thermometer column, weather data, and heat input data for Experiment 2.
E.3 Conducted experiments

E.3.3 Experiment 3 – A multi-room experiment

The third experiment was carried out with closed doors between the rooms inside the apartment and with different PRBS signals controlling the floor heating in each room.

The experiment was started on January 16 2013 at 8:11:00 PM UTC and ran until February 2 early in the morning. It was interrupted two times by opening of doors, and the last of these incidents was on February 1st at 11 AM UTC, where the data set can be cut. That will give about 15 days of data.

Figure E.7 shows the temperatures and floor heating inputs for each of the rooms. All the temperature signals have many fluctuations, and the between-rooms temperature differences are quite large. Where the temperatures are between 13 and 20°C in the small, the large room, the corridor and the living room, the bathroom is significantly warmer – 21 to 27 °C. The bathroom has no cold walls, and it is neighboring the boiler room.

<table>
<thead>
<tr>
<th>Table E.5</th>
<th>Options used for generation of the PRBS signals used in the design of Experiment 3.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>n</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
</tr>
</tbody>
</table>

Table E.5 provides an overview of how the PRBS signals used for Experiment 3 were created. Two different signals are used. One generating a relatively quickly alternating signal, one giving slower alternations. After creation of a long sequence, five sequences are taken out of this by stepping $\frac{2^n-1}{3}$ indexes every time. This ensures not to have cross correlated signals Godfrey 1980. Again after doing this, the individual signals were scaled by doubling the sequences of heating in order to maintain reasonable indoor temperatures during the experiment. The resulting heating signals are seen in Figure E.7. The control signal is now fluctuating between 0 and 32. This is a binary representation of the signals in the different rooms. See Andersen et al. 2013e on experiment design for documentation of this.

Experiment 3 provides good data for fitting a model where the individual rooms are described by individual states. The heat exchange between the living room and the corridor may require special attention since they are not separated by walls or doors and because of the level difference between them.
Figure E.6 Mean temperature measurements, weather data, and heat input data for Experiment 3.
Figure E.7  The floor heating signals from Experiment 3.
E.3.4 Experiment 4 – repetition of Experiment 1

This is a repetition of Experiment 1 at the same time of year – one year later. An overview of the obtained data is shown in Figure E.8. The ventilation heating system seems to have been off most of the period. The experiment was start on February 22nd and ended on March 10th. However, from March 5th, a quite some disturbances are seen. Still it gives more than 11 days of data without disturbances.

Temperatures in the individual rooms and floor heating signals are plotted in Figure E.9.
Figure E.8  Overview of data from Experiment 4.
Figure E.9  The floor heating signals from Experiment 4.
E.3 Conducted experiments

E.3.5 Experiment 5 – A multi-room experiment with electric radiators

Experiment 5 includes the use of radiators. This provides a heat input that can be assumed to be fully known. Whereas the floor heating power is estimated from flow and temperature measurements in the floor heating system, and calculations using assumed properties of the liquid in the floor heating system, the power supplied to the radiators have been measured directly.

Radiators of about 700 W each were placed in the middle of the small room, the large room, the corridor, and in the living room. They all followed different PRBS signals, uncorrelated with the floor heating signals.

<table>
<thead>
<tr>
<th>Signal</th>
<th>n</th>
<th>$R_0$</th>
<th>$\lambda$</th>
<th>start</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>85</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The radiators followed PRBS signals generated as described in Table E.6. The signal was made into 4 uncorrelated signals as described in Section E.3.3.

The floor heating signals were scaled not to overheat the apartment. However, as seen from Figure E.10, the temperatures of the two rooms reached 40°C which is the limit of the measurement range of the thermal sensors. For this reason, these two radiators were turned off during the experiment (leading to disturbance of the experiment). However, the data can still be used if only the (output) points of temperatures of 40°C are removed.

The ventilation seems to have been turned off the whole period. Moreover, there are many interruptions of the experiment.
Figure E.10  Indoor air temperature, weather measurements, and total heating for Experiment 5. Ventilation heating was turned off by the inhabitants of the building. Moreover, the experiment has been interrupted several times.
### E.3 Conducted experiments

#### Figure E.11
The heating signals and temperature measurements for the individual areas of the apartment in Experiment 5.
E.4 Conclusions

A new data acquisition system was installed in the low-energy house in Sisimiut in April 2011. A brief summary has been given of the data collected since related to modeling of heat dynamics of the building. 5 experiments have been carried out where the heat input has been varied. The data obtained in these experiments have been reported in more detail. Three of these experiments are intended for modeling of the guest apartment as one large room, 2 of them for multi-room models. The last experiment was with electric radiators. This enables estimation of the power capacity of the floor heating.

Some sensors seem to periodically stall. This was especially seen in the overview of all the data collected. It is advised to survey the data acquisition during experiments in order to make sure that everything works as expected.
References


Parameter Identifiability in Grey-box Models of Heat Dynamics of Buildings

Authors:
Philip Delff Andersen¹, Rune Juhl¹, Uffe Høgsbro Thygesen², Henrik Madsen¹.

To be presented at:
IEA EBC Annex 58 meeting, Hong Kong, October 2013.

¹DTU Compute, Matematiktorvet, Building 322, DK-2800 Kgs. Lyngby, Denmark
²DTU AQUA, National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund Slot, Jægersborg Allé 1, 2920 Charlottenlund
Abstract

Successful modeling of heat dynamics of buildings requires application of both physical knowledge and statistical analysis. Physical knowledge is often used in model formulation and parameter interpretation. Reliable parameter estimation requires unique parameter estimates of the model given data. The ability to recover the underlying parameters is called identifiability. Identifiability of a model is crucial for the parameter estimates to be physically interpretable. This paper explains and illustrates issues of identifiability related to heat dynamics of buildings. This is closely related to experimental planning. It contains simulation examples of gray-box simple models illustrating pitfalls in model formulation and guidelines for how identifiability can be ensured. The paper is relevant both for experimental planning and model data analysis.

F.1 Introduction

When modelling a physical system, a statistician looks for a “simple” model that describes the system “well”. The simplicity of the model is a strength in itself. However, depending on the experiment, it may also be the only possibility if the model parameters are of interest and if the model estimate will be used on simulations or forecasts using new input data. This has to do with identifiability of the system.

Observability means that for a given set of parameters, $\theta$, and known inputs and outputs, it is possible to estimate the state of the system over time.

Identifiability is related to the inference on the parameters of the system. For a model to be identifiable, different values of the parameters must lead to different probability distributions of the output of the system. In other words, the joint likelihood function of all the parameters must have a unique global maximum given the output of the system. If this is theoretically the case for the system, the system is said to be structural identifiable. Whether the system is practically identifiable has to do with the dynamics of the inputs as well.

This note focuses on systems on linear time-invariant stochastic differential equations, i.e. models on the form

$$dX_t = (AX_t + BU_t)dt + \sigma_\omega d\omega_t$$
$$Y_k = CX_{tk} + DU_{tk} + e_k$$
$$e_k \sim N(0, \Sigma_y)$$  \hspace{1cm} (F.1)

where $X_t \in \mathbb{R}^n$ is the state matrix, $U_t \in \mathbb{R}^u$ is the known input vector, $Y_t \in \mathbb{R}^m$
is the output of the system. $\omega_t$ is a vector of Wiener processes Øksendal (2007), and $\{e_k\}$ is a discrete-time white noise process. $A, B, \sigma_\omega, C,$ and $D$ are matrices of appropriate dimensions. The identification problem concerns inference on parameters in all but the input, output, and noise processes.

Such a system is observable iff the matrix

$$\mathcal{O} = \begin{pmatrix} C \\ \vdots \\ CA^{n-1} \end{pmatrix}$$

has full rank, where $n$ is the number of states, i.e. the length of $X_t$.

Whether the system is structurally identifiable has to do with the relation between the transfer function from the input to the output and the parameters of the state-space model. In an identifiable state-space model, a given transfer function can only be generated by one combination of state-space parameters. The practical identifiability is rather related to identifiability of the transfer function.

A system is identifiable if different parameter combinations lead to different probability distributions for the output (Ljung 1987). This means that profile likelihood can be used for studies of systems that are only partially identifiable. Only if the maximum of a profile likelihood curve of a parameter is well-defined, this parameter can be identified. This method will be applied as well as analytic methods.

### F.1.1 The transfer function

During an experiment an input is applied to excite a system to generate an output (Madsen 2008; Ljung 1987). For any linear and time-invariant continuous-time system the output is the convolution of the input and an impulse response function, $h$:

$$y(t) = \int_{-\infty}^{\infty} h(u) x(t-u) du$$  \hspace{1cm} (F.3)

Introduce the Laplace transform (Lathi 1998)

$$\mathcal{L}\{x(t)\} = X(s) = \int_{-\infty}^{\infty} e^{-st} x(t) dt$$  \hspace{1cm} (F.4)
In the Laplace domain eq. (F.3) becomes a product of the transfer function and the Laplace transform of the input:

\[ Y(s) = H(s)X(s) \] (F.5)

\( H(s) \) is the transfer function of the system. For a linear and time-invariant system the transfer function is the compact description of the dynamical relationship between the input and the output. The state-space formulation in eq. (F.1) is a convenient and intuitive representation of a system, but it is important that all parameter combinations for a particular state-space model lead to different dynamics or transfer functions. If not, the parameters cannot be identified from the dynamics that can be measured.

### F.1.2 Obtaining the transfer function

The Laplace transform is a useful tool to find the transfer function of a dynamical system in continuous time. The Laplace transform maps the differential equation into an algebraic equation which then is easily manipulated algebraically to obtain the transfer function.

The bilateral Laplace is used for the analysis. It corresponds to the unilateral Laplace assuming \( x(0) = 0 \) which is like assuming the system is off at \( t = 0 \). Two transformations are required to know (Lathi 1998):

1. Linearity: \( L(a \cdot x(t) + b \cdot y(t)) = a \cdot X(s) + b \cdot Y(s) \), \( a, b \in \mathbb{R} \)
2. Differentiation: \( L\left(\frac{dx(t)}{dt}\right) = sX(s) \)

The general state space formulation given in eq. (F.1) is first written on the innovations form (Ljung 1987) given in eq. (F.6). The derivations of the Kalman filter and hence the components required for the innovations form can be found in (Jazwinski 1970). Note eq. (F.6) is all in continuous time including the observations. While it is possible to discretise the stochastic differential equation and arrive at a discrete time transfer function it will only complicate the notation.

\[
\frac{d\hat{X}(t)}{dt} = A\hat{X}(t) + BU(t) + K\epsilon(t) \tag{F.6}
\]

\[
Y(t) = C\hat{X}(t) + DU(t) + \epsilon(t) \tag{F.7}
\]
Start by applying the Laplace transform to the system eq. (F.6)

\[ sX(s) = AX(s) + BU(s) + K\epsilon(s) \]  

and isolate \( X(s) \)

\[(sI - A)X(s) = BU(s) + K\epsilon(s)\]

\[ X(s) = (sI - A)^{-1}BU(s) + (sI - A)^{-1}K\epsilon(s). \]  

(F.9)

Now apply the Laplace transform to the observation eq. (F.7)

\[ Y(s) = CX(s) + DU(s) + \epsilon(s) \]  

(F.10)

and insert eq. (F.9) into eq. (F.10)

\[ Y(s) = C(sI - A)^{-1}BU(s) + C(sI - A)^{-1}K\epsilon(s) + DU(s) + \epsilon(s) \]

\[ = (C(sI - A)^{-1}B + D)U(s) + (C(sI - A)^{-1}K + I)\epsilon(s). \]  

(F.11)

The output \( Y(s) \) is the sum of two transfer functions \( H_i(s) \) and \( H_n(s) \) acting on the input and noise respectively. In the remainder of this paper the focus will be on the input acting transfer function

\[ H_i(s) = C(sI - A)^{-1}B + D \]  

(F.12)

and the structural properties, i.e. does an inverse function mapping the transfer function back to the state space formulation exist.

F.1.3 Design of input signals

While structural identifiability has to do with the problem if the state space parameters can be derived from a given transfer function, practical identifiability rather addresses if the transfer function can be well estimated from data. Data must excite the system at enough frequencies to cover the desired transfer function. In this paper all data has been simulated using PRBS signals (Godfrey 1980) to ensure all frequencies are present in data thus avoiding the practical issues.

F.2 Example: An insulated wall

The first example is a simple RC-network describing the heat dynamics of a wall. The identifiability of the system will be studied using different assumptions about the model used and the measurements available.
An insulated wall is subject to an input temperature, $T_a$ and an output temperature, $T_i$. The output can be thought of as the temperature in a room surrounded by infinitely well insulated walls and the wall considered here. It can also be interpreted as the envelope of a building without heating. We now impose a lumped parameter model on the wall. In the model, it has one heat capacity, $C_w$ (with one temperature, $T_w$) surrounded by two resistances, $R_{aw}$ between the input and $T_w$, and $R_{wi}$ between $T_w$ and the output.

![RC-diagram of an example of an insulated wall and no heating indoors.](image)

This gives the following coupled first order stochastic differential equations:

\[
\begin{align*}
\dot{T}_w &= \frac{1}{C_w} \left( \frac{T_a - T_w}{R_{aw}} + \frac{T_i - T_w}{R_{wi}} \right) dt + d\omega_1(t) \\
\dot{T}_i &= \frac{1}{C_i} \left( \frac{T_w - T_i}{R_{wi}} \right) dt + d\omega_2(t)
\end{align*}
\] (F.13)

(F.14)

The state $T_i$ is measured with noise

\[Y_{tk} = T_{i,t_k} + e_{tk}, \quad e_{tk} \sim N(0, \sigma_y^2)\] (F.15)

The state vector is $T = [T_w \ T_i]^T$ and the input vector $U = T_a$. On matrix form the system is then given by

\[
A = \begin{bmatrix}
-\frac{1}{C_w R_{aw}} & \frac{1}{C_w R_{wi}} \\
\frac{1}{C_i R_{wi}} & -\frac{1}{C_i R_{wi}}
\end{bmatrix} \quad B = \begin{bmatrix}
\frac{1}{C_w R_{aw}} \\
0
\end{bmatrix} \\
C = \begin{bmatrix}
0 & 1
\end{bmatrix} \quad D = 0
\] (F.16)

The number of parameters is 4. The question is now how much of the system than can be estimated reliably. Applying eq. (F.12) leads to the transfer function:

\[
H_{input}(s) = \frac{\frac{1}{C_i C_w R_{aw} R_{wi}}}{s^2 + \frac{1}{C_i C_w R_{aw} R_{wi}} + \frac{1}{C_i C_w R_{aw} R_{wi}} \cdot s}
\] (F.18)
The transfer function eq. (F.18) has two poles, no zeros and a gain. Comparing eq. (F.18) to the standard zpk form results in 3 equations:

\[ K = \frac{1}{C_i C_w R_{aw} R_{wi}} \quad (F.19) \]
\[ a_1 = \frac{R_{aw} C_i + C_i R_{wi} + R_{aw} C_w}{C_i C_w R_{aw} R_{wi}} \quad (F.20) \]
\[ a_0 = \frac{1}{C_i C_w R_{aw} R_{wi}}. \quad (F.21) \]

Only eqs. (F.20) and (F.21) are independent thus up to two parameters can be identified. This will be investigated through simulations and analysis of likelihood functions in the following.

### F.2.1 Simulating data

The model was simulated using CTSMR using

\[ R_{wi} = R_{aw} = 0.1\text{K/W}, \quad C_i = 1\text{J/K}, \quad C_w = 10\text{J/K} \quad (F.22) \]

1000 realizations were randomly sampled. A confidence interval of the process is seen in fig. F.2.

### F.2.1.1 What if all parameters are estimated?

What if the modeller is unaware of structural problems of the model and attempts to estimate all four parameters? Solve eqs. (F.20) and (F.21) for all four parameters \( C_i, C_w, R_{aw}, R_{wi} \).

\[ C_w = -\frac{C_i}{C_i^2 R_{wi}^2 a_0 - a_1 C_i R_{wi} + 1} \quad (F.23) \]
\[ R_{aw} = -\frac{C_i^2 R_{wi}^2 a_0 - a_1 C_i R_{wi} + 1}{C_i^2 R_{wi} a_0} \quad (F.24) \]

\( C_i \) and \( R_{wi} \) can be chosen freely thus giving \( C_w \) and \( R_{aw} \) to obtain exactly the same transfer function. The denominator in eq. (F.23) is the same as the numerator in eq. (F.24). This must be positive if restricting the parameters to physical values. The valid region is governed by the second order polynomial equation. For the chosen parameter values the range is

\[ \frac{6 - \sqrt{26}}{10 R_{wi}} < C_i < \frac{6 + \sqrt{26}}{10 R_{wi}}. \]
Table F.1
<table>
<thead>
<tr>
<th>Time</th>
<th>Ta</th>
<th>Yi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 22</td>
<td>15.0</td>
<td></td>
</tr>
<tr>
<td>Feb 24</td>
<td>17.5</td>
<td></td>
</tr>
<tr>
<td>Feb 26</td>
<td>20.0</td>
<td></td>
</tr>
<tr>
<td>Feb 28</td>
<td>22.5</td>
<td></td>
</tr>
<tr>
<td>Mar 01</td>
<td>25.0</td>
<td></td>
</tr>
<tr>
<td>Mar 03</td>
<td>17.5</td>
<td></td>
</tr>
<tr>
<td>Mar 05</td>
<td>20.0</td>
<td></td>
</tr>
<tr>
<td>Mar 07</td>
<td>22.5</td>
<td></td>
</tr>
<tr>
<td>Mar 09</td>
<td>25.0</td>
<td></td>
</tr>
</tbody>
</table>

Figure F.2 The ambient temperature is simulated with a PRBS signal shown in the upper panel. The lower panel shows realizations of the indoor temperatures. The simulations are performed with very little noise.

Figure F.3 shows the problem of lack of identifiability using the likelihood function. The value of the likelihood function is the same in the whole region shown. The maximum likelihood estimate of $C_w$ is plotted in contours as function of chosen values of $R_{wi}$ and $C_i$. So, for any of the shown combinations of $R_{wi}$ and $C_i$, there is a value of $C_w$ leading to the same likelihood function value. It clearly shows the issue that based on data, there is a continuum of parameter estimates resulting in the exact same likelihood value, i.e. the parameters cannot be estimated. The area plotted is the area of positive values of the resistances and capacities.

eq. (F.23) restricted to physical values. The valid range is the same for $R_{aw}$ and the surface similar. This strange surface lies in the 4D parameter space resulting in equally good likelihood values everywhere. Therefore the results only depends on where the optimiser is initialized.

F.2.2 First solution: Two known parameters

Assume the resistances $R_{wi}$ and $R_{aw}$ are known leaving the two capacities to be estimated. Equations (F.20) and (F.21) are solved again for $C_w$ and $C_i$ using the true values $R_{wi} = R_{aw} = 0.1$ K/W and the transfer function parameters
Figure F.3 $C_w$ as function of $C_i$ and $R_{wi}$ disregarding unphysical values. The surface all collapses into the same transfer function.

$a_0 = 10$, $a_1 = 12$. Surprisingly, two solutions exists thus knowing the resistances does not make the model identifiable. The solutions are

\[(C_{i,1}^*, C_{w,1}^*) = (5 \text{ J/K}, 2 \text{ J/K})\]
\[(C_{i,2}^*, C_{w,2}^*) = (1 \text{ J/K}, 10 \text{ J/K})\]

Figure F.4 shows the negative log-likelihood surface (for one realisation) with two minima with the same likelihood value. The solution obtained by an optimization algorithm will depend on the initial guess.

It turns out knowing just the right combination of parameters can make the model identifiable. Knowing $C_i$ and $R_{wi}$ allows estimating $C_w$ and $R_{sw}$ and vice versa. An example of this is shown in Figure F.5. Only a capacity and a resistance is estimated, and the likelihood function has a well-defined maximum. All other combinations will have two solutions.

Rather than having to know half of the true parameters to ensure identifiability the model should be modified to fix the problem.
F.2.3 Other solutions - briefly

Measuring the inner surface heat flow of the wall will add a row to the $C$ matrix and thus one more transfer function. In total there are two poles and a gain which are all independent. With 4 parameters the model remains unidentifiable over the entire parameter set. One could reparametrise the resistances as $R_{wi} = R_{aw} = R/2$.

Adding outer surface heat flow of the wall turns out to be an identifiable model. In fact one can skip the temperature measurements and still identify all parameters.

Control the indoor temperature and let it be an input. Only model $T_w$ as a state and let heat flow measurements on both sides of the wall be the outputs. This model is identifiable, but more interesting: Skip the outer surface heat flow measurements and the model remains identifiable. Thus only one heat flow measurement is required here.
F.2.4 Summary

For an RC-model with one state and two distinct thermal resistances applied to heat dynamics of a wall, it was seen that if the temperature of the one side of the wall is modeled by a state driven by the heat flux through the wall, the heat flux must be measured on both sides of the wall to enable estimation of both thermal resistances and both thermal capacities. If only one heat flux is available, only three parameters can be estimated.

If the temperatures on both sides of the wall are inputs, only one measurement heat flux is needed for the system to be structurally identifiable.

F.3 Example: The envelope of a heated building

We now turn to the situation of a building being excited by internal heating and outdoor temperature which are both inputs. This part of the analysis closely relates to experiments for estimation of heat dynamics of buildings. In these cases temperature measurements are normally outputs. This will be the case...
throughout this section.

**F.3.1 One state**

The first system to be considered has one single state - the indoor temperature:

\[
dT_i = \frac{1}{C_i} \left( \frac{T_a - T_i}{R_{ai}} + c_h P_h \right) dt + \sigma_i d\omega_i(t) \tag{F.25a}
\]

\[
Y_k = T_i(t_k) + \epsilon_k \tag{F.25b}
\]

\[
\epsilon_k \sim N(0, \sigma^2_o) \tag{F.25c}
\]

So the state vector is \(T = T_i\), and the input vector is \(U = (T_a, P_h)^T\). Notice that the input is scaled by an unknown constant \(c_h\). The matrix formulation is given by

\[
A = \begin{bmatrix} -\frac{1}{C_i R_{ai}} \end{bmatrix} \quad B = \begin{bmatrix} \frac{1}{C_i R_{ai}} & \frac{c_h}{C_i} \end{bmatrix} \tag{F.26}
\]

\[
C = \begin{bmatrix} 1 \end{bmatrix} \quad D = 0 \tag{F.27}
\]

The number of parameters is 3. Apply eq. (F.12) to obtain the two transfer functions

\[
H(s) = \begin{bmatrix} \frac{1}{s + \frac{1}{C_i R_{ai}}} & \frac{c_h}{s + \frac{1}{C_i R_{ai}}} \end{bmatrix} \tag{F.28}
\]

Both transfer functions share a pole which is also the gain in the first transfer function. From these two transfer functions there are only 2 independent pieces of information both in the second transfer function,

\[
K = \frac{c_h}{C_i} \tag{F.29}
\]

\[
a_0 = \frac{1}{C_i R_{ai}} \tag{F.30}
\]

Equations (F.29) and (F.30) cannot be solved uniquely. Thus it is required to know one of the parameters. \(c_h\) can be interpreted as the efficiency of the heater. For the remainder of the paper it is considered known.

**F.3.2 Two states**

The first model is extended to model the temperature inside the envelope of the building. Heat will be applied either in the envelope or the indoor temperature.
F.3 Example: The envelope of a heated building

F.3.2.1 Heating the envelope

The second unobserved state, the envelope $T_e$, is introduced.

$$dT_i = \frac{1}{C_i} \frac{T_e - T_i}{R_{ei}} dt + \sigma_i d\omega_i(t) \quad (F.31a)$$

$$dT_e = \frac{1}{C_e} \left( \frac{T_a - T_e}{R_{ae}} + P_h + \frac{T_i - T_e}{R_{ei}} \right) dt + \sigma_e d\omega_e(t) \quad (F.31b)$$

$$Y_k = T_i(t_k) + \epsilon_k \quad (F.31c)$$

$$\epsilon_k \sim N(0, \sigma_o^2) \quad (F.31d)$$

The system is first given on matrix form by

$$A = \begin{bmatrix} -\frac{1}{C_i R_{ei}} & -\frac{1}{C_e R_{ae}} \\ -\frac{1}{C_e R_{ae}} & -\frac{1}{C_e R_{ei}} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad D = 0$$

Again, apply eq. (F.12) to obtain the two transfer functions

$$H(s) = \frac{\frac{1}{C_e C_i R_{ae} R_{ei}}}{s^2 + \frac{C_e R_{ae} + C_i R_{ae} + C_i R_{ei}}{C_e C_i R_{ae} R_{ei}} s + \frac{1}{C_e C_i R_{ae} R_{ei}}}$$

The denominators in appendix F.3.2.1 are the same. The gain in the first transfer function is not independent information. Thus only the second transfer function contains independent expressions,

$$K = \frac{1}{C_e C_i R_{ei}} \quad (F.32)$$

$$a_1 = C_e R_{ae} + C_i R_{ae} + C_i R_{ei} \quad (F.33)$$

$$a_0 = \frac{1}{C_e C_i R_{ae} R_{ei}}. \quad (F.34)$$

Only 3 equations, but 4 parameters leaving the model unidentifiable as is.

One possible solution would be to know one of the parameters like in the wall example. But again one must be careful as eqs. (F.32), (F.34) and (F.34) has two solutions if either $C_i$ or $R_{ei}$ are known. If $R_{ae}$ is known the system remains unidentifiable. In fact both capacitances can be written as functions of $R_{ei}$, thus knowing $R_{ae}$ does not help at all. Finally knowing $C_e$ means the system is uniquely solvable.
F.3.2.2 Heating indoors

Now the heating $P_h$ is moved from the envelope to the inside of the building. The system equations are modified as

$$\frac{d}{dt} T_i = \frac{1}{C_i} \left( \frac{T_e - T_i}{R_{ei}} + P_h \right) dt + \sigma_i d\omega_1(t) \quad (F.35a)$$

$$\frac{d}{dt} T_e = \frac{1}{C_e} \left( \frac{T_a - T_e}{R_{ae}} + \frac{T_i - T_e}{R_{ei}} \right) dt + \sigma_e d\omega_e(t). \quad (F.35b)$$

Moving the input does not change the dynamical part specified by the $A$ matrix. Only the $B$ matrix changes to

$$B = \begin{bmatrix} 0 & 1 \\ \frac{1}{C_e} & 0 \end{bmatrix}. \quad (F.36)$$

After applying eq. (F.12) the two transfer functions are

$$H_1(s) = \frac{1}{s^2 + \frac{C_e}{C_i} + \frac{R_{ae} + C_i R_{ae} + C_i R_{ei}}{C_e C_i R_{ae} R_{ei}} s + \frac{1}{C_e C_i R_{ae} R_{ei}}} \quad (F.37)$$

$$H_2(s) = \frac{1}{s^2 + \frac{C_e}{C_i} + \frac{R_{ae} + R_{ei}}{C_e C_i R_{ae} R_{ei}} s + \frac{1}{C_e C_i R_{ae} R_{ei}}} \quad (F.38)$$

Disregarding the repeated terms it is clear only the second transfer function is required. The independent set of equations are

$$b_0 = \frac{R_{ae} + R_{ei}}{C_e C_i R_{ae} R_{ei}} \quad (F.39a)$$

$$b_1 = \frac{1}{C_i} \quad (F.39b)$$

$$a_0 = \frac{1}{C_e C_i R_{ae} R_{ei}} \quad (F.39c)$$

$$a_1 = \frac{C_e R_{ae} + C_i R_{ae} + C_i R_{ei}}{C_e C_i R_{ae} R_{ei}} \quad (F.39d)$$

The 4 parameters in the 4 equations in eq. (F.39) are uniquely solvable, thus the model is identifiable.
### F.3.3 Three states - modelling floor heating

The system is now extended with a third state representing the floor heating system. It is connected to the indoor temperature. The new system is

\[
\begin{align*}
    dT_i &= \frac{1}{C_i} \left( \frac{T_f - T_i}{R_{fi}} + \frac{T_e - T_i}{R_{ei}} \right) \, dt + \sigma_i d\omega_i(t) \quad (F.40a) \\
    dT_e &= \frac{1}{C_e} \left( \frac{T_a - T_e}{R_{ae}} + \frac{T_i - T_e}{R_{ei}} \right) \, dt + \sigma_e d\omega_e(t) \quad (F.40b) \\
    dT_f &= \frac{1}{C_f} \left( P_h + \frac{T_i - T_f}{R_{fi}} \right) \, dt + \sigma_f d\omega_f(t). \quad (F.40c)
\end{align*}
\]

Again, eq. (F.40) is written on matrix form as

\[
A = \begin{bmatrix}
    -\frac{1}{C_i R_{ei}} & -\frac{1}{C_i R_{ei}} & \frac{1}{C_i R_{ei}} \\
    \frac{1}{C_e R_{ae}} & \frac{1}{C_e R_{ae}} & 0 \\
    \frac{1}{C_f R_{ei}} & 0 & \frac{1}{C_f R_{ei}}
\end{bmatrix}, \quad B = \begin{bmatrix}
    0 \\
    0 \\
    \frac{1}{C_e R_{ae}}
\end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \quad D = 0.
\]

The model has 6 parameters. The 2 inputs results in 2 transfer functions. The transfer functions for eq. (F.40) are not small and omitted here. The principle remains the same. Apply eq. (F.12) to obtain the transfer functions and identify the independent terms. Combining terms from both transfer functions gives 6 equations

\[
\begin{align*}
b_{1,1} &= \frac{1}{C_e C_i R_{ae} R_{ei}} \\
b_{2,0} &= \frac{1}{C_e C_i C_i R_{ae} R_{ei} R_{fi}} \\
b_{2,1} &= \frac{1}{C_f C_i R_{fi}} \\
a_{2,0} &= \frac{1}{C_e C_i C_i R_{ae} R_{ei} R_{fi}} \\
a_{2,1} &= \frac{1}{C_e C_i C_i R_{ae} R_{ei} R_{fi}} \\
a_{2,2} &= \frac{1}{C_e C_i C_i R_{ae} R_{ei} R_{fi}}
\end{align*}
\]

This seemingly complicated set of equations is in fact solvable as is. The model is therefore identifiable.
F.4 Conclusions

The identifiability of the physical parameters in dynamical systems on state space form has been analysed through examples from heat dynamics of buildings. Two examples were analyzed with different numbers of states and different measurement equations. The method used was to calculate the transfer functions from inputs to outputs and see whether the physical parameter could be deduced given that the transfer functions were known. Also, studies of likelihood functions calculated from simulations have been studied, illustrating the problem of parameter estimation in unidentifiable systems. Especially for more complex systems where transfer function analysis is tedious, profile likelihood is a strong method for analysis of identifiability issues.

The first example dealt with a wall between indoors and outdoors inspired by a test setup for experimental analysis of the thermal properties of an opaque wall. The wall temperature was modeled as a state of the system. It was found that if the temperature on the inside of the wall is modeled as a state, a heat flow measurement on the outer surface of the wall was crucial to estimate the heat capacity of both indoors and the wall and the thermal capacities between indoors and the wall, and between the wall and outdoors. If only a heat flow on the inside was measured, only three parameters could be identified, and if only the indoor temperature was measured, only two parameters could be identified. If the indoor temperature was a known input, only a measurement of the heat flow on one side was needed for the three parameters of the wall to be identifiable.

The second example was about a heated building and its heating system with the outdoor temperature as a known input. The indoor temperature was a state in all the considered setups. With only a measurement of this state, it was seen that the system was unidentifiable if the heating input affected the envelope directly. The thermal energy entering the system from the heating must be known in order to identify both thermal resistances and capacities. Accordingly, the system is not identifiable if the heating affects the envelope which exchanges thermal energy with both outdoors and indoors. When the envelope was modelled as a state, identifiability was obtained when the indoor state was heated directly by known thermal power only by reducing the description of the envelope to consist of one capacity and one resistance.
References


