Stochastic Modelling of Building Integrated Photovoltaic Modules

Nynne Friling

Kongens Lyngby 2006

Supervisor: Prof. Henrik Madsen - Technical University of Denmark, IMM
External Supervisor: Ing. J. J. Bloem - Joint Research Center, Italy
Dr. M. J. Jiménez - CIEMAT, Spain

IMM-M.Sc.-2006-80

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Technical University of Denmark DTU

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

Summary

The present thesis deals with mathematical modelling of the heat transfer in buildings and building integrated photovoltaic modules. One of the purposes of this thesis is to extend the knowledge about the performance of building integrated photovoltaic modules.

Due to the limited oil resources it becomes more and more important to focus on renewable energy. The main purpose of the modules is to produce electricity, and as a spin-off the air behind the modules can be heated and used for heating the building. This is two ways in which to produce renewable energy.

The point of turn of the mathematical methods applied is stochastic differential equation, state space models, maximum likelihood estimation, and the extended Kalman filter. Subsequent the model estimation various analyses of the residuals and tests of the models have been carried out in order to uncover the reliability and usefulness of the model. The estimated models are both linear and non-linear models. The non-linear effect to be investigated is the infrared radiation and the wind speed.

The thesis emanates from the article 'Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules'[Jiménez et al. 2006]. The basics of the applied models of the photovoltaic module ensue from the article. Introductory models of how to describe the heat transfer in buildings are outlined and developed. Technics and methods such as RC-modelling and lumping are touched on. After the description of the applied theory the rest of the thesis is related to the analysis of the photovoltaic module. The module temperature is highly negative correlated with the efficiency of the module. This makes it important to be able to predict the

temperature of the module in order to determine the production of electricity and to investigate how the temperature of the module can be decreased. Introductorily an in-depth analysis of how to describe the temperature of the module is carried out. This is done both in the sense of identifying the best describing variables, and also in order to discover the best fitting models. Both single and multiple state space models are estimated. This makes it possible to examine if it is necessary to split up the module temperature in two states.

The results of the thesis

The best description of the data is obtained by using an extended single state model containing non-linear influences. The best results are found by applying the temperature at the top of the module as the output variable. This finding is supported by thermal images taken of the module where the measured top temperature covers most of the module temperature. The analyses show improved performance for applied variables calculated from the electrical flow of the module compared to the similar measured variable. One of the main aims has been to identify the influence of the ambient wind speed. It was beforehand expected that a filtered version of the wind speed was influencing the heat transfer between the air and the module. The analyses have revealed that a filtered wind speed having nearly the same fluctuations as the measured wind affects the module.

After identifying the preferable model, data where variations of the set-up have been made, are analyzed. The result of the testing reveals that the model is able to discriminate the variations from each other. It is identified that higher forced velocity behind the module and obstacles turning the laminar air flow into a more turbulent flow raise the heat transfer coefficient between the ambient air and the module. This leads to the desired decrease of the module temperature.

Resumé

Denne eksamensopgave omhandler matematisk modellering af varmeoverførelse i bygninger og bygningsintegrerede solcellemoduler. Et af formålene med denne opgave er, at opnå en øget viden om, hvordan bygningsintegrerede solcellemoduler fungerer, og hvad der påvirker dem.

Det bliver mere og mere vigtigt at fokusere på vedvarende energi grundet, de begrænsede olieressourcer. I denne opgave er et bygningsintegreret solcellemodul med en tvunget luftstrøm bag modulet analyseret. Hovedfunktionen for solcellemodulet er at producere strøm. Som en sideeffekt kan den tvungne luftstrøm bag modulet anvendes til opvarmning af bygningen. På denne måde fremstilles der to forskellige former for vedvarende energi.

De matematiske metoder, som anvendes i denne opgave, er stokastiske differentialligninger, state-space modeller, maximum likelihood estimation og det udvidede Kalman filter. Efter at have estimeret modellerne gennemføres forskellige residualanalyser og test af modellerne. Dette gøres for at afdække, om modellerne er pålidelige og brugbare. De estimerede modeller dækker både lineære og ikke-lineære modeller. Vind og infrarød stråling undersøges som ikke-lineære indflydelser.

Denne eksamensopgave udspringer af artiklen: 'Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules'[Jiménez et al. 2006]. De anvendte modeller for solcellemodulet stammer fra denne artikel. Indledende er modeller til at undersøge, hvordan varmeoverførelse kan beskrives skitseret. Teknikker og metoder, såsom RC-modeller og lumpning, er berørt. Efter beskrivelse af den anvendte teori omhandler resten af opgaven modellering af solcellemoduler. Modultempera-

turen er negativt korreleret med effektiviteten af modulet. Dette betyder, at det er vigtig at kunne prædiktere modultemperaturen for at kunne bestemme produktion af strøm. Ydermere giver dette mulighed for at undersøge, hvordan modultemperaturen kan reduceres. Indledningsvis er en dybdegående analyse til at kortlægge modultemperaturen gennemført. Dette gøres både for at identificere de bedst beskrivende variabler, men også for at finde den bedste model. State-space modeller med både en og flere tilstande er identificeret. Dette gør det mulig at undersøge nødvendigheden af at opdele beskrivelsen af modultemperaturen i flere områder.

Opgavens resultater
Den bedste beskrivelse af data fås ved at bruge den mest avancerede ikke-lineære state-space model med en tilstand. De bedste resultater opnås ved at anvende den temperatur, som er målt i toppen af modulet. Dette resultat understøttes af varmebilleder af modulet, hvor den målte toptemperatur repræsenterer størstedelen af temperaturen i modulet. Analyser har også vist en øget grad af beskrivelse, når variabler, som er estimeret ud fra den elektriske strøm, anvendes frem for de tilsvarende målte variabler. Et af hovedformålene har været at identificere udendørsvindens indflydelse på modulet. På forhånd var det forventet, at en form for filtreret vindhastighed ville have indflydelse på modulet. Analysen har afsløret, at den filtrerede vindhastighed, som påvirker modulet, kun afviger meget lidt i variation fra den målte vindstyrke.

Efter at have identificeret den foretrukne model, er data, hvor der er varieret på test-opsætningen, anvendt og analyseret. Resultaterne af dette forsøg viser, at det er muligt for modellen at skelne mellem variationerne i opsætningen. Det er i denne forbindelse fundet, at en øget ventilationshastighed bag modulet og modstande, som ændrer strømningen fra laminær til turbulent i luftrummet bag modulet, øger varmeoverførelseskoefficienten mellem udendørsluften og modulet. Dette fører til den ønskede reduktion af modultemperaturen, som betyder øget strømproduktion.

Preface

This master's thesis was prepared at Informatics Mathematical Modelling (IMM), Technical University of Denmark (DTU) in fulfilment of the requirements for acquiring the Master degree in Engineering. The thesis represents 40 ECTS points out of the required 300 ECTS point. The project has run from September 2005 to the end of August 2006.

The thesis deals with stochastic modelling of a photovoltaic module. The purpose is to identify the physical factors influencing the temperature and efficiency of the module. One of the cornerstones of the report is to pinpoint the non-linear influence from the outdoor wind speed. The basis of the thesis is the article 'Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules'.

The thesis is supervised by Professor Henrik Madsen from IMM. External supervisors are Ing. J. J. Bloem, European Commision Joint Research Center, and Dr. M. J. Jiménez, Research Centre for Energy, Environment and Technology in Spain.

Lyngby, 30^{th} of August 2006

Nynne Friling

Introduction

The background of this thesis is the article 'Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules' by M. Jiminez, H. Madsen and H. Bloem. The models treated in the article are reused and developed futher. Also some of the applied data overlap. In the present thesis more data and combinations of data are investigated.

In order to make this thesis get off properly a few introductory items have to be clarified.

- All the models have been estimated by means of the software CTSM. [Kristensen & Madsen 2003*b*] gives a short introduction to the modelling in CTSM. In this guide also images are included. This may help understand some of the statements and explanations in the thesis.
- The x-axis of the plots over time has to be expound. The value is stated in minutes. Value zero is in time accordance with midnight. Observation 720 corresponds to 12 a.m. and so forth.

Acknowledgements

I wish to thank my Professor Henrik Madsen who during the whole period has helped and guided me through the preparation of this thesis. I would also like to thank Henrik Madsen for giving me the oppertunity of paticipating in the Dynastee Conference in Athens in October 2006 and the stay at the Joint Research Center in Ispra, Italy.

I would like to thank Ing. J. J. Bloem from the European Commision Joint Research Center in Ispra for the rewarding week at the Center. I appreciated the visit at the test site since it gave me some hands-on experience. Furthermore I wish to thank PhD. Marìa José Jiménez, CIEMAT, Research Centre for Energy, Environment and Technology, in Spain, for background information on the models. Thank you both very much for inputs and corrections during the process of writing this thesis.

During the project I have also had meetings with researchers from Risø National Laboratory, Danish Technological Institute and Department of Civil Engineering at Technical University of Denmark, who have taken the time to help me in the process. The researchers have helpfully provided me with new knowledge and articles.

I would also like to thank my fellow students, Anna, Stig, and Søren, for their help, support, and ideas during this work.

Finally I would like to thank Annalise Dühring Wiberg for her helpful inputs on the linguistics of the thesis.

Contents

Chapter 1

Renewable Energy

1.1 Renewable energy methods for dwellings and mathematical modelling

Today around one-third to half of the energy consumption in the industrialised countries is used for lighting and making the thermal condition of dwellings comfortable [Prasad & Snow 2005]. Furthermore it is estimated that half of the world's oil resources have already been used. If no better insulation and new-thinking energy production methods are taken into consideration in a wider sense, the remaining oil resources are estimated to cover the energy needs for 30 years. For the moment there is no direct substitute for oil. Beside the limitation of resources the renewable energy production can also reduce the level of pollution and global warming which in certain areas is a large scale problem. The renewable energy sources are not depending on oil ressources, unless in the production phase, and are for this reason less contaminating.

Favourable regulations and taxations are a way to encourage the application of renewable energy methods. These actions can take place on both international and national scale. For the moment only few actions have been taken.

Today the most common renewable energy production methods are wind en-

ergy, solar heating, biofuel, and photovoltaic technology, which is the topic of this thesis. The first step towards a larger proportion of renewable energy is an efficient further development of equipment and methods. This needs to be followed up by continuously improvements of the equipments, the materials, and the methods. At present, new thinking photovoltaic cells are under development. It is expensive, both in terms of energy and economics, to produce the cells made of silicon which are the preferred ones at the moment. Therefore new materials such as organic material are being researched as a substitute material.

Another important issue in connection with renewable energy is to obtain acceptance from the consumers. It is of great importance to provide reliable information on the energy-performance as well as the overall preformance of alternative energy sources to get a higher acceptance from the consumer of renewable energies. This is for instance the case when a consumer has to decide if a building integration of photovoltaic module will be beneficial. Can mathematics be of any benefit in relation to the issue? To obtain a generally better understanding of the renewable energy methods, mathematical modelling can be applied. Mathematical modelling can as an example, help to determine which materials are most efficient. The optimal operational level may also be determined. Furthermore, the mathematical modelling may give an understanding of which external factors, such as wind speed, are influencing the equipment. Sometimes it is even possible to reveal former unknown influences.

Chapter 2

The Mathematics Behind Heat Dynamics of Buildings

2.1 Introduction

This section is an introduction to the mathematics and the physical laws behind modelling of heat dynamics of buildings. Furthermore, the advantages of using statistical and physical knowledge simultaneously, called grey-box modelling, will be discussed. In the models used in this chapter it is assumed that the heat transfer relationships are linear.
Since the mathematical description of heat dynamics of buildings can be very complicated, methods to overcome this will be discussed. Lumping is used as a way to decrease the complexity of the models. The models found are so-called RC-models. In extension of the lumping method RC-models of four different types of houses are outlined. Some of the descriptions and models are also presented in [Madsen 1985] and [Madsen et al. 1994].

2.2 Modelling method

When searching for a representative model for a physical phenomenon it is obvious to use the physical laws. This approach to the problem is called white-box modelling or the deductive approach. At the other end of the scale is the black-box modelling where only statistical models are used. This way of analyzing data is also called an inductive approach. This approach requires no preceding knowledge about the data and its nature. This is for instance the case when neural networks are used for modelling data. In between these two methods is the grey box modelling. This approach uses the knowledge from both physics and statistics, resulting in better models. Grey-box modelling in relation to stochastic modelling is treated in the articles [Kristensen et al. 2003*a*] and [Kristensen et al. 2003*b*]. By using grey-box modelling it is possible to test if each of the parameters in the models applied is significant, and thereby it is often possible to reduce the number of parameters in the final model. This also implies that it becomes easier to interpret the model. Another reason why the grey-box models are preferable is that the final models may sometimes show physical relationships which were not known in advance, [Andersen 2001] and [Madsen 2001].
Previous research has shown that the deductive approach is only able to describe the long-time variations of the heat transfer, which implies that it is not possible to get the short-time variations and fluctuations modelled [Madsen 1985]. Short-time variation is when the heat transfer changes due to fast changes in the weather conditions, for instance due to a cloud passing by the sun. By using the inductive approach with two or more time constants it is possible to model both the short- and the long-time dynamics, which raises the level of description the models provides.

2.3 Lumped parameters

When dealing with heat dynamics both the time and the spatial coordinates are seen as stochastically independent variables. This implies that it is necessary to introduce partial differential equations to describe the correlations. Since partial differential equations can often be very complicated to solve, the lumped parameterization method will be used to simplify the mathematics. The method turns the partial differential equation into ordinary differential equations, which are much easier to handle and solve. Lumped processes are defined as: Lumped processes are all processes in which the spatial dependence of the variables under consideration can be neglected [Kecman 1988]. The use of lumped parameters is an approximation to the real problem, because the heat properties are still

changing in the space dimensions. It is therefore always important to investigate if the method is reasonable to use in the specific case. It is well documented that the use of lumping is an acceptable way to model the dynamics of buildings if the building is not influenced by non-linear phenomena [Madsen et al. 1994]. The use of lumped parameters is most frequently a good way to analyze cases, which otherwise would become very complicated. Figure 2.1 shows an example of how lumped parameterization can be used when modelling a simple wall. This kind of modelling is called RC-modelling, because the system is built of resistances and capacitances similar to analogous electrical systems. From Figure 2.1 it can be seen that one of the approximations is that the heat transfer is only distributed horizontally. Therefore, this approach is only valid in this case if it is known that the temperature of the walls are homogeneously distributed. If temperatures are measured vertically at several places in the wall, the lumping can be worked out in layers. This approach can be implemented by adding more state equations to the existing system.

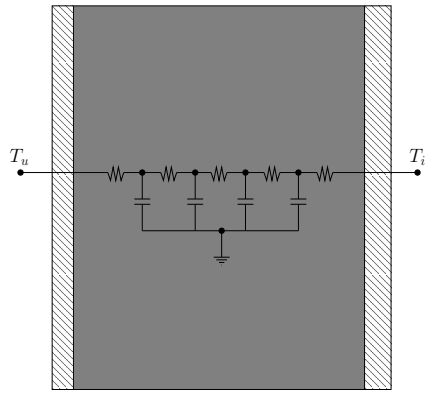

Figure 2.1: Illustration of a lumped system in a wall [Madsen 1985]

2.4 Different kinds of heat transfers

There are three ways in which heat can be transferred:

- Conduction
- Convection
- Radiation

One definition of heat transfer is that it is normally transferred from a higher temperature object to an object with a lower temperature [hyperphysics 2006]. One of the physical reasons why it is difficult to make an analytical description of the heat transfer is that the three different transmission ways are contributing to the process simultaneously.
The Equations 2.1-2.8, are found in [Madsen 1985].

2.4.1 Conduction

Conduction means energy transmitted through random splice between atomic particles. In relation to heat dynamics of buildings this kind of energy is observed as heat transfer through the walls, floors, and roof.
The mathematical formulation of conduction is built on an empirical law, [Both & Christiansen 2002],

$$\frac{dQ}{dt} = -\lambda A \frac{dT}{dx} \tag{2.1}$$

where $\frac{dQ}{dt}$ is the heat per time unit. λ is the thermal conductivity constant which varies quite a lot depending on which materials are used for the construction of the wall. Physical factors, such as temperature, have also impact on the thermal conductivity, but since the there are no large variations in the physical factors when dealing with heating of buildings, only the influence of the materials used is considered. A is the cross section of the conductive solid, in this case the wall, perpendicular to the direction of the heat current. $\frac{dT}{dx}$ is the temperature gradient.

If the part of the building considered does not contain neither heat sources nor sinks, it is possible to set up an ordinary diffusion equation which describes the time rate of the temperature

$$\frac{\partial T}{\partial t} = \frac{-\lambda A}{c\rho}\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) \tag{2.2}$$

where c is the heat capacity and ρ is the density of the material. x, y, and z are the spatial coordinates. In the cases where walls consist of only one material, it is possible to make an approximation to the problem by reducing Equation 2.2 to be only one-dimensional:

$$\frac{\partial T}{\partial t} = \frac{-\lambda A}{c\rho}\frac{\partial^2 T}{\partial x^2} \tag{2.3}$$

Equation 2.3 becomes even simpler if the temperature gradient is seen as constant which is the case if the walls are homogeneous. The thickness of the walls

is stated as d.

$$\frac{dT}{dt} = \lambda A \frac{T_2 - T_1}{d} \tag{2.4}$$

T_1 and T_2 are the temperatures at the boundaries of the layer. To make Equation 2.4 valid for the entire thickness of the wall Equation 2.5 can be used. T_o and T_u denotes the outside and the inside surface temperature of the wall respectively

$$\frac{dT}{dt} = UA(T_o - T_u) \tag{2.5}$$

where U, the transmission coefficient, is given as

$$\frac{1}{U} = \sum \frac{d_i}{\lambda_i} \tag{2.6}$$

i denotes the number of homogeneous layers. This makes it possible to make the earlier described lumped description. As stated throughout the section, the wall must consist of homogeneous parallel layers if the one-dimensional description is to be sufficient. In cases where an approximation for an inhomogeneous wall is satisfying the one-dimension method above can be used. Furthermore, it is also possible to apply the latter procedure when making approximate calculations of ordinary constructed walls when conditions are non-stationary.

2.4.2 Convection

The second contributor to heat transfer of buildings is convection. This is the kind of heat transfer where a fluid is in motion as a result of difference in temperature. The standard equation for convection is seen below

$$\frac{dQ}{dt} = hA\Delta T \tag{2.7}$$

where h is the convective heat transfer coefficient which is depending on factors such as compressibility, viscosity, temperature, the velocity and the profile of the flow in the medium, and the distance between the layers [Both & Christiansen 2002]. Furthermore, this implies that the convection coefficient varies in value depending on whether the wall is horizontal or vertical. A denotes the area which the heat is passing through. ΔT is the temperature difference between the wall and the main body of the fluid.

2.4.3 Radiation

The last of the three ways in which heat transfer can take place is by radiation, which is mainly the heat transfer from the sun. Radiation is seen when two

bodies with different temperatures are in optical contact with each other. This is typically the situation when the heat transfer is transmitted through windows. The heat is exchanged from for instance the sun to the flooring on the other side of the window. In cases where one of the bodies is surrounded by the other Equation, 2.8 can by applied to solve the problem.

$$\frac{dQ}{dt} = A\epsilon\sigma(T_s^4 - T^4) \tag{2.8}$$

A is the surface area of the surrounded body, ϵ is the emissivity, and σ is the Stefan-Boltzmann constant, $\sigma = 5.67 \cdot 10^{-8} Wm^{-2}K^{-4}$. As in the case of convection radiation is often difficult to calculate in practice. The heat transfer by radiation inside buildings is rather limited, but it is necessary to include the radiation in the model for rooms with windows. It can be seen from Equation 2.8 that the calculations for the radiation are heavily dependent on the temperature differences.

2.5 RC-models

Taking the previous sections into consideration it is obvious that the heat dynamics description of entire buildings can be very complicated. A good approximation of modelling the heat dynamics of buildings is to use lumped RC-models. The choice of building materials has a great influence on the heat balance equations and thereby the costruction of the models. Four models with different combinations of building materials can be found in the sections 2.5.1-2.5.4. The models are simplified since there is only one room, one window and no variations in type of walls and floor in the individual models. In the sections only difference between light constructions and heavy constructions is made. This is due to the fact that the heat transfer is transmitted faster through a light construction than through a heavy construction. Previous research [Hansen 1985] has shown that it is necessary to include one or more time constants to describe for instance a solid wall. In extension to this it is obvious that it is necessary to use different RC-models to describe various kinds of buildings ranging from greenhouses to solid old churches. The first two models contain two time constants whilst the two latter examples contain three time constants. It is obvious that it is possible to extend the models made and create new models. Comments on extentions and improvements are suggested in a section after the describtion of the four models.

The most important simplifications for all four models are:

1. The heat capacity is concentrated in a thin layer inside the wall

2. All surfaces, apart from the window, are considered to have the same temperature
3. Radiation transfer as a mechanism for heat transfer between T_i, T_o, and T_m is not considered.
4. The heat capacity of the room air is neglected.

The temperatures mentioned in bullet number three above are: T_i is the room temperature, T_o is the temperature of the wall surfaces, and T_m is the temperature of the heat accumulating layer. To create the model the heat balances for the room air, the surface, and the heat accumulating layers need to be identified. Therefore it is necessary to establish individual heat balance equations concerning each of the four models. All the general priciples are explained in relation to the first model only.

2.5.1 Light walls and solid floor

The characteristics of this first model are that the dominating heat capacity is located in the floor and in the ground under the floor while the walls are thermally light. A typical building with these characteristics could be a building constructed of wooden walls with a concrete floor. In Figure 2.2 both a model and the analogous electrical system for the building are outlined. T_u is a pure temperature source, r symbolizes the resistance which is responsible for the exchange of heat. C is the capacitance. The ϕ-values are called pure heat sources [Davies 2004].

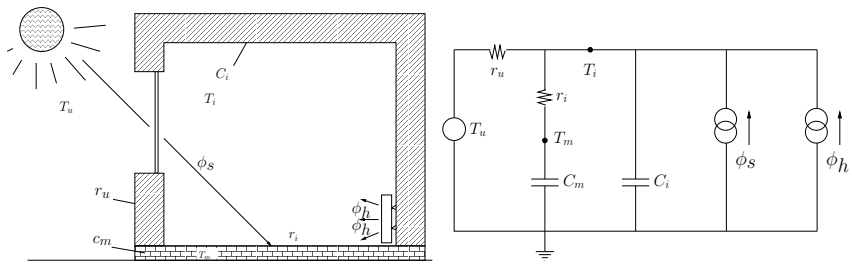

Figure 2.2: Illustration of a building with light walls and the dominating heat capacity in the floor

In the following the heat balance equations for the building described above will be outlined. From the analogous electrical system shown in Figure 2.2 it is possible to see how some of the heat balance equations are found. The three heat balances are:

- The room temperature

$$\frac{dQ_k}{dt} + \sum hA_h(T_o - T_i) + Gc_a(T_l - T_i) = 0 \tag{2.9}$$

- The surface

$$\frac{dQ_r}{dt} = \sum k_1 A_1(T_o - T_u) + \sum k_2 A_2(T_o - T_m) + \sum hA_h(T_o - T_i) \tag{2.10}$$

- The heat accumulating layer

$$c_m \frac{dT_m}{dt} = \sum k_2 A_2(T_o - T_m) \tag{2.11}$$

The explanations of the notation are: T_u is the temperature of the outside surface, T_l is the temperature of the ventilation air, $\frac{Q_k}{dt}$ is the convection part from persons, light, radiators, etc. $\frac{Q_r}{dt}$ is the heat transferred directly to the surface as radiation from the sun, light, radiators etc. G denotes the quantity of the ventilation air and c_a is the specific heat of the air. T_o, T_i and T_m denote respectively the temperature of the surface, the room, and the heat accumulating layers. A denotes areas. By insulating T_m, T_0, and T_i in the three heat balance equations, Equations 2.9-2.11, it is possible to formulate a deterministic linear state space model in continuous time with two steady state equations like the one shown in Equation 2.12. It can be seen that only the first equation has a dynamic impact.

$$\left.\begin{array}{lcl} \frac{dT_m}{dt} & = & AT_m + \mathbf{B}U \\ T_0 & = & C_1 T_m + D_1 U \\ T_i & = & C_2 T_m + D_2 U \end{array}\right\} \tag{2.12}$$

$$\mathbf{U} = (T_u \;\; T_l \;\; \frac{Q_k}{dt} \;\; \frac{Q_r}{dt})^T \tag{2.13}$$

T_m is the state vector and U the input vector. U contains all the external factors which have influence on the system, see Equation 2.13. A, C_1, and C_2 are constants while B, D_1 and D_2 are vectors consisting of constants. The specific expression of the constants can be found by isolating T_m, T_0, and T_i in the heat balance equations.

The model formulated above is not sufficient for describing the heat dynamics of a building, because it is only able to describe slow changes equivalent to long-time dynamics [Hansen 1985]. It is furthermore known that the heat transfer

to a building can be influenced by short time dynamics such as a cloud passing by the sun. Adding one or two time constant, a model where both long- and short-time dynamics can be modelled is identified [Hansen 1985]. Therefore an extra time constant is added in the model below. When having more than one time constant it can be of advantage to express Equation 2.12 by the use of matrices, which is done in Equation 2.14

$$\begin{bmatrix} c_m \frac{dT_m}{dt} \\ c_i \frac{dT_i}{dt} \end{bmatrix} = \begin{bmatrix} \frac{-1}{r_i} & \frac{1}{r_i} \\ \frac{1}{r_i} & -(\frac{1}{r_u} + \frac{1}{r_i}) \end{bmatrix} \begin{bmatrix} T_m \\ T_i \end{bmatrix} + \begin{bmatrix} 0 & 0 & A_w P \\ \frac{1}{r_u} & 1 & A_w(1-p) \end{bmatrix} \begin{bmatrix} T_u \\ \frac{dQ_t}{dt} \\ \frac{dQ_s}{dt} \end{bmatrix} \tag{2.14}$$

where c_m and c_i are the total heat capacities of the walls and the air, respectively. r_i is the resistance against the heat transfer between the large heat accumulating medium and the room air. r_u is the resistance of the heat transfer to the outdoor air. The details of the expressions of these resistances can again be found by solving the heat balance equations. Furthermore, the system also contains the heat supplies for the radiators, $\frac{dQ_r}{dt}$, and the sun, $\frac{dQ_s}{dt}$. A_w is the area of the window, p indicates the effective part of the window where the solar radiation is having influence on T_m.

2.5.2 Solid walls and light floor

In this model the dominating heat capacity is found in the outer walls, while the floors are considered to be light and well isolated. A typical building with these characteristics is a brick house with wooden floors. Since it was outlined in the latter section how to find and isolate the balance equations, the equations for the heat transfer in this case are found to be:

$$\begin{bmatrix} c_m \frac{dT_m}{dt} \\ c_i \frac{dT_i}{dt} \end{bmatrix} = \begin{bmatrix} -(\frac{-1}{r_i} + \frac{1}{r_u}) & \frac{1}{r_i} \\ \frac{1}{r_i} & \frac{-1}{r_i} \end{bmatrix} \begin{bmatrix} T_m \\ T_i \end{bmatrix} + \begin{bmatrix} \frac{1}{r_u} 0 & 0 & A_w P \\ 0 & 1 & A_w(1-p) \end{bmatrix} \begin{bmatrix} T_u \\ \frac{dQ_t}{dt} \\ \frac{dQ_s}{dt} \end{bmatrix} \tag{2.15}$$

In figure 2.3 is a drawing of the building and the concordant analogous electrical system. The models above can be identified on the basis of the analogous electrical system.

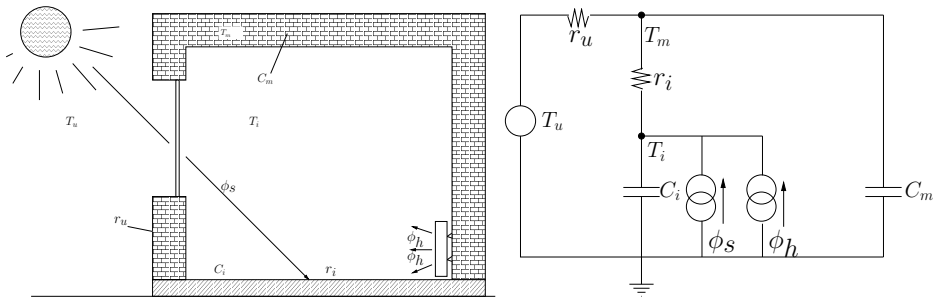

Figure 2.3: Illustration of a building having the dominating heat capacity in the outer walls and a light floor

2.5.3 Light walls and double solid floor

This model describes the situation where the dominating heat capacity is placed in the floor. The walls are considered to be light. Compared to the former models this model has three time constants. The reason why two time constants have been chosen for the description of the heat transfer in the floor area are due to heat sensitivity. This third model has proved to be an excellent way of modelling a greenhouse[Nielsen 1996]. This type of model has proven to be able to estimate reliable and accurate physical parameters [Nielsen 1996]. It was also found that two time constants are not sufficient in order to describe the heat transfer of greenhouses. The model is outlined in Equation 2.16 and in Figure 2.4.

$$\begin{bmatrix} c_{m_1}\frac{dT_{m_1}}{dt} \\ c_{m_2}\frac{dT_{m_2}}{dt} \\ c_i\frac{dT_i}{dt} \end{bmatrix} = \begin{bmatrix} -(\frac{-1}{r_i}+\frac{1}{r_j}) & \frac{1}{r_j} & \frac{1}{r_i} \\ \frac{1}{r_j} & \frac{-1}{r_j} & 0 \\ \frac{1}{r_i} & 0 & -(\frac{-1}{r_i}+\frac{1}{r_u}) \end{bmatrix} \begin{bmatrix} T_{m_1} \\ T_{m_2} \\ T_i \end{bmatrix} + \begin{bmatrix} 0 & 0 & A_w P \\ 0 & 0 & 0 \\ \frac{1}{r_u} & 1 & A_w(1-P) \end{bmatrix} \begin{bmatrix} T_u \\ \frac{dQ_t}{dt} \\ \frac{dQ_s}{dt} \end{bmatrix} \tag{2.16}$$

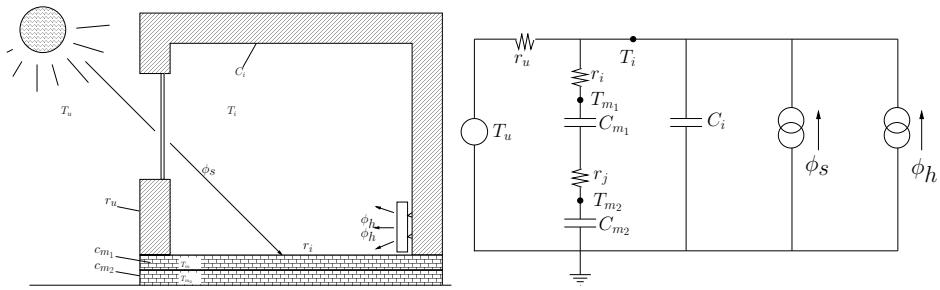

Figure 2.4: A building having the dominating heat capacity in the floor while the outer walls are light constructions

2.5.4 Double solid walls and light floor

The dominating heat capacities in this case is inside the building. Two heat accumulating layers for the walls are included in the model because of the thick outerwalls. The model is shown in Equation 2.17. Figure 2.5 shows a scheme of the building along with the analogous electrical system. This model can for instance be used when modelling old churches since thick walls are often present in these constructions [Madsen et al. 1994].

$$\begin{bmatrix} c_{m_1}\frac{dT_{m_1}}{dt} \\ c_{m_2}\frac{dT_{m_2}}{dt} \\ c_i\frac{dT_i}{dt} \end{bmatrix} = \begin{bmatrix} -(\frac{-1}{r_i}+\frac{1}{r_v}) & \frac{1}{r_v} & \frac{1}{r_i} \\ \frac{1}{r_v} & \frac{-1}{r_v} & 0 \\ \frac{1}{r_i} & 0 & \frac{-1}{r_i} \end{bmatrix} \begin{bmatrix} T_{m_1} \\ T_{m_2} \\ T_i \end{bmatrix} + \begin{bmatrix} 0 & 0 & A_w P \\ 0 & 0 & 0 \\ 0 & 1 & A_w(1-P) \end{bmatrix} \begin{bmatrix} T_u \\ \frac{dQ_i}{dt} \\ \frac{dQ_s}{dt} \end{bmatrix} \tag{2.17}$$

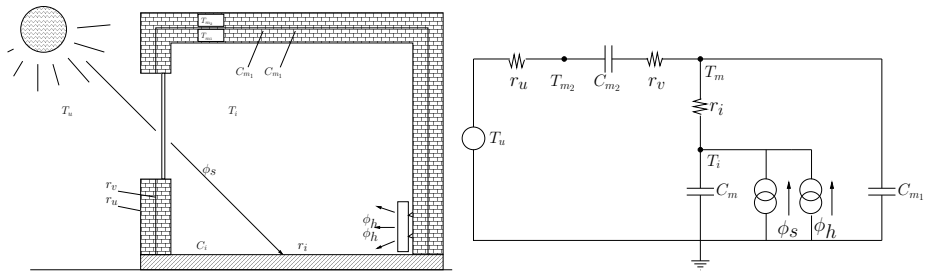

Figure 2.5: Two dominating heat capacity layers in the outer walls, while the indoor floor is a light construction

2.6 Model extentions and data improvements

The four models give a guideline for how to create models for heat transfer of simplified buildings. By using the RC-modelling it is possible to make extentions which suit a particular context. The extentions could be other combinations of the heat accumulating layers or more complicated buildings for instance having more rooms. However, it is important to keep in mind that it is only possible to model linear scenarios with the RC-models.

Besides the development of describing models improvements could also be made in relation to the input variables. A contributor to heat transfer which is not taken into consideration is the temperature of the soil, both underneath the building and in the surrounding area. One reason why the soil temperature in previous models is not taken into consideration could be that the temperature underneath the building is considered to be approximately constant. This will not change the total heat transfer of the building.

It could be interesting to add the heat transfer caused by the surround soil both in sense of the difference in temperature and the reflection to the models. The reflection is of specific interest for instance in areas where snow is common [Technologies 2006].

In relation to reflection this is a measurement which is not available in all cases. The software used in this thesis is able to estimate the influence of unmeasured variables. This is done by adding an extra state in the set of system equations or by including the parameter in the model. Later in the thesis the long-wave radiation to the surroundings is estimated.

Chapter 3

Non-linear Heat Transfer Phenomena in Buildings

In the previous chapter different RC-models have been examined. As mentioned RC-models can only be used to model heat transfer of buildings that are influenced by linear heat transfer phenomena and in cases where approximations are adequate. The aim of this chapter is to identify non-linear heat transfer phenomena that causes the necessity of introducing non-linear modelling methods.

Very little research has been made in the field of non-linear heat transfer. This can be due to the widespread use of RC-models. Another reason is a low level of knowledge in general about non-linear influences on heat transfer. In this thesis the phenomenon is called non-linear heat transfer, in principle this kind of heat transfer can be designated by other terms, such as advanced or unknown ways of heat transfer. This implies the risk that the phenomena actually investigated were not discovered in the research.

The research has shown that radiation and moisture are two of the most well-described and best known causes of non-linearities in relation to heat transfer in buildings. Another less described phenomenon, which can be assumed to influence the heat transfer in buildings with non-linear effects is the outdoor wind-load.

Over time different non-linear terms seem to be important in the modelling, due to the change in choice of constructions and designs of buildings. The dominating building materials in the past decades have been bricks and concrete which are both vulnerable of moisture. One of the new trends of building envelopes is the façades primarily made of glass. These types of façades are expected to be more influenced by the wind conditions and the non-linearities in the case of radiation. Furthermore the double envelopes and shading can contribute to the non-linearities.

3.1 Details of the phenomena

The following sections give some explanations and reasons why the non-linearities occur and in some cases ways to express the phenomena mathematically.

3.1.1 Moisture

The interaction between heat transfer and moisture in buildings is an area where a lot of research has been done. In spite of the effort definition of the problem is still lacking [Xiaoshu 2002]. [Liesen & Pedersen 1999] mention moisture as the most non-linear capacitance depending primarily on vapour density and temperature.
It is possible for the moisture to enter the building envelope since most buiding envelopes are made of porous materials such as bricks [Xiaoshu 2002]. When the moisture enters the bricks several complicated processes will start. It is difficult to describe the boundary conditions when moisture has percolated into the bricks. All degrees of moisture encompassed from dry to saturated influence the material properties such as thermal conductivity, density and heat transfer. In [Liesen & Pedersen 1999] it is noted that especially the heat transfer is influenced. This fact often leads to the necessity of simplifying the description by approximations. When specimens are taken it is very difficult to determine experimentally any joint conclusions, since there is no pattern for the spread of the moisture. These circumstances lead to the nessecity of non-linear modelling. Moisture in the outer walls is not only influenced by the weather conditions, it is known that walls in new buildings have a higher level of water content than old walls. Furthermore, the moisture is not spread uniformly in the walls[Häupl et al. 1997]. This is also an indication that the applied heat transfer coefficient varies over time. This is an important aspect in relation to time series analysis. The following articles treat the topic of moisture in buildings in a non-linear perspective; [Häupl et al. 1997], [Deru & Kirkpatrick 2001], [Liesen & Pedersen

1999] and [Xiaoshu 2002].

3.1.2 Wind speed

The influence of wind load has an impact on the heat transfer. In the article, [Jiménez et al. 2006], on which this thesis is based, it is shown that the ambient wind speed affects the heat transfer from the photovoltaic module to the ambient air. In the article the wind speed is raised to an unknown power, W^k, which makes the influence non-linear. In relation to the materials affected by the wind in a non-linear way, it can only be ascertained for certain for glass, since the photovoltaic module mainly consists of glass. This area of research is interesting in relation to many newly-built buildings having façades dominated by glass. In [Troelsgaard 1981] it is stated that the ambient overall coefficient of the heat transfer, α_u, is a function of the wind speed raised to the power of 0.78. α_u does only have influence in relation to the heat transfer through glass. Empirical results have proven that the calculation of α_u has to be divided in two separate equations, which are found in Equation 3.1, [Troelsgaard 1981].

$$\alpha_u = \begin{cases} 10.8 + 3.9W & W \leq 5m/s \\ 5.0 + 7.15W^{0.78} & W \geq 5m/s \end{cases} \tag{3.1}$$

Comparing the statement of the nessicity of dividing the calculation of the overall coefficient of the heat transfer to the finding in [Jiménez et al. 2006], there is a divergence. From background knowledge about the conditions of the wind at the test site it is known that the measured winds are below 5 m/s, and still the non-linear influence of the wind is tested to be significant. This underlines the strength of estimation methods where it is possible to determine if it is necessary to apply the influence of the wind as non-linear or if it is sufficient to assume linearity.
No articles or analyses have been identified or have described this behaviour for for instance bricks or other porous materials.

3.1.3 Radiation and the solar angle of incidence

The infrared radiation, cf. Equation 2.8, is also a non-linear influence which is needed to be taken into consideration especially, due to the temperature raised to the fourth power. In the later analyses of this thesis it is shown that several infrared radiation contributors are significant.
The relationship between the heat transfer and the angle of incidence is non-linear [Melgaard 1994]. The research has shown that not only the heat transfer and the angle are non-linear, but the number of glass covers has also an impact

on this non-linear relationship. It has been shown that buildings with mechanically ventilated double envelope façades cannot be modelled by RC-modelling since the shading and thereby the incidence angle of the radiation are non-linear [Manz et al. 2004].

Chapter 4

The Mathematical Methods Used in the Modelling

In this chapter the applied mathematical methods will be outlined. The first part describes theory concerning estimation of models while the second part describes methods of analysis of output data of the model.
The strength of the mathematics mathods applied is that it is possible to test the models and estimate not measured variables. A very important aspect of this approach is that the estimation of the models incorporate prior knowlegde, which increases the quality of the models.

4.0.4 Estimation methods

The models applied in the thesis are stochastic differential models, which are differential equations where one or more of the therms are a stochastic process. To estimate parameters in stochastic differential models, two essential issues have to be considered in advance. Primarily if there are any prior knowledge the parameters in the model or if all the parameters have to be estimated. Secondly it is also necessary to decide if the model is going to contain non-linear terms. These two points have impact on which mathematical methods are to be used for estimating the parameters.

4.1 Continuous-discrete stochastic state space models

Continuous-discrete stochastic state space models are built up by two kinds of equations: a set of system equations, (4.1), and a set of measurement equations, (4.2). The system equations are stochatic differential equations and are defined in continuous time whereas the measurement equations are in discrete time. This reflects the fact that the measurements are collected discretely, while the description of the data is modelled to be stated in continuous time. The function $\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})$ is called the drift term and $\boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})$ is the diffusion term. If the diffusion term is not included in the model it will reduce into a state space model based on ordinary differential equations. In this case it is not possible to identify the uncertainty of the system equations. Stochastic state space models enable the possibility to model physical systems where random fluctuations in the states are present.

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t \tag{4.1}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k \tag{4.2}$$

Where $\boldsymbol{x}_t \in \mathcal{X} \subset \mathbb{R}^n$ is a vector of state variables, $\boldsymbol{u}_t \in \mathcal{U} \subset \mathbb{R}^m$ is a vector of input variables, $t \in \mathbb{R}$ is the time variable, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ is a vector of parameters, $\boldsymbol{y}_k \in \mathcal{Y} \subset \mathbb{R}^l$ is a vector of output variables. $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$, $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$ and $\boldsymbol{h}(\cdot) \in \mathbb{R}^l$ are possible non-linear functions; $\{\boldsymbol{\omega}_t\}$ is an n-dimensional standard Wiener process, also known as a random walk, and $\{\boldsymbol{e}_k\}$ is an l-dimensional white noise process with $\boldsymbol{e}_k \in N(\boldsymbol{0}, \boldsymbol{S}(\boldsymbol{u}_k, t_k, \boldsymbol{\theta}))$. $\boldsymbol{\sigma}(\cdot)$ is the gain of the increments of the Wiener process. It is assumed that $d\boldsymbol{\omega}_t$ and $\boldsymbol{e}_k$ are mutually uncorrelated.
The stochastic term, $d\boldsymbol{\omega}_t$, in Equation 4.1 changes the equation from being a deterministic state space model to a stochastic state space model. The applied measurement values may have a deviation from the true values. This deviation will then be contained in the noise term. Another reason for introducing the noise term is that the model can be deficient due to variables not considered in the model having an influence on the system. This can be caused by the following circumstances: either due to a lack of knowledge about the variables' effect on the system, or because measurements of the variable are not available. The measurement error, $\boldsymbol{e}_k$, contains errors in the output signals which are caused by measurement noise [Madsen 2001] and [Andersen 2001].
It is obvious that the aim is to have values of both $d\boldsymbol{\omega}_t$ and $\boldsymbol{e}_k$ as low as possible since this means that the model describes the physical system sufficiently.

4.1.1 Parameter estimation

The purpose of the parameter estimation is to estimate the optimal set of parameters. This is obtained by maximizing the likelihood function of the model. In this thesis two methods of parameter estimation are described, namely the maximum likelihood estimation and the maximum a posteriori estimation. Maximum likelihood estimation is applied when there is no prior knowledge about the parameters whereas maximum a posteriori estimation is perferable if any prior knowledge on the parameters is available. In this thesis only the maximum likelihood estimation will be applied. The mathematics behind the two methods are quite similar since the maximum likelihood estimation is a special case of the maximum a posteriori estimation. It is therefore decided to outline the theory of the maximum a posteriori estimation.

Considering a vector $\mathbf{Y}$ containing S stochastically independent continuous sequences, Equation 4.3.

$$\mathbf{Y} = [\mathcal{Y}^1_{N_1}, \mathcal{Y}^2_{N_2}, \ldots, \mathcal{Y}^i_{N_i}, \ldots, \mathcal{Y}^S_{N_S}] \tag{4.3}$$

Each of the $\mathcal{Y}$'s is defined as in Equation 4.4.

$$\mathcal{Y}^i_{N_i} = [\boldsymbol{y}^i_{N_i}, \ldots, \boldsymbol{y}^i_k, \ldots, \boldsymbol{y}^i_1, \boldsymbol{y}^i_0] \tag{4.4}$$

Furthermore $p(\theta)$ needs to be introduced as the prior probability function of the parameters, θ.
The point estimates in the state space model can be found as the vector of parameters, θ, that maximize the posterior probability density function, Equation 4.5:

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto \left(\prod_{i=1}^{S} p(\mathcal{Y}^i_{N_i}|\boldsymbol{\theta}) \right) p(\boldsymbol{\theta}) \tag{4.5}$$

The expression in the large brackets is the likelihood function.
When applying Bayes rule, $P(A \cap B) = P(A|B)P(B)$, and the principle of products of conditional probability density functions, $p(\mathcal{Y}^i_{N_i}|\boldsymbol{\theta})$, in Equation 4.5 can be written as:

$$p(\mathcal{Y}^i_{N_i}|\boldsymbol{\theta}) \propto \left(\prod_{k=1}^{N_i} p(\boldsymbol{y}^i_k|\mathcal{Y}^i_{k-1}, \boldsymbol{\theta}) \right) p(\boldsymbol{y}^i_0|\boldsymbol{\theta}) \tag{4.6}$$

It is necessary to use the conditional densities to form the likelihood function since the residuals of the ordinary differential equation part of the stochastic differential equations are correlated.
Equation 4.5 and 4.6 lead to the following expression of the posterior probability

density function:

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto \left(\prod_{i=1}^{S}\left(\prod_{k=1}^{N_i} p(\boldsymbol{y}_k^i|\mathcal{Y}_{k-1}^i,\boldsymbol{\theta})\right)p(\boldsymbol{y}_0^i|\boldsymbol{\theta})\right)p(\boldsymbol{\theta}) \tag{4.7}$$

The diffusion term, $d\boldsymbol{\omega}_t$, in Equation 4.2 is assumed to the independent of all the state variables, $\boldsymbol{x}_t$, and furthermore it is driven by a Wiener process. The increment of a Wiener process is Gaussian distributed, which makes it possible to approximate the conditional probability density function by Gaussian densities. This assumption creates the opportunity to apply the Kalman filter technique for estimating the mean and covariance of the conditional probability density function in order to be able to calculate the likelihood function. The Gaussian distribution is completely described by its mean and covariance. The technique will be described in detail in Section 4.1.2. The stochastic differential equations are driven by a Wiener process. The characteristics of the Wiener process are that the increments are Gaussian It is well-known that the density function for the multivariate normal distribution is given as:

$$f(\boldsymbol{x}) = \frac{\boldsymbol{e}xp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T bs\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \tag{4.8}$$

$\hat{\boldsymbol{y}}_{k|k-1}^i$ is the condititional mean, $\boldsymbol{R}_{k|k-1}^i$ represents the condititional covariance and ϵ_k^i is the residual. Equation 4.9-4.11

$$\hat{\boldsymbol{y}}_{k|k-1}^i = E\{\boldsymbol{y}_k^i|\mathcal{Y}_{k-1}^i,\boldsymbol{\theta}\} \tag{4.9}$$

$$\boldsymbol{R}_{k|k-1}^i = V\{\boldsymbol{y}_k^i|\mathcal{Y}_{k-1}^i,\boldsymbol{\theta}\} \tag{4.10}$$

$$\boldsymbol{\epsilon}_k^i = \boldsymbol{y}_k^i - \hat{\boldsymbol{y}}_{k|k-1}^i \tag{4.11}$$

The parameters are assumed to be Gaussian distributed, which leads to:

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} \tag{4.12}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = V\{\boldsymbol{\theta}\} \tag{4.13}$$

$$\boldsymbol{\epsilon}_{\boldsymbol{\theta}} = \boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}} \tag{4.14}$$

When taking these assumptions into account the posterior probability density function can be formulated as:

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto \left(\prod_{i=1}^{S}\left(\prod_{k=1}^{N_i}\frac{\exp\left(-\frac{1}{2}(\boldsymbol{\epsilon}_k^i)^T(\boldsymbol{R}_{k|k-1}^i)^{-1}\boldsymbol{\epsilon}_k^i\right)}{\sqrt{\det\left(\left|2\pi R_{k|k-1}^i\right|\right)}}\right) \times p(\boldsymbol{y}_0^i|\boldsymbol{\theta})\right) \times \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}\right)}{\sqrt{\det\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)}\left(\sqrt{2\pi}\right)^p} \tag{4.15}$$

The last density function term shall only be included if the there is prior knowledge about the parameters. If the term is included the parameter estimation is a maximum a posteriori estimation, while an exclusion of the term leads to a maximum likelihood estimation. The parameter estimates can now be determined by further conditioning on the initial conditions:

$$\mathbf{y_0} = [\boldsymbol{y}_0^1, \boldsymbol{y}_0^2, \ldots, \boldsymbol{y}_0^i, \ldots, \boldsymbol{y}_0^S] \tag{4.16}$$

and applying nonlinear optimization to find the minimum of the negative logarithm of the resulting posterior probability density function, i.e.:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \{-\ln\left(p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{y_0})\right)\} \tag{4.17}$$

In the general case the estimation method implied by Equation 4.17 is maximum a posteriori, but, if no prior information about the parameters is available, it declines to maximum likelihood.

4.1.2 The extended Kalman filter

As stated in the previous chapter the Kalman filter can be applied to calculate the conditional mean and covariance for the assumed Gaussian conditional densities. There are two variations of the Kalman filter. The ordinary Kalman filter is applied when solving linear systems and the extended Kalman filter is applied for non-linear systems. Though both linear and non-linear systems, are present in the thesis, only the theory of the extended Kalman filter will be presented, since the extended Kalman filter decreases into the Kalman filter for linear systems.
The recursive algorithm of the extended Kalman filter consists of two main categories of equations: prediction and update. The prediction equations predict the state and observations. The update equations update the state predictions with the last measurement.
The review of the theory of the extended Kalman filter follows the description in [Jiménez et al. 2006].
The prediction equations of the extended Kalman filter of the output variables are stated in Equation 4.18 and 4.19

$$\hat{\boldsymbol{y}}_{k|k-1}^i = \boldsymbol{h}(\hat{\boldsymbol{x}}_{k|k-1}^i, \boldsymbol{u}_k^i, t_k^i, \boldsymbol{\theta}) \tag{4.18}$$

$$\boldsymbol{R}_{k|k-1}^i = \boldsymbol{C}\boldsymbol{P}_{k|k-1}^i\boldsymbol{C}^T + \boldsymbol{S} \tag{4.19}$$

The residuals are calculated in the innovation equation below.

$$\boldsymbol{\epsilon}_k^i = \boldsymbol{y}_k^i - \hat{\boldsymbol{y}}_{k|k-1}^i \tag{4.20}$$

In each iteration of the extended Kalman filter the Kalman gain is estimated. The Kalman gain determines the level of influence of new observations. If the measurement covariance of an observation is large, the Kalman gain will not attach importance to the observation, and the updated state will be approximately the same value as the predicted one. The Kalman gain equation is:

$$\boldsymbol{K}_k^i = \boldsymbol{P}_{k|k-1}^i \boldsymbol{C}^T (\boldsymbol{R}_{k|k-1}^i)^{-1} \tag{4.21}$$

After estimating the values of the prediction, innovation and Kalman gain equations the mean and the covariance need to be updated applying Equation 4.22 and 4.23.

$$\hat{\boldsymbol{x}}_{k|k}^i = \hat{\boldsymbol{x}}_{k|k-1}^i + \boldsymbol{K}_k^i \boldsymbol{\epsilon}_k^i \tag{4.22}$$

$$\boldsymbol{P}_{k|k}^i = \boldsymbol{P}_{k|k-1}^i - \boldsymbol{K}_k^i \boldsymbol{R}_{k|k-1}^i (\boldsymbol{K}_k^i)^T \tag{4.23}$$

where

$$\hat{\boldsymbol{x}}_{k|k}^i = E\{\boldsymbol{x}_k^i | \mathcal{Y}_k^i, \boldsymbol{\theta}\} \tag{4.24}$$

$$\boldsymbol{P}_{k|k}^i = V\{\boldsymbol{x}_k^i | \mathcal{Y}_k^i, \boldsymbol{\theta}\} \tag{4.25}$$

and $\hat{\boldsymbol{x}}_{k|k}^i$ is the conditional mean and $\boldsymbol{P}_{k|k}^i$ covariance of the state vector.
Aside from the prediction equations of the output variables, also prediction equations for the state variable have to be run through. The state prediction equations are found in Equation 4.26 and 4.27. Initial conditions for the extended Kalman filter are $\hat{\boldsymbol{x}}_{t|t_0}^i = \boldsymbol{x}_0^i$ and $\boldsymbol{P}_{t|t_0}^i = \boldsymbol{P}_0^i$. The extended Kalman filter is sensitive to nonlinear effects and therefore it may be necessary to linearize the extended Kalman filter about the current mean and covariance, equal to the subsampling. This is actually one of the characteristics of the extended Kalman filter. This linearizing of the filter makes it possible to obtain a better approximation, the time interval $[t_k^i, t_{k+1}^i[$ is sub-sampled, $[t_k^i, \ldots, t_j^i, \ldots, t_{k+1}^i[$. In Equation 4.26 and 4.27 the simplified analytical solution to the corresponding linearized propagation equations are written.

$$\frac{d\hat{\boldsymbol{x}}_{t|j}^i}{dt} = \boldsymbol{f}_0 + \boldsymbol{A}(\hat{\boldsymbol{x}}_t^i - \hat{\boldsymbol{x}}_j^i) + \boldsymbol{B}(\boldsymbol{u}_t^i - \boldsymbol{u}_j^i) \tag{4.26}$$

$$\frac{d\boldsymbol{P}_{t|j}^i}{dt} = \boldsymbol{A}\boldsymbol{P}_{t|j}^i + \boldsymbol{P}_{t|j}^i \boldsymbol{A}^T + \boldsymbol{\sigma}\boldsymbol{\sigma}^T \tag{4.27}$$

for $t \in [t_j^i, t_{j+1}^i[$, where the notation:

$$\boldsymbol{A} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_t}\Big|_{\hat{\boldsymbol{x}}_{j|j-1}^i, \boldsymbol{u}_j^i, t_j^i} \;,\; \boldsymbol{B} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}_t}\Big|_{\hat{\boldsymbol{x}}_{j|j-1}^i, \boldsymbol{u}_j^i, t_j^i}$$

$$\boldsymbol{f}_0 = \boldsymbol{f}(\hat{\boldsymbol{x}}_{j|j-1}^i, \boldsymbol{u}_j^i, t_j^i, \boldsymbol{\theta}) \;,\; \boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{u}_j^i, t_j^i, \boldsymbol{\theta})$$

has been applied. The analytical solutions are:

$$\begin{aligned}\hat{\boldsymbol{x}}^i_{j+1|j} &= \hat{\boldsymbol{x}}^i_{j|j} + \boldsymbol{A}^{-1}\left(\boldsymbol{\Phi}_s - \boldsymbol{I}\right)\boldsymbol{f}_0 \\ &\quad + \left(\boldsymbol{A}^{-1}\left(\boldsymbol{\Phi}_s - \boldsymbol{I}\right) - \boldsymbol{I}\tau_s\right)\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{\alpha}\end{aligned} \tag{4.28}$$

$$\boldsymbol{P}^i_{j+1|j} = \boldsymbol{\Phi}_s \boldsymbol{P}^i_{j|j} \boldsymbol{\Phi}_s^T + \int_0^{\tau_s} e^{\boldsymbol{A}s}\boldsymbol{\sigma}\boldsymbol{\sigma}^T e^{\boldsymbol{A}^T s} ds \tag{4.29}$$

where $\tau_s = t^i_{j+1} - t^i_j$ and $\boldsymbol{\Phi}_s = e^{\boldsymbol{A}\tau_s}$, and where:

$$\boldsymbol{\alpha} = \frac{\boldsymbol{u}^i_{j+1} - \boldsymbol{u}^i_j}{t^i_{j+1} - t^i_j} \tag{4.30}$$

has been introduced to allow assumption of either zero order hold ($\boldsymbol{\alpha} = \boldsymbol{0}$) or first order hold ($\boldsymbol{\alpha} \neq \boldsymbol{0}$) on the inputs between sampling instants. Zero hold order is when the value of the input is held constant, while first hold order method estimates a linear interpolation between the inputs.

4.2 Testing of the model

When a model has been estimated in CTSM, it is possible and necessary to test the model in order to determine if the model is adequate. The testing can be divided into two main areas. First, the model has to be adequate in relation to the CTSM modelling procedure. This implies investigating whether the estimated parameters are inside the estimation interval, and whether they are significant. Furthermore the correlation coefficients of the parameters have to be examined. Secondly, the residuals of the models that fulfil the first assumption have to be analyzed. This analysis will be completed in both the time domain and the frequency domain.

4.2.1 Check of the estimated model

4.2.1.1 T-test of the parameters

The t-test is used to test whether a parameter can be rejected to be equal to zero. One of the assumptions of applying this test is that the estimators are asymptotically Gaussian distributed. When using maximum likelihood estimation the parameter estimations can be assumed to be asymptotically Gaussian distributed with mean $\boldsymbol{\theta}$ and covariance $\boldsymbol{\Sigma}_{\hat{\theta}}$ in accordance with the central limit

theorem. The use of the maximum likelihood estimation makes it possible to find both the prediction error estimates of the parameters and the variance. The covariance is equal to $\boldsymbol{H}^{-1}$. The matrix $\boldsymbol{H}$ is approximately given by:

$$h_{ij} \approx -\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\left(p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{y_0})\right)\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

for $i,j=1,\ldots,p$. h_{ij} is the i,j-element in $\boldsymbol{H}$. The expression in the equation above is the Hessian matrix evaluated at the maximum of the likelihood value, in CTSM denoted as the negative objective function. Since the aim is to identify the uncertainties of the individual estimates of the parameters, it is necessary to make a decomposition of the covariance matrix, like in Equation 4.31

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}}\boldsymbol{R}\boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}} \tag{4.31}$$

$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ is the diagonal matrix of the standard deviations of the parameter estimates and $\boldsymbol{R}$ is the correlation matrix. The asymptotic Gaussianity described above can be applied when testing for significance of the estimated parameters. The null hypothesis of the test is:

$$H_0: \quad \theta_j = 0 \tag{4.32}$$

and the alternative hypothesis is:

$$H_1: \quad \theta_j \neq 0 \tag{4.33}$$

This kind of hypothesis, where it is tested whether a parameter can be assumed equal to a specific value, leads to a two-tailed test. This implies that the null hypothesis is rejected for values of the test statistic placed at either tail-end of its sampling distribution. The Z-value is calculated as:

$$Z(\hat{\theta}_j) = \frac{\hat{\theta}_j}{\sigma_{\hat{\theta}_j}} \tag{4.34}$$

$\hat{\theta}_j$ and $\sigma_{\hat{\theta}_j}$ are both calculated in CTSM. Z belongs to a t-distribution with the degree of freedom equal to the number of observations, the length of the timeseries, subtracted the number of estimate parameters. Having calculated the Z-value it is possible to estimate the probability, p-value, of the parameter being equal to zero.
The theory is also outlined in [Madsen & Holst 2000] and [Kristensen & Madsen 2003*a*].

4.2.2 Stable parameter estimates

When CTSM estimates the parameters several values are estimated in parallel. These values make it possible to check whether the parameters are free of the

limits. In order to estimate the models in CTSM it is necessary to define an interval by minimum, initial and maximum values. The limits are the minimum and maximum values. If the parameter estimate is well free of the limits, it means that the parameters are satisfactorily estimated. The values that have to be investigated in order to determine if the parameters are adequate, are stated in the three points below [Kristensen & Madsen 2003*b*].

- The value of the penalty function has to be significant compared to the value of the objective function
- $\frac{dF}{dPar}$, the derivative of the objective function has to be close to zero
- $\frac{dPen}{dPar}$, the derivative of the penalty function has to be significant compared to $\frac{dF}{dPar}$

If the three points above are not fulfilled it is necessary to loosen the intial limits. When the new limits are set the model has to be reestimated.
Also the correlation matrix of the parameter estimates gives indications about the model estimation being applicable. When analyzing the correlation values in CTSM as a rule of thumb correlation values up to 0.96 are acceptable. Results containing correlation values above 0.96 is an indication of the model being overparameterized. In this case it can be necessary to remove one or more of the parameters from the model. Since the parameters are correlated it is always preferable to remove one parameter at a time starting with the parameter having the highest p-value. In special cases prior knowledge about the parameters may leed to another procedure.

4.2.3 Analysis of the residuals

After estimating and testing the parameters the last analysis that has to be carried out to verify the applicability of the model is the residual analysis. The analysis of the residuals can be used in two ways:

- To validate a model
- To give input as to how to further develop the model in order to give an improved description of the data

In this section several different analyses of the residuals are examined. The residuals are defined as the difference between the true value and the estimated

value of the output variable, see Equation 4.35.

$$\epsilon_t(\theta) = Y_t - \hat{Y}_{t|t-1,\theta} \tag{4.35}$$

If the model is adequate the residuals must be white noise. White noise is defined as being random mutually uncorrelated identically distributed stochastic variables with mean value 0 and constant variance, σ_ϵ^2 [Madsen 2001]. Different analyses identify various interpretations of the residuals. This implies that it is not sufficient to perform one or two of the analyses to get a characteristic line of the residuals. If the analysis of the residuals proves that the residuals can be assumed to be distributed like white noise, and it can be concluded that the model fits the data well. In the opposite situation the residuals can give guidelines of how to expand or change the model.

The initial analysis of the residuals is to plot the residual versus the time. This plot may reveal non-stationarities and potential outliers. When analyzing the residuals further, there are two main approaches: a test in the time domain and a test in the frequency domain. Different methods in the two domains will be outlined in the two next sections.

4.2.3.1 Residual analysis in the time domain

Test of the autocorrelation function
In connection with the time domain tests the most dominant test is to plot the estimated autocorrelation function, $\hat{\rho}_\epsilon$, with the approximate 95% confidence interval for the time lags. The autocorrelation function may reveal if some of the variations in data are not described in the model, for instance periodical tendencies. The limits are found as the $\pm$ 2 standard deviation. The mean and variance of the autocorrelation are given as:

$$\mathrm{e}[\widehat{\rho}(k)] \simeq 0; \quad k \neq 0, \tag{4.36}$$

$$V[\widehat{\rho}(k)] \simeq 1/N; \quad k \neq 0. \tag{4.37}$$

If the residuals are white noise the autocorrelation function is defined as:

$$\hat{\rho}_\epsilon(k) = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \ldots \end{cases}$$

In most cases not every value of the estimated autocorrelation function, $\hat{\rho}_\epsilon$, for $k > 0$ is exactly zero. This is the reason why the 95% limit is used for deciding whether the residuals can be assumed to be white noise. Below is stated how

to calculate the autocorrelation function.
First the autocovariance function needs to be introduced, Equation 4.38.

$$\gamma(k) = \mathrm{Cov}[X(t), X(t+k)] \tag{4.38}$$

$X(t)$ is a stationary process, t is the time indicator and k denotes the time lag. When the autocovariance function has to be worked out, Equation 4.39 can be used for defining the autocorrelation function.

$$\rho(k) = \gamma(k)/\gamma(0) \tag{4.39}$$

From Equation 4.39 it is obvious that $\rho(0) = 1$, cf. the definition of the autocorrelation function above.
Using the theory above the autocovariance function is calculated in the following way:

$$C(k) = C(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} (Y_t - \overline{Y})(Y_{t+|k|} - \overline{Y}) \tag{4.40}$$

for $|k| = 0, 1, \ldots, N-1$. Furthermore, $\overline{Y} = (\sum_{t=1}^{N} Y_t)/N$.
Based on the estimated autocovariance function the estimated autocorrelation function can be calculated as:

$$\widehat{\rho}(k) = r_k = C(k)/C(0), \tag{4.41}$$

Test of the partial autocorrelation function
Similar to the autocorrelation function the partial autocorrelation function can be used for determining if the estimated models are adequate. The partial autocorrelation function is favourable to uncover the necessity of adding an extra state to the model. The partial autocorrelation can be estimated when few values need to be estimated by using the Yule-Walker equations. A prevalent numerical method is the recursive method described in the Appendix of [Madsen 2001].

Test of change in the signs
If the residuals are assumed to be white noise the mean will be close to zero. On this basis it must be assumed that in average there will be a change in the sign of the residual every second time, therefore $p = \frac{1}{2}$. Since there are only two possible outcomes the change in sign test is binomially distributed. As mentioned the probability of change in sign will be located close to $\frac{1}{2}$, and when the numbers of residuals are high it is possible to approximate the binomial distribution by using the normal distribution. As a rule of thumb this approach can be applied when np and $n(1-p)$ are above 15. When p has to be close

to $\frac{1}{2}$, the number of residuals just have to be above 30. The Z-value can be calculated as:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{4.42}$$

To reject the null hypohesis $p = \frac{1}{2}$ Z shall either be larger than $z_{\alpha/2}$ or less than $-z_{\alpha/2}$. α is the level of significance. The theory can be found in [Madsen 2001].

Portmanteau lack-of-fit-test Apart from the test of the autocorrelation function the Portmanteau lack-of-fit test can reveal if the values of the autocorrelation function are not complying with the random error.

$$Q^2 = (\sqrt{N}\hat{\rho}_{\epsilon(\hat{\theta})}(1))^2 + (\sqrt{N}\hat{\rho}_{\epsilon(\hat{\theta})}(2))^2 + \cdots + (\sqrt{N}\hat{\rho}_{\epsilon(\hat{\theta})}(k))^2 \tag{4.43}$$

which can be reduced to:

$$Q^2 = N\sum_{i=1}^{k}\rho(i)^2 \tag{4.44}$$

N is the number of observations. k is the number of considered autocorrelations. Textbooks suggest that the appropriate value of k lies within 15 and 30. The calculated value of Q^2 has to be weighed against the χ^2-distribution with $m-n$ degrees of freedom. It is assumed that the distribution of Q^2 is approximately χ^2-distributed. n is the number of estimated parameters in the model [Madsen 2001].

4.2.3.2 Residual analysis in the frequency domain

Test in the cumulated periodogram
All the previously mentioned analyses analyze the residuals in the time domain. The normalized cumulative periodogram is used for testing the residuals in the frequency domain. This test can reveal if there are any specific areas of the frequency domain when the residuals are over-represented. Such an over-representation could be due to seasonal or periodic skew behavior of the residuals. The variation of white noise is uniformly distributed, which implies that no frequencies ought to be over-represented. Similar to the autocorrelation function the periodogram contains are 95% limit band. A straight line from (0,0) to (0.5,1) inside the limit in the periodogram indicates that the residuals are white noise.
The periodogram is calculated in the following way: The equation for the periodogram for the residuals, $I(\nu_i)$, can be seen in Equation 4.45, having the frequencies $\nu_i = \frac{i}{N}$, where i is defined in the following interval; $i = 0, 1, \ldots, N/2$. N denotes the total number of observations.

$$\hat{I}(\nu_i) = \frac{1}{N}\left[\left(\sum_{t=1}^{N}\epsilon_t cos 2\pi\nu_i t\right)^2 + \left(\sum_{t=1}^{N}\epsilon_t sin 2\pi\nu_i t\right)^2\right] \tag{4.45}$$

$\hat{I}(\nu_i)$ denotes the amount of variation of ϵ_t related to the frequency ν_i The normalized periodogram is

$$\hat{C}(\nu_j) = [\sum_{i=0}^{j} \hat{I}(\nu_i) / \sum_{i=0}^{\frac{N}{2}} \hat{I}(\nu_i)] \tag{4.46}$$

The confindence limits for the periodogram are calculated by $\frac{\pm K_\epsilon}{\sqrt{q}}$, where q is $q = \frac{(n-2)}{2}$ for n even and $q = \frac{(n-1)}{2}$ for n odd. The approximated coefficient K_ϵ for the probability limit 5% is 1.36 [Box & Jenkins 1976].

4.3 Filtering methods

One of the purposes of the thesis is to identify how the outdoor wind influence on the building integrated photovoltaic module. An anemometer has measured the actual wind speed during the testing. There are scattered opinions as to whether it is the actual wind or a filtration of the wind that have impact on the module. Four different versions out of these are three filtered. The theory of the filters is described below.
The first and simplest filtration is a moving average based on a certain length of observations. The more observations the calculation of the moving average is based on, the less of the original variation is found in the filtered version. In this thesis the moving averages are estimated from intervals of 20 observations.
Lowpass filtering is the second method applied. This filter provides the opportunity of filtering away the high frequencies. The lowpass filter is calculated by the software tool Splus, which uses complex demodulation to identify the lowpass filtrated variable. This method is outlined in [Bloomfield 2000]. In short it is supposed that the variable containing the wind speed can be rewritten as:

$$W_t = R_t cos 2\pi(f_0 t + \phi_t)$$

In order to calculate the lowpass filter it is necessary to determine a cutoff frequency, f_c, which is defined as $f_0/2$. Only the frequencies below the cutoff value are kept through the filter. The smaller f_c, the less variation is lead through the filter. In agreement with the statements above the lowpass transfer function is defined as:

$$H(f) = \begin{cases} 1 & 0 \leq \mid f \mid \leq f_c, \\ 0 & f_c < \mid f \mid \leq \frac{1}{2} \end{cases}$$

Aside the moving average and the lowpass filter, it is also attempted to apply a causal filter. The causal filter is applied in the models as an extra state space equation, $W_f = aW_f - aW$. This means that CTSM estimates a in such a

way that the influence of the actual wind speed, W, and the filtrated wind speed, W_f, is optimal in order to describe the data. The expression above is the relationship is written in continuous time. An easier way in which to understand the idea behind the causal filter, is by rewriting the expression in discrete time, which can be written as $W_f = aW_f + (1-a)W$. This expression shows that the objective of the equation is to weight the influence of the filtered wind and the measured one. Comparing the causal filter with the other filter and the way they are applied in the thesis, it is evident that the causal filter is able to match the specific data, while the other filters are more stationary.

4.4 Model validation

The performance of a model will nearly always increase when adding more terms to the model. The model validation in this thesis is the likelihood ratio test, which can clarify statistically if the improvement of the model is significant.

4.4.1 The likelihood ratio test

The likelihood ratio test makes it possible to test if the full model has a significantly better performance than a sub model. The sub model is a similar model consisting of fewer terms. The two hypotheses are:

$$H_0 : \theta \in M_0$$

$$H_1 : \theta \in M_1$$

θ represents the parameters. M_0 is a subset model of the full model M_1. The likelihood function, $L(\theta|\mathcal{Y}^i_{N_i})$, is given as $p(\mathcal{Y}^i_{N_i}|\boldsymbol{\theta})$. The likelihood ratio is

$$\Lambda(\mathcal{Y}^i_{N_i}) = \frac{sup\{L(\theta|\mathcal{Y}^i_{N_i} : \theta \in \Theta_0\}}{sup\{L(\theta|\mathcal{Y}^i_{N_i} : \theta \in \Theta\}}$$

The test statistics of the likelihood ratio test are calculated as:

$$L_{ratio} = -2 \cdot log(\Lambda(\mathcal{Y}^i_{N_i})) = 2 \cdot log(\frac{sup\{L(\theta|\mathcal{Y}^i_{N_i} : \theta \in \Theta_0\}}{sup\{L(\theta|\mathcal{Y}^i_{N_i} : \theta \in \Theta\}})$$

$$= -2 \cdot (log(sup\{L(\theta|\mathcal{Y}^i_{N_i} : \theta \in \Theta_0\}) - log(sup\{L(\theta|\mathcal{Y}^i_{N_i} : \theta \in \Theta\}))$$

The value of L_{ratio} is asymptotically χ^2-distributed. The degrees of freedom of the χ^2 distribution are estimated as $(dim(M_1) - dim(M_2))$. dim denotes the

number of estimated parameters in the models. If L_{ratio} is below the χ^2 value the h_0 hypothesis is accepted. In the other case the test stresses the necessity of the full model [Madsen & Holst 2000].
CTSM estimates the objective function of the model. The objective function is the negative log-likelihood function.

Chapter 5

Modelling in CTSM

The parameters for the constructed models are estimated by using the free-download application CTSM, which is an abbreviation of Continuous Time Stochastic Modelling.
This section is meant to give some hints about how to estimate models in CTSM. A user guide [Kristensen & Madsen 2003*b*] can be downloaded from the CTSM homepage - www2.imm.dtu.dk/ctsm. The work in relation to this thesis has in different situations given rise to some problems. The purpose of this chapter is to avoid difficulties for future users of CTSM.

5.1 Preparations for the modelling

It is important to identify the type of model that is to be estimated. This will influence the way the model has to be typed-in in CTSM. If the model is non-linear it is straightforward to type it in, while linear models have to be sorted by a state and an input vector. This entails that it is not possible to transfer a matrix notation directly to CTSM.

The dataset to be applied has to be set-up in a .csv-file in the following way:

- The time variable
- The input variables
- The output variable(s)

It is necessary to satisfy all three categories in order for CTSM to proceed the calculations.
In relation to the input variables in the file for estimation purpose it is very important to mention that the variables need to appear in the same order as the columns in CTSM. It is not possible to have headers of the variables for recognition purposes.

5.2 The actual modelling

The following gives some advise of how to use CTSM. The advices below have appeared during the modelling process and from the CTMS manual.

- It is of great importance not to use too widespread limits for both states and parameters. If too wide limits are used it is not possible for CTSM to find a model. Wide limits can in some cases be in the magnitude of 10^{-3}. The wider the limits are, the longer the calculations will last, because CTSM will have a larger set of possible outcomes. On the other hand it is also important not to tighten the limit too much, so that CTSM is inable to find a feasible solution. It may take a couple of attempts to find workable limits for the estimation.
- When some workable limits have been found, the p-values and the penalty functions indicate if the initial values are acceptable in order to describe the data satisfactorily.
- It can be an advantage to use the exponential function of the noise, since this enables the values of the parameters to be both positive and negative. While doing this it is important to make the limit cover both positive and negative values.

- In CTSM it is not possible to weigh the output variables. A way in which to solve this problem is by adding extra state space equations. This is attempted in Chapter 9.
- It is possible to change the time unit in the estimation. For instance, if the time interval needs to be changed from minutes into seconds. This is done by multiplying the time variable by 60. Somehow this has a significant extending influence on the estimation time.
- Sometime when a function, like the logistic function, is added in a model CTSM is only able to find a local minimum, which implies that the parameter estimates due to a low number of iterations, are estimated close to the intial value of the parameter. If this occurs new initial guesses can remedy the problem.

5.3 After the modelling

No troubles have been identified in this phase of the modelling.

Chapter 6

Photovoltaic in General

6.1 Introduction

The following chapter is an introduction to photovoltaic technology; what is it and how does it work? Photovoltaic is often denoted PV. Also test standards and prior analyses will be discussed. This is done in order to be able to stress the basis for building up the models. Prior knowlegde and testing may reveal important issues.
PV technology has a huge potential even though the efficiency of the equipment today and in general has some limitations. In less than an hour the sun delivers enough energy to cover the energy demand of one year of the world's population[Markvart 2000]. The efficiency of PV cells today range from 5% to 25% depending primarily on the applied materials and methods. The efficiency expresses how much energy is converted into electricity. The rest is mostly transformed into heat in the module or reflected.
From 1992 to 2002 the cumulative installed PV capacity in the International Energy Agency countries raised from 109.9 MW to 2596 MW. This indicates a market on the rise[IEA 2006].

6.2 What is PV technology?

PV is an abbraviation of photovoltaic, which literally means light-electricity. 'Photo' dates from Greek 'phos' meaning light, and 'Volt' from the Italian scientist Volta. The most widespread PV solution are the stand-alone solutions, which are inclined and surrounded by air. These modules can either be installed at the ground or at roofs. PV cells produce electricity on the basis of solar radiation. Another method by which to produce renewable energy from the energy deliverred by the sun is solar heating where for instance water and air are heated. The main purpose of PV modules is to produce electricity, but the PV modules can also satisfy other functions. The PV module analyzed in the thesis is meant for integration in building façades. BIPV is an abbreviation of building integrated photovoltaic.One of the ideas of the building integrated photovoltaic, BIPV, module analysed in this thesis is to generate heated air in the ventilated gap behind the module. This preheated air can be utilized as heating for the building. Meanwhile the PV module is cooled down. Since the efficiency of the module is negative correlated with the temperature of the module a double gain is obtained. This is described in [Bazilian et al. 2002] and [Andresen 2002]. A residential-scale ventilated BIPV system is tested in Sydney [Mai et al. 200x]. Also the comfort of the building has to be taken into account. This implies that the indoor climate and the inflow of light to the building do not decrease unacceptably due to the installation the modules. Beside heating the building, the modules can be installed for shading purposes. This can be seen as a cooling-down factor[Andresen 2002]. There is a strong negative correlation between the efficiency of the module and its temperature. For this reason it is beneficial to have the module cooled down by the ventilation.

6.3 The construction of a PV module

Before having an operational PV module, an extensive procedure has to be carried out. A PV module is composed by cells. The procedure below is a short description of how a cell is made. The prevailing material for PV cells is silicon, which can be extracted from sand. There are for the moment three dominating technics and final products. At this point of time research is carried out in order to identify less expensive PV matrials. The production of the PV equipment is expensive in the perspective of both finance and energy. Therefore, factors such as cost and energy consumption in the production phase, the efficiency of the cells, and the physical flexibility of the cells and modules must be taken into consideration in the phase of decision. These factors are listed, for the three products, in the bulleted summary below.

- **Monocrystalline silicon**
  - The production of the monocrystalline silicon is the most expensive and time consuming process. This is due to the fact that it takes more time to grow a mono crystal compared to a poly crystal. The long production time and procedure leed to a high level of energy consumption.
  - Today the efficiency of the commercial monocrystalline silicon cells is about 12-16%.
  - The flexibility is rather limited, because the cell cannot resist even medium angles of bending. If the cell is bent too much it will simply break and become dysfunctional. Furthermore, it is not possible to repair a broken cell.
- **Polycrystalline silicon**
  - The production of the polycrystalline silicon cells is faster and less time-consuming than the monocrystalline silicon. This results in lower economical expenses compared to the monocrystalline.
  - The efficiency appears in the area of 9-12%.
  - Polycrystalline has the same limited flexibility as the monocrystalline silicon.
- **Amorphous silicon**
  - Often referred to as thinfilm cell.
  - The expenses in relation to the production of amorphous silicon is in the same area as polycrystalline. It is hoped that productions can be carried out at far lower expenses.Furthermore, the amount of material needed is less than 1 μm in thickness. In comparison to the two other technics the thickness of amorphous silicon is about 200 times less.
  - One of the drawbacks of amorphous silicon is that the efficiency is only in between 3-8%.
  - Amorphous silicon can be sputtered on curved surfaces and also over large areas.

The facts above are found in [Laukamp et al. 1998] and [Andresen 2002].
To sum up the bullets above the decision process is a cost benefit analysis where the costs, needs and functions are weighted against each other. For the moment the polycrystalline silicon PV modules are desirable due to an acceptable efficiency and lower expenses in relation to production and investment.
For both the mono- and polycrystalline silicon the crystals are grown in blocks.

In short, the block is cut into thin sliced cells, of 200-300 μm. The cells are then doped and anti-reflection coated in order to increase the efficiency of the cells. Finally the cells have to be screen printed. After the treatment the most common procedure is to connect the cells in series to construct a module. The series of cells are placed between materials that can resist all kind of weather conditions. The are two dominating materials: glass and tedler. In Figure 6.1 the structure of the module analyzed in this report is illustrated. The PV cells are polycrystalline silicon. In the phase of production it is important that all air gathered between the sheets of glass is removed, otherwise this would lead to lower efficiency of the module. Another very important issue is that, in accordance basic theory of electricity, the weakest cell in a series will determine the efficiency of the module. Due to this all cells are tested and classified.

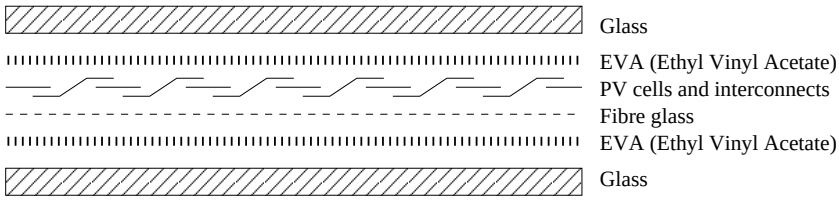

Figure 6.1: Illustration of the structure of the PV module

6.4 Standards for PV modules

A survey among the European Members States reveals that standards of calculation do not cover the area of renewable energy [Bloem n.d.*a*]. In different parts of the world different generel standards are applicable, but in an over-all perspective the standards are rather identical. The standards cover both the module set-up, the testing, and the measuring methods. The data applied in this report are measured in accordance with the European Standard, IEC61215.

The reason why it may be profitable to investigate the standards is for instance that guidelines are set out for which observations are necessary in relation to modelling or analyzing the performance of a module. It is important to keep in mind that the efficiency of the module and its temperature are closely linked together. The higher the temperature of the module becomes, the less efficient the module. The standards also have regulations of how to carry out testing. The present European standard, IEC61215, is based on the following standard

reference environment(SRE), when identifying the nominal operating cell temperature (NOCT):

- Tilt angle: At normal incidence to the direct solar beam at local solar noon
- 800 $\frac{W}{m^2}$ irradiance
- 20 ^oC ambient temperature
- Wind speed: 1$\frac{m}{s}$
- Open-rack mounted module

The problem concerning the standard is that operational PV modules, especially the BIPV modules, and the environment rarely satisfy the conditions stipulated in the standard. For the moment, international standard organisations are developing standards suitable for the growing market of BIPV and on the demand from building designers. In relation to BIPV most of the conditions stated in the present standard is not fulfilled. The irradiance of 800 $\frac{W}{m^2}$ is the irradiance at noon for a PV module in the optimal position. BIPV modules will nearly always be placed vertically in the facade. Furthermore there is no free air flow at the rear side as for an open-racked module. The main problem about the BIPV modules not fulfilling the standard, is that the factors listed above have considerable influence on the module temperature. In a Dutch calculation norm for the Energy Performance of dwellings and residential buildings it is estimated that the efficiency of the modules is reduced by 0.7 for roof installations [Bloem n.d.*a*]. This is mainly due to the raise in temperature. The temperature has impact on the electrical efficiency of the module of about 0.5 $\%/^oC$ referring to the conditions at 25^oC. In Figure 6.2 two curves show the relationship between the irradiance and the efficiency at different temperature levels. 45^oC is not an unusual module temperature. The highest temperature in the top of the module analyzed in this thesis is about 50^oC. Without the forced ventilation the module temperature can get even higher [Christ 2001].

In extension to these obvious deviations a test containing four different scenarios has been carried out on a specific day at JRC in Ispra. The test maps out the difference in efficiency by examining the temperature difference between the module and the ambient temperature. In short, the constructions of the four modules are: free-rack, façade application, insulated and a roof application.

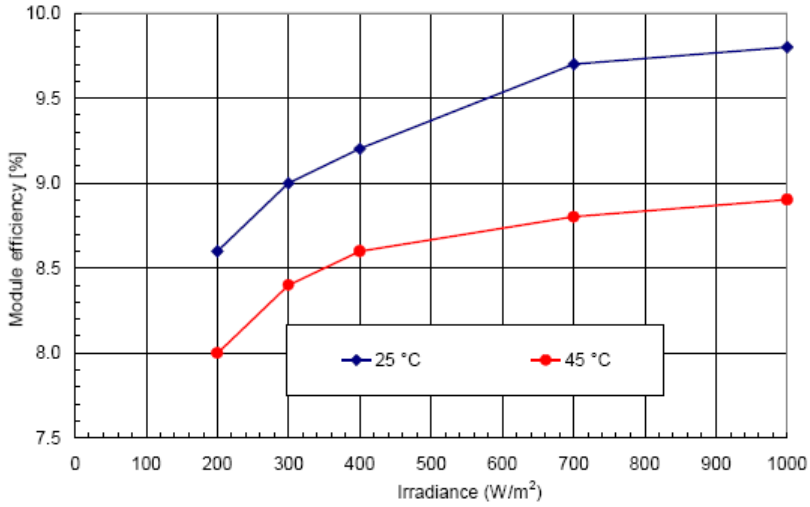

Figure 6.2: This plot shows the influence of the module temperature on the efficiency referring to the conditions at $25^{o}C$ and $1000W/m^2$ irradiance [Christ 2001].

In Figure 6.3 both the theoretical notion and the result of the test are illustrated. The test is described in several articles, e.g. [Jiménez et al. 2006] and [Bloem n.d.*b*]. It can be seen that the line of the free-racked has the smallest slope. This implies that the module temperature of the free-rack is the lowest of the four from the test construction. The plots also reveal that the line referring to the façade integrated module is placed just below the line of the insulated module. This gives a guidance that the BIPV modules have higher module temperatures which leads to a lower efficiency. The fact that the lines do not follow the same line forth and back is due to the heating of the module during the day hours.

Apart form relating to the test environment, the standard also contains guidelines for how to carry out testing and guiding as to which variable must be measured and collected. Further, it is stated in the standard what kind of equipment that has to be used in the testing. The standard rules that the interval of the measurement of the variables must be of maximum 60 seconds. As a minimum irradiance, ambient temperature, cell temperature, wind speed and wind direction must be collected.

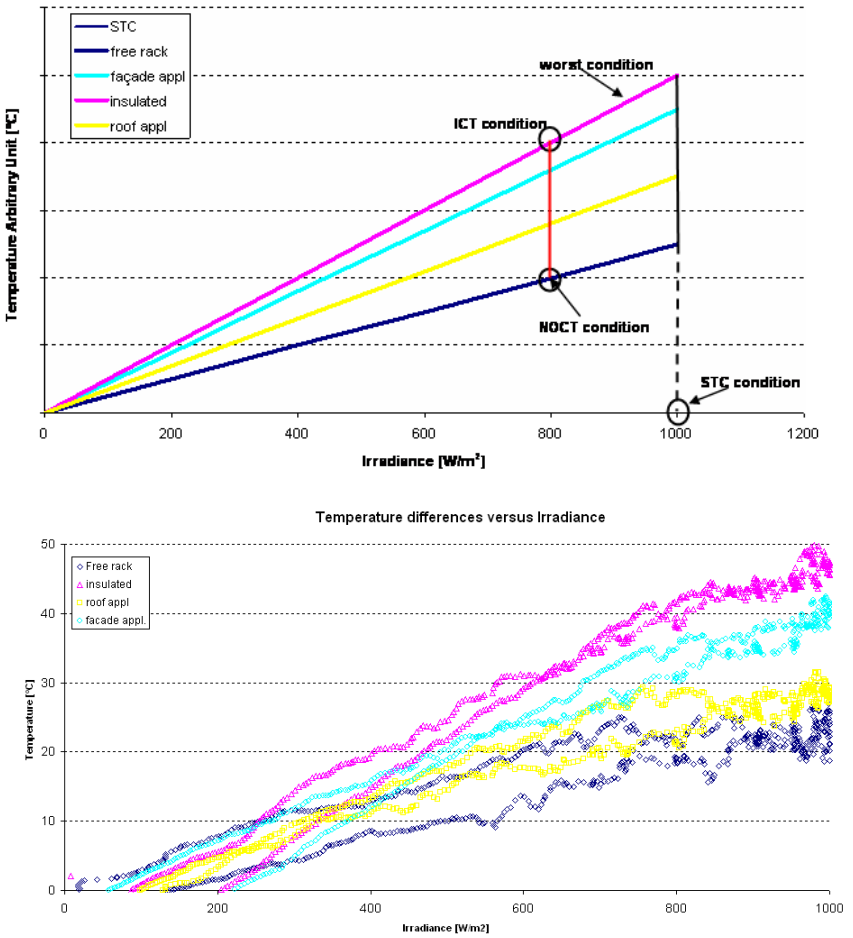

Figure 6.3: The relation between the difference between the module and the ambient temperature versus the irradiance [Jiménez et al. 2006]. ICT denotes the insulation cell temperature. The upper plot show the theoretical relationships, while the other plot arise from measurements.

6.4.1 Prior research in the field of modelling of PV modules

In the search for previous mathematical modelling of BIPV or analysis of similar tests of BIPV, it has become clear that not much research has been done. More research has been carried out on standard PV free-rack mounted modules. This research can also be of interest, since the most fundamental difference between the PV and the BIPV is the ventilated air-gap behind the module and the deviations from the standard. The primary model of interest is the thermal models explaining the efficiency of the modules.

The mathematical method, utilizing stochastic differential equations, applied in this report is only recognized in [Jiménez et al. 2006], from which this thesis arises. In both cases the estimation of the models is carried out in CTSM. One of the big advantages of CTSM is that it is possible to estimate models containing non-linear terms. This article presents models for BIPV modules and is based on the same data as applied in Section 8. In one article neural networks and linear regression are applied to estimate a model[TamizhMani et al. 2003].

Previous research may give some guidelines as to determining which variables have to be considered in the models. In [Luque & Hegedus 2003], it is mentioned that the humidity may influence the performance of the PV module. Besides that, the ambient temperature, the irradiance and the wind speed the humidity and the direction are included in the introductory models in [TamizhMani et al. 2003]. After reducing the model, it is found that neither the humidity nor the wind direction are significant. The data are measured at the present test site, but these variables are not included in the data sets. In a further analysis it could be interesting to investigate if the findings above can be proven to be correct.

Chapter 7

The Applied Variables and the Measuring Methods

This chapter gives an introduction of the test site and the data collected and applied.
All data analyzed in the present report have been collected at the Test Reference Enviroment, TRE, at the Joint Research Center, JRC, in Ispra, Italy in 2002

7.1 The Test Reference Environment

The main idea of TRE is to make the set-up as identically to a building façade as possible in order to be able to investige the behaviour of module. Therefore the module is placed in an insulated box and an air gap between the module and the wood board acts as the façade. The present test configuration is inspired by earlier test site, where the PV module was integrated in a test cell[Bloem n.d.*b*]. One of the improvements made is that the tubes for the inlet and outlet air are moved from the front to the rear side of the test set-up. This is done to reduce preheating of the inlet air.

Two drawings of the test set-up is shown in Figure 7.1. The drawing to the left shows the energy flows influencing on the module. The overall energy bal-

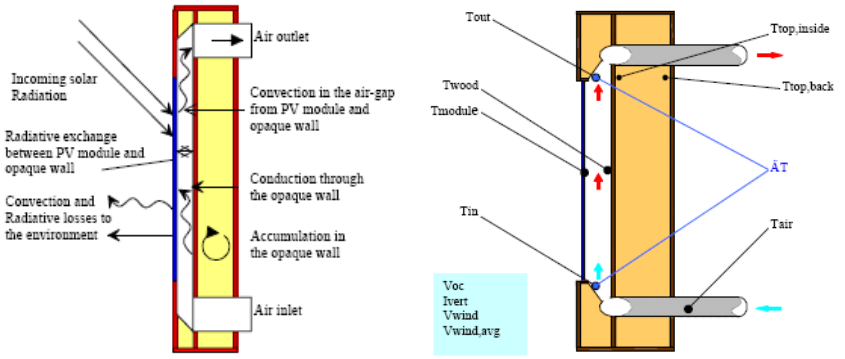

Figure 7.1: Overview of the test set-up [Bloem n.d.*b*]

ance of the BIPV system is very complex. The models are only based on the convective and radiative energy flows, since these two heat transfer methods are the most dominant. The heat transfer through conduction is neglected in accordance to the thickness of the module, which is only less than two centimetres thick. TRE is constructed in such a way that the heat transfers can be measured accurately. The conduction through the opaque wall, consisting of the wooden wall and insulation material, have not been taken into consideration due to the fact that the wall is well insulated. On the right side of the figure is a drawing showing where the variables are measured. In addition to the drawing a $T_{moduletop}$ and a $T_{modulebottom}$ are measured seven centimetres from the edge of the module, as in the case of the top temperature of the wooden board shown in the drawing. Some facts about the Test Reference Enviroment are shown below.

- The dimensions of the module are 120x120cm
- The module contains 121 PV cells.
- Ten centimetres mechanically ventilated air bag between the module and the wooden board.

A more detailed description can be found in for instance [Bloem n.d.*b*].

7.2 Introduction to the measurements

This section contains some general remarks on the measurements. All the variables are measured once every minute according to the maximum interval specified in the standard. One of the biggest problems in relation to this test is that some of the measurements have systematic errors. For instance some equipment can be installed in a wrong angle. In the upcoming section the size of the errors attached to the different measurements and its influence on the results will be discussed. It is necessary to distinguish between random error and systematic error. Since the models are based on a large data material, the random errors will not have a depreciating influence on the results. It is more complicated to deal with the systematic error, since the same size of error will be applied in the entire model

7.3 Output variable

The purpose of this thesis is to model and predict the PV module temperature. The module temperature is therefore the only output variable.

7.3.1 The module temperature

The measurement of the module temperature is rather tricky. The first reason as to the difficulties of measuring the accurate temperature of the module is due to absence of the possibility of placing a thermo couple inside the cell to measure the temperature. To make the best possible measurements two thermo couples are placed seven centimetres from the plate edge at the bottom and the top of the module in between the two glass plates. The measured temperatures and the average temperature are found in Figure 7.2. At first thought, the obvious temperature of the module to apply in the models from a statistical point of view is the average of the bottom and the top temperature. By analysing the two thermal images shown in Figure 7.3, it becomes clear that the measured top temperature represents between 50-90% of the module temperature. The module temperature is lower in the bottom of the module due to the forced ventilation in the air gap behind the module. The calculated average temperature therefore is not in any sense a representative temperature, but more an artificial temperature which is unlikely to present in the module. Unfortunately no thermal images have be taken in the period 16th-18th of August, which is the data the models are based on. Figure 7.3 reveals that the heat distribution

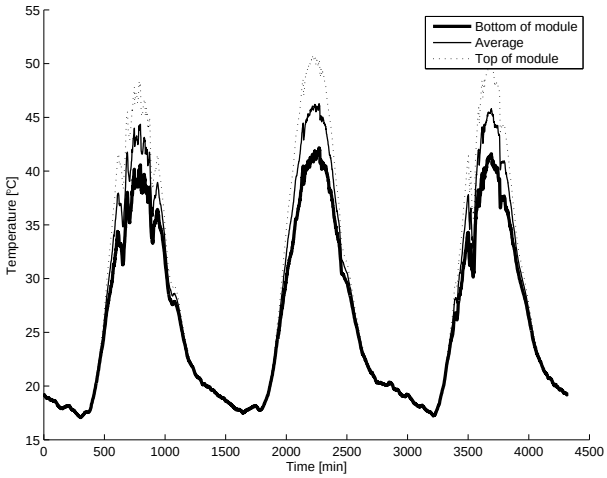

Figure 7.2: Plot of the temperature of the module

of the module varies much.

The thermal images were discovered during the estimation of the models. The introductory analyses will therefore be based on both the average and the top temperature. The intention is to investigate which of the temperatures delivers the best results.

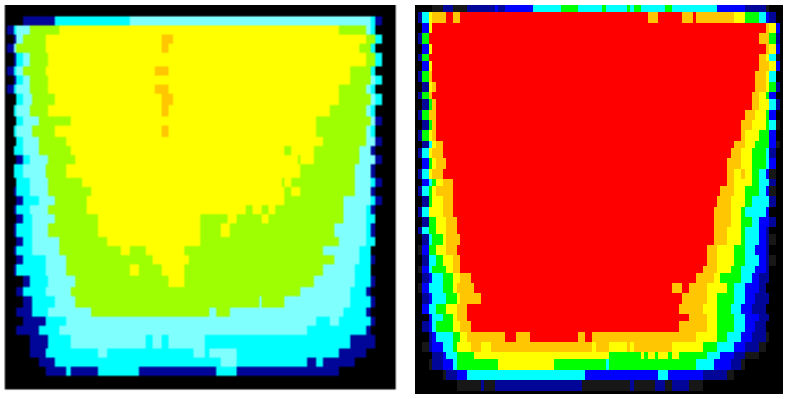

Figure 7.3: Two thermal images of the module from different days

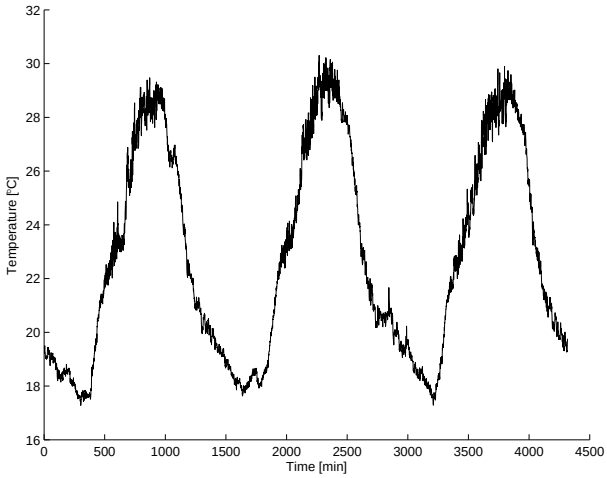

Figure 7.4: Plot of the ambient temperature

7.4 Input variables

All the remaining variables in the analysis are input variables. The use of these variables will make it possible to predict the module temperature. The analyses will help to identify the influence from different variables on the module temperature.

7.4.1 The ambient temperature

It is obvious that the ambient temperature has an influence on the performance of the PV module, since the whole front area of the module is in contact with the ambient temperature. In Figure 7.4 the temperature measurements are shown. Like the past variables there is a clear 24 hours periodicity. The ambient temperature has a low variation compared to the measured wind speed shown in Figure 7.7. Around noon the largest fluctuations are found.
The ambient air temperature is measured by using a ventilated double shielded PT100 sensor It is constructed in such a way that the temperature is measured inside the equipment. This implies that the temperature is not affected by preheating or movements in the air.

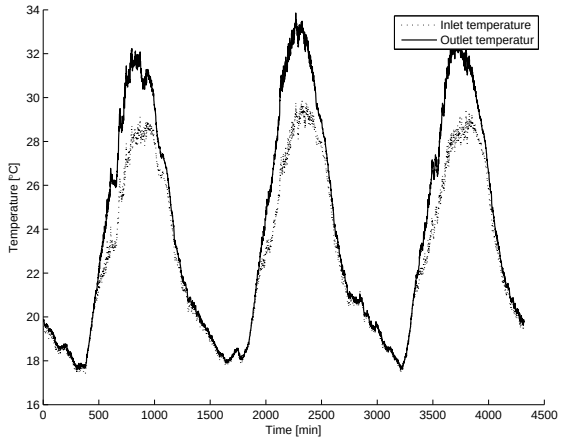

Figure 7.5: Plot of the inlet and outlet temperature

7.4.1.1 The temperature of the air flow behind the module

Between the PV module and the opaque wooden board is a mechanically driven air flow. This entails that the flow can be controlled and measured accurately, since it is possible to get a constant flow in contrast to naturally ventilated systems. The air flow can be set at four different constant velocity levels. Figure 7.5 shows the two temperatures, the in- and outlet temperatures, on which the calculation of ΔT is based. In the data set both a calculated and a measured ΔT is available. The two variables are plotted in Figure 7.6.
To measure the air flow a Swemair 300 instrument is applied.
The air flow behind the PV module cools down the module. The thermal system and thereby the electrical system of the PV module becomes more efficient the higher the forced air flow becomes [Gandini et al. n.d.].

7.4.1.2 The measurement of the wind speed

In Figure 7.7 a plot of the measured wind speed can be seen. The figure reveals that the wind conditions at the site fluctuate. A moving average wind speed with measurement intervals of 20 minutes is present. The reason why this transformation of the wind speed is taken into consideration is that the PV module, in theory, might be more influenced by the more constant wind conditions. A low-pass filtration of the wind speed is also presented in Figure 7.7.

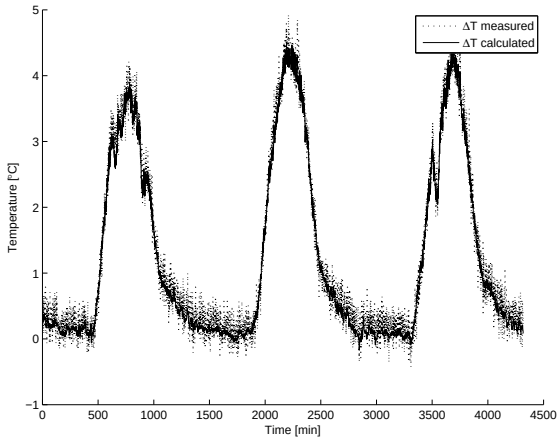

Figure 7.6: Plot of the temperatures of the ventilated air gap behind the module

An anemometer has been used to measure the wind speed. The anemometer was placed on the roof of the pent just above the module. When measuring wind it is important to place the equipment close to the site, since the wind conditions fluctuate over short distances.
One major problematic nature of the test site is that the wind in Ispra is rather limited. The wind speed is seldom above 2 m/s. This can be a limitation of the scope of the models.

7.4.1.3 The irradiance on the PV module

The most important variable when dealing with PV modules is the irradiance. The irradiation is represented in the data set in two versions. A pyranometer is applied to collect the irradiance data. The pyranometer is placed about 20 centimetres from the module. There is a risk that a systematic error is attached to this measurement. The pyranometer is placed on a wooden pillar at a hinge. If the screws are not fastened properly, a twist of the equipment may lead to an incorrectly measured irradiance compared to the irradiance of the module. A more precise way to describe the irradiance is by calculation, since the short circuit current, I_{sc}, of the module over a wide range is proportional to the solar radiation. In the models both measures are applied in order to find the best describing variable. In Figure 7.8 the two variables are plotted.

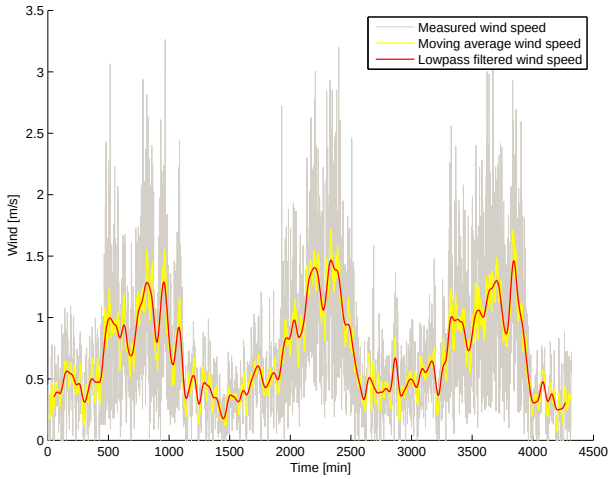

Figure 7.7: Plot of the measured wind speed, a calculated moving average version of the wind and a lowpass filtration of wind speed

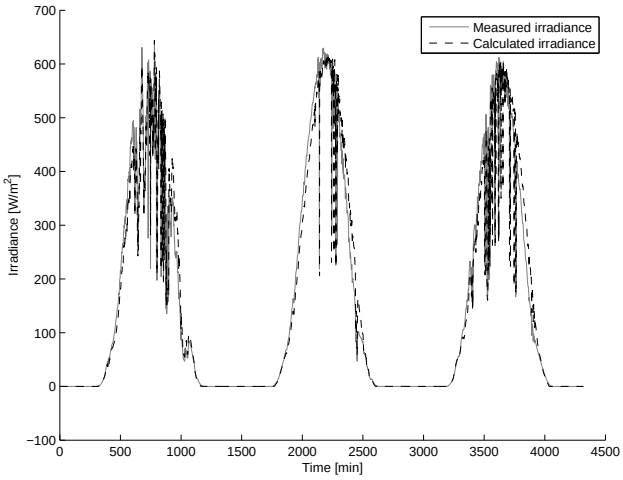

Figure 7.8: The measured and calculated irradiance of the module

Chapter 8

Single State Models - Model identification

8.1 Thermal models

The purpose of this chapter is to identify the best model to predict the module temperature. All the models are constructed in accordance with the three heat transfer principles decribed in Section 2.4. Introductorily a description of the models is outlined. All the models in this chapter are estimated both for a 24-hour period, the 16th of August 2002, and for the three-day period, 16-18th of August 2002. This offers a basis for comparing the influence of the number of observations on the residuals, the uncertainties of the model and the parameter estimates. In the search for a satisfactory model several different models, both linear and non-linear, have been estimated and tested. Subsequently the results of the models will be analysed and compared in order to determine which model or models will suit the data the best. The best model will also be applied in some of the following chapters.

The models applied in this chapter is developed in connection to the article by [Jiménez et al. 2006].

8.2 The applied data

The applied data of this section are gathered at the test site - TRE - during a three day-period, 16th-18th of August 2002, at JRC in Ispra, Italy. The measuring started at midnight the 16th of August. TRE is described in Chapter 7.

The forced ventilation in the air gap was unfortunately changed from velocity level 7 to level 10 at 8.00 a.m. the 16th of August. During the rest of the period the forced velocity was kept at level 10, which is equal to 2.49 m/s. Assessment of the plots of the variables and the prediction of the module temperature reveal no conspicuous deviation due to the change in velocity. On the other hand it is later proven that it is possible for the models to differentiate between velocities of the forced ventilation. The divergences are however most visible in relation to the heat transfer coefficients.

8.3 The design of the models

In order to predict the dynamic temperature of the module it is required to investigate which energy flows, and thereby heat transfers, are having influence on the module. The influence from conduction is insignificant in relation to the BIPV set-up, due to the thickness of the module. For this reason conduction is not included in the models. Equation 8.1 and 8.2 represents the full model, corresponding to the extended non-linear model, including the units of the terms. This is done in order to justify the structure of the models.

$$\overbrace{C_{pv}\frac{dT}{dt}}^{\frac{J}{K}\frac{K}{s}=W} = \overbrace{A\sigma\epsilon_{pv_1}(T_{rad}^4 - T^4)}^{m^2\frac{W}{m^2K^4}K^4=W} + \overbrace{Ah_cW^{k_{airwind}}(T_{air}-T)}^{m^2\frac{W}{m^2K}K=W} + \overbrace{pw_{gap}S\rho c_p\Delta T}^{\frac{m}{s}m^2\frac{kg}{m^3}\frac{J}{kgK}K=W} + \overbrace{\frac{A\sigma}{\frac{1}{\epsilon_{pv_2}}+\frac{1}{\epsilon_w}+1}(T_{wood}^4 - T^4)}^{m^2\frac{W}{m^2K^4}K^4=W} + \overbrace{A\alpha W^{k_{irradwind}}I}^{m^2\frac{W}{m^2}=W} + dw \tag{8.1}$$

$$T_m = T + e \tag{8.2}$$

The terms containing temperatures raised to the fourth power are radiative heat transfer. The remaining terms holding temperature differences are the convective heat transfers. Last terms represent the absorption of the vertical solar radiation affecting the module.

As it can be seen the final unit and all the terms in equation 8.1 are stated

in Watts, the SI-unit for energy. On the left side of the equation one of the unit measures is given as $\frac{K}{s}$, but since the measurements are carried out at one-minute intervals, it is necessary to multiply all the terms on the right side by 60. The model could also be estimated in CTSM by using the time interval in seconds, but this leads to significantly extended estimation time.

8.4 Why use heat transfer models?

According to the Institute of Electrical and Electronics Engineers former standard, IEEE PAR 1479 "Recommended Practice for the Evaluation of Photovoltaic Module Energy Production" it is necessary to know the temperature of the PV module to predict the production of energy from the PV module [TamizhMani et al. 2003]. It is also outlined in [Bloem n.d.*a*]. To support the reasonability of the standard Figure 8.1 shows the relationship between the energy production and the average temperature of the module. The two varables are positively correlated. This plot only shows that the variables vary together, but it does not support the knowledge that the efficiency of the module drops when it becomes warmer. The plot shows that if it is possible to predict the temperature of the module, it is also possible to obtain knowledge about the production of electricity of the module. Not surprisingly both variables follow a 24-hour period. Furthermore it can be seen that the delivered energy varies quite a lot. This is a result of change in the short time variations of the weather conditions, for instance due to clouds passing by the sun. Moreover the plot reveals a slight delay of the average temperature of the module compared to the delivered amount of energy. By using stochastic differential equations this circumstance is taken into consideration.

It is important to be aware of the interconnection between module temperature and produced energy. An increase in the temperature of PV cells results in reduction in the open circuitvoltage, which in turn reduces the power output. In [Jiménez et al. 2006] it is stated that there is a negative temperature effect of about $0.5\%/^oC$ on the maximum power. This shows the necessity of modelling the production of energy based on the temperature of the module, which is also outlined in [Born 2001].

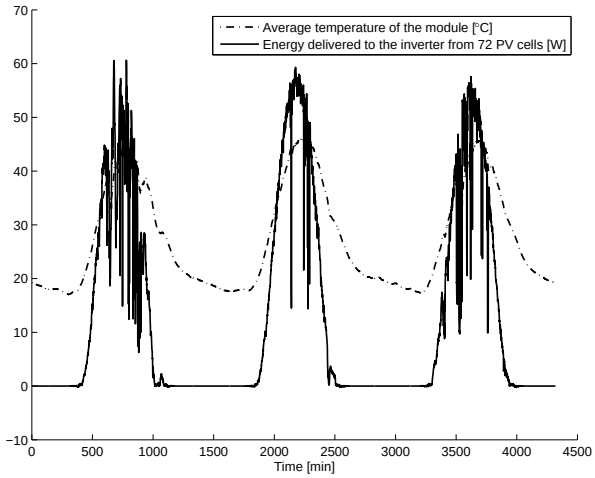

Figure 8.1: Plot of the average temperature of the module and the delivered energy

8.5 The procedure for the modelling and the analysis

To investigate which physical factors are influencing on the temperature of the PV module, different model configurations are applied. There are two main types of models in this analysis, linear and non-linear state space models. It is chosen to estimate one linear model and two non-linear models. Several of the variables, both input and output, are available in more than one version. To minimize the number of possible model combinations it is decided to calculate the possible combinations for the linear models first. This is due to the knowledge that the majority of the variation from the data in the non-linear models is described by the linear terms. In addition the linear models can be estimated faster in CTSM than the non-linear models. Both the number of calculations and the calculation time are more comprehensive in the case of non-linearities. In the applied data set four of the variables are present in several variations; either by measurement, calculation or filtration. The original data set the temperature difference in the ventilated air gap and the irradiance provides both as a measured and a calculated variable. The calculations are made on basis of the electrical flow measurements. The filtration is only relevant in relation to the wind speed. The wind speed is only present in the non-linear models. The

analysis will for this reason contain the possible different combinations of the variables in order to investigate which of the variables represents the temperature of the BIPV module the best. The variables found most suitable in the linear cases will also be applied in the non-linear models. During the analysis phase it has become clear that the first thought as to applying the calculated average temperature of the module as the output variable is probably not the best choice, according to the thermal images in Section 7.3.1. It is therefore decided to estimate all the models using both the average and the measured top temperature as the output temperature. One of the purposes of the chapter is to make an in-depth analysis to find the most representative variables and models to apply in the coming chapters. Above only the temperature in the top of the module is mentioned. In the preliminart analyses it has also been tested if the temperature measured at the bottom of the module could produce better predictions. Both the results and the thermal images confirm that this choice gives the models a worse performance. When interpreting the thermal images a solution for a better estimation could be to apply weigthing on basis of calculations of the area of the module in accordance to the heat distribution in the module registrated at the thermal images. This has also been attempted, by a 80/20 distribution between the top and the bottom temperatures. The analysis of the residuals reveals a slightly poorer result compared to the models where the temperature in the top is applied. Two arguments for not using this approach are that the heat distribution of the module change over time, due to external factors such as the wind and the irradiation and that CTSM does not possess the possiblity of estimating a weighted output variable.

The method to determine how well the output variable of the model fits the data, is an analysis of the values from CTSM expressing the goodness of fit and the residuals of the model.

Each of the estimated models will be analyzed individually. First the values from CTSM explaining the goodness of fit will be evaluated. Subsequently the average and the standard deviation of the residuals of the models will be analyzed. The residuals are calculated on the basis of the one-step prediction error. Eventually the values of the estimated parameters will be compared to each other. The comparison is not only between the different models but also for the models estimated on data for one day and three days.

It is chosen to compare the autocorrelation, the partial autocorrelation, the cumulated periodogram of the residuals for the best linear models, and the best of each of the two diffenrent non-linear models. Furthermore a change in sign and the portmanteau lack-of-fit-test will be carried out. This will be done in separate sections at the end of this chapter.

8.6 The linear models

The first and simplest of the state space models, Equation 8.3 and 8.4, contains the convective influence from the ambient temperature and the convective influence by the temperature difference in the air gap. Furthermore the convective influenced from the temperature of the wooden board and the radiation is included. In the model it is presumed that the insulation behind the wooden board is dense. This implies that the chosen heat balance system is not influenced by the ambient air temperature from the rear side.

$$dT = k_{air}(T_{air} - T)dt + k_{delta}\Delta T dt + k_{wood}(T_{wood} - T)dt + k_i I dt + dw \quad (8.3)$$

$$T_m = T + e \quad (8.4)$$

In order to perform the following analyses it is necessary to investigate whether the estimation in CTSM is adequate. In Table 8.1 and 8.2 it can be seen that the models based on the calculated delta temperature, ΔT_{calc}, and irradiation, I_{calc}, and the top temperature of the module have p-values equal to 0.0000. P-values lower than 0.05 means that the parameter estimates can be considered to be different from zero. For some of the other models the p-values are higher than 0.05 assuming a 95% confidence interval, which means that some of the estimations of the parameters could be equal to zero. This needs to be investigated if it is chosen to further apply such models. If a term is insignificant it can be assumed to be equal to 0. The term can then be removed from the model, and after that the models need to be reestimated. In order to maintain the heat balance system it is chosen to avoid this kind of considerations in relation to the linear models. It is also chosen to include the derivatives of the penalty function in the analysis, since this value gives an indication of whether the parameter is well inside the fixed interval limits. All the derivatives of the penalty function values shown in Table 8.1 are acceptably low, indicating that the parameters are well defined.

The correlations between the parameters are listed in Table 8.2. Prior estimations have shown that correlation values up to around 0.96, do not influence how adequate the model is. It can therefore be concluded that most of the values are satifactorying. The model based on the calculated variables has too high correlation value, 0.9753.

Table 8.1: Summary of the highest p-value and the derivative of the penalty function of the linear models

16th of August				
The model	p-value		dPen.func	
	$T_{moduleavg}$	$T_{moduletop}$	$T_{moduleavg}$	$T_{moduletop}$
$\Delta T_{meas} I_{meas}$	0.0166	0.0551	0.0007	0.0006
$\Delta T_{meas} I_{calc}$	0.0099	0.1677	0.0007	0.0006
$\Delta T_{calc} I_{meas}$	0.1114	0.2911	0.0006	0.0006
$\Delta T_{calc} I_{calc}$	0.0000	0.0000	0.0006	0.0006
$\Delta T_{divided}$	0.2991	0.1297	-0.0158	0.0239
16th-18th of August				
$\Delta T_{meas} I_{meas}$	0.0000	0.0000	0.0007	0.0006
$\Delta T_{meas} I_{calc}$	0.0765	0.3092	0.0007	0.0006
$\Delta T_{calc} I_{meas}$	0.1202	0.0177	0.0007	0.0006
$\Delta T_{calc} I_{calc}$	0.7860	0.0000	0.0005	0.0006
$\Delta T_{divided}$	0.8200	0.8980	-0.2677	0.0494

Table 8.2: Summary of the highest correlations between the parameter estimates of the linear models

16th of August		
The model	Correlation	
	$T_{moduleavg}$	$T_{moduletop}$
$\Delta T_{meas} I_{meas}$	0.8983 ($k_{air} - k_{delta}$)	-0.8496 ($k_{wood} - k_{irrad}$)
$\Delta T_{meas} I_{calc}$	0.9037 ($k_{air} - k_{delta}$)	0.9394 ($k_{air} - k_{wood}$)
$\Delta T_{calc} I_{meas}$	0.9678 ($k_{air} - k_{delta}$)	0.9678 ($k_{air} - k_{irrad}$)
$\Delta T_{calc} I_{calc}$	0.9690 ($k_{air} - k_{delta}$)	0.9753 ($k_{air} - k_{wood}$)
$\Delta T_{divided}$	-0.9878 ($k_{air} - k_{out}$)	-0.9908 ($k_{air} - k_{out}$)
16th-18th of August		
$\Delta T_{meas} I_{meas}$	-0.8666 ($k_{delta} - k_{wood}$)	0.8527 ($k_{air} - k_{wood}$)
$\Delta T_{meas} I_{calc}$	0.8877 ($k_{air} - k_{delta}$)	0.9189 ($k_{air} - k_{wood}$)
$\Delta T_{calc} I_{meas}$	0.9550 ($k_{air} - k_{delta}$)	0.9615 ($k_{air} - k_{wood}$)
$\Delta T_{calc} I_{calc}$	0.9666 ($k_{air} - k_{delta}$)	0.9752 ($k_{air} - k_{wood}$)
$\Delta T_{divided}$	-0.9854 ($k_{air} - k_{out}$)	-0.9922 ($k_{air} - k_{out}$)

In most cases the ΔT and k_{air} are highly correlated, see Table 8.2. ΔT is the difference between the inlet and the outlet temperatures. The inlet and the ambient temperatures must be considered to be correlated in a physical sense, since the inlet air is taken from the ambient air. It is therefore decided to introduce an extention of the linear model, which can eliminate potentially the

problem that the correlation between the inlet and the ambient temperature seems to create. This model represents the same heat balance system, but in the design of this model, named *divided* in the table, it is decided to divide the ΔT-term into an inlet and an outlet temperature. Furthermore T_{in} is set equal to T_{air}. This divided model can be seen in Equation 8.5 and 8.6.

$$dT = k_{air-in}T_{air}dt + k_T T dt + k_{out}T_{out}dt + k_{wood}T_{wood}dt + k_i I dt + dw \quad (8.5)$$

$$T_m = T + e \quad (8.6)$$

Table 8.1 and 8.2 reveals that the divided models have higher p-values, derivatives of the penalty function values and correlations compared to the former model.

Table 8.3: Averages and standard deviations of the residuals for the different models calculated on basis of data from the 16th of August and 16th-18th of August

16th of August				
	$T_{moduleavg}$		$T_{moduletop}$	
The model	Average	Std.dev	Average	Std.dev
$\Delta T_{meas} I_{meas}$	$2.445 \cdot 10^{-4}$	$7.162 \cdot 10^{-2}$	$2.500 \cdot 10^{-3}$	$7.171 \cdot 10^{-2}$
$\Delta T_{meas} I_{calc}$	$-8.251 \cdot 10^{-3}$	$4.645 \cdot 10^{-2}$	$-6.860 \cdot 10^{-3}$	$4.211 \cdot 10^{-2}$
$\Delta T_{calc} I_{meas}$	$3.070 \cdot 10^{-3}$	$7.052 \cdot 10^{-2}$	$6.260 \cdot 10^{-3}$	$6.976 \cdot 10^{-2}$
$\Delta T_{calc} I_{calc}$	$-6.979 \cdot 10^{-3}$	$4.631 \cdot 10^{-2}$	$-5.039 \cdot 10^{-3}$	$4.136 \cdot 10^{-2}$
ΔT_{divid}	$-2.876 \cdot 10^{-5}$	$4.400 \cdot 10^{-2}$	$-2.265 \cdot 10^{-5}$	$4.064 \cdot 10^{-2}$
16th-18th of August				
	$T_{moduleavg}$		$T_{moduletop}$	
The model	Average	Std.dev	Average	Std.dev
$\Delta T_{meas} I_{meas}$	$3.328 \cdot 10^{-3}$	$7.014 \cdot 10^{-2}$	$5.868 \cdot 10^{-3}$	$7.014 \cdot 10^{-2}$
$\Delta T_{meas} I_{calc}$	$-7.609 \cdot 10^{-3}$	$4.800 \cdot 10^{-2}$	$-6.159 \cdot 10^{-3}$	$4.136. \cdot 10^{-2}$
$\Delta T_{calc} I_{meas}$	$5.336 \cdot 10^{-3}$	$6.787 \cdot 10^{-2}$	$7.839 \cdot 10^{-3}$	$6.740 \cdot 10^{-2}$
$\Delta T_{calc} I_{calc}$	$-2.337 \cdot 10^{-4}$	$4.617 \cdot 10^{-2}$	$-5.603 \cdot 10^{-3}$	$4.059 \cdot 10^{-2}$
ΔT_{divid}	$-3.462 \cdot 10^{-4}$	$4.643 \cdot 10^{-2}$	$-1.567 \cdot 10^{-4}$	$4.047 \cdot 10^{-2}$

In order to determine which of the models describes the data the best, the residuals have to be analyzed. The averages and standard deviations of residuals for the linear models are presented in Table 8.3. By comparing the averages and the standard deviations of the residuals the different models the following trends are identified:

- In all cases the standard deviation is lower when the calculated irradiance, I_{calc}, is applied instead of the measured irradiance

- The calculated temperature difference, ΔT_{calc} gives the lowest standard deviations for all models
- In the majority of the cases the use of the temperature measured at the top of the module results in lower standard deviations, which indicates a better fitting model.
- The divided model has the lowest standard deviations for the three-day period, but not for the one-day estimations. Since all the CTSM estimated values are higher for this more complicated model, it is decided not to take this model into account.
- There is a small reduction in the standard deviation values when the models are based in three-day data compared to the one-day data.

When taking all the facts above into account, the models based on the calculated variables show through. On the basis of these observations the non-linear models in next section will only be estimated applying the calculated variables of the temperature difference in the air gap and the irradiation. From a physicist's point of view it does make sense that the calculated variables generate better results. The calculated variables are based on data collected inside the module via knowledge about the electrical flows. This way of estimating the variables minimizes the amount of noise in the variables and thereby the estimated models.

In general for the results of the linear models it can be concluded that the correlations of the parameter estimates need to be reduced in order to find an adequate model. Therefore a natural progress is to extend the linear model to a non-linear model.

8.7 Introduction to the non-linear models

As mentioned earlier the wind is the only new variable in the non-linear models. The non-linear models, a simple and an extended one, in this section are extensions where the influence of the wind speed is added to the linear model design. The original data set contains both a measured wind speed and a moving average version of the measured wind speed. In the analysis phase different filtrations are applied to the wind speed variable. This is due to prior knowledge indicating that the high frequencies of the wind do not influence the performance of the BIPV. A filtration can also be advantageous in situations where the place of the measuring is not located next to the module. In this case the anemometer is placed at the roof above the test site. The local variation in the wind speed can be relatively fluctuating, but the average or filtered wind speed is more constant

in a bounded area.

Like the linear model, the variations of the non-filtered and filtered wind speed will be tested in order to find the most suitable model for the BIPV module.

8.8 Simple non-linear model

Equations 8.7 and 8.8 represent a state space model for the first non-linear model. As distinct from the linear model the influence from the wind speed raised to an undetermined power, which is estimated along with the other parameters.

$$\begin{aligned} dT &= k_{air}W^{k}_{airwind}(T_{air}-T)dt + k_{delta}\Delta T dt + \dots \\ &\quad k_{wood}(T_{wood}-T)dt + k_i I dt + dw \end{aligned} \tag{8.7}$$

$$T_m = T + e \tag{8.8}$$

In the case of both the simple and the extended non-linear model it is attempted to make CTSM estimate the optimal filter of the wind. The optimal filter is a causal filter, which estimates how much of the variation has to be removed in order to get the best description of data. The causal filter is incorporated in the model by adding an extra state space equation. The system is outlined in equation 8.9-8.11. dw_2 is put equal to 0, since there should be no noise attached to the filtered wind.

$$\begin{aligned} dT &= k_{air}W^{k_{airwind}}_{filter}(T_{air}-T)dt + k_{delta}\Delta T dt + \dots \\ &\quad k_{wood}(T_{wood}-T)dt + k_i I dt + dw_1 \end{aligned} \tag{8.9}$$

$$dW_{filter} = aW_{filter} - aW + dw_2 \tag{8.10}$$

$$T_m = T + e \tag{8.11}$$

First it is necessary to investigate the value from CTSM indicating the adequateness of the models. Table 8.4 shows that all the p-values of the model based on the temperature at the top of the module are significant. All the derivatives of the penalty function values are acceptably low. A common feature for the correlation coefficients, Table 8.5, is that they are still remarkably high, since most of the correlation values are above 0.96. Furthermore it can be seen that the highest correlations are between k_{air} and k_{delta}, just as was the case in the linear models. The high correlations are an indication that the model is not adequate. But in order to map the development from the linear model to

Table 8.4: Summary of the p-value and the derivative of the penalty function for the simple non-linear model

16th of August				
The model	p-value		dPen.func	
	$T_{moduleavg}$	$T_{moduletop}$	$T_{moduleavg}$	$T_{moduletop}$
Wind	0.0047	0.0007	0.0016	0.0013
Causal wind	0.0000	0.0000	0.0068	0.0046
Lowpass wind	0.0000	0.0000	0.0112	0.0095
MA wind	0.5122	0.0000	0.0100	0.0089
16th-18th of August				
Wind	0.0000	0.0000	0.0004	0.0004
Causal wind	0.0237	0.0000	0.0432	0.0018
Lowpass wind	0.0041	0.0000	0.0016	0.0004
MA wind	0.0000	0.0000	0.0154	0.0004

the simple non-linear model, it is decided to analyze the best of these simple non-linear models in the section of results. Table 8.6 reveals that the standard deviations are the lowest for the causal lowpass filter of the wind for both module temperature measurements. The second lowest standard deviations refer to the models where the measured wind is applied. There are no patterns in the values of the correlations of the parameters in order to determine which of the two output variable, $T_{moduleavg}$ and $T_{moduletop}$, perform the best. It is the same situation concerning the number of observations on which the models are based on. Table 8.6 clearly shows that the standard deviation of the models having the temperature at the top of the module as an output variable are lower than the model estimating the average temperature of the module. Regarding the one day causal filter model the standard deviation drops from $4.389 \cdot 10^{-2}$ to $3.803 \cdot 10^{-2}$. It is similar for the three-day data model. In order to investigate if there is any relation between the measured wind and the causal lowpass filteration of the wind the model prediction of these two are plotted in Figure 8.2. The plot shows how close CTSM estimates the causal wind, represented by the grey line, to the measured wind, represented by the black line. This means that the causal filter removes only very little of the flutuation of the measured wind. Figure 8.2 gives a clear indication that the BIPV module is actually sensitive to fast changes in the wind speed. This could easely be different at locations where higher speed and fluctuations of the wind occur.

Table 8.5: Summary of the correlations for the parameter estimates for the simple non-linear model

16th of August		
The model	Correlation	
	$T_{moduleavg}$	$T_{moduletop}$
Wind	0.9893 ($k_{air} - k_{delta}$)	0.9662 ($k_{air} - k_{delta}$)
Causal wind	-0.8300 ($W_{filter} - k_{wood}$)	0.9051 ($k_{air} - k_{wood}$)
Lowpass wind	0.9782 ($k_{air} - k_{delta}$)	0.9862 ($k_{air} - k_{delta}$)
MA wind	0.9756 ($k_{air} - k_{delta}$)	0.9825 ($k_{air} - k_{delta}$)
16th-18th of August		
Wind	0.9712 ($k_{air} - k_{delta}$)	0.9754 ($k_{air} - k_{delta}$)
Causal wind	-0.9866 ($W_{filter} - a$)	-0.9486 ($k_{delta} - k_{wood}$)
Lowpass wind	0.9689 ($k_{air} - k_{delta}$)	0.9793 ($k_{air} - k_{delta}$)
MA wind	0.9706 ($k_{air} - k_{delta}$)	0.9787 ($k_{air} - k_{delta}$)

Table 8.6: Averages and standard deviations of the residuals for the different simple non-linear models calculated on basis of data from the 16th of August and 16th-18th of August. In the table both the results for average and top module temperature as output variables are displayed

16th of August				
	$T_{moduleavg}$		$T_{moduletop}$	
The model	Average	Std.dev	Average	Std.dev
Wind	$-6.819 \cdot 10^{-3}$	$4.482 \cdot 10^{-2}$	$-4.546 \cdot 10^{-3}$	$3.932 \cdot 10^{-2}$
Causal wind	$-7.540 \cdot 10^{-3}$	$4.389 \cdot 10^{-2}$	$-5.774 \cdot 10^{-3}$	$3.803 \cdot 10^{-2}$
Lowpass wind	$-7.849 \cdot 10^{-3}$	$4.589 \cdot 10^{-2}$	$-6.355 \cdot 10^{-3}$	$4.043 \cdot 10^{-2}$
MA wind	$-7.806 \cdot 10^{-3}$	$4.598 \cdot 10^{-2}$	$-6.264 \cdot 10^{-3}$	$4.044 \cdot 10^{-2}$
16th-18th of August				
	$T_{moduleavg}$		$T_{moduletop}$	
The model	Average	Std.dev	Average	Std.dev
Wind	$-7.640 \cdot 10^{-3}$	$4.652 \cdot 10^{-2}$	$-6.433 \cdot 10^{-3}$	$3.842 \cdot 10^{-2}$
Causal wind	$-7.838 \cdot 10^{-3}$	$4.611 \cdot 10^{-2}$	$-6.878 \cdot 10^{-3}$	$3.748 \cdot 10^{-2}$
Lowpass wind	$-7.557 \cdot 10^{-3}$	$4.784 \cdot 10^{-2}$	$-6.970 \cdot 10^{-3}$	$4.017 \cdot 10^{-2}$
MA wind	$-7.817 \cdot 10^{-3}$	$4.784 \cdot 10^{-2}$	$-7.143 \cdot 10^{-3}$	$3.985 \cdot 10^{-2}$

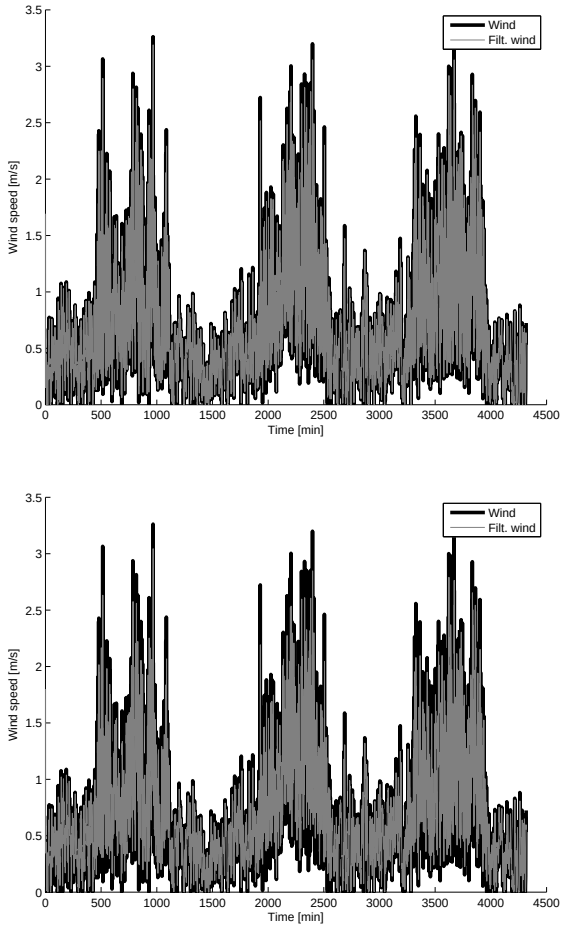

Figure 8.2: The measured and optimal filtered wind in the three-day period. The model used for prediction is based on the $T_{moduletop}$. It can be seen how similar the measure wind and the causal filter are predicted.

8.9 Extended non-linear model

Since high correlations were found for the simple non-linear state space model, it is determined to extend the model further. The state space model is outlined in Equation 8.12 and 8.13. Now it is presumed that the wind speed also has influence on the irradiation. Again the wind speed is raised to an unknown power which is estimated. The long wave radiation to the surrounding is considered to be relative to a mean radiant temperature, T_{rad}, which is estimated as an unknown parameter in CTSM. Moreover the heat transfer between the wooden board and the PV module is considered a radiative effect instead of a convective as in the two previous models. Physically this influence is radiative, so this term in the two previous models can be viewed as a simplification. Due to the temperatures raised to the fourth power it is necessary to use absolute temperatures in the modelling.

$$\begin{aligned} dT = {} & k_{rad}(T_{rad}^4 - T^4)dt + k_{air}W^{k_{windair}}(T_{air} - T)dt + k_{delta}\Delta T dt + \dots \\ & k_{wood}(T_{wood}^4 - T^4)dt + k_i W^{k_{windirrad}} I dt + dw \end{aligned} \tag{8.12}$$

$$T_m = T + e \tag{8.13}$$

Two notable observations are made after estimating the models on the basis of the model described above. For the main part of the models the parameter estimates of the influence from the wind on the irradiation, $k_{windirrad}$, and the temperature difference in the air gab behind the module, k_{delta}, are not significant, see Table 8.7. Looking at the models from a physical point of view the most reasonable is to remove the influence of the wind on the irradiation that cancels the influence in the models from the temperature difference in the gap behind the module, since this is one of the main ideas behind the testing. When reviewing Table 8.7 it is noted that $k_{windirrad}$ also has high p-values. It shall be added that k_{delta} is significant in the simple non-linear models. The insignificance of $k_{windirrad}$ means that the applied irradiation measure is not affected by the wind. This also makes sence since the irradiation is calculated on basis of measurement inside the module, and therefore not affected by the ambient wind. Secondly it has become clear that the fluctuation in the wind leads to very unstable parameter estimates of the wind in the models. To compensate for this uncertainty it is determined to replace the estimation of the parameter of for the wind with a fixed parameter equal to the average wind. The p-values for the parameters are correlated, which means that fixing one parameter can lead to change of the p-values in the reestimation. This can also be seen in Table 8.7 where the p-value of k_{delta} has changed in values. Some of the p-values of k_{delta} are still insignificant, but it has to be kept in mind that k_{delta} is significant for the models based on the measured and the causal wind. These two variants of the wind were analyzed as the best in the case of the

Table 8.7: P-value of the non significant parameters of the extended non-linear model with $T_{moduletop}$ as output variable

	$k_{windirrad}$	k_{delta} before removing $k_{windirrad}$	k_{delta} after removing $k_{windirrad}$
16th of August			
Wind	0.0000	0.0000	0.0000
Causal wind	0.1875	0.0156	0.0132
Lowpass wind	0.2749	0.8160	0.5484
MA wind	0.9880	0.9275	0.8018
16th-18th of August			
Wind	0.0000	0.0000	0.0000
Causal wind	0.2951	0.0000	0.0000
Lowpass wind	0.8449	0.1456	0.2664
MA wind	0.8080	0.0204	0.0286

simple non-linear model. On the basis of the findings above all the models have been reestimated with $k_{windirrad}$ is equal to 0 and W_f fixed to the average of the measured wind. All the results below derive from the reestimation. Table 8.8 shows that most of the models have rather high p-values. Some of the correlations, Tabel 8.9, are still too high, when keeping in mind that the value 0.96 can be used as a rule of thumb. In relation to the models based on either the measured wind or the causal filtered wind all the correlations are below 0.93. The averages and the standard deviations of the model are stated in Table 8.10, where the $k_{windirrad}$ is fixed at the value zero, which results in no influence from the wind on the irradiation. From the table it can be seen, as for the simple non-linear model, that the lowest standard deviations are to be found in the model where the measured wind and the causal filter is applied. It can also be seen that the standard deviation of the residuals are lower when the temperature at the top of the module is applied. No further analyses will be made on basis of the models where the average temperature of the module has been applied. When comparing the standard deviations of the residuals shown in Table 8.10 where the temperature at the top of the module is applied as the output variable, it is clear that the measured wind speed or the causal filter of the measured wind have the lowest values. This means that the temperature and the efficiency of the module are sensitive to the fluctuations in the wind. It shall be taken into account that the test site is rarely affected by high wind speed. For this reason the conclusions cannot be transferred directly to other locations where higher speed and fluctuations are present.

Table 8.8: Summary of the p-value and the derivative of the penalty function for the extended non-linear model

16th of August				
The model	p-value		dPen.func	
	$T_{moduleavg}$	$T_{moduletop}$	$T_{moduleavg}$	$T_{moduletop}$
Wind	0.0000	0.0000	0.0536	0.5441
Causal wind	0.0000	0.0132	0.0068	0.2466
Lowpass wind	0.0043	0.5484	0.0616	0.1996
MA wind	0.2431	0.8018	0.434	0.1939
16th-18th of August				
Wind	0.0018	0.0000	0.0404	0.2566
Causal wind	0.0037	0.0000	0.1790	0.1744
Lowpass wind	0.9577	0.2644	0.1518	0.1758
MA wind	0.7015	0.0286	0.4592	0.1345

Table 8.9: Summary of the correlations of the parameter estimates for the extended non-linear model

16th of August		
The model	Correlation	
	$T_{moduleavg}$	$T_{moduletop}$
Wind	0.9730 ($k_{air} - k_{delta}$)	-0.9000 ($k_{air} - k_{delta}$)
Causal wind	-0.8300 ($W_{filter} - k_{delta}$)	-0.8899 ($k_{air} - k_{delta}$)
Lowpass wind	0.9183 ($T_{rad} - k_{windair}$)	-0.9068 ($k_{air} - k_{delta}$)
MA wind	0.9726 ($k_{windair} - k_{windirrad}$)	0.9092 ($k_{rad} - T_{rad}$)
16th-18th of August		
Wind	0.9325 ($k_{windair} - k_{windirrad}$)	0.8991 ($k_{rad} - T_{rad}$)
Causal wind	0.9070 ($k_{rad} - T_{rad}$)	0.9264 ($k_{rad} - T_{rad}$)
Lowpass wind	0.9006 ($k_{air} - k_{delta}$)	0.0.9783 ($k_{rad} - T_{rad}$)
MA wind	0.9317 ($k_{rad} - T_{rad}$)	0.8989 ($k_{air} - k_{delta}$)

Table 8.10: Averages and standard deviations of the residuals for the extended models calculated on basis of data from the 16th of August and 16th-18th of August

16th of August				
	$T_{moduleavg}$		$T_{moduletop}$	
The model	Average	Std.dev	Average	Std.dev
Wind	$-9.128 \cdot 10^{-5}$	$4.076 \cdot 10^{-2}$	$-1.329 \cdot 10^{-4}$	$3.485 \cdot 10^{-2}$
Causal wind	$-1.933 \cdot 10^{-4}$	$4.096 \cdot 10^{-2}$	$-3.032 \cdot 10^{-5}$	$3.583 \cdot 10^{-2}$
Lowpass wind	$2.606 \cdot 10^{-5}$	$4.243 \cdot 10^{-2}$	$8.371 \cdot 10^{-5}$	$3.782 \cdot 10^{-2}$
MA wind	$-3.094 \cdot 10^{-5}$	$4.261 \cdot 10^{-2}$	$-1.253 \cdot 10^{-5}$	$3.805 \cdot 10^{-2}$
16th-18th of August				
	$T_{moduleavg}$		$T_{moduletop}$	
The model	Average	Std.dev	Average	Std.dev
Wind	$-4.181 \cdot 10^{-4}$	$4.417 \cdot 10^{-2}$	$-1.514 \cdot 10^{-4}$	$3.656 \cdot 10^{-2}$
Causal wind	$-3.341 \cdot 10^{-4}$	$4.393 \cdot 10^{-2}$	$-4.509 \cdot 10^{-5}$	$3.540 \cdot 10^{-2}$
Lowpass wind	$-2.599 \cdot 10^{-4}$	$4.528 \cdot 10^{-2}$	$-5.605 \cdot 10^{-5}$	$3.783 \cdot 10^{-2}$
MA wind	$-3.270 \cdot 10^{-4}$	$4.529 \cdot 10^{-2}$	$-7.904 \cdot 10^{-5}$	$3.763 \cdot 10^{-2}$

8.10 Analysis of the parameter estimates

All the preceding analyses have been made in order to determine the adequacy of the models. The models are estimated for both one day data and three day data. When comparing the standard deviations for a model based on one-day data and three-day data respectively no clear picture of the values appear. In general the standard deviations of the models based on three-day data are slightly lower. The data collected at different days vary. The objective of this section is to investigate the stability of the parameter estimates in the sense of the variation in the parameter estimates. Parameters of two different one-day models and one model based on three day observations. In Table 8.11 parameter estimates of two one day models and the three-day model are listed.

In order to transform the value of a from continuous to discrete time the exponential function, $e^{a\tau}$ is estimated. τ is equal to 1 since the time interval is in minutes.

In Table 8.11 the parameters of the extended non-linear model are found. For the purpose of comparison two one-day data based models are estimated. Moreover, the parameter estimates of the model based on three-day data are also represented in the table. Across the models the parameter estimates are raised to the same power. The comparison of the models shows that the model based on data from the 16^{th} of August and model based on the three-day period are most

Table 8.11: Parameterestimates and the associated standard deviations in brackets for the extended non-linear model where the causal filter is applied

Parameter	16^{th} of August	17^{th} of August	$16-18^{th}$ of August
$\hat{k}_{trad}$	$2.539 \cdot 10^{-9}$ $(3.002 \cdot 10^{-10})$	$1.514 \cdot 10^{-9}$ $(3.376 \cdot 10^{-10})$	$2.114 \cdot 10^{-9}$ $(1.819 \cdot 10^{-10})$
$\hat{T}_{rad}$	284.96 $(8.676 \cdot 10^{-1})$	285.96 $(1.464 \cdot 10^{0})$	284.40 $(7.013 \cdot 10^{-1})$
$\hat{k}_{air}$	$1.869 \cdot 10^{0}$ $(1.714 \cdot 10^{-1})$	$1.474 \cdot 10^{0}$ $(1.389 \cdot 10^{-1})$	$1.923 \cdot 10^{0}$ $(1.596 \cdot 10^{-1})$
$\hat{k}_{windair}$	$1.345 \cdot 10^{-1}$ $(1.309 \cdot 10^{-2})$	$1.685 \cdot 10^{-1}$ $(1.576 \cdot 10^{-2})$	$1.383 \cdot 10^{-2}$ $(8.283 \cdot 10^{-3})$
$\hat{k}_{delta}$	$-1.387 \cdot 10^{0}$ $(5.594 \cdot 10^{-1})$	$-2.733 \cdot 10^{0}$ $(5.069 \cdot 10^{-1})$	$-1.545 \cdot 10^{0}$ $(3.188 \cdot 10^{-1})$
$\hat{k}_{wood}$	$4.844 \cdot 10^{-9}$ $(5.330 \cdot 10^{-10})$	$2.515 \cdot 10^{-9}$ $(4.799 \cdot 10^{-10})$	$4.074 \cdot 10^{-9}$ $(3.246 \cdot 10^{-10})$
$\hat{k}_{irradiance}$	$1.213 \cdot 10^{-1}$ $(1.186 \cdot 10^{-3})$	$1.011 \cdot 10^{-1}$ $(2.299 \cdot 10^{-3})$	$1.191 \cdot 10^{-1}$ $(8.574 \cdot 10^{-4})$
$\hat{a}$	0.242 $(2.613 \cdot 10^{-1})$	0.084 $(5.593 \cdot 10^{-1})$	0.353 $(1.331 \cdot 10^{-1})$

similar, whereas the model based on data from the 17^{th} of August in most cases has lower parameter estimates. Futhermore, is can be seen from the standard deviation of the parameters that the parameters are quite stable. The focus of this thesis is not to evaluate the size of the parameter estimates. In general it can be concluded that in broad outline the estimated models agree, but when looking into the specific values of the parameter estimate, it can be seen that the models are sensetive to the input and output variables.

8.11 Analysis of the residuals

In accordance with the preceding sections it is decided to further analyze the residuals of the best of each of three types of models. The basis of decision is the calculated standard deviations of the residuals. It is decided to analyze only the residuals of the models where the output variable is the temperature measured at the top of the module. This decision is made based on the knowledge that the average temperature applied does not in any way represent the temperature in the module and the fact that the different analyses carried out previously in this chapter gave better results. In the linear case the model containing the

calculated temperature difference in the air gap and the calculated irradiance gave the best results. For both of the non-linear models it is decided to carry out further analyses on the models where the causal filtered wind is applied, in order to get the most steady results.

In the following sections the autocorrelation function, the partial autocorrelation function and cumulated periodogram of the residuals for the models will be plotted and analyzed. Furthermore, the results of the change in sign test and Portmanteau lack-of-fit-test will be evaluated.

When estimating models it is always an advantage to apply as many data as possible, since this provides a more stable description of the data. One major problem about having 1439 and 4320 observations, as in the case of this analysis, is that it is difficult to test the residual. It is well-known that the more data applied, the narrower the confidence limits become. This entails that it is difficult to reject or accept the trend plots and values shown. The cut-off between rejecting and accepting becomes very sharp.

8.12 Comparison of the models

The average and the standard deviation of the residuals of three most adequate models found in the previous sections are stated in Table 8.12. The following trends in the standard deviations are noted:

- There is a clear pattern that the standard deviation of the residuals becomes smaller the more extended the model becomes.
- The comparison of the average temperature and the top temperature as output variables respectively reveals that the temperature in the top of the module gives the model the best performance.
- The models based on three day data have lower standard deviations of the residauls compared to the one-day data models.

These trends seem very reasonable, since model extensions and more data most frequently provide better results.

Due to the statement saying that an extension of a model will lead to a higher degree of description, it can be of benefit to perform a likelihood ratio test. This test may reveal whether the improvement is significant. The basis for carrying out the test is that the reduced model is a submodel of the full model. Held up

Table 8.12: Summary of the characteristics of the best models

16th of August		
The model	Average	Std.dev
Linear $\Delta T_{calc} I_{calc}$	$-5.039 \cdot 10^{-3}$	$4.136 \cdot 10^{-2}$
Simple non-linear	$-5.774 \cdot 10^{-3}$	$3.803 \cdot 10^{-2}$
Extended non-linear	$-3.031 \cdot 10^{-5}$	$3.585 \cdot 10^{-2}$
16th-18th of August		
The model	Average	Std.dev
Linear $\Delta T_{calc} I_{calc}$	$-5.603 \cdot 10^{-3}$	$4.059 \cdot 10^{-2}$
Simple non-linear	$-6.878 \cdot 10^{-3}$	$3.748 \cdot 10^{-2}$
Extended non-linear	$-4.509 \cdot 10^{-5}$	$3.540 \cdot 10^{-2}$

against the three models in this thesis, it is only possible to test the linear model against the simple non-linear model. The two latter models are not submodels of the extended non-linear model, due to the change from a convective heat transfer to a radiative heat transfer from the wood board behind the module. The analysis is carried out in section 8.17.

8.13 Plot of residuals

The plot of residuals generally show a the lack of ability to predict the module temperature in the daytime. This is problematic, since it is in these hours the PV module produce electricity, and it is therefore in these hours it is most important to be able the predict the temperature of the module. The residual plot can be found in Figure 8.3. A closer look at the plots reveals that the size of the highest residual values decreases along with the model extension. The plot of the three-day period models shows that the models have the greatest difficulties predicting the first day compared to especially the second day. This is an indication that the model prediction is sensitive to the weather and the operating conditions the specific days.

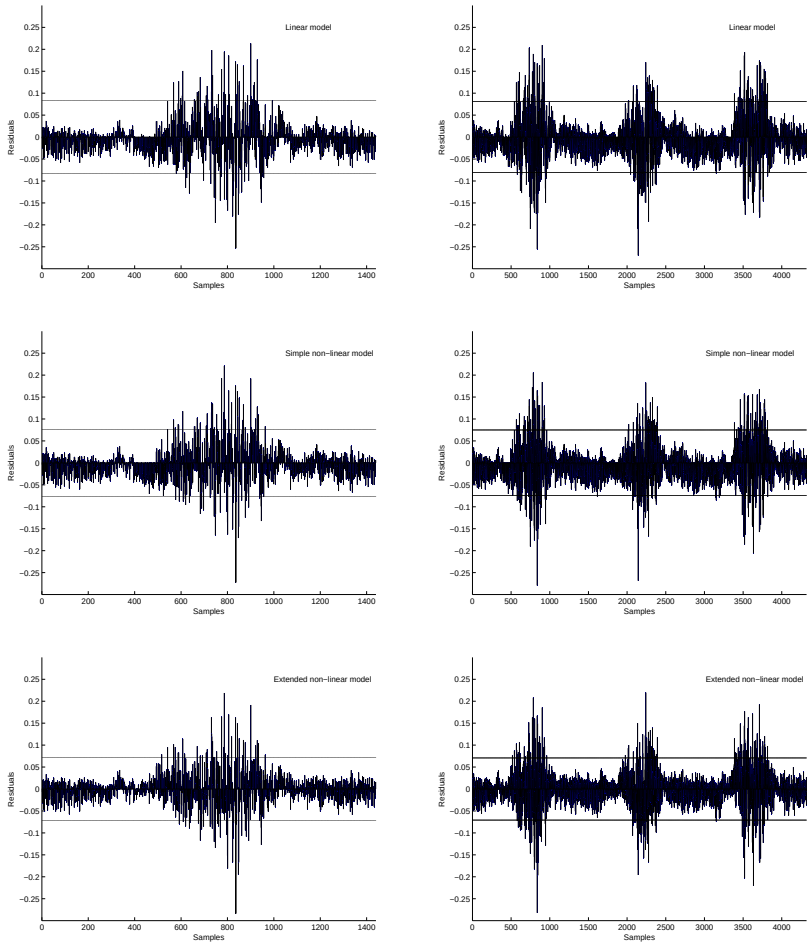

Figure 8.3: The plots of the residual resulting from the models. From the top the linear model and in the bottom the extended non-linear model. The plots at the left-hand side are one-day data and the right-hand side the models based on three-day data.

8.14 ACF and PACF

The autocorrelation function, ACF, and partial autocorrelation function, PACF, plots may help determine two issues. The plots can show if the model can be regarded adequate. Moreover the plots can reveal what to do if the model has a lack of fit. As mentioned the 95% confidence interval in the plot may not be taken too seriously due to the large number of observations. The plots of the autocorrelation function and the partial autocorrelation are respectively found in 8.4 and 8.5. The autocorrelation function plot of the residuals of the one-day data models seems to be more random than the similar models based on three-day data. A common trend of the autocorrelation function plot is that the plot of the extended non-linear model has the smallest and most random correlation values. This supports the trends seen from the standard deviations. The partial autocorrelation plot is favourable with a view to detecting if the model order is adequate. Significant sticks in first and second lag indicate that it can be an advantage to add an extra state in the model. Model extensions and additional terms in the models can also remove significant sticks from the partial correlation function plot. This is the case from the simple to the extended non-linear model. The plots of the extended non-linear model seems to be acceptable random since only the stick in third lag is significant. This can be a motivation for extending the model even further as it will be done in the next chapter, where an extra state is added.

8.15 Periodograms

The only test carried out in the frequency domain is the cumulative periodogram, which may reveal if the residuals are over-represented in specific areas of the frequency domain. A straight line in the periodogram represents white noise and the residuals are then evenly spread. The two periodograms present one-day and three-day models respectively. The periodograms strengthens the picture drawn by the previous analyses. This implies that the residuals of the extended non-linear models form the straightest lines. Though the confidence interval are very tight, the residuals of the extended non-linear model almost manage to stay inside the interval.

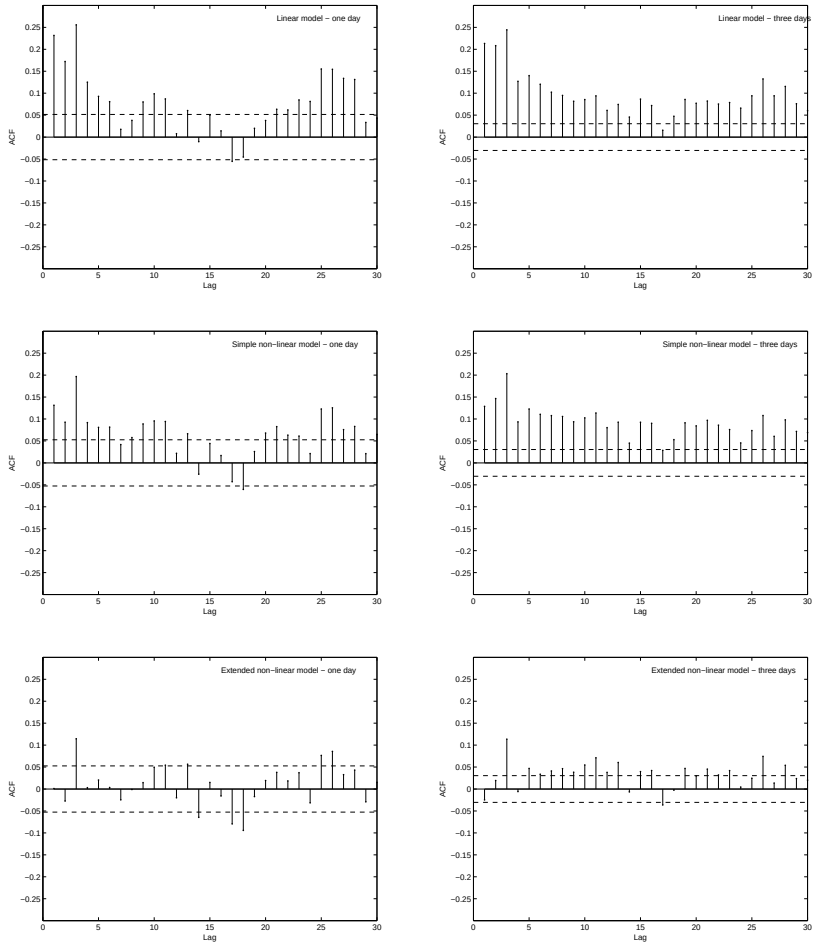

Figure 8.4: Autocorrelation function plot of the residuals of the models. From the top the linear model and at the bottom the extended non-linear model. The plots at the left-hand side are one-day data and the right-hand side the models based on three-day data.

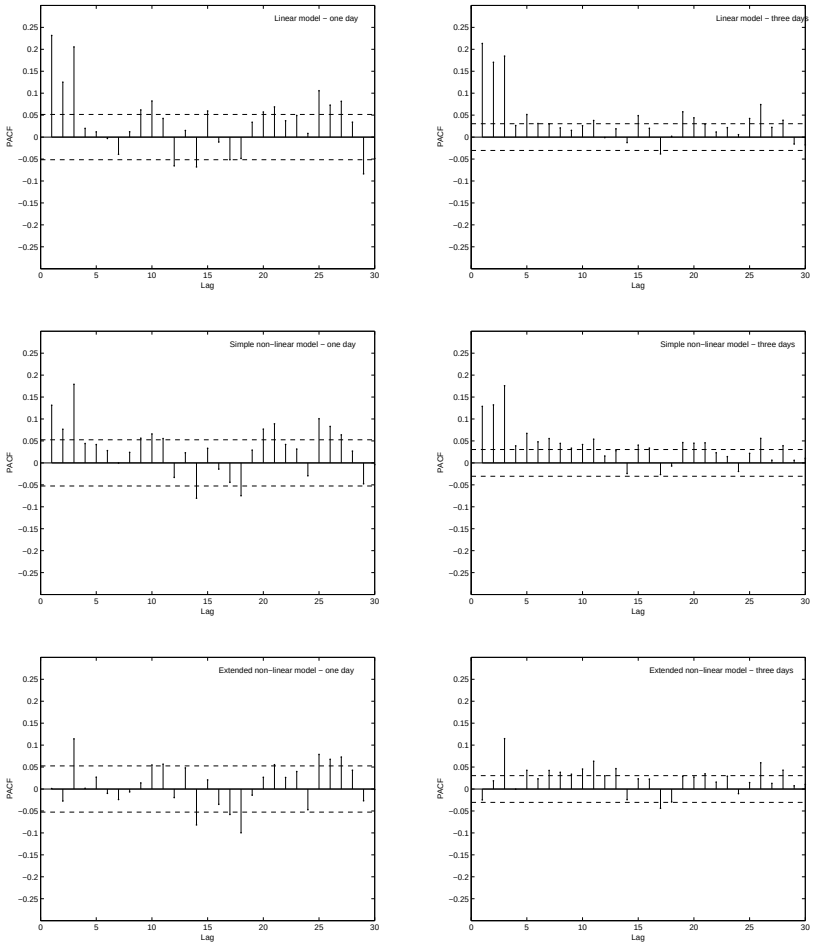

Figure 8.5: Partial autocorrelation function plot of the residuals of the models. From the top the linear model and at the bottom the extended non-linear model. The plots at the left-hand side are one-day data and the right-hand side the models based on three-day data.

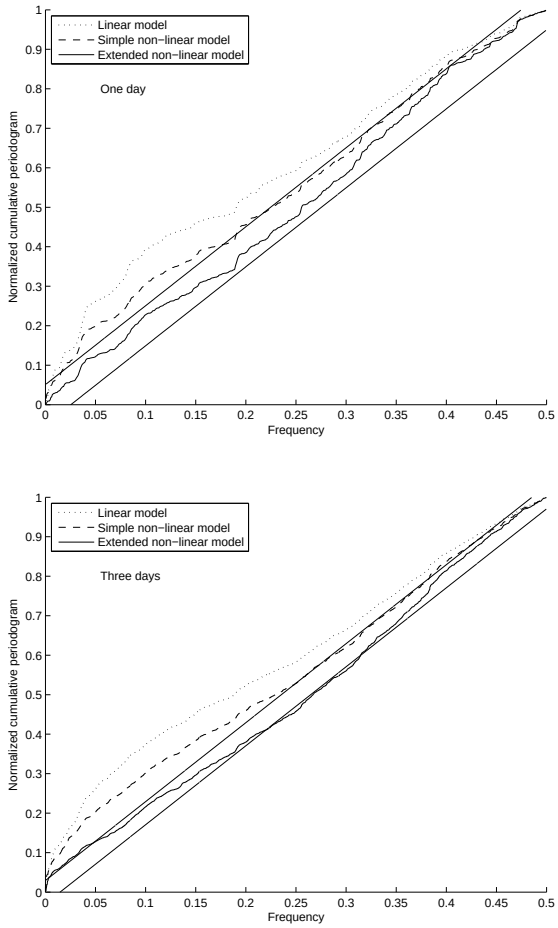

Figure 8.6: This figure shows the cumulated periodograms of the residuals of the models. The plot in the top of the figure is for the models based on one-day observations and the other is for the three-day models

8.16 Tests of the residuals

8.16.1 Change in sign test and Portmanteau lack-of-fit-test

A change in sign test has been carried out in order to support whether the residuals can be assumed to have mean zero. The only way the residuals are able to change sign approximation every second is by having mean zero. The theory is described in Section 4.2.3.1. The number of changes in sign is normally distributed, N($\frac{N-1}{2}$,$\frac{N-1}{4}$) due to the large number of observations. The confidence intervals for the models of one, N=1439, and three days, N=4320, respectively are:

$$N\Big(\frac{1439-1}{2},\frac{1439-1}{4}\Big) = N(719,356.5), \sigma = 18.96 \Rightarrow [700,738]$$

$$N\Big(\frac{4320-1}{2},\frac{4320-1}{4}\Big) = N(2159.5,1079.8), \sigma = 32.86 \Rightarrow [2127,2192]$$

The results are presented in Table 8.13. It can be concluded that the change in sign test of both the linear and the simple non-linear model turns out to be insignificant, meaning that the assumption above is not fulfilled. The extended non-linear model passes the test for both the model based on one-day and three-day data. The change in sign test has also been carried out on the residuals of the extended non-linear models based on the average module temperature. The test showed a tendency of the residuals to have too many changes in sign, 766 and 2237 respectively. This is yet an indication that the average temperature of the module is not representative.

The results of the Portmanteau lack-of-fit-test are not very useful, due to the large number of observations upon which the models are estimated. For this reason it is chosen to have the sum of the autocorrelation function values squared and the Q-value in the table. The Portmanteau lack-of-fit-test is χ^2-distributed with the degrees of freedom (m-n). m is the number of autocorrelation the Q-value is based on and n is the number of parameters in the model. In the theory it is suggested that m is between 15 and 30. Due to the large number of observations m=30 is selected. The level of significance is α=0.05. The χ^2-values:

	# of parameters	χ^2
Linear model	4	38.885
Simple non-linear model	5	37.652
Extended non-linear	8	33.924

Table 8.13: Summary of the change in sign test and the Portmanteau lack-of-fit test of the residuals for the best models

	# change in sign	Significant	Residual sum	Portmanteau Q-value
One day				
Linear $\Delta T_{calc} I_{calc}$	622	No	0.325	466
Simple non-linear	616	No	0.195	280
Extended non-linear	732	Yes	0.0657	95
Three days				
Linear $\Delta T_{calc} I_{calc}$	1827	No	0.358	1548
Simple non-linear	1849	No	0.284	1225
Extended non-linear	2177	Yes	0.058	251

Comparing the χ^2-value with the Q-values, in Table 8.13, it cannot be rejected that there is a lack-of-fit for the models. One positive trend in the test values is that the Q becomes small when extending the models. If for instance the extended non-linear model had been based on about less than 500 observations, the test would have proven that there was no lack-of-fit between the data and the model. Based on these considerations it is noted that the Portmanteau lack-of-fit-test has difficulties of be appropriate for testing models based on a large amount of data. When this circumstance was uncovered it was researched if other similar test were resistant to a large number of observations. In the search for a more suitable method, the Ljung-box test was discovered but similar to the Portmanteau lack-of-fit-test the number of observations has impact on the Q-value when the number of observations increasest. The equation for estimating the Q-value of the Ljung-Box test is given in Equation 8.14.

$$Q^2_{ljung-box} = (N(N+2)) \sum_{i=1}^{k} \frac{\rho(i)^2}{(N-i)} \tag{8.14}$$

The Q-value of the Portmanteau lack-of-fit test as calculated as:

$$Q^2 = (\sqrt{N}\hat{\rho}_{\epsilon(\hat{\theta})}(1))^2 + (\sqrt{N}\hat{\rho}_{\epsilon(\hat{\theta})}(2))^2 + \cdots + (\sqrt{N}\hat{\rho}_{\epsilon(\hat{\theta})}(k))^2 \tag{8.15}$$

It can be seen from the two equations too that there is no difference in the two test when having many observations.

Table 8.14: Likelihood function values and the associated test statistics

	Likelihood function		Test statistics L_{ratio}	
	One day	Three days	One day	Three days
Linear	$2.553 \cdot 10^3$	$7.722 \cdot 10^3$	240	636
Simple non-linear	$2.673 \cdot 10^3$	$8.040 \cdot 10^3$		
Ad. simple non-linear	$2.668 \cdot 10^3$	$7.951 \cdot 10^3$	238	864
Extended non-linear	$2.787 \cdot 10^3$	$8.383 \cdot 10^3$		

8.17 Model validation

To validate the model dimensions the likelihood ratio test is consistent. It is only possible to compare the linear model to the simple non-linear model when taken into consideration that the simplest model has to be a subset model of the full model. Due to the fact that the extended non-linear model in all the preceding analyses has shown through, it is chosen to compare the extended model against an adapted simple non-linear model. As mentioned earlier it is doubtful whether the heat transfer from the wooden board transmits through convection as stated in the simple non-linear model. It is therefore decided to estimate the simple non-linear model where the heat transfer from the wooden board is transmitted through radiation as it is done in the extended non-linear model. The values of the likelihood function and the calculated L_{ratio} is shown is Figure 8.14. In both cases the χ^2-distribution has two degrees of freedom, applying a 5% level of significance the $\chi^2(2) = 5.991$. By comparing the L_{ratio} from the table it is evident that the model extension in both cases give a significantly better description of the data. This means that it is necessary to apply the full model to get the best description.

8.18 Discussion

As a short sum-up it can be concluded that the variables calculated on the basis of the measurements of the electrical flows give the best predictions. A very important finding is that the measured temperature at the top of the module gives the best description to the module temperature. Having in mind that the performance of the module decrease when the temperature raises, it makes sense to apply the top temperature which represents the highest temperatures in the module compared to the temperature at the bottom of the module. The limitation of the efficiency of the module is the lowest performing cell. Due to the correlation between the temperature of the module and the efficiency, the worse case efficient is modelled by having the top temperature of the model as the output variable. Furthermore, it is evident that the extended non-linear model is decisive better than the two other models attempted.

Comparing the performance of the similar models in this chapter and in the article [Jiménez et al. 2006], the performance is improved. The primary reason is the change of output variable from the average to the top temperature. Furthermore it does also influence that the calculated irradiance is applied instead of the measured.

The residual analysis reveals that none of the models give a perfect description of the data. A general conclusion from the analysis of the residuals is that the residuals of the extended non-linear model has the best performance. To name only a couple of the indications that the extended non-linear model has the best performance: the lowest standard deviations, significant change in sign test, and the most straight line in the cumulated periodogram One of the major difficulties for all the estimated model is to obtain acceptable prediction during the day hours. Non of the models got rid of the bulge of high residuals during the day hours.

Another finding in this chapter is that achieved by adding an extra state, where the causal filtered wind is estimated, the best results are attained. The analysis has revealed that only very little of the variation in the wind is removed. Due to this circumstance, it can be discussed if it is necessary to have this extra state. Furthermore a likelihood ratio test of the three-day models based on either the measured wind and the casual filtered wind confirm that the improvement of the model is significant. L_{ratio} is equal to 262. This value has to be compared to $\chi^2_{0.05\%}(1)$=3.841. When taken into account that the wind speed at the site is relatively low it can be assumed that the causal filter can be of even greater importance at locations where the wind is faster and more fluctuating.

The findings above demonstrate that the extended non-linear model has an acceptable performance, but it is still possible to improve the description of the module temperature, which will be attempted in the following chapter.

Chapter 9

Multiple State Models - Top and Bottom Divided Model

In the previous chapter it has become clear that it is difficult to obtain a satisfactory modelling of the temperature of the module. In this chapter there will be attached importance to the fact that some of the measured variables are collected both at the top and at the bottom of the module. The models in this chapter have two system equations: one for description of the temperature at the top and one for the bottom temperature. This makes it possible to estimate the two temperatures separately. This is called a multiple state model. This approach of analyzing the module in several separate sections has been used in earlier analyses. In [Christ 2001] a trisection solution is chosen for the purpose of simulation. The main reason for making this further development is to investigate if it is possible to get a better prediction of the module temperature.

9.1 The model

In the previous chapter it was found that the model where the causal filter was applied to the wind resulted in the best prediction. Unfortunately, despite numerous attempts, it has not been possible to estimate the causal filter for the multiple state models. This can either be the result of too complicated calcula-

tions or unsuccessful initial guesses of the parameter values. The results shown in the previous chapter revealed that in the case of these data the models based on the causal filtered wind and the measured wind generated quite similar results. It is therefore decided to estimate these models based on the measured wind. This will still indicate fairly whether the multiple state models have better or worse performance compared to the single state models.

When taking the available data into account and also the wish for generating models of similar structure to the extended model in the previous chapter, two feasible solutions are identified. The first model is very similar to the extended non-linear model in the previous chapter. The other model is a more advanced one, where CTSM estimates the temperature halfway up the gap. The models will be described in furhter details in the next two sections.

It is necessary firstly to introduce the concept of modelling the top and bottom temperatures of the module separately. A scheme of the estimation set-up is shown in Figure 9.1.

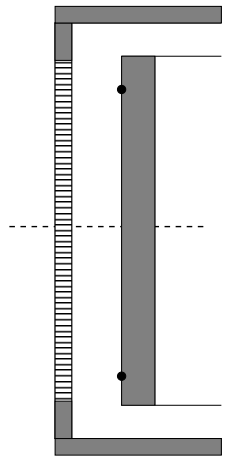

Figure 9.1: Scheme of the set-up with the dotted line indicating the place of lumping

The dotted dividing line between the top and the bottom parts can be viewed as the lumping point. Since the mean radiant temperature, the ambient air temperature and the irradiation are affecting the top and the bottom of the module equally, these contributors are identical in the two system equations in Equation 9.1. ΔT is the temperature difference in the air gap. Two scenarios are conceivable in relation to ΔT. The simplest assumption is to apply the ΔT estimated over the entire height of the gap. In this case ΔT below the dividing

line will be larger compared to the ΔT of the top of the gap. This means that the model is approximative concerning ΔT. In the second scenario it is attempted to estimate the temperature halfway up the air gap. The terms estimating the infrared radiation from the wooden wall to the module are different in the two state space equations. This is done since a temperature at both the top and the bottom of the wooden board is measured. This is an opportunity to eliminate a possible source of error, since there is up to a 5^oC difference between the top and the bottom of the board.

$$\begin{aligned}
dT_{bottom} &= \Big(1 - \frac{1}{1+exp(-f)}\Big)\Big(k_{rad}(T_{rad}^4 - T_{bottom}^4)dt + \ldots \\
dT_{top} &= \frac{1}{1+exp(-f)}\Big(k_{rad}(T_{rad}^4 - T_{top}^4)dt + \ldots \\
df &= dw_3
\end{aligned}$$

$$\begin{aligned}
&k_{air}W_{windair}^k(T_{air} - T_{bottom})dt + k_{delta}\Delta T dt + \ldots \\
&k_{air}W_{windair}^k(T_{air} - T_{top})dt + k_{delta}\Delta T dt + \ldots
\end{aligned}$$

$$\begin{aligned}
&k_{woodbottom}(T_{wood,bottom}^4 - T_{bottom}^4)dt + k_{irrad}I_{irrad}dt\Big) + dw_1 \\
&k_{woodtop}(T_{wood,top}^4 - T_{top}^4)dt + k_{irrad}I_{irrad}dt\Big) + dw_2
\end{aligned} \tag{9.1}$$

$$\begin{aligned}
T_{mbottom} &= T_{bottom} + e_1 \\
T_{mtop} &= T_{top} + e_2
\end{aligned} \tag{9.2}$$

Inspired by the thermal images it is considered necessary to be able to weight the influence of the top and the bottom temperatures of the module respectively. This weighting is denoted f, and is estimated in CTSM. f and $(1-f)$ denote how much influence the top and the bottom system equations respectively shall be have in order to obtain the best description of the module temperature. It is decided to apply f and $(1-f)$ in all the terms in the state equations. The fractions can be seen as indications as to where the dividing line needs to be set. At the introducing stages of estimation, f was entered in the model as a constant parameter. When thinking about how fluctuating the data and thereby the estimated models are, f is not sufficiently dynamical. This can also be concluded from the thermal images presented ealier. It was decided to reestimate the models, where f was applied as a state, only described by a noise term. In this way it is possible to let f vary during the entire period.

The logistic function, $\frac{1}{1+exp(-f)}$, is applied in order to assure that f is limited to be between 0 and 1. Unfortunately it becomes even more intensive for CTSM to estimate, when the logistic function is applied to the model. CTSM does estimate a model, but surprisingly the number of estimations drops from

above 150 to less than ten. CTSM uses iteration to find the solution. The number of iterations is one of the output values in CTSM. This leads to a situation where the estimated parameters are close to or equal to the initial parameter values. The reason for the low number of iterations can be that CTSM has found a local minimum of the objective function instead of the global minimum. In order to find the global minimum it can be attempted to change the initial parameter values in CTSM. This has been attempted by using several different initial values. It has been possible to raise the number of iterations from less that 10 to about 60. Still some of the estimates are close to the intial values, which leads to less trustworthy results.

A general problem in relation to all the models estimated in this chapter is that they are very sensitive to variations in data and changes in the definition of the estimation parameters. This implies that one minor change in the set-up has detectable influence on the rest of the model. The estimation of the models is very time consuming, therefore it is not possible to investigate all possibilities. In the sections below, argumentation for the decisions will be given.

The data applied in the coming analyses are dating from the 16^{th} of August. The previous analyses revealed only very little difference between the results based on one-day and three-day data respectively, which justifies for applying one-day data in the analysis below.

9.2 Simple multiple state model

The model estimated in this section is identical with Equation 9.1 and 9.2.
This simple model is a rough approximation in accordance with the ΔT term. The approximation is rough as ΔT is not identical at the top and bottom parts of the module, since it must be expected that the cold entering air is heated up faster than the preheated air in the upper part of the air gap. Before estimating the residuals it was insured that the estimations in CTSM were adequate. The averages and standard deviations for the residuals of the temperature at the top and the bottom of the module respectively stated in Table 9.1.

Table 9.1: The averages and standard deviations of the residuals of the simple multiple state model

	Average	Std.dev.
$T_{moduletop}$	$2.23 \cdot 10^{-2}$	$9.74 \cdot 10^{-2}$
$T_{modulebottom}$	$-2.14 \cdot 10^{-2}$	$8.51 \cdot 10^{-2}$

Compared to the standard deviations of the single state models these values are more than double their size. In Figure 9.3 are the two residual plots. It can be seen that the model has essential difficulties in predicting the temperature during the day hours. Prediction difficulties during the day hours are also present in relation to the single state models. The residuals seem to be not random in regions of the plot. When comparing the behaviour with the single state models the residuals are larger, which is also stated in the table above. In order to clarify if it is possible to obtain a better description of the module, the autocorrelation and partial autocorrelation plots of the residuals are created. The values of both of the autocorrelation functions are remarkably high. This an indication that the models are not adequate to describe the data. The partial autocorrelation plot, Figure 9.3, related to the temperature at the top of the module indicates that an extra state can increase the level of the description. The rest of the autocorrelation and partial autocorrelation plots show that the model has big difficulties in describing the data.

The plot of f plotted against the time is found in Figure 9.2. It can be seen that f lies between 0 and 1. The curve fluctuates. Surprisingly the temperature at the bottom of the module gives the majority of the description. This is an interesting finding since the single state modelling showed that the temperature in the top provided the best results. In the case of the single state space models the relationship between the top and the bottom temperatures was stationary, which makes it difficult in this situation to compare the performance of the models. When the sun is not influencing the module, the temperature at the top obtains more influence.

As an overall conclusion of the simple multiple state model it is determined to reject further use of the model due to the high standard deviation compared to the more simple single state model. Secondly the lack of stability is also an argument for no further application of the model.

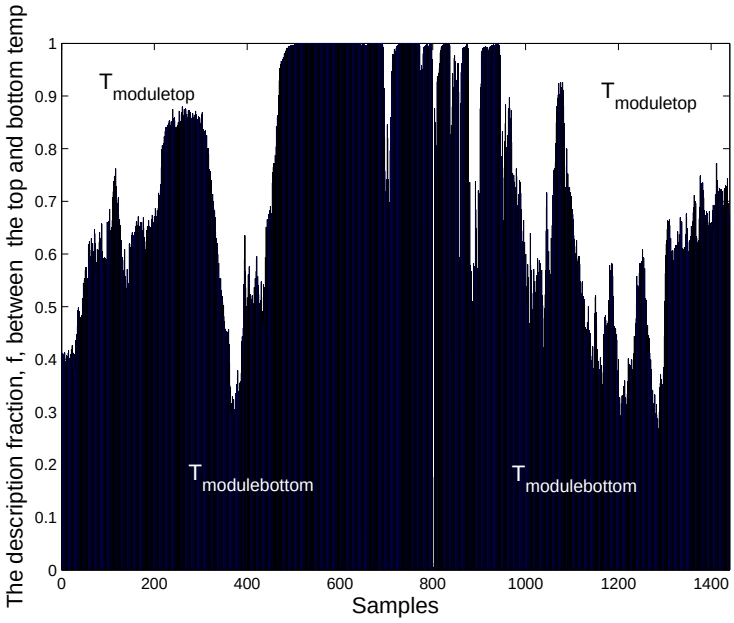

Figure 9.2: The variation in the estimated fraction indicator, f, of the simple multiple model

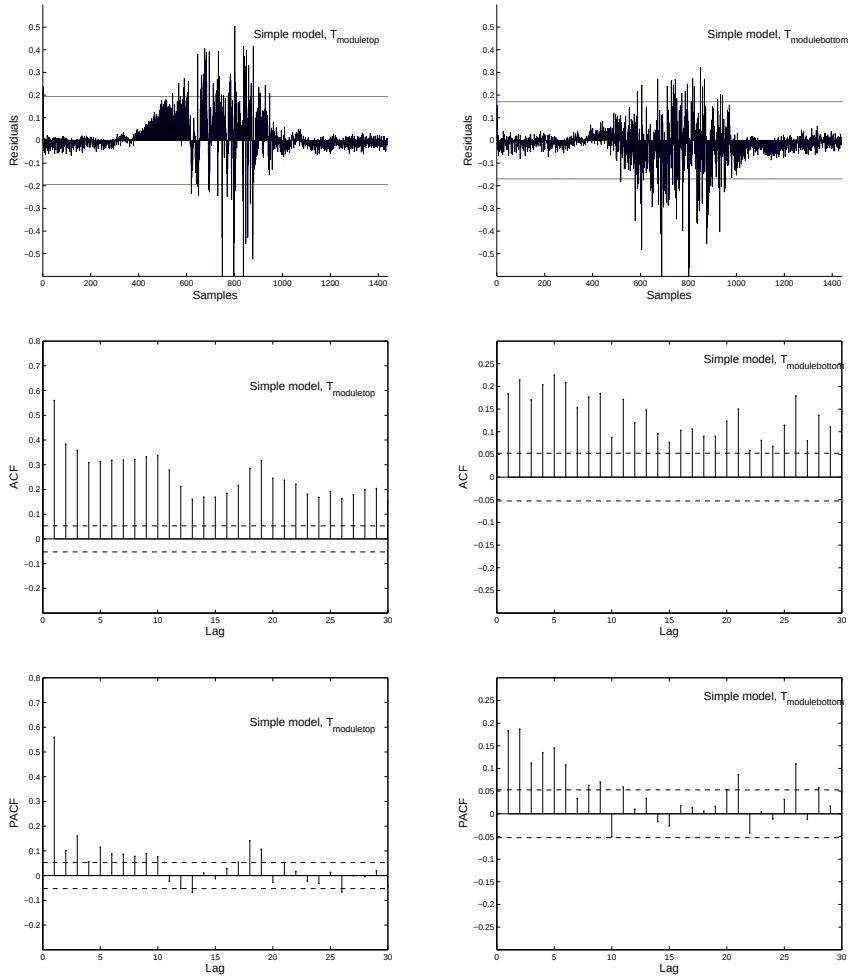

Figure 9.3: The residual, autocorrelation and partial autocorrelation plots of the simple multiple state model. Left-hand side of the plot is plot for the $T_{moduletop}$ and right-hand side is related to the $T_{modulebottom}$

9.3 Advanced multiple state model

Instead of applying ΔT in both system equations it is attempted to divide the convection heat transfer in the air gap into two parts. In the system equation of the temperature at the bottom of the module ΔT is rewritten as $(T_{in} - \hat{T}_{middle})$. In relation to the temperature at the top of the module ΔT is replaced by $(\hat{T}_{middle} - T_{out})$. Since $\hat{T}_{middle}$ is not measured the idea is to estimate the value in CTSM. When estimating the model it becomes clear that this is not possible since the estimated value of $\hat{T}_{middle}$, 18.7°C, is lower than the average of T_{in}, which is 22.5°C. An alternative possibility is then to fix T_{middle} to the average temperature of T_{in} and T_{out}. This is an approximation due to the knowledge about the distribution of heata in the air gap, where the cold air in the beginning of the gap is heated faster than at the top of the gap. Furthermore, it is expected that the distribution will change over time due to temperature changes. Therefore it is decided to use a simple average between the T_{out} and T_{in} as an approximation. In Table 9.2 the averages and standard deviations of the residuals are listed. The standard deviations reveal that the model is better in predicting the temperature at the top of the module than the bottom temperature. Comparing the values with the prior result the model is performing better than the previous multiple state model. Still the performance is worse than the performance of the extended single state model. As for the simple multiple state model the different residual plots are combined in Figure 9.5. The residual plots show more random tendencies compared to the latter model. To substantiate the standard deviation values the plots of residuals reveal that the model has big troubles in predicting the temperature at the bottom of the module. The plots of the autocorrelation functions and the partial autocorrelation functions of the residuals show a need of adding an extra state in order to obtain a better prediction. The added state has to describe a temperature close to the present bottom temperature. In Figure 9.4 the fraction determining the influence of the top and the bottom temperatures respectively appears. For this model the temperature in the top of the module describes the module best. In particular around noon f is very fluctuating. This gives room for speculations if the noise added to f has too large influence. In a further investigation it can be tested if for instance f shall be defined in another way. One method to control the influence of the noise is by fixing the variance of the noise term. Another method is estimating f depending on the previous observation times a coefficient, a, and the noise term. See Equation 9.3 where the equation is stated in continuous time. In Equation 9.4 the equation is written in discrete time. This will extend the complexity of the model and the estimation process.

$$df = a \cdot f + dw \tag{9.3}$$

$$f(t+1) = \phi f(t) + e \tag{9.4}$$

Table 9.2: The averages and standard deviations of the residuals of the advanced multiple state model

	Average	Std.dev.
$T_{moduletop}$	$-2.50 \cdot 10^{-3}$	$4.56 \cdot 10^{-2}$
$T_{modulebottom}$	$-2.31 \cdot 10^{-2}$	$9.88 \cdot 10^{-2}$

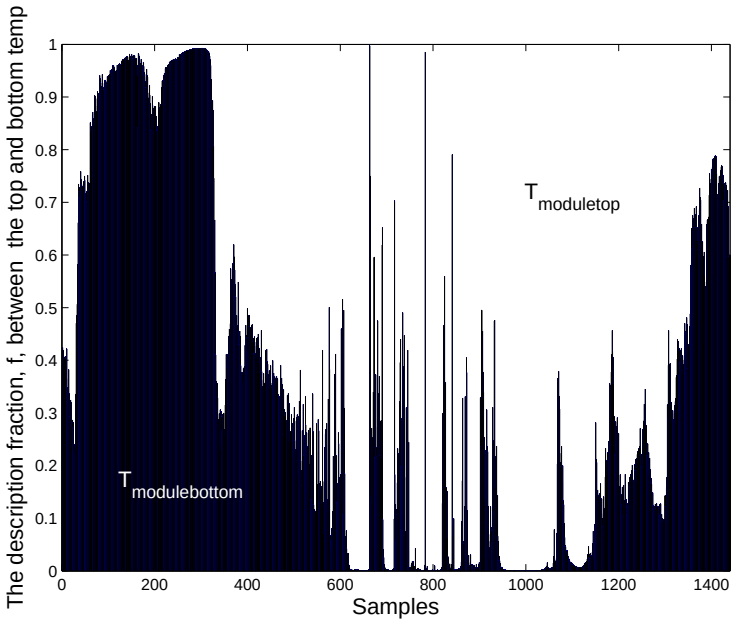

Figure 9.4: The variation in the estimated fraction indicator, f, of the advanced multiple model

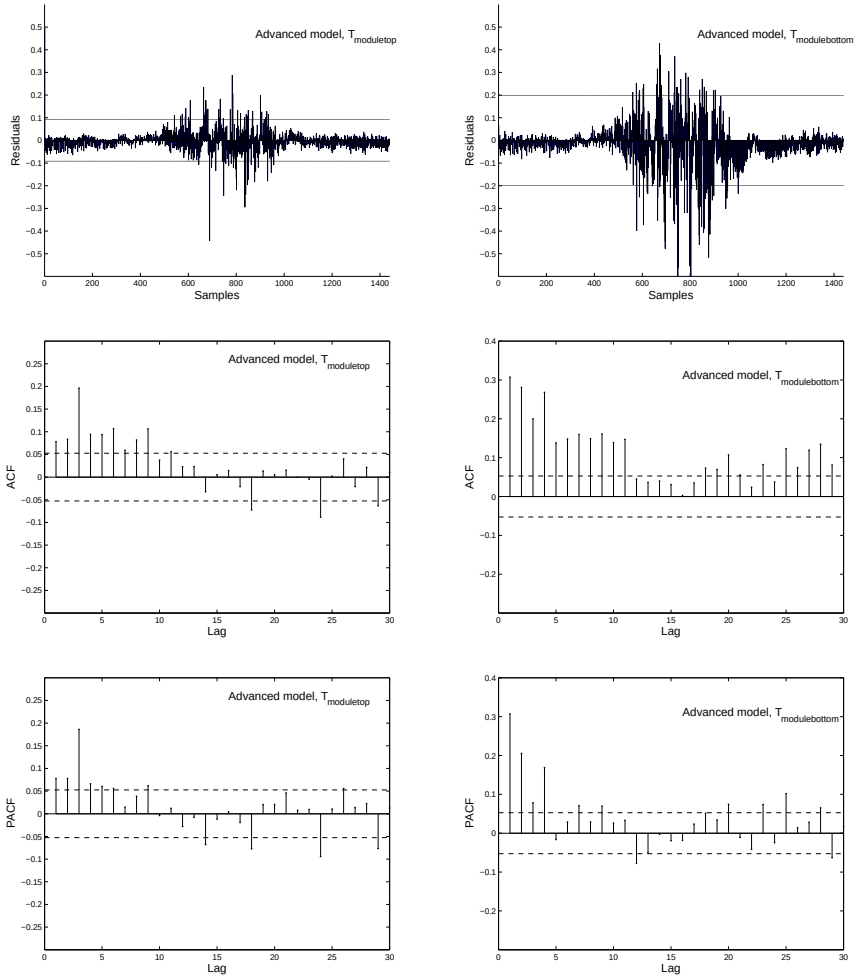

Figure 9.5: The residual, autocorrelation and partial autocorrelation plots of the advanced multiple state model. Left-hand side of the plot is plot for the $T_{moduletop}$ and right-hand side is related to the $T_{modulebottom}$

9.4 Summation

The analyses and results in the sections above show that the tested models have a worse performance compared to the single state models investigated in the previous chapter. It is of no benefit to extend a model and attain a worse result.

As mentioned in the beginning of the chapter the model estimation is unstable. Some of the reasons for this are identified. The models are very sensitive to the input data. This was experienced when data of the following day were applied. Furthermore the CTSM estimation has an influence on the unstable results, due to the low number of iterations before a model is identified. However, if it is assumed that the results make sense, the analysis reveals that extra states added to the models could be favourable. On basis of the available data it is not possible to make such extension of the models. The results point towards undertaking more measurements of the module temperature. This is not an easy task since thermo couples have to be placed inside the module. If it is possible, the extension of measurement can be done at two levels. First it could be an advantage to carry out more horizontal measurements in order to get a better determination of the top and the bottom temperatures respectively. Another improvement could be to measure more temperatures vertically. This will allows to extend the number of state equations and thereby the level of description. When keeping the thermal images in mind this would probably increase the level of description. On the other hand these extensions will increase the complexity of the models in relation to the estimation in CTSM.

It must be concluded that the models estimated on the basis of the available data are not performing satisfactorily compared to the single state models. In the following analysis the extended single state model will therefore be applied.

Chapter 10

Analysis of the Forced Ventilation in the Air Gap

The main purpose of the previous chapters was to identify the best fitting model. In this chapter the aim is to investigate how various conditions in the air gap behind the PV module influence the heat transfer and in general the estimations of the models. This is also an opportunity of testing the reliability of the model. Changes in the conditions can reveal strength and weaknesses of the model. The applied model in this chapter is the extended non-linear single state model predicting the temperatures at the top of the module. This model has proven to be the overall best describing model though it has troubles predicting during the day hours.

10.1 The set-up and the data

The model applied for estimation is the extended non-linear single state model from in Chapter 8. While estimating the models it has become clear that, due to fluctuation in the wind, the causal filtered wind, W_f, needs to be fixed. If W_f is not fixed the standard deviation of the state parameters becomes too high. It is therefore chosen to fix W_f to the average of the measured wind of the specific data set.

The parameters of the models are very sensitive to variations in the wind speed. This means that the models have to be estimated from data where the average wind is quite similar.
Two factors are changeable in relation to the air gap: the forced velocity and the air resistance. The facts of the changeable factors are listed below.

- **The forced velocity**
  Just as in the prior analyses the forced ventilation is held constant in a 24 hour period. In the analysis three velocities are examined. For the purposes of comparison the ambient wind speed is in general below 1 m/s.
  - 1.38 m/s (Level 6)
  - 2.49 m/s (Level 10)
  - 3.43 m/s (Level 13)
- **The air resistance**
  There are two set-ups in order to change the air resistance in the gap:
  - No fins results in a free laminar flow
  - Fins creating a turbulent flow

Figure 10.1: Left: A photo of the fins inside the air gap in the BIPV module[Christ 2001]. Right: A scheme showing the forced air flow in the gap.

The transversal fins, see Figure 10.1 to the left, are placed in the gap in order to make the laminar air flow turbulent. On the right hand side in Figure 10.1 is a scheme of how the flow is expected to behave in the air gap. There is a lack of

theory concerning the behaviour of the air flow in the gap. Prior research has focused on air flow in gaps where the fins are present at both sides of the gap. In the article by [Chin-Hsiang & Wen-Hsiung 1990] the flow patterns in a gap where the fins are in staggered positions at each side of the gap are discussed. Others have made research having the fins placed exactly opposite one another. These studies reveal that the incoming laminar flow becomes more turbulent. The flow in the gap behind the PV module is expected to be less turbulent since a certain amount of the air will succeed in moving along the PV module. In the case of the BIPV module it is not possible to place fins at both sides of the gap, since the fins should then be mounted at the rear side of the module. If fins were mounted at the rear side this might lead to an unwanted influence on the module temperature. In this situation also the heat transfer coefficient, which has to be estimated, will be affected in an unfavorable way. The mounting could affect the functionality of the fragile module.

The set-up containing the fins increases the effective heat transfer area in the air gap. This can lead to an increase of the heat transfer from the PV module to the air in the gap behind the PV module[Christ 2001] and [Bazdidi-Tehrani & Naderi-Abadi 2004]. Compared to the set-up without fins the set-up with fins will lead to a higher air temperature in the gap and moreover a decrease in the module temperature. This effect is favourable, since a decrease in the module temperature raises the electrical performance of the module. The transfer of the heat from the module to the ventilation air is also desirable, if the air is meant to heat the building.

The applied data are all 24-hour data. It is examined that the forced velocity is held constant in 24-hour periods. In the original data there was also a forced ventilation Level 7, but no 24-hour periods were measured. Under these circumstances it was decided not to apply these data. In the light of the previous analyses it is found that 24-hour data are acceptable in order to obtain stable and useful models. The data are measured during a one-month period. In Appendix A.1 a table containing information as to when the information is collected. This may lead to deviation in the external conditions such as temperature and wind speed. According to the logbook of the testing, [Gandini n.d.], all the days were sunny days. The irradiation is one of the important conditions with a view to having equal testing conditions, when dealing with PV modules.

Table 10.1: Averages and standard deviations of the residuals for the models having different velocities in the air gap with and without fins

Fins	Velocity	Average (std.dev.)
Yes	6	$-6.301 \cdot 10^{-4}(3.091 \cdot 10^{-2})$
	10	$-2.292 \cdot 10^{-4}(3.144 \cdot 10^{-2})$
	13	$2.206 \cdot 10^{-4}$ $(3.423 \cdot 10^{-2})$
No	6	$3.477 \cdot 10^{-5}$ $(3.091 \cdot 10^{-2})$
	10	$-1.162 \cdot 10^{-4}(3.332 \cdot 10^{-2})$
	13	$-4.397 \cdot 10^{-5}(3.673 \cdot 10^{-2})$

10.2 The results of the analysis

10.2.1 The residuals

Initially the averages and the standard deviations for the six models are calculated. These values are found in Table 10.1. Compared to the averages and standard deviations of the extended non-linear models, Chapter 8, the values are at the same level; Averages are of the magnitude of $10^{-4} - 10^{-5}$ and the standard deviations are in both cases about $3 \cdot 10^{-2}$. The comparison of the present models reveals that the standard deviations of the model based on data, where the fins were installed, have the lowest standard deviation of the entire thesis. Moreover the standard deviation value in general appreciates along with increased level of velocity.

In Figure 10.2 the residual plots can be seen. When comparing the left and right hand side, respectively no fins and fins, of Figure 10.2 the residuals of the central samples, corresponding to noon, are smaller for the model based on data where the fins were placed in the air gap. The bulge, which has been seen throughout the entire thesis, is nearly erased. In general these plots indicate that the model is better to predict the energy flows of the module, when the fins are applied to the set-up. In the plots of the autocorrelation function, Figure 10.3, a very different picture appears. It is necessary to keep in mind that the confidence limits only can be used as indications due to the large number of observations. The autocorrelation function plots representing data where no fins were placed in the air gap indicate that the models are adequate, whereas the plots representing the data where the fins were placed in the air gap show that the model has a lack of fit in relation to describe the input data. The plots of the partial autocorrelation function, Figure 10.4, supports the findings above. Especially the plot in the upper right corner indicates the necessity of adding two extra states to the models with the fins in the gap. This could be carried

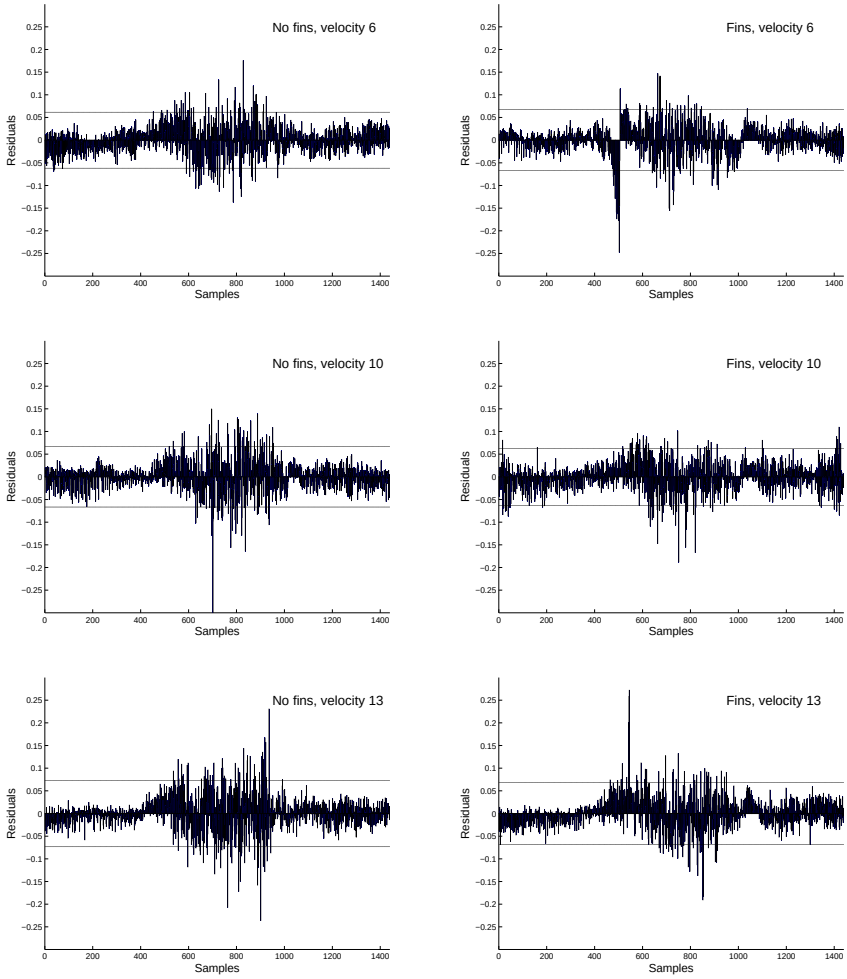

Figure 10.2: Residual plots. To the left the residual of the model based on data, where no fins were placed in the gap. On the right hand side the residual of models where the fins were placed in the gap.

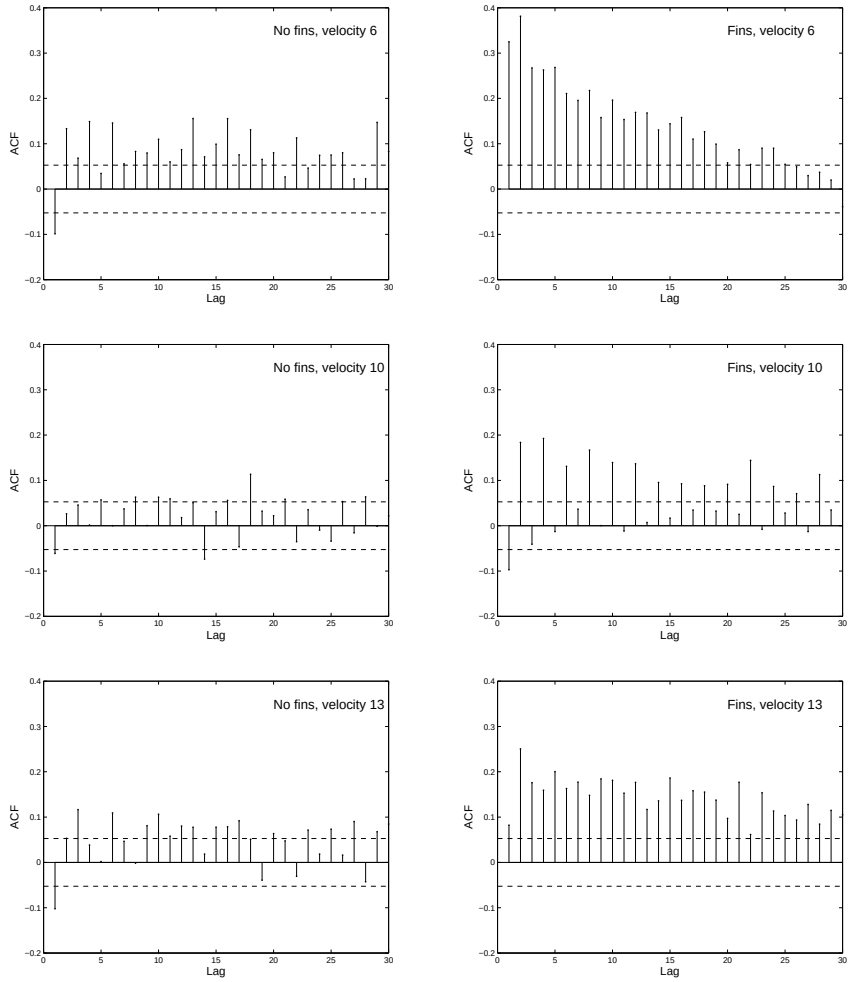

Figure 10.3: The autocorrelation functions of the residuals of the estimated models

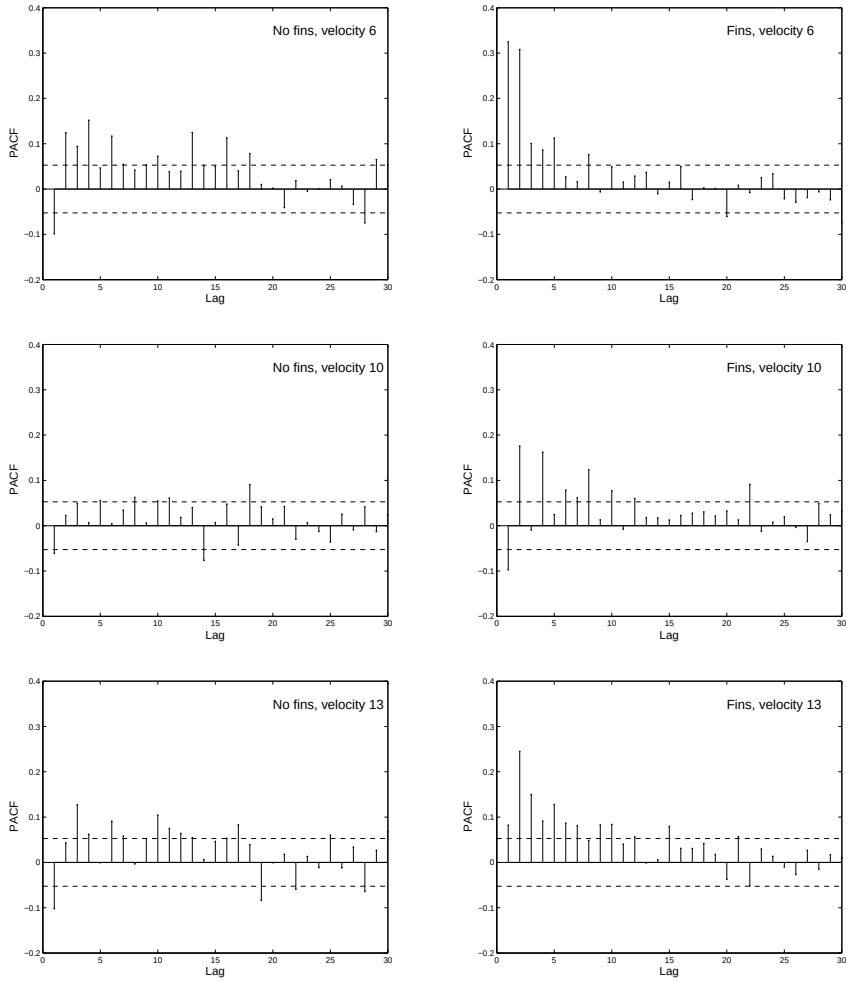

Figure 10.4: The partial autocorrelation functions of the residuals of the estimated models

out if for instance it was possible to divide the estimation of the module into three individual areas. Unfortunately, a sufficient amount of measurements is not available. Some of the other partial autocorrelation plot gives weaker indications of the necessity of adding one or two extra states.

In order to analyze the residuals both in the time domain and the frequency domain the cumulated periodograms of the models have been estimated and plotted. The plots are found in Appendix A.3. The plots reveal that the models where no fins were placed in the gap form nearly straight lines, whereas the other models have a lack of fit for low frequencies.

To make sense out of the results above it seems as if the models based on the set-up with fins can be very powerful if more measurements are available, since the standard deviations of these models are already lower compared to the set-up without fins. An extra state is likely to improve the performance of the model. Such as extension can be tested by the likelihood ratio test in order to determine if the improvement is significant.

10.2.2 Analysis of the parameter values

In Table 10.2 and 10.3 the estimated parameters and the average wind are listed. In the left column signs indicate the trend of the parameters of the different models. $\rightsquigarrow$ denotes that no clear trend is discovered while $\searrow$ and $\nearrow$ indicate declining or rising tendencies of the parameter values. The reason for the two stars in Table 10.2 is that is has been necessary to estimate this model by fixing W_f to an average of the winds of the other models. This decision is taken since the measured wind speed during the trial with no fins and velocity 13 was higher than on the other two days. The tables show that the parameter values are highly dependent on the specific day measurements. It shall be noted that a compared to the other models is very low, 0.085, in the model where the forced velocity is at level 10 and no fins are placed in the air gap. No reason for this has been discovered. In this situation it would have been an advantage if more similar data were available in order to estimate an extra model to verify the result. A common trend of the two tables is that k_{air} rises when the forced velocity is increased and $k_{windair}$ drops accordingly with rising forced velocity in the air gap. These two parameter estimates outline the heat transfer from the module and the ambient air. This relationship will be analyzed in the next section. In the previous section it was mentioned that the heat transfer between the module to the air in the gap would be increased when fins were added. The parameter estimate, k_{delta}, reveals a dominating tendency.

Table 10.2: The parameters of the models without fins in the air gap

	No Fins			
Velocity	6	10	13	Trend
μ_{wind}	0.775	0.714	0.891	$\rightsquigarrow$
k_{rad}	$8.672 \cdot 10^{-10}$ ($1.588 \cdot 10^{-10}$)	$1.519 \cdot 10^{-9}$ ($3.484 \cdot 10^{-10}$)	$3.994 \cdot 10^{-9}$ ($3.908 \cdot 10^{-10}$)	$\nearrow$
T_{rad}	263.42 ($6.556 \cdot 10^{0}$)	285.95 ($1.748 \cdot 10^{0}$)	287.12 ($4.703 \cdot 10^{-1}$)	$\nearrow$
k_{air}	$8.651 \cdot 10^{-1}$ ($1.505 \cdot 10^{-1}$)	$1.478 \cdot 10^{0}$ ($1.491 \cdot 10^{-1}$)	$1.869 \cdot 10^{0}$ ($1.457 \cdot 10^{-1}$)	$\nearrow^{*}$
$k_{windair}$	$2.634 \cdot 10^{-1}$ ($4.265 \cdot 10^{-2}$)	$1.683 \cdot 10^{-1}$ ($1.885 \cdot 10^{-2}$)	$1.349 \cdot 10^{-1}$ ($3.947 \cdot 10^{-2}$)	$\searrow^{*}$
k_{delta}	$-3.259 \cdot 10^{0}$ ($3.295 \cdot 10^{-1}$)	$-2.715 \cdot 10^{0}$ ($4.829 \cdot 10^{-1}$)	$-5.024 \cdot 10^{0}$ ($6.564 \cdot 10^{-2}$)	$\rightsquigarrow$
k_{wood}	$6.288 \cdot 10^{-10}$ ($5.423 \cdot 10^{-10}$)	$2.509 \cdot 10^{-9}$ ($4.838 \cdot 10^{-10}$)	$7.199 \cdot 10^{-9}$ ($4.342 \cdot 10^{-10}$)	$\rightsquigarrow$
k_{irrad}	$1.080 \cdot 10^{-1}$ ($3.740 \cdot 10^{-3}$)	$9.895 \cdot 10^{-2}$ ($2.573 \cdot 10^{-3}$)	$1.131 \cdot 10^{-1}$ ($1.766 \cdot 10^{-3}$)	$\rightsquigarrow$
a	0.335 ($1.305 \cdot 10^{-1}$)	0.085 ($9.143 \cdot 10^{-2}$)	0.327 ($1.650 \cdot 10^{-1}$)	$\rightsquigarrow$

Table 10.3: The table contains the parameters of the models with fins in the air gap

	Fins			
Velocity	6	10	13	Trend
μ_{wind}	0.800	0.899	0.851	$\rightsquigarrow$
k_{rad}	$2.688 \cdot 10^{-9}$ ($1.479 \cdot 10^{-10}$)	$2.116 \cdot 10^{-9}$ ($1.607 \cdot 10^{-10}$)	$2.557 \cdot 10^{-9}$ ($2.447 \cdot 10^{-10}$)	$\rightsquigarrow$
T_{rad}	281.72 ($3.405 \cdot 10^{-1}$)	280.20 ($1.127 \cdot 10^{0}$)	286.32 ($9.324 \cdot 10^{-1}$)	$\rightsquigarrow$
k_{air}	$5.853 \cdot 10^{-1}$ ($6.750 \cdot 10^{-2}$)	$1.338 \cdot 10^{0}$ ($1.465 \cdot 10^{-1}$)	$1.882 \cdot 10^{0}$ ($1.601 \cdot 10^{-1}$)	$\nearrow$
$k_{windair}$	$4.565 \cdot 10^{-1}$ ($4.456 \cdot 10^{-2}$)	$2.577 \cdot 10^{-1}$ ($2.493 \cdot 10^{-2}$)	$1.561 \cdot 10^{-1}$ ($1.454 \cdot 10^{-2}$)	$\searrow$
k_{delta}	$-2.757 \cdot 10^{-1}$ ($1.952 \cdot 10^{-1}$)	$-3.330 \cdot 10^{0}$ ($4.951 \cdot 10^{-1}$)	$-8.661 \cdot 10^{-1}$ ($5.290 \cdot 10^{-1}$)	$\rightsquigarrow$
k_{wood}	$2.816 \cdot 10^{-9}$ ($3.268 \cdot 10^{-10}$)	$1.262 \cdot 10^{-9}$ ($4.099 \cdot 10^{-10}$)	$1.883 \cdot 10^{-9}$ ($4.283 \cdot 10^{-10}$)	$\rightsquigarrow$
k_{irrad}	$8.734 \cdot 10^{-2}$ ($1.912 \cdot 10^{-3}$)	$9.806 \cdot 10^{-2}$ ($3.088 \cdot 10^{-3}$)	$1.009 \cdot 10^{-1}$ ($3.508 \cdot 10^{-3}$)	$\nearrow$
a	0.392 ($7.541 \cdot 10^{-2}$)	0.345 ($1.090 \cdot 10^{-1}$)	0.305 ($1.858 \cdot 10^{-1}$)	$\searrow$

10.2.3 The heat transfer of the BIPV module

The heat transfer between the module and the air in the gap is not analyzed due to no clear tendency of the parameter estimates.
The heat transfer coefficient, h_c, between the ambient air and the module is expected to vary when changing the velocity and the type of flow. The heat transfer coefficient from the module to the air is given as $(k_{air}C_{BIPV}/A)$. C_{BIPV} is the effective heat capacity of the module and A denotes the area of the module. Both C_{BIPV} and A are constant since the same module is applied in all the trials. Due to the lack of knowledge as to the value of C_{BIPV} both constants are not removed in the coming analysis. This means that the values are not directly the physical expressions, but the scaled dimensions of the results of the different models are correct. The scaled heat transfer coefficient from the PV module to the ambient air is from now on denoted the ambient convective coefficient.

In Figure 10.5 two plots are showing the ambient convective coefficient, k_{air}, plotted against the wind speed raised to the power of the estimated coefficient, $k_{windair}$. The plot at the top presents the relations in the interval of the measured ambient wind. The other plot is an extrapolation to illustrate the relationship in the broader interval. There is a risk that the relationships found in this broad interval are artificial. The applied parameter estimates are identical to the values in Table in 10.3 and 10.2. The wind speed raised to an estimated coefficient is not part of the heat transfer coefficient, but the two parameters are closely linked together. In the chapter treating the single state models it is proven that the wind does have an influence on the performance of the module. The upper plot clearly shows that the velocity of the air in the gap has a strong influence on the heat transfer. The higher the forced velocity is, the higher the heat transfer to the ambient air becomes. The influence of the fins compared to the velocity level is very small. The plot actually reveals that the heat transfer for the low velocities is better when the fins are not in the set-up. There may be a link due to the fact that the low velocity levels and the ambient wind speed are very close to each other, which neglect the influence. There is significant difference between fins and no fins in relation to the forced velocity level 6. This difference becomes smaller the higher the forced velocity gets. At the velocity level 13 the ambient convective coefficient is about the same.

The analysis of the models revealed that the models based on the data with fins yielded better predictions. This may have an influence when comparing the parameter estimates of the models with or without fins.

Due to the preceding research showing that the fins should have an increasing impact on the heat transfer it is investigated if other conditions in the trial and the data have influenced the heat transfer of the module. The ambient con-

vective coefficient described the relationship between the ambient temperature and the module temperature. Therefore the difference between the ambient air and the temperature at the top of module is plotted against time. This done for both the models with or without fins in the air gap. The reason why it is chosen to plot the difference between the temperature of the ambient air and the module is that part of the variation of the ambient temperature is removed.

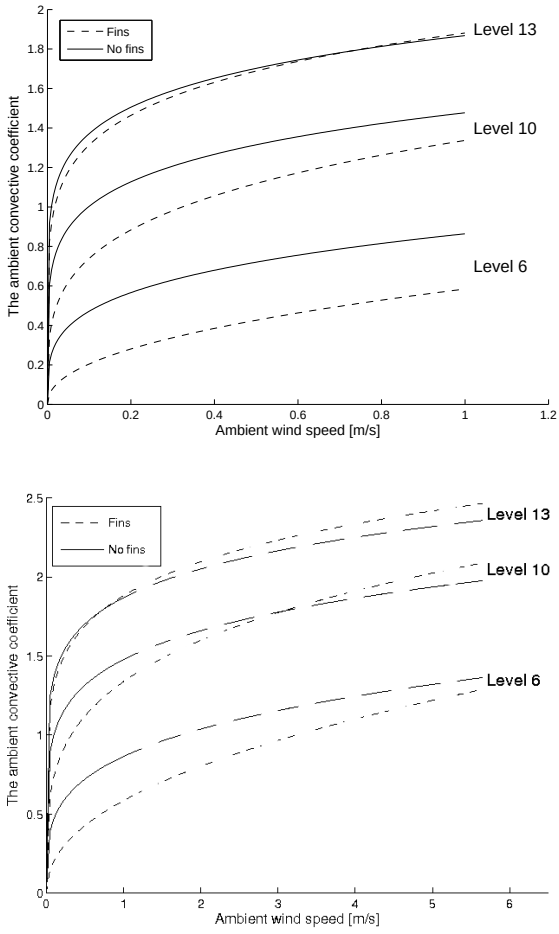

Figure 10.5: The relationship between the ambient wind and the ambient convective coefficient for set-ups with or without fins. Also the forced velocity is varied. The upper plot is the relationships in the interval of the measured wind, while the plot in the bottom is an extrapolation

The plot where the forced velocity is at level 13 is shown in Figure 10.6. It is seen that the difference is smaller in the case of the set-up where the fins are applied. This gives an indication that the fins in the air gap have removed heat from the module. This removed portion of heat may lead to lower ambient conventive coefficients. In Appendix A.2 similar plots for the forced velocity 6 and 10 are found.

These plots underline that the largest temperature difference between the ambient air and the module is for the forced velocity level 13. The bottom plot of Figure 10.5 is an extrapolation which may show a possible behaviour of the ambient convective coefficient for a broader interval of the ambient wind speed. At all three different levels of forced velocity there is a trend that the fins have more influence along with higher ambient wind speeds.

For all three velocities there is a point where the lines representing with or without fins respectively intersect. This means that at a certain point the set-up with fins obtains a larger ambient convective coefficient and thereby a larger heat transfer coefficient. This effect has to be added to the effect that the temperture in the module has already been reduced. The trend is furthermore that the higher the forced velocity is, the faster a change takes place.

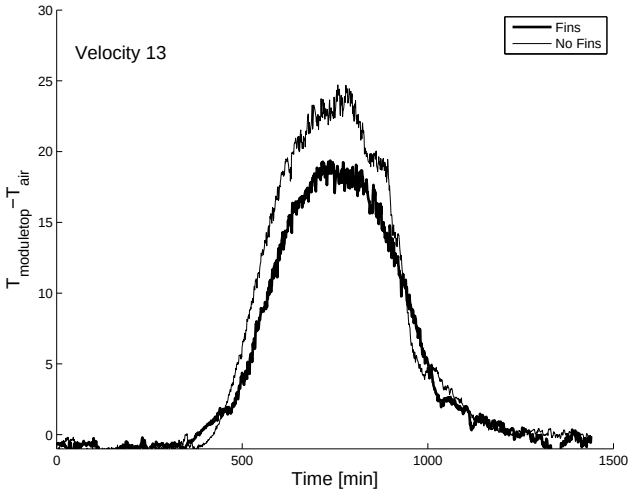

Figure 10.6: The difference between $T_{moduletop}$ and T_{air} plotted against time for set-ups with and without fins in the air gap.

10.3 Summary

This chapter has revealed that the extended non-linear single state model is able to show the difference when the forced velocity and the conditions in the air gap are changed. The residual analysis has shown that the model where the fins are applied in the set-up gives lower standard deviations of the residuals. The important finding is that the prediction in the day hours is improved. This is an essential discovery since the majority of the electricity is produced during these hours, which means that it is of special importance to be able to predict the temperature in these hours. The partial autocorrelation moreover indicates that by adding an extra state to the model an even better description is possible. Due to the lowest calculated standard deviations and the improve predictions during the day hours, it can be concluded that the model where the fins are applied to the set-up are the best performing models of the entire thesis. It has to be added that the performance of the model increases the higher the forced velocity in the air gap is.

The dissimilarities of the model prediction between the two set-ups can be due to either the variations in the measured data or an actually better ability for the model to predict the set-up where the fins are placed in the gap.

The models reveal that the heat transfer is increased when the forced ventilation level is increased. The effect of the fins is not just as contributing, but it was discovered that the difference between the ambient temperature and the temperature of the module were reduced due to the fins. The findings point in the direction that both the fins and a high-forced velocity in the air gap altogehter contributed to an increased heat transfer from the module.

The analysis shows that it is possible to increase the heat transfer from the module. The implementation and the operation of a forced velocity behind the module are energy consuming. In order to determine the gain of the set-up a cost-benefit analysis of the cost referring to the operation of the forced ventilation versus the efficiency gain of the module has to be carried out in order to evaluate if the final output is increased.

Chapter 11

Future Work

When working on a thesis like this, there will always be areas and ideas which cannot be further investigated due to lack of time. In the sections below some of my thoughts are outlined.

The multiple state models are evaluated to be less workable than the single state models, due to high residual values and, in general, the instability of the models. In a future analysis it could be interesting to investigate how to attain a functional and reliable multiple state models. The residuals of the extended non-linear single state model reveal that the worse predictions are found during the day hours, which is the most important period to get a good prediction of since the most electricity in produced in this period. This indications that there is still room for improvements of the model.

The primary focus of the thesis has been to identify models which are able to describe the collected data in a satisfactory way. The analysis of the values and the size of the parameters have been given lower priority. In some future work it could be of great interest to investigate and understand the estimated parameters. This shall be seen in the light of the model type, namely grey-box modelling, where the estimated parameters should actually estimate true physical parameters.

One major stop block in order to extend the models further is the amount of

measured variables. If for instance more measurements of the module temperature were available, it would be possible to construct a multiple model consisting of three states.

With regard to the thermal images it would also be of advantage if thermal images from the specific days of the applied data were taken. Also the statements about the possibly influence of humidity and the angle of the wind speed could be intersting to test in the models.

In relation to the set-up with fins in the air gap a further analysis where the results are compared to the complex empirical equations containing the Reynolds number or the Nusselt coefficient could be carried out. These numbers are can the help description the when dealing with turbulent wind and differencies in temperature, which is the case with these data.

It could also be exciting to carry out a cost-benefit analysis, which determines in energy and economics terms whether there is a benifit of applying the forced velocity of the air gap.

Chapter 12

Conclusion

The analyses of this thesis have proven that it is possible to model the temperature of the module. It is found that the photovoltaic module is complex in many ways. The improvements and findings of this thesis will be discussed in the section below. Due to the subdiscussions and conclusions at the end of the chapters, this chapter will appear as a combined discussion and conclusion.

The stochastic state space models have proven to be excellent with a view to handling the fluctuations of the measured data. Compared to the RC-models first described, the stochastic state models applied are able to model non-linearities. In the modelling process it is proven that the non-linear influence of the wind and the infrared radiation are significant.

The analyses of the models have revealed that the extended single state model had the overall best performance. The likelihood ratio test stressed the finding. Also the analyses of the residuals strengthened that the extended model has the most qualified way of describing the data. Since the analysis of the residuals shown signs of possible improvements, it has been attempted to extend the model from being single state to consist of multiple states. Due to the knowledge as to the changing heat distribution, the dynamical parameter f was implemented in the model to determine the degree of influence of the measured top and bottom temperatures respectively. The results of these models are very unstable. The complexity of the models has great influence on CTSM's ability

to estimate the model. Especially the causal and logistic function give rises to problems, which were hard to solve despite many attempts. It can also be discussed whether f is defined appropriate. Even though the residual values got worse compared to the single state models it is presumed that there is a potential in moving from single to multiple state models. This is stated in the light of the indications of the partial correlations plot for the necessity of extra states in the model. Also the fact that the worse predictions of the module temperatures occur during the day hours, where the most amount of electricity is produced. This makes it essential to improve the model performance. It has to be mentioned that the models where fins turn the forced air flow into a turbulent air flow, do not have the same difficulties of predicting during the day hours.

It can be determined that the performance of the estimated single state models is improved compared to the similar models in the article [Jiménez et al. 2006], which gave raise to this thesis. The primary reason is the change of output variable from the average temperature to the top temperature. Furthermore it does also influence that the calculated irradiance is applied instead of the measured.

In the case of the single state models both the data consisting of one-day and three-day measurement were estimated. When the performance was compared the three-day models had only a little lead. In relation to the residual analysis it was discovered that for some of the tests, e.g. the portmanteau lack-of-fit and the confidence intervals in general, it was a difficult factor that the models were based on that many observations. The analyses of the parameter estimates reveal that the individual set of data has great influence on the size of the parameters.

There are several circumstances in relation to the module temperature that makes it difficult to obtain a satisfactory description. The thermal images were a breakthrough to realize that the heat distribution of the module is complex. The variation over time is significant. Initially the description of the module temperature was an average between the measures at the top and the bottom of the module. Comparing the average temperature with the thermal images it was discovered that the tongues in the heat distribution render the average impossible. In accordance with the thermal images the temperature at the top resulted in the best predictions. It also makes great sense to apply the temperature at the top of the module, since the highest temperatures are found here and thereby the worse performance of the module is described. The temperature of the module, that has to be predicted, is the most important variable to be correct, but also the describing variables can help improve the description of the temperature of the module. The analysis revealed that the models based on the calculated ΔT and on the irradiance obtained the best results. Before this analysis it was believed that a filtered version of the ambient wind speed was influencing the performance of the module. The estimated models clearly show

that the fluctuations of the ambient wind do have influence. The causal filter which estimates the optimal filtration of the wind revealed that only very little of the measured wind has to be removed in order to obtain the optimal description of the module temperature. It can be discussed if it is necessary to apply the filter if only very little improvement is discovered. One strong argument for keeping the causal filter in the model, denoted as the best, is that the wind at the test site in Ispra is limited. It can therefore be expected that the causal filter will be of higher applicability at sites where the wind speed is higher and maybe more fluctuating.

The statements above underline the advantages of the abilities of testing the performance and the fitting of these grey-box models. Introductorily the non-linear influence of the wind was added in both the term of the convection from the ambient air to the module and the irradiance term. CTSM clearly show that the wind does not influence the irradiance into the module. In the investigation it became evident that the irradiance was measured inside the module which means that the ambient wind speed should not influence the irradiance. This entails a reduction of the model, which could also be determined on the basis of physical knowledge. This is one of the strength of grey-bpx modelling.

The last and cheering finding in this thesis is that the model is able to discriminate different velocity levels and set-up, with or without fins in the gap, from each other. The analyses have revealed that the forced velocity in the air gap has a significantly increasing influence on the heat transfer coefficient between the ambient air and the module. Furthermore, the difference between the ambient air temperature and the module temperature was higher for the set-up with fins, where the former laminar air flow is turned into a more turbulent flow. This underpins the theory of the influence of fins in air gaps known in advance. For the set-up with fins in the air gap the prediction problems during the day are nearly removed. This can be seen from the plot of the residuals and standard deviations of the residuals which are the most smooth and the lowest in the entire analysis.

Above several improvements for increasing the level of description of the photovoltaic module are stated. The analyses have proved to be adequate as to to modelling the measure data, though there is still room for further improvements. As an overall conclusion of the models it can be pointed out that, the models based on data where fins are placed in the air gap gave the best results. The standard deviations are the lowest identified, and the residuals are random. Furthermore, the desired increase in the performance of the module is obtained due to the increased heat transfer and the decrease of the temperature of the module.

Appendix A

Appendix to the Chapter Analysis of the Forced Ventilation in the Air Gap

A.1 The dates of the data collection

Table A.1: The dates of the data collection. - denotes that 24 hours data is not available

Velocity of the fan	No fins	Fins
6	14th of August	12 of August
7	-	-
10	17th of August	7th of August
13	28th of August	30th of July

A.2 The difference between $T_{moduletop}$ and T_{air}

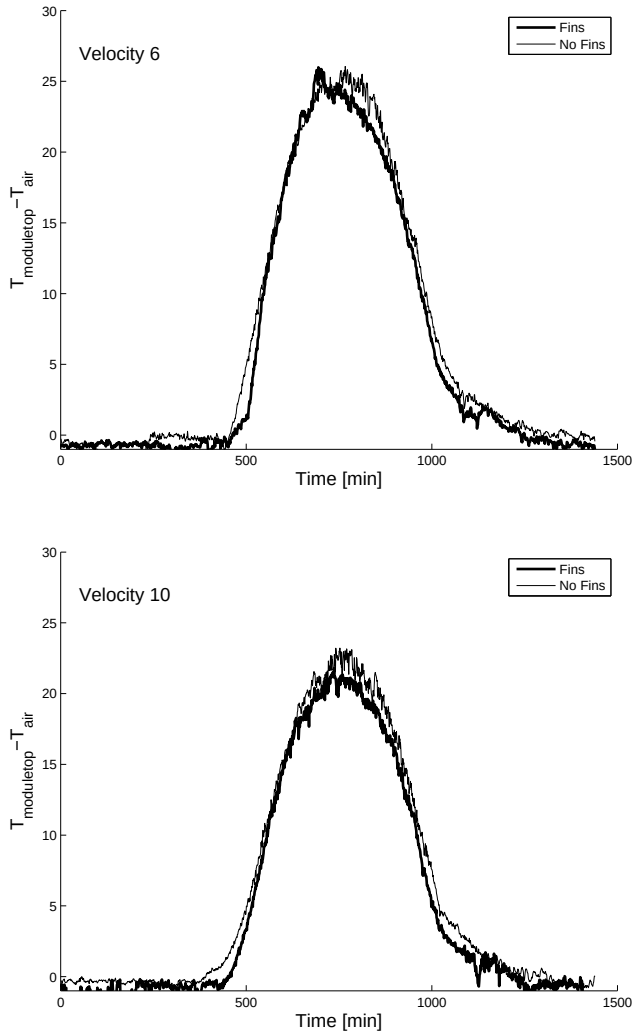

Figure A.1: The difference between $T_{moduletop}$ and T_{air} plotted against time for set-ups with and without fins in the air gap.

A.3 Cumulated periodograms

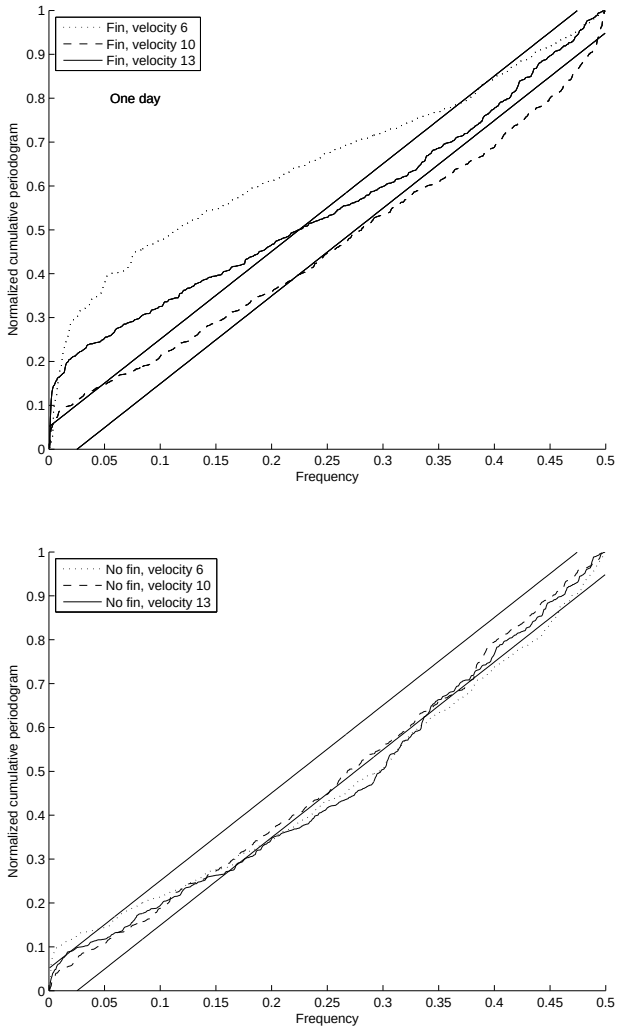

Figure A.2: Cumulated periodograms of respectively models with and without fins add in the gap

Bibliography

Andersen, K. K. (2001), Stochastic Modelling of Energy System, PhD thesis, IMM - Technical University of Denmark.

Andresen, I. (2002), 'Building integrated photovoltaics in smart energy-efficient buildings - a state-of-the-art', www.alaskasun.org/pdf/SolarTechnologies.pdf.

Bazdidi-Tehrani, F. & Naderi-Abadi, M. (2004), 'Numerical analysis of laminar heat transfer in entrance region of a horisontal channel with transverse fins', *Pergamon* .

Bazilian, M. D., Groenhout, N. K. & Prasad, D. (2002), 'Simplified numerical modelling and simulation of a photovoltaic heat recovery system', Article. Private file from Hans Bloem.

Bloem, H. (n.d.*a*), 'Iec 61215 version 1993', Not published - private file from Hans Bloem.

Bloem, J. J. (n.d.*b*), 'Bipv case study for modelling and analysis', Private file from Hans Bloem.

Bloomfield, P. (2000), *Fourier Analysis of Time series - An introduction*, second edition edn, Wiley.

Born, F. J. (2001), Aiding Renewable Energy Integration through Complimentary Demand-Supply Matching, PhD thesis, University of Strathclyde, Energy Systems Research Unit.

Both, E. & Christiansen, G. (2002), *Termodynamik*, Den private Ingeniørfond, Danmarks Teknisk Universitet.

Box, G. & Jenkins, G. (1976), *Time series analysis - forecasting and control*, Holden-Day.

Chin-Hsiang, C. & Wen-Hsiung, H. (1990), 'Numerical prediction for laminar forced convection in parallel-plate channels with transverse fin arrays', *Pergamon* .

Christ, F. H. M. (2001), Comparison of measured and predicted data for pv hybrid systems, Master's thesis, University of Strathclyde, Department of Mechanical Engineering.

Davies, M. G. (2004), *Building Heat Transfer*, Wiley.

Deru, M. & Kirkpatrick, A. (2001), 'Ground-coupled heat and moisture transfer from buildings - part 1: Analysis and modeling', *National Renewable Energy Laboratory* .

Gandini, A. (n.d.), 'Tre logbook', Not published - private file from Hans Bloem.

Gandini, A., Mazzarella, L. & Bloem, J. J. (n.d.), 'Numerical analysis of pv double skin facades', Article.

Hansen, F. M. (1985), *Varmeoverfoerelse til bygninger*, IMSOR, Technical University of Denmark.

Häupl, P., Grunewald, J., Fechner, H. & Stopp, H. (1997), 'Coupled heat air and moisture transfer in building structures', *Pergamon* .

hyperphysics (2006), http://hyperphysics.phy-astr.gsu.edu/HBASE/thermo/heatra.html.

IEA (2006), www.iea-pvps.org.

Jiménez, M. J., Madsen, H. & Bloem, H. (2006), Estimation of non-linear continuous time models for the heat exchange dynamics of building integrated photovoltaic modules. .

Kecman, V. (1988), *State-Space Models of Lumped and Distributed System*, Springer-Verlag.

Kristensen, N., Madsen, H. & Jørgensen, S. B. (2003*a*), 'Parameter estimation in stochastic grey-box models', *Elsevier* .

Kristensen, N. R. & Madsen, H. (2003*a*), 'Continuous time stochastic modelling, ctsm 2.3 - mathematics guide', www2.imm.dtu.dk/ ctsm.

Kristensen, N. R. & Madsen, H. (2003*b*), 'Continuous time stochastic modelling, ctsm 2.3 - user's guide', www2.imm.dtu.dk/ ctsm.

Kristensen, N. R., Madsen, H. & Jørgensen, S. B. (2003*b*), 'A method for systematic improvement of stochastic grey-box models', *Elsevier* .

Laukamp, H., Bopp, G., Hille, G., Laukamp, H., Roth, W. & h. Schmidt (1998), 'Pv systems'. EUCREC Tutorial.

Liesen, R. & Pedersen, C. (1999), 'Modelling the energy effects of combined heat and mass transfer in building elements; part 1,theory', Article. University of Illinois.

Luque, A. & Hegedus, S. (2003), *Handbook of Photovoltaic Science and Engineering*, Wiley.

Madsen, H. (1985), Statistical determined dynamical models for climate processes - part 1+2, PhD thesis, IMSOR.

Madsen, H. (2001), *Time Series Analysis*, IMM.

Madsen, H. & Holst, J. (2000), *Modelling Non-Linear and Non-Stationary Time Series*, IMM.

Madsen, H., Nielsen, A., Saxhof, B. & Wittchen, K. (1994), Models for the heat dynamics of buildings. Laboratoriet for varmeisolering, DTU.

Mai, L., Infield, D., Eicker, U. & Fux, F. (200x), 'Thermal modelling of building integrated ventilated pv facades', Article. www.cibse.org/pdfs/facade.pdf.

Manz, H., Schaelin, A. & Simmler, H. (2004), 'Air flow patterns and thermal behavior of mechanically ventilated glass double facades', *Building and Environment* .

Markvart, T. e. a. (2000), *Solar electricity*, John wiley & Sons, ltd.

Melgaard, H. (1994), Identification of physical models, PhD thesis, Technical University of Denmark - IMM.

Nielsen, B. (1996), Stochastic Modelling for Control of the air Temperature in Greenhouses, PhD thesis, Ministry of Agriculture and Fisheries, Dansih Institute of Plant and Soil Science.

Prasad, D. & Snow, M. (2005), *Designing with Solar Power - A source book for Building integrated Photovoltaic (BiPV)*, Images Publishing.

TamizhMani, G., Ji, L., Tang, Y., Petacci, L. & Osterwald, C. (2003), 'Photovoltaic module thermal/wind performance: Long-term monitoring and model development for energy rating', *National Renewable Energy Laboratory* .

Technologies, S. (2006), 'Solar technologies', www.alaskasun.org/pdf/SolarTechnologies.pdf.

Troelsgaard, B. (1981), Statistisk bestemmelse af modeller for rumlufttemperatur, PhD thesis, IMSOR. In Danish.

Xiaoshu, L. (2002), 'Modelling of heat and moisture transfer in buildings - i. model program', *Elsevier* .