Het karakteriseren van residentieel energiegebruik voor warmte
op basis van slimmemeterdata

Characterisation of Residential Energy Use for Heating
Using Smart Meter Data

Eline Himpe

Promotor: prof. dr. ir.-architect A. Janssens
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: architectuur

UNIVERSITEIT
GENT

Vakgroep Architectuur en Stedenbouw
Voorzitter: prof. dr. ir.-architect A. Janssens
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2016 - 2017

ISBN 978-94-6355-021-5
NUR 961, 955
Wettelijk depot: D/2017/10.500/56

Promotor:
prof. dr. ir.-architect Arnold Janssens

Vakgroep Architectuur en Stedenbouw
Faculteit Ingenieurswetenschappen en Architectuur
Jozef Plateaustraat 22, B-9000 Gent, België
www.architectuur.ugent.be
Tel.: +32 9 264 37 42

Dit werk werd ondersteund door een doctoraatsbeurs
van het Bijzonder Onderzoeksfonds van de Universiteit Gent

Examencommissie:
prof. dr. ir. Luc Taerwe (Universiteit Gent, voorzitter)
prof. dr. ir.-arch. Arnold Janssens (Universiteit Gent, promotor)
prof. dr. ir. Michel De Paepe (Universiteit Gent, secretaris)
prof. dr. Peder Bacher (Danish Technological University)
prof. dr. ir.-arch. Staf Roels (KULeuven)
prof. dr. ir. Chris Develder (Universiteit Gent)
prof. dr. ir.-arch. Jelle Laverge (Universiteit Gent)
dr. ir.-arch. Marc Delghust (Universiteit Gent)

Dankwoord

De eerste woorden van dit boek zijn tevens de laatste waarmee ik mijn tijd als doctoranda afsluit. Een zeer intense periode was het: veelzijdig en divers, diepgaand en inspannend, verfrissend en vermoeiend, aarzelend en volhardend, nieuwsgierig en verwonderd, uitdagend en bemoedigend, kleurig en open, kritisch en lachwekkend, leerrijk en vol ondervinding... Dit woord van dank richt ik aan de mensen die deze naamwoorden hebben toegevoegd aan mijn doctoraatservaring.

Vooreerst dank ik mijn promotor die me na mijn studies de kans bood een doctoraat op te starten in zijn onderzoeksgroep. Hij gaf me de vrijheid en het vertrouwen om dit onderzoek te ondernemen en ondertussen via een aantal boeiende projecten zowel met de (bouw)praktijk als met fundamenteel onderzoek in contact te komen. Arnold, dankjewel! Dank aan de partners bij Eandis, om inzage te verlenen in de slimmemeterdata en me te betrekken bij jullie vragen rond energie feedback. Dank ook aan de mensen van het ECO-Life project voor de talrijke inzichten in de praktijk van het duurzaam bouwen en de fijne samenwerking. Thanks to the partners in IEA Annex 58 for the collaboration and training from the basics to the state-of-the-art on building thermal performance characterisation and dynamic modelling.

I would also like to sincerely thank the examination committee for their careful review of this manuscript and their valuable suggestions to improve it. Bovendien wil ik ook Riet De Baets bedanken voor het enthousiasme waarmee ze delen van dit boek van taalcorrecties voorzag!

Een gigantische dankjewel aan mijn collega's en de medewerkers van de vakgroep voor de hulp en steun in deze laatste lange maanden, om een oogje dicht te knijpen voor mijn andere taken en om mijn bedrukte blik steeds met een bemoedigende glimlach te beantwoorden. Dank allemaal voor de fijne en ongedwongen sfeer (en chocolaatjes) op kantoor, de Brugbabbels - van wetenschappelijk verantwoorde discussies tot hilarische gedachtenkronkels - en grensverleggende team-uitstapjes.

Een doctoraatsonderzoek laat je niet achter op kantoor... dank aan al mijn vriendinnen en vrienden, familie en kennissen, om er voor mij te zijn – ook als ik zelf wat afwezig was. Een kort berichtje, een lange wandeling, een verlossende balletles of een stevige knuffel, ze hebben me stuk voor stuk gesterkt! Een bijzondere dankjewel is bestemd voor mijn ouders, voor de zelfstandige studiehouding die ze stimuleerden, voor de talrijke kansen die ze me gegeven hebben en de vele vormen van steun bij het waarmaken ervan. Aan mijn broer, dank om mij in deze laatste maanden te vergezellen in een wedstrijdje 'hard werken' en onze prestaties tot ongekende hoogten te drijven. Aan oma meter, dank voor de ettelijke liters soep die me deze winter warm hielden! Tenslotte, Raphaël, jij bent het mooiste wat mij overkomen is tijdens deze periode. Dankjewel voor jouw liefde, zorg, geduld en bloemen – een krachtiger motivatie had ik niet kunnen wensen...

The future belongs to those who believe in the beauty of their dreams
(Eleanor Roosevelt)

21 juni 2017

Table of contents

Nomenclature

Hr	Relative Humidity	%
Pq	Precipitation quantity	mm
Q	Energy Use	kWh
Rg	Global horizontal solar radiation	W/m²
Te	Exterior Temperature	°C
Wd	Wind direction	°
Ws	Wind speed	m/s

Greek symbols

ε	Error term

Subscripts and superscripts

e	East
eq	equivalent
l	time lag
n	North
s	South
t	time
w	West

Acronyms

ACF	Autocorrelation Function
AR	Autoregressive term
ARX	Autoregressive models with exogenous inputs
ASW	Average Silhouette Width index

COOK	Cooking
CH	Calinski-Harabasz index
Cov	Covariance
CPGRAM	Cumulative periodogram
CVI	Cluster Validity Index
DB	Davies-Bouldin Index
DHW	Domestic Hot Water
DTRmul	Multiplicatively detrended data
DTW	Dynamic Time Warping (metric)
EN	Energy
ES	Energy Signature
EUC	Euclidean distance (metric)
HDD	Heating Degree Days
HDDeq	Equivalent Heating Degree Days
HS	Heating Season
Is	In-Sample
LM	Classical Linear Regression Model
LOESS	Local regression
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
Os	Out-of-Sample
pp	Percentage point(s)
PRECIP	Precipitation
RADGLOB	Global horizontal solar radiation
RC	Residual Component
RELHUM	Relative Humidity
RMSE	Root Mean Square Error

SC	Seasonal Component
SH	Space Heating
TC	Trend Component
TEMP	Exterior Temperature
TEMPeq	Equivalent Exterior Temperature
TSD	Time Series Decomposition
Var	Variance
VIF	Variance Inflation Factor
WINDDIR	Wind direction
WINDSPEED	Wind speed

Samenvatting

De reductie van het energiegebruik in gebouwen is een belangrijke opgave met het oog op het behalen van de klimaatdoelstellingen. De toepassing van de Europese richtlijn voor de energieprestatie van gebouwen leidt geleidelijk aan tot een toenemende energie-efficiëntie in zowel nieuwe als bestaande gebouwen. Ten minste: in theorie. In de ontwerpfase zijn energieprestatieberekeningen erg nuttig om het energiegebruik in te schatten, maar vaak wijken deze theoretische waarden aanzienlijk af van het reële energiegebruik in gebruiksomstandigheden. De werkelijkheid is immers een stuk complexer dan de theoretische standaardcondities: de uitvoering van de werken is niet altijd perfect, de regeling en werking van de technische installaties kan afwijken, de bewoners interageren met het gebouw en de installaties naargelang hun persoonlijke comfortnoden en activiteiten, die bovendien op hun beurt ook door energie-efficiëntie doelstellingen kunnen gedreven zijn. Om de reële impact van energie-efficiëntiemaatregelen voor een specifiek gebouw te kunnen inschatten, is het dus belangrijk om ook het reële gemeten energiegebruik te bestuderen.

Waar in het verleden het energiegebruik in gebouwen slechts jaarlijks of maandelijks werd genoteerd, worden vandaag steeds vaker slimme energiemeters geïnstalleerd. Bijgevolg wordt voor meer en meer gebouwen het energiegebruik iedere dag of ieder uur bijgehouden. Deze tijdreeksen van het energiegebruik zijn het energetisch eindresultaat van een systeem dat bestaat uit het geheel van het gebouw, de installaties en de gebruikers, en de interactie van dit systeem met omgevingsvoorwaarden zoals het weer - dat in het bijzonder het energiegebruik voor ruimteverwarming beïnvloedt - maar ook de kalender die een invloed heeft op gebruikersgedrag. Teneinde het energiegebruik van dit systeem te karakteriseren aan de hand van slimmemeterdata, moet het wiskundig beschreven worden in functie van deze omgevingsvoorwaarden. De resulterende verzameling parameters die het energiegebruik van het operationele gebouw karakteriseren, wordt ook wel de energiesignatuur van het systeem genoemd. Deze energiesignatuur is het vertrekpunt om het energiegebruik van een operationeel gebouw te vergelijken met dat in een andere periode of met dat van een ander gebouw. Ze laat ook toe om weersafhankelijke en niet-weersafhankelijke types energiegebruik van mekaar te onderscheiden of om een inschatting te maken van het energiegebruik onder specifieke omgevingsvoorwaarden. Energiesignaturen worden gebruikt in toepassingen zoals energie feedback, energie audits en commissioning.

Een eerste doelstelling van dit onderzoek is het evalueren en ontwikkelen van energiesignatuurmodellen die toelaten om het energiegebruik voor warmte in woningen nauwkeuriger te karakteriseren, gebruik makend van data met een tijdsstap van 1 dag. Een tweede doelstelling is het identificeren van energiegebruikspatronen die geobserveerd worden in energiegebruiksdata met een tijdstap korter dan 1 dag, en het evalueren of deze energiegebruikspatronen ook kunnen gebruikt worden om de energiesignatuurmodellen te verbeteren.

De tijdreeksen van het energiegebruik zijn gasmeterdata, per uur gemeten, van 25 woningen in België waar gas gebruikt wordt voor ruimteverwarming. De graaddagenmethode en de energiesignatuurmethode zijn klassieke modellen waarmee men op basis van fragmentarische meteropnames het energiegebruik kan modelleren in functie van het weer, met behulp van lineaire regressiemethodes. Een eerste onderzoeksvraag is of deze klassieke energiesignatuur-modellen (LM-modellen) nog statistisch correct zijn en nauwkeuriger worden wanneer dagelijkse in plaats van wekelijkse meetdata gebruikt worden. We stellen vast dat in de helft van de gevallen de LM-modellen statistisch niet correct zijn doordat de residuen gecorreleerd zijn in de tijd. Een logische stap om de LM-modellen te verbeteren brengt ons bij de tweede onderzoeksvraag, namelijk of de toevoeging van auto-regressieve variabelen aan het model leidt tot een betere modelkwaliteit. Het model wordt dan een ARX-model genoemd en er wordt vastgesteld dat deze modellen wel resulteren in een goede modelkwaliteit voor de meeste gevallen. Wanneer men vervolgens voor statistisch correcte LM- en ARX-modellen het gebruik van dagelijkse en wekelijkse data vergelijkt, kan worden besloten dat de coëfficiënten van de modellen met dagelijkse data tot drie maal nauwkeuriger zijn. De aanbeveling luidt daarom om gebruik te maken van ARX-modellen met dagelijkse data wanneer deze beschikbaar zijn. Verder worden in deze studie ook nog de belangrijkste variabelen voor deze modellen besproken, de nauwkeurigheid van de coëfficiënten en de voorspellingen, en de lengte van de meetperiode.

Tijdreeksen van het energiegebruik met tijdsstappen kleiner dan 1 dag (bv. 2 uur) vertonen vaak energiegebruikspatronen die een effect zijn van gebruikspatronen, gebouw- en installatiekarakteristieken, en die veel meer details prijsgeven dan dagelijkse data. Terwijl lineaire regressiemodellen helemaal niet geschikt zijn voor deze data, kunnen ARX-modellen sommige van deze energiegebruikspatronen wel opnemen. In dit onderzoek wordt echter een meer structurele identificatie van deze patronen beoogd. De derde onderzoeksvraag luidt dan ook hoe deze energiegebruikspatronen wiskundig geïdentificeerd kunnen worden. Er wordt een methode ontwikkeld die toelaat om de energiegebruikspatronen te clusteren op basis van hun gelijkvormigheid en gebruik makend van clusteranalyse. De lange tijdreeksen met patronen kunnen worden samengevat in enkele clusters van gelijkvormige patronen. Wanneer bovendien deze clusters worden afgebeeld in functie van weers- en kalendervariabelen, wordt vastgesteld dat heel wat patronen voorkomen voor specifieke weersomstandigheden of dagen van de week. Een vierde onderzoeksvraag betreft het wiskundig modelleren van de clusters van energiegebruikspatronen in functie van deze variabelen, zodat ook kan voorspeld worden welk patroon zal optreden in bepaalde omstandigheden. Hiervoor wordt een classificatiemethode toegepast, namelijk de multinomiale logistische regressie. We besluiten dat de clustering en classificatie van de energiegebruikspatronen toelaten om een beter inzicht te krijgen in het aan energie gerelateerd gedrag van het bewoonde gebouw. Het resultaat zijn woning-specifieke energiegebruikspatronen voor verschillende weersomstandigheden en verschillende dagen van de week, die onder meer ook het begin en einde van het stookseizoen aanduiden. Wanneer we terug uitzoomen naar dagelijkse energiegebruiksdata, dan kunnen de energiegebruikspatronen worden opgenomen in de LM- en ARX-modellen. De vijfde onderzoeksvraag is of deze extra cluster variabele(n) toelaten om de energiesignatuurmodellen verder te verbeteren. Er wordt vastgesteld dat in vergelijking met de klassieke LM-modellen de determinatiecoëfficiënt gemiddeld stijgt van 0.76 naar 0.85 en de modelkwaliteit goed is. Met de resulterende modellen kan het energiegebruik dat samenhangt met specifieke energiegebruikspatronen worden geanalyseerd.

De toepassing van energiesignaturen wordt uitgewerkt aan de hand van een voorbeeldtoepassing, namelijk de vergelijking van het energiegebruik voor verschillende periodes, bijvoorbeeld voor toepassing in energie feedback. Er wordt besloten dat aan de hand van de energiegebruikspatronen en de regressiemodellen, veranderingen in het energiegebruik beter gevonden en geduid kunnen worden. Tot slot worden de ontwikkelde modellen ook geïllustreerd voor twee datasets waarin de energie voor zowel ruimteverwarming als warm water gebruikt wordt.

Summary

Keeping climate change actions in mind, the reduction of energy use in buildings is a matter of growing interest. The implementation of the Energy Performance of Buildings Directive leads to an increasing energy-efficiency in both new and refurbished buildings, at least in theory. While calculations of the building energy use in the design stage are useful to estimate the building energy performance in reference conditions, these theoretically estimated energy use figures often differ significantly from the actual measured energy use of the occupied building. In reality, construction deficiencies may occur, the actual operation and control of the building services may differ, and the building users interact with the building envelope and services according to their personal comfort needs and activities, which in turn may also be driven by energy-efficiency objectives. In order to estimate the actual and case-specific impact of energy-efficiency measures, there is an urgent need to better understand the actual energy use in occupied buildings.

The application of energy monitoring and smart metering systems leads to an increasing availability of frequent, hourly or daily, and long-term measurements of the actual energy use in occupied buildings whereas in the past only monthly or yearly meter readings were collected. These energy use time series are the outcome of the system consisting of building, building services and occupants (the 'occupied building'), and the interaction of this system with ambient conditions such as the weather, which influences especially the energy use for space heating, but also the calendar, which may influence user behaviour. In order to characterise the energy use of the system starting from the measured data, the challenge is to describe it mathematically in function of these ambient conditions. The resulting set of parameters that characterise the energy use of the system is called the Energy Signature of the system. It is the basis for comparing the energy use of an occupied building in different periods, for comparing different occupied buildings, for separating weather-dependent and weather-independent types of energy use, or for estimating the energy use for a given set of ambient conditions. Energy Signatures can therefore be applied in energy feedback, energy auditing or commissioning.

When occasional energy use measurements and weather data are available, widely applied data-driven models are the Heating Degree Day and Energy Signature methods, using basic linear regression techniques. However with the availability of frequent energy monitoring data, the question rises whether data at daily or sub-daily time resolution allows to improve Energy Signature models. Eventually the sub-daily energy use time series often display energy use time patterns that reflect the effects of behavioural patterns, building and services characteristics in more detail than aggregated measurements do. Other questions are whether these patterns can be mathematically identified and characterised in function of weather or calendar variables, and whether they allow to improve the Energy Signature models with daily data.

The overall objective of this dissertation is to evaluate and develop statistically sound Energy Signature models that allow for a more accurate characterisation of the energy use for heating in residential buildings. The energy use time series in focus are hourly gas meter readings from 25 Belgian dwellings, where gas is used mainly for space heating purposes. A first question is whether the Energy Signature models using classical linear regression techniques (LM-models) are valid and improved when using daily instead of weekly energy use data. It is found that in half of the cases the LM-models are statistically invalid because of autocorrelated residuals. A logical next step leads to the question whether these models are improved by adding auto-regressive terms to the LM-model, which is then called an ARX-model. It is found that the ARX-models result in an overall model validity for the majority of the cases. When comparing valid LM- or ARX-models with daily and weekly data, it is found that the coefficient estimates for the models with daily data are up to 3 times more accurate. It is therefore recommended to use daily data when available and apply ARX-models to characterise the Energy Signatures. Furthermore, the dissertation also investigates the key variables in the model, the accuracy of the model coefficients and predictions and the length of the training period for the model.

In measurements of the energy use at a sub-daily time interval, the diurnal fluctuations in energy use are observed and patterns are recognised. These are called energy use time patterns. While LM-models are invalid for these data, ARX-models deal with some of the energy use time patterns in their auto-regressive terms. The mathematical recognition of the visually observed groups of energy use time patterns in the data is the 3rd research question. A procedure is developed that allows to cluster the energy use time patterns based on their similarity in shape, using cluster analysis techniques with dynamic time warping distance measures. As a result, the long series of patterns are summarised in a limited number of groups. When graphically presenting the clusters of patterns in function of weather and calendar variables, it is found that many of the patterns can be related to specific calendar or weather conditions and allow a more profound insight in the energy use related behaviour of the occupied building. The fourth question is how these clusters of patterns can be mathematically modelled in function of those variables, so they can be used for prediction purposes. A classification procedure using multinomial logistic regression models is used for this purpose. Zooming out to daily data again, the fifth research question is whether the inclusion of these clusters of variables into LM- and ARX-models allows to improve the Energy Signatures with daily aggregated data. It is found that in comparison to classical LM-models, the goodness-of-fit of the LM- and ARX-models with cluster variables increases from 0.76 to 0.85 on average and the models are statistically valid. In the resulting models the energy use associated with the different energy use patterns can be estimated and compared.

One possible implementation of Energy Signatures is the comparison of the energy use for different periods, for detecting changes in energy use in energy feedback applications. It is concluded that using the regression models and the energy use time patterns, changes in energy use can be better identified and understood. Finally, the developed models are illustrated for cases with combined energy use for space heating and domestic hot water generation.

1

Introduction

Reduction of energy use in buildings is a major challenge faced by society today. While calculations of building energy use have become a general practice when designing or refurbishing buildings, the characterisation of the actual energy use in operational buildings – especially in residential buildings - is less common. The increasing application of smart energy meters offers opportunities for the long-term monitoring of actual energy use for specific buildings and their users. Though it requires modelling to characterise the operational energy use from measurement data, that is: to mathematically describe the energy use of the building in function of external variables such as weather or calendar variables. This modelling allows to compare or estimate the energy use in different conditions, for applications in energy use feedback, normalisation, prediction, auditing etc.

The introductory chapter outlines the general context of this dissertation and the specific background of data-driven building energy use modelling. Next, it expands the research objectives and the approach and finally ends up by explaining the structure of the study.

1.1 Building energy performance assessment

1.1.1 Energy use reduction

In the context of climate change actions, increasing energy efficiency and reducing non-renewable energy use is vital for the reduction of greenhouse gas emissions [1]. In the European Union, where about 40% of the energy use originates from buildings, the increase in building energy efficiency constitutes an important measure for reaching greenhouse gas reduction targets. This is reflected in the Energy Performance of Building Directive (EPBD) and the Energy Efficiency Directive, where an increasing energy-efficiency and use of renewable energy is envisaged for both new and existing buildings [2,3]. In the Flemish region of Belgium, the implementation of these directives and the actions taken are described in the Energy Decree of the Flemish Government and the Flemish Energy Efficiency action plans [4,5]. Actions taken to improve energy-efficiency or increase the share of renewable energy in buildings are amongst others: the installation of minimum energy performance levels for new and renovated buildings with building permit, requirements for energy performance certificates and energy audits, the implementation of smart energy meters, public energy-efficiency information programs or financial incentives for energy saving investments. These actions encourage the adoption of a wide spectrum of measures that can be roughly classified in three groups depending on the component the intervention applies to: (1) building construction (e.g. insulation of the building envelope), (2) building services (e.g. application of efficient HVAC) and (3) the use of the building by its occupants (e.g. adjusting indoor temperature settings).

1.1.2 Energy use calculations

In order to identify opportunities for reducing energy use and to assess the energy use and the impact of energy-efficiency measures, estimations of the building energy use are useful, if not required. Often energy use is calculated by use of a mathematical model of the building that is based on the laws of physics, with building data such as the building geometry and material properties, followed by simulating the energy used in predefined boundary conditions such as the weather conditions and user profiles. The approach is called *forward modelling*, and the models are called *descriptive* or *calculation* or *white-box* models [6]. The probably most well-known examples are the simplified calculation methods used in implementations of the EPBD in the European countries and the European standards they are based upon [4,7–9]. These models, developed for energy performance rating, are useful in the design stage, to assess the theoretical performance of the building in reference conditions for legislative purposes, and – to a certain extent – to assess design options and energy-efficiency measures for design purposes. However, these theoretical energy use figures often differ significantly from the actual energy use of the operational building [10–12]. First, the physical models, algorithms, assumptions or inputs used may be insufficient to represent the real complexity of building energy use [13,14]. Secondly, deficiencies may be introduced during the construction phase and the actual control and operation of the building services may differ from the design specifications. Moreover, initial commissioning activities may be insufficiently performed and are usually not continued during the operational phase [15,16]. Thirdly, the weather conditions vary from standard conditions and building users interact with the building envelope and services according to their personal comfort needs and activities, and thus differ from a standard user profile.

For understanding the complexity of the actual energy use, it is important to be aware of the interplay between the building envelope, services, users and the weather conditions. Studies show that, with regard to residential space heating energy use, (1) the measured energy use in energy inefficient houses tends to be lower than calculated according to the EPBD-methods, and the gap increases with decreasing insulation level of the house. (2) For houses with a high insulation level, the measured energy use tends to be higher than calculated, and (3) the average temperature of the house tends to increase with increasing insulation levels [10,11,14,17]. Therefore, calculations of the impact of energy-efficiency measures can be misleading. Possible reasons for this are of different natures (1). One is related to the building envelope and services: the physical temperature take-back, or the slower temperature decay in insulated houses, and therefore higher average temperatures in both heated and unheated rooms (2). Another is merely related to the building services and users: for example in houses with central heating systems (that are often more recent or refurbished, and better insulated) more rooms are heated than in houses with a local heating system, probably because this is the most obvious way of controlling the heating system, and (3) related to the users. It is called direct economic rebound, or the tendency to a less parsimonious heat use as the overall heating cost decreases. This interplay significantly complicates the descriptive modelling of the actual energy use. Attempting to close the gap between calculated and real energy use, various approaches to improve the results from calculation models have been proposed, including the application of curve fitting factors to account for the average change in gap with changing insulation level, the use of more advanced physical models and calculation algorithms, and the diversification of user profiles and inputs [10,11,18].

1.1.3 From design phase to operational phase

When using descriptive models, a rather extensive inventory of building and user data is needed to feed the model. This is a time-consuming task and often requires the intervention of a building specialist. When designing or refurbishing a building, architects and engineers are involved and most of these data are collected anyway. So the forward modelling approach is considered advantageous. However, energy is used during the whole operational lifetime of a building, and the energy use continues to change - possibly but not necessarily - due to energy-efficiency measures. For example, occupants or their presence and activities in the building may change, they may decide to a more conscious energy-related behaviour, or the insulation level may be changed and the building services may be replaced. During the operational lifetime, knowledge of the evolution in energy use is valuable, to ascertain the actual energy use and the actual impact of energy-efficiency measures. It is a basis for making decisions on enhancing the energy efficiency and to stimulate energy-conscious occupant behaviour [19–22]. The application in view is then not the official certification of the building apart from its actual occupants. Rather than that, the actual energy use is assessed for applications such as energy feedback, energy auditing and commissioning. Energy feedback can be used for keeping occupants informed about the actual (changes in) energy use via a variety of media such as letters, web tools or in-home devices. Confronting people with their energy use and confirming their energy savings, is helpful in awareness raising and in stimulating and maintaining changes in energy-related user behaviour [19]. Energy auditing is the inspection of the building energy flows by an energy expert, architect or engineer, for providing the building owner with personalised advice about energy-efficiency measures, usually preceding the building refurbishment. Commissioning is the

examination of the compliance of the performance of buildings or building services with the design or operational targets, which can take place after installation and throughout the operational lifetime [15,23]. During the operational lifetime of the building, the collection of building and user data and the set-up of descriptive models become an obstacle for they are time-consuming tasks that require a specialist, next to the challenge of accurately approximating the actual energy use flows. That is why energy use measurements are proposed as a basis for the actual energy performance assessments of operational buildings in use [24].

1.1.4 From calculations to measurements

Where in the past only yearly, or perhaps monthly, energy meter readings were collected, the application of energy monitoring and smart metering systems leads to an increasing availability of frequent, hourly or daily, and long-term measurements of the actual energy use in occupied buildings [5,25–27]. These energy use time series are the actual outcome of the interplay between the building construction, services, users and weather conditions. Therefore the raw energy use measurements are often used directly in for example energy feedback applications, recent measurements being represented in comparison to historical measurements [28]. However, when looking for feedback on the impact of energy-efficiency measures, an interpretation of the raw data is likely to be compromised by differences in weather conditions. For example, when estimating the energy savings due to the replacement of the gas boiler of a house, one compares the energy use in periods before and after the replacement. Usually the weather conditions in these two periods are different, which makes it hard to decide whether a change in the actual energy use is due to the boiler replacement and/or to the different weather conditions. This is especially the case when the energy use includes energy used for space heating. The European Standard EN15603 specifies that weather corrections are needed if the time period for the assessment is shorter than 3 years [8]. Moreover, energy use meters are typically applied per energy carrier, for billing purposes, and not per energy use function. For example, gas and heat meters often deliver energy for both space heating and domestic hot water use. Using unprocessed data it is uncertain which energy use function is the cause for a change in energy use or which one is more decisive for the energy use of a specific building. For example in energy auditing applications, the separation of energy use for space heating and domestic hot water is useful to decide which type of energy-efficiency measures may lead to the highest energy use reductions in a specific building.

In conclusion, when applying energy use measurements, the challenges are (1) to model the energy use measurements mathematically in function of the weather conditions, and (2) to disaggregate its different constituents. Models that are based on measurement data are called *data-driven* or *experimental* models. Two types can be distinguished. A first modelling approach is called black-box modelling: the model structure and the model parameters are identified from the data only. It is opposed to white-box modelling. In between these two extreme approaches lies a range of hybrid or grey-box models that use a mathematical model of the building (based on physical laws) for which the parameters are identified by fitting the model to the measurements [6].

1.1.5 Energy Signature models

Energy Signature models are widely applied data-driven models that express the heating energy use in function of weather variables. They have been applied at least since the early fifties and are still widely applied today for normalising the heating energy use with respect to the weather conditions [24,29,30]. The measurements consist of heat meter or gas meter readings and weather data typically at monthly or weekly time intervals. Linear regression techniques are typically applied for estimating the parameters of the expression. In a simple case, the energy use Q is expressed as a function of the heating degree days (*HDD*) following a simple energy-balance model:

$$Q = c_1 + c_2 \times HDD + \varepsilon_t \qquad [1\text{-}I]$$

The equation can be visualised in a 2-dimensional graph (Figure 1-1(a)) and informs about the relation of energy use with the heating degree days (slope of the line, coefficient c_2) and the energy use at zero heating degree days (coefficient c_1), which informs about weather-independent energy use. In Figure 1-1 this is illustrated for an example case where weekly measurements of the energy use for space heating in a house are gathered for an entire heating season. If the statistical quality of the regression model is acceptable, the coefficients of the Energy Signature model can be interpreted. The regression line in Figure 1-1(a) presents the relationship between the energy use and HDD and 95% confidence intervals of the mean for this relationship. Assuming that the base temperature for the heating degree day calculation is estimated correctly, the intercept with the y-axis (coefficient c_1) refers to the weather-independent part of the energy use. In this case, where only space heating energy use is considered, this is not significantly different from zero. In Figure 1-1(b), the 95% prediction intervals for the relationship are shown, that are used when predicting the energy use for given HDD-values, in context of energy use predictions or normalisation. Figure 1-1(c) is a fictitious example of the comparison of the Energy Signature models for two different houses. For one house (in blue) the weather-independent energy use is clearly higher than for the other house (in green), but the slope of the regression line is smaller, indicating that on average the weather-dependent part of the energy use is smaller, for example because the house (in blue) is better insulated. Figure 1-1 (d) is a fictitious example of how the Energy Signature of a house indicates that the space heating energy use decreases after a refurbishment.

In general, more complex energy-balance models can be derived from the building energy balance equations, including multiple input variables that are related to the weather or to other influential phenomena. The parameters are estimated from statistical analysis of the measurement data, typically using linear regression methods. According to Hammarsten, the Energy Signature of a building then denotes a set of parameters that describe the energy performance of the building for a given energy-balance model [29]. Note that the term *Energy Signature* is often attributed to one specific parameter of the expression, that is the coefficient of the temperature difference input variable (coefficient c_2 in equation 1-I), and is then also referred to as the building heat loss coefficient [31]. Notice that in this dissertation, the term Energy Signature refers to the complete set of parameters estimated for a certain energy use expression (coefficients c_1 and c_2 in equation 1-I).

Energy Signature models are the most well-known examples of data-driven energy use models that are typically applied when occasional meter readings of the gas or heat use are available. However when evolving from occasional to frequent (daily, hourly) meter readings, it is questioned whether these models are still applicable and whether they can be improved, e.g. for providing more accurate estimates of the Energy Signature coefficients, or for extending the Energy Signatures with extra coefficients, allowing to identify other characteristics of the energy use and thus to withdraw more information out of the data. This leads to the research objectives of this dissertation, that are found in section 1.3. Apart from the Energy Signature models, many other types of data-driven building energy use models have been developed. They are discussed more profoundly in section 1.2.

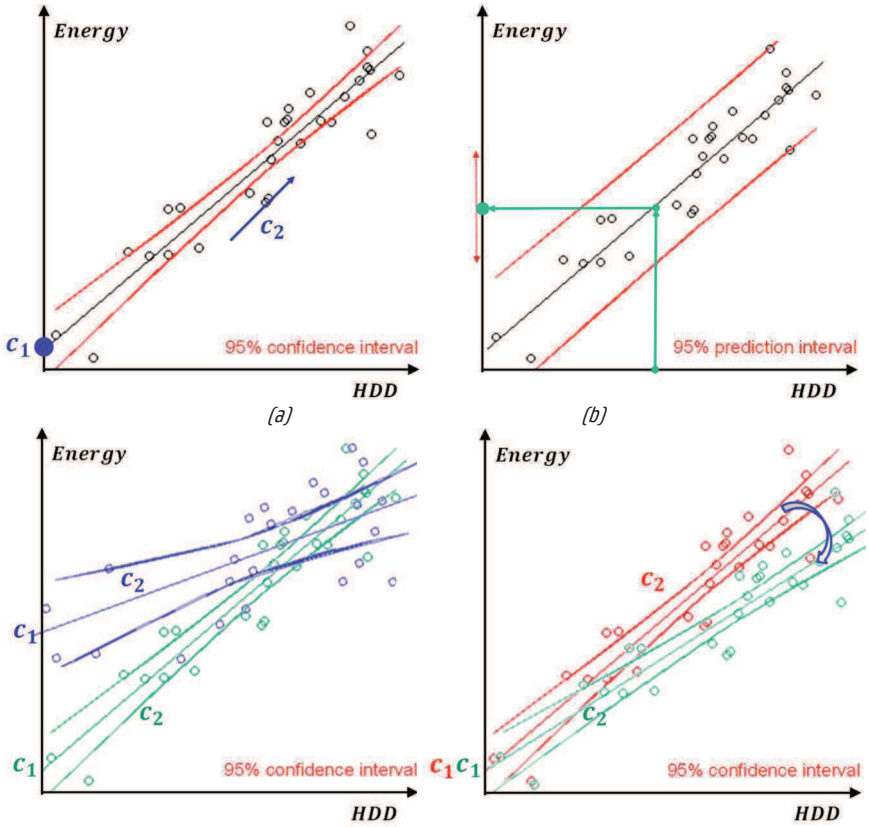


Figure 1-1: Energy Signature of a house – Example

1.2 Data-driven energy use modelling

A variety of data-driven building energy use models exists. In order to organise the different approaches, look at the structure in Figure 1-2. It is a modification of the well-established framework for Life Cycle Assessment as described in ISO14040, of which the general principles are found useful in the context of energy performance assessment [32,33]. The main elements of an assessment method are the definition of the goal and the scope of the assessment, the experiment and data collection, and models used in the assessment. These elements are interrelated, and related to possible interpretation and findings. Finally, there is a relation between the assessment method and possible applications.

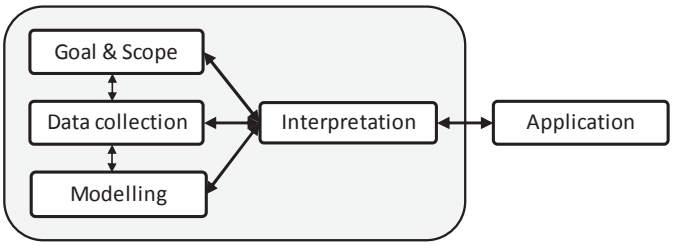


Figure 1-2: Energy Performance Assessment Framework

The goal and scope definition determines the objective(s) of the assessment and the (boundaries of) the system that is assessed. It also defines the assumptions, conditions and requirements for the experiment or data-collection and for the modelling. In section 1.2.1 different objectives of building energy performance assessment are distinguished. In the data collection the experiment is implemented and data are collected, validated, converted and allocated to the correct system processes. From these input data the outputs of the assessment are calculated, which can be interpreted keeping the goal & scope, the data collection and modelling assumptions in mind. Section 1.2.2 deals with several groups of models used for these calculations.

1.2.1 Data-driven energy use modelling: different objectives

Building energy performance assessment is a common phrasing that encompasses different types of assessment relating to different goal and scope definitions and different applications. In the European Standard EN15603 on the overall energy performance of buildings, the building is defined as a system composed of the building construction and envelope, the building services and the users. The energy sources are outside the system boundaries [8]. Because buildings do not use energy per se, but for creating an agreeable indoor environment for their occupants (through space heating, ventilation, lighting), and for sustaining the activities of the user in the building (e.g. bathing, cooking, using appliances…), the user is included in the system boundaries. A first and major possible objective of building energy performance assessment is design energy rating, for the purposes of judging its compliance with the building regulations or design requirements, or for certification. The goal is to assess the characteristics of the building without considering the use of the building by real occupants, and to compare them to benchmarks. In experimental models, the estimation of the characteristics is referred to as the characterisation or identification of the physical performance characteristics of the building. A well-known

example is the identification of the overall heat loss coefficient and the global solar aperture coefficient of the building using a co-heating experiment [34]. During the experimental data-collection the building is unoccupied and sealed and a quasi-stationary homogeneous heating experiment is performed. The models used are similar to the Energy Signature models, and the identified characteristics are derived from the coefficients of the energy use expression. Central is the physical interpretability of the estimated characteristic for comparison with theoretical benchmark values. This should be done with great caution, since the coefficients may be biased [29,31,35,36]. Consequently, influential processes such as the weather and user behaviour should be quantified and included in the modelling step, or they should be excluded from the experiment, which explains the highly specified experimental test set-up and the predictor variables used in the models. Next to the co-heating test, there are a range of static and/or dynamic test protocols that allow to identify characteristics of the dynamic behaviour of the building.

Another main objective of building energy performance assessment is the assessment of the energy use of the occupied building in operational conditions, comparing it to a benchmark, such as the energy use in a previous period, in a similar building, a target energy use or theoretically estimated energy use. Therefore, the characteristics of the energy use of the building in use are to be identified. The users are thus not replaced by a test protocol, but fully interact with the building envelope, services and appliances, when the measurements take place. In these conditions, classical Energy Signature models are typically applied if the measurements include space heating energy use, and then the identified characteristics depict the relationship between the energy use and weather variables for the specific (average) use conditions. However, it is important to remark that even if the models are quite similar to those used for identifying physical performance characteristics, the identified parameters cannot directly be compared to theoretical design characteristics because they are influenced by the actual use of the building. Secondly, since the energy use of the operational building is not limited to space heating energy use and the remaining energy use functions are often less related to the weather and more to the user, the classical energy signature models fall short in identifying their constituents. Note that many other objectives require building energy use modelling, for example the objective of predicting the energy use for application in model predictive control algorithms or short-term energy feedback. In these conditions the parameter estimates and even the model shape are inferior to the prediction quality of the model, which can also be black-box [37,38].

1.2.2 Data-driven energy use modelling: different models

A building is a dynamic system: due to variations in the weather and the use of the building and services, the state of the system changes through time. For example, when turning off the space heating, the interior temperature will slowly decrease. Söderström defines a dynamic system as a system in which the control action at a given time will influence the output at future time instants [39]. Frequent measurements of this system will thus also depend on each other in time. For modelling the system described by these measurements, in principle dynamic models are needed. In buildings, when focussing on the energy use for heating, a lot of dynamics appear at periods shorter than one day, for example depending on the daily weather cycle or on the occupational use cycle. Then there are dynamics at timescales beyond than one day, for example due to the thermal inertia of the building or cycles of use patterns (e.g. weekly patterns).

Static models, such as the Energy Signature models described in section 1.1.5, do not account for the time-dependency of the data. Therefore, they only provide information on the steady-state characteristics and they can only be applied when the time interval of the measurements is sufficiently long to erase the dynamics of the system and assume a static behaviour. In general, it is recommended to aggregate the data to at least one day, but some authors recommend longer aggregation periods. When using time intervals shorter than one day, the use of dynamic models is recommended. They allow a more optimal use of the data and insights in the dynamic characteristics of the system. In the IEA EBC Annex 58 project on the characterisation of physical performance characteristics of buildings, guidelines for the data-driven modelling of the building energy use were developed, starting from the static linear regression models, via the dynamic ARX-models that can be used in stationary and linear models, towards the more complex continuous-discrete time state space models [40]. The following paragraphs provide an overview of the main properties of these models using examples from literature. The focus is on the modelling of the heating energy use, that is mainly the energy use for space heating and possibly also for domestic hot water generation. Additionally, some models that allow for the separation of distinct components of time series are summarised. For these models it is found valuable to look at literature from other fields, such as the analysis of electricity use data.

Classical Linear Regression models

Classical Energy Signature models were mainly investigated in the eighties and nineties. Hammarsten provides a theoretical discussion of the various physical and statistical error sources and the requirements that need to be fulfilled for different applications of the energy signature models [29]. Meier investigates the relationship between the measurement period and prediction accuracy [36], Flouquet demonstrates the importance of adding solar radiation as a model predictor to prevent bias in models with weekly data [31], and Reddy & Claridge detect that principal component analysis can help to overcome multicollinearity between the predictor variables but only in certain situations, when the goodness-of-fit of the linear regression model is poor ànd the correlations between predictor variables is high [41]. The three of them use synthetic data-sets. Wouters and Loncour review the experimental test protocols and models used in various projects from that time period [24]. In the SAVE HELP and STEM&PSTAR project, identification of the main thermal performance characteristics of a building was obtained using very strict heating experiments and quality assurance tests during respectively 3 weeks and 3 days period in unoccupied buildings. In occupied buildings however, strict experiments are not feasible and the results were uncertain and inaccurate. In occupied buildings, Westergren et al. recommend the use of weekly measurements during the heating season when applying static models, because they find that the dynamic properties of houses are not averaged out in a 24h period. They recommend 7 to 10 weeks of measurement duration, and in their experiments, which were done in winter and in the beginning of spring, allowing for enough variations in outdoor temperature and solar radiation [42]. Ghiaus deals with the application of linear regression on sub-daily energy use measurements in an occupied school building by sub-setting only school-time data, thus selecting periods with similar use profiles, and applying robust regression on the sub-daily data between the 1st and 3rd quartile if the variables have the same distribution [43]. This brings us to the pseudo-dynamic regression models for determination of the thermal characteristics of the building, in which a dynamic correction is added to the regression. Bloem applies linear regression on daily heating energy use data with the temperature difference and

solar radiation as predictor variables, after correcting the heating energy use for the estimated ventilation losses and energy stored in the (occupied) building [44,45]. Danov et al. apply dynamic solar gain and heat gain corrections on space heating energy use data, that are estimated respectively from solar radiation data and from electricity and occupancy data [46]. Finally, Mortensen and Nielsen use linear regression energy signature models in a more dynamic way by estimating the coefficients for limited time windows of the daily or 4-hourly energy use time series, thus allowing the coefficients to vary slowly over time [47]. With regard to the weather predictor variables in linear regression models, exterior temperature is the main input, and solar radiation is usually the second important input added. Wind speed variables are included in some models. In some studies they are found to be significant, in others they aren't [42,47].

Auto-regressive models with exogenous inputs

Auto-regressive models with exogenous inputs (ARX-models) are dynamic linear regression type of models that can be estimated using the same methods of classical linear regression, for example the ordinary least squares method. In addition to the (weather) input variables ('exogenous inputs'), time lags of the output variables are added as input variables in the model. These are called auto-regressive input variables and they are used to deal with the dynamical properties of the system [48]. In the IEA EBC Annex 58 project, ARX-models were used to identify the thermal characteristics (overall heat loss coefficient and solar aperture coefficient) of a test box and a test house subject to static and dynamic heating experiments [49–51]. Westergren et al. apply ARX-models on hourly measurements in occupied dwellings. The energy use is modelled in function of the interior and exterior temperature difference and solar radiation, and the first and second time lag of all input and output variables. When comparing to classical regression models with weekly data, it turns out that the estimated energy use for a full heating season is more accurate (smaller confidence intervals) [42]. ARX-models are also successfully applied for real-time forecasting purposes, for example in Soldo et al. for forecasting one-day ahead daily natural gas use [52].

Continuous-discrete time state-space models

The continuous-discrete time state-space models are a more complex class of models that are dynamic models for modelling both linear and non-linear, stationary and time-varying phenomena. The physical models are described in the so-called system equations that describe the dynamics of the states of the system in continuous time by use of stochastic differential equations. When the system is a building or building component, linear RC-network models can be used for the system equations. On the other hand, the measurement equations describe how the discrete-time measurements are linked to the states of the system. The parameters of the model can be estimated using for example the maximum likelihood function. In these grey-box models, the internal physical processes in the system can be described in much more detail than in the ARX-models, which describe the external relations between the inputs and outputs of the system [40,48]. For this reason the ARX-models are often categorised as black-box models, however this author argues that often the structure of the ARX-model is also based on a (simpler) physical model, e.g. the heat-balance equation of a building. Bacher and Madsen identify a grey-box state space model for an unoccupied building that is subject to a dynamic heating experiment for a period of 6 days. The measurements used are 5 minute measurements of the interior and

exterior temperature, solar irradiance and energy use. The estimated parameters include the overall thermal resistance, capacitance and solar aperture of the building [53]. Delff et al. identified a similar model for an unoccupied arctic house where a 16 days dynamic heating experiment was performed [54,55]. In the IEA EBC Annex 58 project and the work of Bauwens, grey-box models as well as ARX-models were identified for unoccupied test buildings subject to dynamic heating experiments for identification of the overall building heat loss coefficient. Reliable results can be obtained using ARX-models and an experiment duration of about 15 days, but using grey-box models reliable and more accurate results can be obtained with shorter experiment durations (about 5 days). However it is noted that the grey-box models have difficulties to identify heat losses to thermostatic controlled boundary zones, and thus a conscious experimental design should make sure that all parts of the system are excited [49,51,56].

Time series decomposition approaches

The basic assumption of the previous models is the relationship between space heating energy demand and weather variables, which is observed most purely in a co-heating experiment. In dynamic experiments, the heating demand is also influenced by the control of the heating system, and in occupied buildings also the interaction of the user with the building is added. Additionally, other types of energy use, such as the energy use for domestic hot water - which is often measured in aggregation with the space heating energy use - are less influenced by weather conditions and more by user behaviour. Disaggregation of the measured energy use in its different constituents has been done in the context of forecasting and data-mining. De Saint-Aubain separates domestic hot water use from space heating energy use applying smoothing techniques, after the assumption that domestic hot water demand causes large spikes in the time series [57,58]. Kilpatrick et al. filtered the domestic electrical demand into different appliance categories (stand-by, cold, heating element spikes and residual demand) also based on the characteristics such as a constant, cyclic or peaking pattern [59]. Basu et al. disaggregate household electricity use into different types of appliances (e.g. electric oven, washing machine, water heater…) using event detection, appliance signature generation and pattern recognition techniques [60]. For both methods, high frequency data with time steps of respectively 10 minutes, 1 minute and 10 minutes, are required, since the properties they are based on, such as spikes and on/off-events, average out in longer time intervals. Another problem is the recognition and classification of diurnal profiles in time series. Huebner et al. apply cluster analysis methods for identifying diurnal interior temperature profiles in living rooms, and so does Aerts for identifying daily residential occupancy patterns [61–63]. Yang and Becerik-Gerber explore different methods for obtaining personalized specific occupancy profiles from measurements of CO2-concentration, door status, interior temperature etc. [64]. The modelling of diurnal patterns in energy use time-series is also used for improving the accuracy of forecasting models. Bacher et al. apply an adaptive linear grey-box model for prediction of the hourly space heating energy use 2 days ahead. In this model, fourier series are used to model the diurnal patterns for weekdays and weekend days [65]. Seem and Braun apply CMAC models (updated look-up tables) to account for the deterministic part of electricity use forecasting, and autoregressive models for the stochastic part of the model [66]. Mestekemper et al. forecast the diurnal aggregated district heating and electricity demand and river water temperatures up to one week ahead, using semiparametric regression smoothing for the deterministic model part, combined with

multivariate time series models for the stochastic part of the problem. In the semiparametric regression smoothing approach, for each hour of the day a regression model is made, allowing for calendar (e.g. day of the week) and weather predictor variables (e.g. exterior temperature) [67,68].

Findings

When characterising the residential energy use for heating using smart meter data, measurements of the energy use of occupied houses are typically available at hourly time steps. Hourly measurements of weather variables, measured at the national or regional weather stations can usually be obtained via the national meteorological institute, if no on-site weather station is available. However measurements of other variables, such as the interior temperature in the building, or information about the building or users, is usually not available, nor is it desirable to perform heating experiments. When applying classical linear regression models on this type of data, in literature various methods were found to deal with the dynamics of the system: for example by aggregating the data to weekly values, or by pre-processing the data through a selection of periods with similar use profiles or applying dynamic corrections. The first option has the drawback that the amount of data points is reduced dramatically and thus on beforehand the model accuracy is reduced. Moreover, for obtaining enough data points, the length of the measurement period needs to be sufficiently long. For the other options more information or inputs are needed, e.g. interior temperature measurements or use profiles. The ARX-models have the benefit that the dynamics in the data can be caught with data at daily or hourly time steps. It is noticed that in most literature examples also the interior temperature is available as a model input. The continuous-discrete time state-space models in general allow for a further increase in model accuracy, reduction in measurement period and more specific parameters to be estimated in comparison to the ARX-models. On the other hand, more different input data series were needed (e.g. interior temperature measurements, heating control…), even if the buildings were unoccupied. Finally, the time series decomposition methods use a different approach to disentangle time series, starting from the patterns that are observed in the data, rather than the relationship with external variables. Their applicability is related to the time step at which certain patterns in the data can be observed. For applications such as the disaggregation of space heating and direct domestic hot water use generation based on spikes in the series, a time step much smaller than 1 hour is needed. On the other hand, for identifying diurnal patterns in the data, hourly time steps are sufficient.

1.3 Research Objectives

The increasing availability of smart energy meters in residential buildings creates the opportunity to assess their actual energy use in operational phase based on high frequency (daily or hourly) and long-term measurement data. Energy Signature models are widely applied data-driven models that allow weather-normalisation of heating energy use measurements using classical linear regression techniques. Apart from the energy use data and weather data, no extra measurements or information on the building are necessary. Therefore these models are applicable in many situations, also when no detailed building data are available. However, in traditional applications only occasional meter readings are available (e.g. monthly or weekly). For these time intervals it may be assumed that the short-term dynamics of the building and weather are averaged out and a steady-state model is sufficient. When changing towards high frequency data, this assumption may be questioned and the applicability of the classical Energy Signature models is required. Therefore, a first research objective is to evaluate the use of classical Energy Signature models with daily and sub-daily data:

1. **Does data at daily time resolution allow to improve Energy Signature models applying classical linear regression techniques?**

This main question breaks up in the following sub-questions:

- Are these models valid from statistical point of view?
- What are the key variables in the model?
- What is the accuracy of the estimated coefficients and of the predictions?
- How long does the measurement period need to be?
- What if data at sub-daily time resolution is used?

While classical linear regression models are widely applied steady-state models, the 'state of the art' for analysing energy use time series are dynamic models. For the continuous-discrete time state-space models additional measurements (e.g. interior temperature) are needed for estimation of the parameters, but this may be avoided when using ARX-models. The second research objective is to evaluate the use of these rather simple ARX-models (auto-regressive models with exogenous inputs) with daily and sub-daily data, following the same sub-questions.

2. **Does the use of auto-regressive models with exogenous inputs (ARX-models) improve the Energy Signatures when data at daily time resolution is used?**

In measurements of the energy use at a sub-daily time interval, the diurnal fluctuations in energy use are observed, and patterns in the energy use are recognised. These are called Energy Use Time Patterns and they provide additional insights in the building energy use. For example in Figure 1-3 the gas use for space heating of a house is presented at 2-hourly time intervals and different Energy Use Time Patterns are observed, especially for weekdays and weekend days. It is also observed that a night setback is applied in the space heating control. The third objective of this study is to mathematically identify the different groups of Energy Use Time Patterns occurring in energy use time series at a sub-daily time step. The fourth objective is to include information from the Energy Use Time Patterns at sub-daily level into linear regression and ARX-models with data at daily level and to evaluate the model performance.

3. **How can groups with similar Energy Use Time Patterns be recognised mathematically?**

4. **How are groups with similar Energy Use Time Patterns characterised in function of specific weather conditions and/or calendar information?**

5. **Does the inclusion of groups of Energy Use Time Patterns into linear regression and ARX-models improve the Energy Signatures with daily aggregated data?**

The overall objective of the study is to evaluate and develop statistically sound Energy Signature models that allow for a more accurate characterisation of the energy use for heating in residential buildings, using daily and sub-daily energy use measurements. The applications in view are energy feedback, energy auditing or commissioning. By means of example one specific implementation of the Energy Signatures is envisaged for application in energy feedback: the comparison of the energy use for two periods and the identification of changes in energy use. Therefore an additional sub-question is:

- How to compare the Energy Signatures for different periods?

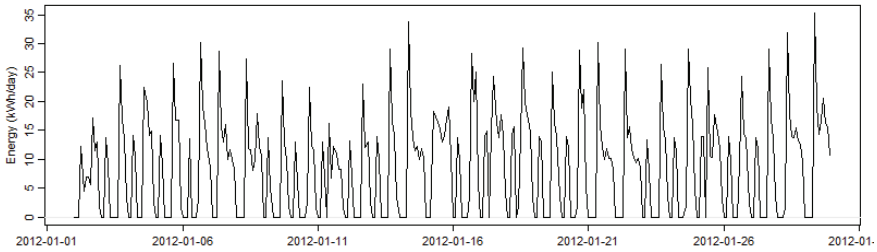


Figure 1-3: 2-hourly gas meter readings of a house in January 2012

1.4 Approach & Structure

1.4.1 Approach

In this dissertation, different types of linear regression models are investigated for the characterisation of Energy Signatures for residential buildings using daily heating energy use data (research questions 1, 2 and 5). The focus is on the energy use for heating, that includes space heating and domestic hot water generation. The system that is modelled is the occupied house, consisting of the building structure and envelope, the building services and the building users, and their interactions. The weather is influencing this system, as does the calendar – since energy use may differ for different days of the week or year, due to occupants' schedules. The relationship between the system and the weather and calendar variables is modelled, starting with classical linear regression models towards more complex and dynamic models. Besides, the characterisation of Energy Use Time Patterns observed in sub-daily energy use data is investigated (research questions 3 and 4). Time series decomposition and cluster analysis techniques are used to recognise groups of similar patterns in the data. For modelling the relationship between these groups of patterns and weather or calendar variables, logistic regression models are applied. The data manipulation and calculations are executed in R, a language and software environment for statistical computing and graphics [69].

The data are energy use time series with hourly gas meter readings from Belgian houses for a period of 2 years, and weather data for the same period. In Belgium, gas is typically used for space heating purposes and often also for domestic hot water generation or cooking. The models are investigated using data from 25 houses where gas is used for space heating only. In the end, the models are illustrated for the situation of combined space heating and domestic hot water use. In view of the applicability of the developed models it is assumed that no other information about the houses or measurements are available. However, in the context of the development of the models in this research, surveys of the occupants were supplied, providing some information of the building and its use. However, since no detailed information is available on the energy performance of the buildings, the models cannot be validated by comparing the Energy Signatures to such information. Instead, validation is performed by assessing the statistical model validity and by cross-validation of the models that are developed on a training data set (year 1) through application on a test data set (year 2). With regard to the cluster analysis, which is an unsupervised learning technique, cluster validity is assessed through relevant indices, but validation with data from surveys is also limited.

The evaluation of the different linear regression models follows a similar approach, addressing the research sub-questions. The models are developed on a training data set consisting of daily aggregated data of one year or heating season. Each time the variables of the model are investigated and conclusions are drawn on the key variables to be included in a base model that suits the majority of cases and/or on an approach for cases-specific variable selection. The statistical validity of the model is discussed, which is necessary for the interpretability of the outcomes. With regard to the outcomes, focus is on the accuracy of the estimated Energy Signature coefficients, since the application example envisaged is the comparison of Energy Signatures for different periods, in this case the 1st and 2nd year of the measurements are compared. Besides, the study also expands upon the accuracy of the predictions of the energy

use for different periods, which is more important in the context of normalisation or prediction of energy use. The effect of reducing the length of the measurement period from one heating season to periods of 1, 2 and 3 months is assessed. For the classical linear regression and ARX-models, the use of sub-daily (2-hourly) instead of daily aggregated energy use data is illustrated by means of an example.

1.4.2 Structure

In chapter 2, the energy use and weather data-sets are described and the results of the occupants surveys are summarised. In chapter 3, 4 and 5 three different types of linear regression models are investigated, following the evaluation approach exposed above. In chapter 3 (research question 1), classical Energy Signature methods using classical multiple linear regression techniques are investigated, the results are compared to those of the traditional situation in which weekly aggregated energy use data are assumed. In chapter 4 (research question 2) auto-regressive terms are added to the linear regression models, which are then called auto-regressive models with exogenous inputs. These state-of-the-art models are evaluated and compared to the linear regression models. Chapter 5 first zooms in on the patterns observed in sub-daily energy use time series and on the mathematical recognition of similar patterns (research question 3). Then the modelling of groups of similar patterns in function of weather and/or calendar data is elaborated (research question 4). In a next part of the chapter, we zoom out again to the daily aggregated data and the Energy Signature modelling (research question 5). The models developed here are based on the classical linear regression models and ARX-models, that are now extended with variables that identify to which group of similar Energy Use Time Patterns each of the days in the data belongs. So in this approach, information from the sub-daily data is processed and entered into the models with daily aggregated data. Again, these models are evaluated and compared to the original linear regression models and ARX-models. Finally, the models are applied to energy use data from houses where heat is used for both space heating and domestic hot water demand. Note that each of the chapters 3, 4 and 5 is completed with a summary and conclusions. In Chapter 6, the general conclusions and perspectives of the study are presented.

2

Data

The data consist of smart meter readings for residential buildings in two Belgian villages, surveys of the occupants of some of these buildings and measurements of weather variables at a nearby weather station. No other measurements of the buildings are available. In accordance with the goals of the study, the focus is on gas smart metering data, that is – in general more than electricity metering data – heavily influenced by weather conditions, since it is typically used for space heating purposes.

In section 2.1, after an exploration of the energy use data and needs for the current study, a sample of the case-study dwellings is selected and the energy use data are described. Then, using the information from the surveys, characteristics of the dwellings and occupants are summarised. In section 2.2, the preparation and properties of the weather data are described.

2.1 Energy use data

The data used in this study belong to the Flemish Distribution System Operator *Eandis*, and was gathered in context of a Proof of Concept (POC) study for Smart Metering during the period 2010-2013. The POC-project took place in the communities of Leest and Hombeek where about 4750 smart meters were installed in residential buildings for registering the gas or electricity use with respectively hourly and 15min time intervals. During the project period two information and feedback campaigns were organised (POC1 and POC2). Before and after each campaign, surveys were sent to the smart metering population in order to gain a basic understanding of the building and its occupants, and to inquire about the occupants reactions to the feedback. Note that all users in the study are anonymous.

2.1.1 Sampling gas meter data

The energy use data considered comes from gas meters. The gas meter data is subdivided into 4 groups, based on the function(s) for which the occupants are using gas: space heating, domestic hot water generation and/or cooking. An overview of the functional groups is presented in Table 2-1. For example, group 2 includes meters that use exclusively gas for space heating and domestic hot water (while not using other energy sources for these two functions) and does not use gas for cooking purposes. A sub-group of group 2 (between parenthesis) includes cases where exclusively gas is used for space heating, and the energy sources for domestic hot water include gas, combined with other energy carriers.

Table 2-1: Gas meter data: groups by function

	1 SH	*2 SH&DHW*	*3 SH&COOK*	*4 SH&DHW&COOK*
Space Heating (SH)	EXCLUSIVELY	EXCLUSIVELY	EXCLUSIVELY	EXCLUSIVELY
Domestic Hot Water (DHW)	NOT	EXCLUSIVELY (INCLUDING)	NOT	EXCLUSIVELY (INCLUDING)
Cooking (COOK)	NOT	NOT	EXCLUSIVELY	EXCLUSIVELY

Table 2-2: Gas meter data: sampling of the population by function

	1 SH	*2 SH&DHW*	*3 SH&COOK*	*4 SH&DHW&COOK*
A-population	74	328 (+10)	21	155 (+4)
B-population	22	99(+3)	4	38(+3)

Based on the response of the surveys sent before each of the feedback-campaigns, the meters are allocated in one of the aforementioned groups. Furthermore, the meters are grouped into the A- and B-population. The A-population includes all cases that completed the starting survey before the first feedback campaign (between April and August 2010) and received feedback from the first feedback campaign onwards (POC1 – February 2011 to February 2012). The B-population received feedback from the second feedback campaign (POC2 – June 2012 to June 2013) and replied to the starting survey (between August 2012 and February 2013). In Table 2-2 the number of meters for each group is mentioned.

2.1.2 Space heating data

A "neat" space heating data set is assembled for the investigation of energy use models in this study. The data set includes cases where gas is used for space heating only, and only gas is used for space heating (no other energy sources), according to the surveys and to observation of the energy use data. Moreover, it includes those cases where no long periods of the data are lacking and where no long periods of absence are present during the first heating season 2011-2012. Apart from a visual inspection of the energy use time series, the following indicators were derived:

- The amount of data-points lacking
- The amount of data-points with 0-value between 1st of May 2011 and 31st of August 2011
- The amount of data-points with 0-value before 1st of May and after 31st of August 2011

In the B1-population, three cases were excluded because more than 100 data-points were lacking, 8 cases because it is not sure they use gas for space heating purposes only (<30 0-values between May and August), and another 5 cases where long periods of absence during winter can be suspected (>50 0-values during winter and intermediate season). The remaining data-set is considered too small for the study. In the A1-population, 7 cases were excluded because more than 50 data-points were lacking, 21 because it is not sure they use gas for space heating purposes only (<20 0-values between May and August + visual inspection of the series). Three cases (case 1,8 and 17) with a low or moderate amount of 0-values during summer time (<60) were not excluded, because further observations revealed a continuous very small energy use, that could be due to a pilot flame, small gas leak or system characteristic. For the remaining cases, those cases with more than 30 0-values before May and After August, the time series were further observed: some of these were excluded because of long periods with no space heating during winter time (e.g. due to long holidays), while others remained, because they only have a shorter heating season. One exception here is case 3, a house that seems to be heated on a regular basis, but only for a few days per week. Finally, for the remaining dwellings, it was checked whether the previous findings were consistent for the full heating season 2011-2012.

As a result, the data sample includes 25 dwellings from the A1-sample. All these cases were included in the POC1 and feedback campaign. However, gas meter data is available only from February 2011 onwards, at the time the feedback loop started. Therefore, it is important to remark, for the remainder of the study, that during the smart metering data gathering, the occupants of the dwellings could have been influenced by the information and feedback they received in the feedback campaign, and no data is available for a "neutral" period. Of course one could question whether populations that not participated in the POC1 campaign, like B1-population, can be considered "neutral" (in terms of their energy-efficiency related behaviour and decisions), since apart from this feedback campaign, many other forms of awareness raising, energy feedback and information on energy efficiency are available in our society. All users included in the data-sample also completed (part of) the POC1 starting survey between 20/04/2010 and 30/06/2010. After POC1, 9 of these cases also completed the final survey, between February and May 2012. Moreover 6 of 25 completed the final survey of the next POC2 feedback campaign in May 2013. These surveys can give us some clues about the (changes in) energy-efficiency of the dwelling and to the effect of the feedback.

Description of the data (daily aggregated)

Figure 2-1 is a time series plot that summarises the energy use data of the 25 cases by use of daily values for the mean, extrema and 1st and 3rd quartiles between 1st of September 2011 and 30th of June 2012. In Figure 2-2 the same period in the next year is presented. In Figure 2-3 the main statistics are calculated from the cumulative energy use figures for each case and for both years. It is found that in general the energy use during the 2nd heating season is higher than during the first heating season.

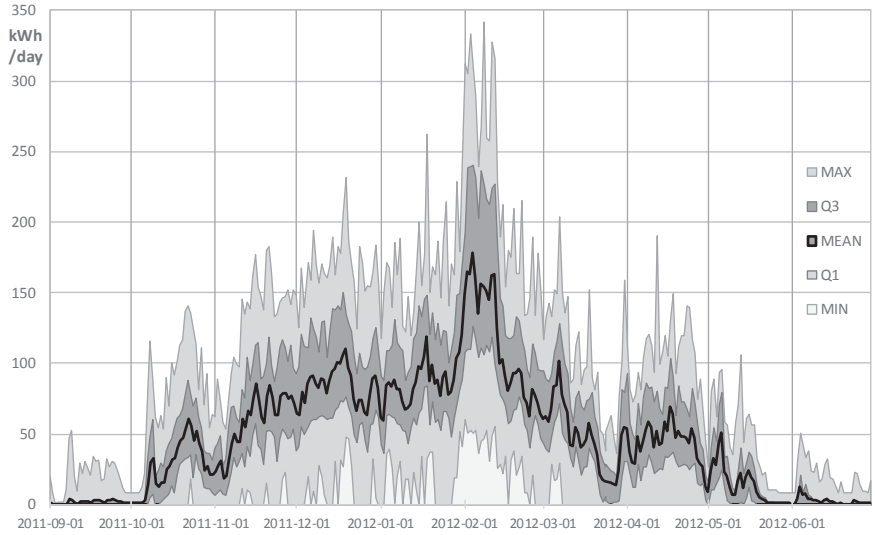


Figure 2-1: Energy use – Sept 2011 to June 2012 – Time Series plot - Mean and Quartiles

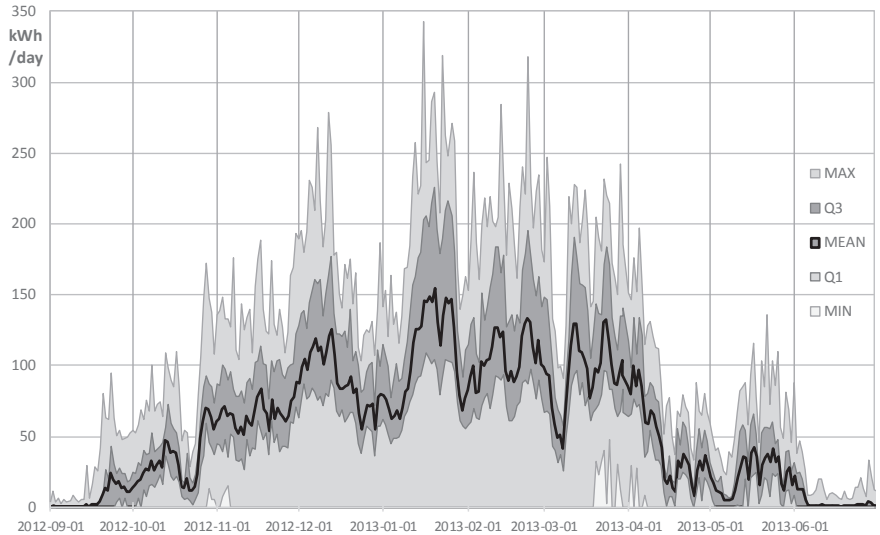


Figure 2-2: Energy use – Sept 2012 to June 2013 – Time Series plot - Mean and Quartiles

Since the energy use is related to space heating only, during summer no energy is used in most of the houses and the heating is turned on for a few days in few dwellings. The heating season 2011-2012 took a clear start between 7 and 12 October 2011 and ended between 1st and 15th of May 2012, for a majority of the cases. The 2nd heating season 2012-2013 started already in September for more than half of the cases, and also lasted longer for a number of cases. During the heating season, space heating is generally active, apart from a few warm days at the end of March. For the purpose of this study, the start and end day of the heating season are specified for each case and heating season individually. Because for some models a a set of complete weeks is required, the individual heating seasons always start on a Monday and end on a Sunday. The first heating season is 20 to 29 weeks long and the second up to 36 weeks. The cases having a shorter heating season, e.g. case 4 and 6, are also among the lowest total energy uses (see Figure 2-4). In Annex A data on the heating season and energy use are listed for each case individually.

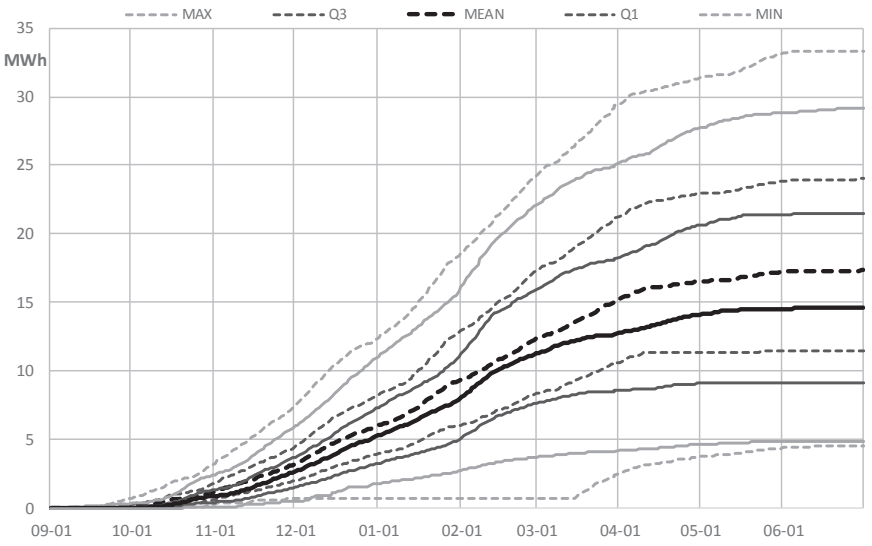


Figure 2-3: Energy use – Cumulative Distribution – Heating Season 1 (__) & 2 (---)

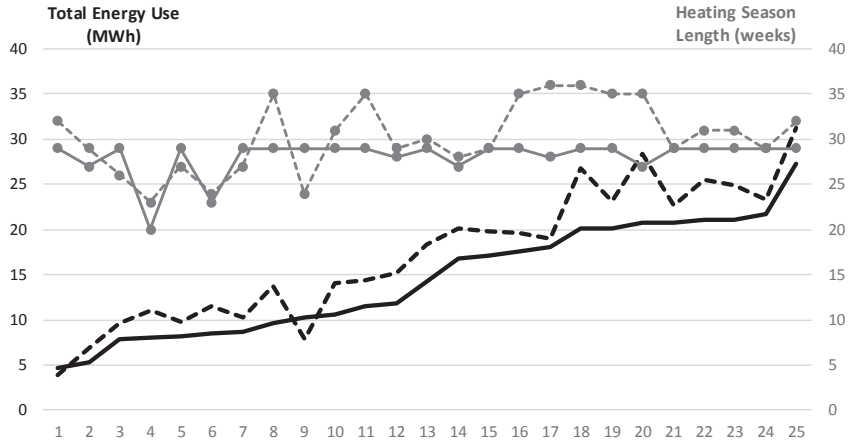


Figure 2-4: Specific Heating Season Length and Total Energy Use – Heating Season 1 (__) & 2 (---)

Figure 2-5 presents the main statistics of the daily energy use for each of the cases during their individual heating season. The dwellings are listed in order of increasing total energy use over the heating season 2011-2012. The average daily energy use ranges between around 20 and 140 kWh/day for both periods.

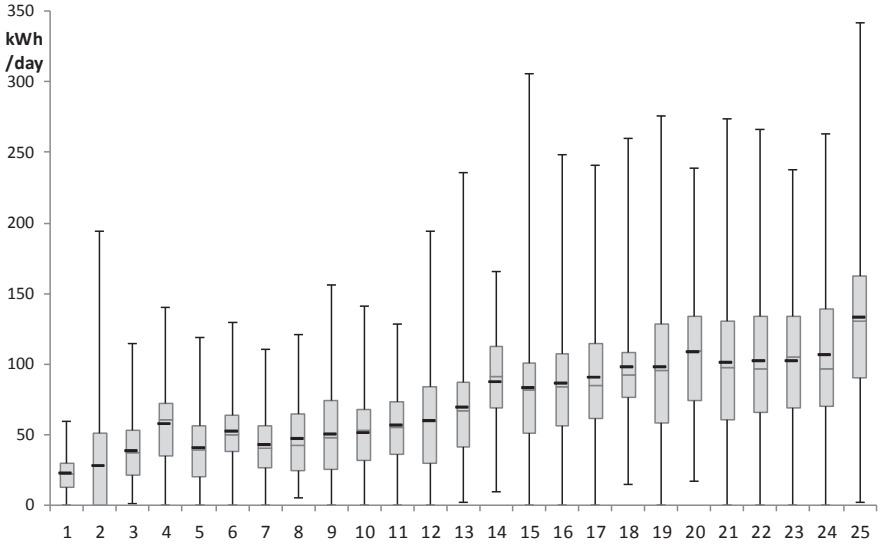


Figure 2-5: Energy use – Heating Season 2011-2012 – Boxplot per case

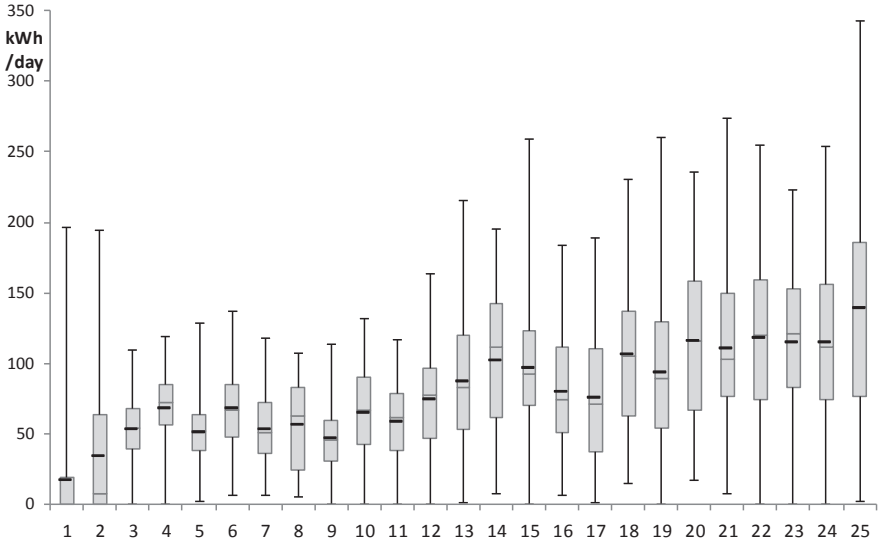


Figure 2-6: Energy use – Heating Season 2012-2013 – Boxplot per case

2.1.3 Documentation of the case-study dwellings

The 25 case-study dwellings can be briefly documented using the surveys. Note that this information is not verified by a building specialist, and may contain incorrect answers. All dwellings are single-family dwellings of different typologies (Figure 2-7). Most terraced houses are amongst the houses with lower energy use and detached houses amongst those with the higher use. The size of the houses is unknown, but could also explain this, since terraced houses are often smaller than detached houses. Regarding the energy-efficiency of the building, the occupants were asked whether roof, wall or floor insulation are applied, whether more than 50% of the windows have high performance glazing and whether a high performance (or gas condensing) boiler has been installed. In Figure 2-8, the bars on top of the y-axis represent the measures installed in the house, and those below are present if they were not. If the answer is unknown, the corresponding bar is not filled. Roof insulation is provided in the majority of dwellings (18/25), wall insulation in 11 cases and a high performance boiler in 10 cases. Floor insulation and high performance glazing are less frequently applied (4 and 3 cases). Since in the majority of the houses only a few insulation measures are applied, it is deduced that the houses were not recently build. In 2 houses (6 and 8) no central heating is available.

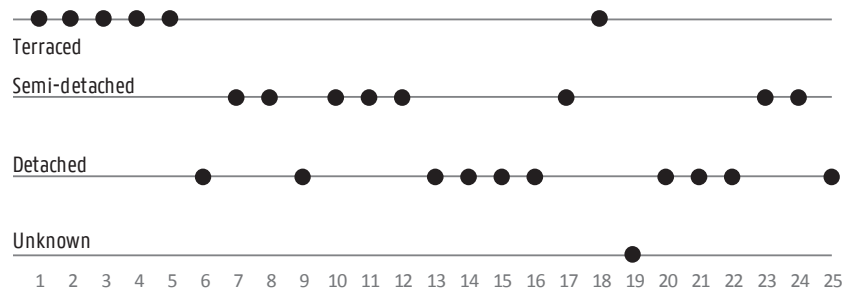


Figure 2-7: Dwelling Typology

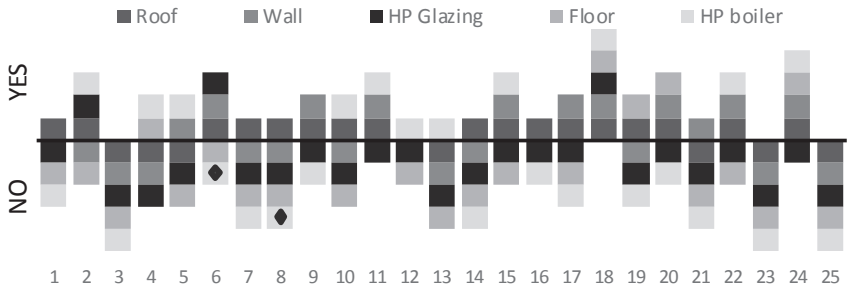


Figure 2-8: Insulation of the roof, wall, floor, high performance glazing (>50% of the windows) and high performance boiler (♦ no central heating)

The number of occupants and whether they are usually at home during working hours are presented in Figure 2-9. The median number is 2 occupants, and most of them are usually at home during working hours. In the larger households, occupants are mostly absent during working hours. Furthermore most of the houses have a programmable thermostat (Figure 2-10)., but none of them declares to have an exterior temperature sensor connected to the thermostat. All occupants declare that they apply night set-back of the space heating system. During daytime, most households declare that they lower the set-point temperature when they are absent (Figure 2-11).

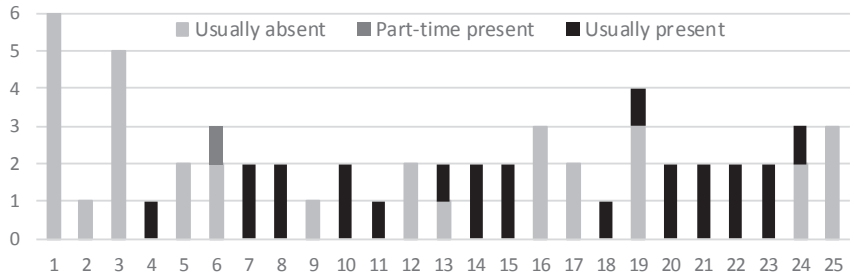


Figure 2-9: Number of occupants that is present during working hours

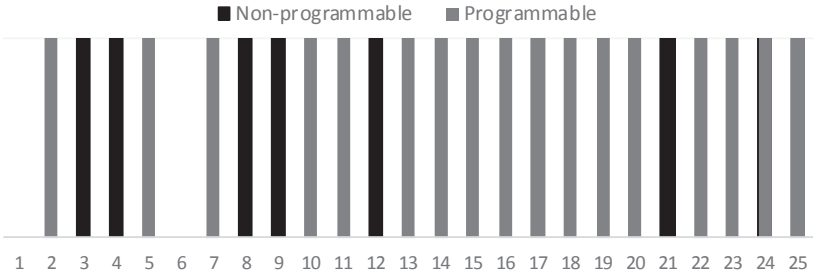


Figure 2-10: Type of thermostat

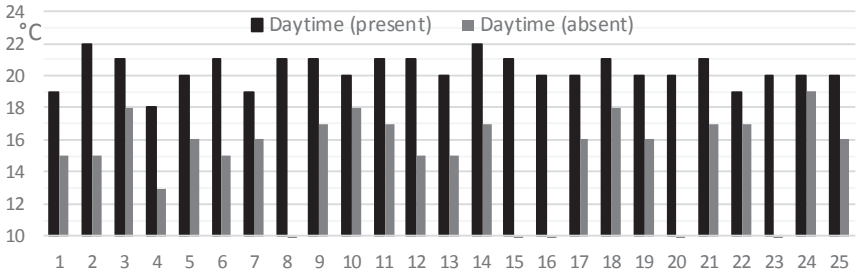


Figure 2-11: Set-point temperatures

2.2 Weather data

2.2.1 Data collection

Weather data are available for the whole period of the study, which is from 1st of February 2011 until 1st of June 2013. They are provided by the Royal Meteorological Institute of Belgium. The data are collected at the weather station in Sint-Katelijne-Waver, at about 10 km distance from the villages of Leest and Hombeek. The data set includes measurements with 10 minutes time steps of the following weather variables: exterior temperature (dry bulb), relative humidity, precipitation quantity, wind direction, wind speed (2m height) and global solar irradiation on a horizontal surface. Between June 29th 2011 at 14h00, and July 7th at 10h00, data are missing from the weather station of Sint-Katelijne-Waver. In this period, hourly observations from the RMI national weather station of Uccle are introduced.

2.2.2 Data pre-processing

First, the weather data are aggregated to hourly values, in line with the time step of the gas use data. All weather variables are aggregated by averaging, except for the precipitation quantity, aggregated by summation. The time is originally indicated in Coordinated Universal Time standards (UTC), and adapted to local time, that is Central European Time (CET = UTC+1) and Central European Summer Time (CEST = UTC+2). By convention, the time variables indicate the start of the time period for which it is valid.

Some secondary wind speed variables are calculated from the hourly aggregated data: the square of the wind speed, the square of the wind speed multiplied with the exterior temperature, the wind speed for the four cardinal directions and the square of the wind speed for these directions multiplied with the exterior temperature. When calculating the wind speed for the cardinal directions, the it is multiplied with the share of each of the cardinal directions Dc (in °) at the same time interval. The share of each of the cardinal directions is calculated using Equation 2-I:

$$\begin{Bmatrix} 0 & \text{if } 1-|Dc-D|<0 \\ \frac{1-|Dc-D|}{90} & \text{if } 1-|Dc-D|>0 \end{Bmatrix} \quad [2\text{-I}]$$

Where D is the hourly average wind direction angle (North = 0° or 360°, East = 90°, South = 180° and West = 270°). For example, if the average wind direction is 225° (SW), then the share of the wind directions south and west is 50% and north and east is 0%. In situations where values around 0° / 360° occur, the equation is corrected.

The primary and secondary hourly climate variables are now further aggregated to 2-hourly, daily and weekly data. For the daily aggregated data, two more secondary variables are calculated, that is the equivalent temperature (Equation 2-II) and the equivalent heating degree days (Equation 2-III):

$$Te_{eq} = 0{,}6 \times T_t + 0{,}3 \times T_{t-1} + 0{,}1 \times T_{t-2} \quad [2\text{-II}]$$

$$HDD_{eq} = \max\left(0; Te_{base} - Te_{eq}\right) \quad [2\text{-III}]$$

T_t, T_{t-1} and T_{t-2} are the exterior temperature of the current day and the previous two days. Te_{base} is the base temperature, and is assigned a standard value of 16,5 in Belgium. It represents the exterior temperature below which space heating energy demand is assumed to exist, on average over the Belgian building stock. The applied equations are consistent[1] with the calculation methods applied in Belgium for gas billing and normalisation of gas use figures [70].

A list of primary and secondary climate variables is provided in Table 2-3.

Table 2-3: Climate variables

Symbol	Name	Description	Unit
Te	TEMP	Exterior temperature (dry bulb)	°C
Hr	RELHUM	Relative humidity	%
Pq	PRECIP	Precipitation quantity	mm
Rg	RADglob	Global solar irradiation (on a horizontal surface)	W/m²
Wd	WINDDIR	Wind direction (angle, 0° N; 90° E; 180° S; 270° W)	°
Ws	WINDSPEED	Wind speed (at 2m height)	m/s
Ws²	WINDSPEEDSQ	Wind speed squared	(m/s)²
Ws²Te	WINDSPEEDSQTEMP	Wind speed squared multiplied with exterior temperature	°C(m/s)²
Ws,e	WINDSPEEDeast	Wind speed multiplied with WINDDIReast	m/s
Ws,w	WINDSPEEDwest	Wind speed multiplied with WINDDIRwest	m/s
Ws,s	WINDSPEEDsouth	Wind speed multiplied with WINDDIRsouth	m/s
Ws,n	WINDSPEEDnorth	Wind speed multiplied with WINDDIRnorth	m/s
Wd,n	WINDDIRnorth	Share of wind direction North	-
Wd,s	WINDDIRsouth	Share of wind direction South	-
Wd,w	WINDDIRwest	Share of wind direction West	-
Wd,e	WINDDIReast	Share of wind direction East	-
Te,eq	TEMPeq	Equivalent exterior temperature	°C
HDDeq	HDDeq	Heating Degree Days (base temperature 16,5°C)	-

[1] Except for the method of estimating the daily exterior temperature. In the official method it is estimated by averaging 13 2-hourly temperature meter readings in the national weather station of Uccle (BE). In this study it is calculated by averaging 10 min values from the weather station in Sint-Katelijne-Waver to daily values.

2.2.3 Data set description

The original data set includes two complete heating seasons, divided in a training heating season (2011-2012) and a test heating season (2012-2013). In Figure 2-12, Figure 2-13 and Figure 2-14 the data for the training heating season (2011-2012) are described. The average exterior temperature is 6.8°C, with a maximum of 17°C and minimum of -12°C in February 2012. The maximum hourly average solar radiation decreases from about 200W/m² in Spring to about 50W/m² in the midst of Winter. It shows a negative correlation with relative humidity, with a correlation coefficient of -0.65, which is the strongest correlation appearing amongst the primary climate inputs. The average wind speed is around 2.4 m/s and maximum values of 7 m/s occur. The prevailing wind directions are S, SW and W, which is typical in Belgium. Regarding the data for the test heating season (2012-2013), for most of the primary input variables, the range of the values is similar to those from the training data set (see Figure 2-15). The exterior temperature has a very similar mean, but more values in the range between 0 and -5°, combined with a higher minimum temperature of -6°C. The average and maximum solar radiation are higher than in the training data set. The remainder of the data for the test heating season 2012-2013 and the correlation plots for the different wind speed variables are presented in *Annex A.*

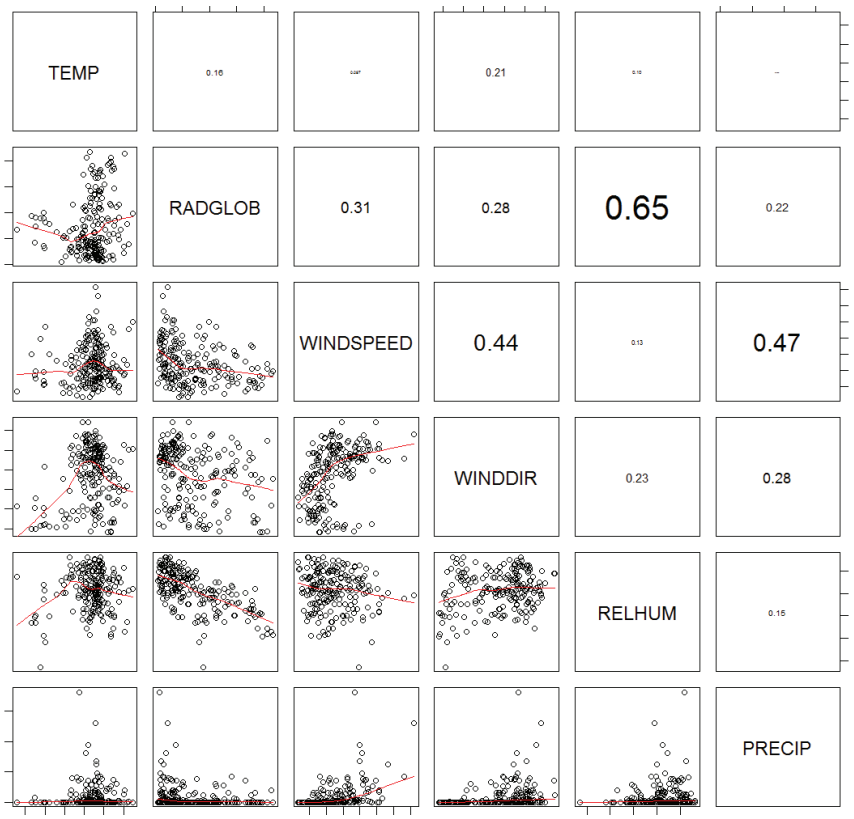


Figure 2-12: Climate variables – Heating season 2011-2012 – correlation plot

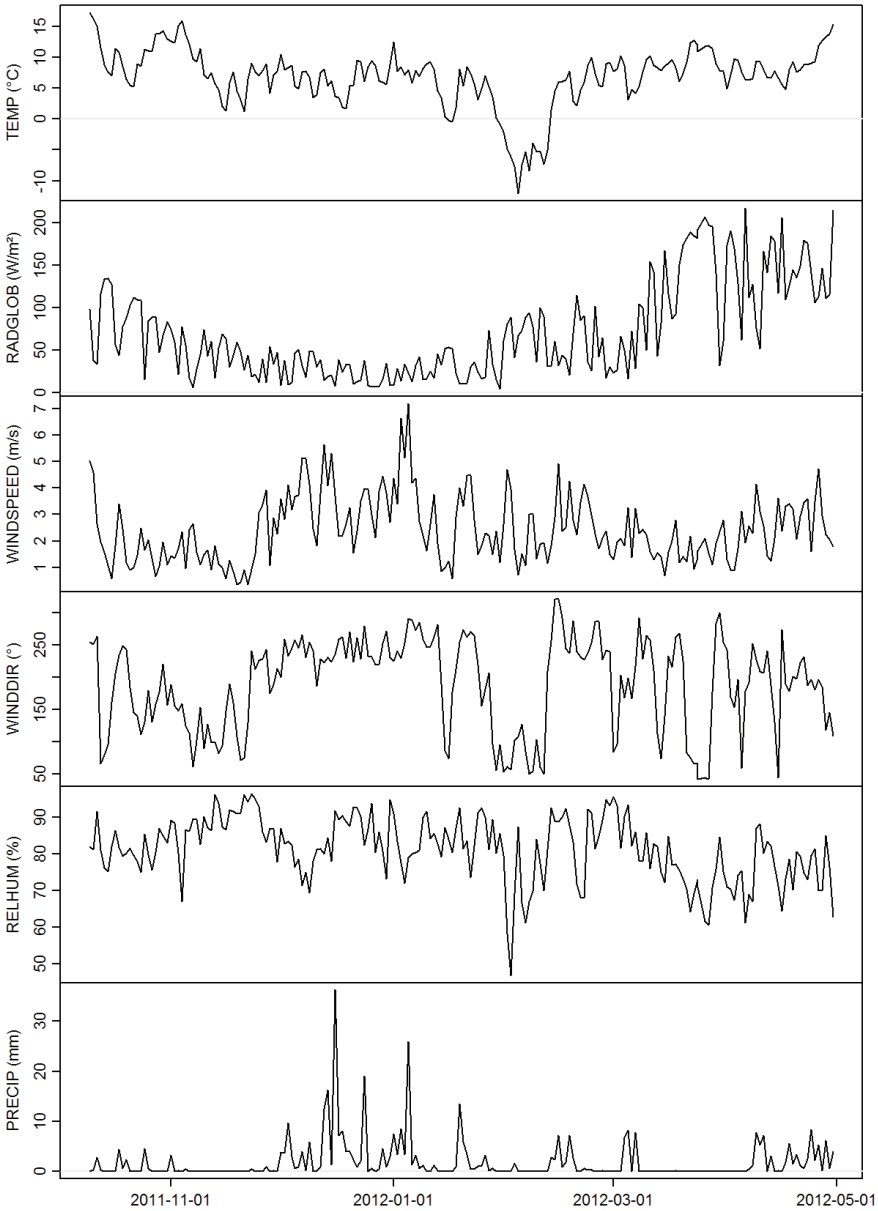


Figure 2-13: Climate variables – Heating season 2011-2012 - time series plot

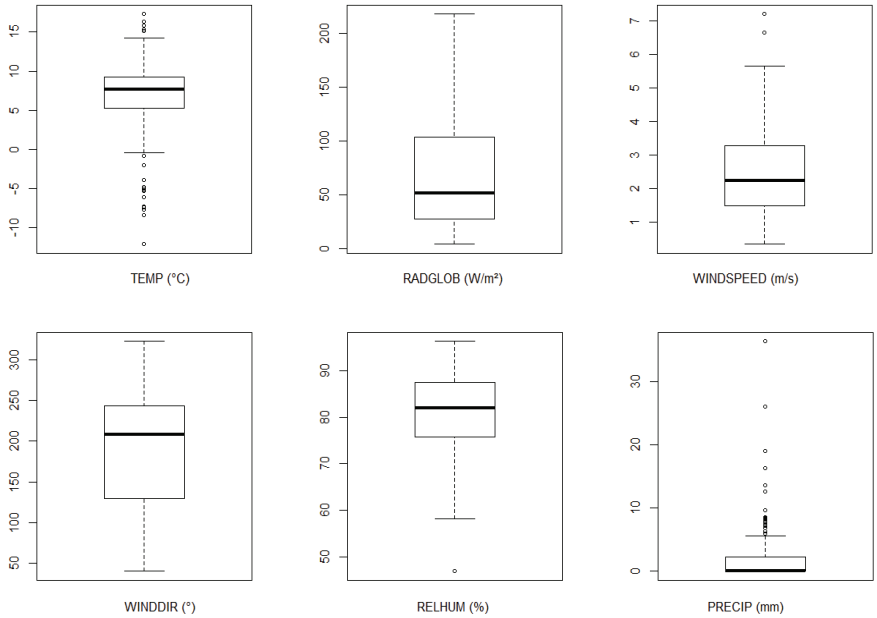


Figure 2-14: Climate variables – Heating season 2011-2012 – boxplot

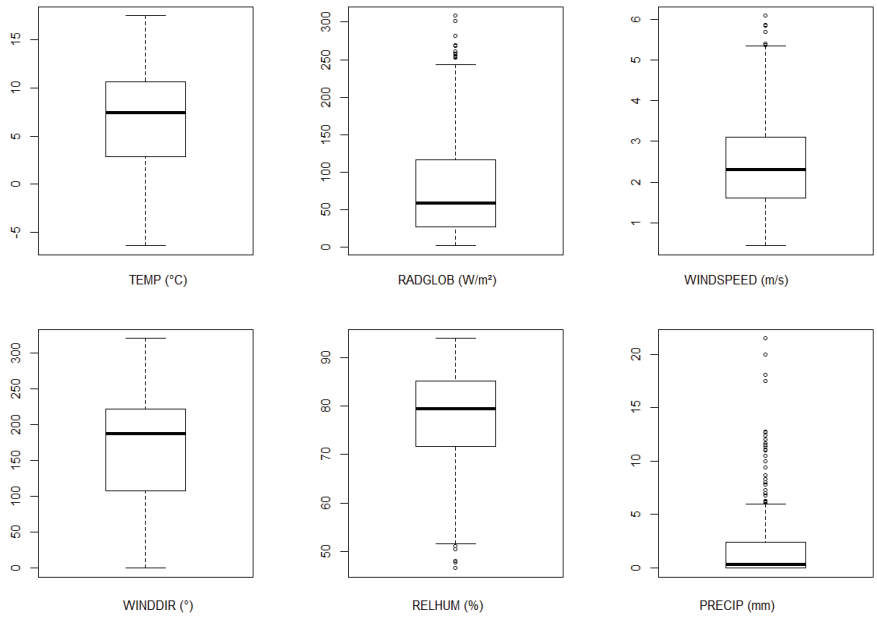


Figure 2-15: Climate variables – Heating season 2012-2013 – boxplot

3

Classical Linear Regression

In the widely applied Energy Signature models, classical linear regression techniques are used to characterise the building energy use in function of weather variables based on occasional (weekly, monthly…) energy use data. In this chapter, these LM-models are confronted with more frequent (daily, sub-daily) data and they are evaluated. The measurement period used for model estimation is one heating season. The following research questions are addressed: **Does data at daily time resolution allow to improve Energy Signature models applying classical linear regression techniques? (Research question 1)**

- What are the key weather variables in the LM-model? (section 3.2)
- Are the models valid from statistical point of view? (section 3.1 and 3.3)
- What is the accuracy of the energy use predictions? (section 3.4)
- What is the accuracy of the estimated Energy Signature coefficients? (section 3.5)
- Can the measurement period be reduced to 1 up to 3 months with limited loss in accuracy? (section 3.6)
- What if data at sub-daily time resolution is used? (section 3.7)

3.1 Statistical methods

A linear regression model expresses the linear statistical relation between a dependent variable (output) and one or more independent variables (exogenous inputs). The relationship is linear: the regression model is linear in its parameters. The relation is statistical, there is a probability distribution of the dependent variable for each level of the independent variable(s). The regression function is here estimated using the method of the least squares, that is by minimising the sum of the squares of the errors, where an error or residual is the deviation between the measured and expected value of the dependent variable. After fitting the model, **the validity of the model** is investigated. By use of regression diagnostics model assumptions are checked:

- Normal distribution of the errors: by use of the Normal Quantile-Quantile probability plot (Q-Q) (see Figure 3-1 (b))
- Constant variance of the errors or homoscedasticity: plot of the (square root of the) errors against the fitted values (see Figure 3-1 (a) and (c))
- Linearity: plot of residuals against predictor variables or response against predictors (Figure 3-2)
- Independent errors: plot against time: auto-correlation function of residuals (see Figure 3-1 (d))

Furthermore, the model is checked for multicollinearity and for outlying and influential observations:

- Multicollinearity is defined as (a high degree of) correlation of the predictor variables among themselves. In practice, it does usually not inhibit to obtain a good model fit or to estimate the mean responses or predictions for new observations, if made within the region of observations. However the estimation and interpretation of regression coefficients is more difficult (and their standard errors increase): they reflect only a partial effect of the predictor on the response variable, depending on other correlated predictor variables in the model [71]. Multicollinearity is detected by use of the Variance Inflation Factors (VIF). These factors measure how much the variances of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related. According to Kutner et al., a maximum VIF among all input variables in excess of 10 is often taken as an indication that multicollinearity may be unduly influencing the least squares estimates and mean VIF values (over all predictors) considerably larger than 1 indicate severe multicollinearity. According to Gareth et al., a maximum VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity [72].
- Outlying and influential observations: observations that are well separated from the main group of the data can have a large influence on the fitted regression function. Outlying observations can be found by large absolute values of the standardised (internally studentised) residuals (indicating outlying Y-observations) and leverage values (indicating outlying values of the independent variables) higher than .2 (moderate leverage) or .5 (very high leverage) [71]. The influential observations can be identified by a high Cook's distance (influence of all fitted values), which shows the effect of a certain observation on all fitted values. They are observed using diagnostic plots (Figure 3-1 (e) and (f)).

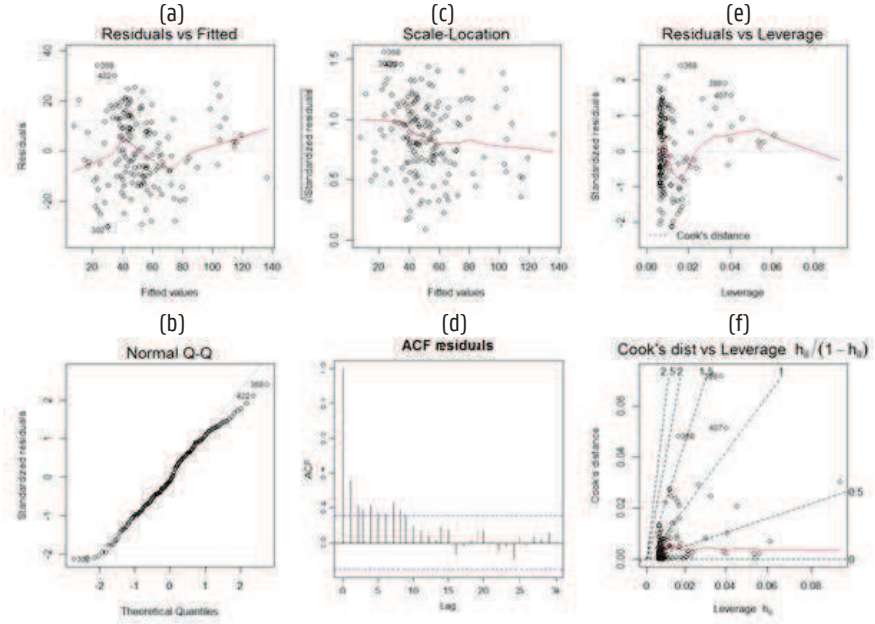


Figure 3-1: Regression diagnostics – Example

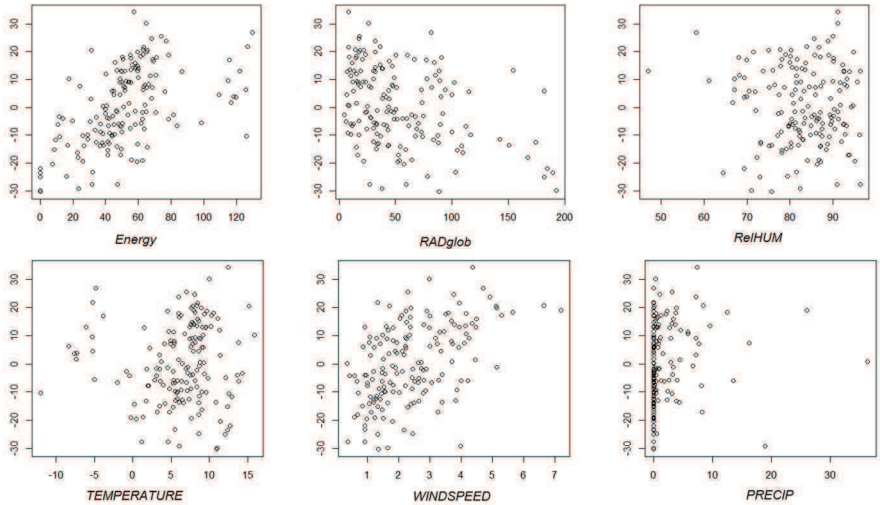


Figure 3-2: Regression diagnostics: Residuals vs Variables – Example

In this study diagnostic plots are used to check model validity. Note that in automated modelling, when no human intervention is desired, the use of statistical tests instead of diagnostic plots could be investigated.

The following **criteria** are used for model selection and/or for assessing the prediction accuracy:

- The coefficient of determination R^2 and the adjusted coefficient of determination R^2adj. These are measures for the goodness-of-fit of the model. The R^2 measures the proportionate reduction of the total variation in the outcome of the model associated with a set of predictors, and takes values between 0 and 1. While the R^2 will only increase by adding more predictors, the adjusted R^2 takes into account the number of predictors through the degrees of freedom [71]. In this study, the focus is on the R^2adj.
- The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are measures of prediction accuracy, that can be applied when comparing different models applied on the same data set. The information provided by the RMSE is equivalent to the information provided by the R^2adj for a given data-set, but the advantage is that the RMSE is on the same scale as the data. It can also be applied out-of-sample, for example to compare the prediction accuracy of different models on one test data set. The MAE can be used for similar purposes, but is less influenced by occasional very large errors. Both RMSE and MAE can also be expressed relative to a benchmark, in which case it will be called relRMSE or relMAE [73].
- The Mean Absolute Scaled Error (MASE) is also a prediction accuracy measure, that has the advantage of being applicable for comparing the prediction accuracy across different models and/or different data sets. The MASE is calculated by dividing the mean absolute error (in-sample or out-of-sample) by the mean absolute error of the one-step naïve forecast method applied on the training data set (in-sample). The naïve forecast method being a forecasting method for time series, whereby the forecast for a future value is set equal to the last observed value [74]. The MASE was proposed by Hyndman and Koehler, because it is independent of the scale of the data, it is only infinite or undefined when all historical observations are equal, and has a good interpretability [73].
- The t-statistic for a particular regression coefficient is used to test the hypothesis whether this particular regression coefficient equals zero, and if not accept the alternative hypothesis that the coefficient is different from zero [72]. Unless stated otherwise a significance level of 5% is used and the p-value of the t-statistic is checked: if it is below 0.05 the null hypothesis is rejected.

In the **model selection process**, the following approach is followed: in an initial round forward stepwise regression procedures are applied to select the input variables and the different stages in the selection are explored. From this exploration a number of models emerges, that are listed in Table 3-1. In a second round, discussed in section 3.2, based on the previous exploration and physical insights, linear regression models are constructed, starting from a simple model and gradually adding more inputs, as well as checking the effect of similar input variables. The outcomes are a base model that includes the key variables for the majority of cases, and findings that can be used when implementing an automated case-specific variable selection.

A full heating season training period is used to select the model variables, so the data set as well as the variation in weather inputs is reasonably large and has good chances to cover the region of observations that can be expected in our climate, in order to enable application of the models over the full heating season (e.g. is not limited to specific weather conditions during shorter periods in the heating season). In section 3.6 the effect of reducing the training period is investigated. The time step of the data is one day, which is the shortest time step that is

reasonable to use when static models are used, even if dynamic effects longer than one day can be present (e.g. building time-constants or behavioural profiles) [24,29]. The effect of reducing the time step will be discussed in section 3.7.

Table 3-1: Linear Regression – Model List

Model	Variables
LM1	HDDeq16.5
LM6	TEMP
LM7	TEMPeq
LM8	TEMP-RADglob
LM13	TEMP-RADglob-RELHUM
LM14	TEMP-RADglob-PRECIP
LM15	TEMP-RADglob-WINDDIR
LM40	TEMP-TEMPl1-TEMPl2
LM50	TEMP-RADglob-WINDSPEED
LM51	TEMP-RADglob-WINDSPEEDSQ
LM52	TEMP-RADglob-WINDSPEEDSQTEMP
LM57	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth
LM58	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-WINDSPEEDnorth
LM59	TEMP-RADglob-WINDSPEEDeast
LM60	TEMP-RADglob-WINDSPEEDwest
LM61	TEMP-RADglob-WINDSPEEDsouth
LM62	TEMP-RADglob-WINDSPEEDnorth
LM63	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest
LM64	TEMP-RADglob-WINDSPEEDwest-WINDSPEEDsouth
LM65	TEMP-RADglob-WINDSPEED-WINDSPEEDSQTEMP
LM66	TEMP-RADglob-WINDSPEED-TEMPl1
LM67	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1
LM68	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-WINDSPEEDnorth-TEMPl1
LM69	TEMP-RADglob-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1
LM70	TEMP-RADglob-WINDSPEEDwest-TEMPl1
LM71	TEMP-RADglob-WINDSPEED-TEMPl1-TEMPl2
LM72	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1-TEMPl2
LM73	TEMP-RADglob-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1-TEMPl2
LM74	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1-RADglobl1
LM75	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1-WINDSPEEDl1
LM76	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1-PRECIP
LM77	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1-RELHUM
LM78	TEMPeq-RADglob-WINDSPEED
LM79	TEMPeq-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth

3.2 Weather variables

The value of different climate inputs in an Energy Signature model are investigated and a linear regression base model is constructed. The training data set consists of daily data for the 25 dwellings during the first heating season 2011-2012.

3.2.1 Exterior temperature

The exterior temperature *Te* is the main weather predictor for the space heating energy use. Because of phenomena such as thermal inertia, today's space heating energy use is also influenced by the conditions of the building and climate of the previous days. One way to include this in the model is to replace the temperature predictor by the equivalent temperature *Te,eq*, that is a weighted average of temperatures of today and the two previous days (see chapter 2.2). Another way is to include the exterior temperature of the previous day(s), also called time lags as separate predictors in the model. The coefficient (or weighting) of each predictor is then estimated in the linear regression model, that is described by for example equation 3-I:

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Te_{t-1} + c_4 \times Te_{t-2} + \varepsilon_t \qquad [3\text{-I}]$$

Figure 3-3 shows spider diagrams of the calculated R²adj and relative RMSE for the models with exterior temperature variables, applied on the 25 case-study data-sets. The relative RMSE is here the RMSE of the model in comparison to the model with *Te* as a predictor (therefore the value is 1 for the *Te* model LM6). The spider diagram shows that for all cases the predictors that take into account previous temperatures, the RMSE reduces (or remains equal), and thus the accuracy of the in-sample predictions is higher. In most cases, the use of the temperature time lags as predictors in the model leads to the highest accuracy since the relationship between the different time lags can now be estimated instead of being imposed. Similarly, the adjusted coefficient of determination R²adj is up to 7 pp higher (better) for the models that include historical values.

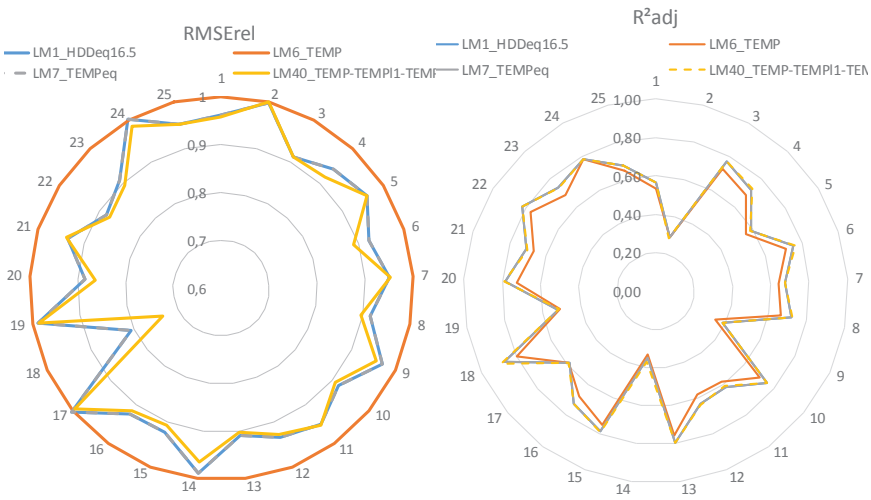


Figure 3-3: Linear regression - Temperature predictors – RMSE and R²adj

Note that exterior temperature predictors are often replaced by heating degree days, for example the equivalent heating degree days HDDeq with base temperature of 16.5°C that are applied in Belgian Gas consumption normalisation (see chapter 2.2). In the settings of this study, where only space heating data is used during the heating season, the use of heating degree days (e.g. $HDDeq_{16.5}$ instead of equivalent temperatures, does not make a big difference.

From the analysis of the exterior temperature variables it is confirmed that the integration of previous days' temperatures into the model is a benefit. The use of time lags as individual predictors in the model often leads to a little higher in-sample prediction accuracy than the use of an equivalent temperature.

3.2.2 Other weather variables

Other available weather inputs are global solar radiation on a horizontal surface *Rg* (RADglob), relative humidity *Hr* (RELHUM), precipitation quantity *Pq* (PRECIP), the wind speed *Ws* (WINDSPEED) and wind direction *Wd* (WINDDIR). First, the addition of these weather variables is explored, starting from a model with exterior temperature *Te* as primary input (LM6). Precipitation quantity and relative humidity coefficients prove to be not insignificantly different from zero in the majority of the models and lead to negligible model improvements. Both solar irradiation and wind speed are the most important additional weather variables, significant at a 5% level in 24/25 cases. Figure 3-4 shows improvements of the adjusted R^2 between 4 and 26 pp and reductions of the RMSErel between 4 and 35% by adding *Rg* and *Ws* to the equation (LM50). Therefore, it is concluded that equation 3-II (LM66) presents a good base Energy Signature model:

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Rg_t + c_4 \times Ws_t + c_5 \times Te_{t-1} + \varepsilon_t \qquad [3\text{-II}]$$

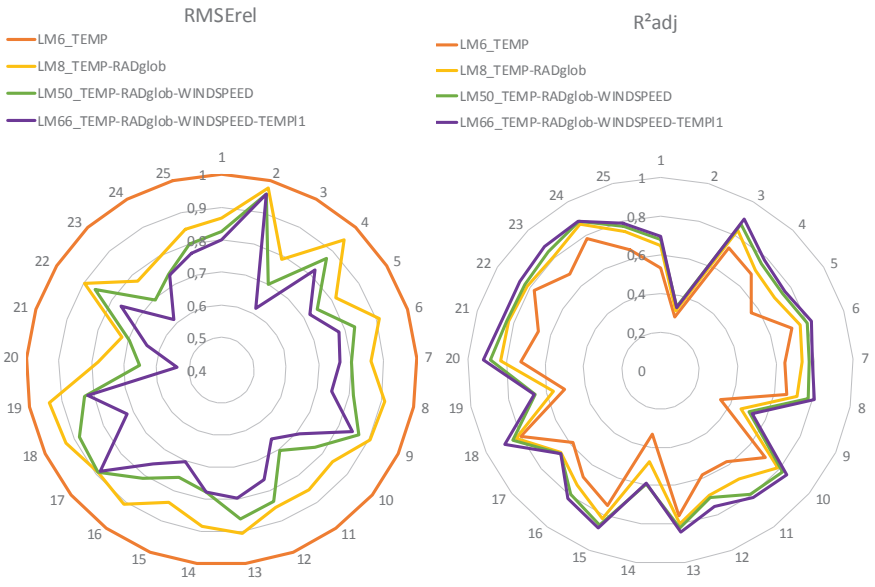


Figure 3-4: Linear regression - Weather predictors – RMSE and R²adj

Wind speed variables

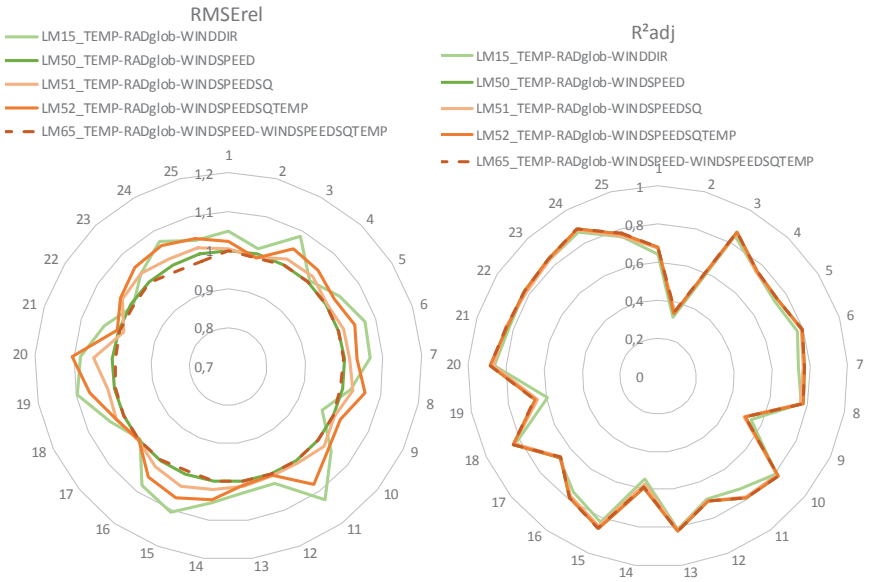


Figure 3-5: Linear regression - Wind speed predictors – RMSErel and R²adj

The heating demand of a building is influenced by uncontrolled air infiltration through the building envelope as well as natural or mechanical ventilation. The ventilation and infiltration heat losses are related to the wind speed, the temperature of the exterior air as well as the wind direction (e.g. due to the setting of the building in the landscape or the orientation of the main air leaks, the effect of wind may be different depending on the wind direction). Now the following alternative wind variables are investigated: *Wd* (WINDDIR), Ws^2 (WINDSPEEDSQ), Ws^2Te (WINDSPEEDSQTEMP), *Ws,n* (WINDSPEEDnorth), *Ws,e* (WINDSPEEDeast), *Ws,s* (WINDSPEEDsouth), *Ws,w* (WINDSPEEDwest) and combinations of the wind speed for specific orientations {*Ws,n*, *Ws,e*; *Ws,s*; *Ws,w*}. In Figure 3-5 and Figure 3-6 models with the different wind variables are compared to a model using simply *Ws* as a variable and RMSErel is now also relative to this model (LM50).

The adjusted R^2 of models using *Wd*, Ws^2, Ws^2Te or {Ws^2Te, *Ws*} is very close to the reference model's R^2adj, for all cases an also the accuracy of the in-sample predictions is similar or slightly lower than for the model with Ws (see Figure 3-5). For the model with both inputs {Ws^2Te, *Ws*} (LM65), the p-value of the coefficient Ws^2Te does not indicate significance on the 5% level for 21/25 cases (and for all of them on the 1% level). Therefore, it is found that the addition of Ws^2Te to a model which already carries *Ws* is not actually valuable, and neither is the replacement of *Ws* by *Wd*, Ws^2 or Ws^2Te.

The inclusion of the wind speed for one of the directions North, East or South leads to lower or equal model fits for all cases. Especially for direction North, the R^2adj decreases and in 22/25 the coefficient is not proven significantly different from zero. On the other hand, the wind speed from the West results in a significant coefficient in 24/25 cases and in an R^2adj close to the reference case with *Ws* and less than 8% increase in RMSE. In Figure 3-6 combinations of wind speed from multiple directions are added to the comparison. It is found that models with only

Ws,w or a combination of {*Ws,w* ; *Ws,s*} on average lead to a similar prediction accuracy and similar goodness-of-fit, meaning that they are slightly worse for some cases and slightly better for other cases, when compared to the reference with *Ws*. The models with wind speed for all 4 wind directions (LM58) or the combination of {*Ws,e;Ws,s;Ws,w*} (LM57) always lead to a slightly higher prediction accuracy than the reference model. The decrease in RMSE is on average 2%. Ws,e and Ws,s are not proven significant in respectively 14/25 and 8/25 cases, while Ws,w is significant in 24/25 cases.

Since in the R²adj the goodness-of-fit is adjusted for the amount of variables used, it seems that the improvements in model fit between the reference model and model {*Ws,e;Ws,s;Ws,w*} are balanced with the net addition of 2 more variables: both models give very similar results, with small improvements in accuracy when wind speed is entered for three wind directions. It is concluded that a model with Ws as input is a good choice, however when the wind direction Wd is available as an input, in general, a model with {*Ws,e;Ws,s;Ws,w*} is preferred, since small improvements in in-sample prediction accuracy can be obtained. If case-specific modelling is allowed, removing Ws,e or Ws,s can be considered too.

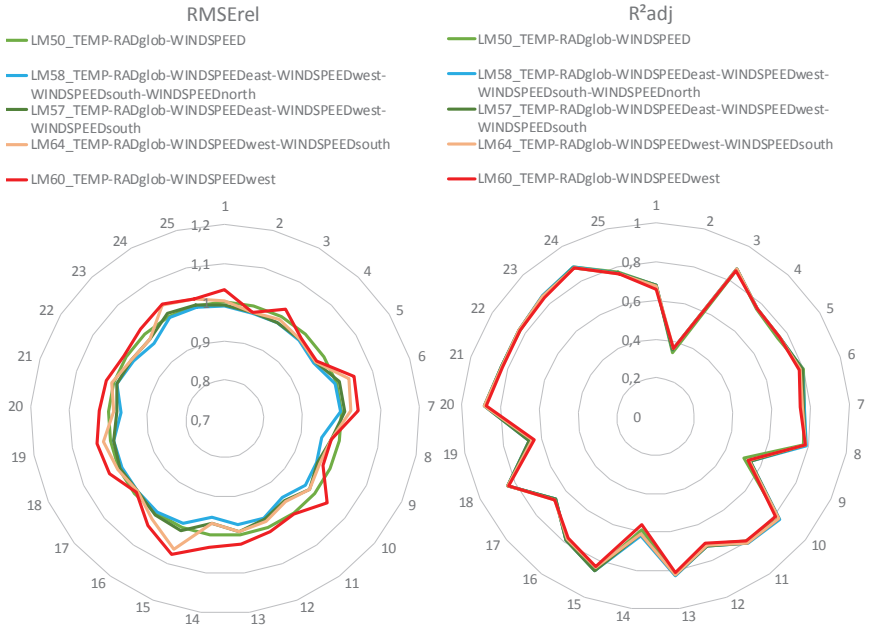


Figure 3-6: Linear regression – Wind speed by Winddir predictors combined – RMSErel and R²adj

Additional weather variables

Further additions to the base model (LM66, equation 3-II) and the alternative base model with wind speed for 3 wind directions (LM67) are now considered. As was mentioned, the addition of precipitation or relative humidity have no or a very limited effect on the model quality and in-sample predictions for most dwellings (see Figure 3-7). Moreover the coefficients are not proven to be significantly different from zero in most cases (23/25 for Pq and 18/25 for Hr[2]). Figure 3-7 also shows the effect of adding of a first time lag of the solar irradiation input or the wind speed input. The influence of these additions on the RMSE and R²adj is negligible. Regarding solar radiation, the variable Rg_{t-1} is significantly different from zero in 5/25 cases, but precisely in these cases it provokes that the significance of Rg_t cannot be proven anymore. Regarding wind speed, the variable Wst-1 is significant in 17/25 cases, but again it provokes lower higher p-values for the other wind speed variables in the model.

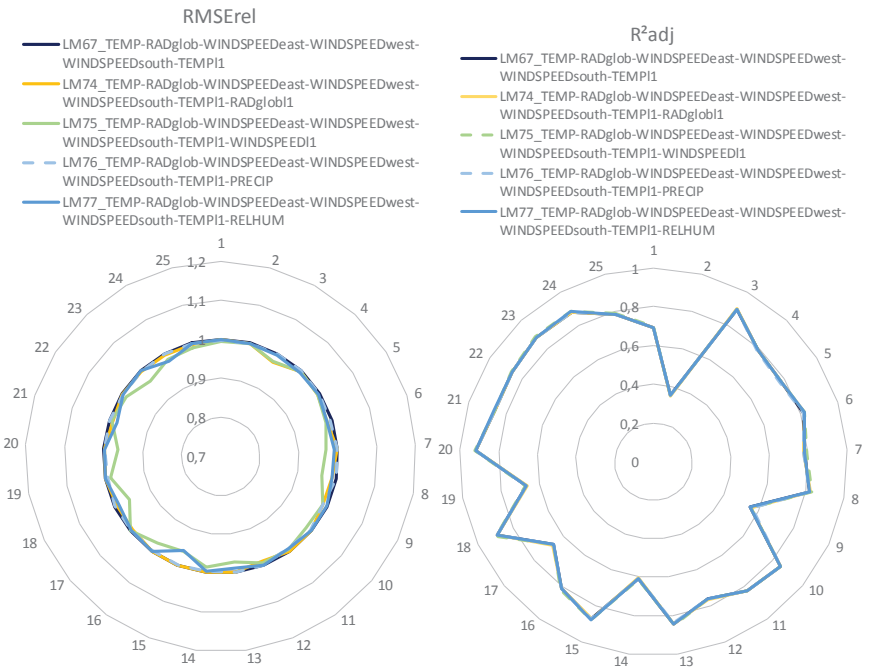


Figure 3-7: Linear regression – Additional weather and lagged predictors – RMSErel and R²adj

[2] The variable Hr is correlated with Rg with a correlation coefficient of 0.65. When Hr is added to a model which already contains Rg, the VIF of Rg rises from 1.3 to about 2.9 and the standard error of the coefficient doubles. VIF of Hr is about 2.5. Therefore the significance of the coefficient for Hr in some cases might be due to this correlation with Rg (which has significance on the 1% level for all cases where Hr is significant).

3.2.3 Validation with out-of-sample data

In this section some model quality and in-sample forecast accuracy measures are summarised for a selection of models. Then the selected models are applied on a test data set and the out-of-sample forecast accuracy measures are discussed. The selection includes the simplest model that uses only exterior temperature Te as an input (LM6), and gradually the most relevant inputs – global horizontal solar irradiation Rg, wind speed Ws,e (East), Ws,s (South) and Ws,w (West), and the first time lag of the exterior temperature Te_{t-1}– are added to this model, resulting in a candidate base model LM67. The most relevant extra input to this candidate model would be the second time lag of the exterior temperature Te_{t-2} (LM72), which results in little higher accuracy and is significant (together with significant Te_{t-1} and Te_t) in 6/25 cases. The equation for the candidate base model is:

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Rg_t + c_4 \times Ws,e_t + c_5 \times Ws,w_t + c_6 \times Ws,s_t + c_7 \times Te_{t-1} + (c_8 \times Te_{t-2}) + \varepsilon_t \quad [3\text{-III}]$$

In case a simpler model is required or if the wind direction data is not available, the parameters representing the wind speed for the cardinal orientations can be replaced by the overall wind speed Ws (LM66 (LM71)), with little loss of in-sample accuracy:

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Rg_t + c_4 \times Ws_t + c_5 \times Te_{t-1} + (c_6 \times Te_{t-2}) \ \varepsilon_t \quad [3\text{-IV}]$$

In case an automated variable selection procedure is desirable, allowing to select the best model for each case specifically, it is proposed to follow a forward selection procedure testing the variables in the order that is described above, and retaining all significant variables.

Training Data

The measures for these selected models when applying to the training data-set (in-sample predictions) are presented in Figure 3-8(b), together with the goodness-of-fit indicator R^2adj of the models (a). In the simplest model, the R^2adj is on average (over all the cases but case 2) 0,64 and takes values between 0,33 and 0,80. The average R^2adj rises to 0,80 for the candidate base models (LM67 and LM66), and takes a minimum of 0,53 and maximum of 0,92. Only in case-study 2 the R^2adj remains between 0,29 and 0,35 for all models considered. This dwelling is somehow an exception, since it is heated very irregularly and often remains unheated during several days, indicating occasional and irregular occupation. Because of this, it is expected that the heat use will be more defined by whether or not the building is occupied than by climate variables.

In Figure 3-8(b) the RMSE is presented relative to model LM67. This graph illustrates the gradual increase in prediction accuracy when adding Te, Rg and Ws to the models, then the very similar results for both candidate models LM66 and LM67, that both include Te_{t-1}, and finally reductions in RMSE for some cases when also Te_{t-2} is added as a variable.

Test Data

The selected models, are now applied to the test data set (heating season 2012-2013). In Figure 3-8(c) the RMSE of the out-of-sample predictions is presented, also relative to model LM67. This graph illustrates whether the order of succession (in terms of accuracy) is maintained when they are applied on the test data set. In contrast to the training data set results, it is found that model LM8 performs worse than the simple model LM6 in a number of dwellings, but in general for the more comprehensive models, prediction accuracy increases. With regard to the candidate base models, the model with one wind speed input Ws (LM66) now performs on average little better than model LM67 using {*Ws,e; Ws,s; Ws,w*}. Addition of the second temperature lag again leads to small improvements in a number of cases.

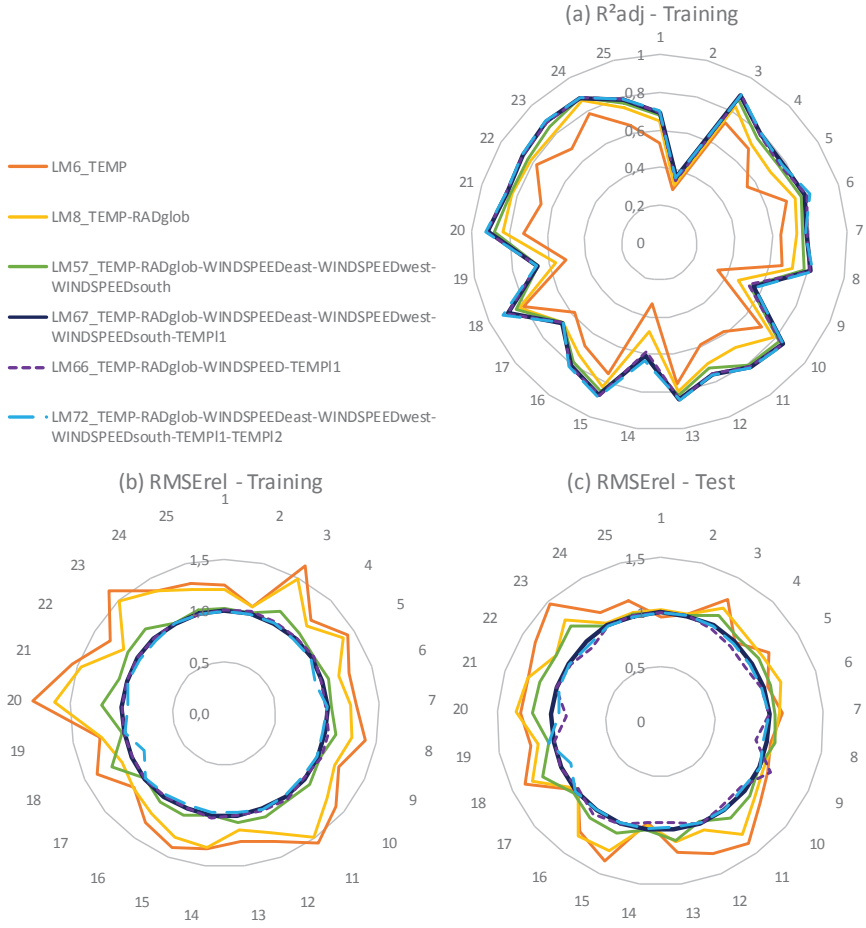


Figure 3-8: Linear Regression – Training & Test data set – RMSErel (LM67) and R²adj

3.3 Model validity

Previously, a process of model selection and evaluation lead to an Energy Signature base model LM66 (eq. 3-II), that has an acceptable structure and simplicity for all studied cases, yet it can be slightly improved or enriched for specific cases, by adding or replacing specific variables. The most valuable additions or replacements for specific cases are: addition of Te_{t-2} (eq. 3-IV), replacement of *Te* with *Te,eq*, or replacement of *Ws* with (a combination of) *{Ws,w; Ws,s; Ws,e}*. Eventually, the additions of *Hr*, *P*, Ws_{t-1} or Rg_{t-1} only lead to minor increase in the goodness-of-fit of the model. In order to make sure that the outcomes of the linear regression model, such as the forecast and confidence intervals, can be trusted, it is important to investigate whether the assumptions of linear regression modelling are met. The main regression diagnostics were summarised in section 3.1. In this section the model validity for the base model will be illustrated in detail by use of an example case and then generalised for all the cases.

3.3.1 Example: case 6

In illustration of its statistical validity, the regression diagnostics of the linear regression base model for one of the cases is discussed in detail. Case 6 is selected for this purpose, which is the 6th dwelling in row when the cases are ordered by increasing total energy use over the test heating season (see Figure 2-4). The dwelling used about 8.5 MWh of energy for space heating during the test year and on average 52 MWh per day. In terms of the statistic measures discussed in previous sections, this dwelling is situated close to the averages. The goodness-of-fit measure R^2adj is 0.82 (average 0.79), the MAE of the in-sample predictions is 9 (average 13) and the nMAE is 7% (average 7%). The MAE of the out-of-sample predictions is 14 (average 16) and nMAE 10% (average 10%). Figure 3-9 presents the regression diagnostic plots for case 6 and model LM66. It is found that in general the model assumptions are sufficiently fulfilled. The *Residuals vs Fitted* and *Scale-Location* plots, and the time series plot of the residuals (Figure 3-10), show rather constant variance over the range of observations and over time, so the assumption of homoscedasticity is accepted. The normal Q-Q plot follows a straight line, with only very small deflections in the outer ends, indicating the errors have a distribution close to normal. The outlying points are found not very influential on the estimated model in this case and they are kept within the data (Plots with Cook's distance and Leverage). Deviations from linearity are detected using the plots of the residuals vs the predictors, when the model LM66 is used, no clear violations are found. Also the autocorrelation function (ACF) indicates no heavy dependency of the residuals against time. It is further discussed in section 3.7.

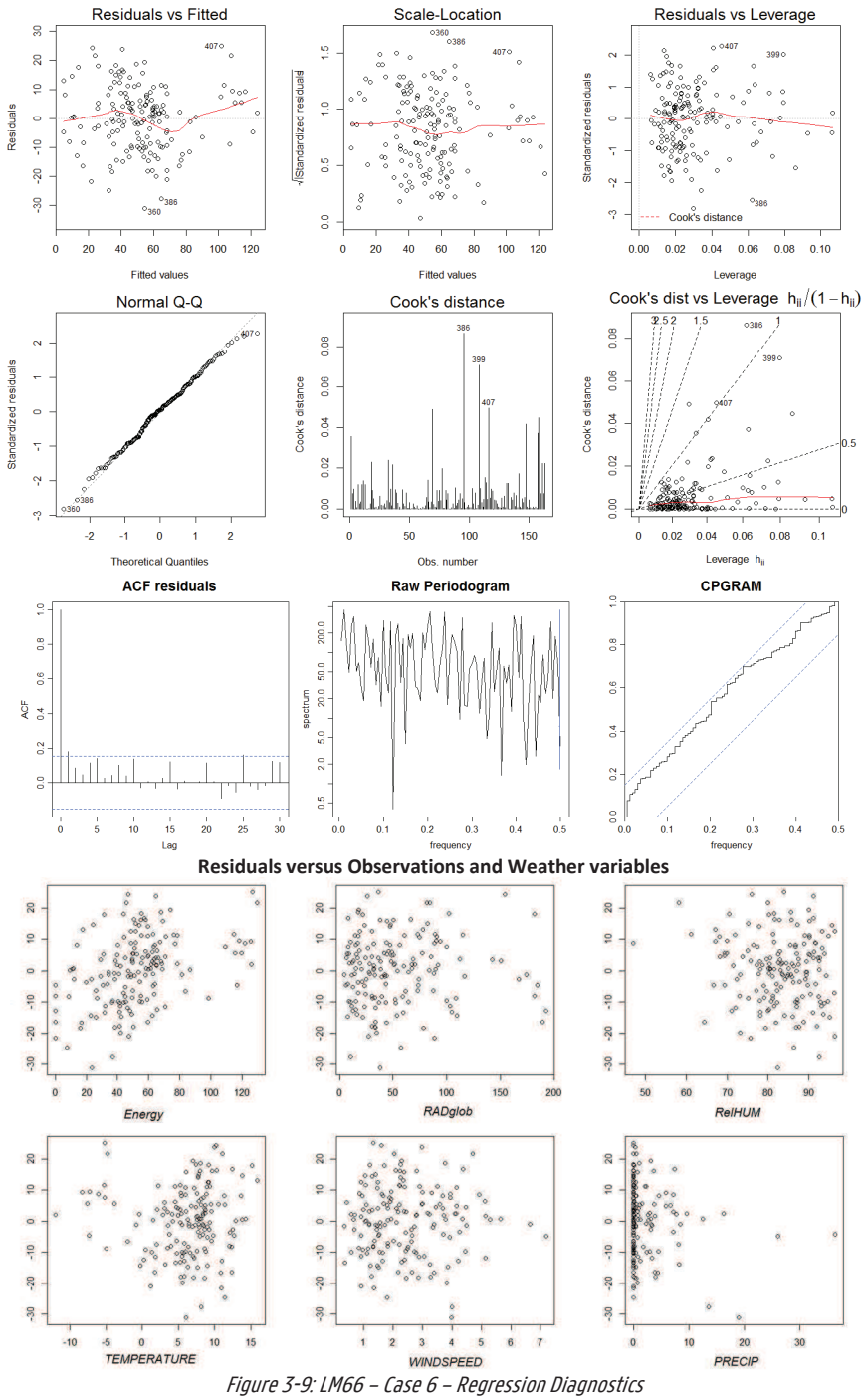


Figure 3-9: LM66 – Case 6 – Regression Diagnostics

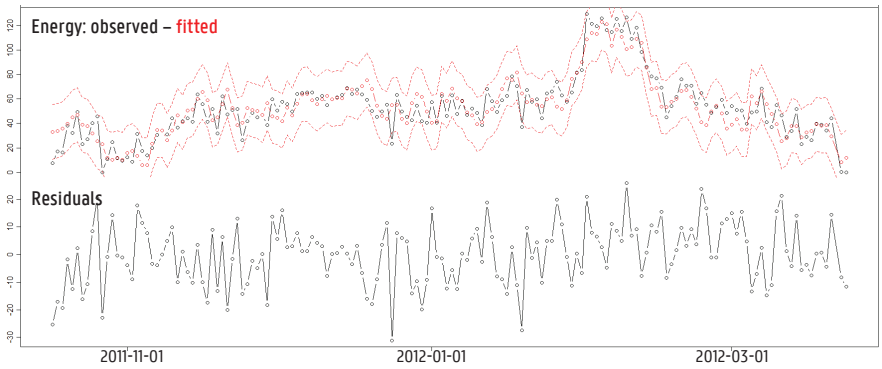


Figure 3-10: LM66 – Case 6 – Training data – Time series of the energy use (with 95% PI) and residuals

3.3.2 All cases

Zooming out to the diagnostics for all the cases when applied to **the training heating season 2011-2012**, the main issue is the appearance of small or larger deflections at the outer ends of the Q-Q plot, typically due to a few outlying observations. When taking a closer look to these observations, two phenomena are distinguished throughout the cases involved. First, influential observations appear when the energy use is zero during one or several consecutive days during the heating season. Secondly, high Cook's distances are sometimes observed around the beginning of February, when the exterior temperature was very low. Thirdly, other non-systematic outliers appear in some cases. The last two groups of outliers could be explained by various reasons related to behaviour or system characteristics, but based on the available information, we cannot assume that they are not belonging to the actual system that is modelled, that is the interaction between the building, system and user during operation. Therefore, it is decided to keep these outliers within the data. On the other hand, the first group of outliers is very probably due to people turning off the heating when they are on holidays or the heating system is broken. This is also observed more severe in case 2, where no energy is used for more than half of the days, probably due to irregular occupation. For this group of outliers, it is decided to remove them from the data, since the goal is to model the energy use during periods of occupancy rather than un-occupancy[3]. This action has effect on 15 cases, where up to 7 outliers are deleted. As a result, normality improves and for 12/25 cases it is accepted that the model assumptions are fulfilled. They are marked in green in Table 3-2. Amongst the remaining cases, marked in red, is case 2, for which now the model assumptions are rather acceptable, but still a very low goodness-of-fit (R^2adj = 0,11) and the coefficients of the weather variables are not significant. For the other remaining cases, the assumption of the independency of errors is violated (this specific assumption will be discussed more extensively in section 3.7).

[3] In order to avoid the deletion of observations with energy use zero that are not due to absence, but rather expected low energy uses due to temporarily high exterior temperatures during the heating season, only those observations are deleted that have energy use of zero and belong to the cases with the 10% highest Cook's distance when the LM66 model is applied.

Most of them also display deviations from normality (deviations from the diagonal at one or both tails in the Q-Q plot), as well as signs of heteroscedasticity in the *Scale-Location* plots (trend-line deviating from the horizontal), showing increasing variance at higher energy use levels (e.g. case 16,25) or specific periods of time (e.g. case 8,18). This is illustrated for case 18 in Figure 3-11.

Table 3-2: Linear Regression - Model Validity Summary

Validity		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2011-2012	*	1	0	1	0	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
2012-2013	°	0	0	1	0	1	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0

Instead of using it as a test period for out-of-sample predictions, the **second heating season 2012-2013** can also be used as a training data set (e.g. to estimate the model coefficients in section 3.5). The model validity then needs to be checked again. It is also summarised in Table 3-2. For many cases, the model validity conclusions are similar for both heating seasons, but for some of them differences appear. For example in case 1, the energy use was continuously zero during more than half of the second winter period, so probably the house was temporarily unoccupied.

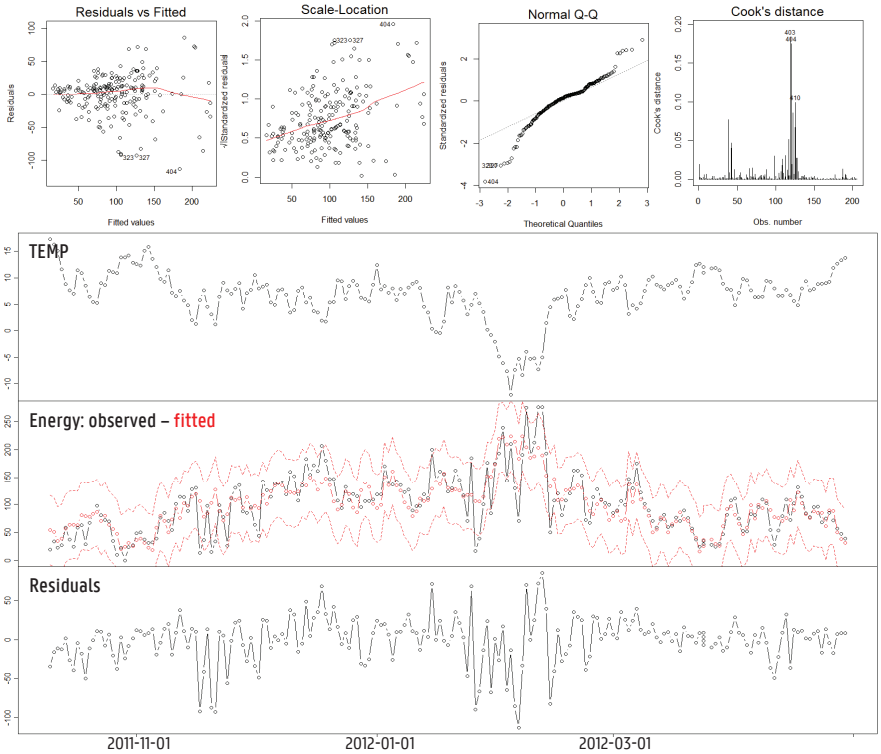


Figure 3-11: LM66 – Case 18 – Diagnostics; Time series of the energy use (with 95% PI) and residuals

3.4 Prediction accuracy

The Energy Signature base model (LM66) can be used to estimate the energy use Q_t for combinations of input values of the predictors. The outcome of this operation is called a prediction. When applied on the training data set (on which the model was fitted), the predictions are called in-sample predictions *Is*, and when applied on a test data set, they are called out-of-sample predictions *Os*. Remark that the model is fitted on a certain training data set in which a certain range of the predictor values appear. Therefore, the model's validity is documented for this range of predictor values, and for example the 95% prediction intervals are valid for predictions within this range. Both training and test weather data sets are described in section 2.2 and Annex A. It is found that the ranges of the main predictor values fairly overlap, only some higher solar radiation values appear in the upper quart of the test data set. The specificity of the model fit for a certain training data set also implies that they are valid for the system (the building-services-user-interaction) during the training data period. If the system is altered by time of the test period, the predictions might be far from the actual observed values in the test data period[4]. In reverse, this characteristic is a key for determining changes in the system through deviations between the predicted and observed values in a test data period.

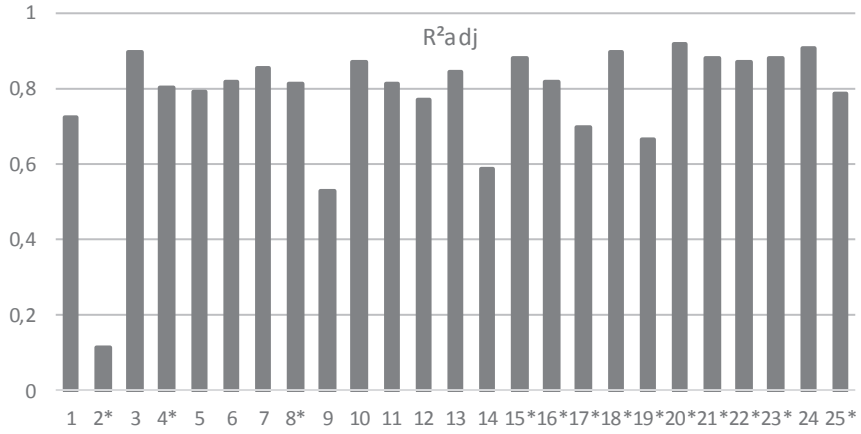


Figure 3-12: LM66 – R²adj

[4] Moreover, if the system is changed importantly for a part of the training data period, this will affect the model itself (e.g. prediction intervals can become larger, residuals for a certain period being different). This might be a reason for some of the cases described in the previous section, where model assumptions were not fulfilled.

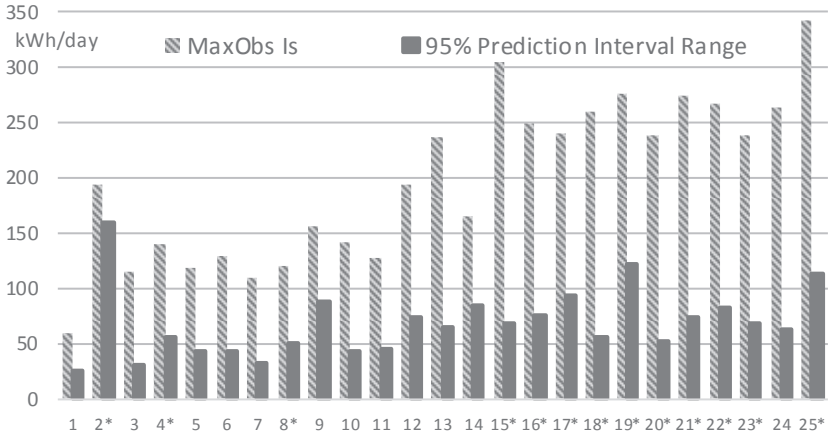


Figure 3-13: LM66 – 95% PI range and in-sample Maximum Observation

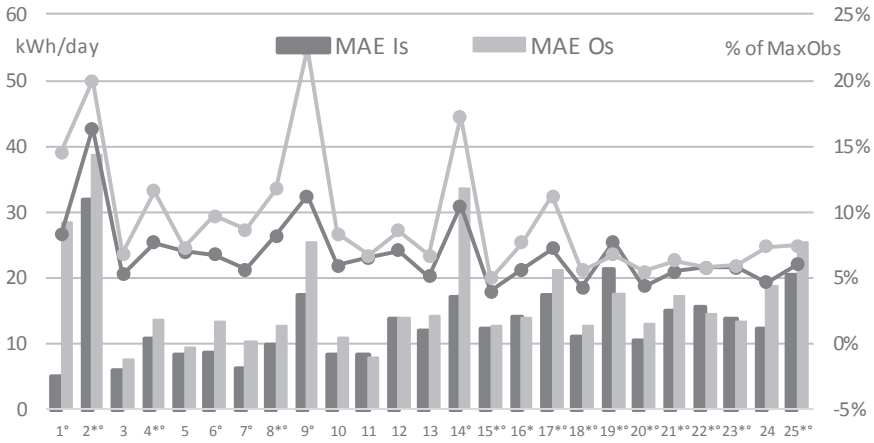


Figure 3-14: LM66 – MAE (bars) and nMAE (lines) for In-Sample and Out-of-Sample predictions

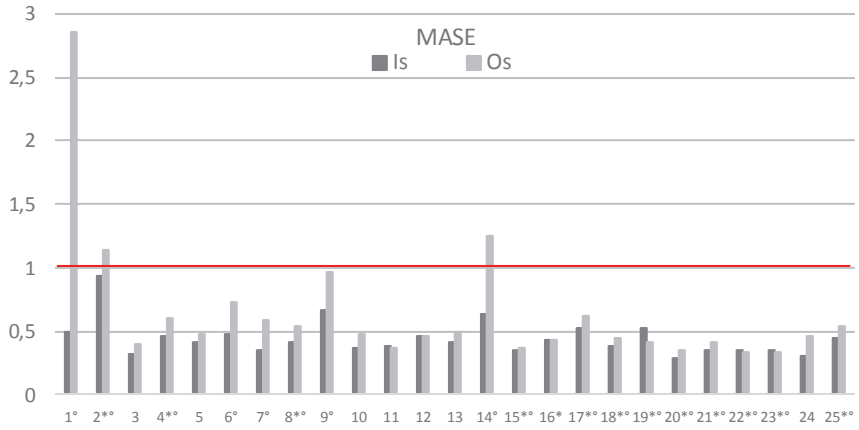


Figure 3-15: LM66 – MASE for In-Sample (Is) and Out-of-Sample (Os) predictions

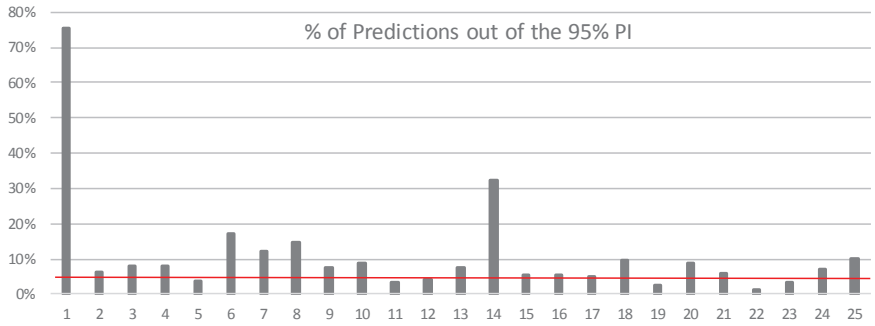


Figure 3-16: LM66 - % of Out-of-Sample predictions out of the 95% PI

For the sake of completeness, the goodness-of-fit of the Energy Signature base model[5] is documented in Figure 3-12 by use of the adjusted R^2. Taking into consideration the 12 cases for which the linear regression model is valid, the average R^2adj is about 0.78 and reaches to 0.91. When a prediction is made, the uncertainty of the prediction can be expressed by use of a prediction interval with a certain probability (in this study 95%), which signifies the range of values between which there is a 95% probability that it contains the actual observed value. Therefore, the ranges of the prediction intervals (as the difference between its upper and lower boundary) indicate at what level of precision differences between predicted and actual values can be determined. The prediction interval ranges are illustrated in Figure 3-13 together with the maximum observed daily energy use values. For the 12 considered cases, the prediction intervals have ranges between 26 and 87kWh/day. In the worst case, a difference of more than 80kWh/day between an observed a predicted value can fall within the 95% prediction intervals. More in general, it can be understood that by use of this Energy Signature model, it could be very hard to identify that there is a significant difference in energy use between training and test data periods of, for example, 10kWh/day. An often used measure for prediction accuracy is the Mean Absolute Error (MAE), which is the mean of the absolute values of the errors and is on the scale of the data (Figure 3-14). The smaller the MAE, the smaller will be the prediction errors, as well as the prediction intervals. For the in-sample predictions of the models valid for training and test periods, the MAE is between 5 and 18kWh/day. It is on average 6.7% of the maximum observed daily energy use in the training data set. For the out-of-sample predictions for the same cases, the average MAE error is a little higher, on average 11.8 kWh/day, or about 7.4% of the maximum observed values in the test data set.

Another measure for forecast accuracy is the Mean Absolute Scaled Error (MASE), that has the benefits that it is independent of the scale of the data, and is valuable when comparing the results of the models applied on different data sets (see section 3.1). In Figure 3-15, for the in-sample predictions, all values are lower than one, indicating that the in-sample forecasts of the Energy Signature model have a better performance than the naïve forecast method. For case 2, where most regression coefficients are insignificant and the adjusted R^2 is very low, the MASE is close to 1. Regarding the test data sets for the cases where the model was found valid in both

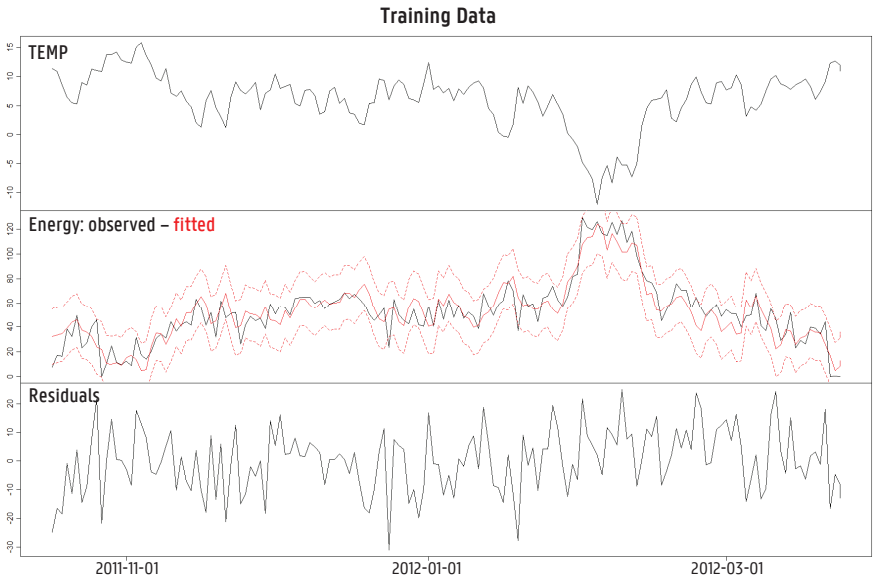
[5] Remark that in this model, the outlying observations with zero energy use are deleted, as discussed in section 3.3.2. The cases for which model validity was not entirely reached, are marked by use of the symbol * (training data) and ° (test data).

training and test periods, the in-sample MASE is between 0.30 and 0.46. The out-of-sample MASE is between 0.36 and 0.49 which is 93% to 153% of the in-sample MASE. The out-of-sample prediction accuracy is between 93% and 153% of the in-sample prediction accuracy (as indicated by MASE), and is for most cases a little lower than the accuracy expected from the in-sample predictions. For some of the cases for which the model assumptions were not valid, the out-of-sample MASE is clearly worse and even worse than the in-sample naïve forecasts. One example is case 1, that is probably unoccupied for more than half of the winter. For other cases, e.g. case 9 and 14, the high difference in MASE goes together with violation of the model assumptions only for the test data set, but research methods used up to know, do not allow to pinpoint a cause for this.

Finally, the amount of out-of-sample predictions that lie out of the prediction interval is counted and presented in Figure 3-16. In theory, we would expect the percentage of points out of the PI to be around 5% when the model fits well and the test and training data set come from the same system. Large percentages appear for again case 1 and 14. Also for case 6, 7 and 8 more than 10% of the predictions lie out of the 95% PI. For example, in Figure 3-17 the time series of actual and predicted data for the training and test sets is presented. When observing the out-of-sample predictions, it is observed that in February and March 2013 about 24 predictions are higher than estimated. This could be due to a change in the building-system-user interaction, but it could as well be due to a model deficiency.

In this section the in-sample and out-of-sample prediction accuracy of the models was documented using different accuracy measures, that is the range of the prediction interval, the MASE, MAE and nMAE, and the percentage of predictions out of the prediction intervals. The 95% prediction interval range and the MAE illustrate the uncertainty that goes together with predictions in the linear regression base model, which is found to be rather large for most of the models. When applied on the test data set, the MASE increases for most cases, but not dramatically (maximum factor 1.5). Furthermore, some cases with very low test prediction accuracy are pointed at.

For most of them, it is difficult to make sure whether this is a model defect (the model needs improvement) and/or an actual difference in energy use between the datasets, nor to direct towards causes for such a difference.



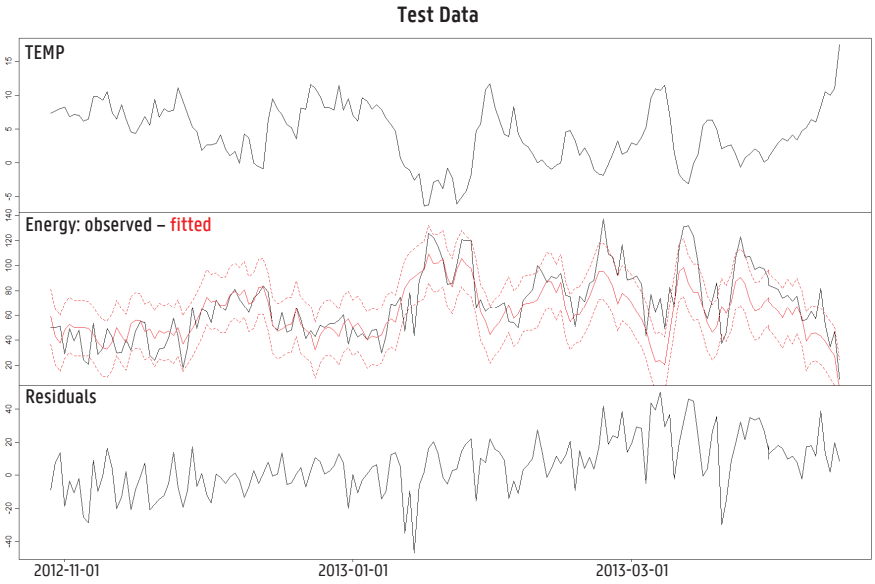


Figure 3-17: LM66 – Case 6 – Training data and Test data – Time series of the energy use (with 95% PI) and residuals

3.5 Energy Signature Coefficients

While in the previous section the focus was on the output of the energy use expression, this section is dedicated to the estimated coefficients or model parameters. If the statistical quality of the model is sufficient, the estimated model coefficients can be understood as the Energy Signature of the building for the given energy-balance model [29]. This set of parameters characterises the relationship between the building's energy use for space heating and the main climate parameters for the observed period. Since in this study the buildings are observed in fully operational conditions, the Energy Signature actually characterises the complete system of the building, the building services and the occupants. Therefore, it is important to mention that the estimates do not equal the parameters of the pure building physical heat loss calculations, such as the heat loss coefficients or solar apertures. Instead the estimates are influenced by the complex interaction of the building with the systems and the users. On the other hand, significant differences in Energy Signature of a certain building over different periods can be used as an indicator for changes in the energy use of the building, due to changes in (the interaction of) the building, the system or the occupants.

In addition to the regression diagnostics discussed in the previous section, now the statistics that are related to the different predictors are illustrated, such as the significance of the coefficients, their estimates and standard errors. Also the issue of multicollinearity between predictor variables is discussed. This statistical phenomenon is important specifically when the model coefficients are considered. In the first part of this section, the coefficients and their statistics are discussed in detail for the example case 6. Then the Energy Signature model is applied separately on the data sets for heating season 2011-2012 (the former training data set) and 2012-2013 (the former test data set) and the coefficients are presented. Finally, it is investigated whether the coefficients of the two data sets are significantly different, by use of an Energy Signature model with interactions.

3.5.1 Coefficient statistics

Table 3-3 presents data on the model coefficients for the Energy Signature base model LM66 when applied to case 6 in the heating season 2011-2012. Statistical validity was found sufficient for the model applied on this case. Based on the p-values of the coefficients it is accepted that the coefficients are significantly different from zero, even if for the solar irradiation parameter this is the most uncertain (the p-value is more close to 0.05). Specifically with regard to the validity of the coefficients, the Variance Inflation Factors (VIF) are listed as indicators for multicollinearity amongst the predictors (see section 3.1). For *Ws* and *Rg* predictors, the VIF is close to 1, meaning that the correlation of these predictors with other predictors is low. On the other hand, the VIF of predictors Te_t and Te_{t-1} is close to 5, indicating multicollinearity issues. More specifically, Te and Te_{t-1} are related variables, since one is the value of the other at the previous time step. Therefore, there is a risk that the precise effect or estimate of the exterior temperature on the energy use is influenced by Te_{t-1} and vice versa, and the standard errors for both coefficients increase. However, if the general relation between the exterior temperature and the energy use is of interest, instead of the specific coefficients c_2 and c_5 in equation 3-IV, this issue can be overcome. As explained in Madsen et al., the steady state temperature coefficient c_{Te} is the stationary gain of the transfer function from the exterior temperature input and the uncertainty

of the obtained coefficient can be calculated by the first order error propagation formula [40,75]. As a result, in LM66 the estimate $c_{Te} = c_2 + c_5$ is 4.79 kWh/°C with a standard error of 0.19. The estimates and standard errors for the other weather variables can simply be found in Table 3-3.

The signs of the coefficients are conform physical logic: an increase in exterior temperature or solar irradiation will generally result in a decrease in space heating energy use and an increase in wind speed (during the heating season) will result in an increase in space heating energy use. Since all coefficients have a different unit, their relative effect on the outcome is hard to see. Therefore in the table the coefficients are also standardised to the variances of the dependent and independent variables. For example, an increase in 1 standard deviation in exterior temperature will have an effect about 4 times larger on the energy use than an increase of 1 standard deviation in wind speed and about 10 times larger than one standard deviation increase in solar radiation. For comparison the coefficients table of LM50 is also presented, which does not include the predictor Te_{t-1}. The obtained coefficients and standard errors are very similar to those obtained with the full base model, also with regard to the steady state exterior temperature parameter that was calculated above.

Table 3-3: Coefficient Statistics – Case 6 – Heating season 1 - LM50 and LM66

LM66		**Coefficients**						**Partial correlations**	
	Unit	Estimate	Std. Error	t value	p-value	Standardised coefficient	VIF	Estimate	p-value
(Intercept)	kWh	74,11	2,84	26,12	0,0000				
RADglob	kWh/(W/m²)	-0,05	0,02	-2,23	0,0272	-0,08	1,2	-0,18	0,0245
TEMP	kWh/°C	-2,89	0,40	-7,27	0,0000	-0,54	4,8	-0,49	0,0000
TEMP.l1	kWh/°C	-1,90	0,40	-4,71	0,0000	-0,36	4,8	-0,38	0,0000
WINDSPEED	kWh/(m/s)	4,48	0,73	6,12	0,0000	0,23	1,2	0,45	0,0000

LM50		**Coefficients**						**Partial correlations**	
	Unit	Estimate	Std. Error	t value	p-value	Standardised coefficient	VIF	Estimate	p-value
(Intercept)	kWh	72,81	3,01	24,15	0,0000				
RADglob	kWh/(W/m²)	-0,05	0,02	-2,29	0,0234	-0,09	1,2	-0,18	0,0234
TEMP	kWh/°C	-4,58	0,19	-23,71	0,0000	-0,86	1,0	-0,88	0,0000
WINDSPEED	kWh/(m/s)	4,61	0,78	5,90	0,0000	0,23	1,2	0,43	0,0000

3.5.2 Model coefficients

The Energy Signature model is now applied separately on the data from heating season 2011-2012 (the former training data set) and heating season 2012-2013 (the former test data set) and the coefficients are estimated. Model validity for the models, when applied to each of the datasets, is summarised in section 3.3.2. In Figure 3-18 the Energy Signature coefficients and their 95% confidence intervals are presented. The steady-state exterior temperature coefficient is calculated from the coefficients of *Te* and Te_{t-1} following the approach described in the previous section 3.5.1. The cases are ordered by increasing total energy use over the heating season 2011-2012. The estimates of the coefficients tend to follow the increasing trend, especially c_{Te} and the intercept coefficient. From physical point of view this is also understandable: for example the coefficient c_{Te} is related to the building heat loss coefficient and mean interior temperature, and in general, buildings with higher heat loss coefficients tend to have higher space heating energy demand. Even if the actual space heating energy demand and c_{Te} are influenced by many other aspects, such as user behaviour, when observing this increasing trend over the whole sample, that confirms the theoretical expectations. The *Te* and *Rg* parameters are mostly negative and the *Ws* parameter mostly positive, which follows physical logic. Exceptions are found for some cases for the Rg and Ws parameters, where the confidence intervals reach both positive and negative values. For these cases, it is concluded that these coefficients are not significantly different from zero. This may be for physical reasons (e.g. a dwelling that is less exposed to wind…). But more probably, it is due to the inability of the linear regression model to describe the data properly, for example because of changes in the system that is modelled during the measurement period, influential "outliers" or exceptions to the normal situation, or un-modelled phenomena.

By comparing the coefficient estimates for the two heating seasons, changes in energy signature between the two heating seasons can be detected. However, since for many cases the model assumptions are not fulfilled, the width of the confidence intervals may be underestimated. Therefore, the comparison between the two heating seasons will be treated in a later chapter.

The standardised coefficients for all cases are presented in Figure 3-19. Coefficients that are not proved to be significantly different from zero are marked with a red dot. Exterior temperature indeed proves to have the highest effect on the energy use. The effect of solar radiation and wind speed is case-dependent (which one has the highest effect), but not negligible in the majority of cases.

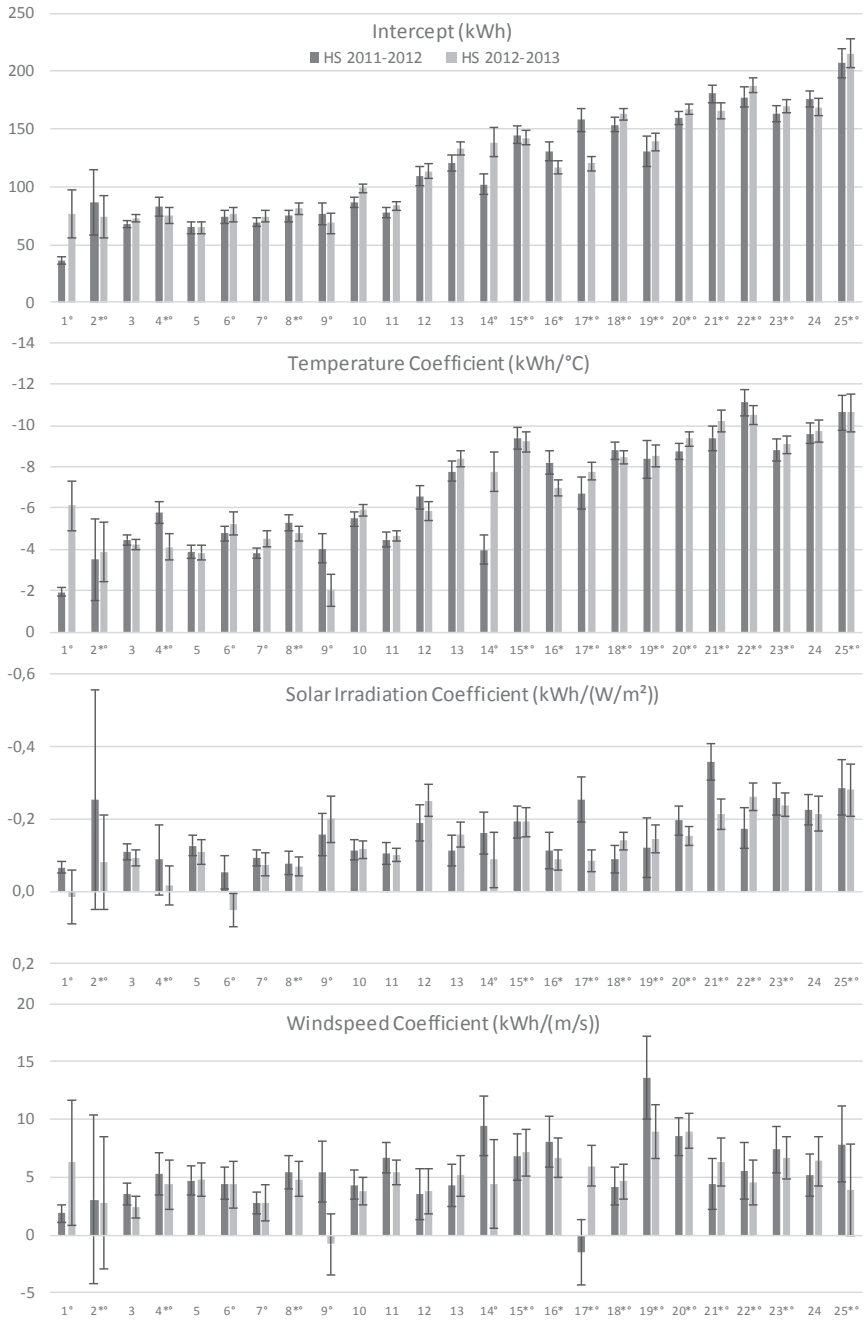


Figure 3-18: Energy Signature Coefficients and 95% CI – Base model – HS1 and HS2

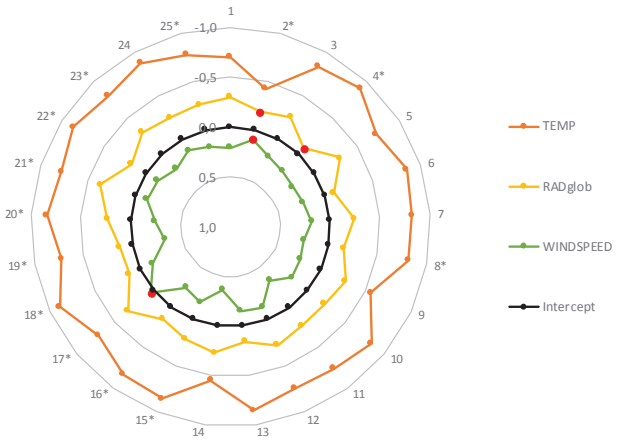


Figure 3-19: Energy Signature: Standardised Coefficients for HS1 – LM50

Preliminary to this analysis, the coefficients and their standard errors were also compared for the selection of models made in section 3.2. For all coefficients it was found that the estimates as well as the standard errors were similar for models LM66, LM71 and LM78. LM50 estimates were also quite close and standard errors only a little higher for some cases, so also LM50 would make a good model for coefficient estimation, by including the main weather predictors *Te, Rg* and *Ws*. Moreover, it is found that these three predictors are essential to a decent model. When *Ws* and or *Rg* are excluded (as in LM8 and LM6), the coefficient estimates diverge from those estimated with the more complex models, and the standard errors increase significantly, making the estimates more uncertain. When *Ws* is excluded from the model (LM8), the estimates for *Rg* are in a different range (see Figure 3-20). Therefore it is remarked that one should be very careful when comparing Energy Signatures that are estimated from different models, and thus based on different energy-balance equations.

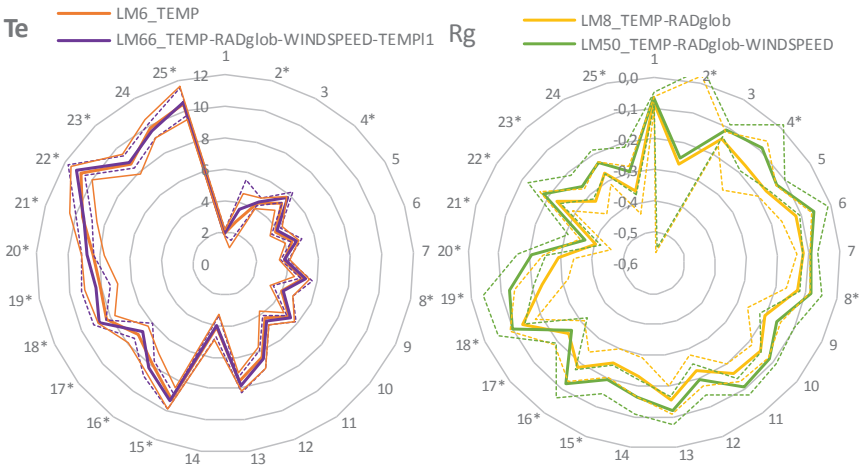


Figure 3-20: Energy Signature: Model coefficients for Te and Rg – different models –HS1

3.6 Training period

When a full heating season of data is used to train the linear regression model, the training data set encloses a rather large range of values for the predictor variables, that is (more or less) representative for the occurring variations in weather conditions during the heating season. Perhaps longer training data periods, containing several heating seasons, could cover these weather variations even better, but unfortunately then the probability rises that also the observed system (building – services – user-interaction) has changed, and model assumptions such as linearity or homoscedasticity, are violated. In the meantime, there are good reasons to look towards shorter – though reliable – training data periods. Not only this would reduce the time needed to collect observations and provide feedback faster, it would also enable to identify changes in energy use that appear over shorter periods, and to identify more precisely when these changes occurred. Therefore in this section, the relation between the properties of the predictors in the training data and the model results, such as the prediction accuracy or the coefficient estimates, is investigated. The Energy Signature base model is now fitted on different data periods, including the entire first and second heating season, as well as all possible combinations of 1, 2 or 3 successive months during these two heating seasons. The fitted models for each of the training data periods are used to predict the energy use of the entire heating season they are part of, and to determine the Energy Signature coefficients. The analysis is limited to those cases for which in the full heating season model the assumptions are found valid, that is 12 cases for heating season 1 and 8 cases for heating season 2.

3.6.1 Prediction accuracy & model coefficients

One research question is how well the energy use of an entire heating season can be predicted by use of a model that is trained with 1, 2 or 3 months data from that heating season, or how much of the prediction accuracy is lost when the training period is reduced. Another question is whether the Energy Signature coefficients can be estimated and with which accuracy? Figure 3-21 presents the Mean Absolute Error of the full heating season predictions, separately for the training data from the first and from the second heating season. It is found that for three training periods, the prediction accuracy is very close to the full heating season prediction accuracy: the MAE increases only 0 to 7% for all cases. The training periods are the three month training periods Jan-Feb-Mar and Feb-Mar-Apr in the first heating season, and Feb-Mar-Apr in the second heating season. They are listed in Table 3-4. For these training periods, also the goodness-of-fit of the model remains the same or improves a little (up to .03 in R^2adj) and the estimates for all model coefficients are not significantly different. The standard errors of the coefficients are also similar or increase up to 80%, thus reducing the accuracy of the estimates for some coefficients. This first group of thee month training periods thus presents very convincing results, enabling a bisection of the duration of the training period, with minor loss of accuracy.

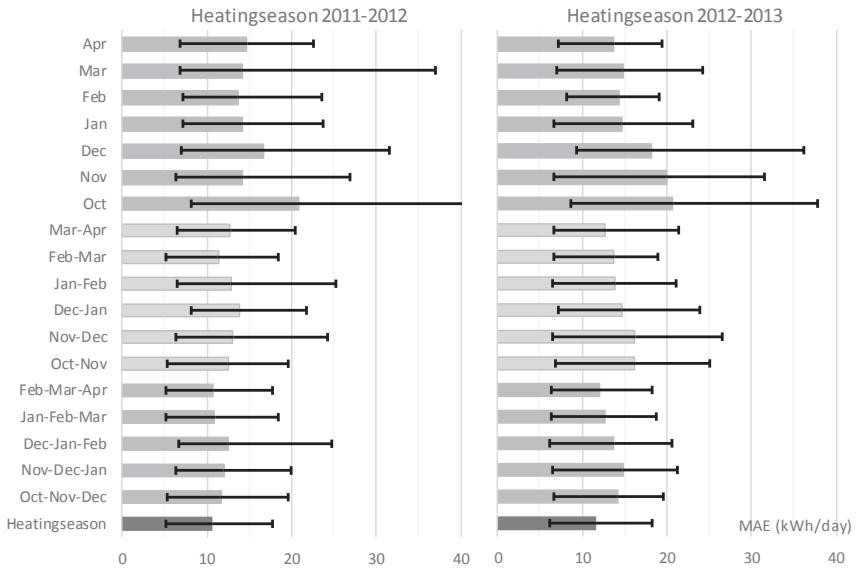


Figure 3-21: Training Period: Prediction accuracy (MAE) for heating season predictions

For a second group of training periods, the full heating season prediction accuracy is lower but still quite acceptable and consistent over all the cases. The MAE rises less than 15% on average, and maximum 25%. The training periods have a duration of two or three months, they are listed in Table 3-4 (row 2). The third group of training periods includes the remaining three month periods except for Nov-Dec-Jan of the 2nd heating season, as well as selected periods of mainly two months (see Table 3-4, row 3). The members of this group have an increase in relMAE (relative to the full heating season training set) below 50%, and an average (over all cases) below 25%. With regard to the coefficients, for selected periods Feb-Mar(1), Dec-Jan-Feb(1) and Jan-Feb(1), all coefficient estimates are not significantly different at the 95% confidence level. For the other periods, there are 1 up to 4 cases where one or two coefficients are significantly different (often the temperature coefficient c_{Te} or the solar radiation coefficient c_{Rg}), all other coefficients remain not significantly different. However, the standard errors of the coefficients are further increasing.

	relMAE (max) increase factor	**relMAE (mean) increase factor**	**Training Periods**
1	< 1.10	< 1.05	Jan-Feb-Mar(1), Feb-Mar-Apr(1), Feb-Mar-Apr(2)
2	< 1.25	< 1.15	Feb-Mar(1), Mar-Apr(2), Oct-Nov-Dec(1), Jan-Feb-Mar(2), Nov-Dec-Jan(1)
3	< 1.50	< 1.25	Mar-Apr(1), Jan-Feb(2), Apr(2), Dec-Jan-Feb(1), Dec-Jan-Feb(2), Nov-Dec(1), Jan-Feb(1), Oct-Nov-Dec(2)
4	> 1.50	> 1.25	Nov-Dec-Jan(2), Oct-Nov(1,2), Nov-Dec(2), Dec-Jan(2), Feb-Mar(2),Oct(1,2), Nov(1,2), Dec(1,2), Jan(1,2), Feb(1,2), Mar(1,2), Apr(1)

Table 3-4: Training Period: Grouped by level of decrease in prediction accuracy

About half of the training periods is included in groups 1, 2 or 3. The remaining periods have a maximum relMAE that is higher than 1.5 (see Table 3-4, row 4),. This group includes most of the 1 month training periods and the remaining two month periods, as well as the period Nov-Dec-Jan in the second heating season. Apart from an increase in prediction accuracy that is higher than 25% on average, here and there cases turn up where the estimates are different from those estimated with the full heating season data, and the standard errors of the coefficients increase more and more. Remark however that even for most of these "worse" training periods, often a few cases exist that perform well, and for example have a relMAE lower than 10% and good estimates for the coefficients. However, this is not applicable to all cases any more.

The standard errors are a measure for the accuracy with which the coefficients are estimated: the higher the standard error, the wider are the confidence intervals of the estimate. In general, the standard errors tend to decrease with increasing model quality, and a good goodness-of-fit tends to go with decent prediction accuracy. However, the last statement is not always true. For example in the case of Nov-Dec-Jan or the month of January (2nd heating season), the R^2adj and standard errors are very good, but the prediction accuracy is relatively high. The model fits the given data-set very well, but is not capable of accurately predicting the energy use over the full heating season. Most probably the training data set is not representative for the full heating season, for example because the range in weather inputs is too specific. Regarding the other training periods, Table 3-5 presents the standard errors for a selection of models from the 1st heating season, for which for all cases and for all model coefficients, it is found that the estimated confidence intervals overlap -and thus the estimates are found not significantly different. In the table the standard error of each coefficient is scaled to its standard error when estimated with the full heating season data. The numbers are the average, minimum and maximum values appearing throughout the 12 cases. For example for the exterior temperature coefficient, on average the standard errors (and thus the confidence intervals) are more than doubled when going to only month training data period (February). In practice, this will probably leave many more cases undecided when comparing the coefficients for different periods of time. Moreover, for some coefficients zero is included in the confidence intervals, and so the coefficients are found not significantly different from zero. This is found especially for the solar irradiation and wind speed coefficients, and appears in many cases when reducing the data period to 1 or 2 months. Amongst the training periods observed, the lowest standard errors are obtained for the training data period Feb-Mar-Apr. For this period, the p-values for all coefficients are also significant. For example for c_{Te}, the standard errors increase between 0 and 30%.

Table 3-5: Coefficient standard errors: training data period / full heating season – mean [min, max]

	B	**Te**	**Rg**	**Ws**
Heating Season	1	1	1	1
Dec-Jan-Feb	1.65 [1.3; 1.9]	1.84 [1.6; 2.1]	3.64 [2.2; 4.3]	1.51 [1.2; 1.8]
Jan-Feb-Mar	1.45 [1.3; 1.7]	1.27 [1.2; 1.5]	1.57 [1.1; 2.0]	1.51 [1.4; 1.8]
Feb-Mar-Apr	1.32 [1.1; 1.4]	1.23 [1.0; 1.3]	1.23 [1.0; 1.3]	1.38 [1.2; 1.5]
Jan-Feb	1.81 [1.4; 2.1]	1.91 [1.6; 2.2]	3.98 [2.4; 4.7]	1.72 [1.4; 2.0]
Feb-Mar	2.09 [1.5; 2.6]	1.38 [1.0; 1.7]	1.90 [1.3; 2.5]	2.37 [1.7; 3.0]
Feb	2.74 [1.7; 3.6]	2.23 [1.5; 2.9]	5.69 [2.7; 7.5]	3.13 [2.1; 4.1]

In conclusion, when the training period for the Energy Signature model is reduced, the prediction accuracy and the accuracy of the coefficient estimates vary for the different 1, 2 or 3 months periods observed. For some well-selected periods of 3 months (group 1) the difference in prediction accuracy when compared with the heating season training data model is negligible, and it is possible to make a similar estimation of the coefficients, with a small loss in accuracy. On the other hand, for other training data periods the prediction accuracy measure MAE more than doubles, the accuracy of the estimates reduces. When comparing these cases for different time periods, for more cases the conclusion will be undecided or insignificant, or might even be wrong.

3.6.2 Input data characteristics

In the previous section it was found that for some training periods the loss in prediction accuracy (when comparing to the prediction accuracy for the full heating season) was lower than for other periods. The question rises what are the characteristics of the training periods with the lower losses in accuracy? So now the accuracy statistics are observed in relation to the characteristics of the different training data sets, in order to deduct under which conditions of the training data sets, a certain level of prediction accuracy is obtained. Since the amount of training periods is not that large, and data periods partially or entirely overlap, a full statistical study is not performed, but nevertheless some general rules-of-thumb can be identified.

The prediction accuracy statistic used is the relMAE[6]. The observed training period characteristics are the amount of data points (excluding outliers and zero-values, as discussed in section 3.3.2), the summary statistics of the weather inputs (minimum, mean, maximum, first and third quartile, range) and the correlations between the weather inputs. First the relationship between the prediction accuracy and the data characteristics is explored. Figure 3-22 presents scatter plots of the relMAE for some key variables. For example, the prediction accuracy tends to increase (and thus the relMAE decreases) with increasing number of data points, but some low values of relMAE are already found when the amount of data points is higher than 45. The range of temperatures ΔTe (between maximum and minimum daily average temperature) shows a clear relationship with the relMAE. The highest ΔTe go together with the lowest relMAE values. This parameter is correlated to the minimum external temperature Te_{min}, which thus also show relationships with relMAE. From statistical point of view, it is said that a model is valid for the range of observations it was trained for, and it is expected that if the range of observations in the training data set reflects the range of observations in the test data set (the full heating season), the predictions will be more reliable. Since the exterior temperature is the most important model parameter and throughout the full heating seasons a range in exterior temperatures of 29 and 24°C appears (for 1st and 2nd heating season), it is understandable that the models with the highest ΔTe are amongst those with the highest prediction accuracy. Similarly, for solar irradiation the highest prediction accuracy is found for the training sets with the highest maximum solar irradiation values. However for the less accurate models, the relationship is not as clear as for the temperature difference. With regard to the wind speed, the variation in the maximum wind speed is not that large. It is seen again that the lowest relMAE appears for the highest wind speed

[6] The relMAE is here the MAE of the prediction of a model on a full heating season data relative to (divided by) the MAE of the in-sample predictions of the full heating season model.

variables, but that could also be because the periods with the highest wind speed are often the colder months, with higher ΔTe. The cross-correlation between Te and Rg ranges between -0.6 and 0.7. For all three correlations between pairs of weather variables, for the more accurate models the absolute values of the correlation coefficients are below 0.4 / 0.5.

In Table 3-6 the characteristics of the training periods are summarised for the four groups from Table 3-4 that were defined by level of decrease in prediction accuracy. It is found that the most important criteria to distinguish between the different groups are related to the variation in exterior temperature and the maximum solar radiation. The plot in Figure 3-22 (e) illustrates the effect of the combination of these two criteria on the prediction accuracy. In terms of duration, it seems that a training period of 3 months is recommended as a shorter training period (although the weather criteria should be checked) and sometimes can be shortened to about two months if the weather conditions are fine. Periods of one month seem to be a little short, both in order to receive enough variation in weather conditions, and to have enough data points for modelling.

It is concluded that based on this examination, some general rules of thumb can be pointed at when deciding on the usability of a given training data period for constructing Energy Signatures. The general idea that the larger range in input variables, the larger the amount of data-points and the lower correlation amongst them is beneficial, is here exemplified by actual numbers for these ranges (see Table 3-6). Note however that this exercise is limited to a limited amount of cases and weather conditions (appearing in 2 years), and could be further investigated in future research using a full statistical analysis, when larger data-samples are available.

Table 3-6 Training Period: Characteristics per group (by level of decrease in prediction accuracy)

Predictions	1	2	3	4
relMAE (max) increase factor	< 1.10	< 1.25	< 1.50	> 1.50
relMAE (mean) increase factor	< 1.05	< 1.15	< 1.25	> 1.25
Data period	3 months	2-3 months	(1)-2-3 months	1-2-3 months
Data points minimum	85	45	30	30
Weather characteristic				
Te Variation	> 23°C	> 15°C	> 10°C	> 8°C
Te Minimum	< -5°C	< 2°C	< 4°C	< 6°C
Rg Maximum	> 200W/m²	> 75W/m²	> 75W/m²	> 50 W/m²
Ws Maximum	> 6m/s	> 4,5m/s	> 4m/s	> 3m/s
Cor Te, Rg (abs)	< 0.4	< 0.4	< 0.6	< 0.7
Cor Te, Ws (abs)	< 0.2	< 0.2	< 0.55	< 0.7
Cor Rg, Ws (abs)	< 0.4	< 0.5	< 0.5	< 0.6

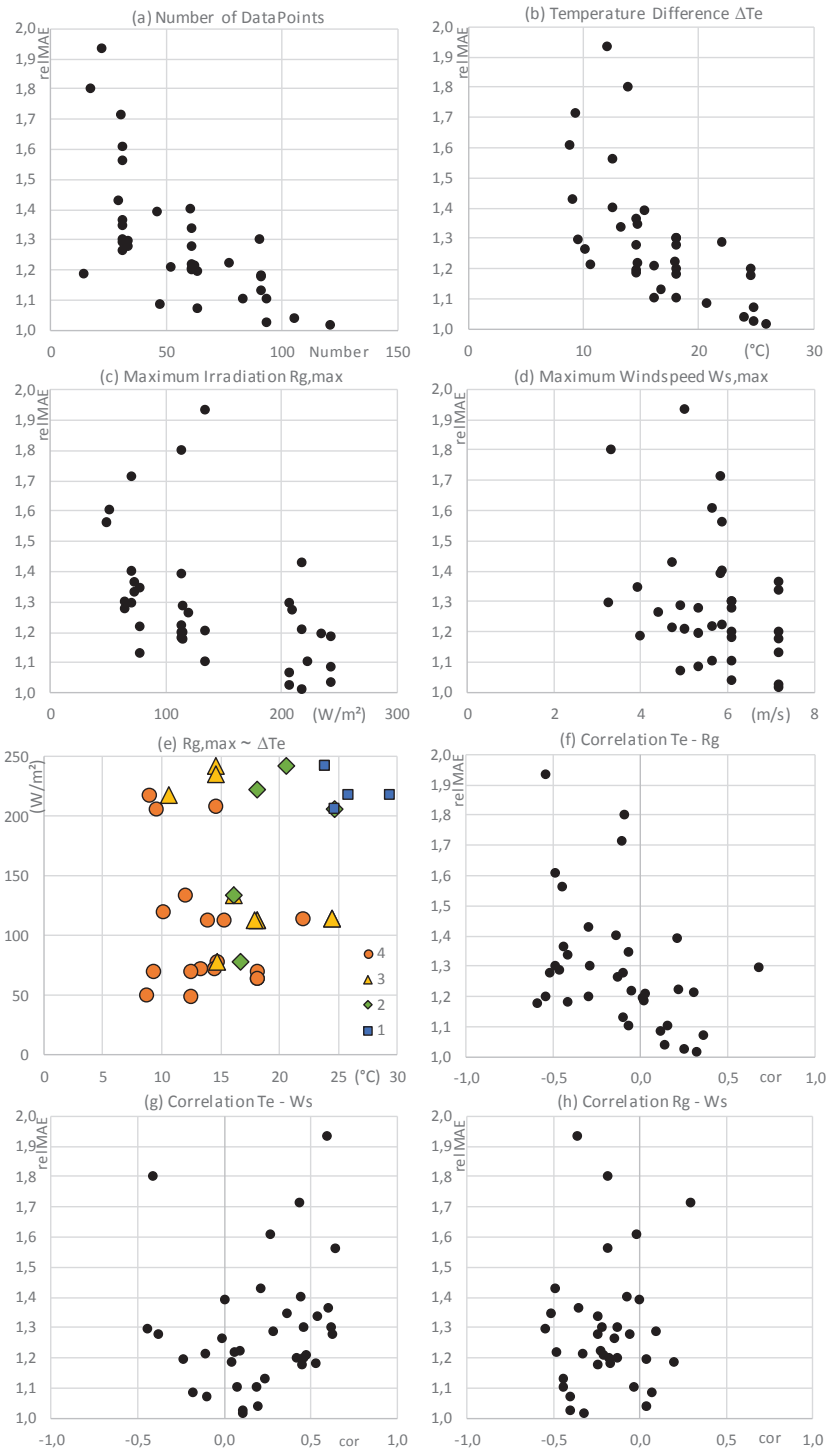


Figure 3-22: Training Period: relMAE & weather variables

3.7 Time step

3.7.1 Autocorrelation in time series data

One of the basic linear regression assumptions is that the error terms are independent or uncorrelated. In particular in the case of time series data, the correlation of error terms *over time* is a point of particular interest. This is the correlation between consecutive errors (or errors following each other at a certain time distance) and is also called autocorrelation or serial correlation. Observe an excerpt of the time series of 2-hourly energy use observations of case 25 in Figure 3-23. The time series is subjected to a daily cycle that is not only a result of the daily cycle in weather conditions (e.g. the daily cycle of the sun), but also to a cycle in the observed system as a result of the interaction of the user, building and services (e.g. the night set-back of the heating system, the start-up of the heating system in the morning). When the linear regression base model is applied to the time series, the goodness-of-fit of the obtained model is very low ($R^2adj = 0.06$) and the in-sample predictions cannot follow the variance in the data. As a result, also the errors or residuals show this daily pattern. The correlation of the residuals in time can be checked by use of the plot of the autocorrelation function (ACF), that represents the correlation of the residuals with themselves at different time lags[7]. The ACF of the example time series points towards severe autocorrelation, especially at the 12th time lag (24 hours). It is said that the errors are autocorrelated. Remark that also non-linear effects are present, as well as heteroscedasticity, violating other fundamental assumptions of the linear regression model. According to Kutner et al. a major cause for autocorrelated error terms is that one or several key variables are excluded from the model, and when the time-ordered effects of such variables are correlated, this will be visible in the errors, since the errors include the effects of the missing variables [71]. In the example case, variables related to the control of the space heating system are not included in the model, however they seem to be very influential on the observed values.

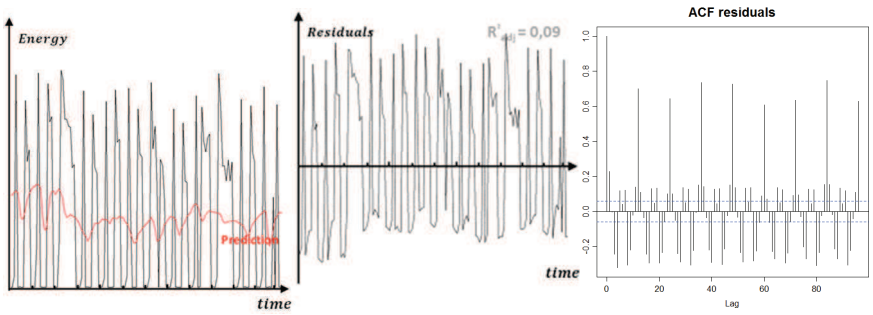


Figure 3-23: Autocorrelation in 2-hourly Time Series data: Example

Important consequences of (positive) autocorrelation in linear regression models estimated by the OLS method, are that the variance of the error terms could be seriously underestimated as well as the calculated standard errors of the coefficients [71]. A remedial measure for autocorrelation is to include the missing key variables in the linear regression model, assuming

[7] Reading of the ACF: at time lag zero, the autocorrelation function is always 1, this is the correlation of the residual with itself. For the other time lags, autocorrelation is low or acceptable when the values are (mostly) in between the dashed lines.

that this variable is available or can be captured. If that is not the case, the linear regression equation can be transformed to include autoregressive terms. The use of autoregressive models will be discussed in chapter 4 of this study. However, in the context of defining Energy Signatures, probably the most applied way to resolve autocorrelation and non-linearities in the linear regression model, is by aggregating the data to time steps of one day or multiples of one day, so that each data point includes one or multiple full 24 hour cycles. The advantage of the aggregation is that the heaviest serial correlations and non-linearities are suppressed, a disadvantage is that the sub-daily information in the data is wiped out too and the amount of data points is reduced. The decision of which level of aggregation to use is thus a balance between these two considerations. In typical Energy Signature applications, the interest is merely in the (lower frequency) relationship of the building with the weather variables, rather than in sub-daily aspects such as those related to the system operation or user behaviour. Therefore the daily time steps are often applied, but also weekly or monthly aggregation is used (mostly if no detailed data is available, since longer data periods are required).

3.7.2 Autocorrelation in the daily models

Observing the regression diagnostics for the linear regression base model with daily data, in section 3.3.2 violation of the independency of the errors assumption was mentioned for a number of cases. From analysis of the ACF plots and the cumulative periodogram (CPGRAM, see Figure 3-9), it was concluded that for 13/25 cases no or moderate serial dependence of the errors appears and the independency model assumption is accepted. For the other 12/25 cases, clear autocorrelation was found. They are marked in red in Table 3-7 and Table 3-8 (last two rows). Indeed, insufficient fulfilment of this model assumption often goes together with violations to one or more of the other linear regression assumptions.

Amongst the cases where obvious autocorrelation is observed, two types of autocorrelation appear. When looking to the first heating season, the first type is found in cases 8, 18, 20, 21, 22 and 23. The most clear example of this type is case 8 and is presented in Figure 3-24. It is recognised by help of the ACF plot, which follows a slow decay. Also the cross-correlations with the weather inputs (e.g. Te or Rg) and with the Energy use appear for a range of time lags. When observing the time series plot of the residuals, it is seen that over time the residuals are not randomly distributed around zero (for some periods most of the residuals are positive, while for other periods they are mainly negative) as if a slow moving trend is still present in the residuals. This is probably due to a phenomenon that influences the energy use, though is not described by the model (e.g. changes in average indoor temperature due to changes in systems, behaviour…).

The second type of autocorrelation appears at specific time lags. For example, Figure 3-25 presents the residual plots for case 16. In the ACF plot peaks are observed primarily at the first and 7^{th} time lag, and at time lags that are multiples of 7. The cross-correlations with weather variables are rather low, but the same 7-day peaks appear in the cross-correlations with the energy use and in the time series plot of the residuals. In this case, the autocorrelation could be related to effects with a duration longer than 24 hours (e.g. effects of thermal inertia related to the intermittent heating of the building) and to weekly cycles in behaviour (e.g. occupancy patterns, space heating profiles that are different in week and weekend days). Similar autocorrelation patterns are recognised clearly in cases 4, 15, 16, 17, 19 and 25 (and more moderately low in cases 5, 9, 10, 11 and 14). Most often the highest autocorrelation appears at the

1st and 7th time lag, but sometimes also at the surrounding time lags (e.g. 3rd, 4th, 8th). The previous observations of the autocorrelation helps to explain why in section 3.2.2 the addition of a second time lag of the ambient temperature (LM72) lead to an improvement in goodness-of-fit for some models, including cases 8,14, 16, 18, 20, 22 and 23, that show autocorrelation in the base model LM66.

From this observation of the residual autocorrelations in the linear regression base model, it is concluded that the frequently applied daily data time step leads to a decent model quality in about half of the cases that were studied, especially if time lags of the weather inputs, such as Te_{t-1} are included. A number of models would also benefit from an addition of Te_{t-2} (although the improvements are small and if auto-correlation existed, it is not entirely removed). Nevertheless, in about 12/25 cases for the 1st heating season and 15/25 for the 2nd heating season, autocorrelations remain, and the model validity is under discussion, which makes the interpretation of the model results (e.g. confidence intervals) unsure. Especially for these cases, other modelling approaches are to be considered. For example for the data sets with weekly autocorrelations, further aggregation to weekly data could be considered. However this will further reduce the amount of data points and prolong the experimental period. Another approach is to include previous time steps of the output as an input into the models, moving towards a model group that is called the autoregressive models with exogenous inputs (ARX-models). This approach is explored in the next chapter, and can also be beneficious for modelling with higher frequency data, such as sub-daily data time steps.

Table 3-7: Linear Regression - Model Validity and Autocorrelation – Training Season 2011-2012

2011-2012	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity	1	0	1	0	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
AC TYPE 1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1	1
AC TYPE 2	1	1	1	0	,5	1	1	1	,5	,5	,5	1	1	,5	0	0	0	1	0	1	1	1	1	1	0

Table 3-8: Linear Regression - Model Validity and Autocorrelation – Training Season 2012-2013

2012-2013	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity	0	0	1	0	1	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0
AC TYPE 1	0	1	1	1	1	0	1	0	0	1	1	1	1	0	1	1	1	0	1	0	0	0	0	1	1
AC TYPE 2	1	1	,5	1	,5	,5	0	1	,5	,5	,5	1	,5	1	0	,5	0	1	0	1	1	1	1	1	0

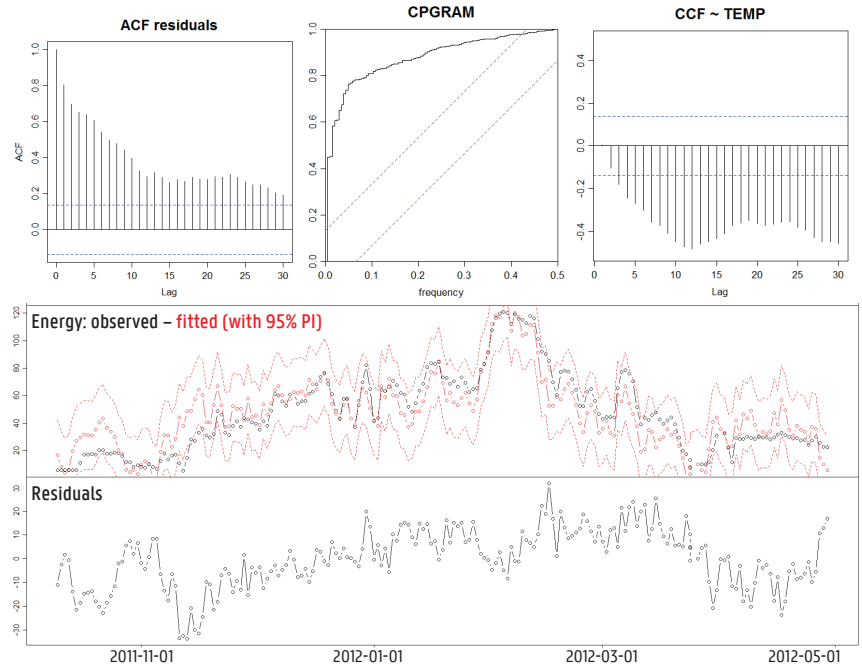


Figure 3-24: 1st type – example case 8 – slow decay

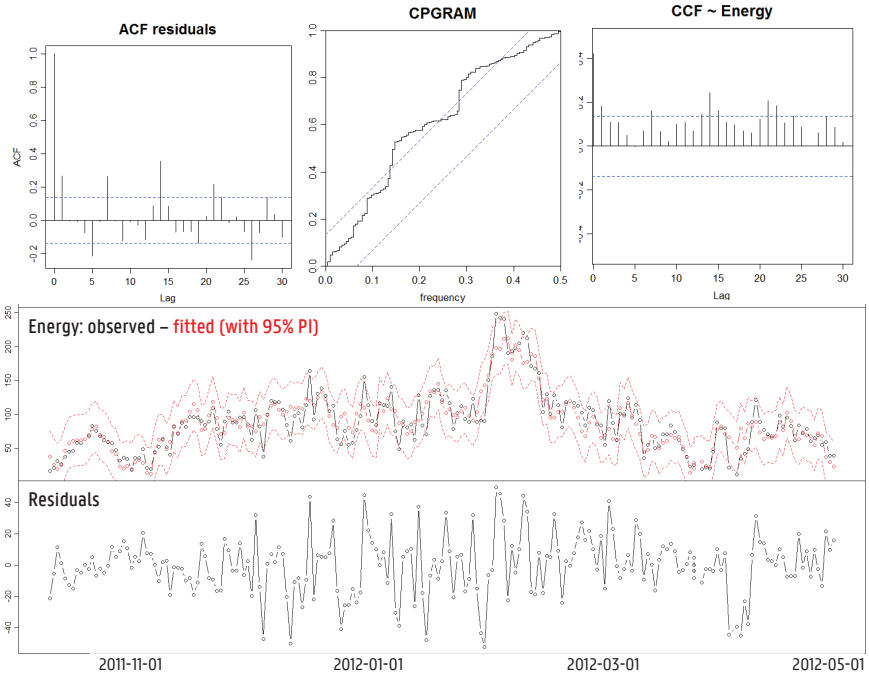


Figure 3-25: 2nd type – example case 16 – specific peaks

3.8 Summary & Conclusions

Using linear regression methods, Energy Signature models are fitted to datasets containing the daily aggregated energy use and weather data for the entire heating season 2011-2012. First the value of different forms of an energy-balance model, including different weather inputs, are investigated. The main weather predictor is the exterior temperature *Te*, followed by solar radiation *Rg* and wind speed *Ws*, that also have a large positive effect on the goodness-of-fit of the model and they are significant at the 5% level in 24/25 of the studied cases. Furthermore, addition of the first time lag of the exterior temperature, Te_{t-1}, is valuable and it is significant in 20/25 cases. These variables are included in the Energy Signature base model, that is described by equation 3-IV.

The base model includes the key weather variables to provide a decent model composition and both in- and out-of-sample prediction accuracy, in perspective to what is reachable with the given weather variables and model type. Yet in case-specific modelling it can be improved by adding, deleting or replacing specific variables, while observing the goodness-of-fit indicator R^2adj, in-sample and preferably also out-of-sample prediction accuracy (to avoid overfitting) and the significance of the coefficient estimates. The weather variables to look at in the first place are: the addition of a second lag of exterior temperature Te_{t-2}, and the replacement of *Ws* by the wind speed for one or a combination of the orientations south, west or east {*Ws,e;Ws,s;Ws,w*}. For other weather variables considered, the effect is for most cases negligible or small: relative humidity *Hr*, precipitation quantity *Pq*, or time lags for wind speed and solar radiation. The equivalent exterior temperature *Te,eq* is an often used predictor in practice, which is a better choice than the simple use of *Te*. However the estimation of separate predictors Te_{t-1} and eventually Te_{t-2} is preferred, because then the weighting of the temperature lags is estimated for the specific case, instead of being imposed.

During the heating season, sometimes one or several days may turn up when no space heating is used despite the wintry weather conditions. This may be due to occupants not being home or a breakdown of the boiler. When applying the Energy Signature model, these days are recognised as outliers in the data and it is found that model quality improves when deleting them from the data. By use of an iteration in the modelling, where the model is re-estimated without cases that belong to the 10% highest Cook's distance and have zero energy use, this pre-processing step can be performed automatically. Note that days with no heating in the warmer periods of the heating season are not recognised as outliers (since it is expected that low or no heating is applied in such conditions), and therefore they are fortunately not excluded.

In order to make sure that the outcomes of the linear regression model, such as the prediction and confidence intervals, can be trusted, it is important that the assumptions of linear regression modelling are met. One of these assumptions is the independency of the errors, more specifically, the independency of the errors in time. In about half of the 25 cases, this assumption is clearly violated when a linear regression model is used. And in many cases where clear autocorrelation appears, it coincides with deviations from normality and/or signs of heteroscedasticity. Therefore an important conclusion from this chapter is that the Energy Signature models, when applied to daily aggregated data, are suitable in about half of the cases only. For the cases where the linear regression model is unsatisfactory, it is proposed to further aggregate the data to weekly time

steps, or to look for other types of models (see next chapter). As for the rest, for data with sub-daily time steps, the autocorrelation issue becomes even more severe and linear regression models are inadequate.

For the cases where the linear regression assumptions are met, the adjusted R^2 is between 0,78 and 0,91. In comparison to the naïve prediction method, the in-sample predictions are between 54 and 70% more accurate, and the out-of-sample predictions between 51 and 64%. For most cases the predictions on the test data set are a little lower than what would be expected from the in-sample-predictions. However for some cases (e.g. case 1, 6 and 9), the predictions on the test data set diverge clearly from the observations, since probably changes in energy use occurred in the second heating season. Finally, the 95% prediction intervals are rather wide for most cases, ranging between 25 and 80 kWh/day. This indicates a rather large uncertainty that coincides with the predictions from linear regression models. With regard to the Energy Signature coefficients, the sign and magnitude of the estimates follow the physical logic.

Amongst all cases considered, there is one case that is from many points of view an exception. For case 2, the uncertainty going together with the predictions as well with the coefficient estimates, is extremely large, solar irradiation and wind speed coefficients are not significantly different from zero, and the adjusted R^2 is low. Its energy use is very irregular in time and is often zero, as if the house is occupied for periods of a few days with irregular periods of un-occupancy in between. For this kind of atypical energy use series, the general modelling approach followed does not work.

A reduction of the training period for the model can be desirable in order to enable faster energy use characterisation and feedback. The training data period is reduced from an entire heating season to periods of 3, 2 and 1 month(s) throughout the first and second heating season. It is found that the prediction accuracy and the accuracy of the coefficient estimates vary for the different training periods observed. For some well-selected periods of 3 months the difference in prediction accuracy when compared with the heating season training data model is negligible, and it is possible to make a similar estimation of the coefficients, with a small loss in accuracy. For other training data periods, the prediction accuracy measure MAE more than doubles, the accuracy of the estimates lowers. When comparing these cases for different time periods, for the majority of cases the conclusion will be undecided or insignificant, or might even be wrong. In order to deduce rules of thumb for the selection of a suitable training period, the model results and the properties of the input variables in the training data are compared. It is found that the most important criteria for selection of a training period are related to the variation in exterior temperature and the maximum solar radiation. For the best training data sets, the variation in exterior temperature (during the heating season) is higher than 23°C, and solar radiation values higher than 200W/m² show up. In terms of duration, it seems that a training period of 3 months is recommended. If the characteristics of the weather conditions are fine, the training period can be shortened to about two months. Periods of one month seem to be a little short, both in order to receive enough variation in weather conditions and to have enough data points for modelling.

4

ARX-Models

Auto-regressive models with exogenous inputs (ARX) are made of one or several auto-regressive terms, that are historical observations of the output (energy use), and exogenous inputs, such as climate variables. They can also be estimated using the ordinary least squares method. These models can be understood as a dynamic variation on a classical multiple linear regression model, in which historical values of the output variable are used as input variables. The lagged energy use predictors can take up regular variations in the energy use that are not per se due to climate, but for example to patterns related to occupational schedules. Therefore, it is expected that these models provide a solution to the problem of the autocorrelated errors, that was detected in the LM-models for about half of the cases studied. In this chapter, ARX-models are constructed for the same data-sets as used in chapter 3, and their performance is evaluated and compared to the classical linear regression models. The key research questions in this chapter are: **Does the use of auto-regressive models with exogenous inputs (ARX-models) improve the Energy Signatures when data at daily time resolution is used? (Research question 2)**

- What are the key weather variables in the ARX-model? (section 4.1)
- Are the models valid from statistical point of view? (section 4.2)
- What is the accuracy of the estimated ES coefficients? (section 4.3)
- Are significant differences in Energy Signature observed when comparing the 1st and 2nd heating season? (section 4.4)
- What is the accuracy of the energy use predictions? (section 4.5)
- What if data at sub-daily time resolution is used? (section 4.6)

4.1 Variables

For the ARX model construction, the training data set consists of the daily aggregated energy use data from the heating season 2011-2012 (that is the same data set as used in section 3.2). Regarding the selection of exogenous inputs, six combinations of the most important weather variables are selected based on the findings in chapter 3.2. They are listed in Table 4-1 and they are referred to with the same number as the corresponding linear regression model. Regarding the auto-regressive part of the model, six combinations of auto-regressive terms are selected, based on experiences from previous studies by the author, in which it is found that – in unoccupied dwellings during a heating experiment - especially the first 1 up to 3 time lags are important [51]. Secondly, in an exploratory exercise on the current data-set, it is found that additionally also the 7^{th} and 14^{th} time lag can be important. In Table 4-1, all of the auto-regressive parts include the first time lag of the energy use Energy.l1, eventually completed with a selection of energy use time lags up to the 14^{th} lag. These model parts are referred to with capital letters A to F (Table 4-1). All AR- and X-parts are combined to create 36 candidate ARX-models. For example the simple model ARX6A has the variables: external temperature $Te_{,t}$ (like linear regression model LM6) and the first time lag of the energy use Q_{t-1}.

For example, the ARX66C-model is described by the following equation 4-I:

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Rg_t + c_4 \times Ws_t + c_5 \times Te_{t-1} + c_6 \times Q_{t-1} + c_7 \times Q_{t-2} + c_8 \times Q_{t-3} + \varepsilon_t \qquad [4\text{-I}]$$

Table 4-1: ARX – Model List

X	WEATHER VARIABLES	AR	ENERGY LAGS
6	TEMP	A	ENl1
8	TEMP-RADglob	B	ENl1 + ENl2
50	TEMP-RADglob-WINDSPEED	C	ENl1 + ENl2 + ENl3
57	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth	D	ENl1 + ENl2 + ENl3 + ENl4 + ENl5 + ENl6 + ENl7
66	TEMP-RADglob-WINDSPEED-TEMPl1	E	ENl1 + ENl7
67	TEMP-RADglob-WINDSPEEDeast-WINDSPEEDwest-WINDSPEEDsouth-TEMPl1	F	ENl1 + ENl7 + ENl14

4.1.1 Auto-regressive terms

First the value of the different auto-regressive parts is analysed. For all considered combinations with exogenous inputs, it is found that the addition of the first time lag of the energy use (AR-part A) has the most significant (and positive) effect on the prediction accuracy (RMSE) and the coefficient of determination (R^2adj), which increases with 3.3 pp on average and up to 16pp. Extra auto-regressive terms (AR-parts B to F) can lead to further improvements for specific cases, but have a negligible effect in the other cases. The R^2adj increases with maximum 4pp, on average less than 1pp, in comparison to the A-model. In Figure 4-1 this is illustrated for the models using X-part 66 and the AR-terms A, D and E. The RMSE is here relative to the corresponding linear regression model LM66. In section 3.7.2, for the linear regression models showing autocorrelation in the residuals, two types of autocorrelation were recognised: for the first type the autocorrelation function shows a slow decay, while for the second type peaks in the autocorrelation function appear at specific time lags. In Figure 4-1 it is seen that for the cases in which the first type of autocorrelation appeared (■) the addition of a first time lag (ARX66A) leads to large improvements in goodness-of-fit and prediction accuracy: the RMSE decreases on average 30%, and even up to 62% for case 8. Improvements after addition of extra time lags are smaller (max. 7%). For the cases in which the second type of autocorrelation occurs (♦ for very clear AC, ◊ for moderate AC), the RMSE reduces on average 6%, and up to 14% for case 15. For this type, the addition of further time lags is clearly useful for many cases, leading to improvements up to for example 11% for ARX66D in case 5. Finally, for the cases in which no autocorrelation was observed in the LM66-model, the ARX66 models lead to the lower improvements, on average 3% for Q_{t-1}, and another 3% for combinations of the further time lags.

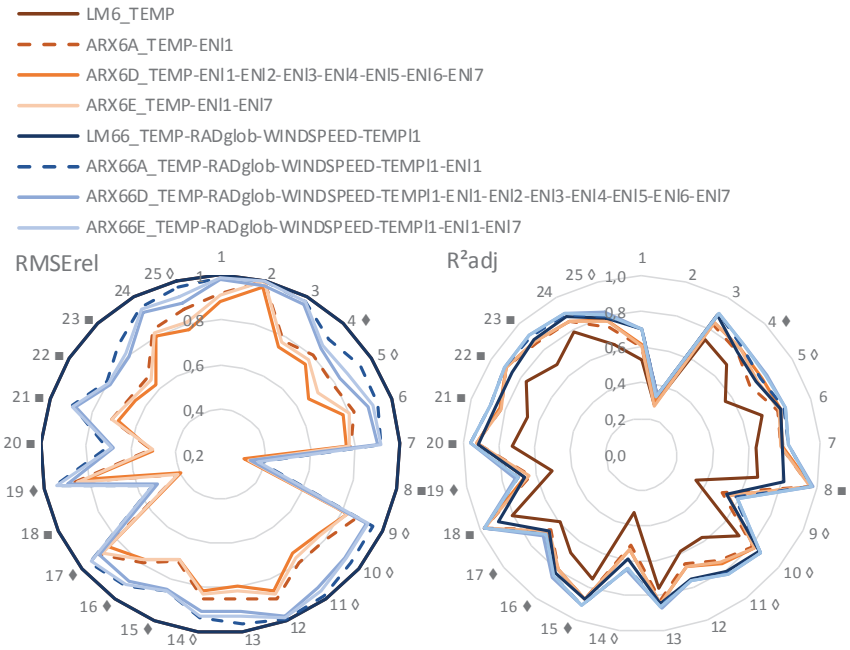


Figure 4-1: ARX – AR-terms for ARX6 and ARX66 – RMSErel (to LM-model) and R^2adj

Considering also the significance of the energy lag coefficients, it is found that Q_{t-1} is significantly different from zero in at least 23/25 cases for the ARX-A models, and mostly stays significant when more energy use lags are added. Amongst the additional energy use lags Q_{t-7} is significantly different from zero in 13/25 cases (ARX66E) up to 19/25 cases (ARX6E) and is the strongest parameter to add to the X-part of the model. When also Q_{t-14} is added (ARX-F), improvements are small and most coefficients are not significant. The models with all 7 energy use lags (ARX-D) lead to the highest prediction accuracy, but the R^2adj more or less remains, and a number of the energy use lags are not significant (especially the 4th, 5th, 6th lag), indicating redundancy. For some cases, e.g. 17 and 23, the combination of the 2nd and 3rd energy lag (ARX-C) leads to good results (better than ARX-E), but the ARX-B models have lower improvements. In conclusion, the most efficient X-part to add is Q_{t-1}, and the most interesting additional parameter is Q_{t-7}, Q_{t-2} or $Q_{t-2} + Q_{t-3}$. The highest prediction accuracy is obtained with a full model, including 7 energy lags (ARX-D), though not all coefficients are significant and the goodness-of-fit remains. Therefore, if a case-specific suitable model is aimed at, it is proposed to start from the model ARX-D, or a model with all auto-regressive terms up to 8, and allow step-by-step deletion of the insignificant AR parameters, by use of a backward stepwise elimination procedure.

Remark that the relative improvements by adding AR-terms to a model with exogenous inputs are higher when the respective base models have lower accuracy or goodness of fit. For example in Figure 4-1 (left) it is seen that the ARX6-models have a clearly lower RMSE relative to the LM6 model than the ARX66-models to the LM66 base model, and (right) the R^2adj also increases more. Figure 4-2 (left) confirms that while the nRMSE is on average 25% lower for the more accurate linear regression model with more weather variables (LM66), the differences in nRMSE between the smaller ARX6 and larger ARX66 model are smaller, 15% on average. It seems that part of the knowledge contained in the weather variables in ARX66, is contributed to the AR-terms in the models that do not possess these weather variables.

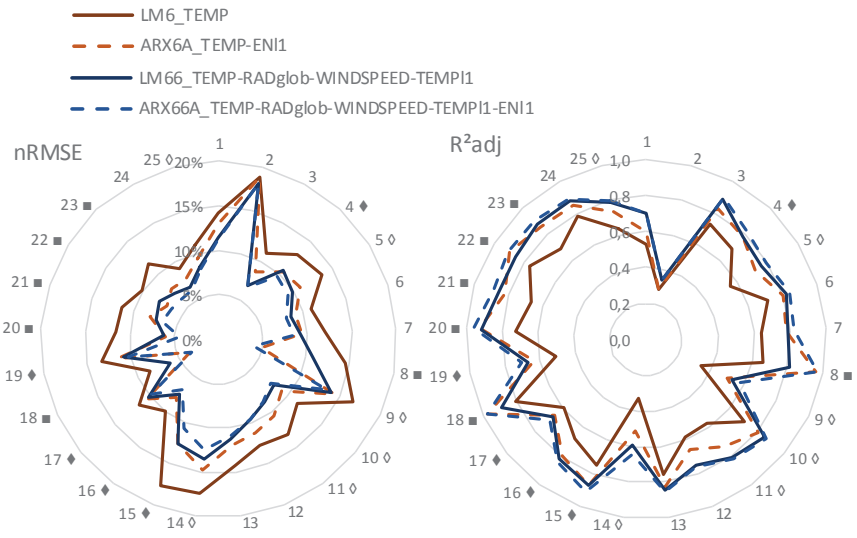


Figure 4-2: ARX – LM and ARX 66A and 6A – nRMSE and R^2adj

4.1.2 Exogenous inputs

Figure 4-3 shows the RMSErel and R²adj for ARX-models consisting of one auto-regressive term Q_{t-1} with the different combinations of exogenous inputs from Table 4-1. It is found that the effect of adding weather variables in the X-part of the model is indeed smaller than in the LR models in Figure 3-4, but still the inclusion of the most important weather variables leads to improvements in the model. For example the average R²adj rises from 0.773 when *Te* is included, to 0.805 when also *Rg* is added and to 0.823 when also *Ws* is added. Replacing Ws by {Ws,e ; Ws,s ; Ws,w} leads to the same results in terms of in-sample RMSE and R²adj. When observing the p-values of the t-test, the conclusions for the significance of the coefficients do not alter in comparison to the linear regression models: the coefficients are significantly different from zero in the majority of the cases (>20/25), except for Ws,e (10/25) and Ws,s (17/25). On the contrary, the addition of the first time lag of the exterior temperature Tet-1 in the ARX-models goes with negligible improvements in goodness-of-fit (R²adj increase of 0.2 pp in ARX vs 2pp in LR models) and the coefficient is proven significantly different from zero in 11/25 cases (instead of 21/25). It seems that part of the knowledge contained in the Tet-1 coefficient is moved to the Qt-1 coefficient in the ARX-models. Therefore, based on the in-sample statistics, the ARX-model ARX50A is found the most efficient base ARX-model, that yields clear improvements to the LM-models for many cases. When this model is applied, the R²adj lies between 0.57 and 0.97, with an average of 0.82. In comparison to the base linear regression model LM66, this is an increase of about 3 pp. The average nRMSE decreases slightly from 10.4 to 8.9% (max 13%, min 3%).

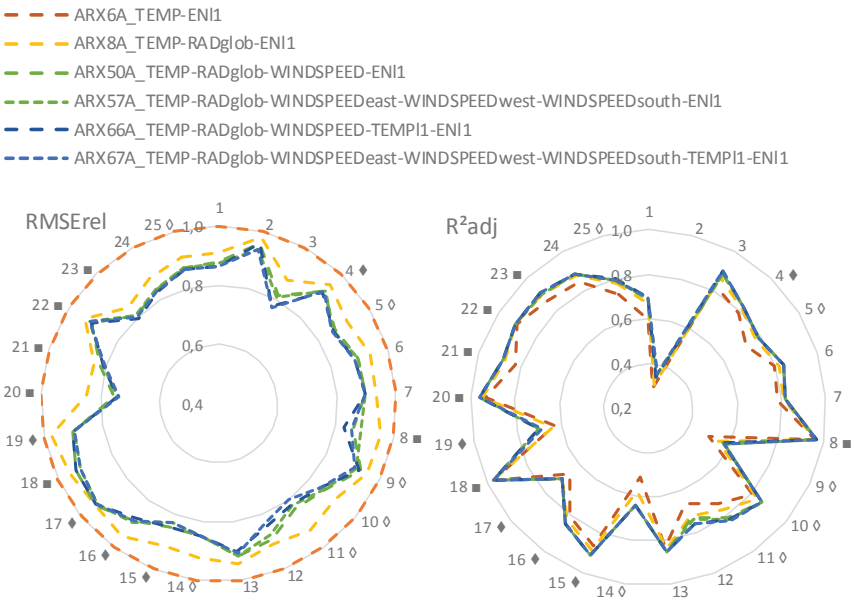


Figure 4-3: ARX – X-terms for ARX-models A – RMSErel and R²adj

4.2 Model validity

For the linear regression models in chapter 3, the statistical model assumption that was most under discussion was the independency of the errors or the occurrence of autocorrelated errors. Autocorrelation was then recognised clearly in 12/25 cases, and more moderately in another 5/25 cases (see section 3.7.2). For a number of them, the violation of the independency assumption went together with the violation of other model assumptions, such as the normal distribution of the errors and/or homoscedasticity. The application of autoregressive terms in the ARX-models is a benefit precisely for the issue of the autocorrelated errors. It is found that when the largest ARX-models with seven subsequent energy lags (ARX50D or ARX66D) are applied, for 23/25 cases no autocorrelation remains. For the other two cases (11,19) moderate autocorrelation appears in the 8^{th} time lag. For these cases the ARX-model could probably be optimised by adding this time lag. For all cases the model with autoregressive part D can be reduced to the smaller models with autoregressive parts A, B, C or E – that contain combinations of the most important energy lags for the majority of cases - while not loosening the independency of errors assumption. For example Figure 4-4 presents the acceptable autocorrelation diagnostics for the AR-C model applied to case 8, that are a very clear improvement to the LM-model (Figure 3-24). In Table 4-2 for each case the model(s) are marked for which the assumption of independent errors is accepted, ànd all auto-regressive terms are significantly different from zero. Remark that for all but one case, the most extended AR-D term is not marked, since one or several of the autoregressive terms in the AR-D model are not significantly different from zero at the 95% confidence interval (and leads to higher standard errors on the coefficient estimates, when they remain in the model). On the other hand, the first auto-regressive term (AR-A) is significant in 22/25 cases, but it is not always sufficient to resolve autocorrelation. The selection of the most suitable model is thus case-specific. Furthermore, remark that the X-50 exogenous inputs (Te, Rg, Ws) are sufficient to find a suitable model for each case where an ARX-model is used, but for a number of cases the ARX66 models are equivalent or better. Finally, in Table 4-3 model validity is summarised. It seems that when the autocorrelation issue that appeared in the linear regression models, is resolved, and the zero-outliers are deleted from the data (see section 3.3.2), for most of the cases all other model assumptions are accepted too. Exceptions are cases 4, 16, 18, 19, 25, where heteroscedasticity is still observed, and about 5 cases where small autocorrelation is still observed for some time lags (which is actually resolved by adding further lags to the selected models). It is concluded that in comparison to the linear regression models the statistical model quality has improved a lot and it can now be considered sufficient for 20/25 cases. The selection of the essential auto-regressive terms is case-dependent. In an automated model selection procedure, it is proposed to perform a backward selection procedure on a full model with up to 8 auto-regressive terms and X-66 exogenous inputs, until a model is reached in which only significant terms are included.

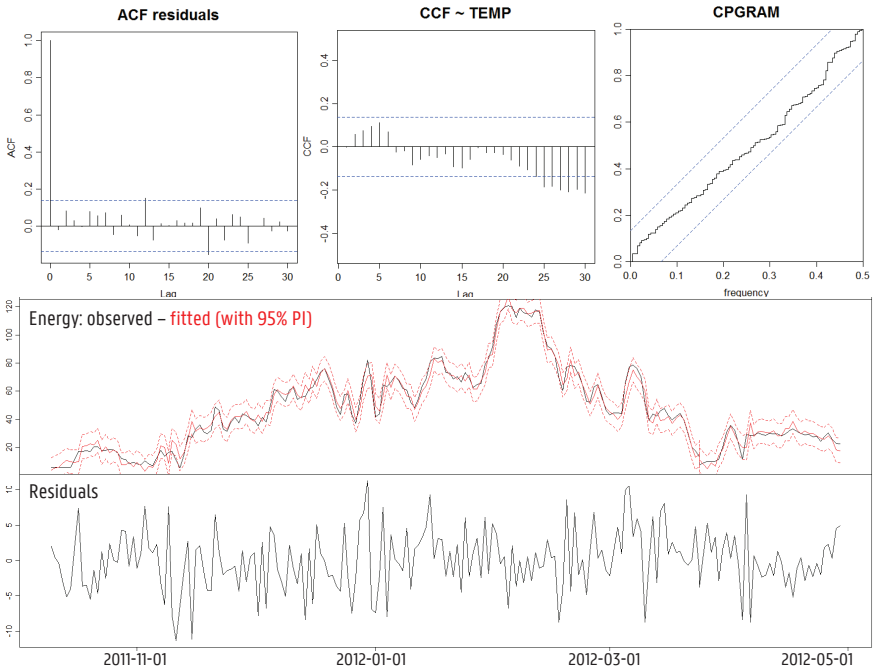


Figure 4-4: ARX66C - example case 8: statistic plots and ls 1-step ahead predictions

Heating Season 2012-2013

For each case, the LM66 and the ARX50-models are fitted and by observing the model validity, it is checked whether the case-specific model structure that applied to the first heating season, is also valid for the data of the second heating season. For 8/25 cases the model assumptions are better fulfilled when applying another model structure, often by adding another auto-regressive term. The original model structures are marked in red in Table 4-4 and the selected models in blue. After the modelling process, there are again about 5 cases in which a small amount of autocorrelation remains at a specific time lag that is not included in the proposed models. Apart from the three cases where the model was found invalid also in the first heating season, 6 more cases are now found not completely valid, often due to slight or clear heteroscedasticity, or violation of the normal distribution of the errors (see Table 4-5).

Table 4-2: ARXselect - HS 2011-2012 - suitable models (X) and selection per case (X)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity				*												*		*	*						*
LM66	X	X	X			X	X					X	X												
ARX50A	X			X			X					X	X	X	X	X			X	X	X	X		X	
ARX50B				X							X											X	X	X		
ARX50C								X									X	X							
ARX50D								X																	
ARX50E					X	X			X	X			X									X		X	X
ARX50F																									X
ARX66A						X							X											X	
ARX66B								X												X	X		X		
ARX66C								X																	
ARX66D																									
ARX66E					X	X				X			X							X				X	
ARX66F																									

Table 4-3: LM and ARX Models - Model Validity and Autocorrelation – Training Season 2011-2012

LM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity	1	0	1	0	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
AC TYPE 1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1	1
AC TYPE 2	1	1	1	0	,5	1	1	1	,5	,5	,5	1	1	,5	0	0	0	1	0	1	1	1	1	1	0
ARX																									
Validity	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	0
AC	1	1	1	1	1	1	1	1	1	0,5	1	1	0,5	1	0,5	1	1	0,5	1	1	1	1	1	1	0,5
AC lag										(8)			(7)		(5)			(4)							(5)

Table 4-4: ARXselect – HS 2012-2013 – selection per case (X) and models no longer suitable (x)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity	°	°		°		°			°						°		°	°	°						°
LM66	x	X	X			x	(x)					X													
ARX50A	X			X		X	X						X	X	X	(X)			(x)	(x)		X		X	
ARX50B											(x)										X		(x)		
ARX50C								X			X						(X)	X	X				X		X
ARX50D																									
ARX50E					X				X	X							(x)			X					x
ARX50F																									

Table 4-5: LM and ARX Models - Model Validity and Autocorrelation – Training Season 2012-2013

LM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity	0	0	1	0	1	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0
AC TYPE 1	0	1	1	1	1	0	1	0	0	1	1	1	1	0	1	1	1	0	1	0	0	0	0	1	1
AC TYPE 2	1	1	,5	1	,5	,5	0	1	,5	,5	,5	1	,5	1	0	,5	0	1	0	1	1	1	1	1	0
ARX																									
Validity	0	0	1	0	1	0	1	1	0	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	0
AC	1	0,5	1	1	1	1	1	1	1	0,5	1	1	1	1	0,5	0,5	1	1	1	1	1	0,5	1	1	

4.3 Energy Signature Coefficients

An example of the Energy Signature coefficients for case 8 is found in Table 4-6, for both models LM66 and ARX66C (equation 4-I). The standardised coefficients indicate that the energy lags, especially the first lag, have a large effect on the energy use. The inputs for which time lags are used (Energy and Te) have high Variance Inflation Factors, which is expectable, but – as was explained in section 3.5.1 – this is overcome when the steady-state coefficients for the weather variables are calculated, and the lagged predictors are aggregated. For example, for the ARX66C:

$$C_{Te,steady\text{-}state} = \frac{c_{Te} + c_{Te,t-1}}{1 - (Q_{t-1} + Q_{t-2} + Q_{t-3})} \qquad [4\text{-II}]$$

When comparing the steady-state coefficients for ARX66 and LM66 (see Figure 4-5), it is found that the confidence intervals for all coefficients are wider for the ARX-model, and they overlap, except for c_{Te}, which is estimated significantly higher (in absolute value) by the ARX model.

ARX66C		***Coefficients***						***Partial correlations***	
	Unit	Estimate	Std. Error	t value	p-value	Standardised coefficient	VIF	Estimate	p-value
(Intercept)	kWh	14,19	1,90	7,46	0,000				
RADglob	kWh/(W/m²)	-0,02	0,01	-3,12	0,000	-0,04	1,24	-0,22	0,002
TEMP	kWh/°C	-1,97	0,15	-13,51	0,000	-0,32	5,2	-0,69	0,000
TEMP.l1	kWh/°C	0,77	0,17	4,39	0,000	0,13	7,4	0,29	0,000
WINDSPEED	kWh/(m/s)	1,78	0,27	0,27	0,000	0,08	1,3	0,41	0,000
Energy.l1	kWh	1,05	0,06	18,38	0,000	1,05	29,8	0,79	0,000
Energy.l2	kWh	-0,48	0,07	-6,51	0,000	-0,48	49,6	-0,42	0,000
Energy.l3	kWh	0,24	0,05	5,09	0,000	0,24	19,6	0,34	0,000

LM66		***Coefficients***						***Partial correlations***	
	Unit	Estimate	Std. Error	t value	p-value	Standardised coefficient	VIF	Estimate	p-value
(Intercept)	kWh	75,05	2,72	27,55	0,000				
RADglob	kWh/(W/m²)	-0,08	0,02	-4,59	0,000	-0,15	1,2	0,30	0,000
TEMP	kWh/°C	-2,69	0,42	-6,45	0,000	-0,44	5,0	-0,40	0,000
TEMP.l1	kWh/°C	-2,57	0,42	-6,19	0,000	-0,42	5,0	-0,40	0,000
WINDSPEED	kWh/(m/s)	5,39	0,74	7,25	0,000	0,23	1,1	0,46	0,000

Table 4-6: Coefficient Statistics – Case 8 – Heating Season 2011-2012 - LM50 and LM66

Figure 4-5 presents the steady-state coefficient estimates for all cases for the data of the first heating season. The estimates are presented for the linear regression model LM66 and a (case-specific) ARX-model (marked in blue in Table 4-2). Note that for some cases where an ARX-model is not needed, this is also a linear regression model. Like in case 8, also for the other cases where an ARX-model is selected, the confidence intervals are wider in comparison to the linear regression model. Knowing that the statistical quality and the goodness-of-fit of the ARX-models is much better than for the linear regression models, it is concluded that the standard errors of the estimates are for many cases underestimated in case of the linear regression model, making the estimates appear more accurate than they actually are. The coefficient estimates for the temperature coefficient have a confidence interval width of 1.5 kWh/°C on average, or about 20% of the estimate.

Figure 4-5: Energy Signature: Coefficients and 95% CI – Training data – LM66 and ARX

4.4 Comparison of Energy Signatures for 2 periods

In the study of energy use monitoring data, a key question is whether there are significant differences in the energy use of a building over time, that are very probably not due to changes in the weather, but to changes in the system consisting of the building, building services and the occupants, or in how this system reacts to the weather conditions. The purpose of this section is to detect changes in energy use between the first heating season 2011-2012 and the second heating season 2012-2013 for each of the 25 case-study houses. This will be investigated using the Energy Signature models that were selected in section 4.1, and include a suitable LM or ARX-model for each case and for each heating season. When the models are fitted to the data for the first heating season, in 20/25 cases it is accepted that the statistical model assumptions are fulfilled. For the second heating season, there are about 10 cases that are considered invalid. Those cases for which the assumptions were not entirely fulfilled are indicated in Table 4-7 (* and °). When drawing conclusions, these cases need to be observed more in detail.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Model Validity *(*, ° or § if not considered valid)*																									
HS 1				*												*		*	*						*
HS 2	°	°		°		°			°						°		°	°	°						°
HS 1+2 (IA)	§	§		§		§			§						§	§	§	§	§						§
Comparing the energy use in the two heating seasons *(X if changes are detected)*																									
Coefficients	X			X			X							X			X								
Interaction	X		X	X		X	X	X	X	X				X		X	X	X			X	X		X	
Sign. Level	0,001	1	0,001	0,01	1	0,05	0,010	0,001	0,001	0,05	1	0,1	0,01	0,001	1	0,01	0,001	0,001	1	1	0,001	0	1	0,01	1
Conclusion	X	/	X	#	//	#	X	X	X	X	//	//	//	X	/	#	X	#	/	//	X	X	//	X	/
	{X = Change; # = Change?, invalid; / = No change, invalid; // = No change, valid}																								

Table 4-7: Energy Signature – Comparison HS1 vs HS2

Coefficients

One way to compare the two heating seasons, is by comparing the Energy Signature coefficients that are found by fitting the model to each of the heating seasons separately. The steady-state coefficients for the intercept, exterior temperature, solar irradiation and wind speed are presented in Figure 4-6 for both heating seasons. If zero is included within the confidence interval, it is not proven that the coefficient estimate is significantly different from zero (at the 95% confidence level). For example for case 1 in the second heating season, this happens for all coefficients. For this case, the energy use is zero during about half of the 2nd heating season (see Figure 4-7). Therefore little input data is left to use for fitting a model, and it describes especially the warmer winter periods, leaving very few data points to describe periods with lower ambient temperature. This results in an invalid model with very high uncertainty.

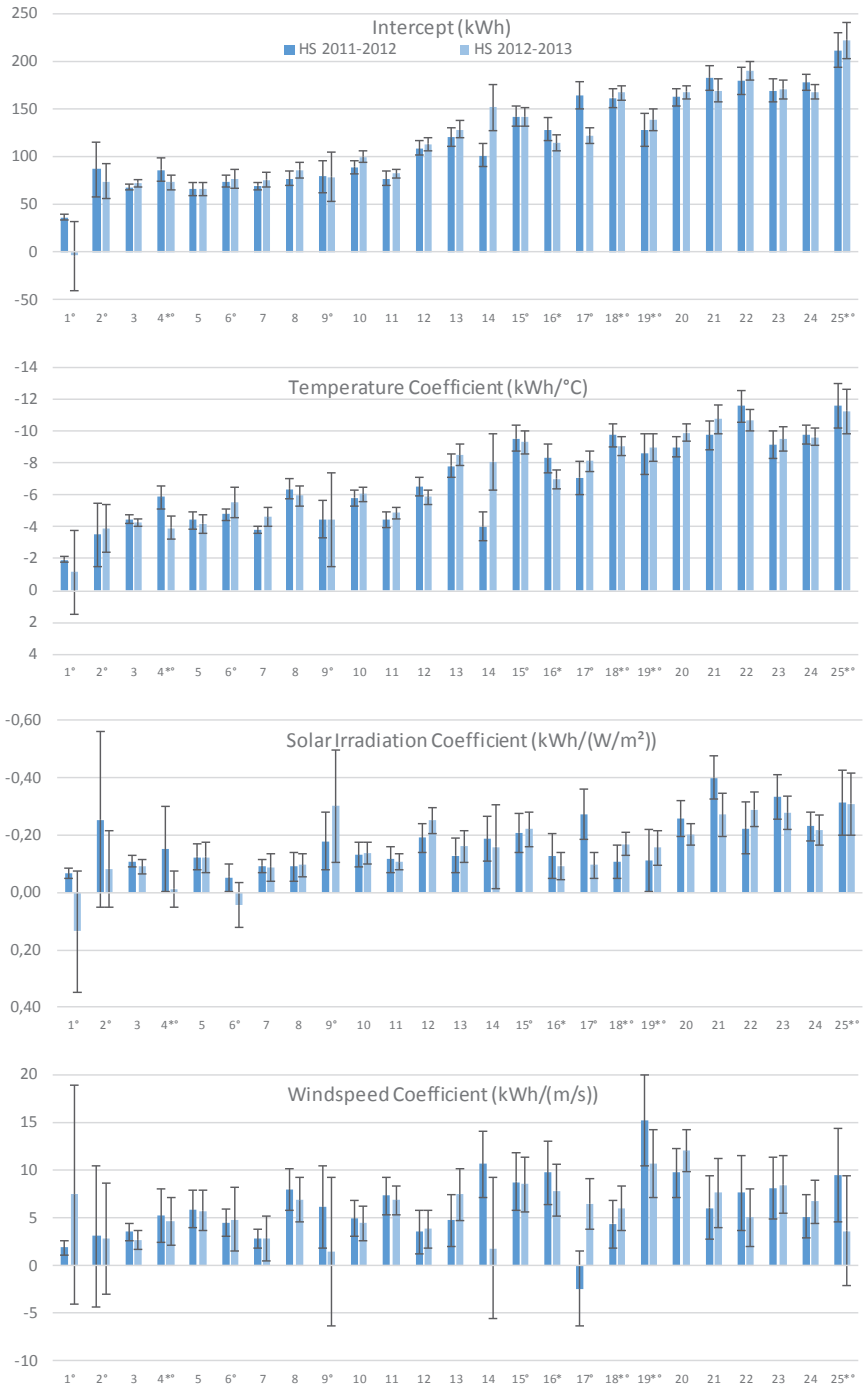


Figure 4-6: Energy Signature: Coefficients and 95% CI – HS2011-2012 and 2012-2013 – ARXselect

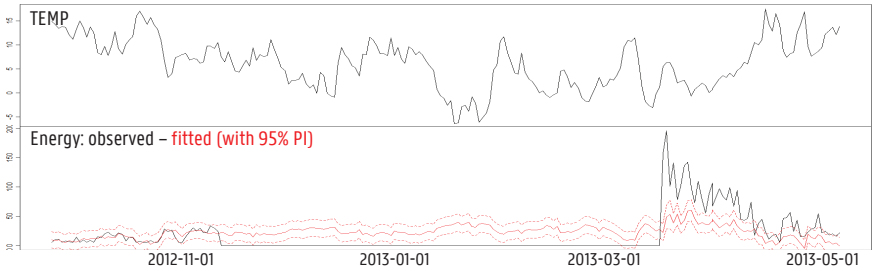


Figure 4-7: Case 1 – Time series of Te and Energy use in HS2 and prediction based on HS1-model

When comparing the Energy Signatures of the 1st and 2nd heating season, the confidence intervals of the coefficients should be compared. If the model is valid and the 95% confidence intervals of a coefficient do not overlap, there is a 95% chance that the coefficients are different, and for these cases it is concluded that the Energy Signatures are different. For the remaining cases, no difference in Energy Signature can be detected using this method. For 5/25 cases at least one coefficient is significantly different (1, 4, 7, 14 and 17). For some of them model validity is questioned, but when observing the data, this merely confirms that a difference appears. Coming back to case 1, because of the invalidity of the model, it is hard to conclude that the difference in coefficient estimates actually means that the energy use during the occupied period in the second heating season is different from the first heating season. However, when the (valid) energy signature that is fitted on the first heating season data is used to predict the energy use during the second heating season (see Figure 4-7, red lines), it becomes clear that the energy use of the building has increased after the period of un-occupancy.

Interaction model

In statistics, a frequently applied approach to identify whether two data sets are well characterised by the same model, is to fit the base model by use of the data from both data sets, while adding a qualitative predictor to the model, that identifies to which of both sets the data point belongs. For example, for the Energy Signature model based on LM66, equation [3-IV] is than transformed to:

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Rg_t + c_4 \times Ws_t + c_5 \times Te_{t-1} + c_6 \times PE_t + \varepsilon_t$$

$$\text{with } PE_t = \begin{cases} 0 & \text{if } t \in \text{HS2011-2012} \\ 1 & \text{if } t \in \text{HS2012-2013} \end{cases} \qquad [4\text{-III}]$$

If the data point belongs to the first heating season, PE_t is zero and the equation is reduced to the original base model. If c_6 is not proved significantly different from zero (by analysis of the p-value), the equation reduces to the original base model, and it is concluded that the Energy Signatures of both data sets are not significantly different. If c_6 is significantly different from zero, there is a significant difference between the Energy Signatures for both data sets and it is here modelled as an additive effect on the intercept coefficient: the intercept coefficient for the second heating season is then $c_1 + c_2$. Because it is possible that the difference between the two data-sets is not additive in the intercept but is specifically related to one or more of the predictors, the model can be expanded to include interactions between the qualitative predictor and the original variables.

For example, equation [4-III] then becomes:

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Rg_t + c_4 \times Ws_t + c_5 \times Te_{t-1} + c_6 \times PE_t + c_7 \times PE_t Te_t + c_8 \times PE_t Rg_t + c_9 \times PE_t Ws_t + \varepsilon_t \quad [4\text{-IV}]$$

Differences between the datasets can now be assigned to specific weather predictors and by analysis of the p-values it can be found which coefficients are significantly different from each other. During the analysis, it is found that if the coefficients of the interaction terms are found significant, this typically occurs for only one or a few terms. In order to avoid that the interaction effects are concealed by either absent or redundant terms (as a result of model misspecification), a stepwise selection procedure is applied in order to define the best suited model for each case specifically. The procedure is a backward selection procedure, with stepwise exclusion of the least significant terms, until only significant terms are included in the model. The initial 'full' model includes the main weather variables Te, Rg, Ws and Te_{t-1}, and the 1st up to 8th AR-term, the qualitative predictor and its interactions with the other variables.

Following this modelling approach, for each case a model is constructed that describes the energy use for the first and second heating season. As with the previously defined models, regression assumptions need to be observed. In Table 4-7 (HS 1+2 (IA)), the cases are marked (§) where model assumptions were not completely fulfilled. The model assumptions that were mostly harmed are constant variance and/or normality, as with the models for the individual heating seasons. Then the model coefficients are observed, and, more specifically, whether or not significant interaction terms (95% confidence) occur (see Table 4-7, *Interaction*). Now the following conclusions can be drawn on the energy use between the two heating seasons:

- For 11/25 cases marked with "X", it is concluded that there is a difference in energy use between the first and second heating season. For 8 of them, the model is valid and the coefficients are trusted. The change in energy use can occur for only a period between the two heating seasons, but it is clear enough to influence the model coefficients. For the 3 other cases (see Figure 4-7 and Figure 4-8), there are issues with model validity in one of the two heating seasons only, and – after a close observation – it is decided that these issues merely confirm that there is a difference between the two heating seasons, but the change in energy use appears somewhere during one of the heating seasons.
- For 4/25 cases marked with "**#**", interaction coefficients are significantly different from zero, but the model is invalid and therefore the conclusions are uncertain.
- For 4/25 cases marked with "/", none of the interaction terms is accepted to be significantly different from zero, but heteroscedasticity or issues with normality are observed. For example for cases 15, 19 and 25, the variance increases with increasing daily energy use, and this happens as well in the separate models for each of the heating seasons, so there is no reason to accept a difference in energy use. Also for case 2, the issues take similar shapes for both heating seasons. It is concluded that no significant difference in energy use can be observed using these models.
- For 6/25 cases marked with "//" the interaction model is valid and no interaction terms are significant. Therefore no difference in energy use is detected. (However it is possible that changes occur, for example smaller changes, but it is not possible to detect them).

Remark that using the approach with the interaction terms, the comparison between the two heating seasons discloses more cases where a change in energy use occurs and leaves less cases undecided than when models are fitted to the separate datasets. For some of them, the change in energy use can be observed visually, but for other cases, they are hard to recognise. In Figure 4-8 two cases are presented where the change can be observed visually. For example in case 9, a difference is observed when comparing the model fit with the observations, or the residuals, before and after beginning of January. Similarly, for case 17, this difference is seen around mid December 2011. The identification of changes could be improved by an approach that allows shorter periods to be compared, so that temporary changes in energy use can be detected more easily, or approaches that allow to compare data at shorter time steps, so changes in energy use patterns could support the decision. Secondly, remark that if this comparison was made using the standard linear regression instead of the case-specific models, much more cases would have remained undecided because of the invalidity of the models.

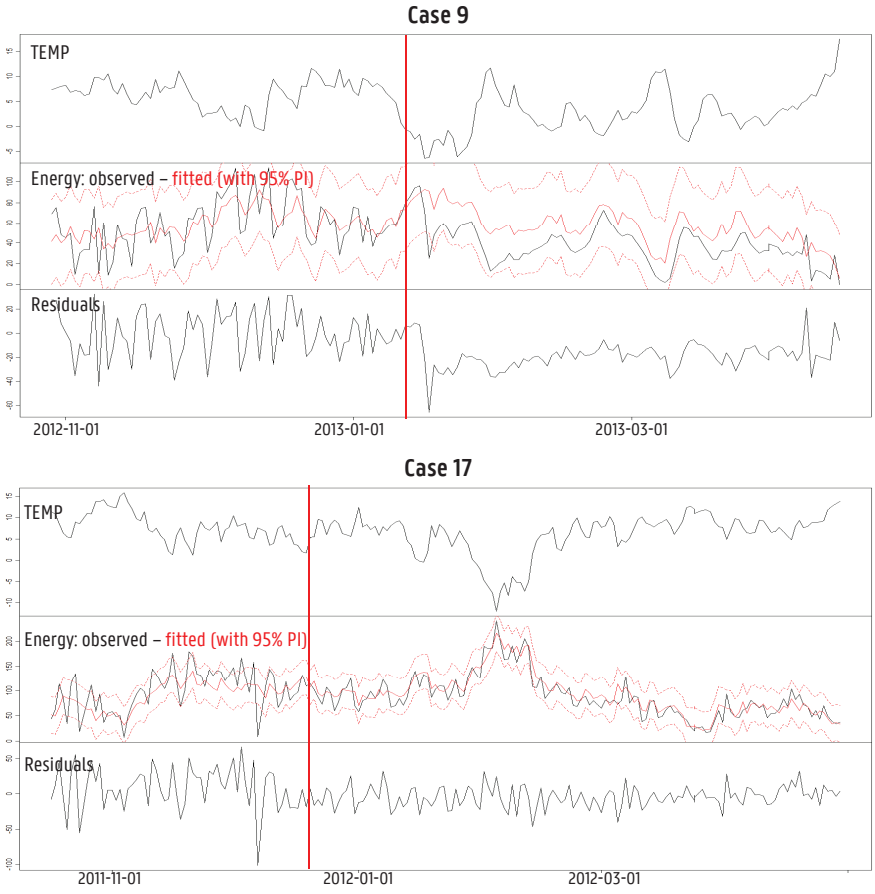


Figure 4-8: Changes in energy use – time series plots - examples

4.5 Prediction accuracy

4.5.1 Heating Season Predictions using Heating Season Training data

The in-sample and out-of-sample prediction accuracy statistics are determined for each case, for a suitable ARX or LM model (ARXselect, in blue in Table 4-2). The selected models are fitted on the full heating season 2011-2012 and the goodness-of-fit measure is presented in Figure 4-9. In comparison to the linear regression base model, the adjusted R^2 increases for the ARX models with on average 0.04 pp, and up to 0.16 pp (case 8).

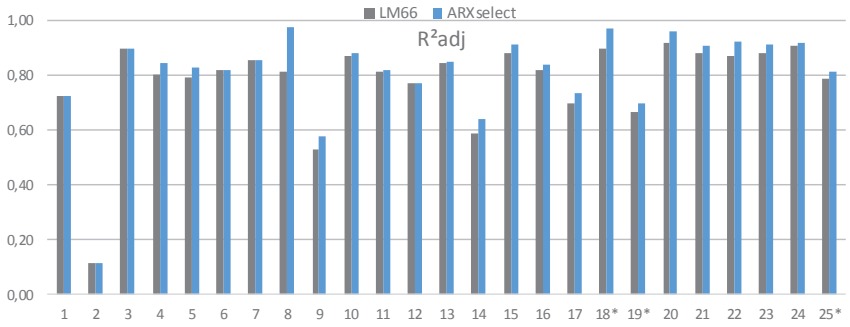


Figure 4-9: R^2adj – LM66 and ARXselect

For predictions with ARX-models, where recent historical values of the output are used as inputs into the model, the predicted outputs for the previous time steps are entered into the prediction equation. For both in- and out-of-sample predictions, the prediction period is the start of the heating season and starting values for the auto-regressive terms are zero. Then the predictions are made chronologically and each time the auto-regressive terms are updated with previously predicted values. The inputs for the exogenous terms, the weather data for the test period, are available (as with the linear regression models).

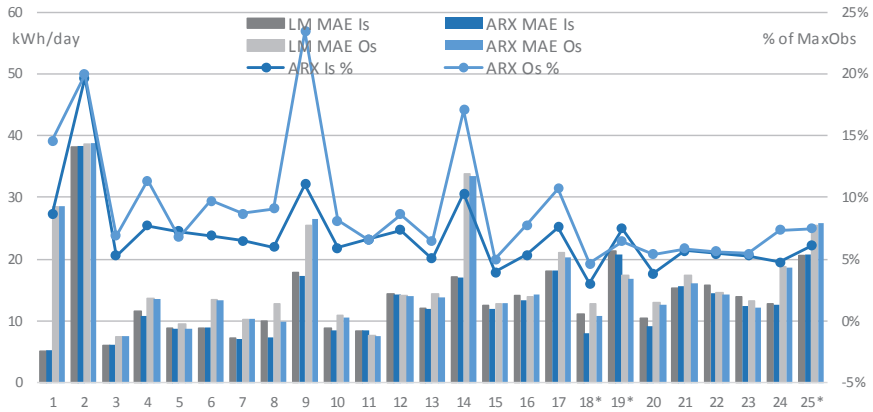


Figure 4-10: MAE and nMAE (%) for In-Sample and Out-of-Sample predictions – LM66 and ARXselect

For the cases where an ARX model is selected, the in-sample MAE decreases on average 6% and maximum 30%. It is on average 12 kWh/day (5 to 21 kWh/day). In about 9 cases, the difference in MAE with the LM66 model is smaller than 2%. For the out-of-sample predictions, with the heating season 2012-2013 as test data set, the findings are similar (Figure 4-10). In comparison to the linear regression models, the prediction accuracy increases only a few percent for most cases, but for some cases (e.g. 8, 18, 21, 23) the predictions are a lot closer to the observed values, with decreases in MAE up to 23%. For example, observe the out-of-sample predictions for case 8 in Figure 4-11, for the LM66 and ARX50C models respectively: the residuals for the ARX50C model are clearly smaller[8]. Now compare the MAE for the in-sample and out-of-sample predictions, and eventually the normalised MAE (the MAE divided by the maximum observation of the data set). It is found that for those cases where no significant difference in energy use was detected between the first and second heating season (see Table 4-7), the MAE for both heating seasons are close. For cases with a difference, often large differences in MAE are found (case 1, 9, 14).

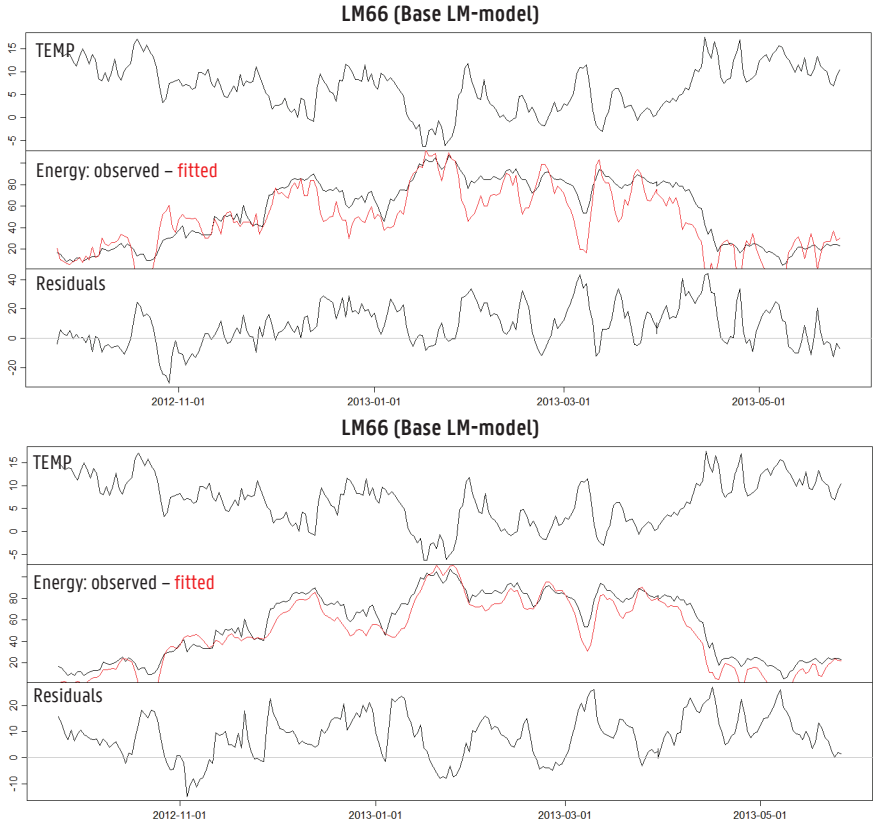


Figure 4-11: LM66 – ARX50C – Case 8 - Test data

8 Remark however that in the 0s-residuals of the better ARX50C model in Figure 4-11, tendencies are observed. It seems that in this case still an slower phenomenon remains, that can be due for example to longer term adaptations to the space heating control. In the 1-step ahead predictions (e.g. the in-sample 1-step ahead predictions in Figure 4-4), these tendencies are less obvious, but also present.

4.5.2 Heating Season predictions using 3 months Training data

When the training data period is reduced to 3 months data, and the goal is to estimate the energy use for the complete heating season, the MAE of the predictions typically increases for the majority of the cases and training data periods. In section 3.6 the effects of reducing the training period was discussed in detail, by analysing the results for 12 cases where the linear regression model assumptions were fulfilled quite well. In Figure 4-12 the results are summarised for all 25 cases (except for case 2) using the case-specific model from Table 4-2 and training periods of 3 months during the first heating season. When the months of Jan-Feb-Mar or Feb-Mar-Apr are used as training data sets, the MAE of the predictions on the complete first heating season increases on average 8% and 6%, with a maximum of 33%, in comparison the same model with a full heating season training period. The prediction accuracy is than on average in the same range as if the linear regression model LM66 was applied to the full heating season training data, and for about 9 cases it is significantly lower. Also in section 3.6, it was found that these periods result in the highest prediction accuracy. For the other combinations of 3 months training data, the MAE increases between 20 and 30% on average. The percentage decrease or increase in MAE in comparison to the linear regression models is also summarised in Figure 4-12 (right). For the full heating season training data sets, the prediction accuracy improves, or at least it remains equal, for all cases. For the three-month training periods cases exist where the prediction accuracy is lower for the ARX-models. For the "better" training periods Jan-Feb-Mar and Feb-Mar-Apr, this happens only in two cases, and is limited to an increase in MAE of 6%. For the other training periods, the situation is worse, both in amount of cases harmed (up to 9 in the period Oct-Nov-Dec and in severity, up to 33% for cases 20 and 23 in the period Nov-Dec-Jan. It is concluded that the reduction to three months data periods goes together with a loss in accuracy, but for the most suitable training periods, the increase remains below 10% for the majority of cases.

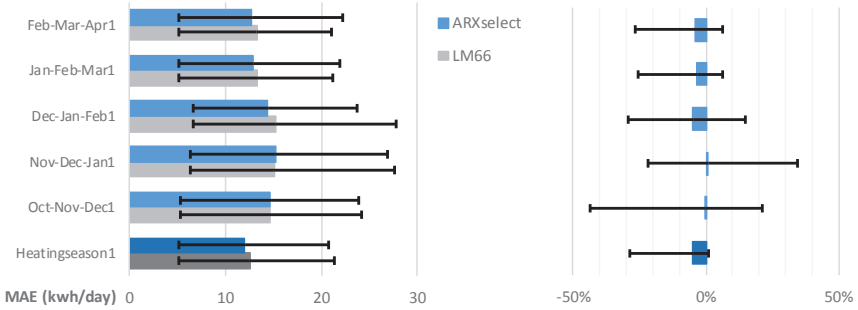


Figure 4-12: LM66 and ARXselect: MAE and relMAE for heating season predictions

4.5.3 Predictions on a Successive test data set

In the previous sections, the ARX-models were applied for long-term predictions over a period that does not follow the training data period, and using the predictions for the previous time steps to complete the autoregressive terms. A different application is the short-term predictions of the energy use, for example the one-step ahead predictions, where tomorrow's energy use is predicted using the observed energy use for today (and eventually previous days) in the autoregressive term(s). For example Figure 4-14 presents the in-sample and out-of-sample one-step-ahead predictions of the energy use for case 8 (using a full heating season training data set and the ARXC-model), that are indeed very close to the observed values, and much closer than when a long-term prediction horizon is used (Figure 4-11).

A data set consisting of the 3 months of Jan-Feb-Mar from the first heating season is used to train LM and ARX-models and for each case the model with the highest in-sample one-step ahead prediction accuracy is selected amongst models LM66, ARX50A, B, C and E. Then the model is applied to different test data sets with shorter and longer prediction horizons and the prediction accuracy is calculated. Now observe Figure 4-13 that presents the summary statistics for the relative MAE (relative to an LM66 model) for the 20/25 cases where an ARX model was selected[9]. As was already discussed in the previous section, the full heating season predictions (0s-HS) are up to 30% better than when a linear regression model was used, and for a few cases, they were worse. The in-sample one-step ahead predictions (Is-1s) are between 1 and 50% better than when a linear regression model is used, which illustrates the capabilities of the models to give better predictions than the linear regression models on the short term. Furthermore, test data sets successive to the training data set are applied, with a length of 3 days, 1 week and 1 month. The autoregressive terms are updated with the previous time step predictions. It is found that for some cases the predictions are much worse than with the linear regression models, but for the majority of the cases, the MAE decreases. In comparison to predictions with the linear regression model, the increase in prediction accuracy is larger as the prediction horizon becomes shorter. For predictions for the one week or 3 days following the training data, the MAE decreases on average 14% and up to more than 50%. This example illustrates the potential of the ARX-models for short-term predictions. For applications where prediction in an on-line setting are needed, where the goal is to provide short-term predictions as accurately as possible, the analysis should be further continued beyond the level of this illustration, for example by application on short-term test data sets for different periods, so the results are less affected by test periods with exceptional or irregular energy use. Furthermore, the ARX-model and the prediction process could be further optimised, e.g. by adding the most recent observations to the training dataset and re-estimate the model at every time step, so the model can adapt to the most recent or short-term evolutions in energy use, or to elaborate on the model structure, e.g. by modelling the residual terms.

[9] In 5 cases a linear regression model was the best suitable model fort his training data set and these cases are thus not considered in the summary statistics.

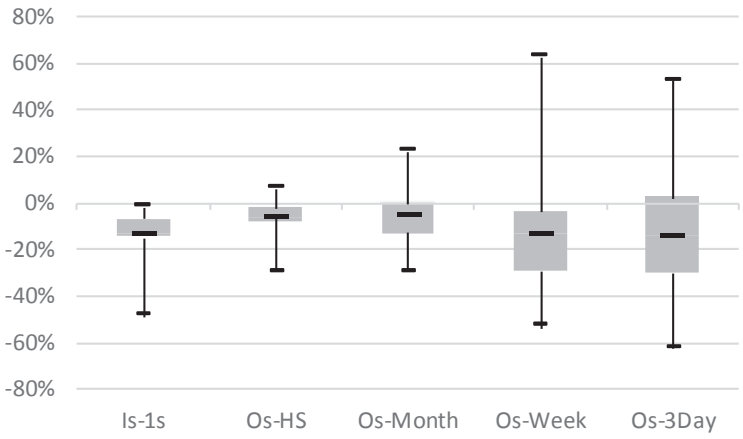


Figure 4-13: Training data Jan-Feb-Mar: relMAE (ARXselect / LM66)

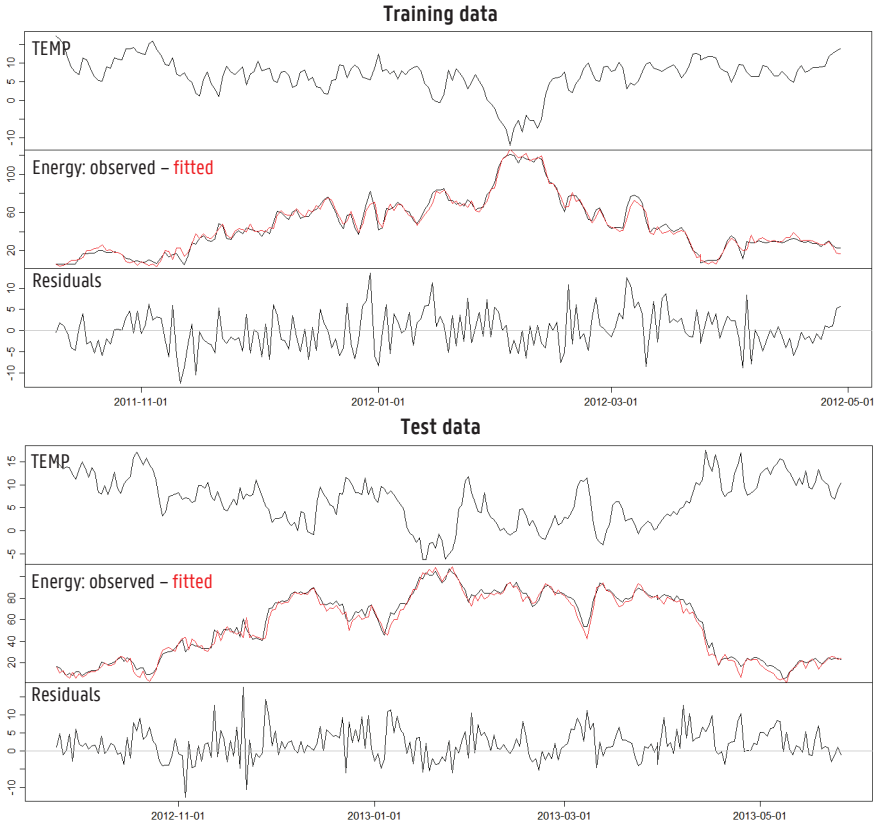


Figure 4-14: ARX50C – Case 8 – One-step ahead predictions - Training data and Test data

4.6 Time step

So far in this chapter, autoregressive models have been applied on daily aggregated data. Now, the data is aggregated to respectively 2-hourly and weekly values.

4.6.1 2-hourly aggregated data

In datasets with shorter, sub-daily time steps, the issue of autocorrelated residuals is typically more severe and by use of linear regression models a poor performance is obtained (see section 3.7.1). By way of illustration, for case 25, the 2-hourly time series from January, February and March 2012 are used to train models with and without auto-regressive terms. The data contains a clear daily cycle. When a linear regression model (without autoregressive terms) is constructed, the only weather variable for which the coefficients are significantly different from zero, is the exterior temperature. The goodness-of-fit of the model is very low ($R^2adj=0.09$) and the predictions are not very accurate. The MASE for this model is 0.95, meaning that the prediction accuracy (in terms of MAE) is not much better than if the naïve prediction method would be used. If now autoregressive terms are added (see equation 4-I), the first energy lag is significant, but the model fit improves tremendously as soon as the 12^{th} energy lag (= 24 hours lag) is added. Furthermore, since also a weekly pattern is observed in the data, the 84^{th} time lag is valuable. The autocorrelation function for the linear regression and ARX-model are plotted in Figure 4-15. It seems that the main inputs required to reduce the autocorrelation in the data are now present in the ARX-model, however further optimisation is possible by fine-tuning the model construction. The goodness-of-fit has now increased to reasonable values, e.g. an adjusted R^2 of 0.76. In the graph on the right top, the observed and predicted energy use for a 12-day period in January 2012 are drawn, and now the in-sample predictions are able to follow the variance available in the data (daily cycle) as well as the weekly pattern.

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Q_{t-1} + c_4 \times Q_{t-12} + c_5 \times Q_{t-84} + \varepsilon_t \qquad [4\text{-V}]$$

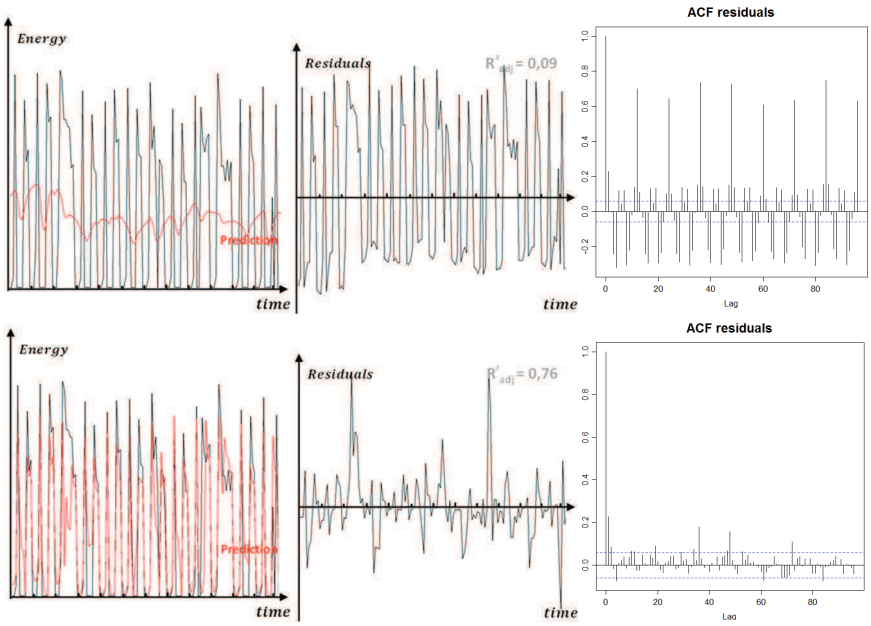


Figure 4-15: Autocorrelation in 2-hourly Time Series data: LM (top) and ARX-models (bottom)

4.6.2 Weekly aggregated data

In traditional applications of Energy Signature models, monthly or weekly energy use meter readings are available, or it is recommended to aggregate the data to at least weekly values for averaging out the dynamics. For illustrating the impact of the use of daily versus weekly aggregated data, the linear regression base model with the main weather variable, the exterior temperature, is now applied to weekly aggregated data of the first heating season. The model coefficients are then compared to the model coefficients of the selected LM- and ARX-models applied on daily aggregated data. It is found that the confidence intervals for the exterior temperature coefficient are between 1.2 and 2.8 times smaller compared to weekly aggregated data (disregarding case 2). For the intercept coefficient the standard errors are between 1.5 and 3.3 times smaller (disregarding case 2).

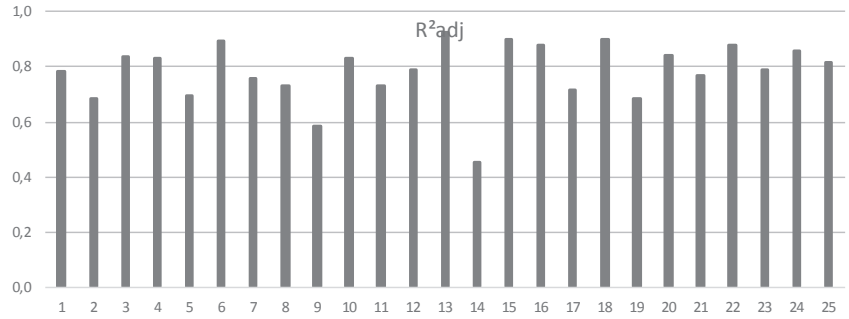


Figure 4-16: LM-model for weekly aggregated data - R²adj

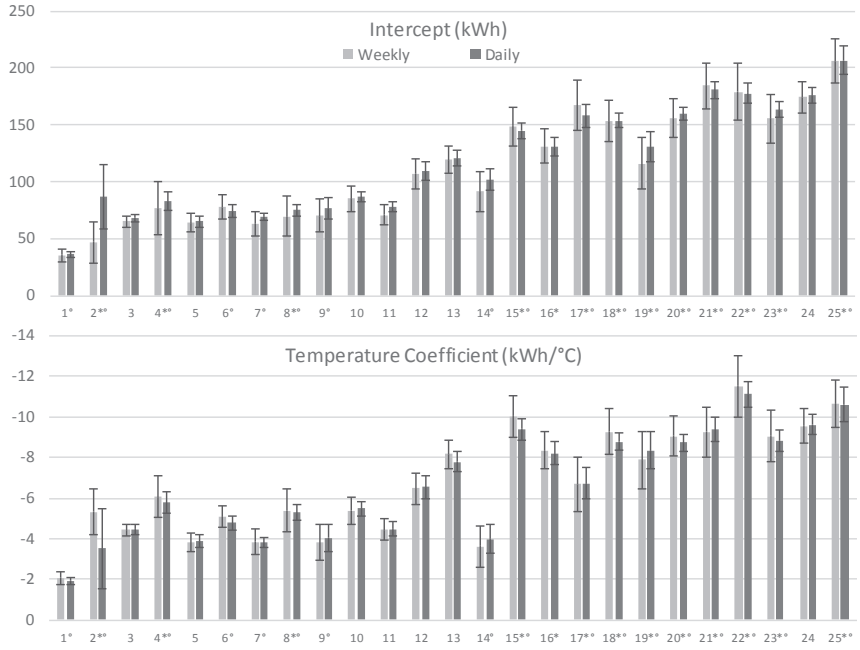


Figure 4-17: Energy Signature: Coefficients and 95% CI – HS2011-2012– Weekly and daily aggregated data

4.7 Summary & Conclusions

An ARX-model is composed of one or several auto-regressive terms and exogenous inputs. First ARX-models are fitted to daily aggregated data for the entire heating season 2011-2012. The key predictors for the ARX-model are the first auto-regressive term Q_{t-1}, and the three main weather predictors in the exogenous terms: exterior temperature *Te*, solar irradiation *Rg* and wind speed *Ws*. These predictors lead to clear improvements in the goodness-of-fit of the model and the coefficient estimates are significantly different from zero at the 95% confidence level for a majority of cases (for example 23/25 cases for Q_{t-1}). Note that the predictor Te_{t-1} when applied in this ARX-model, is not proven significantly different from zero in 14/25 cases, which happened in only 4/25 cases in the linear regression Energy Signature base model. For these cases, part of the knowledge that was allocated to Te_{t-1}, is now preferentially included in Q_{t-1}.

The ARX-model with the four key predictors is applicable and a benefit for many cases where the linear regression model was not suitable. It is also an improvement for some cases where the linear regression model was accepted. It is however not always sufficient to entirely resolve the autocorrelation in the errors: often more autoregressive terms are needed. For the given cases, the most effective additional predictors are the weekly time lag Q_{t-7}, or more recent energy use lags Q_{t-2} or $Q_{t-2} + Qt_{-3}$, or the lagged external temperature Te_{t-1}. While for the linear regression models, the additional parameters can be useful but not essential, for the ARX-model the additional predictors can be requisite in order to fulfil the regression assumptions. The selection of these predictors is case-specific. When an automated procedure is applied, it is proposed to apply a backward stepwise elimination procedure, starting from a model that includes the 3 main weather predictors Te, Rg and Ws, the first temperature lag Te_{t-1}, and the autoregressive terms 1 up to 8. Eventually additional weather predictors could also be added, but the effect on the selection procedure has not been investigated here.

The model selection has been performed manually, and for each case a suitable model is selected for which the linear regression assumptions are fulfilled and the coefficients are significant. For 6 cases the linear regression base model (LM66) is selected and for 9 cases the ARX-model with the key predictors (ARX50A). For the 10 remaining cases, additional autoregressive terms are used: Q_{t-2} in 3 cases (ARX50B), $Q_{t-2} + Qt_{-3}$ in another 3 cases (ARX50C) and Q_{t-7} in 4 cases (ARX50E). For these case-specific models, the statistical model assumptions are now accepted in 22/25 cases, which is a clear improvement compared to the linear regression models. The goodness-of-fit is between 0.58 and 0.98, on average 0.84. In the remaining 3 cases heteroscedasticity is observed. In 5 cases small autocorrelations remain at a specific time lag, which will probably be resolved in an automatic model selection procedure.

The coefficient estimates and prediction errors are then investigated for the case-specific models (including 6 linear regression and 19 ARX-models for the 1st heating season). When comparing the estimates from the ARX-models with those from the LM-model, applied on the same cases, it is found that the confidence intervals for the estimates are larger for the ARX-models, and so the uncertainty of the coefficients was underestimated in the (not-accepted) LM-models.

In the study of energy use monitoring data, a key question is whether there are significant differences in the energy use of a building over time, that are very probably not due to changes in the weather, but to changes in the system consisting of the building, building services and the

occupants, or in how this system reacts to the weather conditions. By use of the estimated energy use coefficients for the first and second heating season separately, the energy uses can be compared to each other, and for about 5 cases a significant change in Energy Signature is detected. A more precise way to compare the two heating seasons is by investigating a model for the combined data for both heating seasons that includes interactions between the variables and a qualitative predictor that indicates to which heating season each data point belongs. If one or several of these interaction terms are significantly different from zero, it is concluded that there is a significant difference in energy use between the two periods. Otherwise, no change in energy use is concluded. In reality, there could still be a change in energy use, e.g. if it is smaller than the model allows to identify, or occurs only a small part of the observed period. Following this approach, for 11/25 cases a difference in energy use is identified between the two heating seasons. For 3 cases there are also indications that there is a change, but the conclusions are less certain. For the remaining 10 cases there is no reason to accept a change in energy use.

The prediction accuracy is investigated for different situations. In a first situation the in-sample and out-of-sample predictions are made for a full heating season, using the first heating season as training data set. It is found that in comparison to the linear regression models, prediction accuracy (MAE) increases, but only with a few percent for most cases. However, for some specific cases with severe autocorrelations, the predictions are a lot closer to the observed values, with MAE decreasing up to 23%. It is also found that for those cases where no significant difference in energy use was detected between the first and second heating season, the MAE for the first and second heating season are close to each other. For the cases where a difference was detected, often large differences in MAE are found, confirming the conclusions.

In a second situation the training data period is reduced to 3 months. For the better training data sets Jan-Feb-Mar and Feb-Mar-Apr, the MAE increases in comparison to the same model with a full heating season training data, but for most cases the increase remains below 10%. In comparison to the linear regression models, for all but 2 cases, the prediction accuracy increases, up to maximum 30%. In the third situation, short-term predictions are observed. The training data set is the three-month period Jan-Feb-Mar and the prediction periods stretch over 3 days, 1 week and 1 month successive to the training data set. For some cases the predictions are much worse than with the linear regression models, but for the majority of the cases, the MAE decreases. In comparison to predictions with the linear regression model, the increase in prediction accuracy is larger as the prediction horizon becomes shorter. For predictions for the one week or 3 days following the training data, the MAE decreases on average by 14%. This example illustrates the potential of the ARX-models for short-term predictions, for example predictions in an on-line setting.

The application of auto-regressive models to a data-set with 2-hourly aggregated data is illustrated by use of an example case. For sub-daily data the issue of correlated residuals is typically more severe and linear regression models are not suitable at all. By use of ARX-models the adjusted R^2, indicating the goodness-of-fit of the model, increases from 0.09 to 0.76, which is a decent result. The most important autoregressive term is the one that corresponds to the 1-day time lag, due to daily cycles that exist in user behaviour as well as in the weather. Furthermore, also further lags like the 7-day lag that also turn up in the daily aggregated data are important, such as the 7-day lag. As with the daily aggregated data, the ARX-model structure will be case-specific.

Finally, for the valid linear regression and ARX-models, the accuracy of the models has improved in comparison to the linear regression models with weekly aggregated energy use data. For example the confidence intervals for the exterior temperature coefficient are between 1.2 and 2.8 times smaller.

5

Linear regression with energy use time patterns

In measurements of the energy use at a sub-daily time interval, the diurnal fluctuations in energy use are observed, and patterns in the energy use are recognised. These are called energy use time patterns and they provide additional insights in the building energy use, since they enable to recognise energy use profiles, stand-by of the heating system, night set-back etc. The first objective of this chapter is to mathematically identify the different groups of energy use time patterns occurring in energy use time series at a sub-daily time step. Two main questions are:

- **Research question 3: How can groups with similar Energy Use Time Patterns be recognised mathematically? (section 5.2 and 5.3)**
- **Research question 4: How are groups with similar Energy Use Time Patterns characterised in function of specific weather conditions and/or calendar information? (section 5.4)**

Days with different energy use time patterns may be understood as days in which the system is in a different state, and may explain a part of the variation in the daily energy use data that is not explained by weather variables. The second objective of this chapter is to evaluate whether the daily LM- and ARX-models from chapters 3 and 4, can be improved by adding a variable that identifies to which group of energy use time patterns each day belongs:

- **Research question 5: Does the inclusion of groups of energy use time patterns into linear regression and ARX-models improve the Energy Signatures with daily aggregated data? (section 5.5)**

Finally, the Energy Signatures for two periods are compared (section 5.6) and the models are applied on a dataset with energy use for space heating and domestic hot water (section 5.7).

5.1 Introduction

The space heating energy use in occupied houses is a result of the interaction between the building, building services and users, and their interaction with weather variables (see Chapter 1). In classical Energy Signature models, where for example monthly or weekly aggregated data are used, it is accepted that most of the variations in the interactions of the user with the building or weather variables are averaged out, and only long-lasting user-related changes (e.g. when people move house) remain to be detected. Therefore, the classical Energy Signature models can focus on the relationship between the energy use and the weather variables and linear models can be used. In previous chapters it was demonstrated that when daily energy use data are used, this assumption is no longer correct. For many cases the presence of auto-correlated errors hints towards user-related aspects that are not 'aggregated' in the data, for example the auto-correlations occurring at weekly intervals that suggest weekly occupancy schedules. In the previous chapter, auto-regressive terms were added to the model in order to deal with this time-dependency in the residuals and to linearize the system again. When focussing on the auto-correlations of the 2nd type, this works especially for regularly recurring patterns.

In the situation where sub-daily data are available, diurnal fluctuations in the energy use can be observed more closely. In this study, the patterns that occur are called Energy Use Time Patterns (time patterns of the energy use). For example in Figure 5-1 (top) the 2-hourly energy use time series for the month of January 2011 are plotted for case 18. In this very clean example, a repetitive daily pattern can be recognised, consisting of a period of zero energy use at night, a peak in the morning and moderate energy use during the rest of the day. This resembles the effect of the nigh set-back of the space heating system, followed by the morning start up and the continued space heating demand during the day. In the second example in Figure 5-1, a similar pattern is recognised during weekend days, while during weekdays the space heating is turned off during the day and presents peaks in the morning and evening, also on Wednesdays afternoon the heating is operational. Thus a weekly schedule is present, that suggests the schedule of working adults with children. In the third example the patterns are more irregular, but also days with heating throughout the day and days with heating only in morning and evening are present. Looking back to the survey responses, indeed house 18 is occupied by 1 person who is usually at home during the working hours, while house 12 and 17 are occupied by 2 people who are usually absent during working hours, and apply a lower set-point temperature when absent. When observing the 25 cases, for which the 2-hourly series are presented in Annex A of the study, these patterns are recognised in the majority of the cases, but the shape and regularity of the patterns differs from case to case. The variations in these patterns throughout the heating season are also influenced by weather variables, for example the situation in which the space heating system shuts down on a sunny afternoon, when solar gains are sufficient to maintain indoor temperature.

Thus by mere observation of the sub-daily energy use time series plots, insights in the energy-related behaviour of the observed system are acquired. These insights may be informative for the occupant, for learning when he is using high amounts of energy, or to energy auditors and commissioners, e.g. for a better understanding of the energy use of the house, to detect anomalies in for example the system control, to detect changes in energy use… However the visual inspection of often long energy use time series and the search for recurring patterns and

anomalies is a difficult and time-consuming task for people to do. If instead a computer could recognise these patterns and group them according to their similarity, then the characteristic energy use time patterns of the house can be summarised, for example in function of the days of the week they occur, or for which weather conditions they occur. This leads to the 3rd research question of this dissertation, that is how groups with similar energy use time patterns can be recognised mathematically. One step further is not only to visualise or summarise the patterns in function of calendar or weather variables, but to express this relationship in a mathematical model. That model then would enable to predict which pattern is most likely to occur for given calendar and/or weather variables. This leads to the 4th research question of this study, that is how groups with similar energy use time patterns are characterised in function of specific weather conditions and/or calendar information.

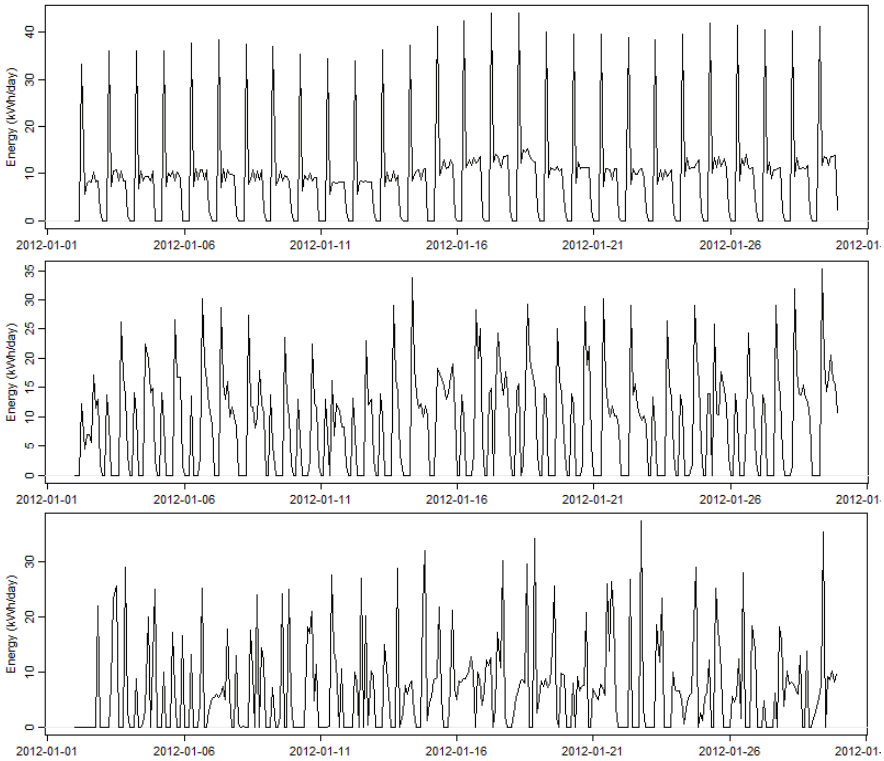


Figure 5-1: 2-hourly time series – Month of January 2012 - Case 18, 17 and 12

The energy use time patterns observed in sub-daily data can also help to explain some of the variation that is found in the daily aggregated data. In the second example in Figure 5-1, it is expected that the daily energy use will be different for weekdays and weekend days (even if the weather conditions are similar), because the space heating is active for a much longer part of the day during weekend days. In the classical linear regression models (Chapter 3), the variation due to the different 'states' of the occupied building is not explained, since it is not related to the explanatory (weather) variables of the model. In the ARX-models (Chapter 4), this variation is (at least partially) described by the auto-regressive terms. For example the 7th time lag of the energy use takes into account the similarity between days that appear every week.

However the variation related to less regular recurring patterns may be more difficult to deal with when using AR-terms. Therefore in this chapter the LM- and ARX-models are extended with an input that is derived from the energy use time patterns, and that indicates which days belong to a group of similar patterns. Thus, zooming out again from the sub-daily to daily aggregated data, knowledge from the energy use time patterns is integrated in the regression models. This leads to the 5th and last research question, that is whether the inclusion of groups of energy use time patterns into linear regression and ARX-models improves the Energy Signatures with daily aggregated data.

5.1.1 Approach

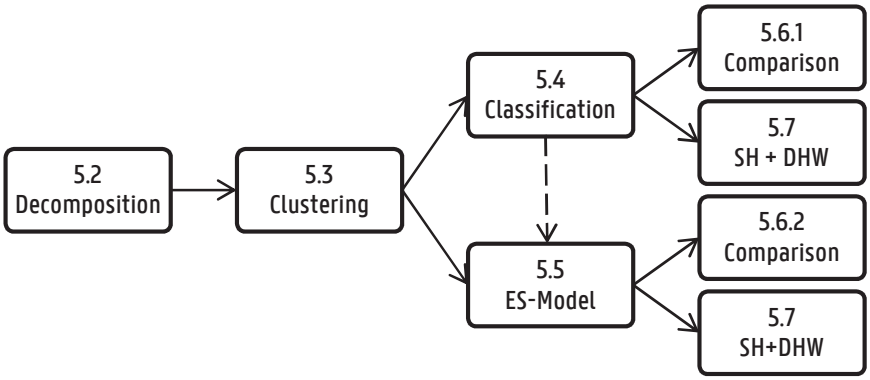


Figure 5-2: Modelling (with) energy use time patterns: approach and structure

The third research question is dealt with in section 5.2 and 5.3. In section 5.2 the energy use time series are decomposed into a seasonal component that deals with energy use time patterns occurring every day or week, a trend component that deals with the slower variations in the data throughout the year, and a residual component. Existing time series decomposition methods are applied and compared. The seasonal components are examined for exploring the prevailing patterns in energy use time series and their properties. However, none of the applied time series decomposition methods allows to differentiate the seasonal component for all patterns that might occur: regularly recurring patterns, occasionally occurring patterns, patterns related to specific weather conditions or to a change in energy use behaviour. Therefore in section 5.3 the goal is to mathematically group the patterns with a similar shape, and to analyse and present those groups of patterns in function of the external variables. The full length energy use time series are first detrended (the series is divided by its trend component) and cut into day-length time series. Each day-length time series represents a one day energy use time pattern. After pre-processing, these day-length time series are clustered into groups of patterns with similar shapes, using cluster analysis methods. As a result, the large amount of patterns is summarised in a limited number of clusters each having one representative pattern. The clusters of patterns are also graphically represented in function of the weather and calendar variables, which allows to visually explore the conditions in which certain patterns appear.

The mathematical relationship between the clusters of patterns and the weather variables is the subject of the 4th research question and is treated in section 5.4. The clusters or energy use time patterns are modelled in function of the weather and calendar variables. The results are case-specific energy use time patterns for different weather conditions and different days of the week, that indicate start and end of the heating season, stand-by of the system etc. The models allow to predict which pattern will occur in certain conditions, or to detect whether the occurrence of certain patterns has changed when comparing them for two periods (see section 5.6.1).

Section 5.5 deals with the 5th research question of this dissertation. Zooming out from sub-daily to daily aggregated data, for each day of the series the 'cluster' variable describes to which cluster of energy use time patterns it belongs. Thus the outcomes of the clustering in section 5.3 are entered into a categorical variable with a daily time step. This variable is added to the linear regression and ARX-models from chapters 3 and 4, as an additional term in the model, or in a combination of interaction terms of this cluster variable and the other regression variables (weather variables and AR-variables). If these terms are significant, than the regression model can be interpreted as a combination of equations for the different clusters. In a two-dimensional graph, for each (significantly different) cluster a different regression line would be obtained. The results of this method are compared to the results of the classical linear regression and ARX-models. The structure of this section is similar to the structure of the chapters 3 and 4, where the model variables and validity and the energy signature coefficients are discussed.

Section 5.6 elaborates on the comparison of the energy use time series of the 1st and 2nd year of the datasets, thus coming back to a possible implementation of the developed methodology, that is the detection of changes in energy use over different periods, for example for applications in energy feedback. The series are compared by analysis of the energy use time patterns, as well as by comparing the Energy Signatures.

While the models in this study were developed using datasets from houses where energy is used only for space heating purposes, the developed models can also deal with the energy use for a combination of space heating and domestic hot water demand. In section 5.7, this application is illustrated by use of two examples.

5.2 Time Series Decomposition

5.2.1 Time series decomposition methods

By use of time series decomposition methods the patterns that are visually distinguished in the time series plots can be separated numerically. This type of statistical methods allows to split the time series into several components, each having a typical characteristic or pattern [74,76,77]. Hyndman et al. mention four types of time series patterns: the trend, seasonal, cycle and irregular component. The seasonal and cycle components are repeating patterns. While the seasonal component has a fixed and known period, the cycle has an unknown and changing periodicity. The trend component is defined as the long-term direction of the series and the irregular (or random or error or residual) component contains the irregularities in the data, that are not explained by the other components [74,76]. Hyndman and Athanasopoulos further clarify that the seasonality in the data is associated with some aspect of the calendar (e.g. the day of the week, the quarter of the year), and that the cycle is typically longer than the length of the seasonal component, and its magnitude is more variable (e.g. a business cycle). Moreover, when the observed time series is relatively short, what is observed as a trend, may actually be part of a cycle that is longer than the length of the observed time series [74].

Relying on the prevailing literature sources, time series decomposition is often applied in econometrics, that is the field of economic modelling based on both theory on economic relations and empirical data, using statistical methods. The data-sets are for example time series of unemployment rates, stock market figures, sales or production volumes [76,78–80]. Therefore, the periods or lengths of the components are often referring to the time frame of economic studies, where the period of the seasonal component is for example one year (including quarters or months as time steps). However, time series decomposition methods can be applied to other groups of time series as well, as for example the electricity or natural gas use of entire cities or regions, the wind speed or atmospheric CO_2-concentrations [80–87]. There, diurnal fluctuations in the data can be relevant, leading to seasonal components with a much shorter period, e.g. one day. This is also the case with the energy use time series that are the subject of this study. For example, Figure 5-3 presents the decomposition of the time series from Figure 5-1 (case 18) into a seasonal component with period of one day, a slowly decreasing and increasing trend component, and an irregular or residual component, which is the difference between the observed series and the sum of seasonal and trend components (the 'fitted' values). Remark that the trend component tends to increase and decrease inversely to the slower movements in ambient temperature. In multi-year time series, it would probably be considered a cyclic component. The seasonal pattern determines the regular variations in energy use, here with a period of one day. They are the result of some properties of the building and building services (e.g. thermal inertia, insulation level, thermostatic control), as well as the energy use profiles and habits of the users (day/night cycles, indoor temperature settings…).

In the previous example, the three time series components are combined in an additive way. However different ways exist to estimate and compose the time series components. Kendall and Stuart describe two classical methods, that is the purely additive and purely multiplicative seasonal decomposition using moving averages [88]. Hyndman expounds that actually each of the components can enter the decomposition equation by addition or multiplication, and De

Livera selects the most commonly applied out of them: the log-additive decomposition and the pseudo-additive decomposition [87,89]. In the STL-method (*Seasonal-Trend decomposition based on LOESS*), Cleveland et al. apply the LOESS-smoother in a procedure with an additive nature, making it possible to capture changes in the seasonal component over time [81]. Each of these traditional methods allow for one seasonal component with one fixed period. Often seasonality is a more complex phenomenon, with multiple levels of seasonality, for example when both daily and weekly patterns are available. For these complex seasonality, which may also include non-integer seasonality and dual calendar effects, De Livera proposes the TBATS exponential smoothing formulation [86,87].

The decomposition of time series in components with specific properties is first of all a way of analysing time series for understanding their constitution by uncovering inherent structures. The graphical representations of these components can provide insights in the behaviour of a time series and changes in the components can give hints towards physical causes for variation in the time series. Secondly, once the seasonal components are obtained, they can be used to pre-process the data. On the one hand, in applications where the seasonal variations are not the main interest, but rather the underlying tendencies, seasonal adjustment is a frequently applied technique. The seasonal pattern is then removed from the observed data. On the other hand, time series can also be detrended in order to focus the analysis on the fluctuations around the trend.

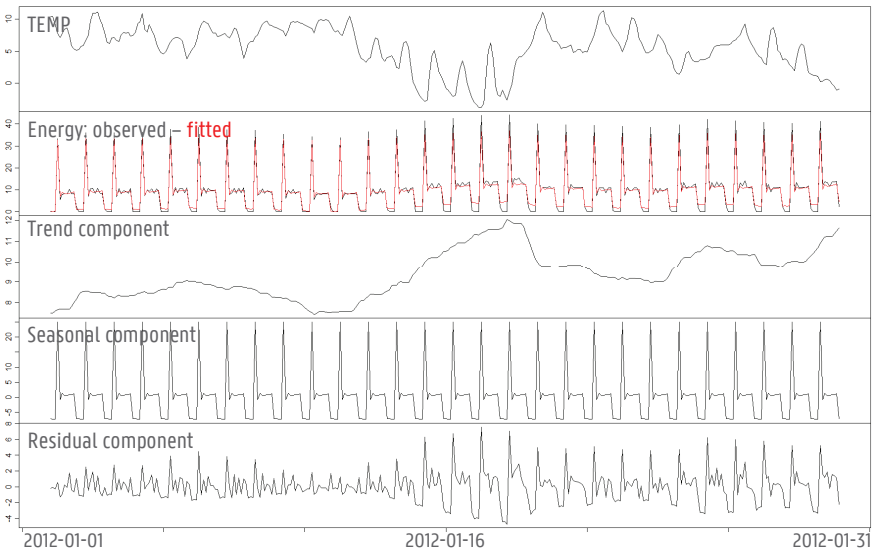


Figure 5-3: Time Series Decomposition of 2-hourly time series – case 18

5.2.2 Decomposition of energy use time series

The decomposition of time series in components with specific properties is a way of analysing time series for understanding their constitution. By separating the different components, the inherent structures in the data show up more clearly, and, by observing the residual component, exceptions to these structures are highlighted. In this section, some typical decomposition methods are applied to the energy use time series for the 25 cases for the heating season 2011-2012. The time step of the series is 2 hours and the seasonal pattern has a daily period (= 12 time steps). The suitability of the methods to the time series is investigated by observing the residual components and the proportion of the variance in the observed series that is explained by the combination of the trend and seasonal components. Through observation of the proportion of variance explained by each of the components, the presence of seasonality and/or trends in the data is mathematically described.

Additive Decomposition by Moving Averages

The additive seasonal decomposition by moving averages is a traditional and commonly used time series decomposition method. According to De Livera it originates from the 1920s and has formed the foundation for most of the decomposition techniques that are known today [87]. It is described in the handbook of Kendall and Stuart in 1976 [88]. The method is implemented in the R environment as a function *decompose,* which is used in this study [69,90]. In a first step, the trend component is determined by taking a moving average of the time series and it is subtracted from the time series. In this study a symmetric window with equal weights is used for calculating the moving average. For calculating the moving average at the beginning and end of the heating season, the data points right before and after the selected period are included, that have value zero. The de-trended time series can be considered a sequence of seasons with a fixed period, consisting of a number of time steps. The seasonal figure is determined by taking the average over all the periods for each time step, followed by centering the string of averages. The period of the seasonal component defines the length of the seasonal component. Typically, the same period is used for smoothing the trend component. Following the observations of diurnal patterns in the data, due to diurnal use and weather patterns, a period of one day is chosen for the seasonal and trend component calculations. Finally, the error component is estimated by subtracting the trend and seasonal component from the original time series.

$$Q_t = TC_t + SC_t + RC_t \qquad [5\text{-I}]$$

For each of the 25 cases, the full heating season energy use time series are decomposed into trend (TC), seasonal (SC) and residual components (RC). The variance in the observed time series (Q) can be expressed as a function of the (co)variances of the components as follows:

$$Var(Q) = Var(SC) + Var(TC) + Var(RC) + 2\times Cov(SC, TC) + 2\times Cov(SC, RC) + 2\times Cov(TC, RC) \qquad [5\text{-II}]$$

In the case of additive decomposition, the calculated covariances are negligible so the equation reduces to the Bienaymé formula and the variance of the original time series is simply the sum of variances of its components:

$$Var(Q) = Var(SC) + Var(TC) + Var(RC) \qquad [5\text{-III}]$$

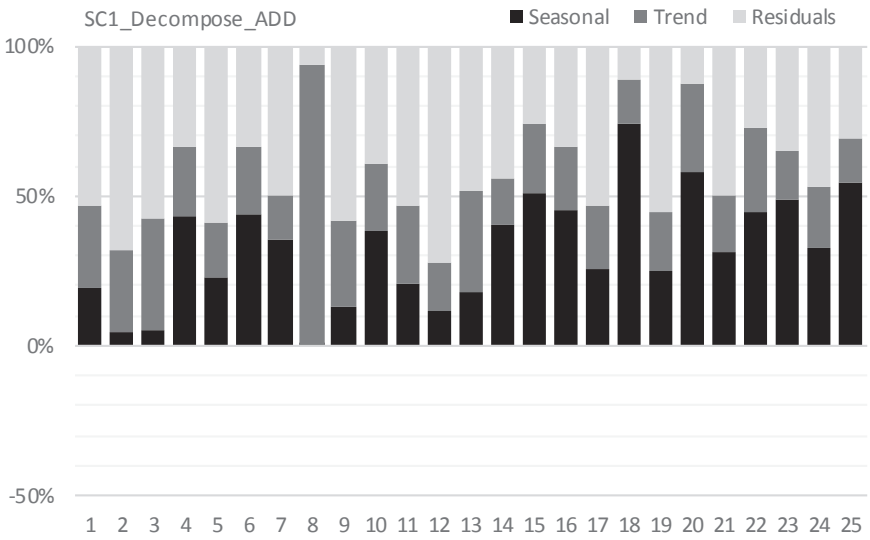


Figure 5-4: Additive TSD – Daily Period – Explained / Observed Variance (%)

In Figure 5-4 the share of the variances of the components in the variance of the original time series exhibits clear differences between the cases. In about 10 cases, the seasonal component explains more than 40% of the variance in the data, and, together with a trend component, between 60% and 90% of the variance is explained. The clearest example is found in Figure 5-3, that represents the additive decomposition of case 18, which has a very clear seasonal pattern. In 14 other cases, the seasonal and/or trend component are moderate or weak, and together they explain between 28 and 60% of the variance in the data. A moderate example is case 17 (Figure 5-1), that actually displays a seasonal component, though with a clear difference between weekdays and weekend days, that is not distinguished by a seasonal component with a daily period. As a result, the seasonal component fits nor weekdays nor weekend days very well, the trend component exhibits a weekly pattern and the residuals are rather high. Case 12 (Figure 5-1) shows a more irregular energy use and is the example with the highest random component.

Finally, case 8 is the exception in this sample, that has a very strong trend component that explains 94% of the variance in the data. As seen in Figure 5-6, the remaining variance around the strong trend component is not explained by the seasonal component (that is negligible in magnitude), but is part of the residual component.

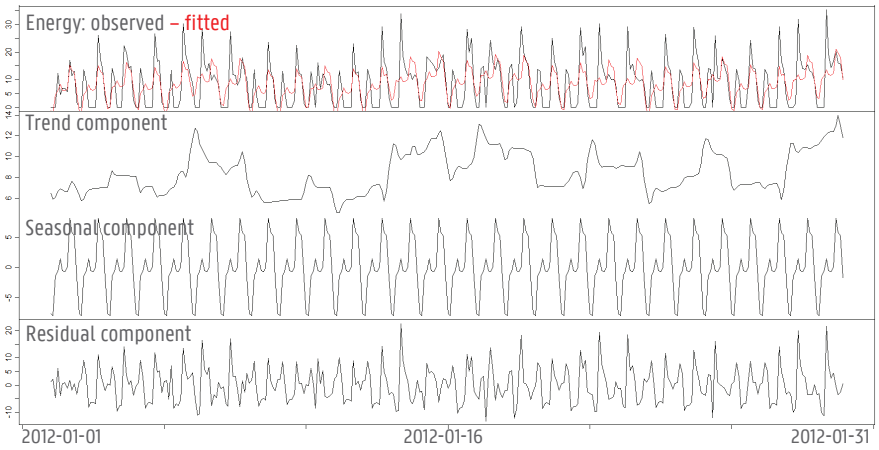


Figure 5-5: Time Series Decomposition of 2-hourly time series – Additive – case 17

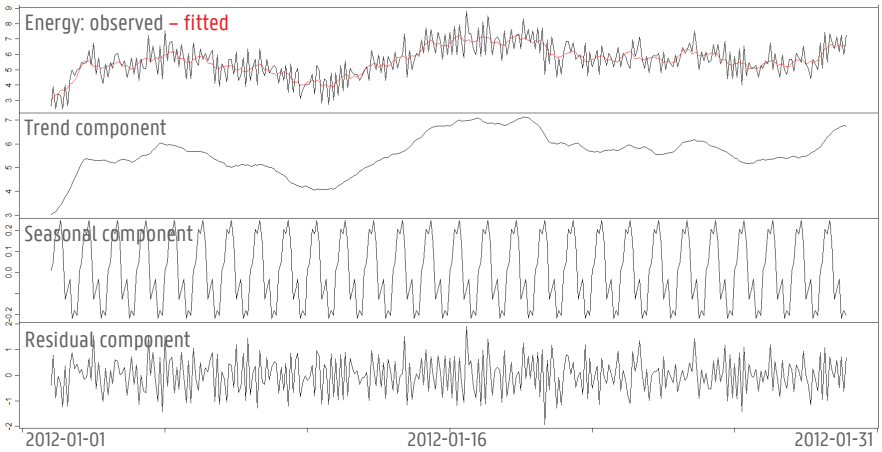


Figure 5-6: Time Series Decomposition of 2-hourly time series – Additive - case 8

When observing the residual component for case 18 (Figure 5-3) more in detail, it is seen that the magnitude of the residuals decreases and increases as the level of the trend component decreases and increases. This phenomenon gives away a disadvantage of the additive decomposition method in the analysis of the energy use time series. In many of them, the magnitude of the seasonal component increases (or decreases) when the trend increases (or decreases). In the example case, the energy use is always zero at night, independent of the outdoor conditions, but the energy use during the day does vary. By simply adding the seasonal and trend component, this variation in magnitude is not taken into account. As described by Hyndman and Athanasopoulos, the additive model is most appropriate if the magnitude of the seasonal fluctuations does not vary with the level of the time series [91]. When it does vary with the level of the time series, one can transform the data to make the variation stable over time (e.g. using a logarithmic transformation), or apply a multiplicative model.

Multiplicative Decomposition by Moving Averages

Multiplicative seasonal decomposition by moving average is an alternative to the additive decomposition, that is mentioned by the same authors, and is also available in the R environment through the *decompose* function [69]. The method is similar, except that the components are multiplied instead of being added:

$$Q_t = TC_t \times SC_t \times RC_t \qquad [5\text{-IV}]$$

For making the interpretation similar to the previous section, in the following example the residuals are also constructed in an additive way, and equation 5-IV transforms to:

$$Q_t = TC_t \times SC_t + RC_t \qquad [5\text{-V}]$$

Figure 5-7 presents the decomposed time series for case 18, with an additive random component. The residuals are smaller and the variance is more constant than in the additive decomposition. What is left in the residuals is the smaller variations against the seasonal component, that occur during the course of the day.

The variances are then calculated as follows for the sum of fitted values and residuals:

$$Var(Q) = Var(SC \times TC) + Var(RC) + 2 \times Cov(SC \times TC, RC) \qquad [5\text{-VI}]$$

The trend component is equal to the trend component in the additive decomposition, and so is its variance. In Figure 5-8 the difference in variance between the fitted values (that is the combination of the estimated trend and seasonal component) and the trend component, is attributed to the seasonal component. In this representation, the variance explained by the seasonal, trend and residuals make up a 100% of the total variance, and the covariance between fitted values and residuals is explained as a percentage of the total variance and has a negative sign.

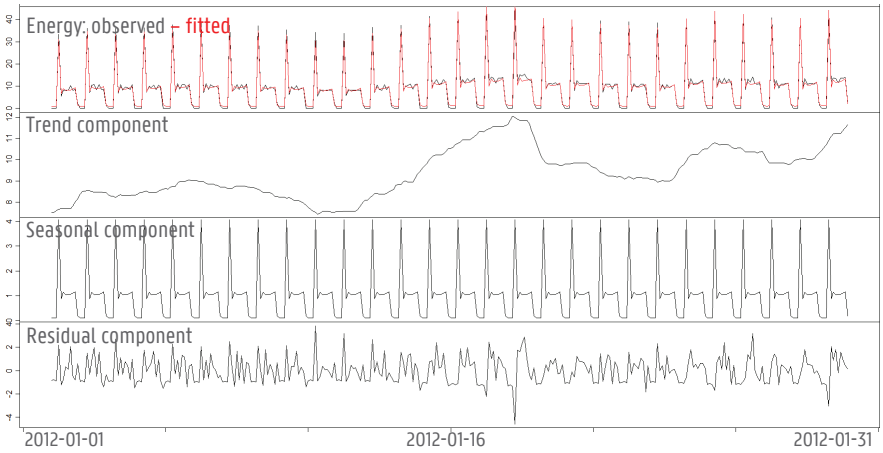


Figure 5-7: Multiplicative TSD with added RC – 2-hourly data - case 18

The covariance between the fitted values and residuals is no longer negligible in many cases. One reason for this is the variation in the residuals during the course of a day. In cases with night set-back, the seasonal component is zero at night, which perfectly fits the data, while during the day, the actual variations in energy use vary more in relation to the seasonal component. For understanding a second reason, observe the time series components for case 18 for the entire heating season in Figure 5-9. It seems that in the beginning of the heating season, and in the coldest winter period, the daily energy use time patterns differ from the rest of the heating season. This aspect is not reflected in the decomposition, where a fixed seasonal component is constructed. Then due to the multiplication, this effect is significantly enlarged, especially for the coldest winter period, where the residuals are largest in magnitude and have a negative sign when the fitted values are close to zero. The covariance thus actually gives away the shortcomings of the seasonal component estimation.

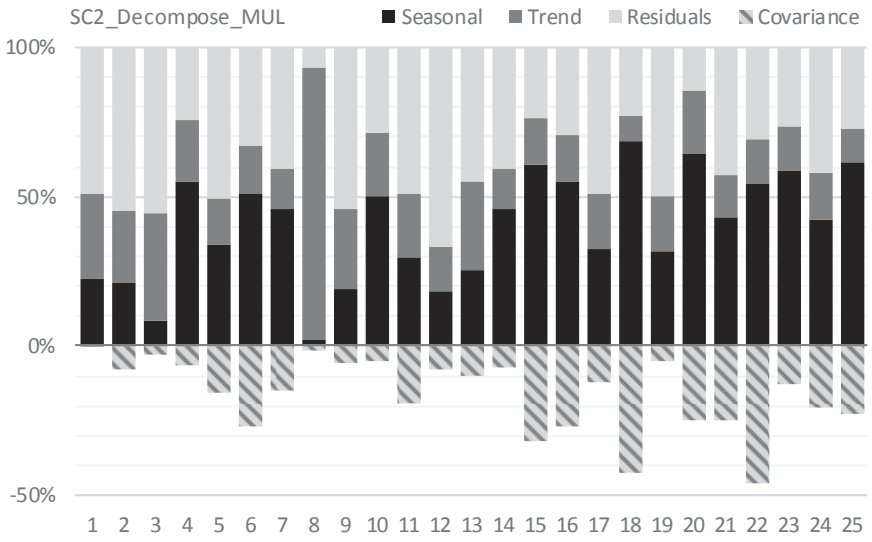


Figure 5-8: Multiplicative TSD – Daily Period – Explained / Observed Variance (%)

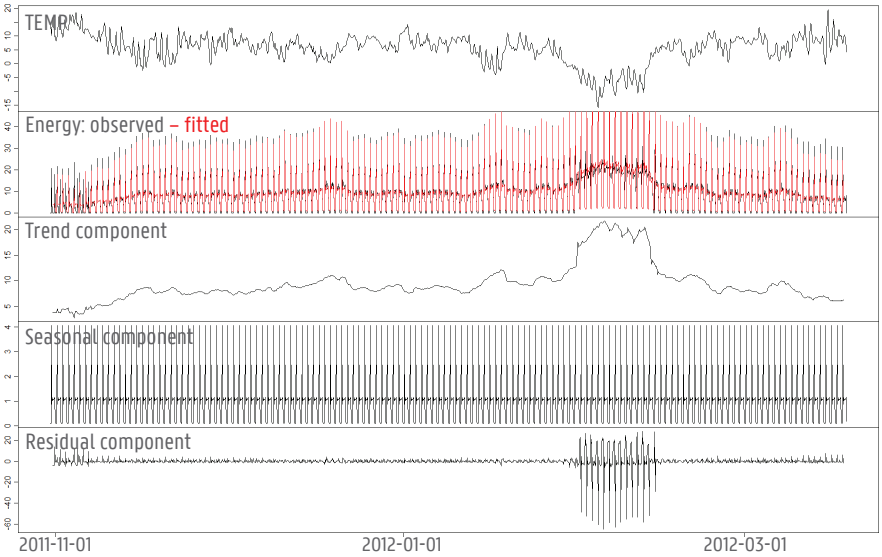


Figure 5-9: Multiplicative TSD with added RC – 2-hourly data – case 18 – entire heating season

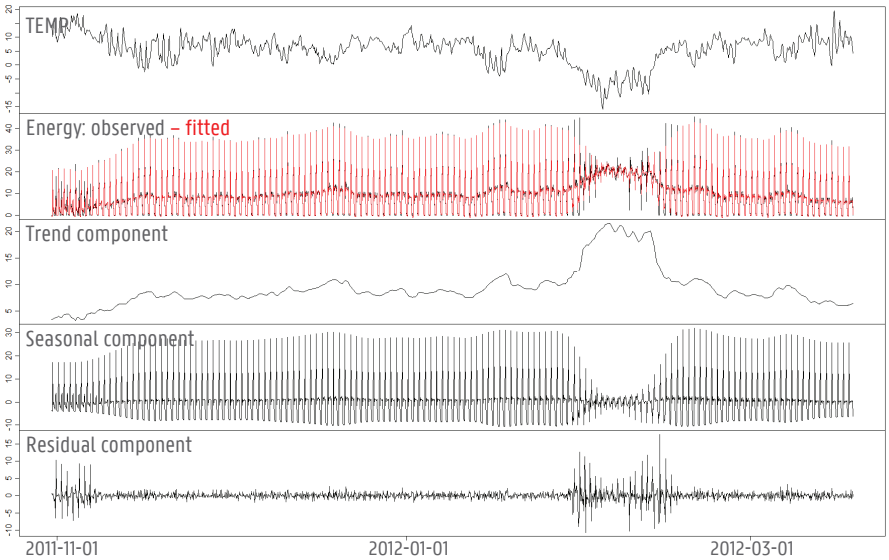


Figure 5-10: STL decomposition – 2-hourly data - case 18 – entire heating season

Decomposition based on Local regression (STL)

A third common time series decomposition method offers the possibility of estimating time-varying seasonal components. The method is called Seasonal-Trend decomposition by LOESS (STL) and was developed by Cleveland et al. The seasonal component is first defined by local regression (LOESS) smoothing of the seasonal sub-series (instead of taking the mean) and the trend is defined by smoothing the deseasonalised series, after which the overall level of the seasonal component is added to the trend component. The procedure is iterative and for a full explanation, the reader is referred to Cleveland et al. [81]. In Figure 5-10 the STL decomposition of the heating season time series for case 18 is presented. Remark that the seasonal component slowly varies and takes up differences observed in the beginning and coldest period of the heating season. It is an additive type of composition, so the trend component only takes up the change in magnitude of the seasonal component around its mean. The change in magnitude near the outer points of the seasonal component, is here included in the time-varying seasonal component itself. The variances are calculated according to eq. 5-VII and presented in Figure 5-11. Remark that the covariance is not negligible, but smaller than for the multiplicative decomposition. It is positively correlated: the higher the fit, the higher the possible residual.

$$Var(Q) = Var(SC) + Var(TC) + Var(RC) + 2\times Cov(SC+TC, RC) \quad [5\text{-VII}]$$

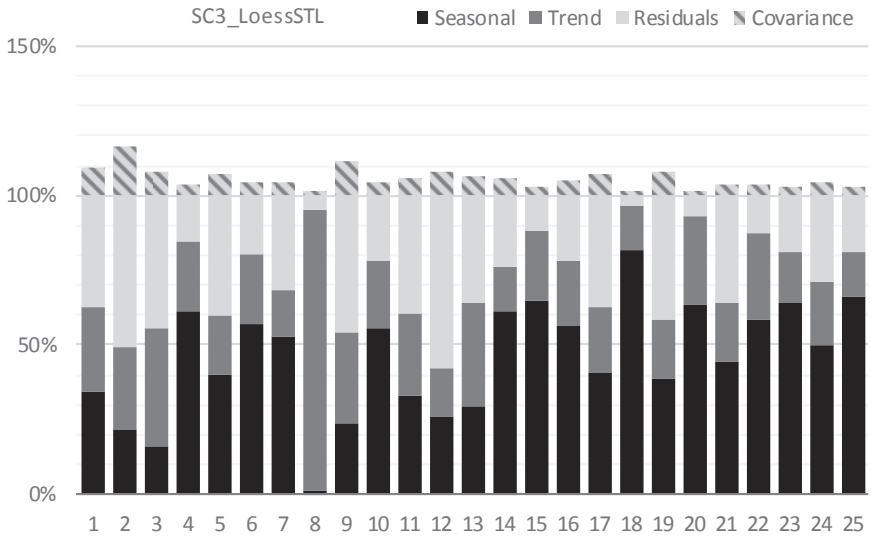


Figure 5-11: STL TSD – Daily Period – Explained / Observed Variance (%)

Comparison

For most cases, the seasonal component explains a significant portion of the variance in the data, and is considered a useful component. About 50% to 75% of the variance explained by the fitted values, is explained by the seasonal component, and the remainder by the trend component. Exceptions are a case where no daily pattern could be distinguished (case 8) or where there are large variations in patterns from day to day (e.g. case 3, case 12).

In Figure 5-12 the variance explained by the combination of seasonal and trend components is presented for the three decomposition methods discussed, when applied to the heating season time series. The co-variance is not displayed. It is found that for all cases the variance explained using STL TSD is the highest, with a minimum of 42% and average of 72%. For the majority of cases, the variance explained is also higher for the multiplicative TSD (average 62%), than for the additive TSD (average 58%). The main reason that the multiplicative TSD fits the data better, is that through multiplication of season and trend components, the magnitude of the seasonal fluctuations varies with the level of the trend, which is a characteristic of the time series considered. The main reason for the higher proportion of variance explained when using STL, is that the seasonal component is time-varying and thus captures variations in seasonal component throughout the heating season.

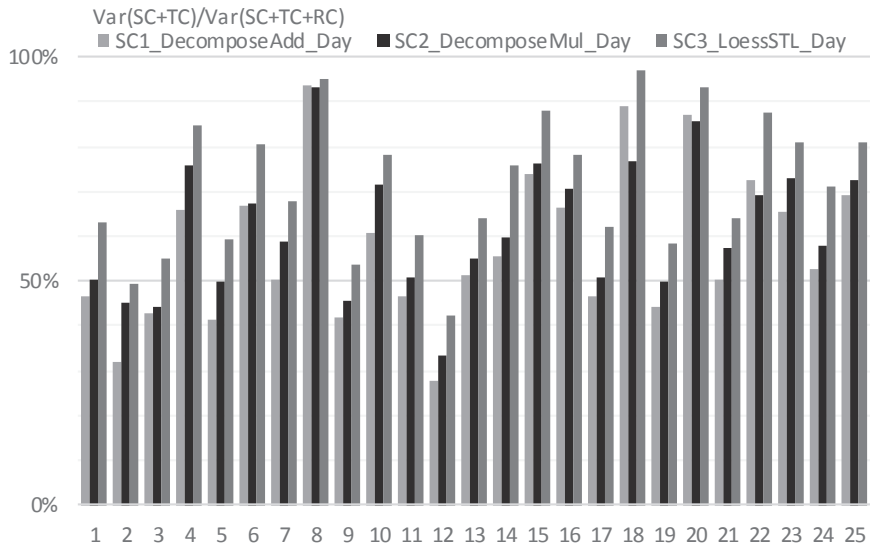


Figure 5-12: TSD Methods – Daily Period – Explained / Observed Variance (%)

5.2.3 Time step

So far in this dissertation, when sub-daily time steps were considered, the data were aggregated to 2-hourly time steps, in order to soften the influence of small time shifts in the energy use in the diurnal variations. In this section, hourly energy use time series are decomposed, and the results are compared to those with 2-hourly data. In Figure 5-13, the difference in variance explained between hourly and 2-hourly data is presented for the different components. It is calculated as $1- Var_{2\text{-hourly}}/Var_{1\text{-hourly}}$ for each part of the decomposed elements (seasonal, trend and residuals). Negative values indicate that the variance explained is lower for hourly data than for 2-hourly data, and vice versa for positive values. The results are presented for the multiplicative TSD, but the conclusions are similar for the other two TSD-methods considered.

First it must be mentioned that when hourly data are used, the overall variance in the energy use time series is higher, because the data are less aggregated. It is found that when using hourly data, the percentage of the observed variance that is explained by the seasonal and trend components is typically lower for hourly data than for 2-hourly data, and the variance explained by the residuals is thus higher. Therefore, in the remainder of this dissertation, 2-hourly data will be used.

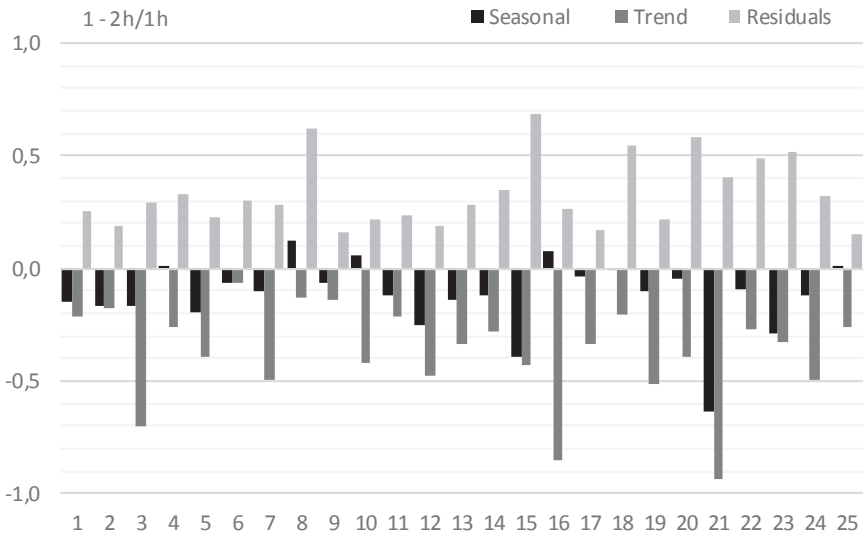


Figure 5-13: Multiplicative TSD – 2-hourly versus hourly data – Difference in Variance Explained/Observed per component

5.2.4 Period of the Components

Following the observation of diurnal patterns in the data, a period of one day for the seasonal and trend component was selected in the previous sections. However, as illustrated in Figure 5-1, the diurnal patterns can be different for different days of the week (e.g. case 17), and, as illustrated in Figure 5-9, they can be different for different periods of the heating season. For taking into account the variations in diurnal patterns for different days of the week when decomposing the series, the use of a period of one week instead of one day, is a possible solution. Figure 5-15 presents the portions of the variance in the data explained by the different components for a multiplicative TSD. In Figure 5-16, the variances for a one day and one week period are compared. It is found that when a weekly period is used, the portion of the variance explained by the seasonal component is mostly similar or higher than when a daily period is used (positive values in Figure 5-16). For example for case 17 (see Figure 5-14), that displays a clear weekly pattern, this is noticed. On the other hand, the variance explained by the trend component is typically lower: the trend is more smooth than with a daily period, so smaller variations in the trend are not captured. Finally, the variance explained by the residuals is often similar or higher than for the daily period. This comparison thus merely illustrates the importance of diversifying the seasonal component.

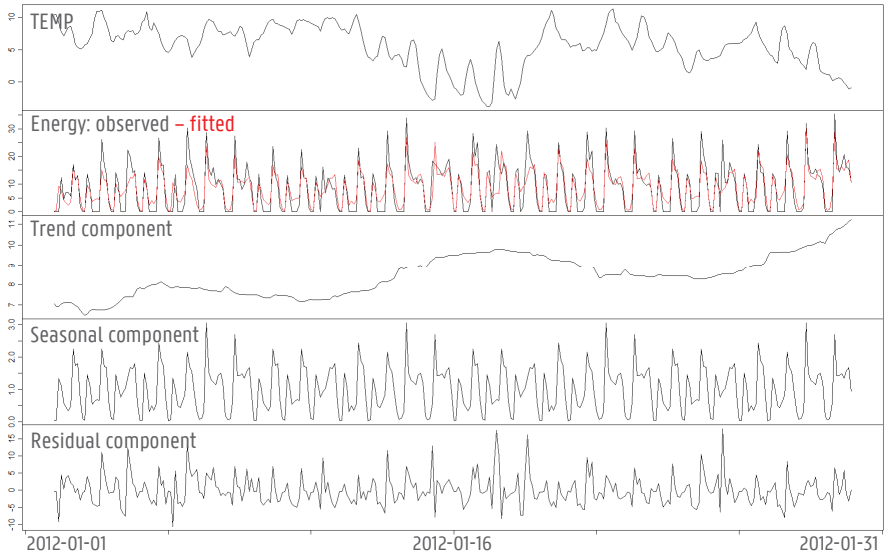


Figure 5-14: Multiplicative TSD – 2-hourly data - case 17

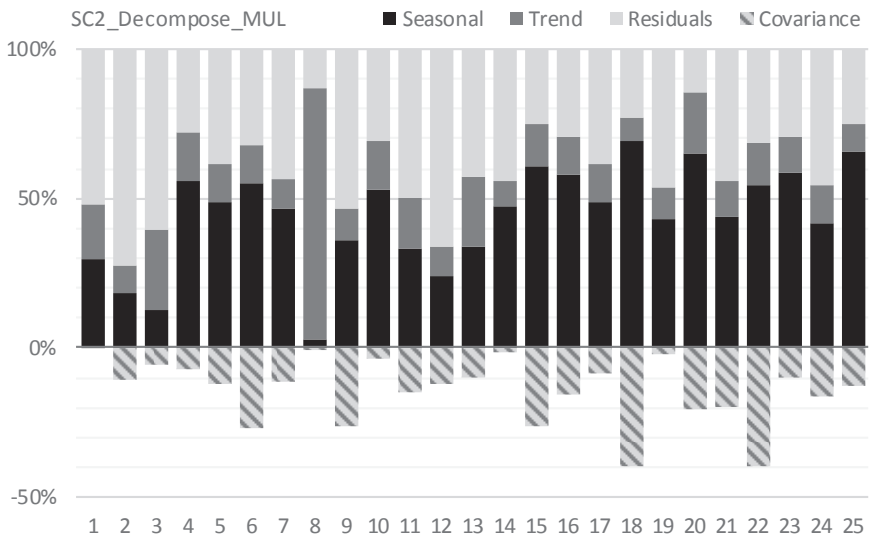


Figure 5-15: Multiplicative TSD – Weekly Period – Explained / Observed Variance (%)

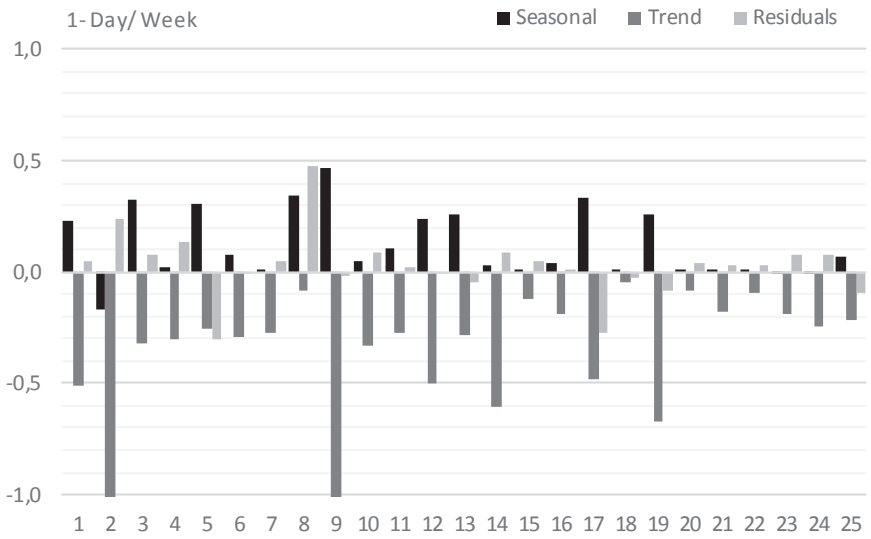


Figure 5-16: Multiplicative TSD – daily versus weekly period – Difference in Variance Explained/Observed per component

5.2.5 Findings

The decomposition of the energy use time series in seasonal, trend and residual components supports the understanding of the time series and is found a useful approach for exploring and analysing them. The level of the trend component tends to change with the (smoothed) weather conditions such as the exterior temperature (e.g. Figure 5-9 and Figure 5-14). Since the data is smoothed for period of one day (or one week), it contains information similar to daily (or weekly) aggregated values, that were used in the linear regression analysis in previous chapters, and where indeed the relationship with weather variables was modelled. When now comparing the daily and weekly smoothed trend components for case 17 in Figure 5-5 and Figure 5-14, it is noticed that this relationship with the weather is not seen for the daily smoothed values, but becomes visible for the weekly smoothed values, because of the weekly patterns in the data. This illustrates why in the linear regression analysis on daily data, the coefficient of determination for case 17 was among the lowest obtained.

From the different time series decomposition methods applied, it is confirmed that the energy use time series that are subject of this study are better described using multiplicative composition of the components. The magnitude of the diurnal patterns varies with the level of the trend component. This can be understood from physical point of view: on colder days (when the level of the trend component is higher), more power will be needed to heat the house, not only on average, but also during the heating hours. On the other hand, when the heating is turned off (e.g. night set-back), the energy use always falls back to a minimum of zero, without variations. Therefore the sub-daily energy use varies between a fixed value of zero and a maximum heat use that varies with the trend.

The fixed seasonal component obtained in the additive or multiplicative decomposition roughly characterises the energy use time pattern that occurs in the data. The time-varying seasonal component obtained using the STL method, confirms the presence of different types of energy use time series patterns in different stages of the heating season. In case 18, Figure 5-10, next to the dominating pattern, at least two different patterns occur: one at the beginning and end of the heating season (the warmer periods) and another in the coldest winter period. The application of multiplicative TSD with a weekly period illustrates that in a number of cases, the diurnal patterns also vary for different days of the week, and are thus characterised by a weekly returning pattern (e.g. Figure 5-14). Besides, when zooming out from the selected period of the heating season towards a full year of measurement data, e.g. Figure 5-17, also summer the summer season displays a different pattern than the rest of the year. In the considered cases with only space heating energy use, this summer pattern is often zero, but thinking of combined gas use data for space heating and domestic hot water, a different pattern can be expected in summer, corresponding to the domestic hot water use.

From these observations, it is concluded that a variation in energy use time patterns occurs over the course of the year, and that a part of this variation can be explained by calendar variables (e.g. the day of the week) and/or by weather variables. The aforementioned time series decomposition methods allow to identify regular weekly patterns, and to visually discern time-varying patterns, but not to identify irregular patterns (e.g. during holidays, due to behavioural changes…) and to mathematically describe the relationship between the energy use time pattern and the weather or calendar variables. This is the goal of section 5.3.

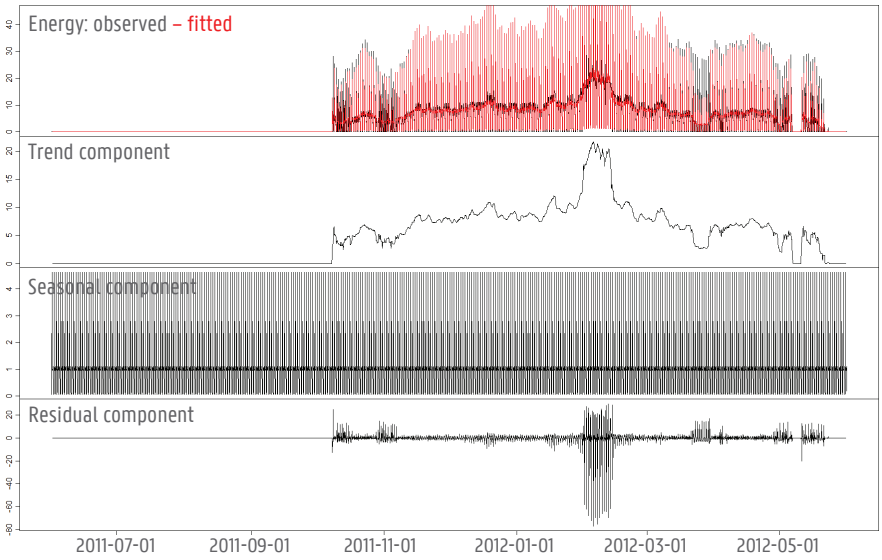


Figure 5-17: Multiplicative TSD – 2-hourly data - case 18 – full year

Recently, more advanced time series decomposition methods have emerged to move beyond the limitations of the traditional TSD methods, for describing more complex seasonal patterns. De Livera et al. propose the TBATS[10] method for modelling multiple seasonality (e.g. a combination of daily and weekly patterns), non-integer and dual-seasonality (e.g. dual calendar effects, moving holidays) using a state space approach [86,87,92]. Dokumentov and Hyndman propose a seasonal trend decomposition based on regression, that also allows to take into account seasonality with a complex topology and allows regressors to be taken into account, possibly with time-varying or seasonal coefficients [93]. However, also for these more complex methods to be applicable, one needs to assume the structure of patterns that occur (e.g. to define the complex topology of successive patterns), or to identify which days do have similar patterns.

Finally, because in the following sections the energy use time series of an entire year will also be observed, instead of the heating season data, the results for the multiplicative time series decomposition of the entire year time series are reported in Figure 5-18. In comparison to the heating season data set (Figure 5-15), the variance explained by the trend component and by the fitted model (SC+TC) is higher when the entire year is concerned, while the variance explained by the seasonal component is lower. This is explained by the summer period, where the model fit is maintained by the trend component that is constantly zero. The variance explained by the fitted model is between 43% and 97% of the variance in the data, and between 1% and 58% of the total variance is explained by the seasonal component.

[10] TBATS is an acronym for the key features of the model: Trigonometric representation of the seasonal components based on Fourier series, Box-Cox transformation, ARMA errors, Trend and Seasonal components.

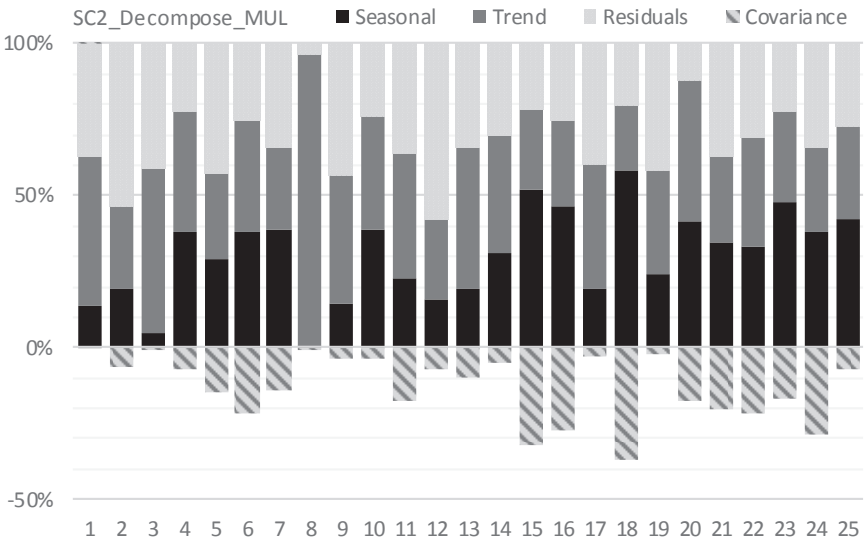


Figure 5-18: Multiplicative TSD – Daily Period – Full Year - Explained / Observed Variance

5.3 Clustering energy use time patterns

The decomposition of the 2-hourly time series of the energy use demonstrated that a variation in energy use time patterns occurs throughout the year, and that part of this variation is related to variations in calendar and/or weather variables. The aim of this section is to mathematically identify groups with similar energy use time patterns that appear in the energy use time series of specific houses, by use of cluster analysis techniques. This statistical approach is introduced in section 5.3.1. The important step of pre-processing of the energy use time series for clustering purposes is described in section 5.3.2 and the cluster analysis and results are found in section 5.3.3.

5.3.1 Cluster Analysis

"*Cluster analysis is the art of finding groups in data*", as pointed by Kaufman & Rousseeuw [94]. It comprises the grouping of a collection of objects into subsets or clusters, such that the objects within a cluster are more similar to each other than those found in different clusters [95]. Since it is unknown to which subset an object belongs (the data are 'unlabelled', the correct answers are unknown), clustering is a pattern recognition technique that is based on unsupervised learning, or 'learning without a teacher'. As such, it is opposed to supervised learning approaches, such as regression or classification, where a system is trained using labelled training data. An important consequence of the lack of labelled data, is that there is no direct measure of success for evaluating the quality of the results. Therefore, according to Hastie et al., one must resort to heuristic arguments for motivating both the algorithms and the results of a cluster analysis [95].

Central to all clustering methods is the notion of a degree of (dis)similarity between the individual objects that are clustered. If the objects are described by variables that are all quantities (quantitative or interval-scaled variables), then a familiar way of estimating the similarity, is by calculating the Euclidean distance between the different variables. By use of such a metric or distance measure, the distance between each pair of objects in the data is calculated and stored in a 'dissimilarity' matrix that counts as many rows and columns as there are objects in the data, and that is the input for the clustering process. The pre-processing of the data and the distance measure used in this study, are discussed in section 5.3.2.

Kaufman & Rousseeuw describe two kinds of algorithms that represent the majority of data clustering methods available, and for each kind they discuss a few methods that cover the majority of applications. The first group of clustering algorithms is called the partitioning methods. They divide the data objects into a fixed amount of k clusters in the 'best' possible way, where the amount of clusters is imposed by the user. Each cluster possesses a centre. One of the most popular partitioning methods is the k-means method, that can be applied to interval-scaled variables, and where the average squared Euclidean distance to the cluster means or centroids is minimised [95]. As a result of using the squared distance, this method is less robust to outliers. A similar method that is more robust, and is also applicable on a dissimilarity matrix (variables do not need to be interval-scaled), is the k-medoids method, or Partitioning Around Medoids (PAM). The method tries to find k representative objects or medoids that are central in their clusters, by minimising the average distance to the medoid. Apart from the PAM-method, Kaufman & Rousseeuw discuss a variant for Clustering LArge Applications (CLARA), and a

partitioning method based on the fuzziness principle (FANNY), where the objects can be assigned to different clusters by a degree of membership in order to avoid the 'hard' clustering of objects [94]. A second group of clustering algorithms are the hierarchical methods. The primary output of a hierarchical method is a dendrogram or tree structure that organises all the objects in a hierarchical way. The clustering can be performed through a step-by-step merging of objects into clusters, until all objects are incorporated, or vice versa, through a step-by-step separation of the whole cluster of objects into smaller clusters, until there are as many clusters as there are objects. In Kaufman & Rousseeuw, a hierarchical method that is applicable to ordinal-scaled variables is called AGglomerative NESting (AGNES) and an example of the 2nd type is the DIvisive ANAlysis method (DIANA) [94]. Through the tree structure, hierarchical methods provide clustering results for all possible amounts of clusters obtainable with the provided data, in contrast to the partitioning methods, that provide precisely *k* clusters. However, the clustering results for *k* clusters obtained by a hierarchical method, are not necessarily the best way of clustering into *k* clusters, because of the inability of hierarchical techniques to correct for erroneous decisions in previous steps of the merging or dividing process [94]. In contrast, partitioning methods are trying to find the 'best' possible clustering with *k* clusters. In this study, the k-medoids method (PAM) will be employed, which also got good results in the study of Paparrizos, who claim that the method is important choice [96].

Cluster Validity Indices & Number of Clusters

Unless the goal of the clustering is to unveil some hierarchical structure in the data, the problem of deciding the amount of clusters *k* remains for both hierarchical and partitioning methods. For hierarchical methods, there is the search for the amount of clusters for which a certain level of 'similarity' is obtained, for which many different approaches, criteria or rules exist. For the partitioning methods, the clustering analysis can be repeated for different values of *k*, and there is the search for which value of *k* a certain clustering criterion is optimal. The selection of the amount of clusters is therefore part of the general task of cluster performance evaluation, more specifically the internal cluster validation, or assessment of the cluster purity. Internal validation is opposed to external validation, where the clustering would be compared to the 'true' clustering of the data. But, as was mentioned before, clustering is an unsupervised procedure and the 'true' clustering of the data is not known.

Table 5-1: Cluster Validity Indices (Source: Arbelaitz et al. [97])

CVI	*Cluster Cohesion*	*Cluster Separation*	*Combination*
Average Silhouette Width	Distance from the clusters' points to all other points in the same cluster	Distance from the clusters' points to all other points in the nearest neighbour cluster	Summation
Calinski-Harabasz	Distance from the clusters' points to its centre	Distance of the centres to the global centre	Ratio
Davies-Bouldin	Distance from the clusters' points to its centre	Distance between the centres	Ratio

Many Cluster Validity Indices (CVIs) have been proposed and their suitability often depends on various characteristics, such as the type of data, assumptions on the cluster structure etc. [98]. The study form Milligan & Cooper in 1985 is considered a reference work on internal CVIs, and

was extended by Arbelaitz et al. in 2013 [97,99]. In their comparison of 30 CVIs, they found that some CVIs perform significantly better than others, and the Silhouette index generally performs best, followed by the Davies-Bouldin and Calinski-Harabasz indices. All three of them are based on a combination (summation or ratio) of an estimate of the cohesion of the cluster and an estimate of the separation of the cluster(s). The calculation principles are summarised in Table 5-1. For calculating the distances, the same distance metric is to be used as in the clustering algorithm itself. For example the Calinski-Harabasz, also called a variance ratio criterion, and the Davies-Bouldin index use the Euclidean metric, and therefore they need to be adapted when other metrics are used. If the clustering data is noisy, the study of Arbelaitz shows that the Silhouette width is less sensitive to that than the other indices mentioned. While the Calinski-Harabasz and Davies-Bouldin compare the distances of the objects to the centres, and come up with one central value that can be used for evaluating the overall clustering quality, a Silhouette width does not relate to the cluster centre and it is available for each single object ànd as a general validity index for evaluating the overall clustering validity. It is applicable for partitioning as well as hierarchical methods. It also has the benefit of being standardised between -1 and +1, and so the CVIs for different data-sets can be placed next to each other. For these reasons, the Silhouette index will be applied in this study, for assessing the number of clusters in the partitioning, the overall cluster validity and how strong individual objects belong to their clusters.

The Silhouette was proposed and documented by Rousseeuw [94,100]. For each object the average dissimilarity *a(i)* between the object and all other objects in its own cluster is compared to the dissimilarity *b(i)* between the object and those in the nearest cluster. The Silhouette or Silhouette width *s(i)* is then:

$$s(i) = \frac{b(i) - a(i)}{\max(\{a(i), b(i)\}}$$

Silhouettes close to +1 indicate that the object is well-clustered in its cluster, while values of -1 indicate that the object lies closer to the neighbouring cluster and would be misclassified. Values around zero indicate that it is not very clear to which of the two clusters the object belongs and it is considered an 'intermediate case'. The results of the individual objects can be summarised in a Silhouette plot, and they can be averaged per cluster to evaluate how 'well-clustered' the different clusters are. Finally, the average of all Silhouettes of the dataset is called the Average Silhouette Width or the Silhouette Coefficient (ASW or SC) and allows validating the overall cluster validity and, by comparing for different values of *k*, the amount of clusters. From their experience with the index, Kaufman & Rouseeuw propose a subjective interpretation for the Silhouette Coefficient, which is incorporated in Table 5-2.

Table 5-2: Subjective Interpretation of the Silhouette Coefficient (Source: Kaufman & Rousseeuw [94])

ASW	Proposed Interpretation
0.71 – 1.00	A strong structure has been found
0.51 – 0.70	A reasonable structure has been found
0.26 – 0.50	The structure is weak and could be artificial
< 0.26	No substantial structure has been found

5.3.2 Data Pre-processing

The 2-hourly time series data for an entire year are cut into time series with length of 1 day, or sequences of 12 values. They are organised in a data frame with 366 rows (amount of days) and 12 columns, which is called the original data frame. In this section, three pre-processing steps are explained: the selection of the rows (days) for clustering, the transformation of the data and the calculation of distance measures. Then also the cluster validity measures are discussed, which will be used to select the optimal number of clusters when applying the Partitioning Around Medoids clustering method. Finally, the relevant pre-processing options are applied while clustering and the results are discussed.

Data Selection

Within the large variety of energy use patterns, two typical patterns can easily be separated from the dataset in a deterministic way, and can easily be understood in a physical way. A first pattern is called the zero-pattern, occurring on days that no energy is used at all, so the energy use at all time steps is zero. Since zero is also the minimum energy use occurring, the deterministic condition is simply that the sum of all energy uses over the day equals zero. The second pattern is called the standby-pattern and consists of a continuous very low energy use that remains for consecutive days or even weeks. Based on the observations of the 2-hourly energy use in the different cases, it is determined in two steps. First all days are selected for which the maximum energy use is higher than 0 and smaller than 1 kWh/2hours ("small energy use"), ànd the difference between the maximum and minimum value of the day is below 0.2 kWh/2h ("continuous"). In a second step, the days are selected for which the differences in daily average energy use for consecutive days are smaller than 1 kWh/2h and it is counted for how many consecutive days this condition is true.

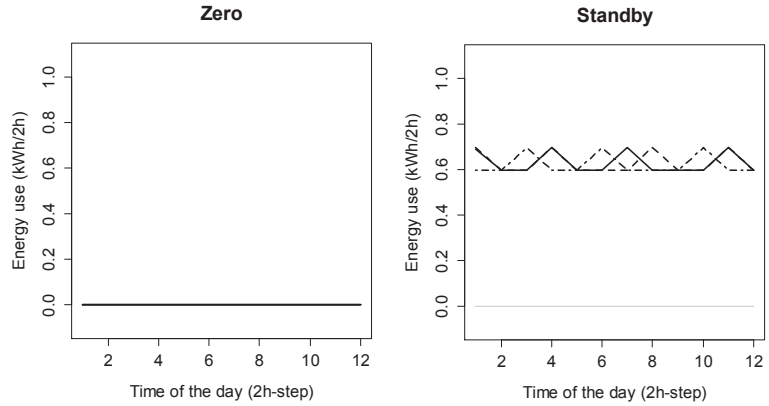


Figure 5-19: Zero-Pattern & Standby-Pattern

Amongst the 25 cases in the space-heating data sub-sample, the zero-pattern appears between 101 and 216 days of the studied year. It typically occurs at higher exterior temperatures, during spring, summer and autumn, when there is no demand for space heating. It occurs occasionally during winter time, probably because people are on holiday, or the heating system is broken. A few exceptions are case 2 (the house where gas is not used frequently) where it appears for 286 days per year, case 9 where it appears for only 15 days and case 25 where it appears for 75 days.

The so-called standby-pattern occurs in 7 cases for 3 up to 56 days. It occurs in spring and/or autumn at the verge between the heating and summer season. It is understood as the standby-losses of the space heating system, which is usually completely shut down during summer. However, such a standby-pattern could also point at a gas leak or deficiency with the gas meter. For case 25, the Zero-, Standby- and InUse-patterns are illustrated in Figure 5-20, and their occurrence is plotted in function of the equivalent exterior temperature, solar radiation and time. For this case, the standby-pattern occurs for 56 days of the year, and the energy use is zero for only few days in summer. In total, the days allocated to the Standby-pattern are responsible for about 162kWh of gas used during one year. For case 9, in addition to the 15 days of zero energy use, only 3 days of Standby are detected. For the rest of the Spring, Summer and Autumn, small energy uses between 0 and 1.2kWh/2hours are detected, that are not continuous (they are zero at night and vary during the day), so most probably they are related to the use of gas for an actual energy use function, such as domestic hot water use or underfloor heating.

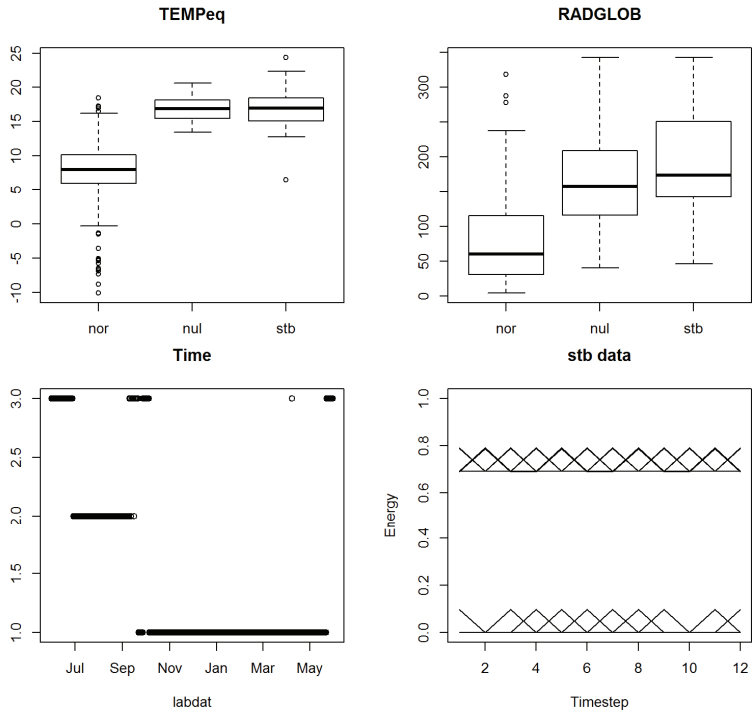


Figure 5-20: Zero-pattern (nul), Standby-Pattern (stb), In-Use Patterns (nor): Case 25

For the data in this case-study, it is not found necessary to place extra conditions on the definition of the Standby-pattern. However, in cases where small gas uses can be expected during summer, e.g. for cooking or domestic hot water use, which might require a storage tank, the conditions should be reconsidered. Since the zero- and standby-patterns are rather common and can be identified in a deterministic way, the days at which they occur are excluded from the dataset that is subject to the clustering algorithm. Note that the boundary conditions are set based on a rather limited amount of cases using gas only for space heating. They could be further checked or refined on larger data-sets.

Transformation / Standardisation

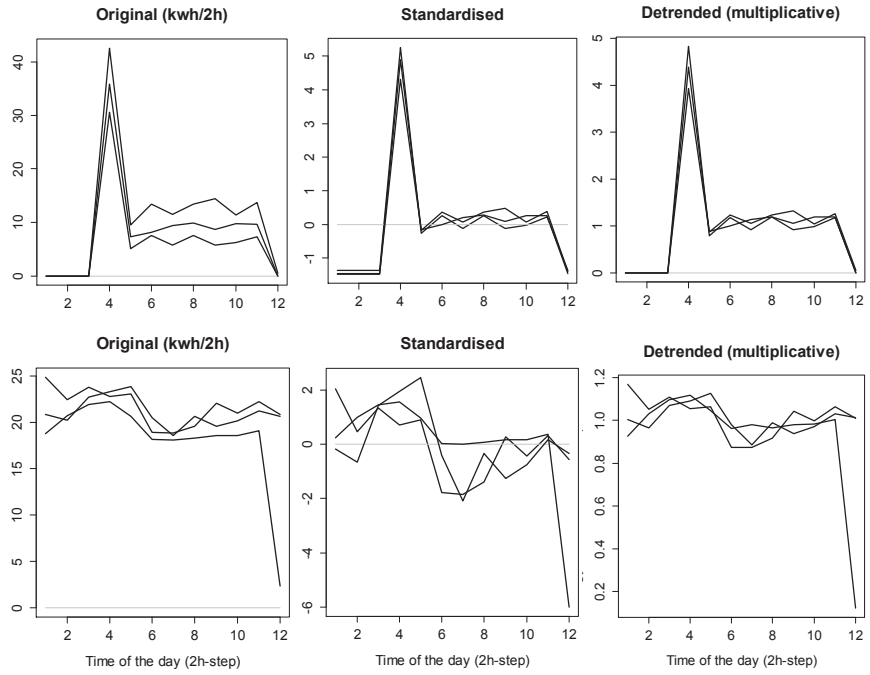


Figure 5-21: Transformation of the energy use series: example series (case 15)

Since the goal of the clustering analysis is to group energy use time series having *similar shapes* over the course of the day, rather than having similar absolute magnitudes, a transformation of the original data is found useful. Take for example the energy use time series in Figure 5-21 (left). The three original series have a very similar shape, but they differ in magnitude. During the clustering process, series having similar shape but different magnitude could be separated. As a result, multiple clusters with similar shapes emerge. A common solution is to standardise the data, by centering the data points around the mean and scaling them using the standard deviation [98]. However, Kaufman & Rousseeuw propose the use of the mean absolute deviation instead of the standard deviation as a measure of dispersion, because it is less sensitive to the effect of large outliers [94]. Remark that in the aforementioned literature sources, standardisation is applied to the different types of variables measured (i.e. the columns of the data table), while in this study there is only one variable type (energy use measurements) and each of the objects is standardised (i.e. the rows of the data table). This way, from each observation the daily mean is subtracted and then it is divided by the mean absolute deviation for that day. As a result of the standardisation, the data with similar shapes are now closer grouped (Figure 5-21, top middle), but remark that the data-points with zero-values tend to come separated a little, and the shape of series with a rather constant energy use during the day, come very close to series with absolute zero values (bottom middle). In the type of data that is studied here, the zero-value is meaningful and stands for the very important distinction between using and not using energy. Days with a total energy use of zero were already separated in the previous pre-processing step, but still due to the standardisation zero-periods will be more easily mixed

up with other periods of constant energy use. Therefore, an alternative to the standardisation process is the use of multiplicatively detrended data (see section 5.2.2), which is similar to scaling the data around the mean (where the mean is a moving average of the concatenated series) (Figure 5-21, right). The main benefit of this action is that zero-values remain zero, and other continuous energy uses will be situated around the value of one. Both standardised and detrended energy use series will be compared in the next sections.

Distance measure or Metric

In order to define how similar or dissimilar the observed series are, the distance between each pair of series is calculated using a distance function or metric. These pairwise distances are then stored in a square matrix with the number of observed series as length, which is called the dissimilarity matrix [98]. The **Euclidean metric** is the classical metric, where the distance between two elements is the straight-line distance between two points. The distance between two time series *a* and *b*, each counting 12 time steps *t*, is then calculated by applying the formula of Pythagoras in the 12-dimensional Euclidean space:

$$\text{distance}(a,b) = \sqrt{\sum_{t=1}^{12} (a_t - b_t)^2}$$

When applying the Euclidean metric, values of corresponding time steps are compared to each other, making the measure sensitive to time shifts in the data. Imagine two energy use series that are similar in shape, but subject to a distortion in the time-axis (Figure 5-22 (a)). The Euclidean distance between the two series will be rather big, while the two patterns might actually be understood as variations of a similar main shape. Remark that when the time step of the data further decreases to an hourly time step, dealing with these time shifts will become even more important. The sensitivity of the Euclidean distance metric to distortions in the time axis has long been known in literature on time series clustering, and gave rise to the development of a technique that deals with similarity of time series that are shifted in time, called Dynamic Time Warping. This class of algorithms originated in the 1970s in the domain of speech processing, in order to deal with speaking rate variations in automatic speech recognition [101–104]. Up until today it is widely used in various disciplines, for example for aligning RNA expression data, fingerprints, electrocardiogram patterns, handwriting and signatures, music and video motion etc. [105,106].

The principle of **Dynamic Time Warping** is that two time series are stretched and/or compressed locally, so that they resemble each other. In other words, elastic or non-linear alignments between the time series are allowed [105,107]. Take again the shifted peak example, for which the DTW-mapping of one series upon the other is exemplified by the dashed grey lines in Figure 5-22 (c), and where the shifted peaks are now mapped upon each other. The alignment of two series is often graphically presented using an alignment plot with warping matrix. Each element of the warping matrix contains the alignment, which is the squared distance, between a point on one series and one on the other series. The best mapping of one series upon the other is represented by the warping path through the matrix that minimises the total cumulative distance between the two series. The path that minimises the warping cost is found using dynamic programming, and its distance is called the DTW-distance. The warping path for the

shifted peak example is presented in Figure 5-22 (d). Remark that the warping path corresponding to the Euclidean distance follows the diagonal of the warping matrix (Figure 5-22 (b)). The mathematical description and programming are extensively described in literature, for example by Itakura [101], Sakoe & Chiba [102], Rabiner & Juang [104], or in a more recent handbook by Muller [108]. A summary is found for example in the works of Keogh & Ratanamahatana [105,106]. The implementation in the R software is documented by Giorgino [107] and Sarda-Espinosa [109].

When searching for the optimal warping path, typically not all possible warping paths are considered, primarily in order to avoid 'pathological', and secondly, for speeding up the computations [106]. In Figure 5-23 (a,b) the warping path for two rather dissimilar time series displays some far-fetched steps, for example the mapping of distant points (that are up to 6 time steps (or 12 hours) away from each other) and the concentration of many points from one series upon a single point from the other series, so that a short pattern from one series is matched with a relatively long pattern from another series. During the clustering process, such far-fetched liaisons between series could result in an unrealistic grouping of the series. Therefore constraints are commonly applied as to better control the possible routes of the warping paths [106,108,109]. The constraints arise from observations of the nature of acceptable paths through the matrix, such as the monotonic condition (the path cannot go backward in time), the continuity condition (every point from both series must be used) and boundary conditions (start and end of the warping path must be maintained). The constraints originate from observations on speech data, but they are found relevant for the energy use time series as well. These conditions are by standard applied in the algorithm used in this study [107].

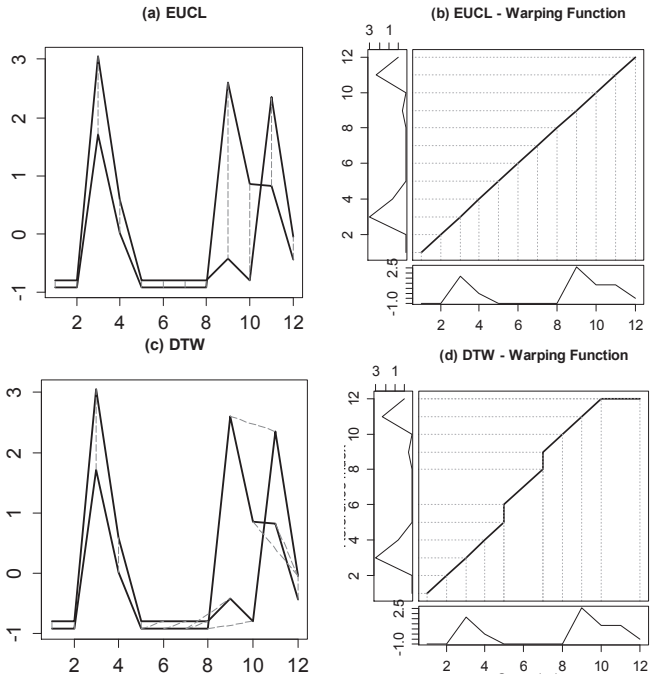


Figure 5-22: Shifted peak example (standardised) - Euclidean and DTW metric

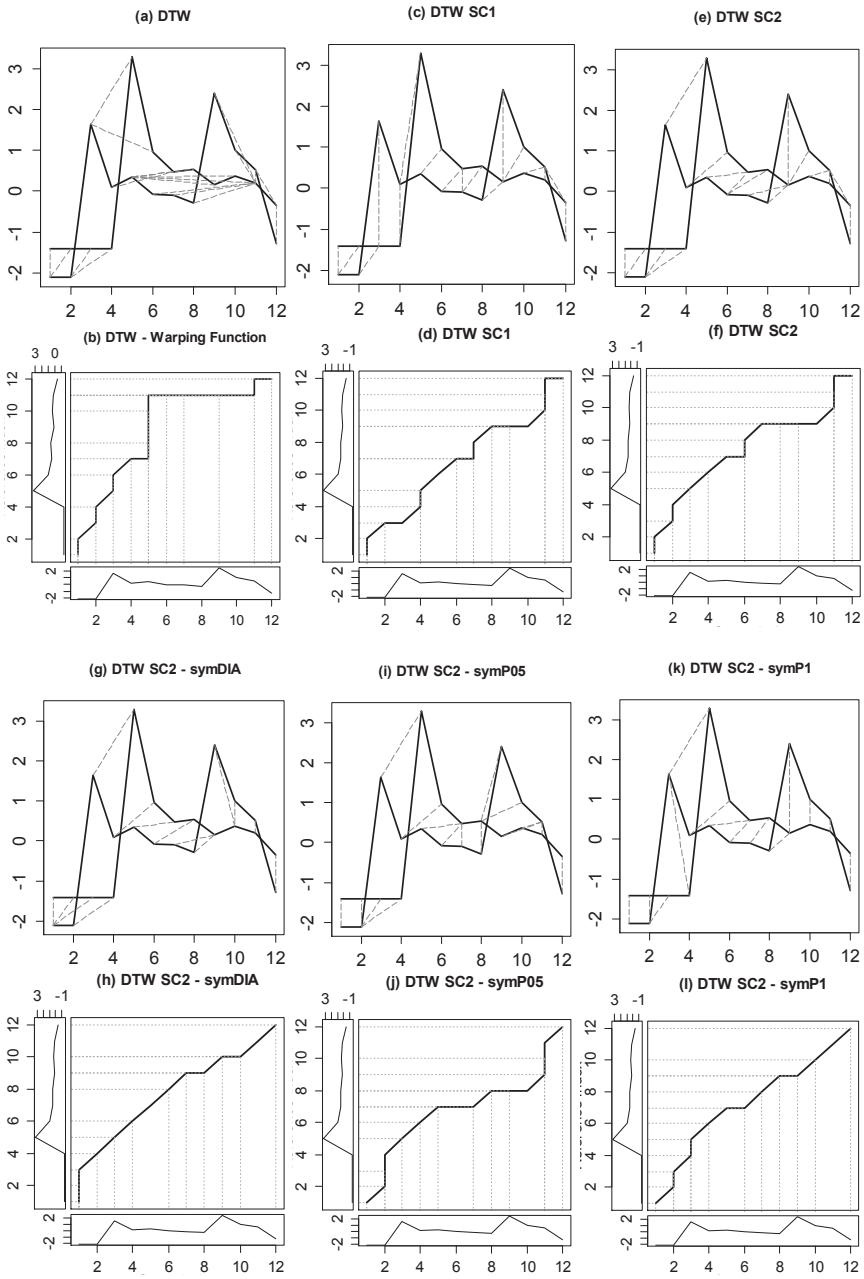


Figure 5-23: Week & weekend example (standardised) – DTW metric constraints

Then, there are global constraints or window constraints, that limit the region of the warping matrix where the warping path can enter (e.g. preventing that the path drifts very far away from the diagonal). Various window types exist, but the most common types are the Sakoe-Chiba Band, for which the allowed region is a band along the diagonal of the warping matrix, and the Itakura Parallelogram, which is a parallelogram along the diagonal of the matrix (see Figure 5-24) [101,102,106,109]. The Itakura parallelogram describes a region that constrains the slope of the warping path, so actually it is the result of local slope constraints on the path, rather than a global constraint on the matrix [108]. The Sakoe-Chiba Band is defined by its window length, that defines the width of the band around the matrix diagonal and corresponds to the maximum time deviation that is allowed in the matching of elements. Ratanamahatana and Keogh notice the customary use of a window length of 10% of the length of the series amongst scientists, however their investigations show that the classification accuracy peaks at even smaller band sizes [106]. In this study, where the series are only 12 points long, a window size smaller than 1 would prevent the warping from working effectively, so only sizes of 1 and 2 are explored, corresponding to an allowable time shift of 2 up to 4 hours between two elements. In Figure 5-23(c-f) the Sakoe-Chiba band with size 1 (c,d) and 2 (e,f) is applied to warp two rather dissimilar time series.

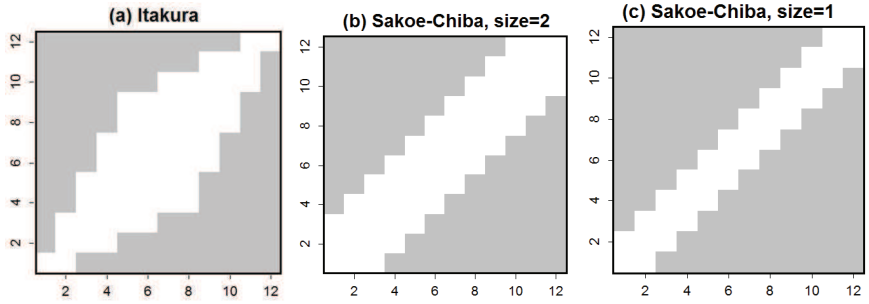


Figure 5-24: DTW Constraints: Itakura Parallelogram, Sakoe-Chiba Bands

Next to the global constraints, there are local constraints that are implemented by way of step patterns. Those put limits on the allowed transitions between matched pairs of the two series and may constrain the slope of the warping path. As explained by Sakoe & Chiba, when the slope of the warping function (in the matrix) is very steep, a (possibly unrealistic) correspondence between a very short pattern in the first series and a relatively long sequence in the second series occurs. Many different step patterns exist as well as different classification systems [107]. Sakoe & Chiba classify according to the intensity of the slope constraint, expressed by the P-measure, and the (a)symmetry of the step weighting. Regarding the slope constraint, the smaller P, the steeper the slopes that are allowed, and the more discrimination between different patterns is degraded. The larger the value of P, the more rigid the warping function is, until it is restricted to the diagonal of the matrix. In their study, P-values of 0, 0.5, 1 and 2 are evaluated and P1 is recommended [102]. P0 corresponds to no constraints (as applied in example Figure 5-23 (a,b)), an alternative with P0 is called symDIA and favours the diagonals in the warping function (e.g. Figure 5-23 (g-h)). P2 corresponds to very strong restriction when applied in our dataset, not allowing for a match between elements that are elongated more than 1 time step. P1 and P05 are in between the extremes (Figure 5-23 (i-j-k-l)). Therefore in this study, the cases of P0 and P1 will be further compared, in combination with the aforementioned global constraints. The (a)symmetric properties of the step pattern refer to the weights given to the possible steps in

the step pattern. In the asymmetric cases, it is possible that some steps obtain weight zero and thus some parts of the time series are not treated equally, which seems a rather dangerous action, and, as told by Sakoe & Chiba, should be avoided as long as possible. So in this study, only symmetric forms of the step patterns are used. As a final remark, the use of constraints and the variety of constraints available, might seem objectionable to the reader, but, as the vast amount of literature illustrates, and Sarda-Espinosa states, it is usually agreed that using no constraint is a poor choice [109].

In this brief section on dynamic time warping, the author has tried to give an overview of the most important aspects of DTW in the context of the current study. A considerable amount of further research on the DTW algorithm relates to further speeding up the calculations, which is a bottleneck in the application of this widely applied and well-established method on large datasets. The use of constraints already has the desirable side-effect of speeding up calculations, but continued pruning off numbers of sequences that could not possibly lead to the best match, has led to the lower bounding techniques proposed by Keogh, Ratanamahatana-Keogh and Begum, and an interactive multiscale approach is proposed by Zinke [82,105,106,110,111]. With the *k*-Shape method, Paparrizos & Gravano recently proposed an alternative distance measure (SBD) and clustering approach (*k*-shape), that presents similar accuracy than the use of the constrained DTW-measure in the k-medoids clustering method applied in this study, but has the benefit of a significantly faster computation [96]. These approaches are not applied in this study, but could be consulted in situations where computational speed is crucial, for example for applications on large datasets.

Findings

Different pre-processing options are selected for further assessment. The data are transformed using either standardisation or multiplicatively detrended data, after the days with Zero and Standby patterns have been removed from the data. The selected metrics include the Euclidean metric and the (unconstrained) Dynamic Time Warping metric, as well as three types of constrained DTW's, using the Sake-Chiba band (size 1 or 2) as a global constraint, and two symmetric step patterns (P0 and P1) as local constraints. Clustering analysis is applied to the 25 cases using the Partitioning Around Medoids method (PAM), allowing for the selection of 3 up to 10 clusters. For selecting the optimal amount of clusters, two criteria are considered, that is the Average Silhouette Width (ASW) and the criterion from Calinski-Harabasz (CH).

Table 5-3: Cluster analysis: preprocessing options

Metric	Constraint		Transformation	
	Global	Local	Standardised	Detrended (Multiplicative)
Euclidean	-	-	STAN_EUC	DTRm_EUC
DTW	-	-	STAN_DTW	DTRm_DTW
DTWc	Sakoe-Chiba 1	Symmetric P0	STAN_DTW-SC1-P0	DTRm_DTW-SC1-P0
DTWc	Sakoe-Chiba 2	Symmetric P0	STAN_DTW-SC2-P0	DTRm_DTW-SC2-P0
DTWc	Sakoe-Chiba 2	Symmetric P1	STAN_DTW-SC2-P1	DTRm_DTW-SC2-P1

DISTANCE MEASURE

In Figure 5-25 and Figure 5-26 the ASW-criterion and the obtained number of clusters are presented for the multiplicative detrended data. When comparing the Euclidean and unconstrained DTW, the Average Silhouette Width is sometimes better for one metric, and sometimes for the other. But it is found that the constrained DTW measures are most often similar or better than both Euclidean and DTW measures. Among the constrained DTW measures, the most stringent measures SC1-P0 and SC2-P1 are typically better. The Euclidean metric often leads to a higher amount of clusters selected than the DTW, since patterns with a similar shape, that are shifted in time, are more likely to end up in different clusters. When comparing the unconstrained and constrained DTW, the constrained DTW limits the time shifts allowed during the warping, which can (in some cases) lead to extra clusters in comparison to the unconstrained DTW, or to make objects that are too much shifted in time end up in different clusters. For example, Figure 5-28 presents the clustered objects (in black lines) and medoids (bold and coloured lines) for Case 6 for three measures. When the Euclidean distance is applied (a), clusters of similar shape are separated if they are shifted in time, for example clusters 1,3,4 and 8 have similar shapes, as well as 2, 5 and 7, and 6 and 9. In the unconstrained DTW (b), they are roughly grouped together, but the objects of clusters 6 and 9 are split over clusters 1 and 2, while in the constrained DTW (c) they are more likely to end up together in the 1st cluster. Thus the Euclidean distance metric allows for a more detailed clustering, and the constrained DTW allows for a rougher clustering.

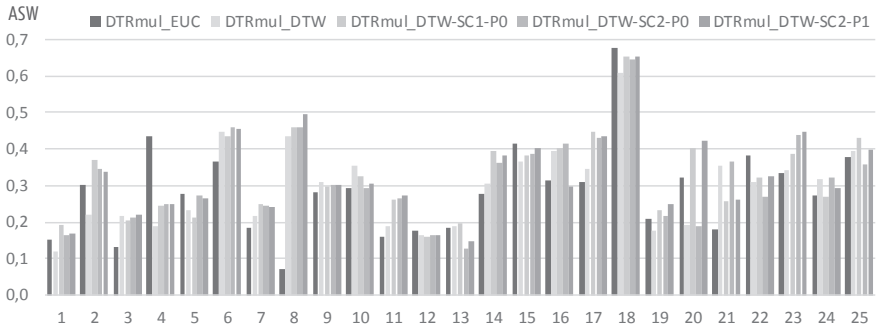


Figure 5-25: Distance measures compared for DTRmul data using ASW-criterion

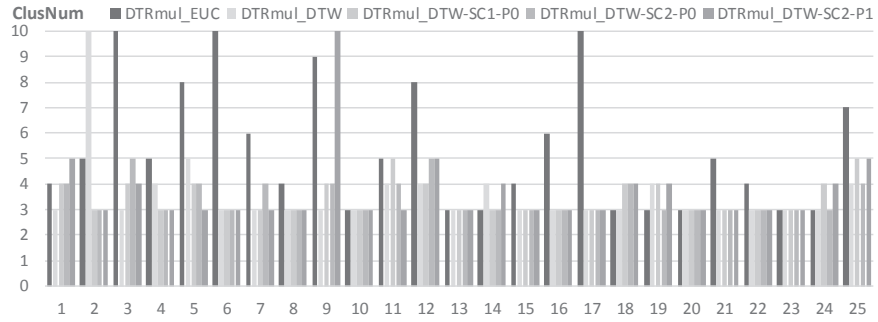


Figure 5-26: Distance measures compared for DTRmul data by number of clusters

Comparing the distance measures on standardised data-sets, or evaluating the clusters using the Calinski-Harabasz Index, can lead to differences in the particular results. Using the CH Index, the amount of clusters selected is typically lower than when the ASW index is used, and the DTW measure performs better than the Euclidean, but still the constrained DTW metrics outperform the other ones. Therefore it is concluded that the constrained DTW-measures SC1-P0 (or SC2-P1) are beneficial for the studied data.

TRANSFORMATION

Clustering using the multiplicative detrended data leads (in general) to better results than when standardising the data, for example the ASW increases around 6pp on average for the 25 cases. See for example the ASW for clustering with the constrained DTW measure for both standardised and detrended data in Figure 5-27. For only about 4 cases, the standardised data lead to higher ASW. The findings are similar when the other distance measures are applied. Also when the CH-index is used, the detrended data lead to an increase between 15 and 30%. Therefore, it is decided to use multiplicative detrended data for the clustering analysis.

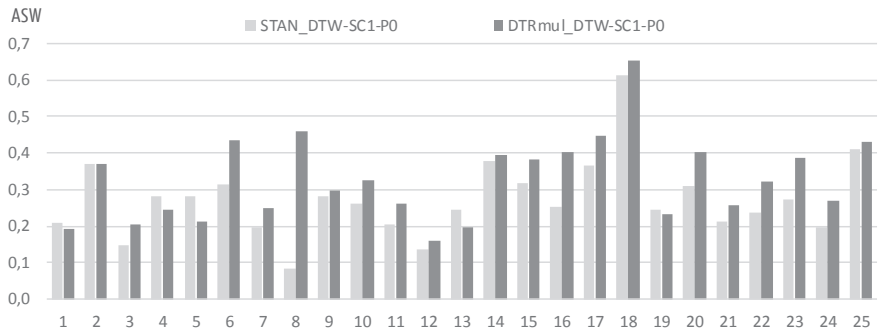


Figure 5-27: : Clustering pre-processing (transformation) compared using ASW-criterion

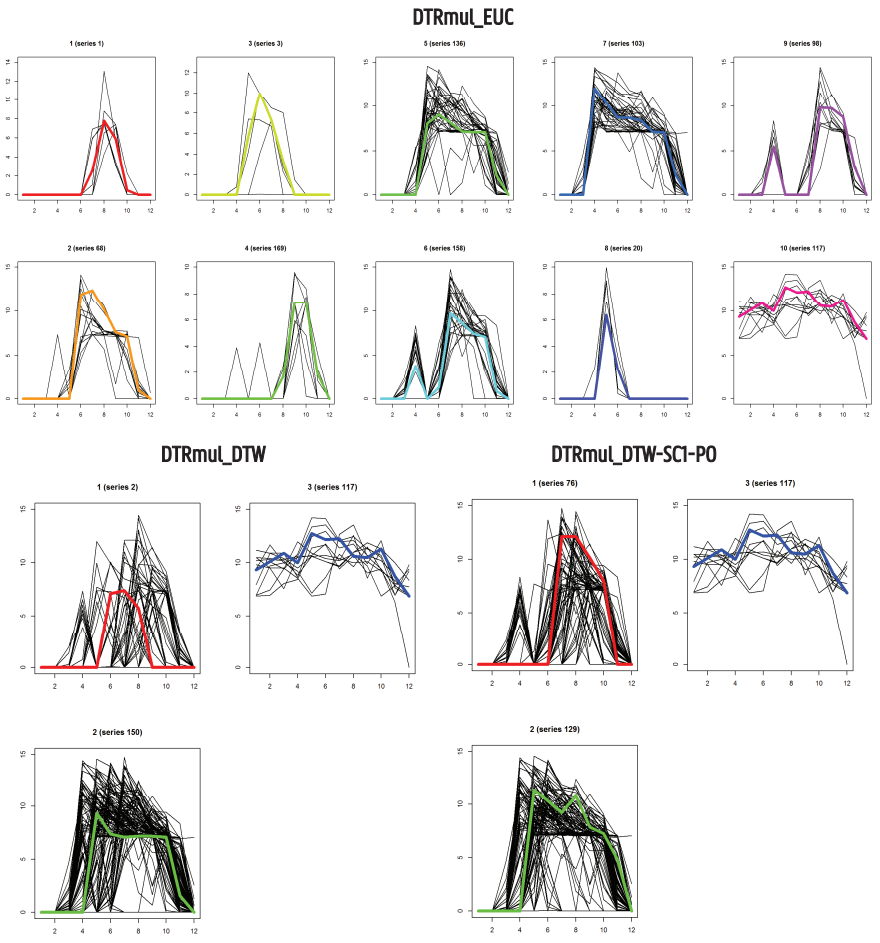


Figure 5-28: Clusters: patterns and medoids; for Euclidean, DTW and constrained DTW: Case 6

5.3.3 Clustering energy use time patterns

Cluster analysis is now applied to the data for the 25 cases and the results are illustrated in terms of cluster validity and number of clusters selected. In a first approach, the clustering algorithm is simply applied to the data from the entire year. In a second approach, the year is split into a summer period, a winter period and an 'intermediate period'. First clustering is applied separately to the objects of the different periods and the representative objects are selected. In a second step, the clustering is applied to the data of the entire year, using the proposed representative objects, and the clustering is further optimised. For both clustering approaches, the resulting seasonal components are estimated by averaging, and the share of the variance in the original data, that is explained by the seasonal components, is calculated. The approaches are compared to each other, and to the reference case, where the seasonal components are estimated by use of multiplicative time series decomposition using moving averages (see section 5.2).

All Year Cluster Analysis

The cluster analysis method applied is the Partitioning Around Medoids (PAM) method. The function is implemented in the R environment in the cluster- and fpc-packages [112,113]. The data is pre-processed according to the conclusions from section 5.3.2: the Zero- and Standby-patterns are excluded from the data, the data is transformed using multiplicative detruding, the distance measure applied is Dynamic Time Warping with a Sakoe-Chiba band of size 1 and step pattern P0, or with a Sakoe-Chiba band of size 2 with step pattern P1.

CLUSTER VALIDITY INDEX & NUMBER OF CLUSTERS

The PAM-algorithm is applied for different numbers of clusters *k*, and the final number of clusters is selected for which the Cluster Validity Index is optimised. As was explained in section 5.3.2, the Silhouette Coefficient or Average Silhouette Width (ASW), Davies-Bouldin (DB) and Calinski-Harabasz (CH) are in the top 3 of best performing CVIs in the study of Arbelaitz et al. [97]. However, even if no significant statistical difference is found in the performance of these three CVIs in general, they do not per se lead to the same amount of clusters selected for specific datasets. Therefore it was decided to perform a small experiment with the three of them for some datasets from this study and applying the PAM-method. First, values of *k* between 2 and 20 are allowed and the optimal number of clusters is selected. At the first sight, it is found that both low and high values of *k* are selected, and often one CVI selects a very high value of *k* while the other ones select a rather low value for the same dataset. When taking a closer look, it is found that for the ASW- and DB-criterion, very often there are a few near-optimal values of the criterion close to each other, but for quite different values of *k*. For example the optimal value may appear for k=20, while the near-optimal value is for k=3. And so for these two indexes, the optimal value for one criterion may coincide with a near-optimal value of the other criterion, and a (near)-optimal *k*-value lower than 10 exists. For the CH-criterion, the results are more different than for the other two criteria: the optimal value for k is often close to the minimum allowed (2 or 3) and as k rises, a slow decay in the criterion appears. As mentioned in section 5.3.2, this criterion would also be more sensitive to noise in the data, which is present in the studied datasets, and so the CH-criterion is not selected. In the remainder of the study, the silhouette coefficient will be used to as a CVI, also for the reasons mentioned in section 5.3.2. It is mentioned that in future research, the use of a combination of 2 or more CVIs, e.g. ASW-criterion and DB-criterion, could be investigated for the selection of *k*.

In the context of this study, the clusters resulting from the cluster analysis will be modelled in function of external variables using logistic regression. If the amount of clusters would be high, let's say higher than 10, then typically also very small clusters will be included with few objects that may not adequately represent the distribution. As a result, the small clusters may be completely excluded from the model, or they may incorrectly be allocated to some very specific conditions of the external variables. Therefore, a maximum of 10 clusters is allowed for. On the other hand, from physical point of view, it is found reasonable that at least 3 different energy use patterns appear throughout the year, which is the proposed minimum amount of clusters. In summary, the PAM-algorithm is applied for a different number of clusters *k* between 3 and 10, and the partitioning with *k* medoids is selected for which the Silhouette Coefficient is maximised.

RESULTS: ASW AND NUMBER OF CLUSTERS

Figure 5-29 presents the selected number of clusters and the Silhouette Coefficients for the 25 case-study dwellings, when using the Euclidean Distance, the DTW distance with Sakoe-Chiba band size 1 and P0, and the DTW distance with Sakoe-Chiba band size 2 and P1 step pattern. For the Euclidean distance, the amount of clusters selected is on average higher than for the DTW distances, where it is between 3 and 5 for most cases. For 6/25 cases the Euclidean distance leads to the highest ASW, while for 9/25 cases the DTW-SC1-P0 and for 10/25 cases the DTW-SC2-P1 has the highest ASW. The ASW is on average 0,28 for the Euclidean distance and 0,33 for the DTW distances. The interpretation of these values will be discussed at the end of the next section.

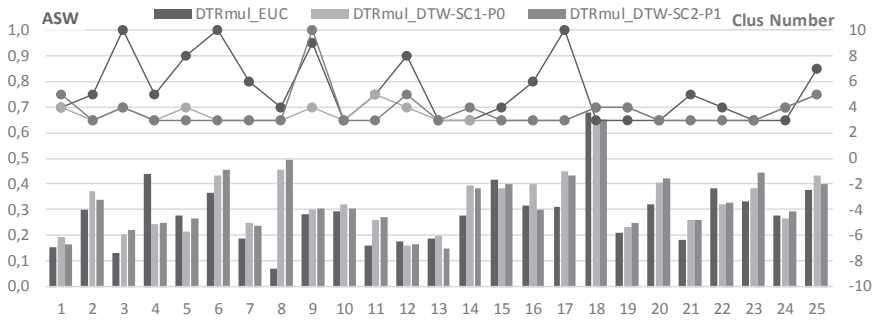


Figure 5-29: Cluster analysis with different distance measures: ASW and number of clusters

Cluster Analysis with Seasons

As stated by Hastie et al., cluster analysis is an unsupervised learning procedure, and for motivating the algorithms and results of clustering, the researcher should resort to heuristic arguments [95]. When observing the All Year Cluster Analysis, it is found that when using the DTW-metric in combination with the PAM-method, the amount of clusters selected tends to go towards the minimum amount of clusters allowed. In the PAM-method, smaller nearby-clusters tend to be absorbed into larger clusters [98]. Besides, the comparison of different clustering criteria and the use of the Euclidean distance metric illustrated that for the same datasets, often a higher number of clusters was selected. In order to get more grip and understanding on the cluster selection results, an alternative cluster analysis approach, that starts from some basic insights in the case study data, is tested.

The basic assumption behind this approach is the variability of the space heating demand through the year, because of its relationship with the weather conditions. For occupied dwellings in the Belgian climate, it can be expected that every year there are days where no energy use is requested for space heating (in summer), days with space heating energy use (winter), and days where a limited space heating demand is required (intermediate season). This is also true for cases where gas is used for space heating and other functions such as domestic hot water demand or cooking, that require energy use also during summer time. This approach is thus also applicable to that kind of data.

First the day-length energy use time series are grouped into three seasons: the summer period (from mid May to mid-September), the winter period (from mid November to mid-March) and the intermediate periods. Cluster analysis is applied to each of the three seasons separately and the representative objects or medoids of the clusters are selected. For winter and summer time it is assumed that between 1 and 5 clusters are selected. For the intermediate season between 3 and 8 clusters are allowed, since it most probably contains some typical summer patterns, typical winter patterns and intermediate patterns. In a second step, the clustering is applied to the data of the entire year, using predefined representative objects, that is all the representative objects from the three seasons, and the clustering criterion is estimated. Then an iterative optimisation algorithm is started: one of the representative objects that originates from the intermediate season is deleted, the cluster analysis is repeated and the clustering criterion value is calculated; this step is repeated for every representative object from the original set. The representative object for which the deletion gives rise to the highest increase in the clustering criterion value (typically one of the representative objects which is similar to another one) is definitively deleted from the set of medoids. Then for each of the clusters the medoid is redefined. Then the whole process is repeated for the new set of medoids, until no further increase in the clustering criterion value is observed or until there are no medoids left from the original intermediate season. The idea behind this approach is that more attention is possibly given to smaller but dense clusters. For the case in which the energy use in summer is zero or stand-by for most of the days, and is thus excluded in the pre-processing step, the summer and intermediate season are merged. This situation applies to the current data-set, where energy is used only for space heating.

RESULTS: ASW AND NUMBER OF CLUSTERS

The ASW and number of clusters are compared for the 'AllYear'-approach and the 'PerSeason'-approach, for the preprocessing option DTW-SC1-PO. Remark that for case 2, the dwelling that uses gas only for limited days of the year, the 'PerSeason'-approach is not applied, because there are almost no observations in the summer and intermediate season. The amount of clusters for the PerSeason-approach is equal or higher than for the AllYear-approach (except for 3 cases) but lower than for the AllYear-approach with the Euclidean distance metric. The Silhouette Coefficient is on average 0,31 which is slightly lower than for the AllYear-approach. For about half of the cases the AllYear-approach leads to higher SC, and for the other half the PerSeason-Approach is higher.

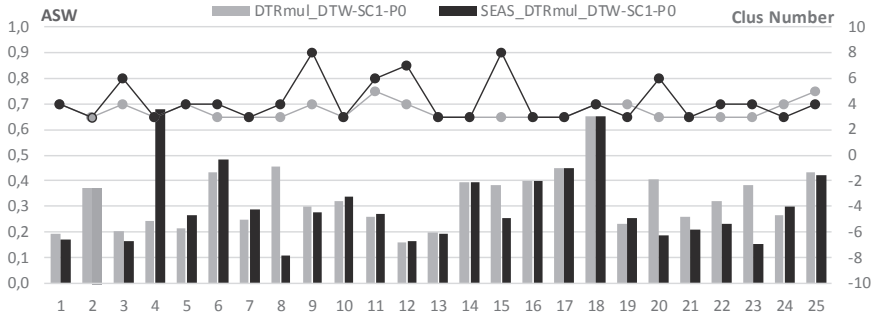


Figure 5-30: : Cluster analysis for all year or with seasons: ASW and number for clusters

Looking closer to the ASW-values for the individual cases for both approaches, for the majority of the cases, the ASW is between 0.26 and 0.50, which would mean that a weak structure is found in the data, according to the subjective interpretations by Kaufman & Rousseeuw (see Table 5-2). For 7 up to 9 cases, the ASW is lower than 0.26, which they interprete as 'no substantial clustering structure' and for one or two cases, case 4 and case 18, the ASW is higher than 0.6, indicating a 'reasonable' structure has been found. However, since the proposed classification is subjective, it will be further observed for this specific data set in section 5.4.

EXAMPLE CASES:

In Figure 5-28 the clusters for case 6 are illustrated when All Year cluster analysis is applied, using different distance measures. For the constrained DTW, three clusters were obtained (see Figure 5-28 (c) and Figure 5-31 (a)). When now the cluster analysis with seasons is performed using the same constrained DTW measure, one additional cluster is maintained. For this case, the silhouette criterion increases. For case 20, using the clustering approach with seasons, the amount of clusters increases from 3 to 6 clusters, and the silhouette criterion decreases. In addition to the three existing clusters, 2 clusters from the intermediate season are maintained (red and yellow) and in the winter period, an extra cluster is separated.

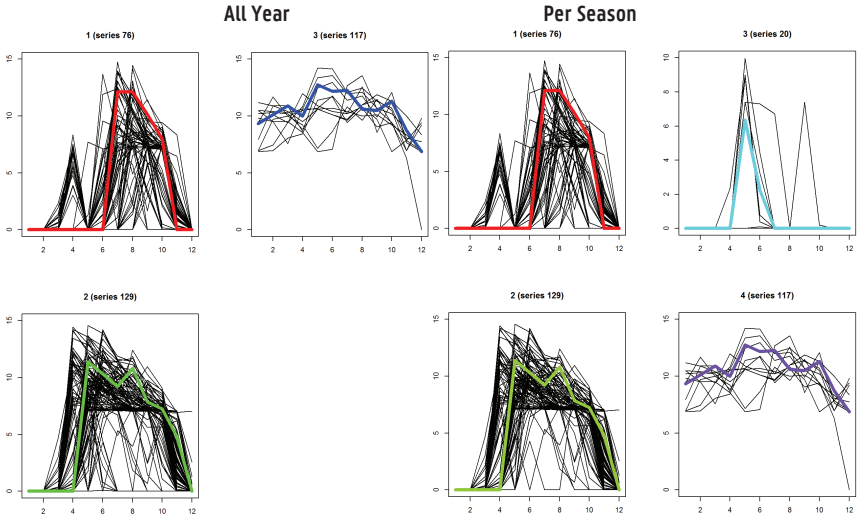


Figure 5-31: Clusters and medoids for Case 6: DTRmul_DTW-SC1-P0 measure: All Year, Per Season

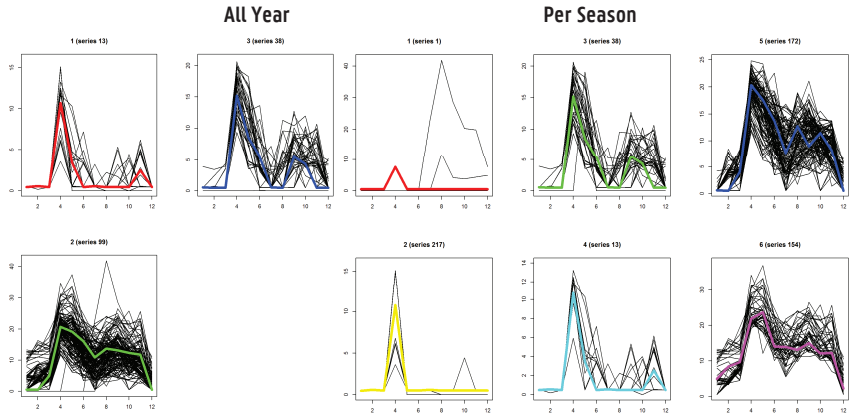


Figure 5-32: Clusters and medoids for Case 20: DTRmul_DTW-SC1-P0 measure: All Year, with Season

5.3.4 Seasonal Components

The observation of the Silhouette Coefficients and number of clusters in previous sub-sections are not sufficient to draw final conclusions on the selection of the clustering approach. In the end, what counts is how well the seasonal components fit the original data. The seasonal components are calculated from the clusters assigned in the two cluster analysis approaches (the All Year and with Seasons clustering), and they are compared to the original non-differentiated time series decompositions on the same data-set (see section 5.2.5). The decomposition method applied is still the multiplicative classical decomposition using moving averages. The trend component remains equal for all situations (it is subtracted from the data before modelling the seasonal component) and so the variations in the fitted data, which is obtained by multiplying seasonal and trend components, are entirely due to the seasonal components. While the differentiated seasonal components were obtained by taking for each time step of the daily period the average for all days in the time series, the differentiated seasonal components are obtained by averaging for all days that belong to the same cluster, and thus the seasonal components differ for each of the clusters. Figure 5-33 presents the variance of the original data explained by the trend, seasonal component, residuals and covariance, for the differentiated seasonal components. The results for the non-differentiated components were presented in Figure 5-18. In Figure 5-33 is observed that the covariances in the differentiated models are clearly lower than in the non-differentiated models, indicating that the fit and the residuals are less correlated. Besides when comparing the results for the All Year cluster analysis for different distance measures used, it was found that the results are similar for the two options using DTW-distance (DTW-SC1-P0 and DTW-SC2-P1). For the options with constrained DTW-metric, on average 76% of the variance in the model is explained by the combination of seasonal and trend component. For individual cases, the differences are below +/-7 pp. For the Euclidean distance, where the number of clusters was typically higher, the variance explained by the combination of seasonal and trend component rises to an average of 81%. For individual cases, an increase with up to 20pp is seen (Figure 5-33). The All Year approach can be compared with the Cluster Analysis with Seasons. It is found that on average, the results are similar, and for specific cases, the differences were below +/- 7pp.

When comparing to the non-differentiated seasonal components, where on average 69% of the variance was explained by the model fit, the application of differentiated seasonal component is an improvement. The Mean Absolute Error (MAE) is another measure applied when comparing the original to the fitted data. It is found that in comparison to the non-differentiated time series decomposition, the MAE reduces between 5% and 50% after the seasonal components have been grouped by use of clustering analysis. In Figure 5-34 the time series components with differentiated seasonal components are plotted for case 18 and can be compared with non-differentiated seasonal components in Figure 5-17. During the summer time the seasonal components are Zero-patterns, and for different periods of the heating season, different seasonal components appear. As a result, the residuals are significantly reduced in comparison to the non-differentiated seasonal components, which is especially observed in the mid-winter period. For the summer period, no effect is seen on the residuals, since the trend component takes also a value of zero.

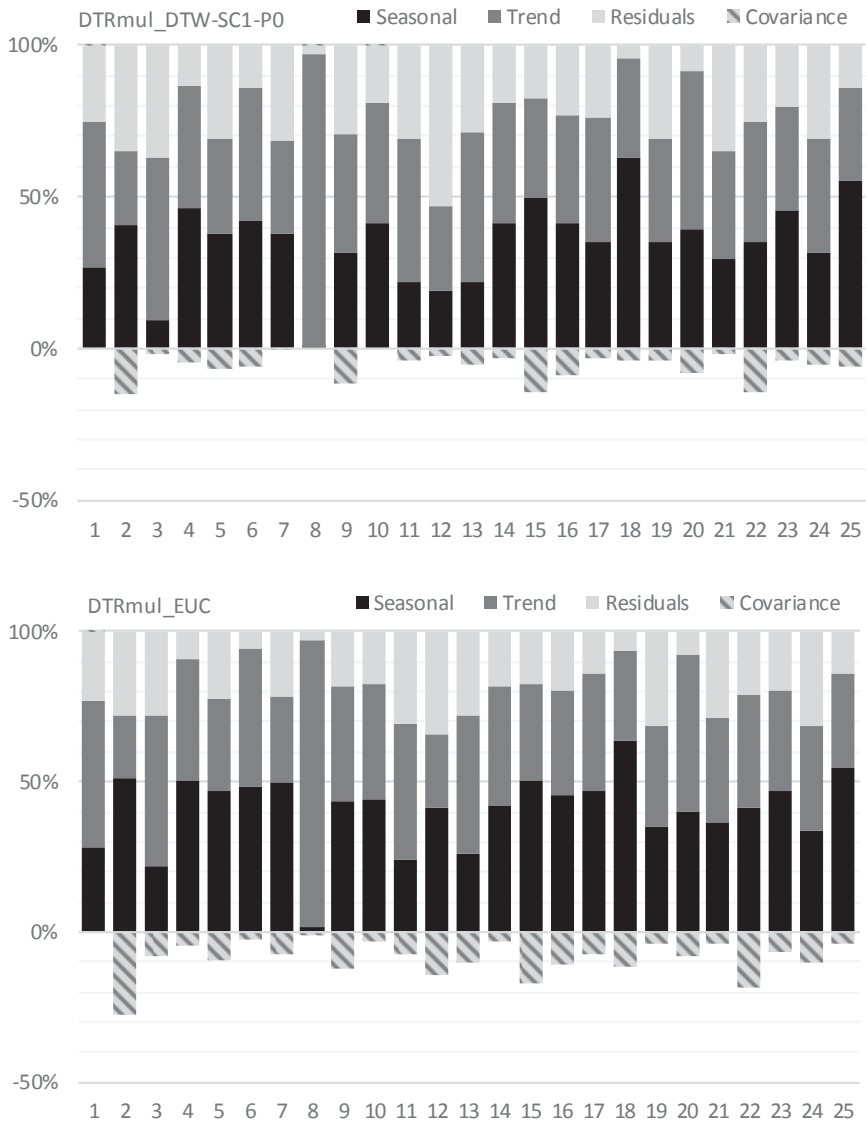


Figure 5-33: Composition with clustered data - Full Year - Explained/Observed Variance

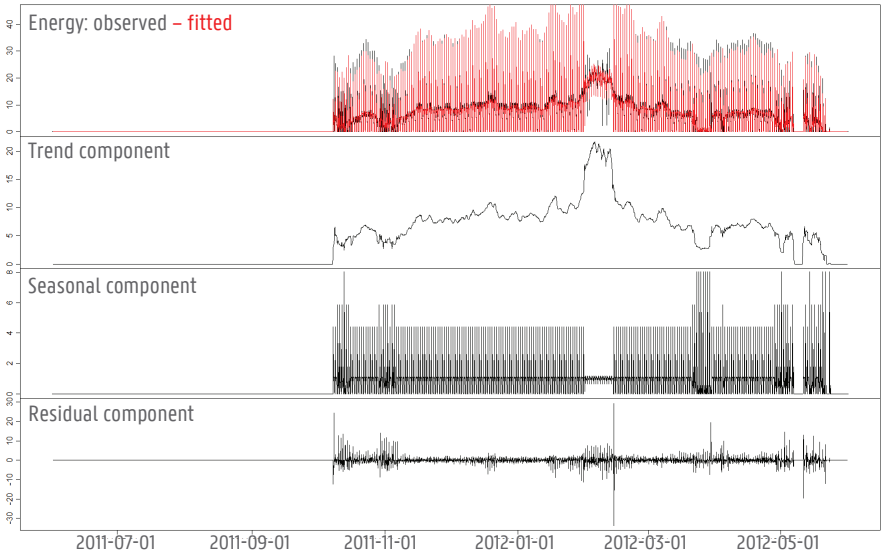


Figure 5-34: Multiplicative TSD with seasonal component differentiated by clustering – case 18 – full year

5.3.5 Examples

The clusters of energy use time patterns most often appear for specific ranges of values of the weather conditions or calendar variables. This is illustrated in the following figures, where each time the following results are plotted:

- Top left: time series plots of the cluster medoids (the representative objects for each cluster)
- Top right: the cluster objects, sorted per cluster (y-axis) in function of the day of the year (x-axis)
- Bottom left: box-plot for the ranges of equivalent exterior temperature for which the clusters appear
- Bottom right: bar plots of the clusters occurrence per day of the week

In Figure 5-35 two situations for case 6 are displayed when the All Year cluster analysis is performed using the constrained DTW (a) and the Euclidean metric (b). The plots of the clusters and clustered objects for these situations are found in Figure 5-28(c) and (a). When observing Figure 5-35 (a), it is found that the third cluster, where no night set-back appears, is characteristic for conditions when the equivalent exterior temperature is below -2°C. Clusters 1 and 2 appear for the rest of the heating season. They are in the same temperature ranges, but cluster 2 appears more on Saturdays, Sundays and Wednesdays than Cluster 1. When observing the clustering with the Euclidean metric, the cluster without night set-back is recognised in cluster 10 (pink). Then it is seen that clusters 2, 5, 6, 7 and 9 are in the same temperature range, and that cluster 5 (green) appears significantly more on Saturdays and Sundays, while clusters 6 and 7 and 9 (blue and turquoise and purple) appear clearly more on week-days. When observing the cluster medoids, it is found that in clusters 6, 7 and 9 the heating starts after six in the morning, while in cluster 5 it only starts after 8 in the morning, which can be explained by different settings of the heating system during the weekend. Then the remaining cluster appear especially at higher exterior temperatures. In Figure 5-36 the All Year cluster analysis (a) and the cluster analysis with seasons (b) is represented for Case 20, using the constrained DTW-metric. The plots of the clusters and clustered objects for these situations are found in Figure 5-32. Using the All Year cluster analysis, three clusters are found that appear in rather different conditions of the equivalent exterior temperature. The cluster 2 that appears at the coldest temperatures (green), is split in two clusters in the cluster analysis with seasons, each occurring at rather different (but slightly overlapping) conditions of the exterior temperature. Also here, in the coldest cluster, the energy use does not drop to zero at night. Furthermore, now three small and rather similar clusters appear at the warmest period of the heating season.

These cases also illustrate an advantage and disadvantage of the Euclidean metric and the DTW metric. On one hand the higher level of detail recognised by the Euclidean metric allows to discern the different energy use time patterns more precisely, and thus to discover relationships that were less clear when the constrained DTW metric was applied. On the other hand, using the Euclidean metric, the energy use time patterns can be so much differentiated, that it is difficult to allocate all of them to different conditions of the external variables. Moreover, the cluster size can become pretty small, and may thus start to represent exceptions or small variations in energy use that are not representative for the data-set.

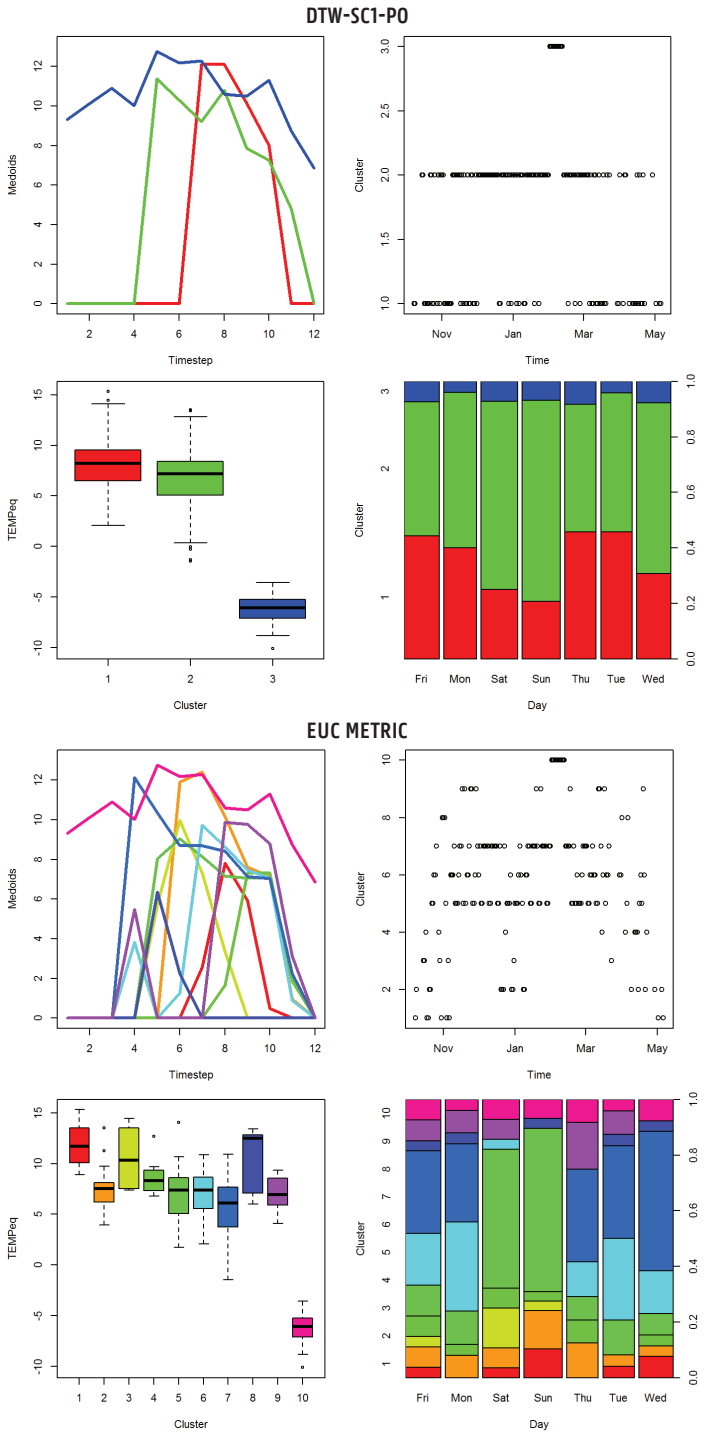


Figure 5-35: Cluster plots with external variables: All Year: Case 6

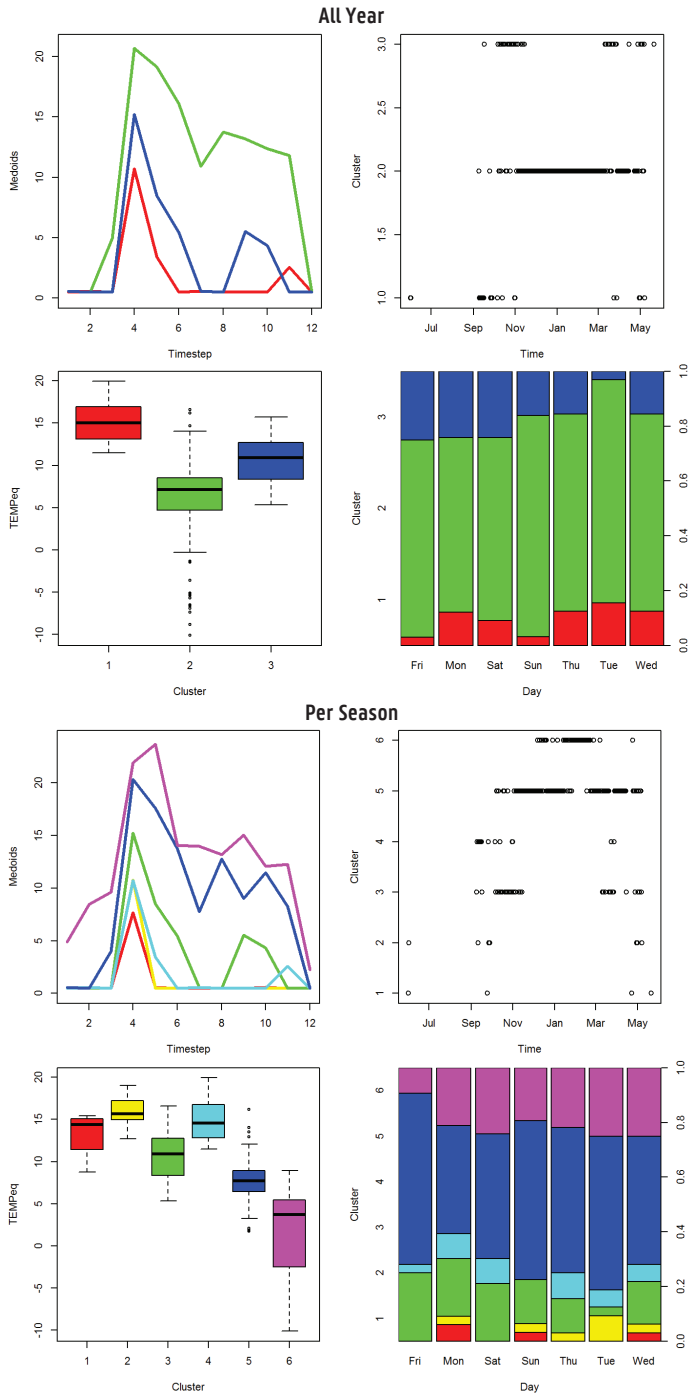


Figure 5-36: Cluster plots with external variables: DTW-SC1-P0 metric: Case 20

5.4 Classification of energy use time patterns

The aim of this section is to mathematically describe the relationship between the groups of energy use time patterns, obtained from the clustering in section 5.3, and external variables, that are weather and calendar variables. The resulting model enables to better understand the energy use time patterns, and to estimate or predict energy use time patterns for other data-sets. The relation between the clusters and external variables is described using multinomial logistic regression analysis, which is often applied in statistical classification methods. This method is explained in section 5.4.1. In section 5.4.2 the model building process is developed. Section 5.4.3 documents the classification analysis findings and results and section 5.4.4 discusses the obtained seasonal components after classification. Finally, in section 5.4.5 the classification results are interpreted and illustrated by use of some examples.

5.4.1 Multinomial Logistic Regression

Logistic regression is a type of regression analysis that is applicable when the dependent variable is a categorical variable, that is a variable with a measurement scale that consists of a set of categories, that can be ordered (ordinal variables) or unordered (nominal variables). The predictor variables may be categorical or continuous variables. In this study the categories are clusters of energy use patterns that are given a number as name. The clusters are not ordered, for example a cluster with number 2 is not per se closer to cluster number 3 than it is to cluster number 5. The predictor variables are weather variables (continuous) and calendar variables (categorical).

Logistic regression is a linear method, meaning that the boundaries for deciding whether an object belongs to one category or another category, are linear. In the simplest case, where there are two categories for the dependent variable (binary: “success” or “failure”), the parameter for the binomial distribution is p, the probability of “success” at a certain value of the predictor variable. The logistic regression model is a linear model for the logarithm of the odds of success $p/(1-p)$, also called the log-odds. When there are more than two categories for the dependent variable, a baseline category is selected arbitrary and the model is constructed of pairwise log-odds of each of the categories with the baseline category. The probability distribution function for the number of outcomes of the categories is the multinomial distribution, and the generalised logistic regression models are referred to as multi-category or multinomial logistic regression models. The models are usually fitted by maximising the likelihood function, which expresses the probability of the observed data as a function of the observed parameter [95,114,115].

As with linear regression models, if multiple predictor variables are available, a variable selection process is performed and the final model is evaluated. In this study a stepwise forward selection process is constructed, that uses preliminary insights in the data. For testing the significance of a predictor added to a model, the likelihood ratio test, Wald’s test and Lagrange multiplier test are available significance tests for nested models. In this study the likelihood ratio test will be used, which is slightly more computationally intensive, but more trustworthy than the two alternatives [114,116].

5.4.2 Data pre-processing and model building

Pre-processing

The dependent variable is the categorical variable that includes the groups of energy use patterns. It is composed of the clusters resulting from the cluster analysis (*Clusters 1,2,3…*), and the clusters with Zero-patterns (*Cluster 0*) and Standby-patterns (*Cluster 20*) that were separated before applying the clustering analysis. For each of the 7 cases where the Standby-patterns appear, the corresponding ranges of the main weather variables (exterior temperature, solar radiation) considerably overlap with those from the Zero-patterns, and so they are difficult to distinguish in a classification exercise, probably leading to lower classification quality and meaningless cluster boundaries. Therefore it is decided to transfer *Cluster 20* to *Cluster 0*. From physical point of view, this is logical, since in both situations, no actual Space Heating energy use is requested. Together, these two patterns cover the time of the year that is not the heating season, and when no gas use is expected. The boundaries between *Cluster 0* and the other clusters therefore define the boundaries of the heating season in houses where gas is only used for space heating, except for cases where the heating system is not working during the heating season, for example due to un-occupancy or system failure.

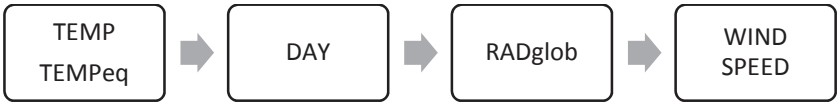
External Variables

The candidate predictor variables are the external variables, including the main continuous weather variables: exterior temperature *Te* (TEMP), the equivalent temperature Te_{eq} (TEMPeq), the global solar radiation *Rg* (RADGLOB) and the wind speed *Ws* (WINDSPEED). Secondly, the main calendar external variables (categorical variables) are: *Day 7* - the seven days of the week (7 levels), *Day 2* – weekdays and weekend days (2 levels), *Day 3* – weekdays, Saturdays and Sundays (3 levels), *Day 4* – Saturdays, Sundays, Wednesdays and other weekdays (4 levels), *Holi We* – weekdays and weekend days + official holidays.

Model Selection

The model selection procedure is a forward stepwise regression procedure: starting from a small model with only one external variable, other variables are gradually added. The model selection is a case-specific automated procedure. Only the variables for which the coefficients are significantly different from zero, remain in the model. The significance test applied is the likelihood ratio test for nested models, which returns a p-value. For accepting or rejecting whether or not a variable is significantly different from zero, confidence intervals of 99% and 95% are tested. At each step of the selection procedure, the suitable models are stored, and the final model is selected by selecting the model with the lowest deviance (Deviance=-2×Loglikelihood).

In the stepwise regression procedure following variables are tested consecutively:



Both the exterior temperature and equivalent exterior temperature are tested, and the model with the lowest deviance is selected. Likewise, for the DAY-variable, the different options are tested and evaluated.

Model Evaluation

In the evaluation of the final model, following aspects are addressed: (1) the misclassification error of the in-sample predictions, which is the percentage of objects misclassified when comparing the original and predicted clusters or categories, (2) the cluster size, for uncovering very small categories that cause the data to be unbalanced, and thus limit the number of predictors for which te effects can be estimated precisely [114], and (3) the variance explained by the seasonal components obtained when implementing the in-sample classification results in the energy use time series decomposition, and the MAE when comparing the original and fitted energy use time series.

5.4.3 Classification of energy use time patterns

The classification methodology is first applied using the results of the All Year clustering analysis as an input, with a constrained DTW-distance (DTRmul_DTW-SC1-P0). For all cases, the classification model reaches convergence. With regard to the external variables, the use of the equivalent exterior temperature leads to lower Deviance than the exterior temperature, and it is significantly different from zero for all 25 cases (99% confidence interval) according to the likelihood ratio test. From the model selection process, four of the candidate models emerge in the results: Te_{eq} + *Rg* (9/25 cases), Te_{eq}+*Rg*+*Ws* (7 cases), Te_{eq}+*Day7* (1 case) and Te_{eq}+*Rg*+*Day7* (8 cases). The main external variables are the equivalent temperature, solar radiation, the days of the week and wind speed. With regard to the calendar variable, *Day7* emerged, which is more open to variations between days than *Day2*, *Day3* and *Day4* and *Holi We*.

After the original clusters have been classified using the classification procedure with multinomial logistic regression, for each object (day) of the data, an original cluster and a 'fitted' or 'predicted' (in-sample) cluster exists, and thus the misclassification error can be determined. The misclassification error for the in-sample predictions is between 11% and 36%, with an average of 25%. This means that for on average 75% of the objects the original and fitted cluster are the same. The training data set consists of 4, 5 or 6 clusters (including *Cluster 0*), but in more than half of the cases one or two empty or very small clusters (<4 objects) appear. These typically originate from rather small original clusters, that appear in Spring or Autumn, and that are swallowed up by the surrounding larger clusters. In the following paragraphs, particular findings of the classification analysis are exposed. Then three options for improving the classification are tested.

Cluster Analysis Approach: All Year or With Seasons?

In section 5.3.3 no definitive conclusions were drawn on the use of the All Year cluster analysis (DTRmul_DTW-SC1-P0) and the cluster analysis with seasons (SEAS_ DTRmul_DTW-SC1-P0). It was found that when the seasonal components were calculated, both of them lead to similar residual percentages. When comparing them after classification, it is possible to check whether for one of them the obtained clusters are more easily related to external variables. It is found that the misclassification error for the clustering with seasons is in the same range as for the All Year

clustering, with an average of 26%. For individual cases, there can be differences between the two approaches up to 18pp for cases 8 and 23 (see Figure 5-37). For the clustering with Seasons, which starts in many cases with a higher amount of clusters, there are empty or very small resulting classes in about half of the cases.

When the seasonal components are calculated and the MAE of the fitted models are compared, it is found that for both models, the results remain very similar and differences are only a few percent, except for a few individual cases, e.g. case 5, where the cluster analysis with season leads to 12% higher MAE, and case 20 and 22 where it leads to an 8% lower MAE. Therefore it is concluded that the cluster analysis with seasons does not lead to significantly better results than the all year clustering. Since the cluster analysis with seasons is also a more complex and more computationally intensive approach, it is considered more advantageous to use the all year cluster analysis in the remainder of the study.

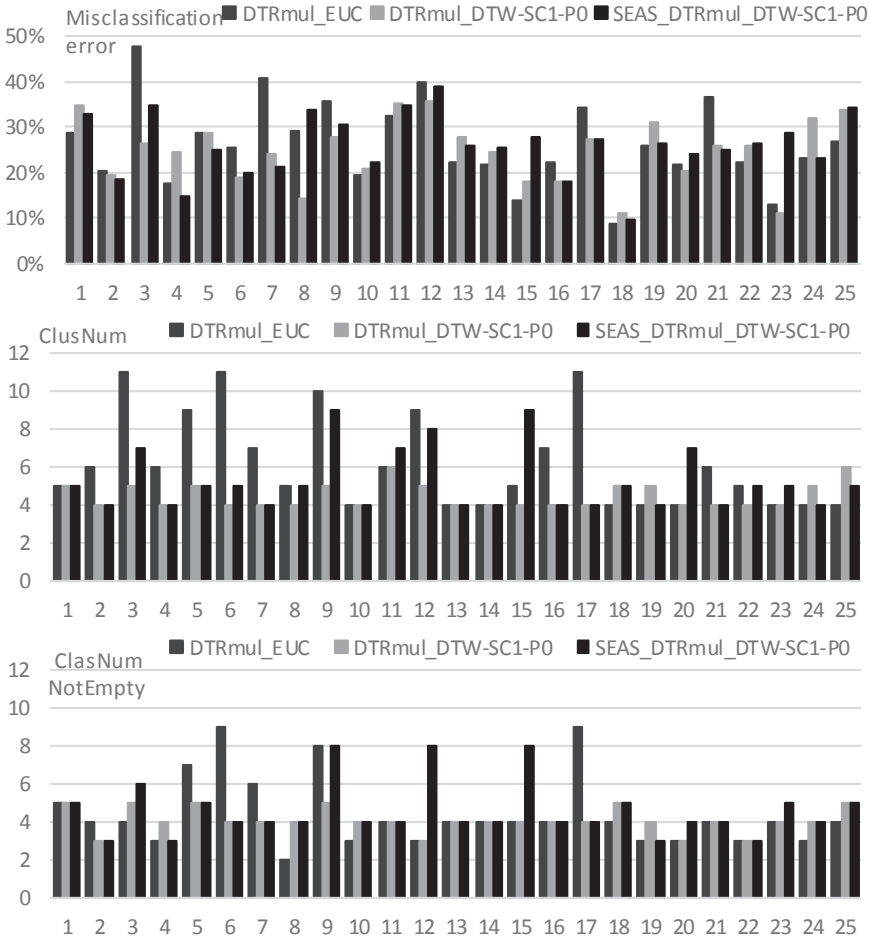


Figure 5-37: Classification after cluster analysis: misclassification error and number of clusters

Cluster Analysis distance metric: Euclidean or constrained DTW?

In the Cluster Analysis it was found that the use of the Euclidean metric leads to a higher amount of clusters than the constrained DTW metric (DTW-SC1-P0) for many cases. As a result the separation of different Energy Use Time Patterns is more detailed, and the resulting seasonal components allow to describe a larger part of the variance in the data. When it comes to the description of these clusters in function of external variables, the question is whether this higher amount of clusters can also be related to specific ranges of the external variables and thus lead to a more detailed model. Figure 5-37 presents the total number of clusters (ClusNum) available as an input to the classification model, or the maximum amount of classes that can be obtained for the given clustering results. Remark that there are at least four clusters, that is the minimal of three clusters obtained from the cluster analysis, and one cluster containing the Zero and Stand-by clusters. Figure 5-37 (ClasNum NotEmpty) presents the amount of non-empty classes obtained after classification. For the clustering with constrained DTW (DTRmul_DTW-SC1-P0), one or two empty clusters are detected in 8/25 cases. For the clustering with Euclidean metric in 19/25 cases one up to seven empty clusters occur. Besides, for clustering with seasons and DTW, the results are in-between the aforementioned results: in 13 cases one up to three empty clusters are detected. Moreover, especially for the cases with a large amount of remaining classes, also some very small remaining classes are detected, where only a few observations are located. For example in case 7 (Figure 5-38), from the original 7 clusters, 6 classes remain, from which two classes absorb most of the objects and three contain less than 7 objects. The clustering with constrained DTW for the same cases, leads to only 4 clusters and all of them remain after classification (Figure 5-39).

The occurrence of empty or small classes does not always lead to higher misclassification errors, since often they are derived from rather small clusters, and thus the effect may be small. However, for those cases where several empty or small clusters occur, the misclassification error peaks to more than 40% (e.g. case 3, 7, 12 using the Euclidean metric). A preliminary observation of the seasonal components and the MAE of the energy use time series fit, indicates in general only small differences between the different approaches. For a few cases, the Euclidean metric clusters lead to a clear improvement in MAE, e.g. for case 6 (see Figure 5-35 and Figure 5-40), 9 and 17, where the relation between certain (large) clusters and calendar variables is more clearly discerned. However, it is found that especially for the small clusters, the model tends to link the smaller clusters wrongly to very specific conditions for the external variables (e.g. to a specific day of the week). Before definitive conclusions are made on the distance metric, a few options for improving the classification method are explored.

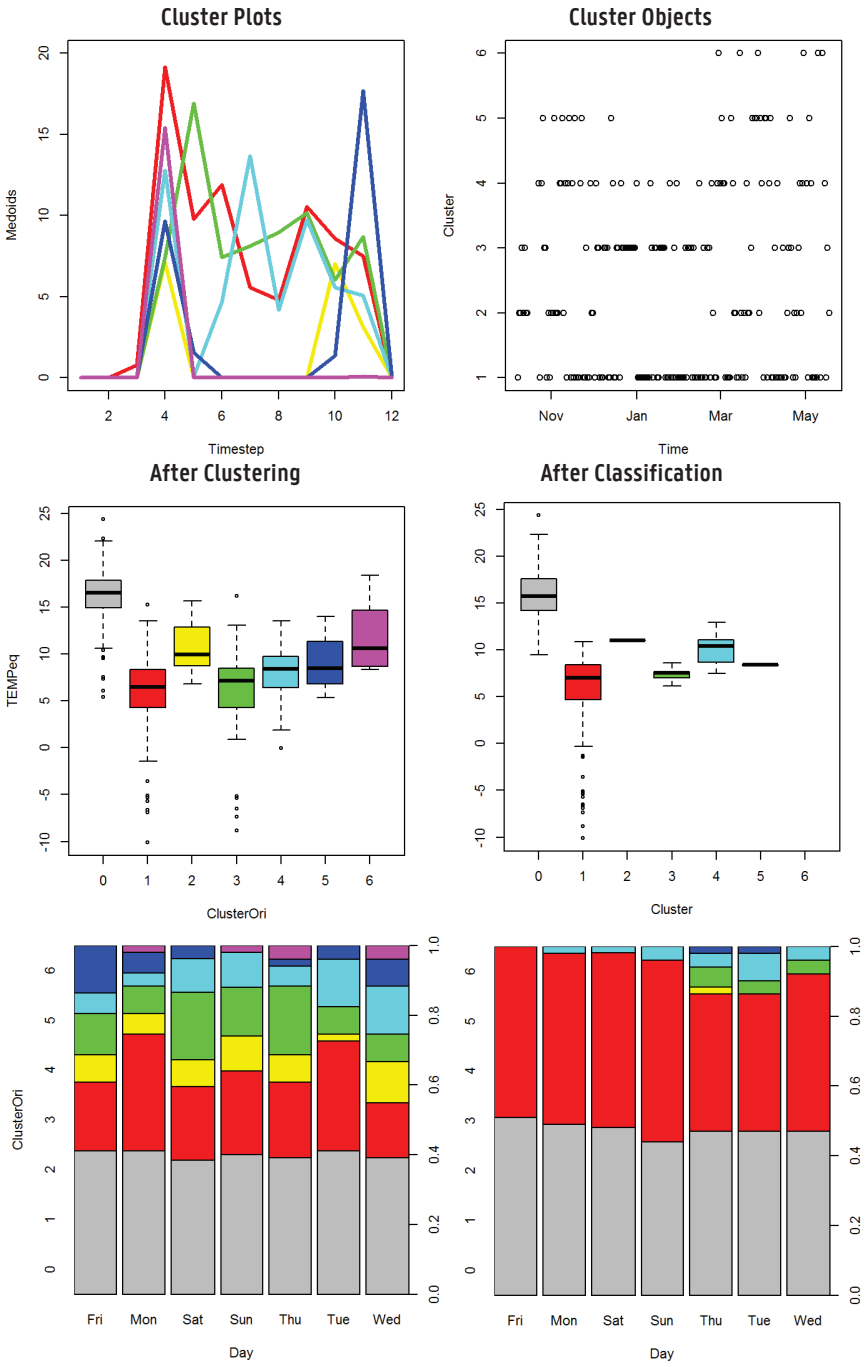


Figure 5-38: Case 7 – plots of clusters with DTRmul_EUC before and after classification with external variables

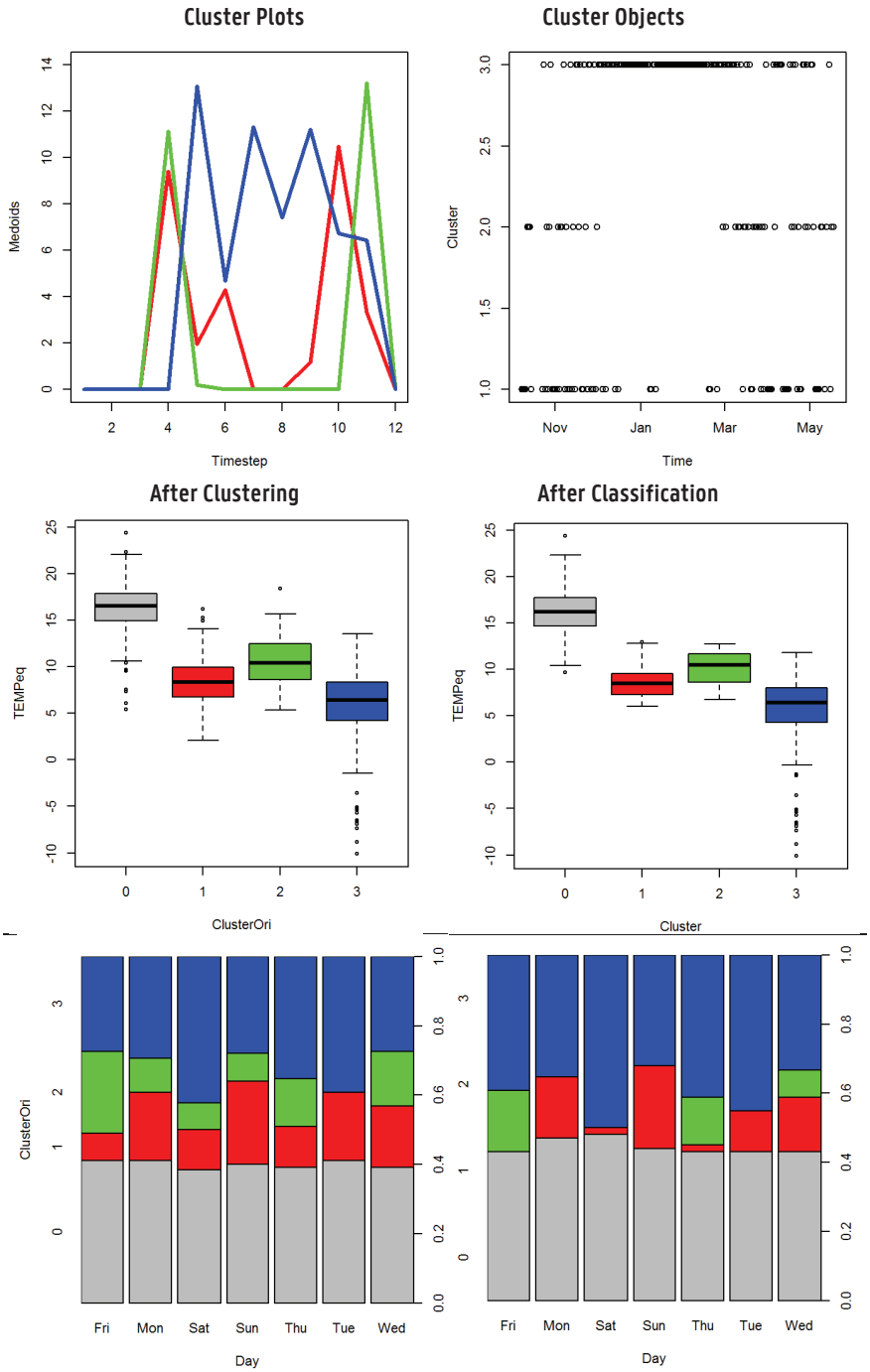


Figure 5-39: Case 7 – plots of clusters with DTRmul_DTW-SC1-P0 before and after classification with external variables

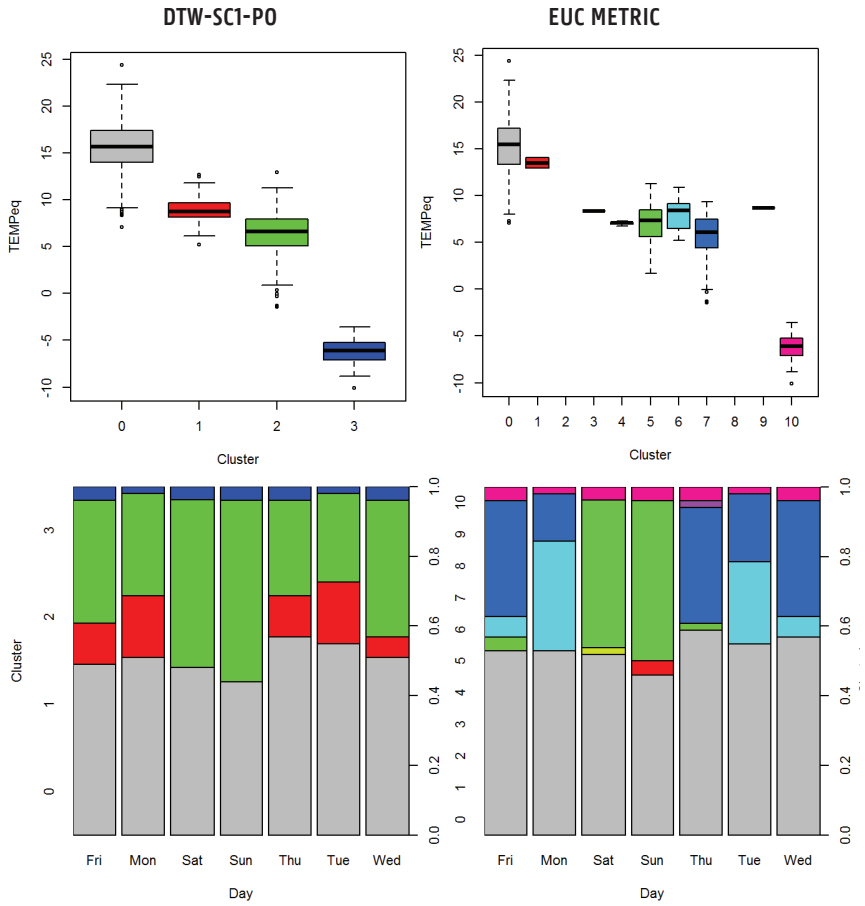


Figure 5-40: Case 6 – Class plots with external variables for DTW-SC1-PO and EUC metric

Classification Method Option 1: Treatment of Empty and Small Clusters

It is observed that after classification some categories contain very few or no objects. A first possible explanation is that the original categories (before clustering) contain few objects relative to other categories, therefore making the data unbalanced. Agresti refers to Peduzzi et al. for an approximate guideline on this matter, who suggest that ideally there are at least 10 outcomes of each type for every predictor. This means that if a certain category occurs only 20 times in the data, the model should not have more than 2 predictors [114]. The guideline is an approximate guideline, and results depend on the situation. For example in case 6 (compare Figure 5-35 and Figure 5-40), category 10 (that contains only 12 objects) is well explained by the model: it is correctly assigned to objects that appear at very low external temperatures. A second explanation is that two categories appear in the same range of all external variables in the model. For example in case 7 (Figure 5-38) categories 2, 5 and 6 contain between 6 and 27 objects in the original clusters and they appear in a similar range of exterior temperature conditions (as well as solar radiation and calendar variables), that also overlaps with the ranges observed for

the other larger categories. The explanatory variables thus cannot explain the different categories, and the objects are allocated to the most likely category, which is often the largest one. Thus category 6 is depleted in the classification process, and categories 2 and 5 contain only two objects, representing a very small range of the exterior temperature. This shows that the two reasons can reinforce each other. When observing the plots of the cluster medoids for case 7 in Figure 5-38 (top), it is found that clusters 1, 2 and 5 are more similar in shape than the other clusters to which their objects are allocated. Therefore a strategy should attempt to bring these clusters together into larger clusters and thus make the data more balanced.

The adaptation of the methodology for the treatment of empty and small clusters is as follows: after a first classification phase, categories with less than 4 objects are detected and each object is assigned to its neighbouring cluster. If more than one cluster is detected, the emptiest cluster is selected or (if both are equal in size) the cluster with the smallest average silhouette width. Then the classification is repeated with one cluster less. This process is iterated until no small clusters remain. In case iterations are needed, it is possible that for certain objects their original cluster as well as their neighbouring cluster are rejected. Then the clustering algorithm is repeated with predefined medoids (those from the remaining original clusters) to allocate each object to the nearest a cluster.

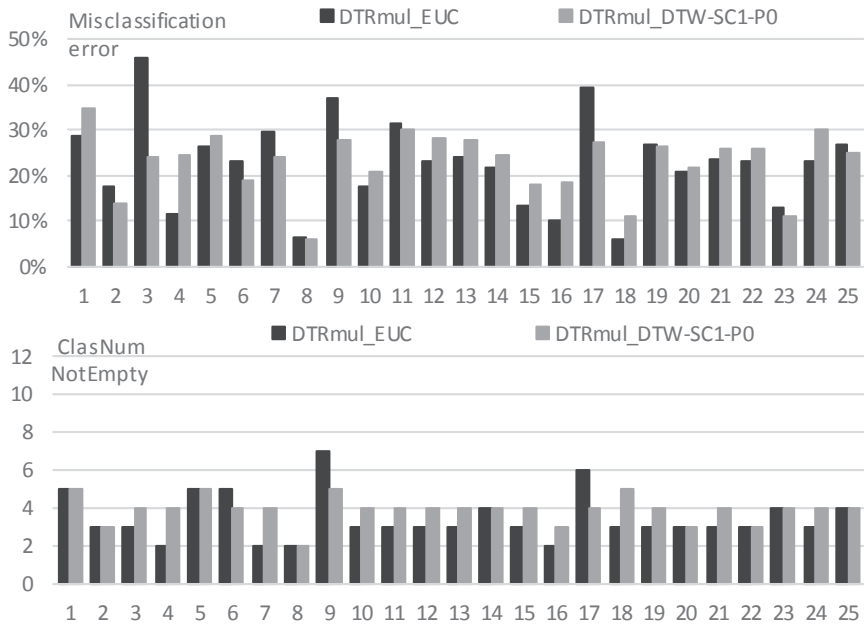


Figure 5-41: Classification with Option 1: misclassification error and number of clusters

For the All Year clustering analysis with constrained DTW, where 1 or 2 small clusters were detected in 8 cases, this leads to a lower deviance and a reduction of the misclassification error with 5 pp on average for the affected cases (Figure 5-41). When the Euclidean metric is used, in 21 cases the amount of clusters was reduced. In 16/25 cases the misclassification error reduced with on average 7pp. In the other 5 cases, it increased on average 2pp, maximum 5pp. For both metrics, the average misclassification error has reduced from 26% to 23% after this modification

of the classification method. Coming back to the example case 7 (Figure 5-38 and Figure 5-39), when applying the adapted methodology to the clustering with Euclidean metric, the amount of clusters is reduced to two: Cluster 0 in Summer and Cluster 1 in Winter. Thus the clusters from the intermediate season (2, 5 and 6) are wiped out. The reason is that the Euclidean metric does not consider the shapes from the intermediate season similar and thus also allocates the objects from the intermediate season to the winter season. On the contrary, the constrained DTW does recognise the similarity in shape and immediately identifies two 'intermediate season' clusters that contain more than 40 objects each, and that are maintained in the classification process.

It is concluded that in general this adaptation of the classification method helps to balance the data for the classification process, leading to more consistent results. The effect of this adaptation is the largest for the Euclidean metric, because the initial amount of small and empty clusters there is intrinsically higher. The median of the final amount of clusters for the Euclidean metric is 3, where it is 4 for the constrained DTW metric (Figure 5-41). And since the relocation of the objects of erased clusters is not based on similarity in shape when the Euclidean measure is used, in general the results for the clustering with DTW metric are preferred. Only for some cases, e.g. case 6, 9 and 17, the Euclidean metric leads to an improvement, for it recognises patterns that are not separated using DTW metric.

Finally, it is noticed that in this study categories with less than 4 objects were deleted. When observing the results, it is found that in some cases, small clusters remain, for example in case 9 there are two clusters with 7 observations, and in case 17 there is a cluster with 8 observations for which the range of external variables overlap with other elements. On the other hand, there are some pretty small clusters with 10-20 observations for which the shape of the energy use time patterns is very different from other clusters, and that are correctly classified in function of the external variables. These are for example cluster 10 in case 6, that appears at very low external temperatures. Deletion of such clusters because of their limited amount of data-points would lead to worse classification results. In conclusion, it is found that for the cases in this study, categories with up to 10 observations could be considered small clusters and absorbed in neighbouring clusters. However, this condition should be further investigated before being implemented on other datasets, and probably related to other properties of the classification results to avoid exclusion of well-classified small clusters.

Classification Method Option 2: Exclude negative Silhouette widths

For each object resulting from the Clustering Analysis, the Silhouette width and neighbouring cluster are calculated. When the Silhouette width is close to zero or negative, the object is located somewhere in between two clusters, and it is not very clear to which cluster it belongs. Though each object has the same weight in the classification method, including vague or indistinct patterns. The objects with silhouette width lower than zero are now excluded from the classification process, in order to investigate their influence on the results. As a result, between 1% and 13% of the objects are excluded from the sample, dependent on the case. The misclassification error for predictions on the reduced sample decreases with 2 pp on average, and for individual cases it can decrease up to 7 pp. For prediction on the entire sample, the misclassification error remains similar to classification with all the objects, or it increases slightly. Finally, the MAE of the predictions on the entire sample also remain similar. Therefore it is decided to not apply this to the final classification method. When also objects with a small but

positive silhouette width (e.g. 0.05; 0.1; 0.15 and 0.20) are excluded from the data, the misclassification error on the reduced data-set decreases considerably to an average of 16%, but the misclassification error on the entire data-set slightly increases to about 28% on average.

Classification Method Option 3: Likelihood ratio tests with p-value 0.01 vs 0.05

The model selection procedure is repeated with a 95% confidence interval, instead of a 99% confidence interval, for the likelihood ratio test. This effects the selected model in 6/25 cases, where secondary variables (Ws, Rg or Day7) are added to the models. This leads to slight improvements in the misclassification error, but the risk of overfitting the model also increases, as is seen in the few examples. Therefore it is decided to keep 99% confidence intervals.

5.4.4 Seasonal Components

Based on the findings in the previous section, the classification method described in section 5.4.2 is modified with a treatment for small and empty clusters (Option 1) and applied to the All Year cluster analysis data for both Euclidean and constrained DTW measures. The in-sample predictions from the classification model are then used as an input for modelling the seasonal components: all the day-length time series that belong to a certain category are averaged, and all obtained category averages are combined to a seasonal component time series. The trend component is obtained separately, by taking a moving average of the data with daily period, and the components are combined through multiplication. The results of the composition can thus be compared to the basic multiplicative classical decomposition with non-differentiated seasonal components as applied in section 5.2.5, and secondly, to the multiplicative decomposition with seasonal components differentiated after clustering analysis in section 5.3.4.

In the original non-differentiated decomposition, the variance explained by the fitted components (trend + seasonal component) was on average 69%. When using the seasonal components differentiated after clustering analysis, the explained variance increased to on average 76% for the constrained DTW-metric and 81% for the Euclidean metric. Now when using the seasonal components differentiated after clustering and classification, the explained variance decreases to on average 71% for both the Euclidean and DTW-metric. This seems hardly better than the original non-differentiated components, but remark that the actual difference between differentiated and non-differentiated seasonal components is larger, since in the multiplicative decomposition, the fit of the seasonal component on days without energy use, is concealed by the zero trend component. This is illustrated when comparing Figure 5-42 and Figure 5-34 to Figure 5-17. Apart from the seasonal component that fits also in the summer period, also the decrease in residuals in the cold winter period in the beginning of February is noticed. The residuals in the intermediate season are also smaller after clustering, but slightly increase again after classification, due to inaccuracies in the prediction of the seasonal component based on external variables.

The Mean Absolute Error of the model fit (trend + seasonal component) is presented in Figure 5-43 for the non-differentiated and differentiated seasonal components. After differentiation using clustering, the MAE decreases for all cases with 5 up to 50% (on average 22%). When comparing non-differentiation with differentiation after classification, the MAE decreases on average 10% and up to 38%. For a number of cases it remains similar and for case 8 (for which

the seasonal component is very small since no clear daily patterns can be found), it increases with 16%. After classification, the resulting MAE is between 0,3 kWh/2h and 3 kWh/2h.

With regard to the distance metric used in the clustering analysis, the findings from Figure 5-44 confirm that the higher level of detail found when clustering with the Euclidean metric, has disappeared after classification, and the variance explained by the different components is slightly lower. Only for a few cases (6, 9 and 17) the Euclidean metric is beneficial over the DTW metric.

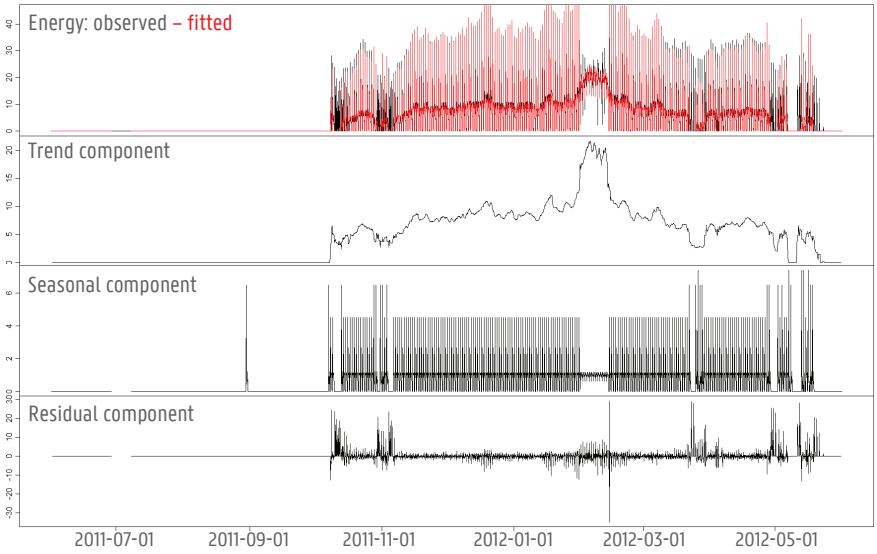


Figure 5-42: Multiplicative TSD with Seasonal Component differentiated by clustering and classification – case 18 – full year

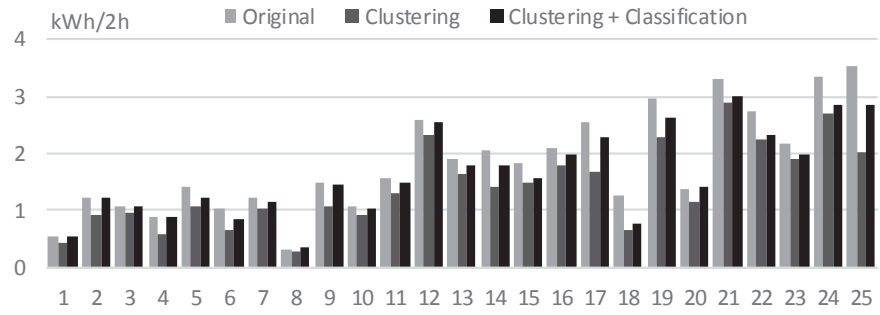


Figure 5-43: Composition with original, clustered and classified SCs - MAE

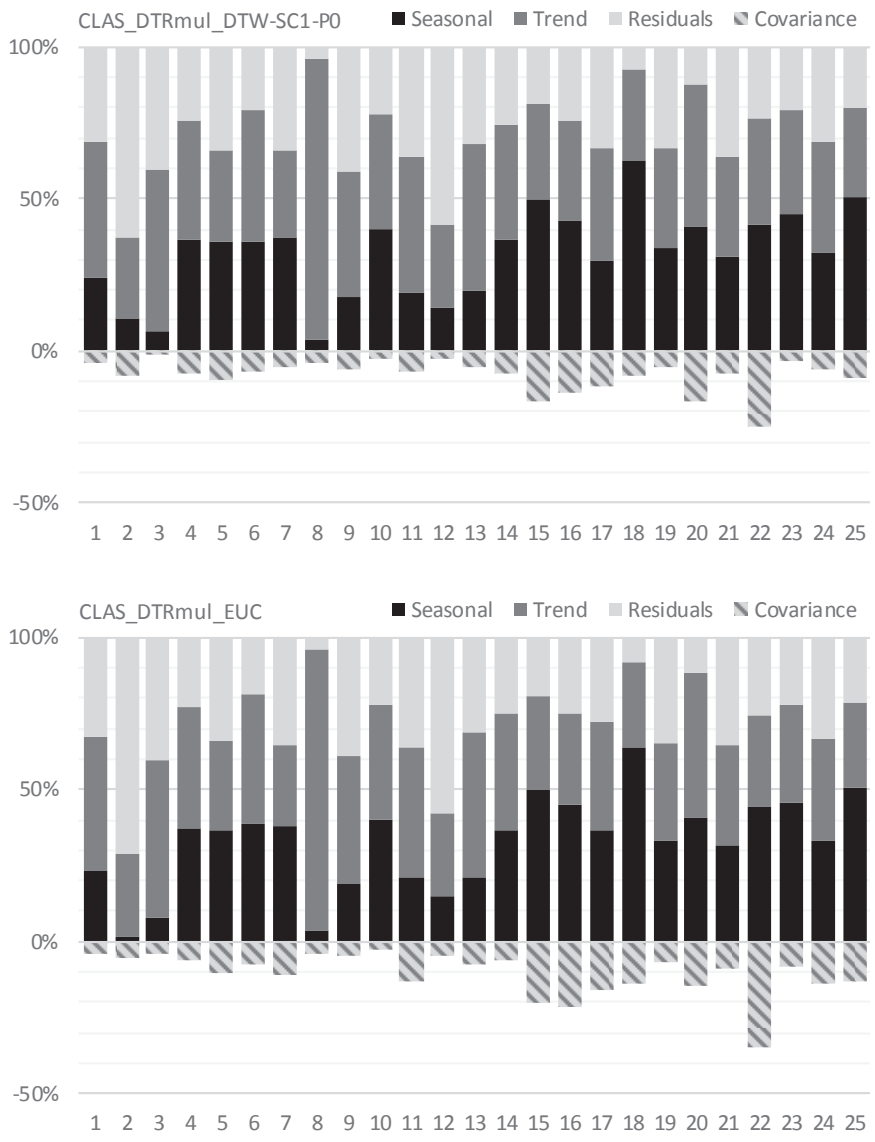


Figure 5-44: Composition with classified data – Full Year – Explained/Observed Variance

5.4.5 Examples

Based on the classification of the Energy Use Time Patterns (using constrained DTW) the 25 cases can be allocated to three groups (see Table 5-4). The first group contains the cases where the Energy Use Time Patterns are characterised by the weather variables, mainly exterior temperature and solar radiation. An example with a very regular Energy Use Time Pattern and well-defined clusters is case 18 in Figure 5-45, where the 5 clusters are characterised by the exterior temperature and solar radiation. Since cluster 3 contains only 5 objects after classification, the deletion of this cluster could reinforce cluster 2 and make the sample more balanced. One of the most irregular examples of this group is case 12 (see also Figure 5-1) for which the clustering and classification is presented in Figure 5-46. The cases in the second group are characterised by both weather and calendar variables, for example case 17 in Figure 5-47. Cluster 1 and 2 represent Weekend days and Wednesday, and Cluster two represents the other Weekdays. This is one of the cases where the Euclidean metric leads to improved results, by further separating cluster 1 in clusters that are more likely to occur on Saturday+Sunday from those on Wednesday. The third group contains two exception cases: case 2 and 8. For case 2 (Figure 5-48), which is irregularly occupied, the Energy Use Time Patterns can be clustered roughly, but the relationship with the only significant external variables is weak and unreliable. For case 8 (Figure 5-49), both clustering and classification fail. This is the case where no clear Energy Use Time Patterns can be observed. For these cases, the use of clustering and classification approaches is not advantageous.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Te	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	X
Rg	x		x	x	x	x	x		x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x
Ws				x								x										x	x	x	
D7	x				x	x	x		x				x				x		x						x
	=	≡	−	−	=	=	=	≡	=	−	−	−	=	−	−	−	=	−	=	−	−	−	−	−	=

Table 5-4: Classification of Energy Use Time Patterns: External Variables

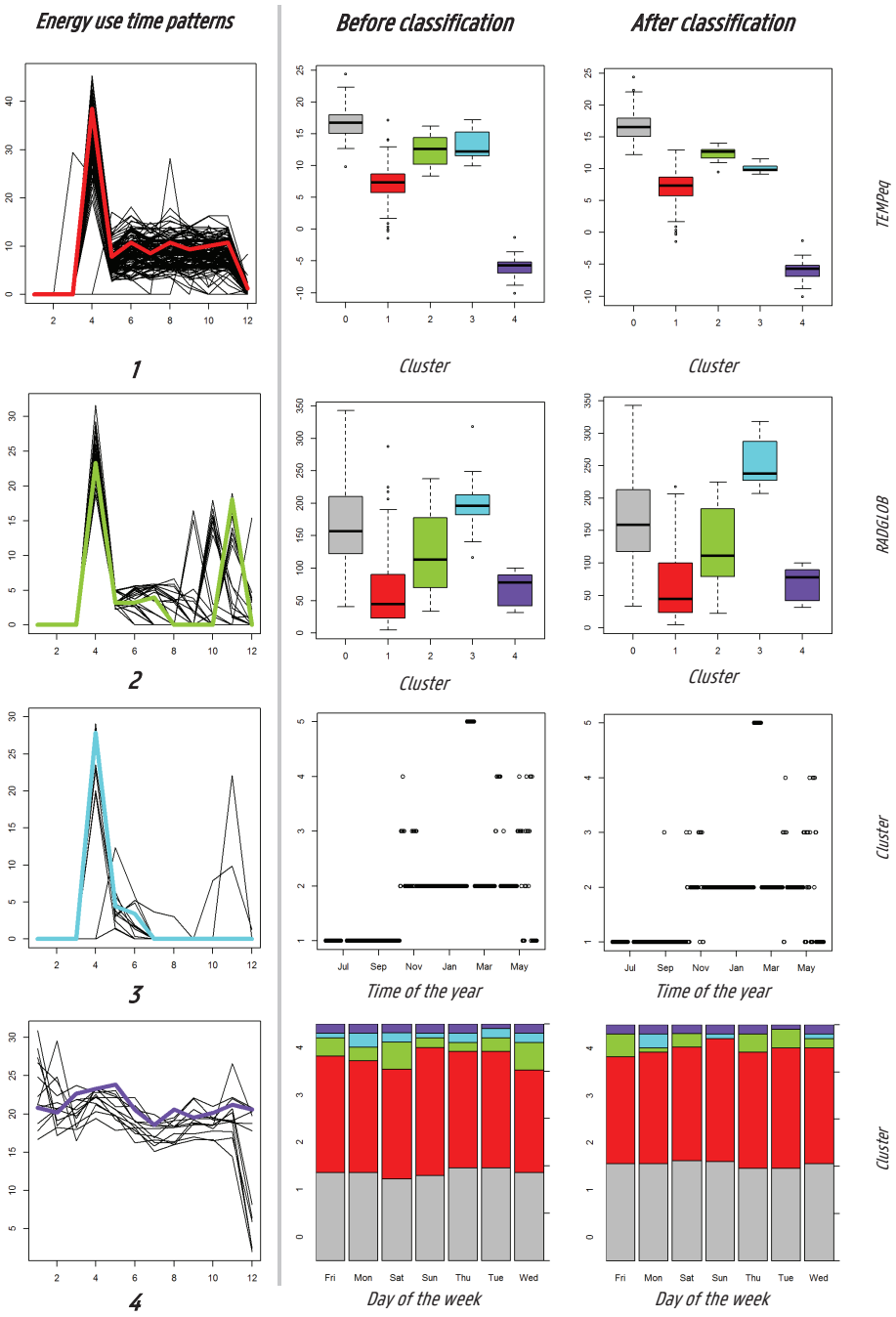


Figure 5-45: Case 18 – Energy Use Time Patterns Classification

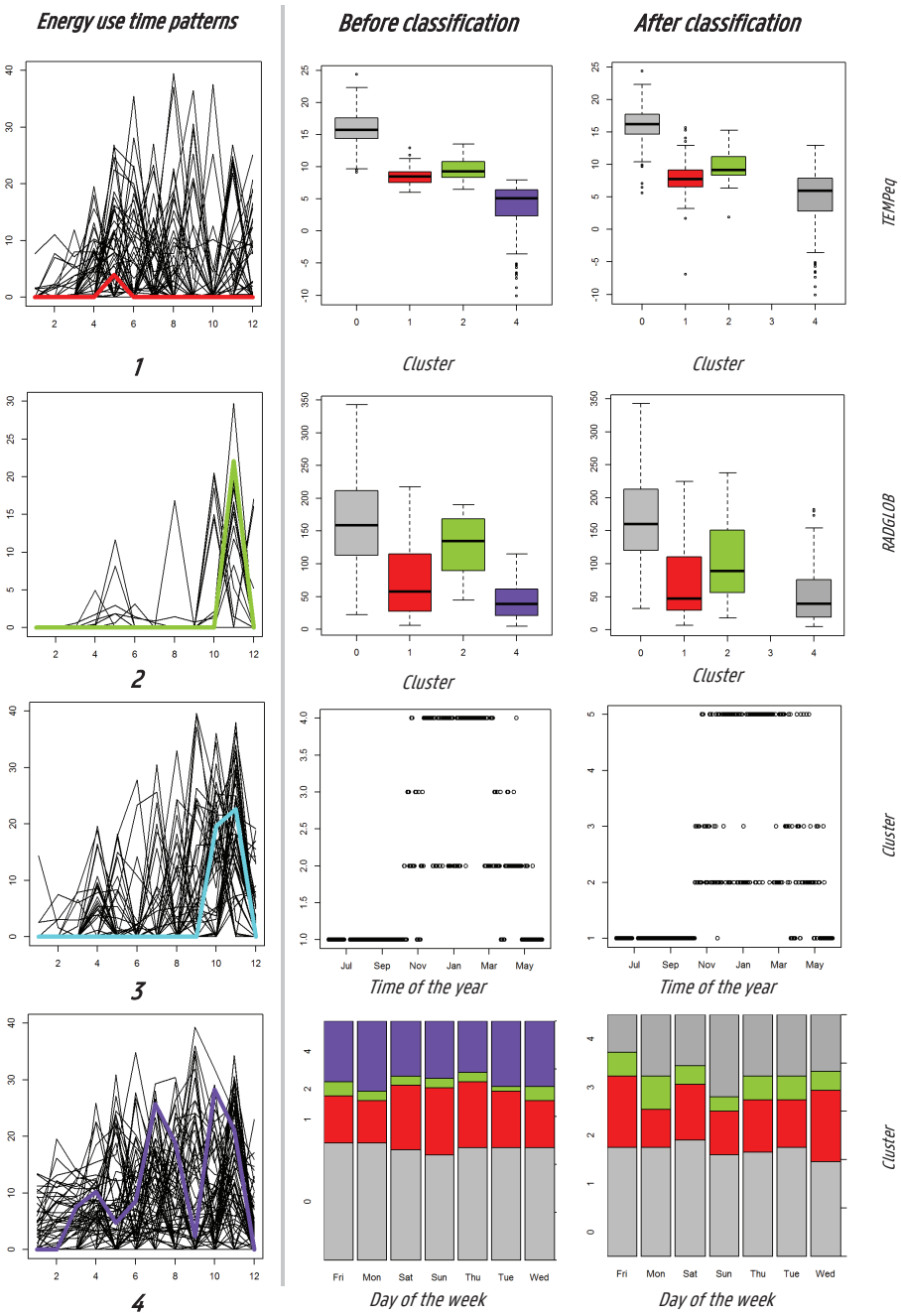


Figure 5-46: Case 12 – Energy Use Time Patterns Classification

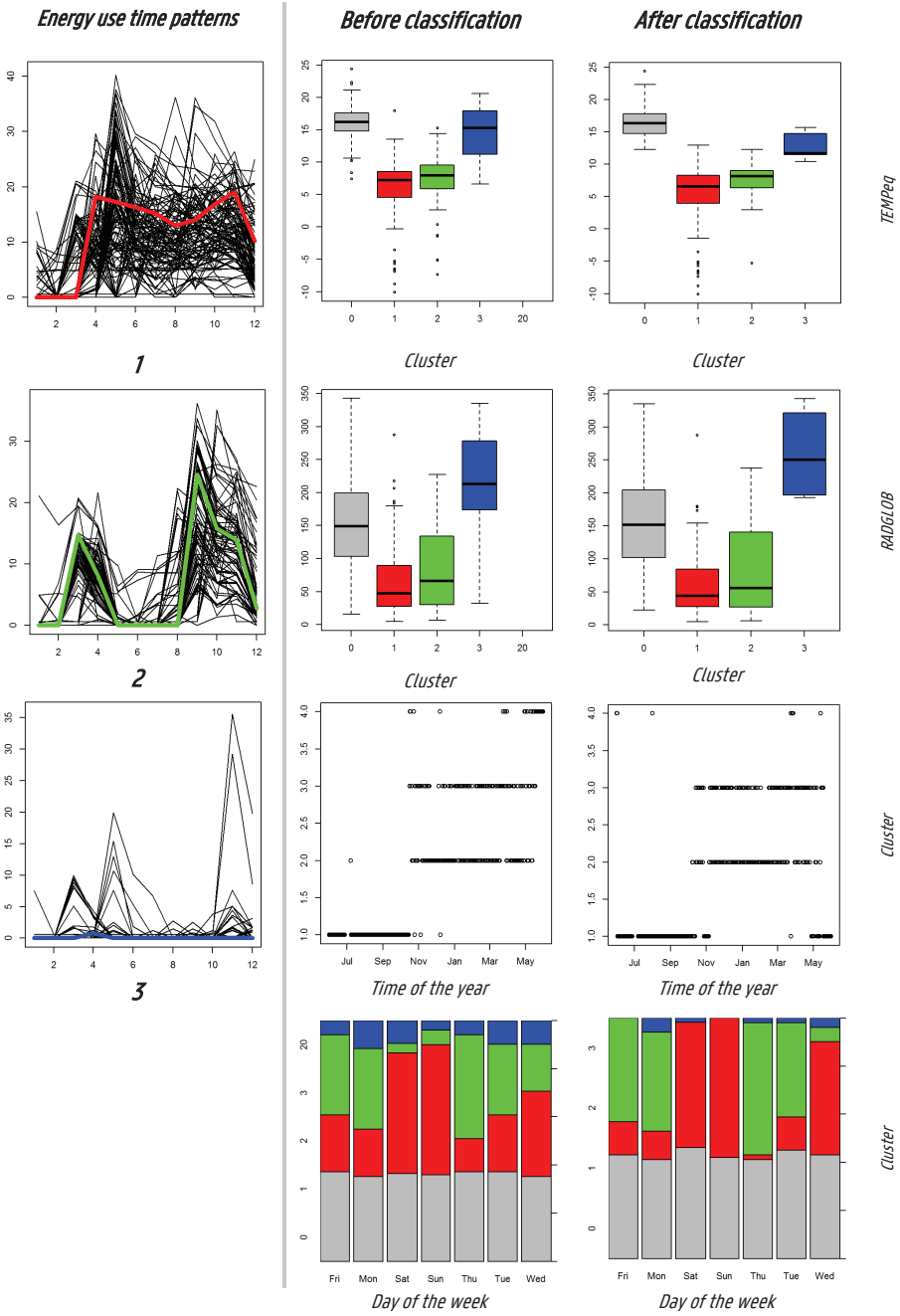


Figure 5-47: Case 17 – Energy Use Time Patterns Classification

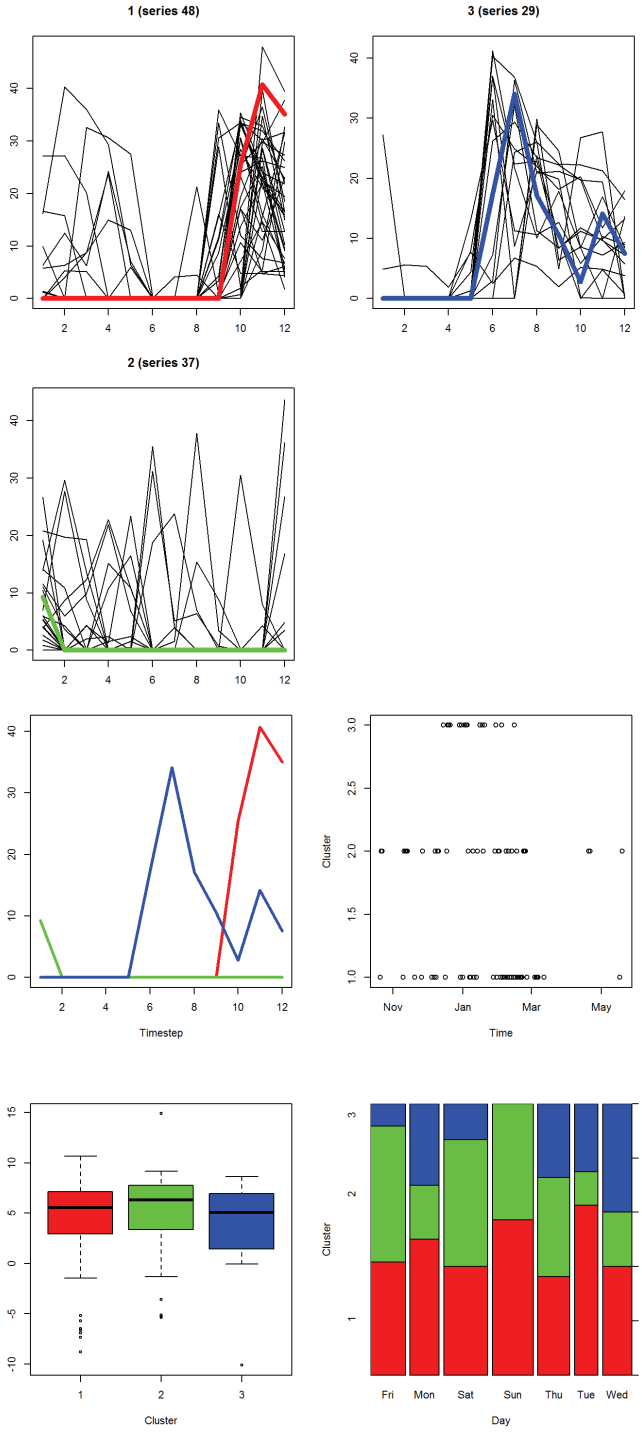


Figure 5-48: Case 2 – Clusters of Energy Use Time Patterns

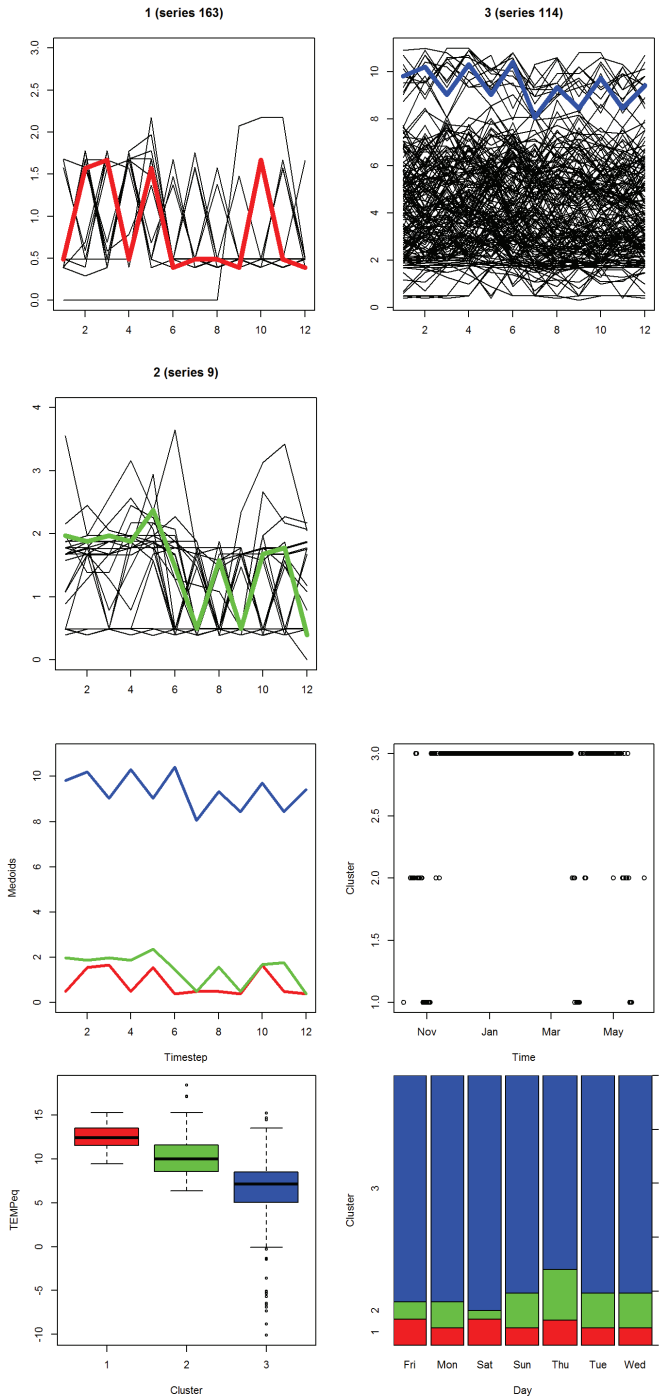


Figure 5-49: Case 8 – Clusters of Energy Use Time Patterns

5.5 Linear regression with clusters of energy use time patterns

Changes within the system that is composed of the building, services and users can influence the relationship between this system and the weather variables when looking to the daily aggregated energy use data. For example variations in the settings of the heating system control that appear for different days of the week, cause different aggregated energy use figures, independent of the weather conditions. Some of these internal differences in the system can be identified by use of the clusters of energy use time patterns that were defined in section 5.3. By use of a categorical variable, these clusters can be taken into account in the linear regression model, and allow the coefficients of the model to be diversified for the different clusters. Note that not all clusters of energy use time patterns will cause relevant distinctions in the overall model. For example the clusters in the intermediate season (e.g. Case 17, cluster 3) may simply be explained by the higher exterior temperatures or higher solar gains that cause shorter times required for the space heating system in order to maintain the indoor temperature. Note also that not all internal differences in the system will be covered by the identified clusters of energy use time patterns. For example if the indoor temperature is decreased a few degrees, or the windows are opened more or less regularly, this might not be observed in the patterns. In this chapter, the inclusion of these clusters of energy use patterns as predictors in linear regression models is investigated, and compared to the linear regression and ARX-models.

5.5.1 Variables and Validity

In order to enable the comparison with the linear regression and ARX-models in Chapters 3and 4, the heating season 2011-2012 is selected from the daily aggregated energy use data for Year 1 (June 2011-May 2012). The linear regression model consists of weather variables, cluster variables and possibly auto-regressive variables. The weather variables include those selected in the linear regression base model (equation 3-III): *Te*, *Rg*, *Ws* and Te_{t-1}. The auto-regressive term considered is the 1st order auto-regressive term Q_{t-1}. The clusters of energy use time patterns are entered into the linear regression model by use of categorical variables. The clusters are obtained from the Cluster Analysis using multiplicatively detrended data and the Dynamic Time Warping measure with constraints (DTRmul_DTW-SC1-P0). The original clusters are used (before classification in function of external variables), obtained from clustering on the data for Year 1. They can be entered as intercept terms or as interaction terms with the other variables. The model validity and model selection are guided by the same measures as used in the linear regression and ARX-models, as introduced in section 3.1.

Reference LM- and ARX-models

As discussed in section 3.3.2, the daily energy use data for the heating season may contain a number of days on which the energy use is zero, for example because the occupants are on holiday, or because of a warmer period during the heating season. In the linear regression models, all Zero-values that are considered outliers (based on the Cook's distance) were deleted from the data. In this section, given the Zero- and Standby clusters from the cluster analysis, it makes sense to exclude both clusters from the data. Figure 5-50 shows the adjusted R^2 values for both situations for the reference linear regression model, that includes *Te*, *Rg*, *Ws* and Te_{t-1}.

For most cases, there is no difference between the two models, or it is below 0.01 pp, but for a few cases, e.g. case 4 and 9, more than 10 zero-values were additionally deleted in the new situation, and the coefficient of determination decreases with 0.08 and 0.06pp. For these two cases, the slope of the regression line may have been influenced by these zero-points. For the ARX-models, the same conclusions are drawn. The coefficient of determination is on average 0.77 for the reference LM-models and 0.80 for the reference ARX-models.

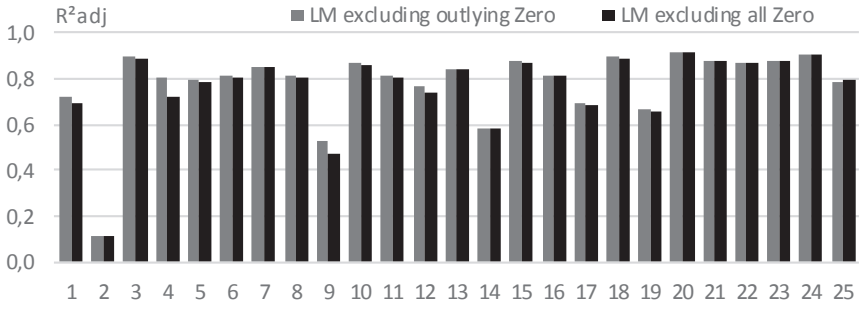


Figure 5-50: LM-models: effect of excluding outlying or all zero-values

Linear Regression and ARX-models with Clusters (LMC and ARXC)

The linear regression models in which the results of the clustering of energy use time patterns are used, or abbreviated: LMC-models, consist of a selection of the four base weather variables and the cluster variable, which is a categorical variable. The selection of weather variables is case-specific, so that only weather variables significant at the 95% confidence level, are included in the model. In comparison to the base model with all base weather variables, this has a negligible effect on the coefficient of determination. Since the Zero- and Stand-by clusters are eliminated from the data, the cluster variable takes values from 1, 2, 3... to *n*, the total number of clusters identified. One level of the cluster variable is used as a reference level or cluster, and the other *n-1* levels are dummies in the regression model. In this study, the cluster that represents the highest number of objects, is selected as reference cluster.

LMCI-MODEL

In a first LMC-model, called LMCI, the cluster variable is simply added as a new variable to the model, leading to equation 5-VIII. Thereby the effect of the cluster variable is an effect on the intercept term of the model. For all cluster levels that differ significantly from the reference level, the intercept coefficient will change, while the coefficients for the weather variables remain. In a 2-dimensional situation (e.g. with Q expressed in function of Te) the intercept of the regression line will change, while the slope of the regression line remains equal.

$$Q_t = c_1 + c_2 \times Te_t + c_3 \times Rg_t + c_4 \times Ws_t + c_5 \times Te_{t-1} + c_{6\rightarrow(n+6)} \times C_{1\rightarrow n} + \varepsilon_t \qquad [5\text{-VIII}]$$

For the LMCI models, the cluster variable is accepted to be significantly different from zero (at the 95% confidence level) for at least one cluster level for 24/25 cases, that is for all cases but case 9. For these 24 cases (see Figure 5-51), the coefficient of determination increases with on average 0.06 pp.

LMCIS-MODEL

A second LMC-model, called LMCIS, enables diversification of both the intercept coefficient and the other (weather) coefficients for different cluster levels. This is achieved by adding interaction terms to the model, that are a combination of the cluster variable with the weather variables. At least one of these terms is significantly different from zero (at 99% confidence intervals) in 11/25 cases. For these cases, the coefficient of determination increases with .01 to .03 pp when compared to the LMCI-model.

ARXCI- AND ARXCIS-MODEL

Next, it is also possible to add auto-regressive terms to the LMCI- or the LMCIS-model, which are then called the ARXCI- and ARXCIS-model. The AR(1)-term is accepted to be significantly different from zero in about 20/25 ARXCI-models, and in 12/25 ARXCIS-models, including three for which the interaction terms were not found significant in the LMCIS-model (but they do in the ARXCIS-model). For these cases, the ARXCI and ARXCIS-models have an equal or higher coefficient of determination than the ARX-models (on average +0.03pp) and then their respective LMCI- and LMCIS-models (on average +0.04). The average values for the coefficient of determination for all 25 cases are summarised for the different model types in Figure 5-52. When comparing the original LM- and ARX-models to the alternatives with clustering variables, the adjusted R^2 increases on average 0.05 and 0.06 pp (for both CI and CIS-terms). For individual cases, the differences may be small or may rise up to 0.17 pp (when disregarding case 2, where changes up to .40 pp are seen).

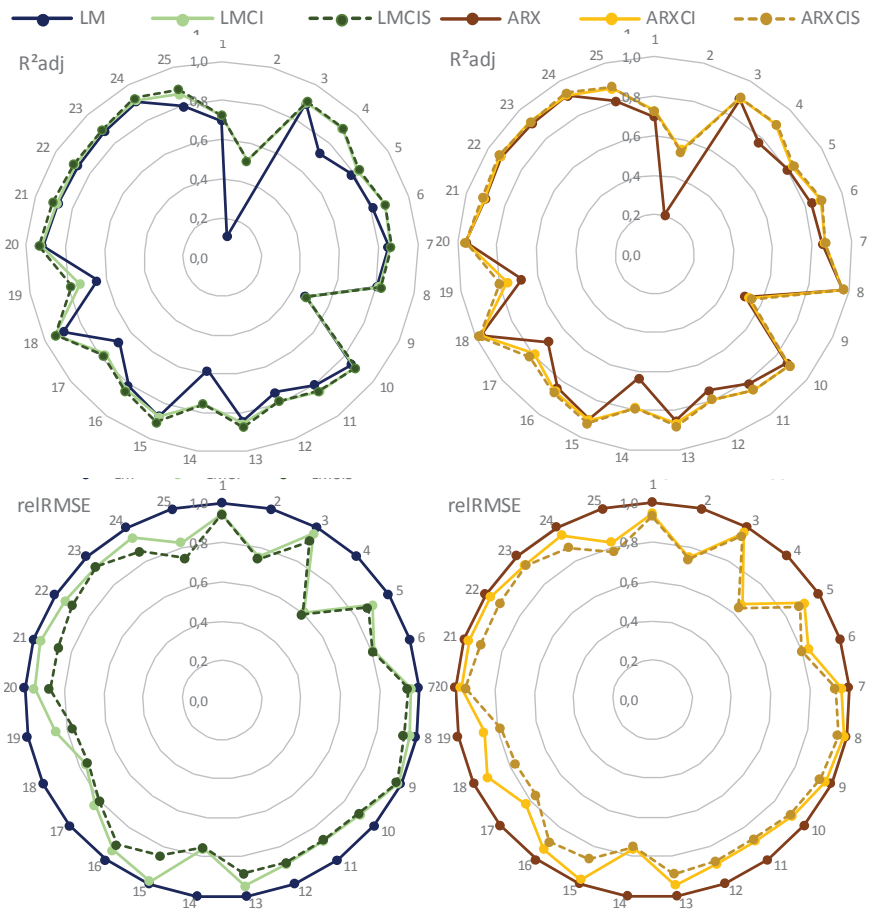


Figure 5-51: LM , ARX and LMC-models – relRMSE and R²adj

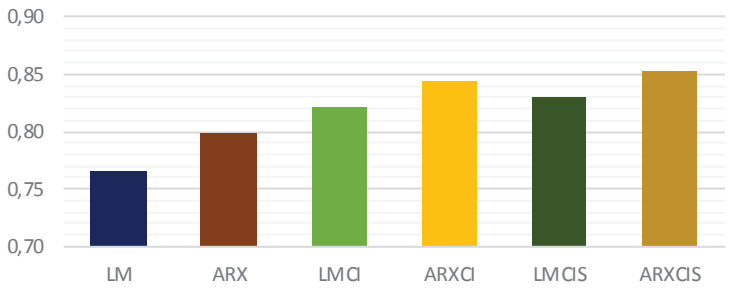


Figure 5-52: LM, ARX, LMC and ARXC-models – average R²adj

Model Validity

By checking the model validity using the statistical diagnostics plots (see section 3.1), the effects of the different model types can be better understood. In section 3.7.2 and 4.2, the findings on the validity of the LM- and ARX-models are summarised, and two types of auto-correlation of the residuals of the LM-models were determined. The first type of auto-correlation goes together with a slow decay in the auto-correlation function, and for the second type of autocorrelation, peaks in the ACF occur at specific time lags. The cases can thus be allocated to groups where the 1st type is dominant, the 2nd type is dominant, and a group where no or very weak autocorrelation appears. It is found that conclusions on the (dis)advantage of the different model types against each other, can be related to these properties. In Table 5-5 and Table 5-6 the selected models and model validity findings are summarised for all model types.

Table 5-5: Model Selection – HS 2011-2012 – suitable models (X) and selection per case (X)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity																•		*	*						*
LM66	X	X	X			X	X					X	X												
ARX50A	X			X			X					X	X	X	X	X			X	X	X	X		X	
ARX50B				X							X										X	X	X		
ARX50C								X									X	X							
ARX50D								X																	
ARX50E					X	X			X	X			X									X		X	X
ARX50F																									X
LMCI	X	X	X	X	X	X	X			X	X	X		X											
ARXCI		X				X			X	X	X			X	X	X		X	X	X		X	X		
LMCIS			X		X		X						X											X	X
ARXCIS								X					X		X	X	X	X	X	X	X	X		X	
GROUP	I		I	I	I	I	I	II	I	I	I	I	I	I	I	I	I	II	I	II	II	II	II	I	I

Table 5-6: LM, ARX, LMC and ARXC-models: Model Validity and Autocorrelation – HS 2011-2012

LM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Validity	1	0	1	0	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
AC TYPE 1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1	1
AC TYPE 2	1	1	1	0	,5	1	1	1	,5	,5	,5	1	1	,5	0	0	0	1	0	1	1	1	1	1	0
ARX																									
Validity	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	0
AC	1	1	1	1	1	1	1	1	1	1	,5	1	1	,5	1	,5	1	1	,5	1	1	1	1	1	,5
AC lag											(8)			(7)		(5)			(4)						(5)
LMC																									
Validity	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	0
AC	1	1	1	,5	1	1	1	1	1	1	1	1	1	1	1	,5	,5	,5	1	1	,5	1	1	1	,5

GROUP I: CASES WITH AC TYPE 2 OR WEAK AUTOCORRELATION

First observe the cases where the 2nd type of autocorrelation appears, together with the cases where no or weak autocorrelation appears. Together, 19/25 cases are considered. For these cases, it is found that the LMCI or LMCIS-model performs equal or better than the selected ARX- or LM-model with regard to the model validity and goodness-of-fit. For most cases no or weak autocorrelation remains after applying the LMCI or LMCIS-model, and an additional AR(1)-term is not needed, though it can slightly optimise the results. For some of the cases, including a number of cases with clear autocorrelation of the second type (e.g. case 4, 25, 14, 17 and 19), autocorrelation is not entirely removed and additional autoregressive terms (AR(1) to AR(3)) are useful for the validity of the model.

This group includes the cases where in the ARX-models an AR(7)-term was used (ARX50E) was used and all cases where in the classification of energy use time patterns the *Day*-variable was included in the logistic regression model. These cases result in the highest improvements in the goodness-of-fit of the LMC/ARXC-models when compared to the ARX- or LM-models, for example case 14, 17, 19 and 25 result in an increase between .08 and .17 pp. (Remark that case 14 was not classified with the Day-variable, although in the original clusters, differences in patterns over different days were found).Apart from these, it also includes cases where the energy use time patterns were found mainly dependent on the weather conditions, but for these groups, the increase in adjusted R^2 is below .05 pp.

One exception in this group is case 9, where the AR(1)-term contributes more to the goodness-of-fit of the model than the cluster variables. When observing the clustering results for this case, it is found that this is one of the cases where only a limited amount of the variation in the data was explained by the seasonal components for the constrained DTW-measure (Figure 5-33), while there are considerable differences in energy use time patterns for different days. Like cases 6 and 17, it is also one of the cases where by use of the Euclidean metric more different energy use patterns are recognised and classified (Figure 5-41).

GROUP II: CASES WITH AC TYPE 1

For the 6/25 cases where clear autocorrelation of the 1st type was recognised in the linear regression models, it is found that the ARX-models perform better than the LMC-models: the auto-regressive terms are needed to take care of the autocorrelation in the model, and the models lead to a better goodness-of-fit. For some cases (e.g. 8 and 20), the cluster variables are of no use for improving model quality. The most clear example is case 8, where no clear energy use time patterns appear (see Figure 4-4 and Figure 5-6). For other cases, the combination of cluster variables into ARXCI or ARXCIS-models is valuable, for example for case 18.

VALIDITY

Table 5-6 summarises the findings on model validity for the selected LM, ARX and LMC/ARXC models. Looking to the auto-correlations for some of the ARX-models, limited autocorrelations remain after applying the ARX50A to ARX50F-models, that include limited AR-terms, for example not the AR(4) and AR(5) terms. Though these can be excluded by adding more terms to the models. For the ARXC-models, only the AR(1)-term was applied, and in the validity table, all cases are mentioned where this is not found sufficient. For those cases, addition of especially AR(2) and AR(3) terms is expected to take away the remaining autocorrelations. Remark that for many other cases, where several AR-terms were needed in the ARX-model, the auto-regression is now removed by use of the cluster variables.

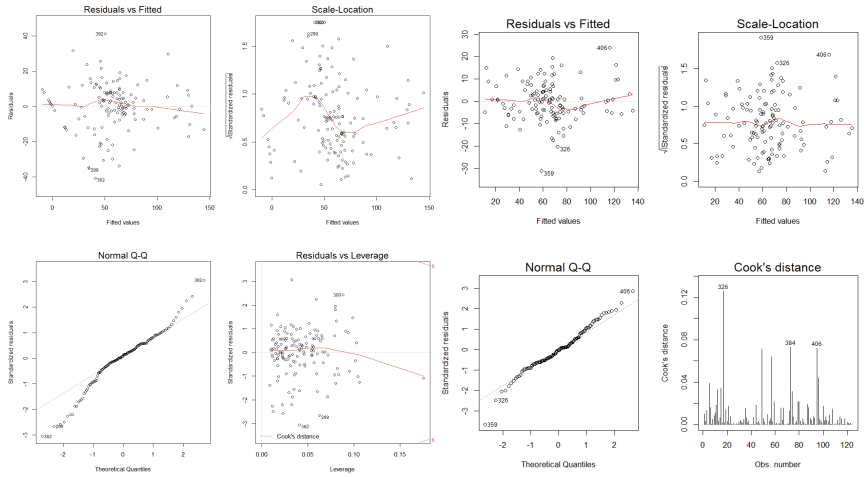


Figure 5-53: Case 4: ARX50B (left) vs LMCB (right) – Diagnostic plot

When using LM- or ARX-models for case 4, the distribution of the residuals was not normal and non-constant variance was observed. These issues disappear after using the regression with cluster variables (Figure 5-53). For cases 16, 18, 19 and 25 non-constant variance was also observed for all models considered. For cases 18 and 19 it appears that in the coldest winter period (and for case 19 also in the month of November), the variance in the residuals is higher than in the rest of the winter (Figure 3-11). For case 16 and 25 the variance is also higher during (a longer) mid-winter period.

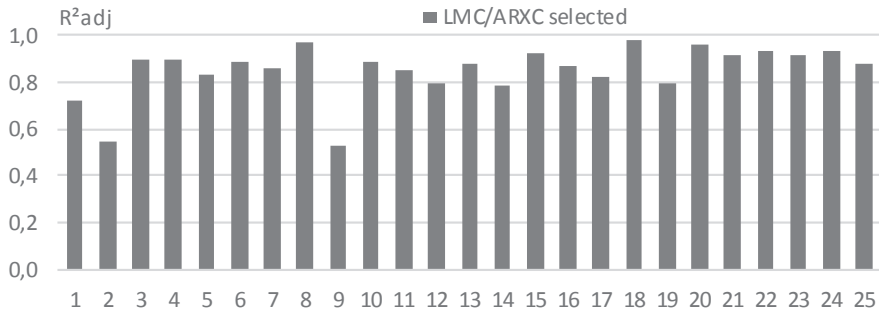


Figure 5-54: R²adj – LMC/ARXC selected

In Figure 5-54 the adjusted R^2 is presented for the case-specific selection of an LMC or ARXC model. In general, it is concluded that for model where the first type of autocorrelation dominates (group II) it is necessary to use auto-regressive terms and thus start from an ARX-model. For some of these models, additional clustering variables may optimise the goodness-of-fit. For the majority of the cases in this study, where the 2nd type of autocorrelation dominates, or weak autocorrelation appears, it is found that the models with clustering variables are preferred, although auto-regressive terms may be required or useful for reasons of model validity or for optimisation of the model.

Cluster variable

In many cases the cluster variable is only significant if added as an intercept term to the model, and for most cases, it is not found significantly different from zero for all the cluster levels. If an interaction term is found significant, it is typically for the interaction with Te or the 1st auto-regressive term. Only in one case it is for the solar radiation (case 4), and wind speed (case 25). By way of example, the results are summarised for 2 cases for which the clustering results were documented in section 5.3.3.

In case 6 (Figure 5-31 and Figure 5-35), where three clusters were detected and the 2nd cluster is the main cluster, it is found that both clusters 1 and 3 have a different intercept. The energy use time patterns in cluster 2 appear on days that the space heating is used for most of the daytime, which seems more likely to appear on weekend days in this case. For the energy use time pattern 1, the space heating is (temporarily) off before midday, which is more likely to appear on weekdays. The 3rd cluster's energy use time pattern signifies periods in which the heating is turned on day and night, which appears only in very cold periods. These different ways of controlling the space heating system, will influence for example the indoor temperature conditions in the dwelling during the day and night, or how heat is stored and released from the thermal mass of the building. In this case, only in the intercept term significant differences are found.

In case 18 (Figure 5-45) the first cluster is the main cluster, that appears most time of the winter, and where a night set-back period is recognised. The 2nd and 3rd cluster appear in the warmer periods of the winter, where the heating is activated only in the morning and/or the evening. When observing the plots of the clusters in function of the weather conditions, it is expected that this is not because of absence of the occupants, but very probably because during daytime, the exterior temperature and solar radiation are sufficient to heat the house, and thus the heating system turns off automatically. It is even recognised that cluster 3 appears typically on days with higher solar radiation, and thus no space heating is needed in the evenings either. In the linear regression with clustering, nog significant differences are found between these three clusters, so the differences are explained by the main model coefficients. Like in case 6, during the coldest winter period, the night set-back is not activated (or not reached), and for this situation, a significant difference in the intercept term, the temperature coefficient and the auto-regressive term is recognised (in contrast to the situation in case 6).

5.5.2 Energy Signature Coefficients

EXAMPLE: CASE 25

Case 25 is one of the cases where the use of the LMC-model leads clearly improves model quality and goodness-of-fit of the model. When observing the clusters of energy use time patterns in Figure 5-55, obtained by using the clustering algorithm with the constrained DTW-measure on the data where Zero- and Standby clusters are excluded, three energy use time patterns are observed. The first one appears at relatively high exterior temperatures and solar radiation values, while the second and third one appear in the same ranges for the observed weather variables, but cluster three is more likely to appear on Wednesday, Saturday and Sunday.

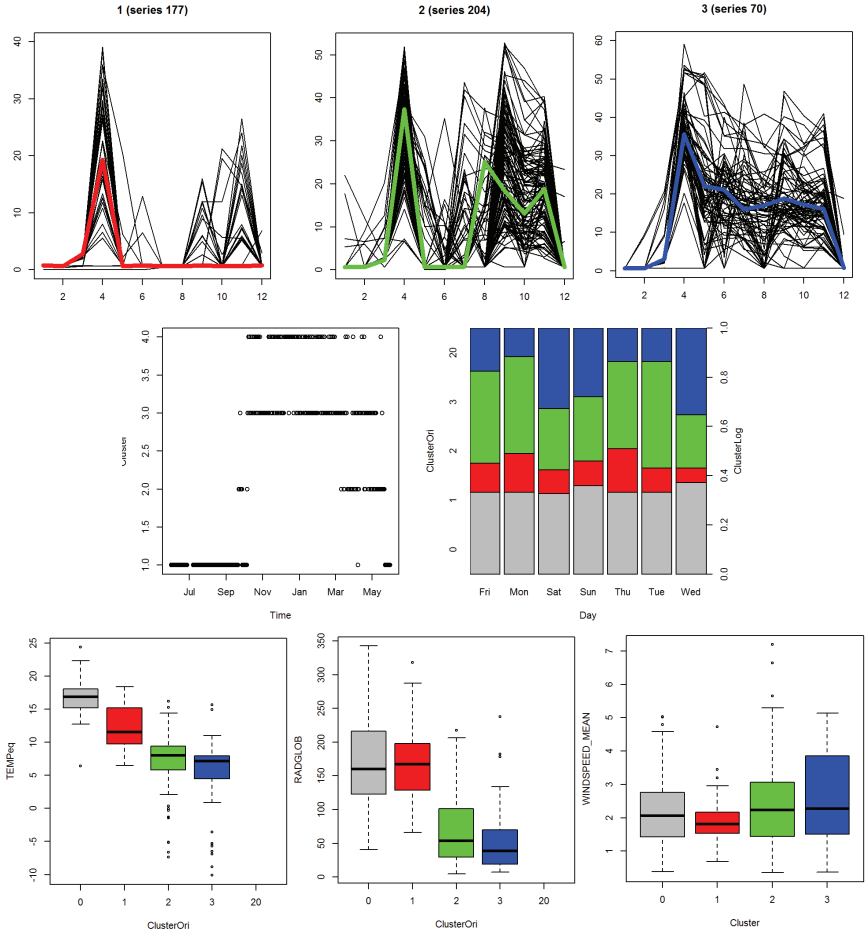


Figure 5-55: Case 25 – Energy Use Time Patterns Clustering

Table 5-7: Coefficient Statistics – Case 25 – Heating Season 2011-2012 – ARX and LMCIS

ARX		Coefficients						Partial correlations	
	Unit	Estimate	Std. Error	t value	p-value	Standardised coefficient	VIF	Estimate	
(Intercept)	kWh	135,69	12,17	11,15	0,000				
RADglob	kWh/(W/m²)	-7,43	0,63	-11,84	0,000	-0,18	1,3	-0,34	0,000
TEMP	kWh/°C	-0,20	0,04	-5,12	0,000	-0,57	2,5	-0,64	0,000
WINDSPEED	kWh/(m/s)	6,07	1,62	3,74	0,000	0,12	1,2	0,24	0,000
Energy.l1	kWh	0,18	0,05	3,49	0,000	0,18	3,0	0,23	0,000
Energy.l7	kWh	0,17	0,04	4,54	0,000	0,18	1,6	0,31	0,000

LMCIS		Coefficients						Partial correlations	
	Unit	Estimate	Std. Error	t value	p-value	Standardised coefficient	VIF	Estimate	
(Intercept)	kWh	176,76	6,16	28,70	0,000				
RADglob	kWh/(W/m²)	-0,17	0,03	-5,11	0,000	-0,15	1,2	-0,45	0,000
TEMP	kWh/°C	-4,20	0,78	-5,36	0,000	-0,32	2,5	-0,46	0,000
TEMP.l1	kWh/°C	-4,07	0,69	-5,86	0,000	-0,32	2,2	-0,28	0,000
WINDSPEED	kWh/(m/s)	6,55	1,66	3,94	0,000	0,14	1,4	0,32	0,000
C1	kWh	-29,17	27,13	-1,07	0,284	-0,14	3,6		
C3	kWh	37,63	7,66	4,91	0,000	0,30			
TEMP x C1	kWh/°C	3,36	2,44	1,38	0,170	0,17	3,1		
TEMP x C3	kWh/°C	-3,50	0,67	-5,25	0,000	-0,23			
WINDSPEED x C1	kWh/(m/s)	-15,09	4,95	-3,05	0,003	-0,16	2,3		
WINDSPEED x C3	kWh/(m/s)	5,48	2,48	2,21	0,029	0,13			

The coefficient statistics for case 25 are presented for the ARX-model with AR(1) and AR(7) and for the LMCIS-model in Table 5-7. In the ACF of the ARX-model small auto-correlations remain at the 5th time lag, and non-constant variance and (small) deviations from normality were observed at the tails of the Q-Q plot. In the LMCIS-model, small auto-correlations remain around the 7th lag, normality improved and the variance increases less (but still) with increasing energy use. The LMCIS-model has a goodness-of-fit of 0,88 (adjusted R^2), while for the ARX-model it was 0,81. The second cluster is the largest cluster and thus the reference cluster of the model. When adding cluster and cluster interaction variables to the model, it is found that the variables for the cluster and its interactions with Te and Ws are significant variables. More specifically, it is found that the p-values for the coefficients related to the third cluster are smaller than 0.05, so they are accepted to be significantly different from the estimates of the reference cluster. Since also the same ranges of weather variables appear for cluster 2 and 3, it is concluded that both clusters are described by a regression function with different coefficient values for the intercept, Te and Ws. The variables for the 1st cluster are not significant at the 95% confidence intervals, except for the interaction with wind speed Ws, for which the range of observations in cluster 1 lies within the lower part of the range for the reference cluster. While for cluster 2 and 3 the wind speed coefficients indicate increasing energy use with increasing wind speed, the coefficient for the 1st cluster (-8.55) is negative and indicates a decrease in energy use with increasing wind speed. Since the 1st cluster is valid for lower amounts of wind speed in combination with relatively higher exterior temperatures, this change in coefficient might be explained by the fact that the

impact of wind speed on the energy use for space heating is related to the temperature of the air which infiltrates in the building due to the wind pressure. The steady-state coefficients for the ARX-model and for the reference cluster of the LMCIS-model are presented in Figure 5-56. The confidence intervals for the coefficients of the LMCIS-model are smaller and thus more accurate.

All Cases

Figure 5-56 presents the steady-state coefficients for the selected LM- or ARX-model and the selected LMC- or ARXC-model. Since the estimates for the LMC-or ARXC-model are only for the reference cluster, the estimates cannot be compared to the estimates of the LM- or ARX-model, and so it is not an issue that they differ in many cases. For example in case 25, the estimate for the intercept in the ARX-model is in between the estimates for clusters 2 and 3 of the LMCIS-model. The confidence intervals of the estimates are often smaller than for the LM-and ARX-models: in 19/25 cases for the intercept coefficient, in 11/25 cases for the temperature coefficient, in 17/25 and 22/25 cases for the solar radiation and wind speed coefficients.

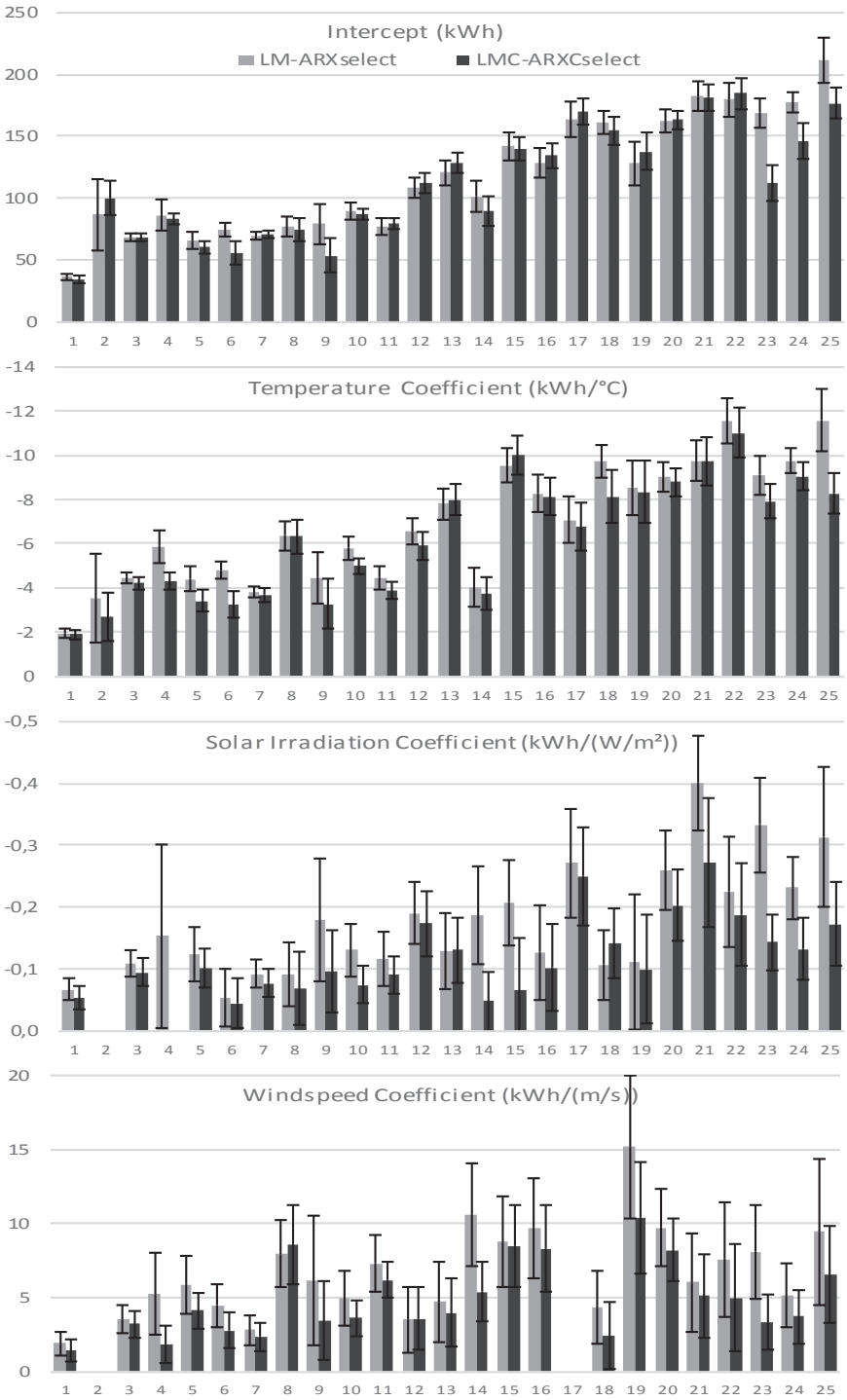


Figure 5-56: Energy Signature: Coefficients and 95% CI

5.6 Comparison of 2 periods

By use of the results of the linear regression with energy use time patterns, the differences in energy use between two periods are compared on two levels: the classified clusters of energy use time patterns, and the coefficients of the linear regression model. For both cases, an interaction model is created, with a qualitative predictor variable PE, that identifies to which period each data point belongs (see section 4.4). In this section, the energy use for Year1 of the data (June 2011-May 2012) is compared to the energy use for Year2 (June 2012-May 2013). The results are compared to the findings from section 4.4, where the heating seasons for these years were compared using selected LM- or ARX-models.

Table 5-8:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
***Model Validity** (§ if not considered valid; § if smaller deflections)*																									
LM / ARX HS 1+2 (IA)	§	§		§		§			§						§	§	§	§	§						§
LMC / ARXC Y1+2 (IA)	§	§					§		§						§	§	§	§	§						§
***Comparing the energy use in the two heating seasons for selected LM/ARX-model** (X if changes are detected)*																									
Coefficients	X			X			X							X			X								
Coefficients Interaction	X		X	X		X	X	X	X	X				X		X	X	X			X	X		X	
Sign. Level	0,001	1	0,001	0,01	1	0,05	0,010	0,001	0,001	0,05	1	0,1	0,01	0,001	1	0,01	0,001	0,001	1	1	0,001	0	1	0,01	1
Conclusion	X		**X**	**#**		**#**	**X**	**X**	X	**X**				**X**		**#**	X	**#**			**X**	**X**		**X**	
Comparing the energy use in the two heating seasons for selected Classification and LMC/ARXC-models																									
Patterns Interaction	X		X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Visual	X		X						X	X		X		X		X	X	X		X	X	X			
Coefficients Interaction	X		X			X	X	-	X	X		X	X			X	X	X	X	X	X	X		X	
Sign. Level	0,001	1	0,001	1	1	0,001	0,001	-	0,001	0,001	1	0,01	0,01	1	1	0,001	0,001	0,001	0,01	0,001	0,01	0,05	1	0,05	1
Conclusion	**X**		**X**			**X**	**X**	-	**X**	**X**		**X**	**X**			**X**	**X**	**X**	**#**	**X**	**X**	**X**		**X**	

5.6.1 Comparison of Patterns

According to the method described in section 4.4., the energy use time patterns are classified in function of external variables by use of logistic regression models. When comparing the data for two periods, the clustering method and classification method are applied on the entire dataset that includes both periods. After a suitable classification model is identified (with selected weather and calendar variables), the Period predictor variable PE is added to the model as an extra term (intercept) and in interaction with the other predictor variables. For each extra term added to the model, the likelihood ratio test (for nested models) is applied, and only significant terms are kept in the model. If a 99% confidence level is maintained for the selection, in 22/25 cases the period predictor or its interaction with another predictor variable is found significant (see "Patterns Interaction"). Since in the visual inspection of the data for some cases only

negligible differences in clustering results are observed for both periods, these results are doubted, and can be a subject of further investigations. For the remainder of this study, the findings from the significance tests are combined with visual observations of the regression cluster plots and the difference in misclassification errors in the models with and without period variable. In the next section, for drawing definitive conclusions, these findings will also be compared with the findings from the comparison of the linear regression coefficients.

As an intermediary result, in 12 cases the presence of a difference in energy use time patterns between the first and second year is confirmed. For case 9 and 17, a change in energy use is observed within one of the two years. In 4 cases (1, 16, 21 and 22), a change in energy use related to the periods with Zero- or Standby-patterns is found. In case 1, the energy use is zero during a long period of the second heating season, in case 16, 21 and 22 a difference between the Zero- and Standby periods is observed. For cases 18 and 20 the relationship with the weather variables seems different, and for case 10, 12 and 14 a difference in calendar variables (and maybe weather variables) is suspected. For these 12 cases, the misclassification error is reduced between 4 and 40%, for the remaining cases where no clear differences could be observed, the misclassification error is reduced by 0 to 13%.

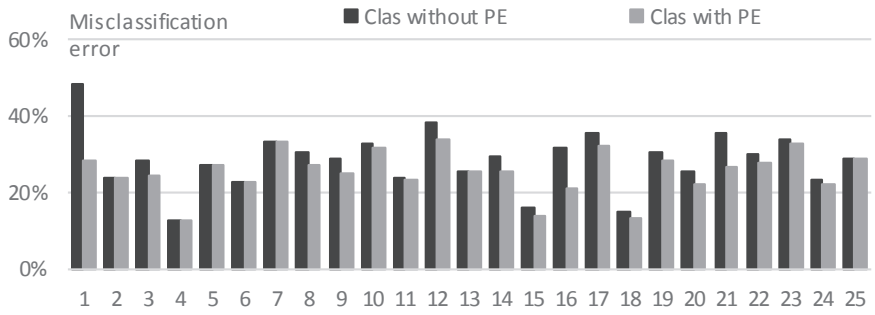


Figure 5-57: Classification for Year 1 and 2, with and without period variable: misclassification error

5.6.2 Comparison of Energy Signature Coefficients

Linear regression models are applied to the daily energy use data, using the clusters of energy use time patterns as additional model variables. Now, the linear regression model is estimated for the two years of data. First a suitable base model is searched for, starting from the model that was selected for the 1st heating season, and improving where possible. Then, the addition of the period variable and its interactions, is investigated. The model selection takes into account model validity, the significance of coefficients and the goodness-of-fit measure R^2adj. The cases for which a model with significant period variable was estimated, are marked in Table 5-8 ("Coefficients Interaction"). For 16/25 cases the period variable (or its interaction) was found significant at the 95% confidence level. For 8/25 cases no significant difference in the energy use model could be found. Finally, case 8 was excluded from the investigation, since it does not present a clear seasonal pattern. For the cases where a change in energy use model was found for the two periods, the model validity generally improves for the selected model with period variable(s), as the deviation of the energy use in a certain period is often causes the residuals of the model to diverge from normality or constant-variance. In total, 10 cases remain where the model assumptions were not completely fulfilled (including 5 cases where the deviations from

the assumptions are relatively small when comparing to the LM- and ARX-models), often in a situation where the variance changes with increasing energy use numbers. For 7 of these cases the period variable was a significant variable, and for those cases, further observation of the results is needed to confirm the difference, for example by observation of the energy use time patterns.

5.6.3 Findings

For most cases, if a difference is observed by comparing the energy use time patterns, this was also observed when comparing the coefficients. For some cases (e.g. case 4, 14), the comparison of patterns indicates a difference between the 1st and 2nd year, which is not detected in the comparison of coefficients. However, this difference is visually confirmed in only one case (case 14). For four cases, according to the comparison of linear regression coefficients, a difference occurs, and in three of them also a difference in patterns was identified, however it was not visually confirmed. The agreement between the findings from both comparisons is a mutual confirmation that differences in energy use appear between the 1st and 2nd year of the measurements. However, such a confirmation is not necessary for accepting a difference in energy use. On one hand, if a certain cluster of patterns becomes more apparent in certain regions of the variables (e.g. on another day of the week), this may remain unnoticed by the linear regression model. On the other hand, the patterns may remain similar and unnoticed, while the relationship of the daily energy use to the weather variables changes, for example due to a change in insulation level or temperature settings.

Using the linear regression models with energy use time patterns, and the classification of the patterns themselves, more differences between the two years are detected than when LM- or ARX-models were used, and there are more indications towards the understanding of what are the occurring differences. In the next section, this is illustrated by use of an example case. Secondly, for most of the cases where a change in energy use was detected in the LM- or ARX-model, this was also detected using the methods with energy use time patterns – if the model assumptions were fulfilled. For example in case 4 a difference in energy use was observed using the ARX-model, but that model was found invalid with regard to the model assumptions. Using the linear regression model with energy use time patterns, a valid model is obtained, but no difference is observed when comparing the coefficients. When comparing the patterns, a difference is observed by the period variable, but after visual observation of the results, it is found that in the clustering results this difference is related to the 2 weakest and neighbouring clusters, and thus it is probable that the detected difference is related to objects that can be interchanged between the clusters. Therefore it is concluded that when looking for further improvements to the comparison method, the classification method could be addressed, and especially the relationship between the findings from the classification and the clustering quality.

A third comparison option, which is not explored in this study, is the comparison of the start and end of the heating season, through application of a logistic regression model with binary 'heating season' variable and explanatory weather variables. The binary 'heating season' variable can be defined by giving the Zero- and Stand-by clusters a value of 0, and the remaining clusters a value of 1.

5.6.4 Examples

Case 17: Changes within Year 1

In section 4.4 a change in the energy use characteristics of case 9 and 17 was suspected within one of the heating seasons. For case 9 this change is detected in winter 2012-2013 (2nd heating season) and for case 17 in winter 2011-2012. This presumption was based on the visual observation of the time series plots of the original and fitted data (using ARX-models). When now investigating the Energy Use Patterns and LMC/ARXC-models, for both cases the differences can be identified. This is illustrated for the example of case 17. For this case, the presumed change was observed around mid December 2011 (Figure 4-8). The clustering and classification of the energy use time patterns is documented in Figure 5-47. In the clustering plots, it is observed that cluster 1 appears especially in the weekend, and also more on Wednesdays, while cluster 2 appears on Wednesdays, but also on other days of the week. This structure was recognised in the classification process, using a model with *Te*, *Rg* and *Day7* as predictor variables and 27% of the objects were misclassified. Now in Figure 5-58 and Figure 5-59 the original cluster plots (before a classification model is applied) are presented for the 1st period and second period of Year 1, before and after December 15th 2011. When comparing the original cluster plots for the 2 periods, it is found that for the 2nd period the partitioning of the clusters into weekend days, Wednesdays and other weekdays, is more explicit than in the plots for the entire year, and that this partitioning is does not apply for the 1st period, where only on Thursdays and Mondays, the 1st cluster appears less. When now repeating the classification model with an intercept and interactions for PE, it is found that both the intercept for PE and the interaction between PE and Day7 are significant model variables, and lead to a decrease in Deviance and a decrease in the misclassification error from 27% to 17%. Despite some misclassifications between the 1st and 2nd cluster, the classification with PE enables to mathematically identify that there are significant differences in the clusters between the two periods, and to recognise the different calendar-patterns that lie underneath this difference. Remark also that the 3rd cluster has become very small in the 1st Period, and appears in different conditions of the weather variables than in the second period. During the classification process, it is absorbed by other categories in the 1st period (this is also the case in the model without PE), but thanks to the intercept term, it can be better identified in the second period, leading to a lower misclassification error for this cluster too.

As a final note on this example, note that when in the clustering process the Euclidean metric was applied instead of the DTW-metric, a maximum of 10 clusters was identified, and during the classification process, it was reduced to 7 clusters by stepwise deletion of the small clusters. The cluster plots for these 7 clusters are summarised in in Figure 5-60 and gives more insight in the different energy uses during year 1. The 2nd cluster (yellow) now covers most of the heating season days before mid-December (and further appears for some weekend days in the 2nd Period), while clusters 1 (red), 4 (turquoise) and 5 (blue) cover most of the heating season especially in the 2nd Period. While cluster 4 covers the weekdays, cluster 5 covers the weekend days, and cluster 1 appears more on Wednesdays, as well as during the School holidays (e.g. Christmas and Spring holidays). When classifying these clusters, a misclassification error of 40% was achieved, partially because of the small clusters that appear at higher temperatures, but probably also because of the absence of an intercept term that allows diversification between the two periods.

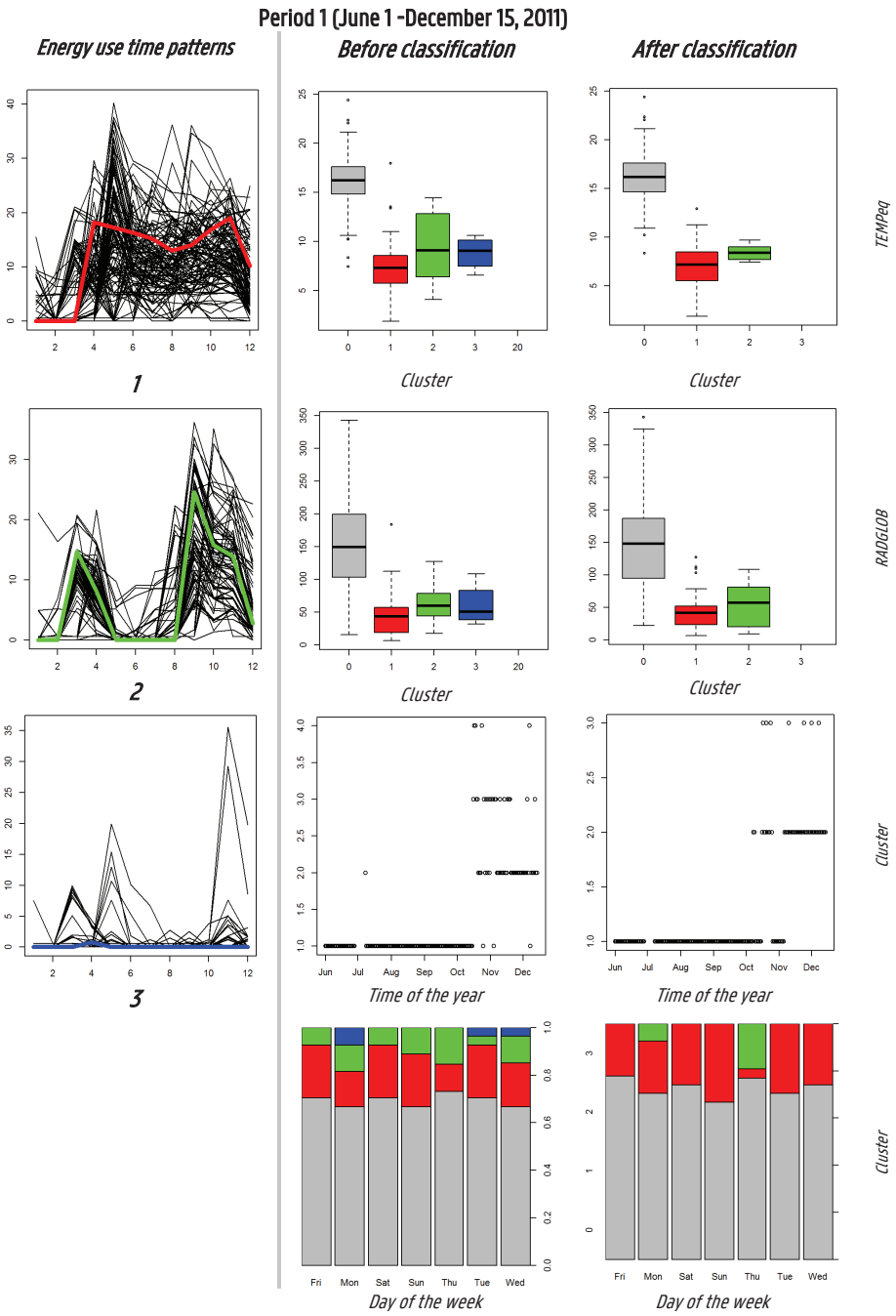


Figure 5-58: Case 17 – Comparison of Clusters for Period 1

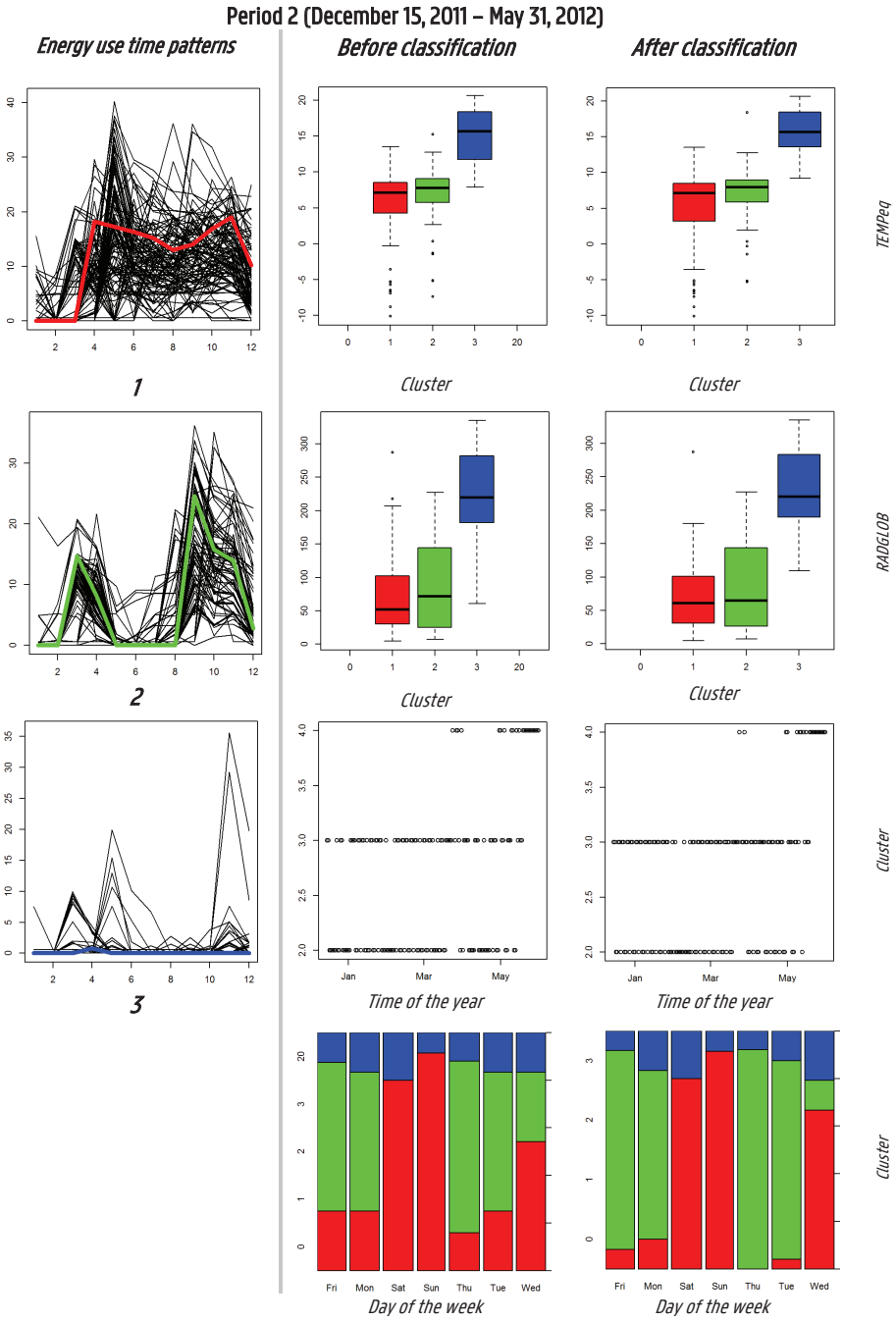


Figure 5-59: Case 17 – Comparison of Clusters for Period 2

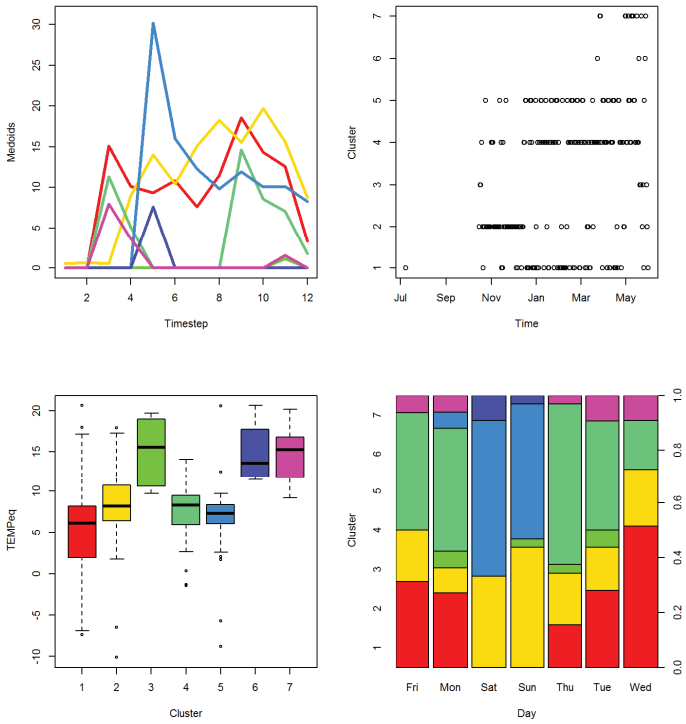


Figure 5-60: Case 17 – Clusters with Euclidean metric

5.7 Application: energy use for space heating and domestic hot water

In many households gas or district heating is used for both space heating and domestic hot water generation, and thus the resulting energy use time series are the result of both energy use functions. The approach developed in this chapter also offers opportunities for this type of energy use time series. The two examples in this section illustrate this application. However it is noted that more profound research is needed to further ascertain its possibilities and limitations. The two example data series originate from the same data-set as the 25 space heating cases, and the data for Year 1 and Year 2 (June 2011-May 2013) are selected.

5.7.1 Example 1

The time series of the energy use of the first example are presented in Figure 5-61 for the months of January, May and August respectively. In the winter period the presence of different patterns for weekdays and weekend days is clearly recognised. When applying cluster analysis with the constrained Dynamic Time Warping measure, four different clusters are identified and the average silhouette width is 0.32. The weekdays and weekend days during the heating season are clearly identified in clusters 1 (silhouette width 0.56) and 3 (silhouette width 0.49). The second cluster contains most of the data from the summer season, where smaller peaks in energy use appear at different moments of the day, and where no differences for different days are recognised. The 4th cluster is the weakest cluster (silhouette width 0.05) and contains data from the intermediate season, but also some 'misplaced' data from the summer season. Then the classification method is applied and the selected model variables are the equivalent ambient temperature, the day of the week and the wind speed variable. After classification, the misclassification error is 24%, and 15% of the misclassified values originate from cluster 4 and go to cluster 2. Thus the classification process somehow 'corrects' for the elements of which the location in cluster 4 was doubtful (remark however that in the classification process all data were used, also objects with a negative silhouette width, which leads to better results in this case).

Finally, linear regression models are applied on the daily aggregated data. When a classical linear regression model is applied (the base model from Chapter 3), the adjusted R^2 is 0.77, all coefficients are significant, but the model is invalid due to auto-correlated residuals and non-constant variance. The weather coefficients are also dominated by the large amount of points in the summer period – the system that is modelled is in fact not linear. Through addition of the cluster variable, and its interaction with the weather variables, which is significant for the exterior temperature, solar radiation and wind speed, linearity is achieved. The model fit improves (the adjusted R^2 rises to 0.93) and fits the data better (see Figure 5-63). The auto-correlation of the residuals is resolved, but the variance is not constant and the residuals are not normally distributed (Figure 5-64). An exploration of the data teaches that this issue is still caused by the difference between summer and heating season: in the summer the variation in energy use is much smaller than in winter, and thus also the variation in the residuals differs.

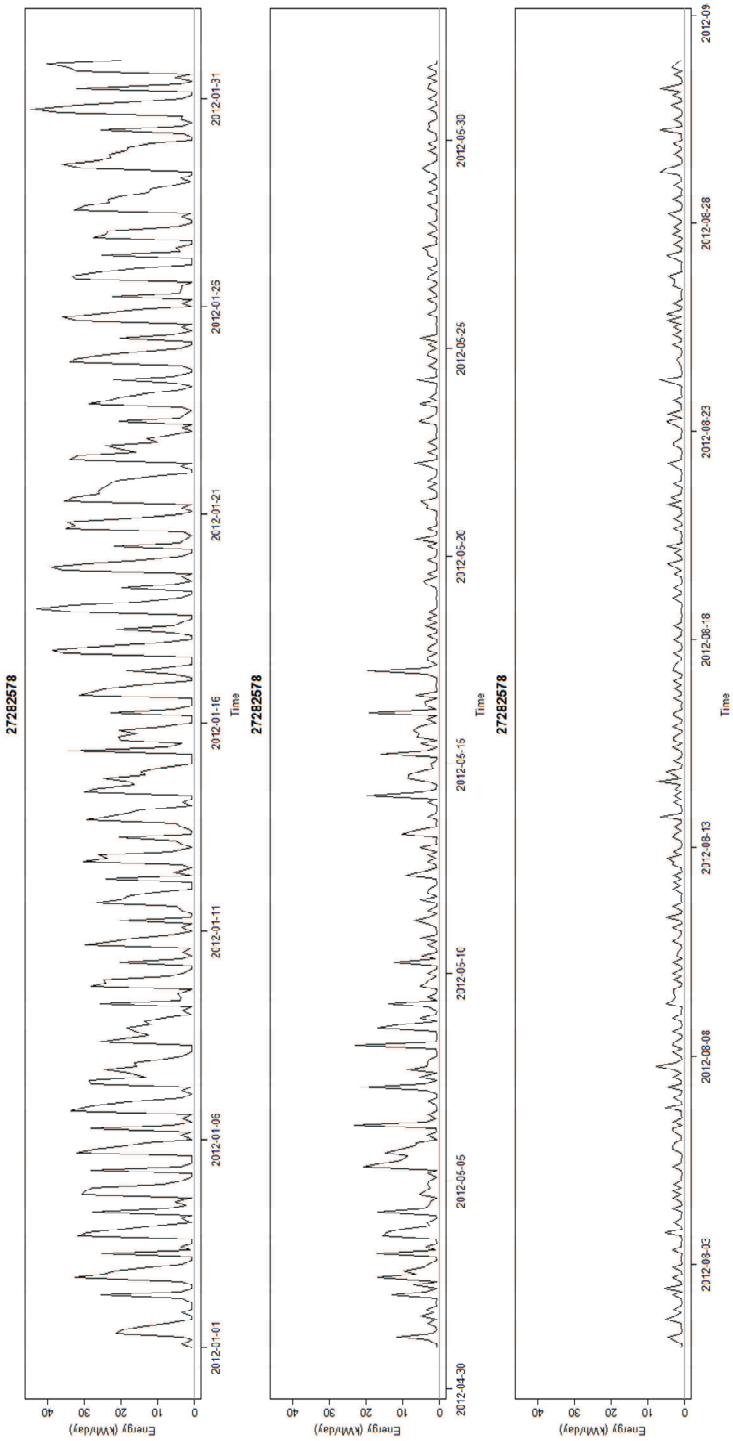


Figure 5-61: Space heating + Domestic Hot water: Example 1: Energy Use Time Series

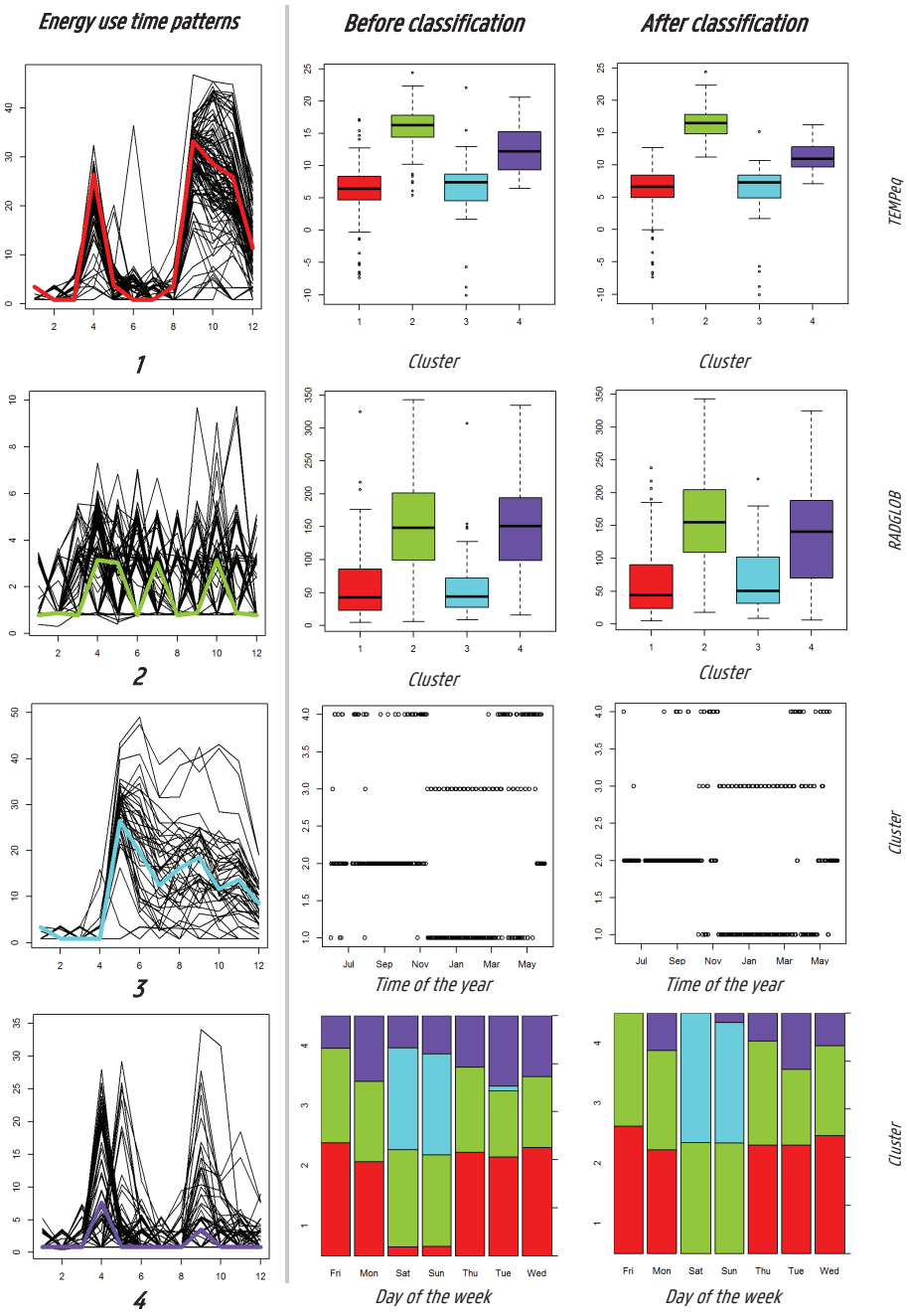


Figure 5-62: Example 1: Energy Use Time Patterns – Clustering and Classification

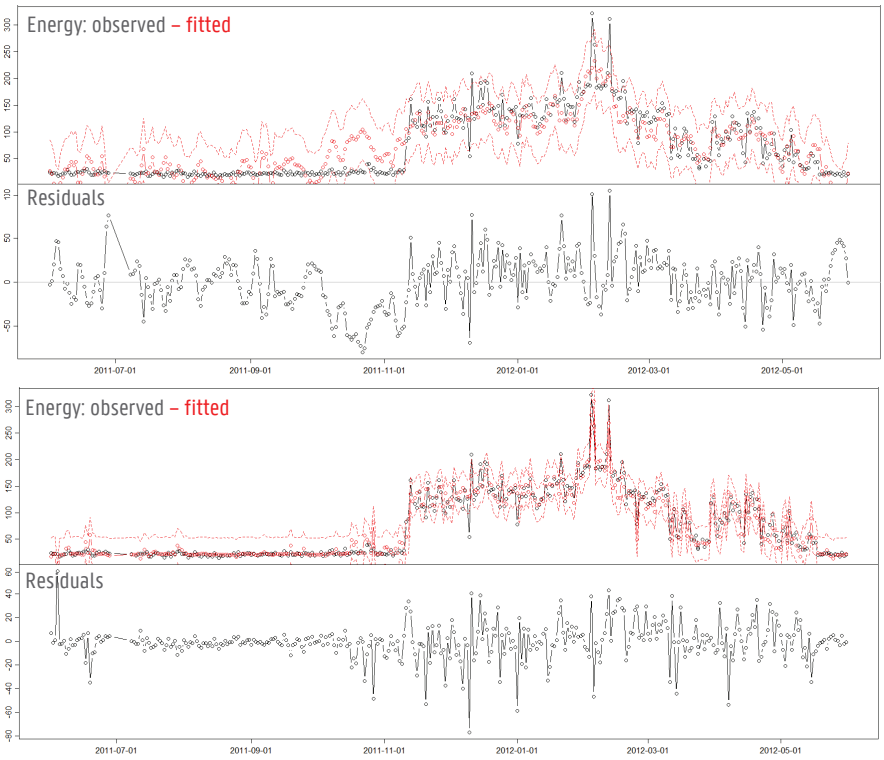


Figure 5-63: Time Series Plots of the energy use and residuals , for the LM(top) and LMCIS-model (bottom)

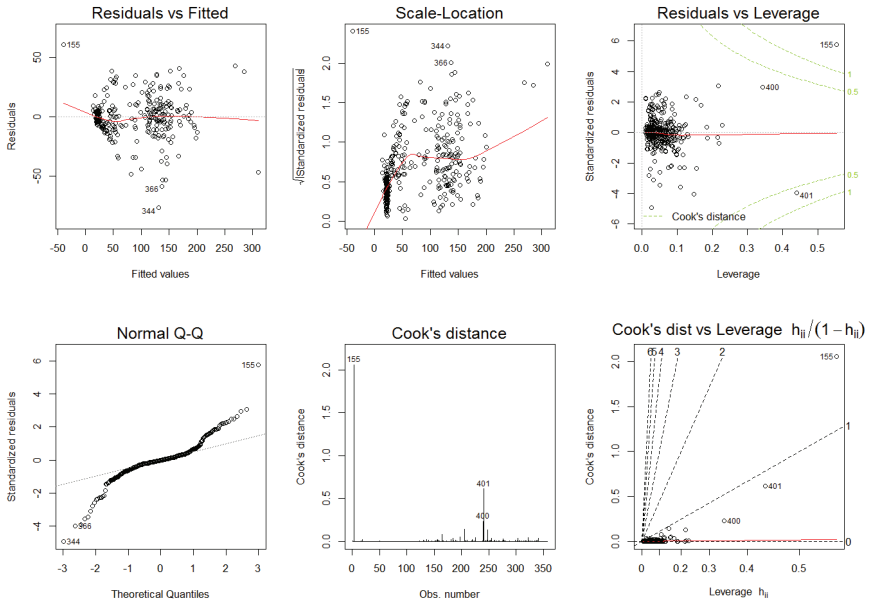


Figure 5-64: Example 1: Model Validity plots

5.7.2 Example 2

For the second example dwelling, the heat use during the winter season is characterised by a heating profile with night setback (Figure 5-65). In the summer period, gas is used only once a week, often on Saturdays. Perhaps domestic hot water is stored in a storage tank, which only needs to be heated after a large domestic hot water demand, such as a weekly bath. Remark that also in the winter month, some small peaks in space heating demand occur at night, however they appear at another time of the day than those in summer time, so it is unclear to what type of energy use they belong to. In the cluster analysis again four clusters are recognised and the average silhouette width is 0.38 (Figure 5-66). Cluster 1 is the 'domestic hot water' cluster during summer time, cluster 2 also appears in summer and in the intermediate season, and clusters 3 and 4 appear during the heating season, at different levels of the exterior temperature. In the classification model, the equivalent exterior temperature, solar radiation and the day of the week are significant variables, and the clusters are quite well classified. The misclassification error is 15%, and especially objects of cluster 2 are misclassified, due to a combination of the unbalance of the cluster sizes together with the appearance of this cluster in the ranges of the larger neighbouring clusters. When a classical linear regression model is applied on the data, the model again suffers from non-constant variance and auto-correlated residuals. A suitable model with clusters is the ARXCIS model, that includes the 1st auto-regressive term, the main weather variables, the cluster variable and its interaction with the cluster variable and the auto-regressive term. The goodness-of-fit of the model rises from 0.91 to 0.98 and the model validity is better, with limited deflections from normality and more constant variance. Thus it is concluded that also for this case, the main energy use time patterns could be recognised and classified, and is considered a clear benefit also for the linear regression modelling of the energy use time series.

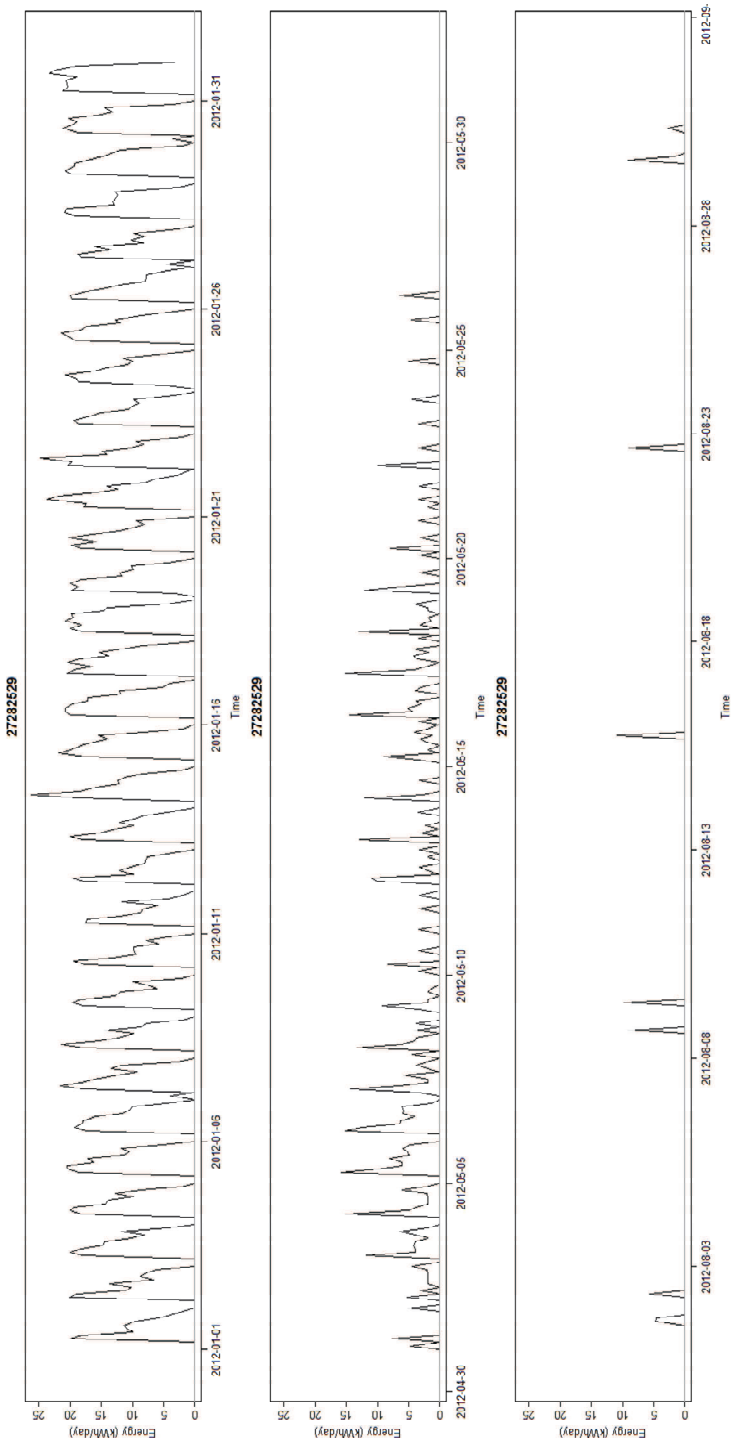


Figure 5-65: Space heating + Domestic Hot water: Example 2: Energy Use Time Series

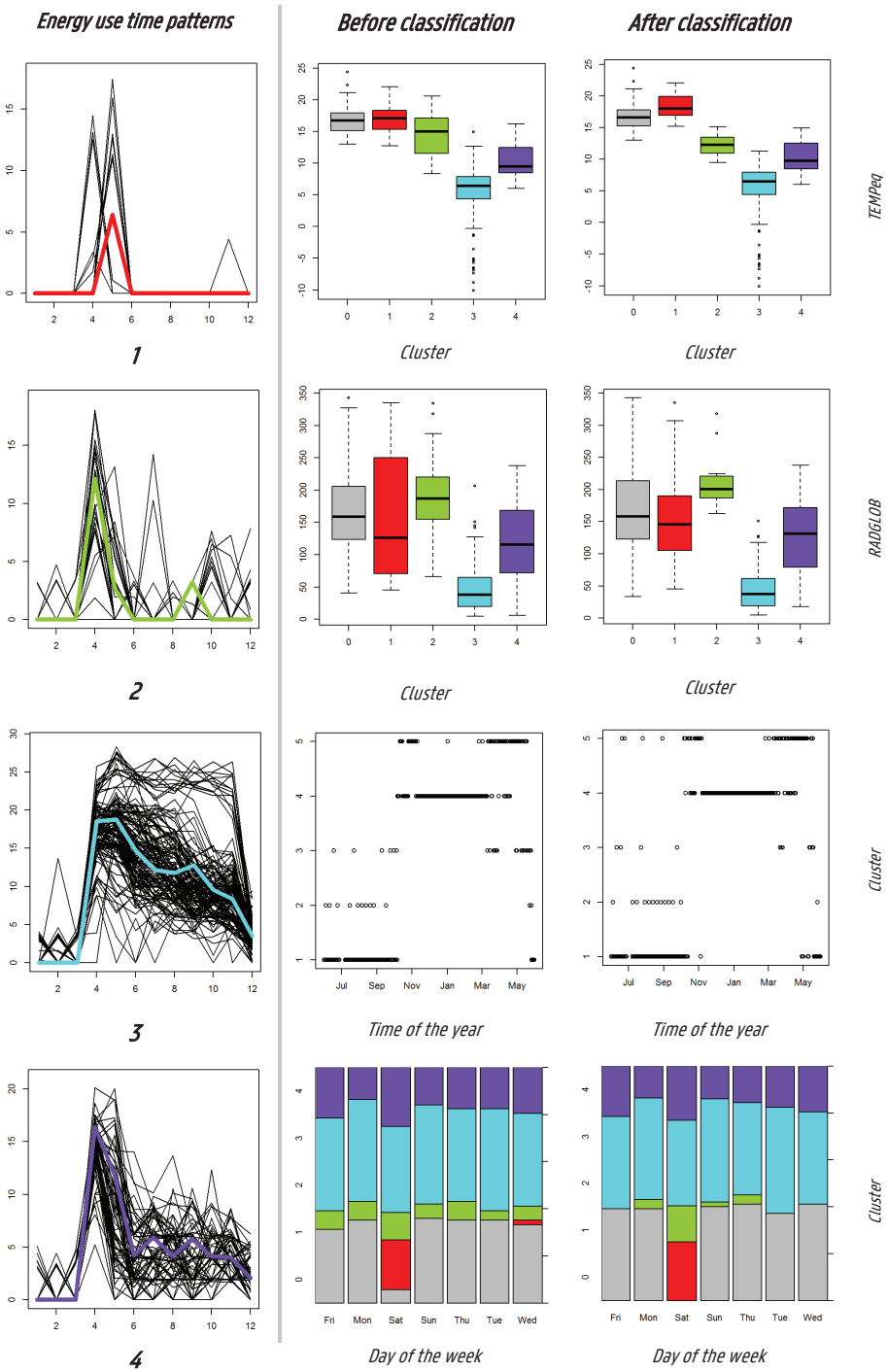


Figure 5-66: Example 2: Energy Use Time Patterns – Clustering and Classification

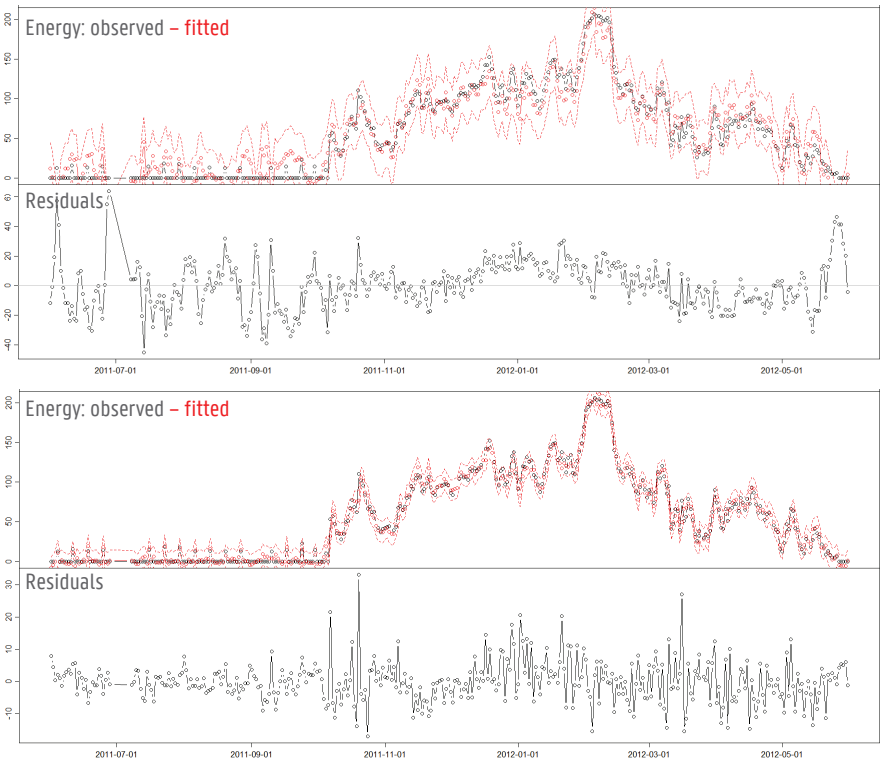


Figure 5-67: Time Series Plots of the energy use and residuals , for the LM(top) and LMCIS-model (bottom)

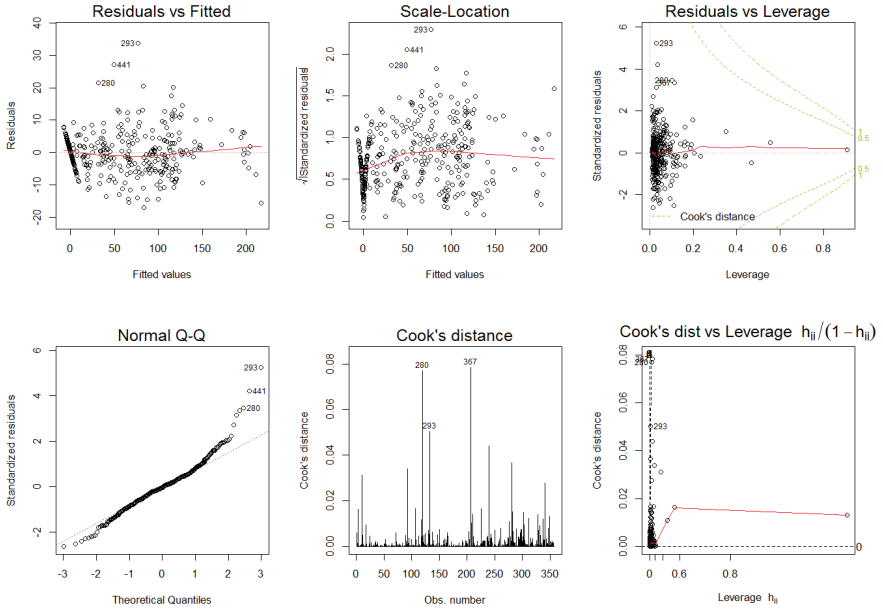


Figure 5-68: Example 1: Model Validity plot

5.8 Summary & Conclusions

The diurnal variations in energy use, observed in time series at an hourly or 2-hourly time interval, often exhibit recurring patterns that are called Energy Use Time Patterns. The subjects of this chapter are (1) the mathematical recognition of similar patterns when observed at 2-hourly time steps, (2) the identification of the relationship between the clusters of similar patterns and the weather and calendar variables observed at daily time steps, and (3) the explanation of the daily energy use figures by means of the weather variables, autoregressive variables and cluster variables, using linear regression models.

First, the energy use time series are decomposed using time series decomposition methods (section 5.2). Those are found useful for exploring and pre-processing the series. The trend component is obtained by smoothing the data and presents the long-term direction of the series that tends to change with (smoothed) weather conditions. The information it holds is similar to the information captured by daily or weekly aggregated energy use data. A second component is called the seasonal component and consists of a repetition of energy use time patterns with a fixed and known period of one day or one week. The third component is the irregular or residual component. For the different cases between 1% and 58% of the variance in the data during one year is explained by a multiplicative and not time-varying seasonal component (see Fig 5-17). This demonstrates that for some cases the energy use time patterns are more influential and regular than for others. Furthermore it is found that when using 2-hourly aggregated data, the variation in the data explained by the seasonal components is higher than when using hourly data, while still the data are detailed enough to understand the patterns. Therefore, throughout this study, 2-hourly aggregated data are used. It is also found that for the energy use time series considered, the magnitude of the diurnal patterns in the data varies with the level of the trend component, and is thus better described by use of multiplicative composition. The findings that the variation in energy use is better explained by use of time-varying seasonal components, and that in a number of cases the energy use is better explained by components with a weekly period, support the conclusion that different energy use time patterns may occur for different days of the week and for different seasons of the year.

Groups of similar energy use time patterns are identified by use of cluster analysis (section 5.3). The *Partitioning Around Medoids* cluster algorithm is applied for partitioning the data into a number of clusters through minimisation of the average distance to the cluster medoid. The Silhouette index is used as a cluster validity index and for selecting the best amount of clusters between 3 and 10. The clusters are grouped based on similarity in shape. Therefore the pre-processing of the energy use time series is required, and it is found that the use of de-trended data (after multiplicative decomposition) is better than the use of standardised data for this goal and type of data. Then the data are cut into series with length of one day, which are the objects that will be clustered. Furthermore, the series that present a Zero-pattern or Standby-pattern are excluded. The similarity or distance between the different objects is calculated by use of a distance measure or metric. The classical Euclidean metric and the constrained and unconstrained Dynamic Time Warping metrics are evaluated. It is found that the constrained DTW-metrics outperform the unconstrained metrics, and the metric constrained with a Sakoe-Chiba band of size 1 and a slope constraint P0 is selected. When comparing the Euclidean and constrained DTW-metric, it is found that for the same data the Euclidean metric allows to

separate patterns of similar shape that are shifted one time step, whereas for the constrained DTW-metric, they are often grouped. This can be a benefit in some cases because it allows a more detailed separation of the patterns. However it often also leads to a higher number of small clusters with patterns of very similar shape and properties, which is disadvantageous in the subsequent classification of the data in function of external variables. The constrained DTW-metric typically leads to a more crude partitioning in 3 to 5 clusters that are better balanced in size, and still permit the recognition of the main energy use time patterns. Therefore, for the subsequent classification of the series, the constrained DTW-metric is selected. The cluster algorithm is usually applied to all data-objects from the entire period considered. From the basic assumption of the variability of the energy use time patterns throughout the different seasons of the year, because of the relationship of the space heating energy use demand with the weather variables, the alternative cluster algorithm is first applied to three seasons (summer, winter and intermediate) separately. Next, the medoids selected for each of the seasons are used as a start for clustering on the data for the entire year, after which the similar medoids are joined step-by-step, and the dissimilar medoids remain. Using this approach, an equal or higher amount of clusters remains, together with small improvements or small deteriorations of the results when observing the variance explained by the deduced seasonal components. Since the differences are relatively small, this confirms the results of the cluster algorithm that is immediately applied to the data for the entire year, which will also be selected after the classification modelling. Based on the cluster results, the seasonal components can be differentiated for the different clusters of objects. As a result, on average 76% of the variance in the data is explained by seasonal components (differentiated using clustering on the entire period or per season), in comparison to 69% for the non-differentiated components obtained by simple multiplicative time series decomposition methods. When the Euclidean distance metric is used and more clusters are selected, the variance explained by the seasonal components rises to 81%. The resulting procedure for clustering of energy use time patterns is summarised in Figure 5-69.

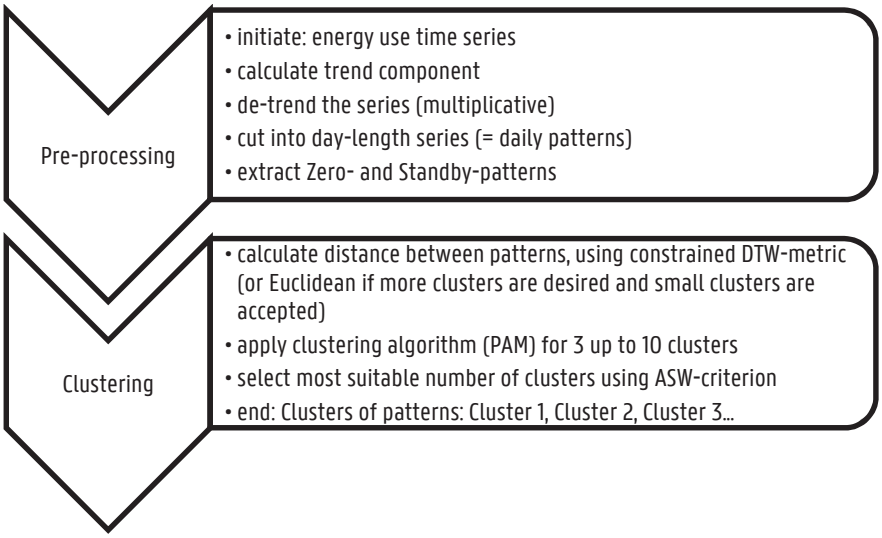


Figure 5-69: Clustering of energy use time patterns: procedure summary

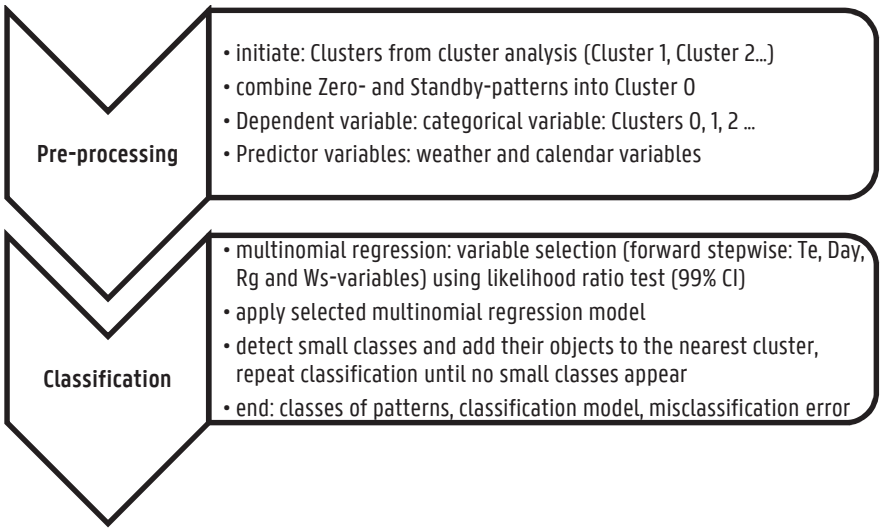


Figure 5-70: Classification of energy use time patterns: procedure summary

The obtained clusters of energy use time patterns occur for specific conditions of weather variables and/or calendar variables. In a next step (section 5.4), the obtained clusters are classified in function of these variables using multinomial logistic regression techniques. The final classification procedure is summarised in Figure 5-70. If Standby-patterns are detected in the data pre-processing, they are first combined with the Zero-patterns into a Zero-Cluster that is also submitted to the classification. The logistic regression model selection is a forward stepwise regression procedure relying on the likelihood ratio test for nested models, with 99% confidence level. The main issue met during the classification process is that clusters containing a relatively small amount of objects and occurring for a range of external variables that overlaps with the range occurring for larger clusters, tend to be absorbed by the larger clusters and remain (almost) depleted – no matter whether the larger cluster is a neighbouring cluster or not. Therefore the classification approach is modified: after a first classification round, the small and depleted clusters are detected and the objects from the smallest cluster are allocated to their original neighbouring cluster; then the classification is repeated and this procedure is repeated for each of the empty or small clusters. As a result, the clusters become more balanced and the overall classification quality improves, and the amount of misclassified objects lowers. This modification is needed for a limited number of cases when the constrained DTW-metric is used, but is especially useful when the Euclidean metric is used. However, after the modified algorithm is applied on the data with the Euclidean metric, often a lower amount of clusters remains than when using the constrained DTW-metric: the median reduces from 4 to 3 clusters. Therefore, it is concluded that the constrained DTW-metric rather than the Euclidean metric is preferable for the majority of the cases. The final models are compared in terms of the Deviance and the misclassification error. In 14/25 cases the energy use time patterns are characterised by weather predictor variables, in 9/25 cases by a combination of weather and calendar predictor variables and in 2 cases clustering and/or classification of energy use time patterns is not advantageous because no clear patterns can be observed or because they cannot sufficiently be explained by the predictor variables. For a few example cases, the results of the clustering and classification

process are visualised. Then again the differentiated seasonal components can be estimated based on the remaining clusters after classification. The variance explained by the seasonal components decreases to 71% on average, which is lower than after the clustering phase, but higher than in the original data. Note that part of the gains of the differentiated seasonal components, e.g. those for the summer season, are concealed by the trend component being zero.

Apart from the modelling of the clusters of energy use time patterns using logistic regression techniques, also the daily aggregated energy use time series are modelled using information from the energy use time patterns, using linear regression models. In order to enable the comparison with the LM- and ARX-models from previous chapters, the data from the heating season are selected for this purpose. The cluster variable is an additional categorical variable that describes to which cluster of energy use time patterns each data point belongs. It is entered into a model with weather variables and optionally also autoregressive terms by use of an extra term (LMCI and ARXCI-models) and possibly also in interaction with the other predictor variables (LMCIS- and ARXCIS-models). The findings from the comparison of LM-, ARX, LMC- and ARXC-models can be related to the type of autocorrelation that is observed when applying a classical LM-model (Chapter 3). It is concluded that in cases with a weak autocorrelation, or with a strong autocorrelation of the 2nd type (with peaks in the ACF at specific time lags), the models with cluster variables (LMC or ARXC) are preferred to the selected LM- and ARX-models. The autocorrelated residuals are often released or diminished; however, for some cases additionally the first AR-term(s) are needed to resolve autocorrelations, in which case an ARXC-model results. This group of cases also includes the cases with clear calendar-related patterns: by use of the cluster variable for each day of the week it can be pointed out specifically whether it differs from some other days. For the 6/25 cases where a severe autocorrelation of the 1st type is detected in the LM-model, it is concluded that an ARX-model is essential, but by adding the cluster variables (ARXCI or ARXCIS-models) it can be optimised for some cases. The overall model validity of the LMC- or ARXC-models is better than for the LM- or ARX-models. The main issue that remains in a few cases is a violation of the assumption of constant variance, since an increase in variance for increasing exterior temperature is determined. The goodness-of-fit of the models improves on average 0.05 pp when compared to the models without cluster variable. The amount of model coefficients is expanded. In case the cluster variable is added as an intercept term, for each of the clusters a specific intercept can be estimated, and when interactions with the other predictor variables are added, also the coefficients for these variables are specific per cluster. Thus it can also be seen for which clusters the coefficients differ significantly from other clusters. It is also noticed that the clusters of energy use time patterns can also be used to select the days on which space heating is applied, and separate them from days with no or standby energy use. As a consequence, the linear regression models with cluster can also be applied to the data for an entire year. Moreover, it can also be applied for cases in which the energy use is not only for space heating functions but also for domestic hot water generation. This application will be illustrated shortly in section 5.7. In Chapter 4, the data of the 1st and 2nd heating season were compared using selected LM- or ARX-models. In this chapter the data for the entire 1st and 2nd year are compared. Three ways of comparing the data are cited: via the clusters of energy use time patterns (the classification model), the coefficients of the energy use model (the LMC- or ARXC-model) and the heating season models. The first two options are implemented and the comparison is made by adding a categorical variable that differs for the two periods compared, and investigating its significance. It is found that when comparing the clusters of energy use time patterns, for almost

every case significant differences can be observed which cannot always be confirmed by visual representations of the two periods. A danger is that the difference observed originates from a difference between small and/or weak neighbouring clusters and may not be significant when considering the wholeness of the clustering and classification process. This danger does not turn up when comparing the energy use models that also tell more detailed to which variable or cluster the difference may be related. For more cases a difference in energy use is detected when comparing to models without cluster variables, but the main advantage of the method is that we can dispose of more detailed insights that can help to understand why a difference in energy use was found.

Further elaborations and optimisations to the modelling approach developed in this chapter could enhance the methods for classifying the clusters of energy use time patterns in function of external variables, and for comparing those clusters for different periods: for example with regard to small or unbalanced clusters, and to the incorporation of more information from the clustering results, e.g. the cluster quality or neighbouring clusters for the objects. Also the incorporation of input data related to different possible holiday periods could be considered, eventually as an input by the user, since holiday periods differ significantly from household to household. Finally it is noted that the clusters of energy use time patterns could also be a stepping stone towards the implementation of more advanced time series decomposition methods.

6

Conclusions & Perspectives

6.1 Conclusions

Reliable and accurate estimations of the actual energy use of occupied buildings are required for assessing the actual building energy use and the actual impact of energy-efficiency measures or changes in the building fabric, the services and the occupants. With the increasing application of energy use smart meters in buildings, frequent and long-term measurements of the actual energy use in occupied buildings become available. Gas or heat meters measure the weather-dependent energy use for space heating and eventually less weather-dependent energy uses for domestic hot water generation or cooking. Energy Signature models are widely applied data-driven models for describing occasional (weekly, monthly) meter readings of the heating energy use of the system that consists of the building fabric, the services and occupants, in function of weather variables. These models not only present energy use estimations, but the Energy Signature coefficients also provide insights in the energy use characteristics, thus leading to a better understanding of the actual energy use. The question rises whether the more frequent (daily or sub-daily) energy use metering data allow to improve the Energy Signature models applying linear regression techniques. In this dissertation, it was investigated by use of a dataset consisting of 2 years of the hourly gas meter data of 25 case-study dwellings where gas is used only for space heating purposes.

A first research question is whether data at a daily time resolution allows to improve the Energy Signature models applying classical linear regression techniques (LM-models). The input variables to the models are weather variables, and the output is the energy use. It is found that when going from weekly to daily aggregated energy use data, the statistical validity of the LM-models is no longer accepted in about half of the investigated cases. For these cases the model coefficients may not be reliable, and the LM-models are not suitable. The main validity issue is the autocorrelation of the residuals or the fact that the errors of the model depend on each other in time (section 3.3 and 3.7). Two main types of autocorrelation are present in the data. In the autocorrelation function one is recognised by a slow decay, and a slow-moving trend is observed in the time series plots of the residuals. The second of autocorrelation is recognised by peaks on specific time lags. A physical interpretation of this issue is that the energy use of today partially depends on the energy use of the previous day(s), because the dynamics of the building are slower than one day (e.g. because heat is stored and released slowly from the building mass), because some phenomena are not modelled (e.g. interior temperature) or because the energy use may be influenced by occupancy patterns (e.g. different energy use behaviour during weekdays and weekend days). This brings us to the second research question, that is whether the Energy Signatures can be improved by use of auto-regressive models with exogenous inputs (ARX-models). In addition to weather variables, auto-regressive terms (these are historical values of the energy use) are used as input variables in the model. It is found that the ARX-models are indeed able to deal with the autocorrelated errors: the models are valid for the majority of the cases (section 4.2). When observing the coefficient estimates, it is found that the width of their confidence intervals was underestimated by the invalid linear regression models and could indeed not be trusted (section 4.2). For the valid linear regression and ARX-models with daily aggregated data, the accuracy of the models improves in comparison to the classical linear regression models with weekly aggregated energy use data. For example the confidence intervals for the exterior temperature coefficient are between 1.2 and 2.8 times smaller (section 4.6.2). Therefore it is concluded that by use of daily energy use data, reliable and more accurate

Energy Signature models are obtained than by use of weekly energy use data. However in order to obtain reliable results, in many cases it is required to use ARX-models instead of LM-models.

Furthermore, the key variables of the LM-models are the exterior temperature, solar radiation, wind speed and the first (and eventually second) exterior temperature lag (section 3.2). For the ARX-models, these weather variables remain as exogenous inputs, but exterior temperature lags are often less important. The key AR-terms are the 1st, 2nd, 3rd and/or 7th lag of the energy use, but their significance depends from case to case (section 4.1). The key variables can be used to create a base model that is suitable for a majority of cases, or they could be entered in a forward variable selection procedure for automatically obtaining case-specific models. Regarding the accuracy of the coefficient estimates, the estimates of for example the exterior temperature coefficient of the valid LM- or ARX-models have confidence intervals with a width of about 1.5 kWh/°C on average, which is about 20% of the estimated value for this coefficient. The prediction accuracy of the ARX-models for predicting the heating season is marked by a mean absolute error between 5 and 21 kWh/day or between 5 and 10% of the maximum energy use that occurs. Especially the ARX-models have potential with regard to short-term energy use predictions, in which case much higher accuracies can be obtained (section 4.5). Another question is on the required length of the measurement period, or whether it can be shorter than an entire heating season. It is found that the measurement period can be reduced to periods of two or three months with a negligible loss in model accuracy, if the (variation in) weather variables is sufficient. For a precise formulation of requirements an investigation on a more extensive dataset is needed, but some rules of thumb are derived in section 3.6.2.

Sub-daily energy use time series allow a more detailed observation of the diurnal variations in energy use that are due to occupational schedules, system control settings, short term weather dynamics etc. (section 5.1). These diurnal variations are responsible for a large part of the variance in the sub-daily data (section 5.2). When using these data in classical linear regression models, the issue of the autocorrelated errors becomes very problematic (section 3.7.1). ARX-models allow a more decent description of the variance in energy use resulting from regularly occurring patterns (section 4.6.1). The detailed modelling of these dynamics is out of the scope of this dissertation, and it is expected that more input variables and dynamic models would be needed for that purpose. However it is found that part of the information contained in the sub-daily energy use data can be used to improve the Energy Signature models with daily aggregated energy use data (research question 5). That information consists of the different groups of energy use time patterns that are observed in the sub-daily energy use time series, and that may explain some of the variations in the daily energy use that are not explained by weather variables, e.g. variations due to occupational schedules.

How these groups of energy use time patterns can be recognised mathematically, is the third research question. For this purpose a procedure is developed that allows to cluster the energy use time patterns based on their similarity in shape, using cluster analysis techniques (section 5.3). The procedure starts with pre-processing of the energy use time series, followed by the calculation of the similarity between the different day length patterns by use of dynamic time warping metrics. Finally, the patterns are grouped in 3 up to 10 clusters. As a result, the long series of patterns are summarised in a limited number of groups that are identified by a representative pattern. The clusters can also be graphically presented in function of weather and calendar variables. Indeed it is concluded that many of the patterns can be related to specific

calendar conditions (e.g. to weekdays or weekend days) or weather conditions (e.g. to the intermediate season or winter time) (section 5.3.5). Moreover during the pre-processing patterns occurring during long periods of standby could be identified. It is concluded that the energy use time patterns allow a more profound insight in the energy use related behaviour of the occupied building, and using the clustering approach large amounts of data can be summarised and presented in an informative way. The fourth question is how these clusters of patterns can be mathematically modelled in function of those variables. A classification procedure using multinomial logistic regression models is used for this purpose (section 5.4). The resulting classes of energy use time patterns can be presented graphically or can be used for prediction purposes, e.g. to estimate which pattern would occur for certain values of the input variables (section 5.4.5). It is found that a part of the variance in the data that was explained by the clustered patterns, is no longer explained after classification of these patterns. The misclassification error for the 25 cases is on average 23%. So a significant part of the patterns in the data is already explained with the current procedure. Although this percentage represents a part of the patterns that cannot be related to the input variables, e.g. because of irregularities in the behaviour or missing input variables (e.g. holidays). However it is also related to the classification method, in which the larger classes tend to absorb small neighbouring classes of energy use time patterns. The improvement of the classification method could therefore be subject of future research.

Regarding research question 5, the clusters of energy use time patterns can be included in the LM- and ARX-models via the cluster variable. That is an additional categorical variable that describes for each day to which cluster of patterns it belongs and that is added to the model as an intercept term, or in interaction with the other variables (section 5.5). The clusters identified in the 3rd research question are used. For prediction with the resulting models, the classification model from the 4th research question can be applied to predict the patterns. It is concluded that for cases in which the 2nd type of autocorrelation appears, or weak autocorrelation appears, regression models with the cluster variable(s) perform better than LM- or ARX-models without cluster variable. In some cases they can sufficiently deal with the autocorrelations, although additional autoregressive terms can be beneficial or necessary for optimising the model. For cases where the first type of autocorrelation appears, still AR-terms are necessary, and an additional cluster variable may optimise the model. In general, when going from classical LM-models to the ARX-models with cluster variables, the goodness-of-fit of the model increases from 0.76 to 0.85 on average. In the resulting models the Energy Signature coefficients are differentiated for each of the significant clusters. The energy use associated with the different patterns can be estimated and compared.

One possible implementation of Energy Signatures is the comparison of the energy use for different periods, for detecting changes in energy use in energy feedback applications (section 5.6). The findings from the comparison of the differentiated Energy Signature coefficients can be combined with findings from the classification of the clusters. It is concluded that using this method differences between the two periods are found for more cases and they can be better understood by use of the Energy Use Time Patterns. Finally, the application of the regression models for cases with combined energy use for space heating and domestic hot water generation is illustrated. Further investigation is needed, but in this situation the groups of energy use time patterns provide insights also in properties of the DHW demand and allow to distinguish more clearly when the heating season starts (section 5.7).

In this dissertation, classical Energy Signature models are revisited and confronted with high temporal (daily, hourly) data that become more and more available. When using daily data, the classical linear regression models are statistically invalid in many cases. Instead, it is recommended to apply ARX-models. These models are still relatively simple to implement in applications such as energy auditing, feedback or commissioning. When sub-daily energy use data are available, a procedure is developed to summarise the energy use time patterns in the data and to present and model it in function of weather and calendar variables. Furthermore, knowledge from these patterns can be used to improve the LM- and ARX-models. This procedure is more complex to execute, but contains many aspects that can be automatized, and could make it more accessible for building specialists (e.g. engineers, architects) in practice.

6.2 Perspectives

In a further development of the proposed methods, aspects of particular interest are the optimisation of the models for dealing with heteroscedasticity, in this case the increase of the variance with increasing exterior temperatures, that remains in a few cases. Secondly, the first type of autocorrelation is dealt with in the ARX-models, but further investigations of the physical causes for this type of autocorrelation are interesting, for example by use of more well documented case-studies where measurements of other energy-related variables are available (e.g. interior temperatures, ventilation flow rates…). Thirdly, dwellings with a very low energy demand for space heating were not represented in the dataset. Since the heat balance of these houses may be quite different from the balance in less energy-efficient houses, further investigation of the applicability of the methods is recommended. Then, the methods applied in this dissertation are automatised to a certain extent, but still need the intervention of the modeller, for example with regard to the visual model validity checks. The use of statistical tests instead could be investigated and the procedure could be further automatised.

In this dissertation, linear regression type of models were explored, starting from classical linear regression models, and gradually adding the auto-regressive and cluster variables. Further elaboration on more complex models could be considered in further investigations. For example, ARMAX type of models can be experimented with in extension to the ARX-models. Also more advanced methods for decomposition of time series could be explored (section 5.2.5).

For the application of the regression models with cluster variables and eventually auto-regressive variables on combined energy use data (for space heating and domestic hot water), a first exploration by use of two examples provides promising results, allowing a better understanding of domestic hot water use, and a better estimation of the start and end of the heating season. In case the moments of energy use for space heating and hot water do not systematically overlap, it might be possible to extract the domestic hot water from the total energy use using the energy use time patterns. Then a better split between these two energy use functions could be made and a better estimation of the share of each energy use function. However, also this is a subject for future research.

Annex: description of the data

Energy use data

Daily aggregated data

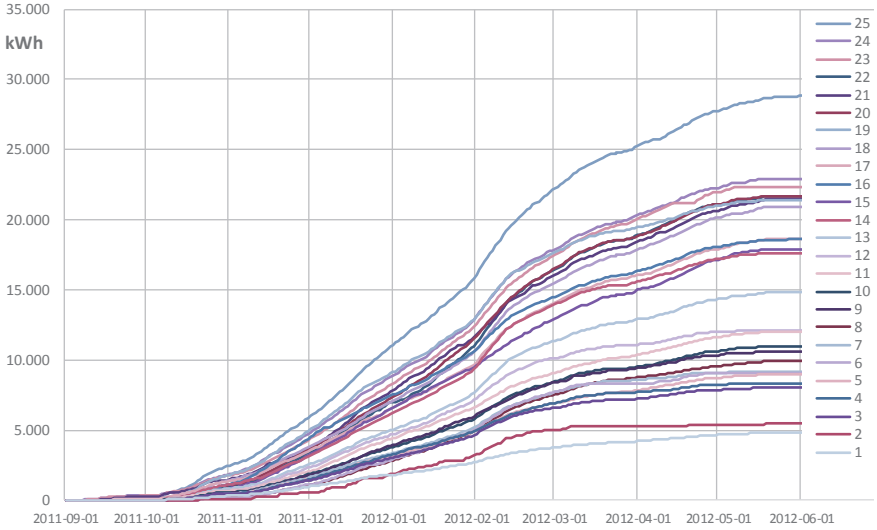


Figure A-1: Energy use per case in HS2011-2012: Cumulative distribution

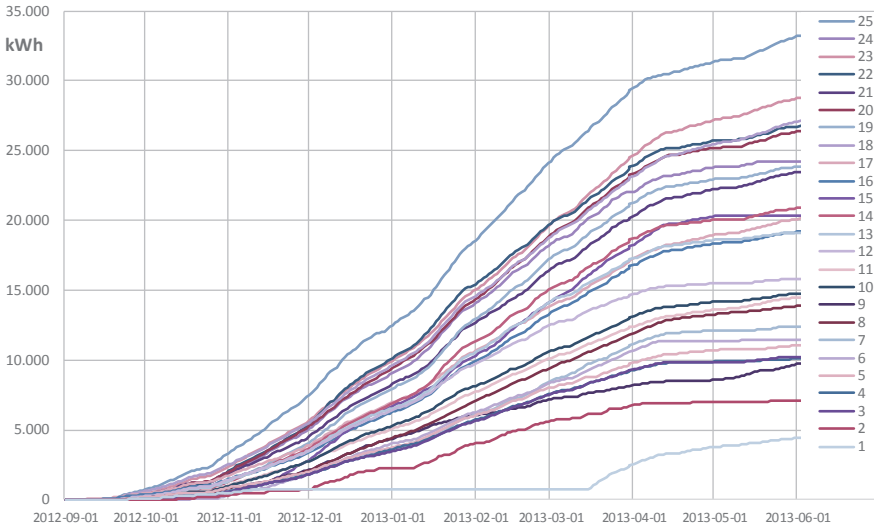


Figure A-2: Energy use per case in HS2012-2013: Cumulative distribution

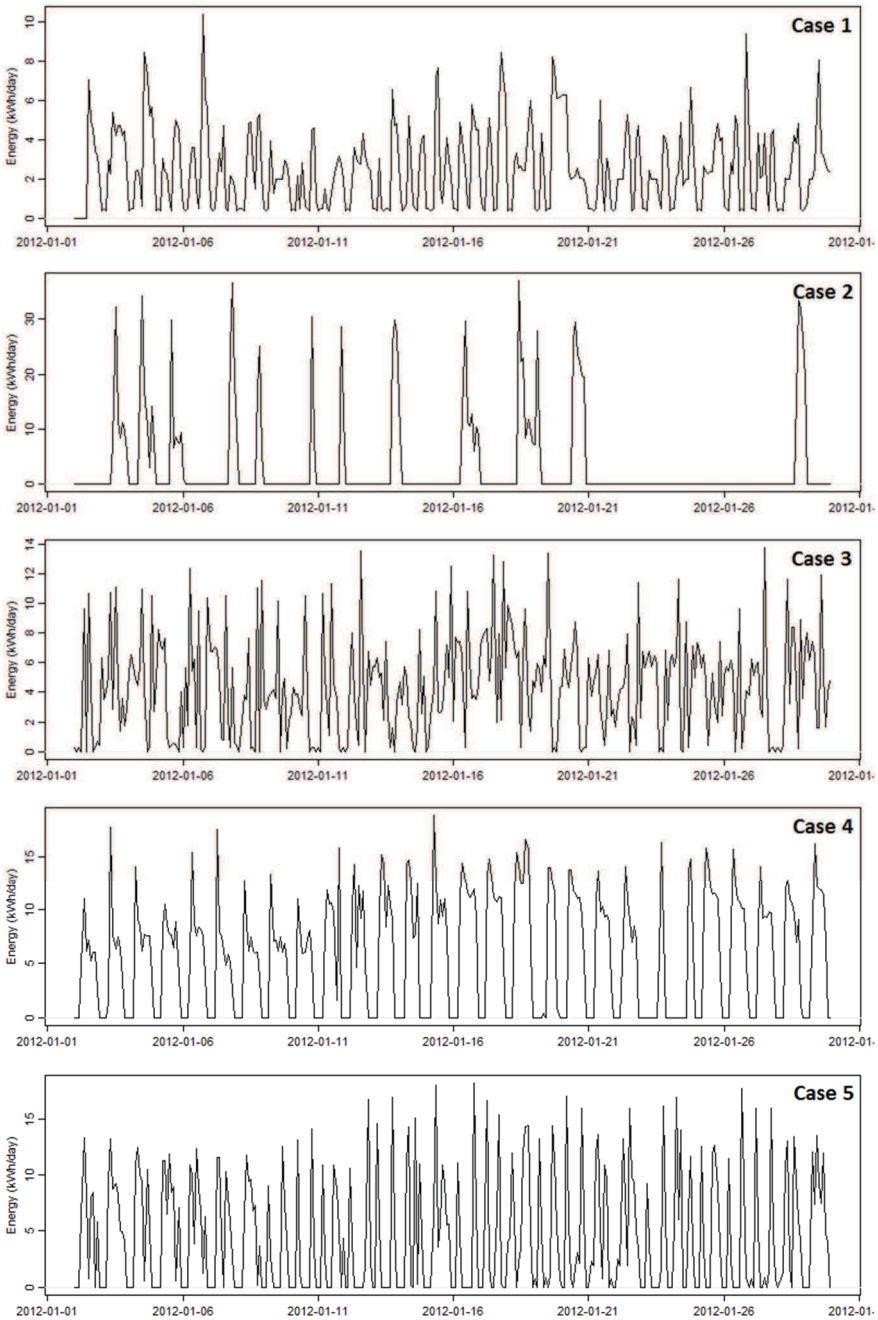
Table A-1: Heating Season 2011-2012: Start and End, Length and Energy Use figures

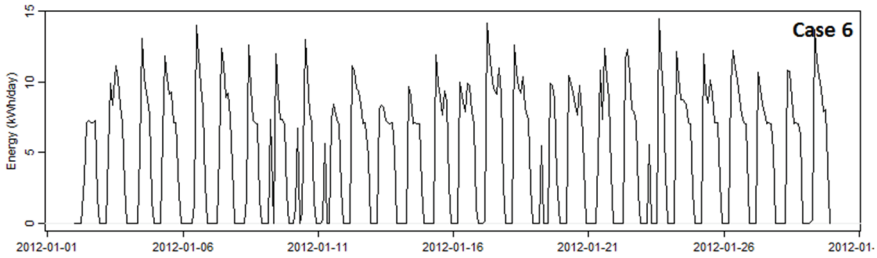
	Start HS 2011		End HS 2012		HS Length		Energy Use (kWh)		
	Day	Weeks	Day	Weeks	Days	Weeks	HS Total	Daily Average	Daily Max
1	8/10	10/10	1/05	30/04	206	29	7853	38	115
2	6/10	10/10	1/05	30/04	208	29	27272	132	342
3	21/10	24/10	1/05	30/04	193	27	5268	27	195
4	7/10	10/10	1/05	30/04	207	29	11527	56	128
5	7/10	10/10	1/05	30/04	207	29	4641	23	60
6	15/10	17/10	24/03	26/03	161	23	8463	52	130
7	8/10	10/10	1/05	30/04	206	29	10322	50	156
8	8/10	10/10	1/05	30/04	206	29	21027	102	238
9	8/10	10/10	1/05	30/04	206	29	20999	102	267
10	8/10	10/10	1/05	30/04	206	29	8700	42	110
11	8/10	10/10	1/05	30/04	206	29	10559	51	142
12	19/10	24/10	1/05	30/04	195	27	16798	88	165
13	7/10	10/10	1/05	30/04	207	29	20755	101	273
14	7/10	10/10	1/05	30/04	207	29	8220	40	119
15	8/10	10/10	1/05	30/04	206	29	20057	97	260
16	1/10	10/10	1/05	30/04	213	29	20172	98	276
17	1/10	10/10	17/04	16/04	199	27	20739	108	239
18	16/10	17/10	1/05	30/04	198	28	18022	91	241
19	1/10	10/10	1/05	30/04	213	29	21768	106	263
20	9/10	10/10	1/05	30/04	205	29	9546	46	121
21	6/10	10/10	1/05	30/04	208	29	17638	86	249
22	12/10	17/10	1/05	30/04	202	28	11890	60	195
23	18/10	24/10	10/03	12/03	144	20	8072	57	140
24	7/10	10/10	1/05	30/04	207	29	14230	69	236
25	6/10	10/10	1/05	30/04	208	29	17025	83	305

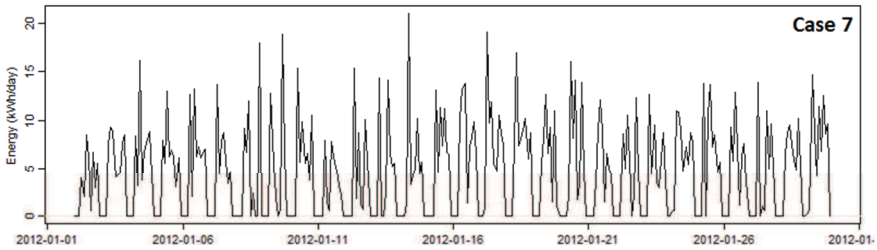
Table A-2: Heating Season 2012-2013: Start and End, Length and Energy Use figures

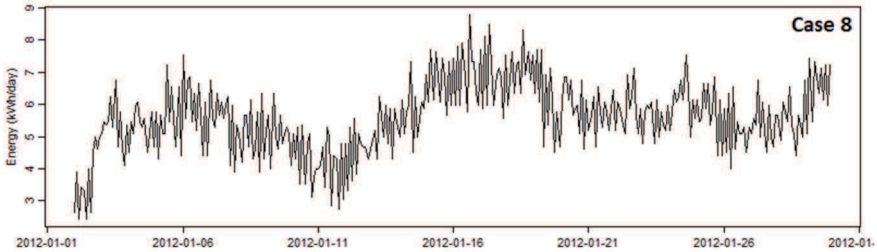
	Start HS 2012		End HS 2013		HS Length		Energy Use (kWh)		
	Day	Weeks	Day	Weeks	Days	Weeks	HS Total	Daily Average	Daily Max
1	21/09	24/09	6/05	6/05	227	32	3823	17	196
2	9/10	9/10	26/04	29/04	199	29	6937	35	194
3	9/10	15/10	14/04	15/04	187	26	9689	52	109
4	28/10	29/10	10/04	8/04	164	23	11017	67	119
5	5/10	8/10	14/04	15/04	191	27	9704	51	129
6	27/10	29/10	16/04	15/04	171	24	11462	67	137
7	21/10	22/10	1/05	29/04	192	27	10177	53	118
8	22/09	24/09	27/05	27/05	247	35	13723	56	107
9	26/10	29/10	14/04	15/04	170	24	7924	47	113
10	24/09	24/09	1/05	29/04	219	31	14094	64	131
11	21/09	24/09	27/05	27/05	248	35	14316	58	116
12	20/09	24/09	15/04	15/04	207	29	15224	74	163
13	1/10	1/10	2/05	29/04	213	30	18386	86	216
14	10/10	15/10	4/05	29/04	206	28	20043	97	195
15	3/10	8/10	2/05	29/04	211	29	19779	94	259
16	19/09	24/09	27/05	27/05	250	35	19658	79	183
17	17/09	17/09	27/05	27/05	252	36	18998	75	188
18	17/09	17/09	27/05	27/05	252	36	26782	106	230
19	20/09	24/09	27/05	27/05	249	35	23055	93	260
20	21/09	24/09	27/05	27/05	248	35	28347	114	236
21	4/10	8/10	1/05	29/04	209	29	22581	108	274
22	25/09	24/09	1/05	29/04	218	31	25573	117	254
23	1/10	1/10	6/05	6/05	217	31	24942	115	223
24	1/10	1/10	24/04	22/04	205	29	23335	114	253
25	19/09	24/09	6/05	6/05	229	32	31204	136	343

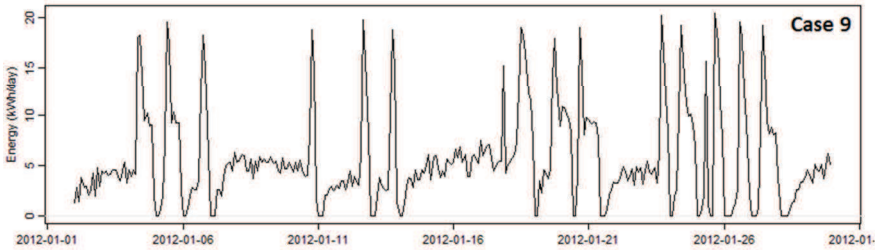
Hourly data (month of January 2012)

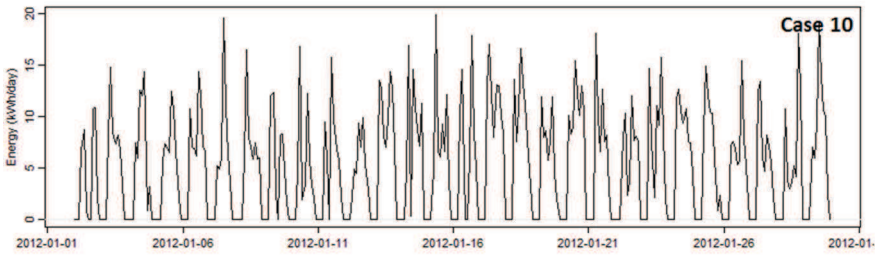


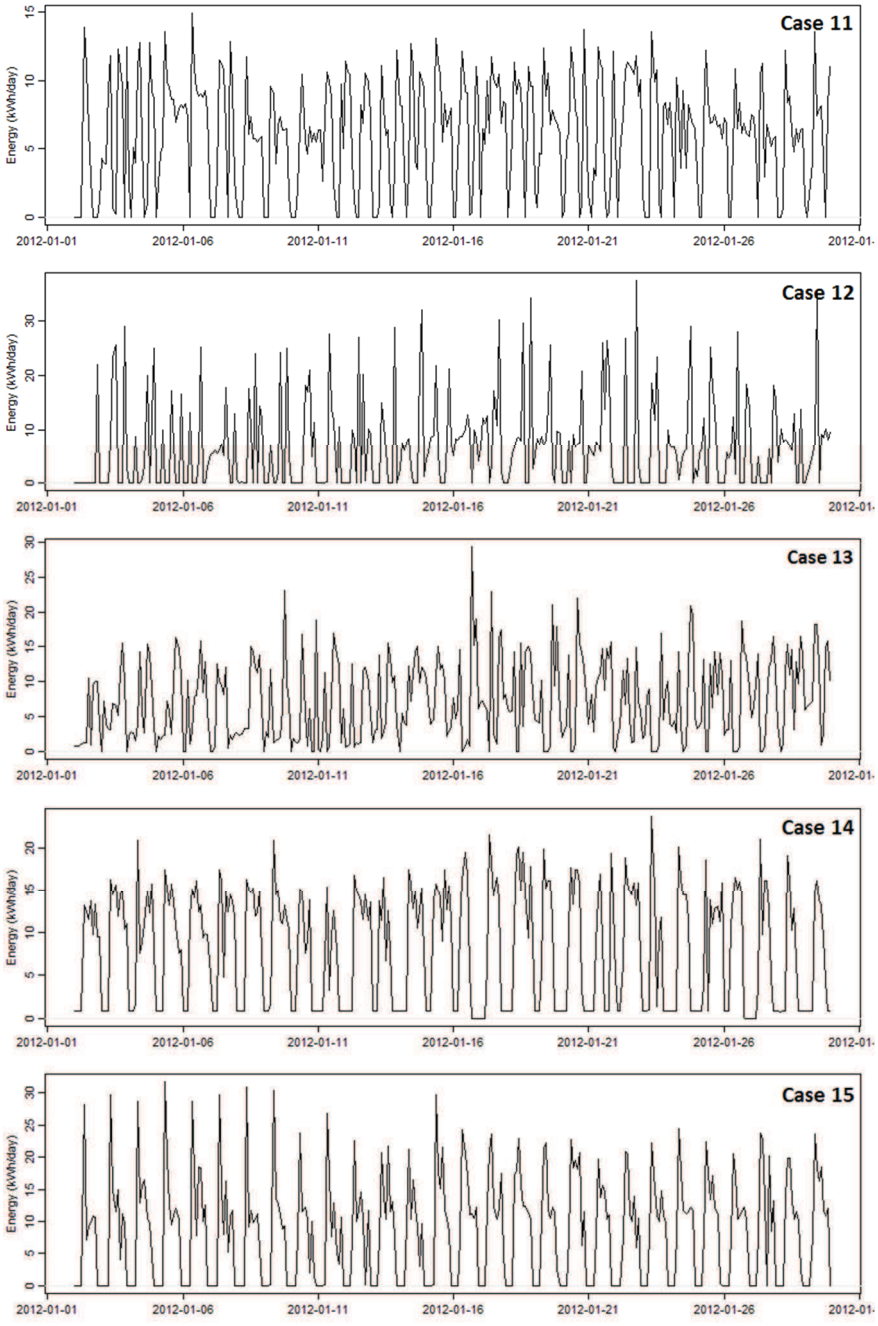

Case 6
Energy (kWh/day)
0
5
10
15
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
2012-01-

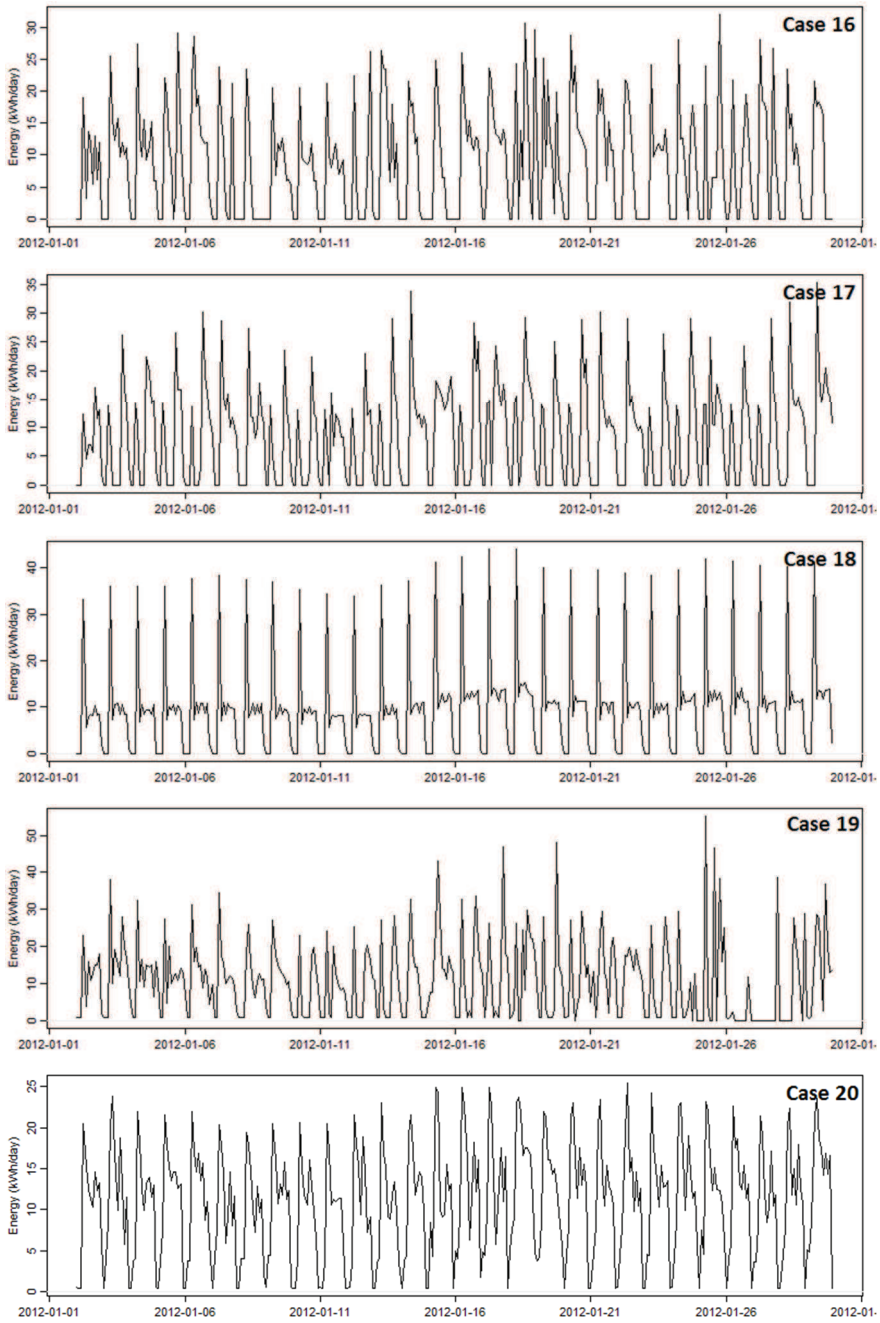

Case 7
Energy (kWh/day)
0
5
10
15
20
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
2012-01-

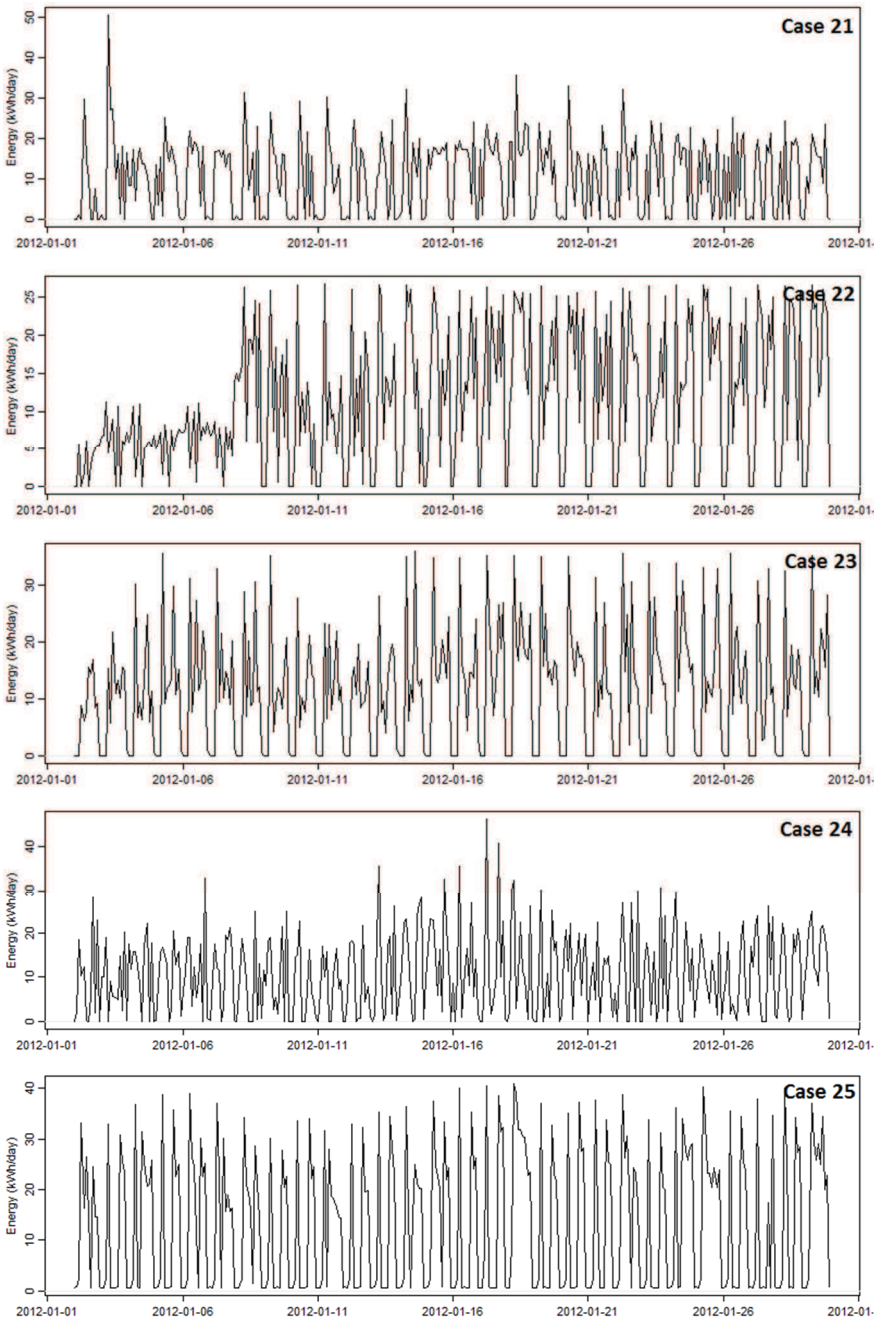

Case 8
Energy (kWh/day)
3
4
5
6
7
8
9
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
2012-01-


Case 9
Energy (kWh/day)
0
5
10
15
20
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
2012-01-


Case 10
Energy (kWh/day)
0
5
10
15
20
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
2012-01-


Case 11
Energy (kWh/day)
0
5
10
15
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 12
Energy (kWh/day)
0
10
20
30
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 13
Energy (kWh/day)
0
5
10
15
20
25
30
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 14
Energy (kWh/day)
0
5
10
15
20
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 15
Energy (kWh/day)
0
5
10
15
20
25
30
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26


Case 16
Energy (kWh/day)
0
5
10
15
20
25
30
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 17
Energy (kWh/day)
0
5
10
15
20
25
30
35
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 18
Energy (kWh/day)
0
10
20
30
40
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 19
Energy (kWh/day)
0
10
20
30
40
50
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 20
Energy (kWh/day)
0
5
10
15
20
25
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26


Case 21
Energy (kWh/day)
0
10
20
30
40
50
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 22
Energy (kWh/day)
0
5
10
15
20
25
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 23
Energy (kWh/day)
0
10
20
30
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 24
Energy (kWh/day)
0
10
20
30
40
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26
Case 25
Energy (kWh/day)
0
10
20
30
40
2012-01-01
2012-01-06
2012-01-11
2012-01-16
2012-01-21
2012-01-26

Weather data

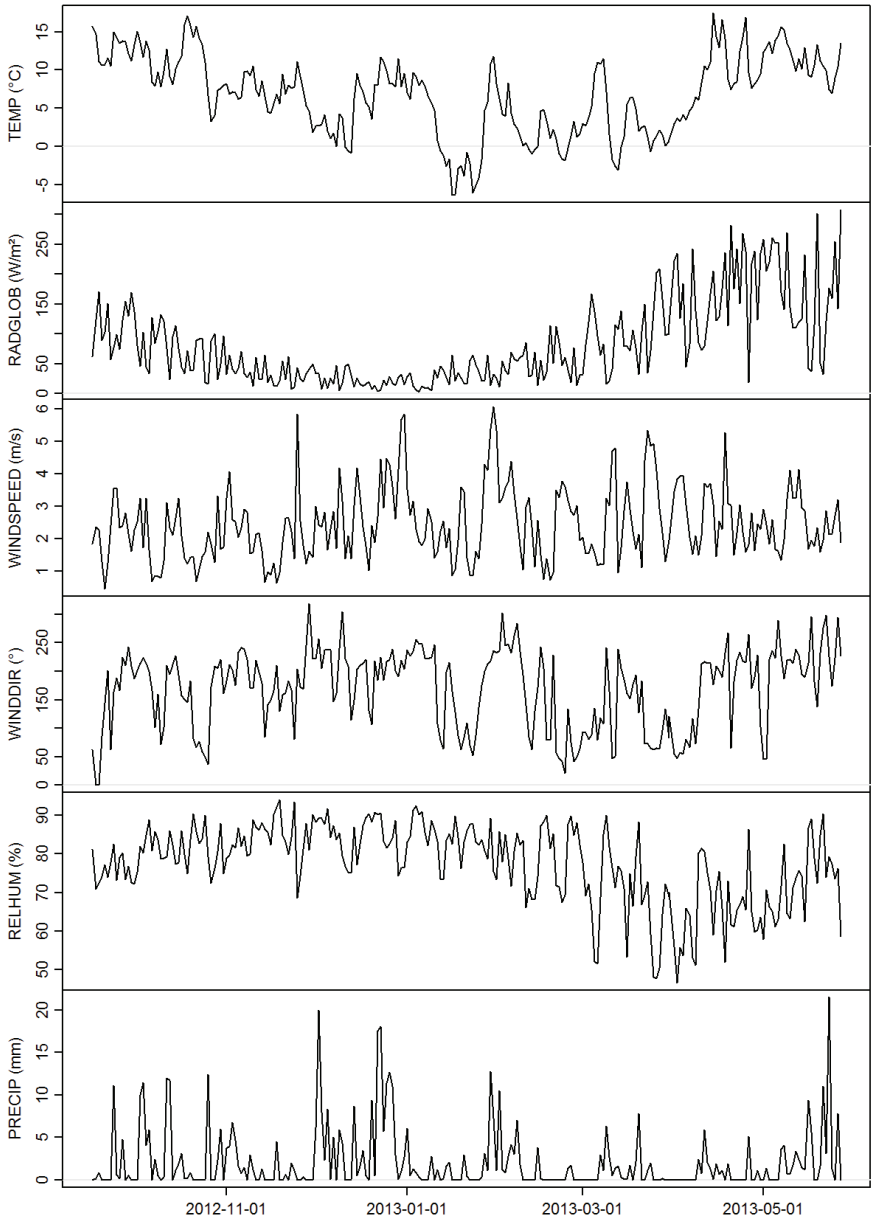


Figure A-3: Climate variables – Heating season 2012-2013 - time series plot

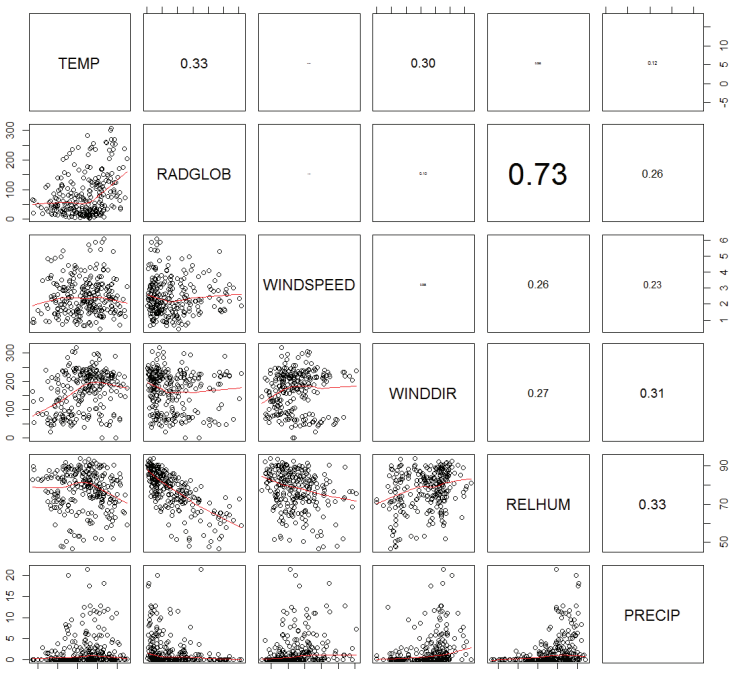


Figure A-4: Climate variables – Heatingseason 2012-2013 – correlation plot

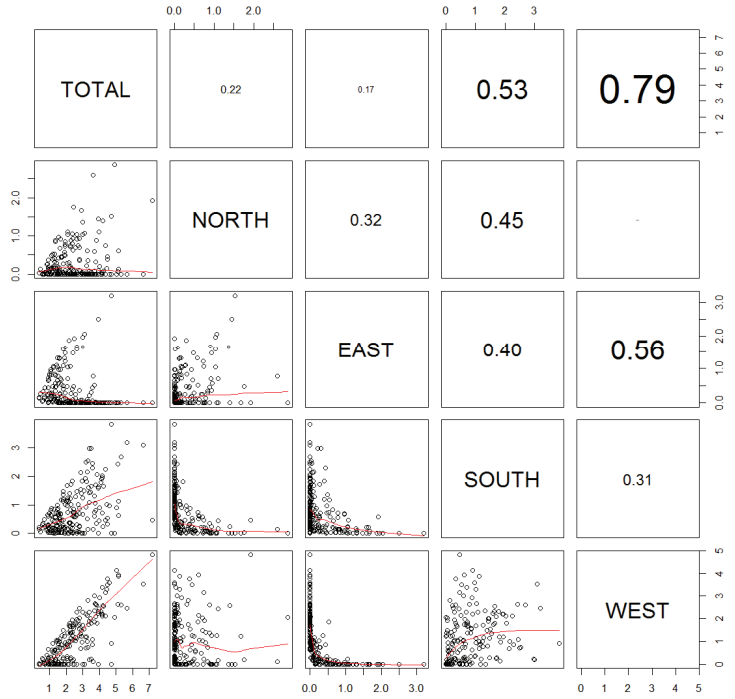


Figure A-5: Windspeed variables – Heatingseason 2011-2012 – correlation plot

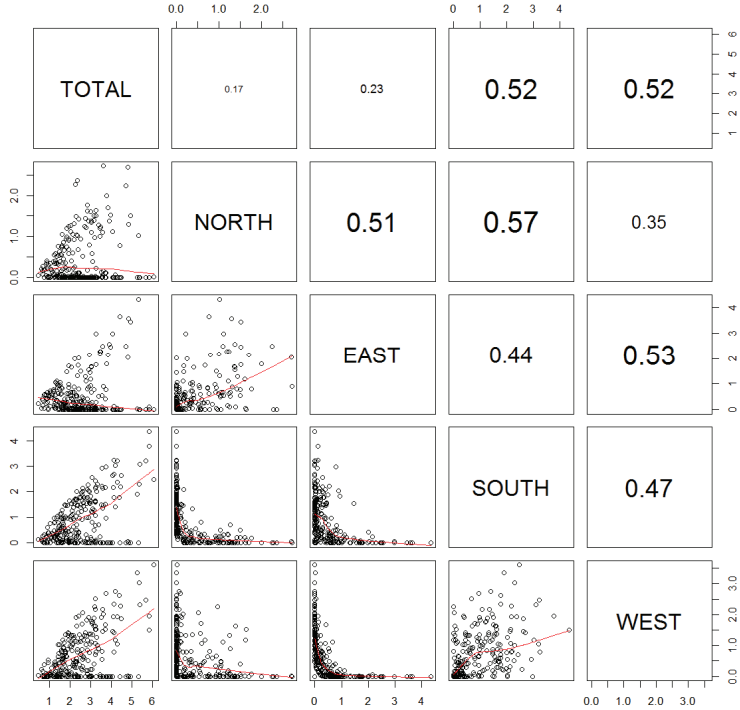


Figure A-6: Windspeed variables – Heatingseason 2012-2013 – correlation plot

Publications

Publications

Journal publications as first author

Eline Himpe, Leen Trappers, Wim Debacker, Marc Delghust, Jelle Laverge, Arnold Janssens, Jan Moens, Marlies Van Holm (2013), *Life Cycle Energy Analysis of a Zero-Energy House,* Building Research & Information, 41(4), 435-449. doi:10.1080/09613218.2013.777329 [117]

Publications in proceedings of conferences as first author

Eline Himpe, Arnold Janssens (2016). *Data-Driven Modelling of the Energy Use in Dwellings Using Smart Meter Data*, CLIMA 2016 - Proceedings of the 12th REHVA World Congress, Vol.6. Aalborg, Denmark: Aalborg University.

Eline Himpe (2015). *Het belang van commissioning in energetisch performante woningen: de CO2-neutrale woonwijk Venning,* Programmaboek van het NZEB-Symposium 2015, 24–25. Gent, Belgium: Passiefhuis-Platform vzw.

Eline Himpe, Arnold Janssens (2015). *Characterisation of the Thermal Performance of a Test House Based on Dynamic Measurements*. Energy Procedia, 6th International Building Physics Conference, IBPC 2015, 78: 3294–3299. doi:10.1016/j.egypro.2015.11.739.

Eline Himpe, Stijn Van de Putte, Jelle Laverge, Arnold Janssens (2015). *Operational Performance of Passive Multi-Family Buildings: Commissioning with Regard to Ventilation and Indoor Climate*,

Energy Procedia, 6th International Building Physics Conference, IBPC 2015, 78: 2983–88. doi:10.1016/j.egypro.2015.11.699.

Eline Himpe, Julio Efrain Vaillant Rebollar, Arnold Janssens (2014). *Heat Losses in Collective Heat Distribution Systems: An Improved Method for EPBD Calculations*, Proceedings from the 14th International Symposium on District Heating and Cooling, edited by Anna Land, 334–41. Stockholm, Sweden: Svensk Fjärrvärme.

Eline Himpe, Julio Vaillant Rebollar, and Arnold Janssens (2013). *Heat Losses in Collective Heat Distribution Systems: Comparing Simplified Calculation Methods with Dynamic Simulations*, Building Simulation 2013 Proceedings, 3432–39. Chambéry, France: International Building Performance Simulation Association (IBPSA).

Eline Himpe, Leen Trappers, Marlies Van Holm, Arnold Janssens, Marc Delghust, Wim Debacker, Jan Moens (2012). *Life Cycle Energy Use of a Zero-Energy House*, I-SUP 2012 Proceedings, 6–10. Bruges, Belgium: VITO.

Book publications as co-author

Henrik Madsen, Peder Bacher, Geert Bauwens, An-Heleen Deconinck, Glenn Reynders, Staf Roels, Eline Himpe, Guillaume Lethé (2016). *Thermal Performance Characterisation Using Time Series Data - Statistical Guidelines.*. In IEA EBC Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements, 83. Subtask 3.2 Report. Belgium: KULeuven.

Publications in proceedings of conferences as co-author

Arnold Janssens, Julio Vaillant Rebollar, Eline Himpe, Marc Delghust, Jelle Laverge (2016). *Transforming Social Housing Neighbourhoods into Sustainable Carbon-Neutral Districts*, MACDES 2016 Cuarto Congreso Internacional, 385–95.

Julio Efrain Vaillant Rebollar, Arnold Janssens, Eline Himpe (2016). *Thermal Comfort and Indoor Air Quality on End-User Satisfaction Level Evaluation in a Nearly Zero Carbon Neighbourhood*. Proceedings of the Indoor Air Conference 2016. Ghent, Belgium: Ghent University.

Julio Efrain Vaillant Rebollar, Arnold Janssens, Eline Himpe (2015), *Energy and Comfort Performance Assessment of Monitored Low Energy Buildings Connected to Low-Temperature District Heating*, Energy Procedia, 6th International Building Physics Conference, IBPC 2015, 78: 3465–3470. doi:10.1016/j.egypro.2015.12.331.

Julio Efrain Vaillant Rebollar, Eline Himpe, Jelle Laverge, Arnold Janssens (2015). *Influence of Recirculation Strategies in Collective Heat Distribution System on the Performance of Dwelling Heating Substations*, Proceedings of the VIII International Conference for Renewable Energy, Energy Saving and Energy Education.

Arnold Janssens, Alcamo Giuseppina, Bacher Peder, Erkoreka Aitor, Flamant Gilles, Eline Himpe (2014). *State of the Art on Full Scale Testing and Dynamic Data Analysis for Building Energy Performance Characterisation*, Proceedings of the Real Building Energy Performance Assessment Seminar, 23–30. Ghent, Belgium: DYNASTEE.

Julio Efrain Vaillant Rebollar, Eline Himpe, Arnold Janssens (2014). *Building Energy Performance Characterisation Based on Dynamic Analysis and Co-Heating Test*, MACDES.2014 Memorias. La Habana, Cuba: CUJAE.

Julio Efrain Vaillant Rebollar, Eline Himpe, Arnold Janssens (2014). *Simulation Models and Performance Assessment of District Heating Substations*, MACDES.2014 Memorias, 292–303. La Habana, Cuba: CUJAE.

Arnold Janssens, Giuseppina Alcamo, Peder Bacher, Aitor Erkoreka, Gilles Flamant, Eline Himpe (2014). *Full Scale Testing and Dynamic Data Analysis for Building Energy Performance Assessment: State of the Art*, MACDES.2014 Memorias, 50–57. La Habana, Cuba: CUJAE.

Julio Efrain Vaillant Rebollar, Eline Himpe, Arnold Janssens (2014). *Performance Assessment of District Heating Substations Based on Dynamic Simulations*. Proceedings from the 14th International Symposium on District Heating and Cooling, 317–24. Stockholm, Sweden: Svensk Fjärrvärme.

Julio Efrain Vaillant Rebollar, Eline Himpe, and Arnold Janssens (2013), *Performance Evaluation of a Low Temperature District Heating System Based on Simulation, Uncertainty and Sensitivity Analysis,* Building Simulation 2013 Proceedings, 3809–3816. Chambéry, France: International Building Performance Simulation Association (IBPSA).

Julio Efrain Vaillant Rebollar, Eline Himpe, and Arnold Janssens (2013), *Potentials of the New Design Concepts of District Heating and Cooling toward Integration with Renewable Energy Sources*, Proceedings of the 12th World Wind Energy Conference & Renewable Energy Exhibition, 490–95. La Habana, Cuba: World Wind Energy Association.

Marlies Van Holm, Eline Himpe, Leen Trappers, Wim Debacker, Arnold Janssens, Marc Delghust, Jan Moens (2012). *The Life-Cycle Energy Consumption of Zero-Energy Houses,* Proceedings of the 15th European Roundtable on Sustainable Consumption and Production, 421–28. Bregenz, Austria: Austrian Institute of Ecology.

References

[1] United Nations Framework Convention on Climate Change, Kyoto Protocol to the United Nations Framework Convention on Climate Change, 1997. http://unfccc.int/kyoto_protocol/items/2830.php.

[2] Directive 2010/31/EU of the European Parliament and of the Council of 19 may 2010 on the energy performance of buildings (recast), 2010. http://data.europa.eu/eli/dir/2010/31/oj.

[3] European Parliament and the Council of the European Union, Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, 2012. http://data.europa.eu/eli/dir/2012/27/oj.

[4] Vlaamse Regering, het Energiebesluit van 19 november 2010, 2010. https://codex.vlaanderen.be/.

[5] Vlaamse Regering, Derde Vlaams actieplan Energie-efficiëntie, 2014. http://www2.vlaanderen.be/economie/energiesparen/beleid/Energie-efficientierichtlijn/Derde-Vlaams_actieplan_energie-efficientie.pdf.

[6] A. Rabl, Application of parameter estimation in buildings: an overview, in: H. Bloem (Ed.), System Identification Applied to Building Performance Data, Office for Official Publications of the European Communities, Luxembourg, 1994.

[7] Concerted Action EPBD, 2016 - Implementing the Energy Performance of Buildings Directive (EPBD) - Featuring Country Reports, ADENE, Lisbon, 2015. www.epbd-ca.eu.

[8] European Committee for Standardization, EN 15603 - Energy performance of buildings - overall energy use and definition of energy ratings, 2008.

[9] International Standardization Organisation, EN ISO 13790 - Energy Performance of buildings - calculation of energy use for space heating and cooling, 2007.

[10] H. Hens, W. Parijs, M. Deurinck, Energy consumption for heating and rebound effects, Energy and Buildings. 42 (2010) 105–110. doi:10.1016/j.enbuild.2009.07.017.

[11] M. Sunikka-Blank, R. Galvin, Introducing the prebound effect: the gap between performance and actual energy consumption, Building Research & Information. 40 (2012) 260–273. doi:10.1080/09613218.2012.690952.

[12] D. Majcen, L.C.M. Itard, H. Visscher, Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications, Energy Policy. 54 (2013) 125–136. doi:10.1016/j.enpol.2012.11.008.

[13] E. Himpe, J. Vaillant Rebollar, A. Janssens, Heat losses in collective heat distribution systems: comparing simplified calculation methods with dynamic simulations, in: Building Simulation 2013 Proceedings, International Building Performance Simulation Association (IBPSA), Chambéry, France, 2013: pp. 3432–3439. http://hdl.handle.net/1854/LU-4124565.

[14] M. Delghust, Y. De Weerdt, A. Janssens, Zoning and intermittency simplifications in quasi-steady state models, Energy Procedia. 78 (2015) 2995–3000.

[15] E. Himpe, S. Van de Putte, J. Laverge, A. Janssens, Operational Performance of Passive Multi-family Buildings: Commissioning with Regard to Ventilation and Indoor Climate, Energy Procedia. 78 (2015) 2983–2988. doi:10.1016/j.egypro.2015.11.699.

[16] J.C. Visier, ed., Commissioning tools for improved energy performance, in: IEA-ECBCS Annex 40: Commissioning of Building HVAC Systems for Improved Energy Performance, IEA ECBCS, 2004. http://www.ecbcs.org/annexes/annex40.htm.

[17] J. Vaillant Rebollar, E. Himpe, A. Janssens, Energy and Comfort Performance Assessment of Monitored Low Energy Buildings Connected to Low-temperature District Heating, Energy Procedia. 78 (2015) 3465–3470. doi:10.1016/j.egypro.2015.12.331.

[18] M. Delghust, Improving the predictive power of simplified residential space heating demand models: a field data and model driven study, Doctoral thesis, Ghent University, 2015.

[19] M. Van Campen, Intelligente systemen voor communicatie over het energiegebruik in een CO2-neutrale woonwijk, Master thesis, Ghent University, 2013.

[20] S. Darby, Social learning and public policy: Lessons from an energy-conscious village, Energy Policy. 34 (2006) 2929–2940. doi:10.1016/j.enpol.2005.04.013.

[21] T. Ueno, R. Inada, O. Saeki, K. Tsuji, Effectiveness of an energy-consumption information system for residential buildings, Applied Energy. 83 (2006) 868–883. doi:10.1016/j.apenergy.2005.09.004.

[22] S. Wang, C. Yan, F. Xiao, Quantitative energy performance assessment methods for existing buildings, Energy and Buildings. 55 (2012) 873–888. doi:10.1016/j.enbuild.2012.08.037.

[23] S. Van de Putte, Energieprestatie van een CO2-neutrale woonwijk: commissioning van de uitvoeringskwaliteit en het gebruik van het gebouw en de installaties (Energy performance of a carbon neutral neighbourhood: commissioning of the workmanship and the use of the building and its installations), Master dissertation, Ghent University, 2014.

[24] M. Santamouris, ed., Energy Performance of Residential buildings - A practical guide for energy rating and efficiency, Earthscan, London, 2009.

[25] European Commission, Commission Recommendation on preparations for the roll-out of smart metering systems, 2012. http://data.europa.eu/eli/reco/2012/148/oj.

[26] Eandis, Infrax, Proof of concept Slimme meters: Eindrapport, Eandis, Melle, Belgium, 2012.

[27] Eandis, Infrax, Piloot Slimme meters: Eindrapport, Eandis, Melle, Belgium, 2014.

[28] I. Vassileva, E. Dahlquist, F. Wallin, J. Campillo, Energy consumption feedback devices' impact evaluation on domestic energy use, Applied Energy. 106 (2013) 314–320. doi:10.1016/j.apenergy.2013.01.059.

[29] S. Hammarsten, A critical appraisal of energy-signature models, Applied Energy. 26 (1987) 97–110. doi:10.1016/0306-2619(87)90012-2.
[30] M.F. Fels, PRISM: An introduction, Energy and Buildings. 9 (1986) 5–18. doi:10.1016/0378-7788(86)90003-4.
[31] F. Flouquet, Local weather correlations and bias in building parameter estimates from energy-signature models, Energy and Buildings. 19 (1992) 113–123. doi:10.1016/0378-7788(92)90005-2.
[32] ISO14040:2006, Environmental management - Life cycle assessment - Principles and Framework, International Organisation for Standardisation, Geneva, 2006.
[33] E. Himpe, L. Trappers, De totale energieconsumptie van een nulenergiewoning: impact van de gebouwschil en de technische installaties, Master Thesis, Ghent University, 2011.
[34] G. Bauwens, S. Roels, Co-heating test: A state-of-the-art, Energy and Buildings. 82 (2014) 163–172.
[35] D.L. Minehart, A.K. Meier, Using synthetic data to explore the usefulness of prism's parameters at inferring causes of changes in normalized annual consumption, Energy. 19 (1994) 135–148. doi:10.1016/0360-5442(94)90055-8.
[36] A.K. Meier, J. Busch, C.C. Conner, Testing the accuracy of a measurement-based building energy model with synthetic data, Energy and Buildings. 12 (1988) 77–82. doi:10.1016/0378-7788(88)90057-6.
[37] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews. 16 (2012) 3586–3592. doi:10.1016/j.rser.2012.02.049.
[38] R. Kumar, R.K. Aggarwal, J.D. Sharma, Energy analysis of a building using artificial neural network: A review, Energy and Buildings. 65 (2013) 352–358. doi:10.1016/j.enbuild.2013.06.007.
[39] T. Söderström, Theory of System Identification: An introduction and overview, in: H. Bloem (Ed.), System Identification Applied to Building Performance Data, Office for Official Publications of the European Communities, Luxembourg, 1994.
[40] H. Madsen, P. Bacher, G. Bauwens, A.-H. Deconinck, G. Reynders, S. Roels, E. Himpe, G. Lethé, Thermal Performance Characterisation using Time Series Data - Statistical Guidelines, in: IEA EBC Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements, KULeuven, Belgium, 2016: p. 83.
[41] T.A. Reddy, D.E. Claridge, Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption, Energy and Buildings. 21 (1994) 35–44. doi:10.1016/0378-7788(94)90014-0.
[42] K.-E. Westergren, H. Högberg, U. Norlén, Monitoring energy consumption in single-family houses, Energy and Buildings. 29 (1999) 247–257. doi:10.1016/S0378-7788(98)00065-6.
[43] C. Ghiaus, Experimental estimation of building energy performance by robust regression, Energy and Buildings. 38 (2006) 582–587. doi:10.1016/j.enbuild.2005.08.014.
[44] J.J. Bloem, A pseudo-dynamic approach to building energy performance assessment - requirements for experimental work in relation to evaluation, in: 2011. http://iet.jrc.ec.europa.eu/energyefficiency/workshop/international-workshop-whole-building-testing-evaluation-and-modelling-energy-assessment.
[45] J.J. Bloem, S. Martin, A pseudo dynamic analysis tool for thermal certification of dwellings, Energy and Buildings. 33 (2001) 207–212. doi:10.1016/S0378-7788(00)00083-9.

[46] S. Danov, J. Carbonell, J. Cipriano, J. Martí-Herrero, Approaches to evaluate building energy performance from daily consumption data considering dynamic and solar gain effects, Energy and Buildings. 57 (2013) 110–118. doi:10.1016/j.enbuild.2012.10.050.

[47] S.B. Mortensen, H.A. Nielsen, Analysis of energy consumption in single family houses, in: 2011. http://iet.jrc.ec.europa.eu/energyefficiency/workshop/international-workshop-whole-building-testing-evaluation-and-modelling-energy-assessment.

[48] H. Madsen, Time Series Analysis, Chapman & Hall/CRC, Boca Raton, Florida, 2008.

[49] M.J. Jimenez, Thermal peformance characterization based on full scale testing - description of the common excercises and physical guidelines, in: IEA EBC Annex 58: Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements, KULeuven, Belgium, 2016: p. 176. http://www.kuleuven.be/bwf/projects/annex58/index.htm.

[50] S. Roels, P. Bacher, G. Bauwens, H. Madsen, M.J. Jiménez, Characterising the Actual Thermal Performance of Buildings: Current Results of Common Exercises Performed in the Framework of the IEA EBC Annex 58-Project, Energy Procedia. 78 (2015) 3282–3287. doi:10.1016/j.egypro.2015.11.726.

[51] E. Himpe, A. Janssens, Characterisation of the Thermal Performance of a Test House Based on Dynamic Measurements, Energy Procedia. 78 (2015) 3294–3299. doi:10.1016/j.egypro.2015.11.739.

[52] B. Soldo, P. Potočnik, G. Šimunović, T. Šarić, E. Govekar, Improving the residential natural gas consumption forecasting models by using solar radiation, Energy and Buildings. 69 (2014) 498–506. doi:10.1016/j.enbuild.2013.11.032.

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

[54] P.D. Andersen, M.J. Jimenez, H. Madsen, C. Rode, Characterisation of heat dynamics of arctic low-energy house with floor heating, in: 2013: p. 24.

[55] P.D. Andersen, C. Rode, H. Madsen, An arctic low-energy house as experimental setup for studies of heat dynamics of buildings, Frontiers of Architectural Research. (2013). doi:10.1016/j.foar.2013.08.003.

[56] G. Bauwens, In situ testing of a building's overall heat loss coefficient: embedding quasi-stationary and dynamic tests in a building physical and statistical framework, Doctoral thesis, KULeuven, 2015.

[57] P.A. de Saint-Aubain, Adaptive Load Forecasting, Master dissertation, Technical University of Denmark, 2011.

[58] P.A. de Saint-Aubain, Non-parametric Method for Separating Hot Water Spikes from Total Heating Consumption Data, (2012) 13.

[59] R.A.R. Kilpatrick, P.F.G. Banfill, D.P. Jenkins, Methodology for characterising domestic electrical demand by usage categories, Applied Energy. 88 (2011) 612–621. doi:10.1016/j.apenergy.2010.08.002.

[60] K. Basu, V. Debusschere, A. Douzal-Chouakria, S. Bacha, Time series distance-based methods for non-intrusive load monitoring in residential buildings, Energy and Buildings. 96 (2015) 109–117. doi:10.1016/j.enbuild.2015.03.021.

[61] G.M. Huebner, M. McMichael, D. Shipworth, M. Shipworth, M. Durand-Daubin, A.J. Summerfield, The shape of warmth: temperature profiles in living rooms, Building Research & Information. (2014) 1–12. doi:10.1080/09613218.2014.922339.

[62] D. Aerts, Occupancy and Activity Modeling for Building Energy Demand Simulations, Comparative Feedback and Residential Electricity Demand Characterisation, Doctoral thesis, Vrije Universiteit Brussel, 2015.
[63] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, F. Descamps, A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, Building and Environment. 75 (2014) 67–78. doi:10.1016/j.buildenv.2014.01.021.
[64] Z. Yang, B. Becerik-Gerber, Modeling personalized occupancy profiles for representing long term patterns by using ambient context, Building and Environment. 78 (2014) 23–35. doi:10.1016/j.buildenv.2014.04.003.
[65] P. Bacher, H. Madsen, H.A. Nielsen, B. Perers, Short-term heat load forecasting for single family houses, Energy and Buildings. 65 (2013) 101–112. doi:10.1016/j.enbuild.2013.04.022.
[66] J.E. Seem, J.E. Braun, Adaptive methods for real-time forecasting of building electrical demand, in: ASHRAE Transactions, 1991.
[67] T. Mestekemper, G. Kauermann, M.S. Smith, A comparison of periodic autoregressive and dynamic factor models in intraday energy demand forecasting, International Journal of Forecasting. 29 (2013) 1–12. doi:10.1016/j.ijforecast.2012.03.003.
[68] T. Mestekemper, M. Windmann, G. Kauermann, Functional hourly forecasting of water temperature, International Journal of Forecasting. 26 (2010) 684–699. doi:10.1016/j.ijforecast.2009.10.010.
[69] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015. http://www.R-project.org/.
[70] SYNERGRID, De graaddagen gebruik in België voor de facturatie en de normalisatie van de gasafgiften, (2006). http://www.aardgas.be/nl/particulier/graaddagen (accessed August 22, 2016).
[71] M.H. Kutner, C.J. Nachtsheim, J. Neter, W. Li, Applied Linear Statistical Models, 5th ed., McGraw-Hill/Irwin, New York, 2005.
[72] J. Gareth, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, New York, 2013.
[73] R.J. Hyndman, A.B. Koehler, Another look at measures of forecast accuracy, International Journal of Forecasting. 22 (2006) 679–688. doi:10.1016/j.ijforecast.2006.03.001.
[74] R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles & Practice, OTexts (www.otexts.org), Melbourne, Australia, 2013. www.otexts.org/fpp (accessed September 2, 2016).
[75] D.A. Finan, F.J. Doyle, C.C. Palerm, W.C. Bevier, H.C. Zisser, L. Jovanovič, D.E. Seborg, Experimental Evaluation of a Recursive Model Identification Technique for Type 1 Diabetes, Journal of Diabetes Science and Technology (Online). 3 (2009) 1192–1202.
[76] R.J. Hyndman, A.B. Koehler, J.K. Ord, R.D. Snyder, Forecasting with Exponential Smoothing: The State Space Approach, Springer, Berlin, 2008.
[77] R.H. Shumway, D.S. Stoffer, Time Series Analysis and Its Applications with R Examples, 3rd ed., Springer, New York, 2011.
[78] D.W. Bunn, A.I. Vassilopoulos, Using group seasonal indices in multi-item short-term forecasting, International Journal of Forecasting. 9 (1993) 517–526. doi:10.1016/0169-2070(93)90078-2.
[79] P. Linde, Seasonal Adjustment, Statistics Denmark, Denmark, 2005. http://www.dst.dk/en/Statistik/dokumentation/metode/saesonkorrektion.

[80] M.K.P. So, R.S.W. Chung, Dynamic seasonality in time series, Computational Statistics & Data Analysis. 70 (2014) 212–226. doi:10.1016/j.csda.2013.09.010.
[81] R.B. Cleveland, W.S. Cleveland, J.E. McRae, I. Terpenning, STL: A Seasonal-Trend Decomposition Procedure Based on Loess, Journal of Official Statistics. 6 (1990) 31.
[82] J. Wang, W. Zhang, J. Wang, T. Han, L. Kong, A novel hybrid approach for wind speed prediction, Information Sciences. 273 (2014) 304–318. doi:10.1016/j.ins.2014.02.159.
[83] M.C. Bas, J. Ortiz, L. Ballesteros, S. Martorell, Analysis of the influence of solar activity and atmospheric factors on 7Be air concentration by seasonal-trend decomposition, Atmospheric Environment. 145 (2016) 147–157. doi:10.1016/j.atmosenv.2016.09.027.
[84] J. Wu, J. Wang, H. Lu, Y. Dong, X. Lu, Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model, Energy Conversion and Management. 70 (2013) 1–9. doi:10.1016/j.enconman.2013.02.010.
[85] E.F. Sánchez-Úbeda, A. Berzosa, Modeling and forecasting industrial end-use natural gas consumption, Energy Economics. 29 (2007) 710–742. doi:10.1016/j.eneco.2007.01.015.
[86] A. De Livera, R.J. Hyndman, R.D. Snyder, Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing, Journal of the American Statistical Association. 106 (2011) 1513–1527.
[87] A.M. De Livera, Modeling Time Series with Complex Seasonal Patterns using Exponential Smoothing, Doctoral thesis, Monash University, Department of Economics and Business Statistics, 2010.
[88] M. Kendall, A. Stuart, The Advanced Theory of Statistics: Design and Analysis, and Time-Series, 3rd ed., Charles Griffin & Company Limited, London, 1976.
[89] R.J. Hyndman, The interaction between trend and seasonality, International Journal of Forecasting. 20 (2004) 561–563. doi:10.1016/j.ijforecast.2004.03.005.
[90] Y. Zhao, R and Data Mining - Examples and Case Studies, 1st ed., Elsevier Academic Press, USA, 2013. www.rdatamining.com.
[91] R.J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts.org/Fpp/. (2013). http://otexts.org/fpp/ (accessed September 2, 2016).
[92] A.M. De Livera, Automatic forecasting with a modified exponential smoothing state space framework, Monash University, Australia, 2010.
[93] A. Dokumentov, R.J. Hyndman, STR: A Seasonal-Trend Decomposition Procedure Based on Regression, Monash University, Australia, 2015.
[94] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, USA, 1990.
[95] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd ed., Springer, New York, 2009.
[96] J. Paparrizos, L. Gravano, k-Shape: Efficient and Accurate Clustering of Time Series, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Australia, 2015: pp. 1855–1870. doi:10.1145/2723372.2737793.
[97] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. Pérez, I. Perona, An extensive comparative study of cluster validity indices, Pattern Recognition. 46 (2013) 243–256. doi:10.1016/j.patcog.2012.07.021.
[98] B.S. Everitt, S. Landau, M. Leese, Cluster Analysis, 4th ed., Arnold (Hodder Education), London, 2001.
[99] G.W. Milligan, M.C. Cooper, An examination of procedures for determining the number of clusters in a data set, Psychometrika. 50 (1985) 159–179. doi:10.1007/BF02294245.

[100] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics. 20 (1987) 53–65. doi:10.1016/0377-0427(87)90125-7.
[101] F. Itakura, Minimum Prediction Residual Principle applied to Speech Recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing. ASSP-23 (1975) 67–72.
[102] H. Sakoe, S. Chiba, Dynamic Programming Algorithm Optimization for Spoken Word Recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing. ASSP-26 (1978) 43–49.
[103] C. Myers, L. Rabiner, A.E. Rosenberg, Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing. ASSP-28 (1980) 623–635.
[104] L. Rabiner, B.-H. Juang, Fundamentals of Speech Recognition, Prentice Hall International Inc., New Jersey (USA), 1993.
[105] E. Keogh, C.A. Ratanamahatana, Exact indexing of dynamic time warping, Knowledge and Information Systems. (2004). doi:10.1007/s10115-004-0154-9.
[106] C.A. Ratanamahatana, E. Keogh, Making Time-Series Classification More Accurate using Learned Constraints, in: Proceedings of the SIAM International Conference on Data Mining, Florida, USA, 2004. http://alumni.cs.ucr.edu/~ratana/.
[107] T. Giorgino, Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package, Journal of Statistical Software. 31 (2009) 24. doi:doi:10.18637/jss.v031.i07.
[108] M. Muller, Dynamic Time Warping, in: Information Retrieval for Music and Motion, Springer-Verlag, Berlin Heidelberg, 2007: pp. 68–84.
[109] A. Sarda-Espinosa, Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package, in: R-Package Version 3.1.0, URL: Https://CRAN.R-Project.org/Package=dtwclust, 2017. https://CRAN.R-project.org/package=dtwclust.
[110] N. Begum, L. Ulanova, J. Wang, E. Keogh, Accelerating Dynamic Time Warping Clustering with a Novel Admissible Pruning Strategy, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 2015: pp. 49–58. doi:10.1145/2783258.2783286.
[111] A. Zinke, D. Mayer, Iterative Multi Scale Dynamic Time Warping, Universität Bonn, Bonn, 2006. http://cg.cs.uni-bonn.de/aigaion2root/attachments/cg-2006-1.pdf.
[112] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, cluster: Cluster Analysis Basics and Extensions, 2015. https://cran.r-project.org/web/packages/cluster.
[113] C. Hennig, fpc: Flexible Procedures for Clustering, 2015. http://CRAN.R-project.org/package=fpc.
[114] A. Agresti, An Introduction to Categorical Data Analysis, 2nd ed., Wiley, Hoboken, New Jersey, 2007.
[115] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, 4th ed., Springer, New York, 2002.
[116] J. Fox, Applied Regression Analysis and Generalized Linear Models, 2nd ed., Sage, USA, 2008.
[117] E. Himpe, L. Trappers, W. Debacker, M. Delghust, J. Laverge, A. Janssens, J. Moens, M. Van Holm, Life cycle energy analysis of a zero-energy house, BR&I. 41 (2013) 435–449.