Dipl.-Ing. Christian Elbe

A Contribution to Residential Energy Disaggregation Based on Appliance-Specific Characteristics

Thesis for the degree of Doctor of Engineering Sciences

Institute of Electrical Power Systems
Graz University of Technology

Supervisor
Univ.-Prof. DI Dr.techn. Lothar Fickert

Reviewer
Univ.-Prof. DI Dr.techn. Lothar Fickert
Graz University of Technology

Reviewer
Univ.-Prof. DI Dr.-Ing. habil. Christian Rehtanz
TU Dortmund University

Head of Institute: Univ.-Prof. DI Dr.techn. Lothar Fickert

8010 Graz, Inffeldgasse 18/I
Phone: +43 316 873 – 7551
Fax: +43 316 873 – 7553
http://www.ifea.tugraz.at
http://www.tugraz.at

Graz, April 2014

To my dear mother.

Statutory Declaration

I declare that I have authored this doctoral thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral dissertation.

Graz, April, 2014

Abstract

It is currently nearly impossible for domestic consumers to recognize which electric devices have a significant share to the total electricity consumption. Reason for this is on the one hand the total number of used electrical appliances in homes and on the other hand that consumers only receive information about whether they pay more or less than in the previous billing period. Hence, different feedback methods and their advantages and disadvantages are examined in this thesis. Additionally, it is investigated in different appliance-specific electricity consumption monitoring methods what includes their applicability and limitations.

This thesis presents a novel unsupervised approach to non-intrusive appliance load monitoring (NIALM). By the term NIALM the energy disaggregation from an aggregated load curve to its sources is meant.
The basic idea of the introduced *fEEDBACk* method is to filter typical load curves of selected electrical devices from the aggregated load curve gathered by smart meters. To realize this, changes with recurring patterns in the load profile are analyzed and recognized by device-specific patterns. Prerequisites of the *fEEDBACk* algorithm are the detection of events as well as the knowledge about the characteristics of different types of appliances. Therefore both mentioned topics are investigated in detail in this thesis.
Typical patterns such as on- and off-duration, active power consumption, and time of use are extracted from electrical appliances (fridge, freezer, washing machine, dishwasher) which are measured in 40 households by the project ADRES. These characteristics can be used to improve the energy disaggregation process by building general appliance models.
It is investigated in the outcome of different event detection algorithms which are analyzed by different metrics. The optimal parameter selection as well as the influence of different sampling periods on the event detectors is studied. For computing the power changes of events (steady state power changes) a novel algorithm is introduced which better computes the true values compared to state of the art algorithms.

The performance of the *fEEDBACk* algorithm is tested under different sampling periods which range from 1 to 20 seconds. It is shown that the performance of the disaggregation process of cooling devices reach a score in the range of 0.87 to 0.99 in the investigated homes. The incorporation of additional features such as the reactive power values further improve the resulted score, in some cases up to 0.06. Above all, the opportunities of the analysis of 15-minute power averages of smart meters are investigated.

Kurzfassung

Für Haushaltsstromkunden ist es derzeit fast unmöglich zu erkennen, welche Elektrogeräte einen bedeutenden Anteil am Gesamtstromverbrauch haben. Grund dafür ist einerseits die große Anzahl an Elektrogeräte in den Haushalten und andererseits, dass die Stromkunden lediglich Informationen darüber bekommen, ob sie einen größeren oder kleineren Stromverbrauch im Vergleich zur vorherigen Abrechnungsperiode haben. Deshalb werden in dieser Arbeit für ein verbessertes Feedback verschiedene Methoden bezüglich Aufschlüsselung des Stromverbrauchs sowie deren Vor- und Nachteile untersucht. Darüber hinaus werden Möglichkeiten zur Messung des gerätespezifischen Stromverbrauchs sowie deren Einsatzmöglichkeiten und auch Beschränkungen untersucht.

Diese Arbeit stellt einen neuen Ansatz zur Lastganganalyse dar. Mit dem Begriff Lastganganalyse ist eine Auftrennung des Gesamtlastganges in die einzelnen Lastgänge der beinhalteten Elektrogeräte gemeint. Die Grundidee der in dieser Arbeit vorgestellten *fEEDBACk* Methode ist es, die typischen Lastgänge von ausgewählten elektrischen Geräten aus dem von Smart Metern unter bestimmten Bedingungen gemessenen Gesamtlastgang zu filtern. Dafür werden wiederholt auftretende Muster im Lastgang analysiert und mit gerätespezifischen Mustern verglichen. Voraussetzungen für den *fEEDBACk* Algorithmus sind einerseits die Erkennung von Änderungen im Lastgang, die durch eine Zustandsänderung von einem Elektrogeräten hervorgerufen werden sowie die Kenntnis der gerätespezifischen Eigenschaften der einzelnen Typen von Elektrogeräten. Deshalb werden in dieser Arbeit diese beiden Themen im Detail untersucht.

Typische Verbrauchsmuster von Elektrogeräten wie die Ein- und Ausschaltzeitdauer sowie die Wirkleistungsaufnahme und auch der genaue Einsatzzeitpunkt von Kühlschranken, Gefrierschränken, Waschmaschinen und Geschirrspülern werden exemplarisch untersucht. Die Messdaten von 40 Haushalten wurden dazu verwendet. Mit Hilfe dieser gerätespezifischen Eigenschaften können Modelle von Elektrogeräten erstellt werden, die die Erkennungsgenauigkeit bei der Lastganganalyse wesentlich erhöhen.

Zudem werden unterschiedliche Detektoren und deren Erkennungsgenauigkeit zur Ermittlung von Zustandsänderungen von Elektrogeräten im Lastgang untersucht. Dabei wird die optimale Parameterwahl sowie der Einfluss von verschiedenen Abtastperioden berücksichtigt. Außerdem wird ein neuartiger Algorithmus zur genauen Berechnung der Leistungsänderungen beschrieben.

Die Genauigkeit des *fEEDBACk* Algorithmus wird bei Abtastperioden zwischen 1 und 20 Sekunden untersucht. Es wird gezeigt, dass die Genauigkeit bei der Lastauftrennung von Kühlgeräten einen F-Maß-Wert im Bereich von 0,87 bis 0,99 erreicht. Die Berücksichtigung der Blindleistungswerte kann eine weitere Genauigkeitsverbesserung um bis zu 0,06 erzielen. Zudem wird auf die Möglichkeiten der Analyse von 15-Minuten-Leistungs-Mittelwerten von Smart Metern eingegangen.

Contents

List of Abbreviations and Symbols

ADRES	Autonomic Decentralized Regenerative Energy System
ATP	Accurate true positive
$a_j(D)$	Number of all extracted power sets $\mathbf{P}_{i,j}^{\ 0}$ for a state j for the device D
$b_j(y)$	Emission probability of state j and output y
$\mathbf{B}_C(D)$	Clustered block from $\mathbf{B}$“(D)
B_m	Power block m
$\mathbf{B}'(D)$	Power blocks assigned to device D
$\mathbf{B}''(D)$	Power blocks assigned to device D with events which not more than s-times used
BLUED	Building-Level fUlly-labeled dataset for Electricity Disaggregation
c_P	Factor for active power of distance function
c_t	Factor for time duration of distance function
d	Distance between two data points
$d_j(u)$	State duration distribution of state j with time duration u
d_P	Distance between two power values
D	Electrical device to be disaggregated
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
D_μ	Euclidean distance for metric μ
$e(t)$	Appliance-specific transition at time step t
$\mathbf{E}'_{all}(D)$	Set of all relevant events $E'_{all,i}(D)$ of device D
$E'_{all,i}(D)$	Relevant event i of device D
E_i	Event i
E_i'	Merged event i
E_k^+	Event k with a positive total active power change $\Delta P(E_i)$
E_l^-	Event l with a negative total active power change $\Delta P(E_i)$
$\mathcal{F}$	False positive events
fEEDBACk	REsidential Energy Disaggregation Based on Appliance-specific Characteristics
FHMM	Factorial hidden Markov model
FN	False negative
FP	False positive
FPP	False positive percentage
FPR	False positive rate
GLR	Generalized likelihood ratio
GOF	Goodness of fit

HAN	Home area network
HDP-HSMM	Hierarchical Dirichlet process hidden semi-Markov model
HMM	Hidden Markov model
HSMM	Hidden semi-Markov model
i	Interest rate
ITP	Inaccurate true positive
J	Number of all states of a Markov chain
K_0	Net present value
K_j	Annual savings for the year j
l_t	Likelihood ratio of point of interest at time step t for event detection
L	Limit for maximum error
$\mathcal{L}$	Log-likelihood
M	Metric for comparison of true and steady state power changes
$\mathcal{M}$	Missing events
$M_j(D)$	Maximum length of the state duration of state j
$M_{on}(D)$	Maximum on-duration of a certain device D
n_0	Minimum sample size
n_E	Number of all events
N_i	Number of power changes which belong to event E_i
NIALM	Non-Intrusive Appliance Load Monitoring
N_ε	Number of elements within the radius ε
O	Computational complexity
p_{ij}	State transition probability from state i to state j
$\mathbf{P}$	Set of active power readings
$P(B_m)$	Continuous active power of block B_m
$P(t)$	Active power reading at time step t
$\hat{P}(t,D)$	Estimated active power value of device D at time step t
$P^0(t,D)$	Ground truth active power value of device D at time step t
P_{230}	Normalized active power value (230V)
$P_{Avg}(t)$	Average active power value within a certain window
$P_E(B_m)$	Continuous active power incorporating the associated events of block B_m
$\mathbf{P}_{i,j}^{\ 0}$	Set i of ground truth active power values $P_j^0(t,D)$ for state j
$P_j^0(t,D)$	Ground truth active power value of device D at time step t for state j
$\hat{P}_j(t,D)$	Estimated active power value of device D at time step t for state j
$P_{k,l}^{\ cont}$	Continuously active power value between event E_k and E_l
PLC	Power line communication

P_m	Measured active power value
$P_{max}(D)$	Maximal continuous active power consumption of device D
P_{min}	Minimum active power change for event detection
$P_{min}(D)$	Minimal continuous active power consumption of device D
$\mathbf{P}_{Post}$	Vector of active power readings of post-event window
$P_{Post}(t)$	Active power reading of post-event window at time step t
$\mathbf{P}_{Pre}$	Vector of active power readings of pre-event window
$P_{Pre}(t)$	Active power reading of pre-event window at time step t
$P_{Pre,i}(t)$	Logic value for pre-event window at time step t
$\boldsymbol{P}'_{Pre}$	Truncated pre-event window
$P_{R,k}$	Residual active power value for event k
P_{thr}	Active power threshold for event detection
$\mathbf{P}_{Total}$	Vector of active power readings of pre- and post-event window
$P_{Total}(t)$	Active power reading of pre- and post-event window at time step t
P_{δ}	Active power threshold for eliminating transients
$\mathbf{Q}$	Vector of reactive power readings
$Q(t)$	Reactive power reading at time step t
R	Resistance
$REDD$	Reference Energy Disaggregation Data Set
s	Maximum number of occurrences of an event
$\hat{s}_m(D)$	Estimated state m of device D
s_t	Test statistic for event detection at time step t
$S(\psi)$	Score function for detector Ψ
$\mathbf{S}_j$	Set of power values for state j
S_t	Latent state of the Markov chain at time step t
t	Time step
$t(E_i)$	Time point of event E_{i}
$t_{end}(B_m)$	End point of power block B_{m}
$t_{off,max}(D)$	Maximal off-duration time of device D
$t_{off,min}(D)$	Minimal off-duration time of device D
$t_{on}(B_m)$	On-duration of power block B_{m}
$t_{on,max}(D)$	Maximal on-duration time of device D
$t_{on,min}(D)$	Minimal on-duration time of device D
$t_{start}(B_m)$	Starting point of power block B_{m}
T	Measurement interval
TN	True negative

TP	True positive
TPP	True positive percentage
TPR	True positive rate
u	Time duration
u_j	State duration of state j
U_{230}	Normalization voltage (230 V)
U_m	Measured effective voltage
v_{thr}	Minimum threshold for labelling events
$vote_{index}$	Index of voting window
w	Width of window
$w(t)$	Disturbance at time step t
w_1	Width of pre-event window
w_2	Width of mask window
w_3	Width of post-event window
WAN	Wide area network
w_c	Width of mask window
w_l	Width of vector $\mathbf{P}_{\text{Pre}}$ as well as $\mathbf{P}_{\text{Post}}$
w_{Post}	Width of post-event window
w_{Pre}	Width of pre-event window
w_R	Window to account power changes of start event
w_{vote}	Width of voting window
X_t	Observable output signal of a Markov chain at time step t
Y	Number of bits
z	Years
$z_{\alpha/2}$	Upper $100\alpha/2$ percentage point of the standard normal distribution
$\hat{z}_m(D)$	Estimated index m for the set of events $\mathbf{E'}_{\text{all}}(D)$
α	Confidence level
$\alpha_j(k,D)$	Forward probability of event k for device D to be in state j
β	Expected inflation rate
$\beta_j(k,D)$	Backward probability of event k for device D to be in state j
δ	Quantity of power averages below limit P_δ
$\Delta P(E_i)$	Total active power change of event E_i
$\Delta P(t)$	Active power change at time step t
ΔP_{Ch}	Maximum threshold for steady power change calculation
$\Delta\mathbf{P}_{e,j}(D)$	Extracted differential power load curve for state j of device D
$\Delta P_{Est}(E_i)$	Estimated steady state power change

ΔP_{FP}	Total power changes of the false positive events $\mathcal{F}$
ΔP_{M}	Total power changes of the missing events $\mathcal{M}$
$\Delta P_{off}(D)$	Measured total active off-power change of a device D
$\Delta P_{on}(D)$	Measured total active on-power change of a device D
ΔP_{Rel}	Threshold for irrelevant power changes
$\Delta P_{True}(E_i)$	True steady state power change
$\Delta P'_{off}(D)$	Estimated total active off-power change of device D
$\Delta P'_{on}(D)$	Estimated total active on-power change of device D
$\Delta Q_{off}(D)$	Measured total reactive off-power change of a device D
$\Delta Q_{on}(D)$	Measured total reactive on-power change of a device D
$\Delta Q'_{off}(D)$	Estimated total reactive off-power change of device D
$\Delta Q'_{on}(D)$	Estimated total reactive on-power change of device D
$\gamma_j(k,D)$	Total probability of event k for device D to be in state j
ε	Radius for DBSCAN
$\zeta_j(D)$	Computed parameter set from the ground truth data for the state distribution of state j
$\zeta'_j(D)$	Estimated parameter set for the state distribution of state j
θ_μ	Point of the perfect detector in metric μ
κ	Factor for computing the normalized active power
λ	Steady state power change algorithm
$\hat{\lambda}$	Best steady state power change algorithm
Λ	Parameter set for steady state power change algorithm
$\mu(\mathbf{P}_{\text{Post}})$	Mean value of vector $\mathbf{P}_{\text{Post}}$
$\mu(\mathbf{P}_{\text{Pre}})$	Mean value of vector $\mathbf{P}_{\text{Pre}}$
μ_w	Mean value of data model
π_j	Initial probability of state j
$\sigma(\mathbf{P}_{\text{Post}})$	Standard deviation of vector $\mathbf{P}_{\text{Post}}$
$\sigma(\mathbf{P}_{\text{Pre}})$	Standard deviation of vector $\mathbf{P}_{\text{Pre}}$
$\sigma'_{off}(D)$	Estimated variation of total off-power change of device D
$\sigma'_{on}(D)$	Estimated variation of total on-power change of device D
σ_w	Standard deviation of data model
τ	Time for merging events caused by transients
$\upsilon_{off}(D)$	Off-duration distribution of device D
$\upsilon_{on}(D)$	On-duration distribution of device D
Ψ	Parameter set of a specific detector
ψ	Specific detector
$\hat{\psi}$	Best specific detector

1 Introduction

1.1 Motivation

Due to the increasing number of electrical devices in households, it is currently not possible for customers to recognize which electric devices have a significant share of the total electricity consumption. At the end of the billing period customers only receive information about whether they pay more or less than in the previous period. The impacts of possible changes in the user's behaviour and of optimization measures in terms of efficiency actions are not traceable as such. As a consequence customers are not able to identify the factors influencing the total electrical energy demand of their electrical devices.

One of the approaches to raise the awareness of consumers is taken by the European Union in form of the introduction of intelligent metering systems. The European Union sets itself the target to reduce the primary energy consumption by 20 % by 2020 compared to projections for 2020. To achieve these reductions the EU's energy efficiency policy relies on five pillars [1]. One of these pillars is the general policy framework to which the Directive 2006/32/EC belongs. Besides other measures member states shall ensure that final customers are provided with "*...individual meters that accurately reflect the final customer's actual energy consumption and that provide information on actual time of use.*" Furthermore "*...billing on the basis of actual consumption shall be performed frequently enough to enable customers to regulate their own energy consumption*" [2].

Through the "third energy package" of the EU, member states shall ensure "*...the implementation of intelligent metering systems that shall assist the active participation of consumers in the electricity supply market*". Until 3rd September 2012 an assessment "*...of all the long-term costs and benefits to the market and the individual consumer or which form of intelligent metering is economically reasonable and cost-effective and which time frame is feasible for their distribution...*" has to be accomplished by the member states. A positive assessment leads to a roll out of smart meters to at least 80 % of consumers [3].

Most of the actions taken by governments to reduce the energy consumption in households, are based on informing consumers about the total power consumption of a specific period. For this reasons customers are not able to identify and monitor the factors influencing their electrical energy demand or even separate the total power consumption by major electrical devices.

In this thesis a novel non-intrusive method for splitting up the total energy consumption to its origins, without the need to install additional meters, is introduced.

1.2 The Research Problem

The research problem addressed in this thesis is the disaggregation of the electricity consumption of a specific domestic consumer to its individual appliances. It can be distinguished between intrusive and non-intrusive appliance load monitoring for accomplishing this task, whereby the latter was introduced by Hart [28] (see Section 2.2). Commercial intrusive solutions are very cost-intensive since they require the installation of additional meters to monitor the energy of all relevant electrical appliances. Non-intrusive appliance load monitoring methods usually work with a specific set of electrical appliances with good recognition accuracy when the algorithm is correctly parameterised. However, if an appliance is replaced or a similar power consuming electrical appliance is used, the algorithm has to be manually re-parameterised by an expert which is time- and cost-consuming. There are several unsupervised non-intrusive appliance load monitoring methods available but their performance is moderate. Especially when unknown electrical devices are used within the home, the algorithms become very complex and the performance decreases. However, in this thesis a new unsupervised method for energy disaggregation is introduced in which electrical appliances can be iteratively disaggregated.

1.3 Objectives

The aim of this thesis is to develop a new method for an unsupervised energy disaggregation that provides domestic customers with the possibility to monitor the electrical energy demand of major electrical appliances in their households. Beyond this, the main focus of the method is to obtain low costs and easy installation with no need for parameterisation.

1.4 Scope of Research

The thesis develops a methodology for energy disaggregation which can be easily integrated in domestic consumers' households without the need for manual parameterisation.
While the basic principles of the method could also be applied to other areas such as for example the agricultural sector as opposed to the domestic sector, a detailed analysis of the characteristics of all consumers is necessary and has to be carried out to successfully fulfil this task.
Not in the scope of this thesis are ways of providing electricity savings because of customers' feedback and data protection.

1.5 Research Methodology

Statistical methods are used for studying the characteristics of electrical appliances. Power measurements in 40 households have been carried out by the project ADRES [66] and are utilised for conducting the appliance-specific characteristics . For carrying out the computations a commercial software, Matlab®, is used. The performance analysis is carried out with another measurement data set from 5 Austrian households which was gathered by the Institute of Electrical Power Systems of Graz University of Technology [72].

1.6 Scientific Contribution

The main scientific contribution in this thesis is a new iterative method (fEEDBACk) for unsupervised energy disaggregation for domestic consumers' households. Prerequisites of the novel fEEDBACk algorithm are the knowledge about the load characteristics of different types of appliances as well as the detection of events. Therefore it is investigated in both mentioned issues in detail in this thesis.

For being able to create appliance specific models, this thesis presents a study of the characteristics of typical electrical appliances which can be found in households. Besides the specific power consumption of these appliances also estimations of typical on- as well as off-durations and time of use are conducted. These additional features increase the disaggregation accuracy.
Beyond this, the thesis analyze event detection algorithms for feature extraction which are a necessity for the fEEDBACk algorithm. The accuracy levels of different event detection algorithms are compared by different metrics. It is also investigated in factors which influence the performance of the event detection algorithms, respectively the sampling period (1, 2 and 5 seconds) and the incorporation of reactive power values. The optimal parameter sets of the investigated event detectors are analyzed in detail. Above all, a comparison of the accuracy of different steady state power change computations by different algorithms is presented. For that purpose a new method for a more accurate power change calculation is proven to be viable.

The outcomes of the described analyses are in a row used for the proposed energy disaggregation algorithm. The fEEDBACk algorithm estimates the parameters for modelling the electrical appliances in a first step and iteratively disaggregates the load curve of a specific device from the aggregated load in a second step. The principal method is based on hidden semi-Markov models and the disaggregation process is based on a modified Viterbi algorithm.

Parts of these contributions are also detailed in the following three already published papers by the author of this thesis:

- "Automated Electrical Energy Analysis for Domestic Consumers Based on Smart Meters" in International Conference on Electricity Distribution (CIRED), Frankfurt, 2011
- "Appliance-Specific Energy Consumption Feedback for Domestic Consumers" in International Conference on Electricity Distribution (CIRED), Stockholm, 2013.
- "Appliance-specific Usage patterns for Load Disaggregation Methods" in Internationale Energiewirtschaftstagung (IEWT), Vienna, 2013.

1.7 Outline of the Thesis

Chapter 1 gives an introduction to the thesis. The motivation, the research problem, the objectives and the scope of the research are presented.
Chapter 2 presents the state of the art of domestic energy monitoring systems as well as previous work. Additionally the specifications of the proposed disaggregation system are presented.
Chapter 3 analyses the characteristics of electrical appliances such as power consumption or time of use.
Chapter 4 gives a detailed study of the accuracy of the event detection algorithms and their optimal parameter sets as well as the computation of the steady state power changes of events.
Chapter 5 presents the novel introduced disaggregation method and contains the main results of the thesis.
In Chapter 6 a performance analysis of the novel algorithm is carried out and the results of the algorithm are presented.
Chapter 7 presents the discussion and Chapter 8 the conclusion of the thesis as well as future work recommendations.

2 Residential Electricity Consumption Monitoring

In this chapter general information about residential electricity consumption is given. Furthermore the advantages of appliance-specific electricity consumption as well as different methods of realisation are figured out. Finally, the literature review of methods for appliance-specific energy monitoring (non-intrusive appliance load monitoring) is carried out.

2.1 Background

2.1.1 Residential Electricity Consumption Trends

"*Final residential electricity consumption accounted for 29.71% of total final electricity consumption in the year 2010* [in the EU-27]. *It was therefore the second most consuming sector after the industry sector with 36.47%, and just before the 29.41% of the services sector.*" [4]
In the year 2010 the residential electricity consumption reached a peak with 840,788 GWh in the EU-27. Table 2.1 shows in detail the residential electricity consumption of the EU-27 member states from 2000 to 2011. Austria's electricity consumption, for example, also reached a new all-time high of 18.052 GWh in the year 2010.
Between 1990 and 2011, the residential electricity consumption increased by approximately 33% in the EU-27 with an average annual growth rate of about 1.38%. Between 1991 and 2011 there were just three years with an annual decrease in consumption compared to the previous year. In 1997 and 2007 the annual decrease was approximately 1% and in 2011 the residential electricity consumption decreased at a maximum level of 4.45%. This can be predominantly related to higher temperatures during the mentioned years which, as a consequence, also reduced the energy use for heating.

"*Although many appliances are becoming more efficient, the number of appliances is rising, appliances are used more often and for longer periods of time, and many appliances have more functions or special features that require more energy.*
The general trend in the residential sector is therefore an increase in electricity consumption. There are, however, important differences between different household electricity end-users. The electricity consumption of residential lighting is, for instance, decreasing. This decrease is to a large extent the result of the phasing-out of less energy efficient incandescent light bulbs. Also the large promotion of

compact fluorescent lamps in many EU Member States even before the phase out period contributed to this success." [4]

Table 2.1: Electricity consumption of households in GWh of the EU-27 member states (source Eurostat)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
EU-27	714,320	738,931	746,837	781,148	795,814	805,494	815,240	807,996	811,951	817,580	840,788	803,337
BE	23,738	24,396	25,921	26,026	26,543	26,007	22,722	21,856	19,982	20,210	20,276	19,257
BG	9,858	9,751	9,306	9,311	8,770	9,046	9,305	9,376	10,027	10,302	10,559	10,912
CZ	13,822	14,239	14,121	14,508	14,525	14,719	15,198	14,646	14,703	14,687	15,028	14,200
DK	10,215	10,159	10,190	10,262	10,332	10,449	10,573	10,349	10,280	10,096	10,389	10,111
DE	130,500	134,000	136,500	139,100	140,400	141,300	141,500	140,100	139,500	139,200	141,700	136,600
EE	1,466	1,585	1,584	1,594	1,618	1,620	1,675	1,773	1,845	1,884	2,023	1,934
IE	6,375	6,728	6,579	6,966	7,346	7,512	8,083	8,063	8,526	8,123	8,546	8,283
EL	14,207	14,546	15,775	16,444	16,852	16,875	17,676	17,957	18,126	18,131	18,130	17,628
ES	43,619	49,685	50,636	54,235	58,046	62,584	67,882	68,214	69,438	71,411	75,679	74,177
FR	128,720	133,887	132,998	141,554	147,088	144,548	147,104	145,755	150,899	151,733	162,470	148,610
HR	5,729	5,560	5,954	5,694	6,072	6,333	6,520	6,392	6,711	6,462	6,651	6,523
IT	61,112	61,553	62,957	65,016	66,592	66,960	67,635	67,220	68,389	68,924	69,550	70,140
CY	1,055	1,042	1,157	1,295	1,316	1,433	1,500	1,608	1,683	1,722	1,738	1,723
LV	1,189	1,239	1,317	1,421	1,467	1,572	1,728	1,794	2,031	2,000	1,938	1,772
LT	1,767	1,818	1,811	1,918	2,090	2,162	2,374	2,489	2,730	2,725	2,590	2,618
LU	792	801	808	822	839	845	831	844	776	904	854	844
HU	9,792	10,130	10,440	11,063	11,032	11,115	11,451	11,250	11,460	11,235	11,202	11,312
MT	559	540	570	629	615	623	659	658	645	570	475	598
NL	21,808	22,111	22,815	23,329	23,531	24,232	24,833	24,294	24,798	24,156	24,703	23,687
AT	**14,962**	**16,209**	**16,730**	**17,275**	**17,119**	**17,489**	**17,471**	**17,301**	**17,543**	**17,723**	**18,052**	**17,814**
PL	21,034	21,376	21,659	24,852	25,476	25,253	26,467	26,369	27,115	27,534	28,615	28,258
PT	10,056	10,625	11,382	11,835	12,432	13,242	13,406	13,863	13,444	14,190	14,522	13,752
RO	7,652	7,724	7,771	8,243	8,043	9,234	9,999	10,389	10,400	11,021	11,329	11,577
SI	2,601	2,675	2,704	3,008	3,012	2,951	3,055	3,021	3,182	3,137	3,219	3,211
SK	5,419	5,222	5,157	5,039	4,817	4,701	4,577	4,602	4,531	4,428	4,370	4,503
FI	18,140	19,373	19,942	20,404	20,338	20,648	21,342	21,491	21,169	22,047	23,589	21,799
SE	42,020	42,180	41,473	41,998	41,375	42,663	41,490	39,638	38,929	40,946	40,422	36,432
UK	111,842	115,337	114,534	123,001	124,200	125,711	124,704	123,076	119,800	118,541	118,820	111,585

The final electricity consumption per dwelling in the EU-27 was 4,146 kWh in 2006 and 4,137 kWh in 2009 [4]. In Austria the average final electricity consumption per dwelling in 2006/2007 was 4,390 kWh and decreased to 3,955 kWh in the period 2010/2011 [6].

2.1.2 Breakdown of Residential Electricity Consumption

As already mentioned before, the number of electrical appliances per household is continuously increasing. Nevertheless electrical appliances which have to cool down or heat up take a major part in the total electricity consumption. A detailed breakdown of the final electricity consumption of the EU-27 can be seen in Figure 2.1. It is worth to mention that on average approximately 10 electrical devices are causing about 60% of the total electricity consumption. In other words, only a couple of electrical appliances are responsible for the bigger part of the total electricity consumption. A similar picture is drawn from a study carried out by Statistics Austria which comes to the conclusion that about 65% of the total electrical energy consumption of an average Austrian household is caused by approximately 10 different major electrical devices as well as standby consumption [6]. Depending on the individual household, the existing equipment of electrical appliances as well as the usage this share can vary.

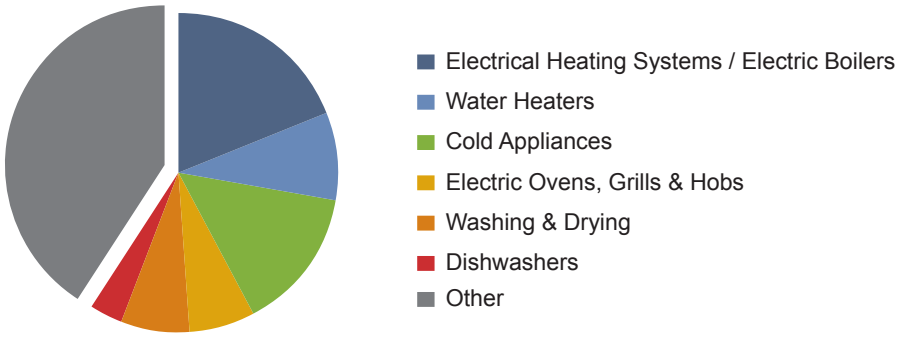

Figure 2.1: Breakdown of final electricity consumption in the residential sector in the EU-27 [4]

The breakdown of the residential electricity consumption of the EU-27 in Figure 2.1 is in line with the average Austrian residential electricity consumption in Figure 2.2. The most significant differences are that the share of electricity consumption of electrical water heaters and dishwashers is greater in Austria than in comparison to EU-27.

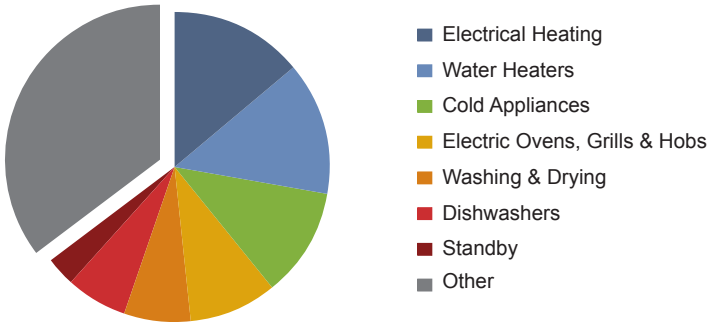

Figure 2.2: Breakdown of final electricity consumption of an average Austrian household [6]

But the increasing amount of small appliances also has an effect on the total electricity consumption which can not be neglected. Figure 2.3 shows the change in electricity consumption of 1990 com-

pared to 2009 for large appliances (televisions, clothes dryers, dishwashers, refrigerators, washing machines) and other small appliances.

"*The breakdown of appliances consumption shows that the strongest growth is recorded for small appliances* (*6.5% / year on average*). *These small appliances more than doubled their share of the total consumption for appliances and lighting, from 18% in 1990 to 39% in 2009. The consumption of large appliances recorded a moderate growth and their share declined from 62% to 44 %. Lighting has a rather stable share* (*about 20%*)." [5]

Since more and more products are introduced into the market it can be expected that the share of small electrical appliances will further increase in the next years.

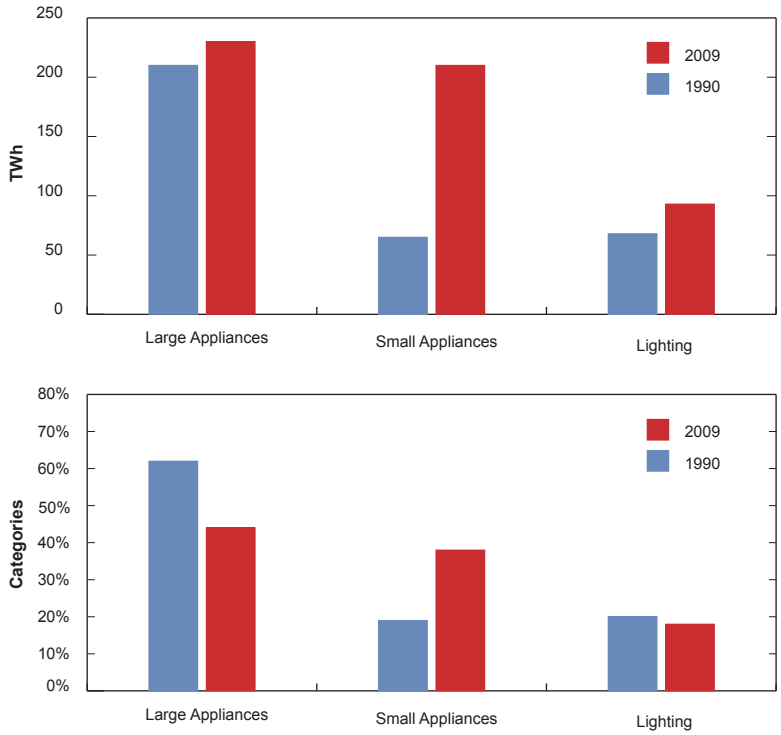

Figure 2.3: Change in total electricity consumption for large and small appliances as well as lighting for EU-27, source: [5]

2.1.3 Variation of Total Electricity Consumption

As already mentioned, the total residential electricity consumption rises due to additional electrical appliances and replacement with bigger devices like large widescreen plasma televisions. Furthermore the total energy consumption varies day by day depending on the users' habits and on the specific individual usage of each electrical device. Influences of a single electrical device on the total energy consumption are not traceable for most consumers nowadays. Even when measures to reduce the total electrical energy consumption are taken, it is not sure that these measures are reflected in the total energy consumption. The reason for that is the dependence on many different

parameters of the total electricity consumption and as a result, such measures are only traceable with special expertise. Besides, the total electricity consumption has seasonal components as well as a usual increase compared to the previous year. As a consequence, the total electricity consumption varies over the year (blue line) and rises yearly (dashed black line: +ΔW) or it can even decrease yearly (-ΔW), as can be seen in Figure 2.4. In addition, the total electricity consumption varies day by day and this makes a simple comparison of these values almost impossible.

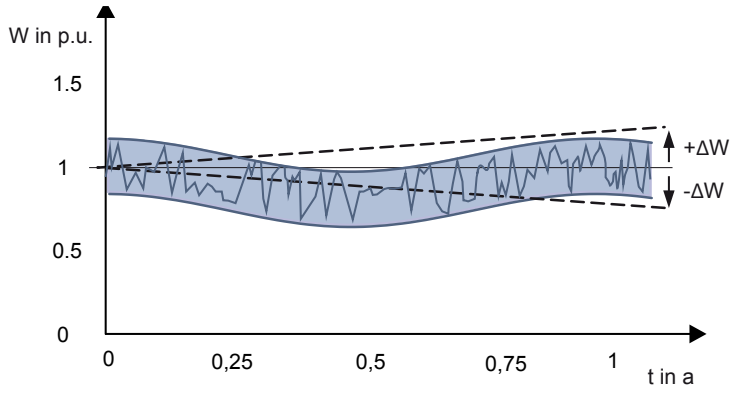

Figure 2.4: Variation of the average total energy consumption during a year

Providing consumers with information about their total electricity consumption during a specific period, like the European Commission is planning to do, is in any case an improvement to the status quo. However, information about the total power consumption does not give consumers the opportunity to differentiate the consumption of single electrical devices. Most of the currently sold smart meters provide the possibility to log load curve. Despite this, the manual observation of such load curves is a time consuming analysis, assuming special expertise about the behaviour of different electrical appliances as well as a diary of the usage of single electrical appliances. Without a handwritten log about the usage of major appliances it is impossible to determine which devices have contributed to the monitored total power consumption. Besides this, a high number of different electrical devices switched on at the same time make such an analysis even more complicated. The regular interval of 15 minutes, which is usually the recorded sampling period, does not allow distinguishing the consumption of most electrical devices. The reason for this is that due to the overlaps in the aggregated load curve a disaggregation to the single appliances is almost impossible (see Section 6.5).

2.1.4 Electricity Consumption Feedback Systems

This section gives an overview of different feedback systems. Furthermore the achievable savings from different studies which provide different feedback methods to residential consumers are listed and compared with each other.

2.1.4.1 Types of Feedback

In general it can be differentiated between *direct feedback* and *indirect feedback* systems to give energy users information about their energy consumption [8]. While direct feedback is provided in real-time, indirect feedback reaches the energy user after consumption. In the latter case, the exact delay can be in the range of several hours up to several weeks or months after consumption. Types for indirect feedback include "*...enhanced utility billing with specific household information and advice; estimated feedback that uses statistical techniques to estimate* (*and potentially disaggregate*) *total household energy usage based on a customer's household type, appliance information, and billing data, and daily/weekly feedback that uses real energy use measures gathered by a utility or third party and presented to the customer via the web, email, or mailed reports.*" [8] Direct feedback can be realised by in-home displays, ambient lighting or also sonification and provide energy consumers with the opportunity for "learning by doing".

The different types of feedback can be expressed in several ways. With [10] a categorisation of direct feedback, indirect feedback, inadvertent feedback, utility-controlled feedback and energy audits was introduced. Parts of this categorization scheme are used by [11] and further extended in six main categories which can be seen in Figure 2.5. Armel et al. [9] added another category named "appliance feedback augmented" where the most efficient energy saving tips for the individual household are automatically provided.

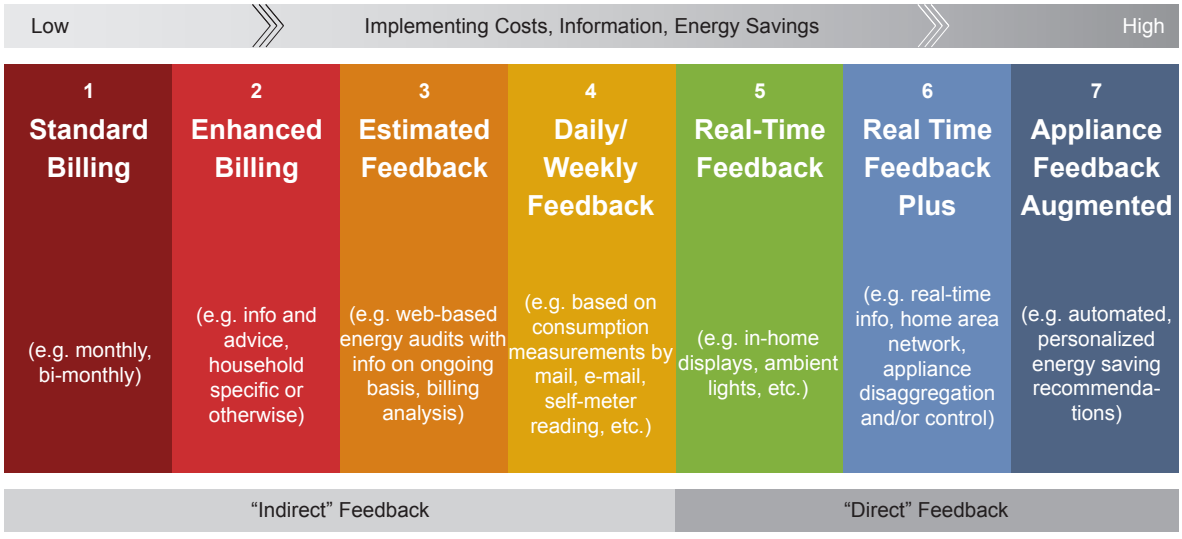

Figure 2.5: Types of feedback, extended from [8], [9]

While standard billing represents the traditional source of feedback that energy consumers get, enhanced billing also provides more detailed information such as comparative statistics, useful to highlight trends in the energy usage. Through estimated feedback the energy bill is broken down

to specific major electrical devices. This is usually done by web-based energy audits which use parameters such as households' members for estimating the specific electricity consumptions. Advanced metering infrastructure allows gathering daily or weekly electricity consumption patterns. However, due to changes in consumption such as purchases of new devices, or a household member who moved in or out, the comparison is not representative anymore (see also above). In-home displays which can be placed anywhere in the home are commonly used for real-time feedback. Besides the actual active power consumption the displays usually also indicate the previous load curve. Another possibility for real time feedback, for example, is ambient lighting [12] or sound.
In contrary to estimated feedback which also provides estimations of appliance-specific energy consumption, in the second to last category the appliance energy consumption is measured and monitored and the feedback about total consumption is provided in real time. This can be realised by additional meters or also by non-intrusive appliance load monitoring algorithms. Also home area networks can be utilized for systems which monitor or control electrical devices. [8]
The last category "appliance feedback augmented" evaluates all measured data from category number six and provides automated energy saving tips for households. In addition to providing notifications for taking actions such as the replacement of a malfunctioned device, appropriate programs and rebates or contract options can also be shown. [9]

2.1.4.2 Achievable Savings

Numerous research studies have already been carried out based on the effect of feedback in the residential sector. The first set of studies was triggered by the oil crisis in the mid-1970s and the second by global warming in the 1990s.
In a meta-review [8] 56 primary feedback studies have been categorised to the above described feedback types. The major part of the studies (34) was carried out in the US and 12 in Europe (the Netherlands, Finland, Denmark, and the United Kingdom). The time span of the studies ranges from 1974 to 2009. However, due to the fact that the energy consuming habits and the number of electrical appliances have changed, only the most recent 36 studies ranging from 1995 are evaluated in Table 2.2. The increase of the average savings from category to category goes in line with the proposed feedback categories above. However, within the single feedback types, the range of achievable savings varies widely, for example, the overall range is from 0.5% - 32%.

A field experiment carried out in Austria in 2010 [18] with a one year duration and the participation of more than 1,500 households suggest that "*feedback provided to the pilot group corresponds with electricity savings of around 4.5% for the average household.*" The provided feedback to the participating consumers was realised by daily, weekly and monthly reports about their total energy consumption and so it best fits within the category "enhanced billing" of [8] and goes in line with

similar savings. In Denmark 1,452 households took part in a survey [19] where randomly selected participants received text messages (SMS) and e-mails about their electricity consumption in 2006 and 2007. An average reduction of about 3% was achieved and this also lies in the range of the last mentioned category of [8].

Table 2.2: Average achievable household energy savings by feedback type, based on 36 studies between 1995 - 2009, source [8]

Type of Feedback	Number of Studies	Average achievable Savings
2 - Enhanced Billing	6	3.8%
3 - Estimated Feedback	3	6.8%
4 - Daily/Weekly Feedback	5	8.4%
5 - Real Time Feedback	18	9.2%
6 - Real Time Feedback Plus	4	12.0%

Another review from Faruqui et al. [16] of 12 research studies from North America and one from Australia and Japan "*...find*[*s*] *that consumers who actively use an IHD* [in-home display] *can reduce their consumption of electricity on average by about 7 percent when prepayment of electricity is not involved.*" Last mentioned savings for real time feedback are 2% lower compared to [8]. [17] presents a similar field experiment which took place in the US in 2010 with a sample size of 1743 households and a duration of 6 months. It found even lower savings of "*...a statistically significant reduction in electricity use of 5.7 percent.*" Due to the large sample size and the experimental design [17] argues that the achieved savings are more realistic than compared with [8].

A very recent study from Northern Ireland [20] exploits a large-scale natural experiment with several thousand of households and data from 1990 to 2009. Prepay consumers are equipped with a keypad where they can see and manage their electricity consumption. In comparison to the control group these customers achieve average savings of about 10.6% in the period from 2000-2005 and even a 17.86% reduction in the period from 1990-2009.

As can be seen, the achieved energy savings in the single studies vary. Several factors such as goal setting, competitions, commitment and social norms strongly influence the results of a study [8]. Beyond these factors, criteria such as the sample size, the study duration, the selection procedure, the evaluation criteria, the frequency of feedback, the rewards as well as the persistence of energy savings also play a big role in the variation of the results [13] - [15].
The influencing factors of electricity consumption are, for example, investigated in [21]. For 2000 Swedish households the electricity consumption patterns are analyzed over a 4 year period. Also

socioeconomic factors have been integrated. [21] compares electricity consumptions of households with different location, characteristics and feedback types. The paper comes to the following conclusion:

"*The analyzes of the consumption patterns showed that the major variations can be found between individual households rather than households groups. Therefore, we consider that individual and specific feedback* (*personalized for each household according to different preferences, characteristics and needs*) *should be provided to the households instead of generalized tips and information applicable to all households.*

Another important feature that any electricity consumption feedback should include is the specific consumption for most of the electrical appliances in the household. That would help the electricity consumers to change their behaviour and/or some of their most consuming appliances. Knowing how much electricity home appliances consume, would increase the overall low level of energy related knowledge existing in most homes and that many similar studies conclude is one of the major impediments for achieving larger domestic energy savings."

2.2 Appliance-Specific Electricity Consumption Monitoring

Since appliance-specific electricity consumption feedback has the biggest impact in energy savings (see above), a more detailed view on this topic is given in this section.

2.2.1 Benefits

Appliance-specific electricity consumption data which is provided daily, weekly or monthly as well as the individual load curve can be a good opportunity for consumers to see what is going on. For energy-interested persons and persons with some corresponding background knowledge these data is useful to get a better understanding of their consumption habits. But also persons with no background related to electricity consumption can highlight their most consuming electrical appliances and take some actions to achieve some energy savings. Measures like the replacement of an electrical device, for example, are easily traceable for consumers. Anyhow, without special expertise it is not possible to achieve the greatest energy savings.

For achieving the most fruitful energy savings, personalised information is necessary [9], [25]. Through automated processing of the data, energy-hungry electrical devices can be identified easily and consumers can find out their level of the standby power consumption. Each household can get individual information on how to save energy and on the most efficient measures in their homes. Parts of the analyzed data can be provided to energy savings consultants for promoting special rebates or programs for relevant households. Automatically, detailed information about further measures and energy savings advice could be made available easily.

Furthermore the power consumption of selected electrical devices can automatically be rated to several parameters such as energy efficiency grade or household size. So, consumers can compare their devices with similar devices from other households and the best available technology with the lowest energy consumption on the market. Through an energy savings evaluation tool the amortization time for different electrical appliances as well as provided measures can be calculated. All the information can be displayed to the individual needs or which are preferred by consumers [26].
Especially, behavioural approaches such as media campaigns or incentive programs will benefit particularly when such an automated analysis is established [24].
Above all also policy makers benefit through appliance-specific consumption information. Programs can be prepared more specifically to the special needs of consumers. But also the effectiveness of such programs could be more easily evaluated. [9]

2.2.2 Data Acquisition and Data Processing

In this section different methods for realising appliance load monitoring are compared against each other.

2.2.2.1 Overview

In general, it can be distinguished between intrusive and non-intrusive appliance load monitoring (NIALM) [42], see Figure 2.6. In both cases a hardware is necessary which measures the current and voltage and consequently calculates the power consumption.

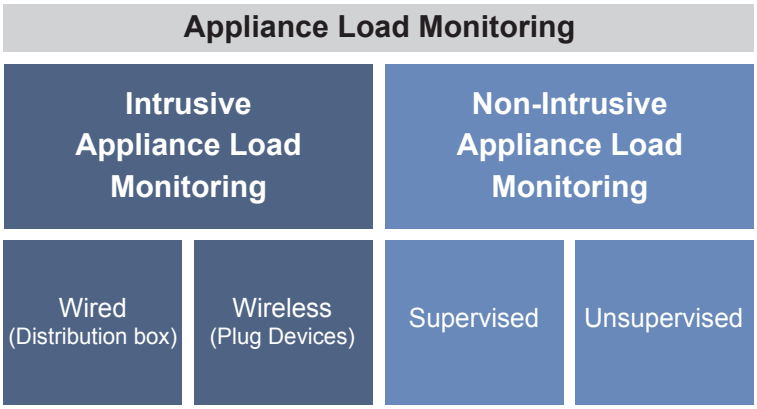

Figure 2.6: Overview of different load monitoring approaches, adapted from [42]

For intrusive appliance load monitoring a separate meter is necessary for each appliance. The meters are sometimes connected to a hub which collects all appliance data. There are easy to install plug-in power and energy monitors available on the market (e.g. Plugwise, Meter Plug, Enmetric, EnergyHub, ThinkEco...). These devices are plugged in an electrical socket and provide one free socket for the device to monitor. While some of them are all-in-one devices which are only able to

monitor one single electrical appliance, other vendors provide a full system with several plug devices and a wireless data collector. This collector can usually be accessed by a web browser or from a mobile device with a provided application (e.g. Plugwise, Meter Plug). Plug-in energy monitors have the advantage that they are very easy to install but cannot measure wired appliances such as water heaters or electrical ovens. Also plugs from appliances which are hard to access will be hardly monitored by energy consumers due to reasons of convenience. Beyond this, especially electrical devices where consumers do not even think of monitoring could go along with a possibly big energy saving potential which could pass unnoticed. So, big energy saving potentials could lie fallow [23]. Apart from this it is worth to mention that only plug-in monitors provide the possibility to control electrical appliances for the reason of home automation.

For monitoring wired devices as well as the minimum base load in home, sensors which are installed in the distribution box are required. (EIB, DigitalSTROM, TED). For installing such meters usually an electrician is necessary which results in additional installation costs. However, this is the only way for measuring wired devices or devices which can not be easily unplugged. It should be noted that just a few electrical circuits only have a single device attached. When there is more than one connected device on a circuit, special expertise is necessary to break up the electricity consumption of the corresponding devices. Usually electrical heaters, washing machines and fridges are connected separately to a single circuit breaker. However, it has been shown that this is not true for each electrical installation and especially not for homes with installations which were installed several decades ago. So, usually a combination of permanently installed wired meters in the distribution box and plug-in devices are necessary to monitor all electrical devices with the biggest share of the total electricity consumption. As already described above, it is not necessary to monitor all devices, since the most effective lever for achieving energy savings can often be achieved by just a couple of devices.

The costs for plug-in energy monitors as well as such systems which are installed in the distribution box lie between 250 EUR - 600 EUR for monitoring at least 8 different electrical appliances. If an in-home display is used additionally, the costs rise by about 40-80 EUR. Per monitored electrical appliance the price range is from approximately 30 EUR to 70 EUR. Lower prices are usually achieved with plug-in monitors.

Above all, integrated low-cost solutions are currently developed for fulfilling energy monitoring and management [22]. Electrical appliances with such functionality are called smart appliances and communicate in a home area network (HAN). However, the ASICs have to be integrated by many vendors to be able to use it. Currently there are no such low cost solutions for this kind of product

nor is there a standard or a policy. Beyond this, a market dominance would take more than a decade because of the standard life times of particular electrical appliances. Furthermore it is not sure if not just only white goods, but also small electrical appliances will be equipped with such an ASIC. The currently available systems for refrigerators, for example, come with additional costs of about 100 EUR. As a consequence, this is still an uncertain option for energy monitoring purposes.

Alternatively, non-intrusive appliance load monitoring (NIALM) [28] can be used to break up the total electricity consumption into its single appliances. To realize this, an electrical meter in the distribution box is necessary which measures the total power consumption on all phases. Additionally, a hardware component is necessary where the algorithm is executed and the information for the consumer is prepared. There exist supervised and unsupervised algorithms for NIALM. The former also requires a parameterisation of the algorithm which can only be done by an expert. In the unsupervised case, the algorithm is capable of machine learning and finds the correct parameter set by itself.

Due to the fact that smart meters are introduced in the European Union in the next couple of years [3], a power measurement device will be available in almost each household by 2020. By using the smart meter data, a cost-effective solution can be available which is already described in [25]. There is no extra cost for the consumer for a power measurement device. The minimum requirement for using smart meters is hardware that is able to read out the power measurement data of the smart meter. These data can be processed in-house by this hardware or transferred to a server farm where the energy analysis is carried out. The installation of the additional hardware can be performed easily. Unfortunately, the provided measurement data of smart meters is limited, which in turn reduces the detection performance of NIALM algorithms (see next section).

An energy disaggregation directly in the household could be performed by a mini computer such as a plug PC which can gather the data directly from the smart meter by one of the home area network interfaces (see below). Such small PCs (e.g.”DreamPlug”) are equipped with more powerful processors, are very cheap (120 - 180 EUR) and just have an auxiliary power consumption of several Watts. These devices can be a good option for running NIALM algorithms. Due to the fact that data measurement as well as the analysis take place directly in the home, data protection issues can be minimised. Such devices can easily be accessed by WLAN from a PC or any mobile device.

Another option is that the data collector reads out all the power measurement data from the smart meter and sends this information to a cloud of a service provider. For realizing this, a simple hardware is necessary with a price that could be in the range of 15 - 50 EUR. The service provider could analyze the data in a server farm and provide energy consumers a web portal where all relevant electricity consumption information is available. They can also provide additional services such as an alarm via text-message or mail when a certain device is consuming more energy than in the preceding periods.

Theoretically the analysis of the power measurement data can also take place directly in the smart meter but this would need hardware changes and can therefore only be integrated in later models. Since smart meters will be rolled out in the next couple of years in Europe and have a life time of a couple of years, this is not an option in the near future.

An overview of all mentioned options can be seen in Table 2.3. The lowest costs of the hardware can be achieved when the measurement data of the smart meter is used and analyzed by a separate computer, either in-house or in the cloud. Since NIALM algorithms require specific measurement data, the next section investigates their specific needs.

Table 2.3: Options for obtaining appliance-specific electricity consumption data for a home, monitoring at least 8 electrical appliances

System	Hardware Costs	Installation Costs	Disadvantages
Plug-in monitor	250 EUR - 400 EUR	Self-installation	• Wired devices can‘t be monitored • Selection of devices to be monitored • Minimum base load is not detectable
Sensors in distribution box	400 EUR - 600 EUR	100 EUR - 200 EUR	• Overlap of energy consumption of devices on the same circuit
Integrated ASIC (smart appliances)	n.a.	Integrated	• Slow adoption rate (> 10 a) • Usually only for white goods • Minimum base load is not detectable
In-house disaggregation of smart meter data	120 EUR - 180 EUR	Self-installation	• Reading out measurement data of smart meters is limited
Cloud-based disaggregation of smart meter data	20 EUR - 50 EUR	Self-installation	• Reading out measurement data of smart meters is limited
Smart meter based disaggregation	n.a.	Electrical utility	• Hardware changes necessary • Slow adoption rate (> 10 a)

2.2.2.2 Data Acquisition by Smart Meters

Smart meters measure voltages and currents with a sampling rate of typically 12 to 15 kHz[1] and the active or instantaneous power is processed internally with a high resolution of a couple of milliseconds for all phases. These data could in theory be used to apply the method of non-intrusive

[1] Lower priced smart meters do have sampling rates to a minimum of about 600 Hz. Since those smart meters do not have the functionality which is required in, for example, Austria those very cheap priced smart meters are out of scope of the investigation.

appliance load monitoring but currently this is not possible. Firmware and, depending on the specific device, also hardware changes would be necessary to make the information available on an interface [9].

For getting good results and minimum detection errors with NIALM algorithms, a load curve with a resolution of about one second up to about 10 seconds is necessary. The greater the resolution in time of active and reactive power the lower is the presence of overlaps in the load curves from different appliances. Recently published NIALM algorithms analyze harmonics as well as turn-on transient energy. These algorithms typically process reactive and active power in milliseconds and current harmonics with several kHz. Though, such algorithms can raise the detection rate and give the opportunity to detect further electrical appliances, also algorithms which use reactive and active power with a resolution of about one second to 10 seconds can achieve reasonable low detection errors.

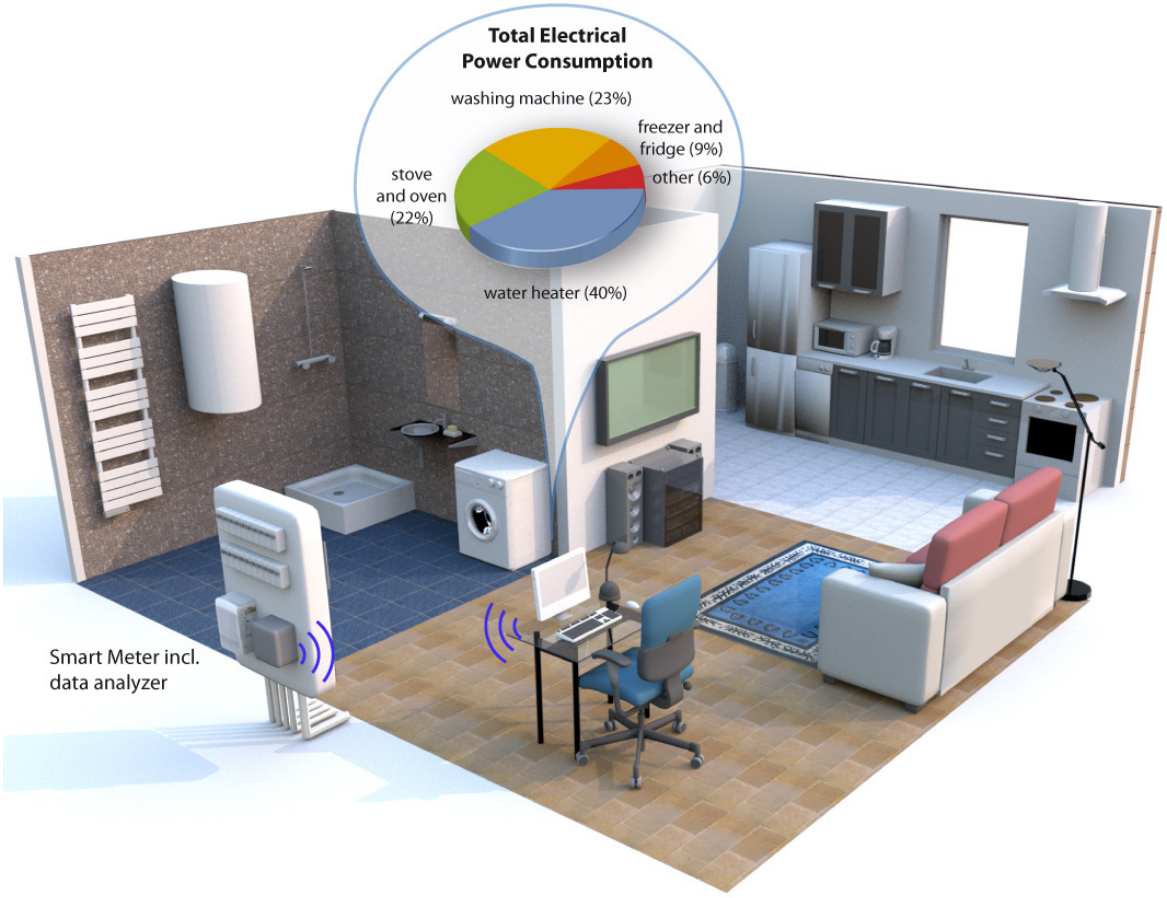

Figure 2.7: Residential electricity consumption monitoring system

Interfaces of Smart Meters

Currently sold smart meters are typically equipped with pulse indication, a multipurpose expansion port, a wire-connected PLC interface and an optical service interface. The multipurpose expansion port allows connections to home area networks (HAN) and through expansions of the multipurpose interface radio ZigBee connections and wireless M-Bus connections can be established. The optical service interface provides the possibility to read out standardized parameters (IEC 62056-21) such as voltage, electrical power consumption per phase or 15-minute averages of active power among other things. ZigBee connections and wireless M-Bus interfaces are predominantly used for in-home displays and connections to other digital meters. Up to now, in collaboration with software developers, a few manufacturers provide off-the-shelf solutions for consumers to display the current total power consumption on PCs every second. Table 2.4 shows an overview of available interfaces of currently sold smart meters in Europe.

Table 2.4: Overview of available interfaces of smart meters by different manufacturers

Interface	Minimum Sampling Interval P/Q per phase	Minimum Sampling Interval Voltage per phase
Service interface (IEC 62056-21)	2-6 seconds	2-6 seconds
Pulse indication	up to 60,000 pulses per kWh or kvar	n.a.
Multipurpose expansion port	P: 2-4 seconds (HAN)	n.a. (HAN)
Direct Line Communication	2-6 seconds, (HAN) 15-minute average (WAN)	n.a. (HAN), 15 minutes (WAN)

So far, it is not possible to read out active and reactive power as well as the voltage per phase every second or even with a greater sampling rate. Although the internal data is processed in milliseconds these data is not provided at one of the interfaces. However, with firmware changes the resolution available on at least one of the HAN interfaces could be increased to the range of about 1 Hz [9] and additionally modified in such a way that also the actual network voltage is transmitted, which would fit the needs of certain NIALM algorithms. These data can be read out and, for example, saved locally in a household.

Due to the limited speed of the direct line communication for the WAN and also due to data protection reasons it is very unlikely that one of the WAN ports of the smart meter is used for transmitting measured data which have a sampling rate of 1 Hz or above.

2.2.3 Costs of the System

To gain market acceptance, the overall costs of the system should be amortised within about 3 to 5 years. The average energy consumption of a household of the European Union is about 4,146 kWh

(see above). With an electricity price of about 18 Eurocent per kWh the total electricity costs are approximately 750 EUR per household.

For the calculation of a cost analysis two different scenarios are used. A moderate case of about 4% net savings and a case with 10% net savings are assumed. In order to make all future cash flows comparable the net present value is calculated. This is commonly used in the discounted cash flow analysis by integrating inflation and returns to get the present value of the money.

The net present value K_0 is the sum of annual savings K_j for all years z and when the interest is constant it can be calculated with [94]:

$$K_0 = \sum_{j=1}^{z} \frac{K_j}{(1+i-\beta)^j} \tag{2.1}$$

A real interest rate $(i-\beta)$ of 3% is used for the calculations where β represents the expected inflation rate. After a period of 5 years the moderate case (4% net savings) results in a capital value of about 140 EUR and the other case in about 340 EUR, see Figure 2.8. This shows that the total costs of an energy saving system have to be very low in order to amortize within a couple of years. Since a computer for in-house disaggregation could cost up to 180 EUR, there can only be a small margin added for a business case. In the case of the cloud computing a very low priced hardware in the range of about 40 EUR can transmit the measured data to a service provider. When a service provider has a large user group of several thousand customers, price ranges of about a few Euros could be achieved for providing such a service [9].

Also a rental service of monitoring devices would result in lower total costs for the consumers. However, a continuous monitoring of the appliance-specific energy consumption is preferable since it can quickly detect increasing consumption caused by the consumer and it can also track measures and give up-to-date advices.

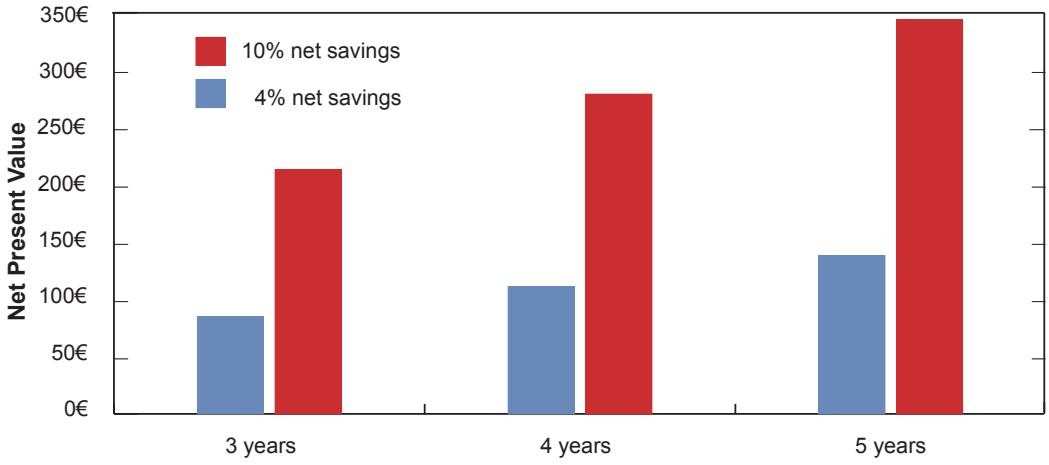

Figure 2.8: Net present value with net savings of 4% and 10% of the electricity costs dependent on different periods (real interest rate (i – β) of 3%) for an average household in the EU

2.3 Literature Review: Non-Intrusive Appliance Load Monitoring

Non-intrusive appliance load monitoring (NIALM) was introduced by Hart around 1990 [28]. The aim of NIALM, which is also called energy disaggregation, is to disaggregate the appliance-specific load curves from the total load curve. In other words, it means breaking up the total electricity consumption to its origins (specific appliances). In the next sections more detailed information about appliance models, different NIALM algorithms including features and inference techniques are given mainly as a summary from [29], [30].

2.3.1 General Description of NIALM

In order to understand the different needs of NIALM algorithm, first of all an overview of an appliance model is given.

2.3.1.1 Appliance Model

In a household there exist a variety of different electrical appliances with different electrical characteristics. For this reason different specific needs for energy disaggregation are given and they depend on the operating states of a device. [28] categorised electrical appliances as follows:

- Type-I: Appliances which consume a constant power in their on-state and no power when they are switched off. These devices only have two different states (ON/OFF), e.g. light bulb, water heater.
- Type-II: Appliances which operate on several different more or less constant power levels. These appliances are also called multi-state appliances and have a finite number of operating states. Such devices can also be expressed as finite-state machines, e.g. three speed fans, washing machines.
- Type-III: Appliances which consume a continuously variable power. Continuously Variable Devices such as dimmer lights fall in this category.

In [29] and [42] an additional category is defined:

- Type-IV: Appliances which consume low power through several weeks or days, e.g. cable modems, electronic clocks.

Type-I appliances are the most simple appliances to detect for NIALM algorithms. Since repeated switches of the power levels usually occur in the Type-II category, the detection effort is a bit greater and lies between simple to moderate. Type-III and Type-IV appliances are the most difficult to detect. Usually greater sampling rates which go along with a greater variety of features are necessary to detect such appliances.
More detailed information to characteristics of electrical appliances is given in Chapter 3.

2.3.1.2 Basics Steps of NIALM

There are a lot of papers addressed to NIALM and especially in the last couple of years there has been an increasing research interest in this field. Reason for this is predominantly the introduction of advanced meter infrastructure which is accompanied with the availability of power measurement data, typically with a sampling period between 15 minutes to one hour. A main focus of recent research work is the field of energy disaggregation for low sampling rates (around 1 Hz) which data can be provided by most smart meters (see above).
Generally speaking, NIALM has three major steps where the last two can be further subdivided (see Figure 2.9). First of all the data acquisition with power measurement devices collects the total load curve on all phases with the desired sampling rate. The selection of the sampling rate is dependent on the next step, appliance feature extraction. While for steady state features such as power changes sampling rates of approximately 1 Hz - 0.02 Hz are sufficient, for transient states analysis such as current harmonics sampling rates greater than 1 kHz are used [30]. In the last step inference techniques and learning algorithms (e.g. hidden Markov models, neural networks, etc.) are applied to the data. It can be distinguished between supervised algorithms which need a human interaction for labelling the extracted data as well for parameterisation and unsupervised algorithms. The latter algorithms achieve the disaggregation task without any interaction by a human being.

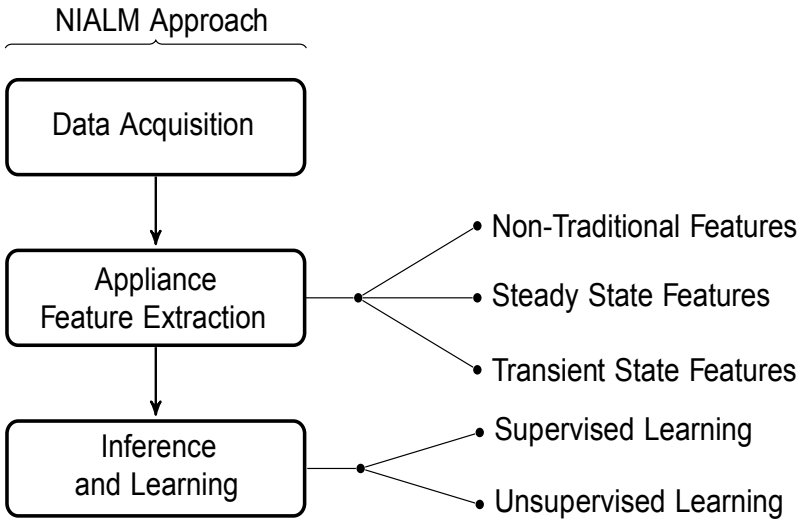

Figure 2.9: Block diagram of NIALM approaches, source: [30]

2.3.2 Data Acquisition

The aim of data acquisition is to gather the active and reactive power as well as the voltage with the desired sampling interval from the mains. The power data is also called load curve or aggregated data. With an A/D converter the electrical quantities are digitalized. These data is used for computing

all relevant electrical quantities such as active power and reactive power. As already described in Section 2.2.2 there exist several commercial solutions for realizing this.

2.3.3 Features Used in NIALM

The subsequent step after the electrical quantities have been computed is feature extraction. At the beginning of NIALM, changes of real and reactive power are used as features to detect state changes of electrical appliances [28]. These state changes are called events and belong to the steady state methods for feature extraction. Recent approaches for example also use current harmonics, transients as well as time of day or correlations between the usages of different devices as features. These features can be classified in three main sections: steady state analysis, transient-state analysis and non-traditional appliance features. A detailed overview can be found in [29] and [30]. Furthermore it is worth to mention that different sampling periods afford different data features which affect the number and types of detectable devices.

2.3.3.1 Steady-State Features

Steady-state analysis implies the assumption that features of an electrical appliance change just when the state of an electrical appliance also changes, otherwise they remain constant. The most commonly used features in steady-state analysis are changes of active and reactive power [28]. Researchers [38] - [41] have tried to disaggregate the load curves of electrical appliances only using the active power values as a single feature. Farinaccio et al. [41] concluded that the recognition of specific devices that have a recurring load profile and high loads such as water heaters or refrigerators works with an error below 15% when the electricity consumption is used for evaluation purposes. Another approach by [38] detected specific major electrical appliances such as stove, oven, geyser, microwave and water kettle by step changes in total active power with a maximum error of 9 % in electricity consumption. However, even when the load profiles of specific electrical appliances can be separated by using active power as a single feature, it also has some disadvantages. These include on the one side that electrical devices which have similar power levels as well as varying or weekly-constant power levels (Type-III and Type-IV) can't be discerned. On the other side, events that occur simultaneously overlap more easily and cannot be tracked back to the corresponding devices.

Beyond active power, a further commonly used feature is the reactive power. Due to the characteristics of electrical devices such as resistive, capacitive or inductive loads, each electrical appliance has its more or less unique power level. The active and reactive power levels vary from device to device and this is shown according to the PQ-plane in Figure 2.10. While electrical devices with major power consumption (e.g. heaters, washing machines) can be more easily differentiated, devices such as laptops, lights or TV sets have similar power consumption and can be hardly distinguished in the PQ-plane anymore.

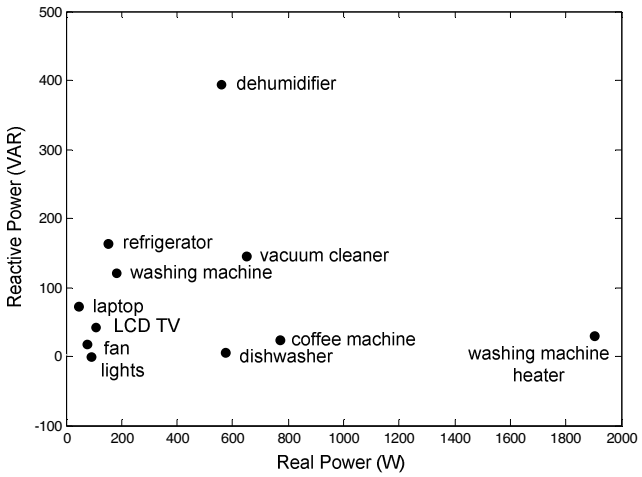

Figure 2.10: Example of a load distribution in PQ-plane from [30], Type-I and Type-II appliances

Beyond the above described power features, researchers also analyzed current and voltage waveforms and extracted features such as peak and Root Mean Square (RMS) values, phase differences (φ) and current harmonics and above all the Power Factor (λ) [31]-[37]. Through these features a better disaggregation accuracy is achieved. However, the extraction of harmonics goes along with the requirement of greater sampling rates of approximately 0.6 kHz and above [9]. The most important advantage of current harmonics is that electrical appliances can also be detected during their on-state because of the presence of their individual current amplitudes. For realizing this, additional training efforts for the algorithms as well as appliance signature databases are necessary.

2.3.3.2 Transient-State Features

Certain electrical devices such as refrigerators have turn-on transients of the active and reactive power. These transients can be used for NIALM since they offer a distinct feature with less overlapping compared to steady state features [30]. The transient power is successfully used for raising the detection accuracy of NIALM algorithms [29], [52] - [55]. In [55] it has been shown that active and reactive transient energy as well as the transient response time lead to higher detection accuracy than steady state features. Beyond these, transients of start-up currents are used by [34], [39], [55] for a better detection accuracy.

Patel et al. [56] uses a single plug-in sensor to monitor wire-bound electrical noise. Electrical devices, especially switch power supplies, generate high frequency electromagnetic interference (EMI). The measured electrical noise is Fourier transformed within a frequency range of 36-500 kHz. Through machine learning techniques a high accuracy in the assignment of events to their corresponding appliances could be achieved by the proposed system. However, this approach only works for ap-

pliances with a switch power supply. Further drawbacks are that the measured noise depends on the wiring architecture as well as sources from the surrounding environment. [30], [56]
Anyhow, a major drawback of all mentioned features in this section is the requirement of high sampling frequencies to gather those transients [30]. But also excessive training efforts for the algorithms are necessary to detect known appliances.

2.3.3.3 Non-Traditional Features

Additionally to the above mentioned features, also non-traditional features are used by NIALM algorithms. A new mapping of the aggregated load curve through rectangles and triangles is proposed by [60]. Rectangles and triangles are best fitted into the load curve neglecting minor fluctuations. Rectangles and triangles are described by several parameters (start time, peak time, peak value, end time and a steady value). These data sets are used for a classification of the events, which leads to less overlaps compared to the traditional power changes used in the steady state features as reported by [60].
Very promising results are achieved by the integration of time of the day, on and off duration distribution as well as the correlation to other electrical appliances by [73], [77].
In [62] inexpensive wireless light intensity sensors are successfully used for the extraction of events when artificial light sources are turned on or off.
[64] and [65] uses audio features from the device usage as a feature for energy disaggregation whereas [63] uses a combination of sound and magnetic sensors which are placed next to electrical appliances.
Generally speaking, the detection accuracy of NIALM algorithms can be raised through additional features. Tough, the installation effort and costs of extra sensors will have to be weighed against an increase in the detection rate.

2.3.4 Inference and Learning

After all features are extracted the next step is a classification of these features to the appliances. As already mentioned above, this can be categorised into supervised and unsupervised algorithms. In this section an overview is given of supervised algorithms that require human interaction by an expert, which is time and cost consuming, and unsupervised algorithms.

2.3.4.1 Supervised Algorithms

For supervised energy disaggregation optimisations and pattern matching methods are mainly used. In the former case a formula which describes, for example, the aggregated load by the power consumption of each appliance is used for the optimization problem. In the optimization process the parameters which result in a minimal error between the measured load curve and the estimated load

curve (which is described by the formula) ought to be found [33]. Several papers have been proposed addressing such an optimization problem (e.g. genetic algorithms [42], integer programming [33]) but it is still a challenge to realise high detection rates.

Pattern matching methods are most frequently used by NIALM algorithms. Features such as active (P) and reactive power values (Q) are extracted from the load curve and matched against already classified features which are available in a database [28]. A new feature vector is classified to the class where the match is the closest possible. So, similar patterns of electrical appliances are classified into the appropriate category. For the matching process distance metrics such as the Euclidean distance are used. But also here simultaneously occurring events and especially electrical devices with similar power consumption cannot be recognized through these approaches. Researches addressing this problem have extended their algorithms, for example, by filtering procedures [41], harmonics of the current [34] or power transients [78].
Alternatively, temporal information such as a specific patterns for using appliances are integrated by Bayesian networks [81] or state transition information by hidden Markov models [81] which are used for energy disaggregation.
Beyond this, some researchers employ classifiers for NIALM algorithms. These classifiers are trained by a database with known signatures of electrical appliances and in a row are used for recognition of the appliances. Various approaches such as neural networks [54], [79], naive Bayes classifiers [80] and support vector machines [56] are used for classification.
However, an approach will only work when the features of the electrical appliances do not overlap and can be differentiated.

2.3.4.2 Unsupervised Algorithms

The research activity for unsupervised NIALM algorithm which disaggregate energy without a-priori information about the number and types of appliances started to grow recently. As already described above through unsupervised algorithms no effort for parameterisation is necessary, which makes energy disaggregation algorithms more applicable. An overview of unsupervised algorithms is given in this section.
[83] uses a blind source separation technique for energy disaggregation. By clustering of the steady state active and reactive power changes and then employing a matching pursuit algorithm, events of the appliances can be assigned accordingly. The matching pursuit algorithm tries to match each event to one or more of the cluster centroids. By this approach, it is attempted to separate overlaps of simultaneously occurring events to its sources. The main issues with the overall algorithm in the case of the clustering procedure are that several appliances couldn't be separated due to similar power consumptions and that the reconstruction of multi-state devices with different events is very

challenging. In the case of the matching pursuit algorithm, missing events as well as the fact that previous results of the matching procedure are not taken into account, lead to most errors.

Another approach from Shao et al. [84] uses motif mining to extract frequent power change patterns and in a row the energy consumption of electrical devices. Motifs are referred to primitive shapes and frequent patterns and in the mentioned approach power difference values are used to distinguish between different devices. The fact that power changes of a particular device usually are just separated by a few power changes of other devices is used to assign power changes to its respective electrical appliances. For frequently used devices such as ovens, refrigerators or microwaves a performance of F-Measure (see below) of around 0.7 could be achieved. The data which was used for the evaluation is the home one of the public dataset REDD [67]. An extension of the motif mining approach was realised by [85] which incorporates statistical data calculation of on- and off-durations of electrical appliances. The addition of state durations allows to better distinguish between similar power events from different electrical appliances. An increase of the F-Measure value compared to [84] is reported through this method.

Also Kim et al. [73] integrated statistical features such as correlation between appliances, state duration probabilities and time of day into a conditional factorial hidden Markov model which has the active power consumption as the input. The power measurement data was gathered in 7 different homes and included between 4 and 10 electrical appliances. In contrary to regular factorial hidden Markov models (FHMM) an increase of the performance could be realised by the statistical features. However, the model performance decreases when the number of electrical appliances to disaggregate rises. It is not reported how unknown appliances affect the model. Above all, another drawback is local optima which can occur by the available inference algorithm for hidden Markov models.

A difference hidden Markov model (HMM) which is tuned by prior models of general appliances is utilized by [74]. In contrary to the mentioned factorial hidden Markov model, this approach allows the disaggregation of a single electrical appliance at once. Further electrical appliances can be disaggregated by an iterative process which subtracts the already disaggregated loads from the total load curve. For the disaggregation of an individual appliance a general model which describes the state transitions is used. The parameters such as power change consumption are extracted by an expectation maximization algorithm. For inference an adapted Viterbi algorithm is used which can filter out observations. Also here, high energy consuming appliances (refrigerator, microwave, clothes dryer and air conditioning) are used for the accuracy evaluation with a sampling period of one minute. A normalised mean error between 21% and 77% is reported for the mentioned appliances.

An improvement of the inference algorithms for hidden Markov models was realised by [75] through a proposed additive factorial approximate inference algorithm. Nine electrical appliances with a power consumption of around 1000 W are modelled with the approach. An average precision of about 87% is reported for those appliances. Kitchen outlets and electronics have a precision below 50%.

In [76] a hierarchical Dirichlet process hidden semi-Markov model (HDP-HSMM) which is a natural Bayesian non-parametric extension of the well-known hidden Markov model is used for unsupervised energy disaggregation. In comparison to hidden semi-Markov models, the HDP-HSMM has the advantage that the number of hidden states of an electrical appliance does not have to be set a-priori. So, the device parameters can be learned during inference. Five electrical devices (refrigerator, lighting, dishwasher, microwave, and furnace) are chosen for disaggregation for 4 different homes with a sampling period of about 20 seconds. An average performance of about 80% was achieved by this approach for the mentioned electrical devices.
An overview of all unsupervised algorithms for NIALM can be seen in Table 2.5. For all approaches the evaluation is carried out with the public Reference Energy Disaggregation Data Set (REDD) [67].

Table 2.5: Overview of unsupervised energy disaggregation algorithms

Approach	Public Dataset	Sampling Frequency	Performance
Blind source separation technique [83]	No	60 Hz	Works moderate and only usable for big appliances
Motif mining [84]	Yes	1 Hz	F-Measure around 0.7 for devices that are used frequently (both consuming low and high power)
Motif mining and incorporation of statistical data [85]	No	1 Hz	F-Measure around 0.96 for a refrigerator
Conditional factorial hidden Markov model [73]	No	1/3 Hz	F-Measure around 0.7 for eight appliances
HMM and prior models of general appliances types [74]	Yes	1/60 Hz	Average mean error between 21% and 77% for four electrical appliances
Additive Factorial HMMs [75]	Yes	1 Hz	Precision of around 87% for big appliances, and below 50% for electronics
HDP-HSMM [76]	Yes	1/20 Hz	Performance of about 80% for four selected appliances

2.3.5 Evaluation Methods

As can be seen in the section above, the performance of the different approaches is reported with different metrics and different sampling frequencies. Beyond this different electrical appliances are evaluated with the mentioned algorithms. This makes a meaningful comparison of the performance of the algorithms nearly impossible.

For the evaluation of NIALM algorithms different metrics have been proposed. The most common metric is the deviation from the true to the predicted electricity consumption of a device (e.g. [28], [41], [86]). However, it has been shown that this metric is not meaningful since the amount of energy can be the same even when the predicted load curve strongly differs from the original one [85]. Alternatively, Kolter et al. [67] proposed a "total energy properly classified" metric where each time step is compared to the ground truth data and devices with a low power consumption have a minor influence in the metric compared to high energy consuming devices.

Also, metrics from the pattern recognition and information retrieval domain ("precision" and "recall") are used for accuracy evaluation [73]. Precision is the fraction of correctly detected samples to all detected samples and recall is the fraction of correctly detected samples to all available (detected and not detected) correct samples. After the computation of both terms in a row F-Measure is calculated by the harmonic mean of precision and recall.

Anderson et al. [45] provides a set of different metrics especially for the purpose of the evaluation of event detection algorithms. The paper proposes to distinguish between event detection and classification. Similarly, Liang et al. [61] suggest splitting the evaluation in three different steps: Detection accuracy, disaggregation accuracy, and overall accuracy.

Above all, one of the public data sets for energy disaggregation (e.g. [67]) should be used to be able to compare the performance to other algorithms. When the NIALM algorithm is capable of an energy disaggregation with different sampling frequencies, the performance in dependence of the sampling frequency should be reported. Additionally, for providing better comparable results, the most frequent evaluation metrics should be used.

2.4 Data Sets for Energy Disaggregation

In this section an overview of public as well as private available data sets for energy disaggregation is given. Furthermore the used data sets within this thesis are presented.

For fostering research in the field of energy disaggregation, public data sets are made available from different sources ([44], [66] - [72]). On the one hand, these data sets allow researchers who have no access to measurement hardware to evaluate their energy disaggregation algorithms. On the other hand, these data sets can be used to compare the performance of different algorithms on the same measurement data. An overview of available data sets can be found in Table 2.6.

The most used data set in recent NIALM algorithms is the "Reference Energy Disaggregation Data Set (REDD)" [67] which was published in 2011. There are also a couple of further data sets available which have similar sampling periods in the range of about one second, see Table 2.6. The sampling rate of the measurement devices which are used for recording the data is usually in the range of several kHz. The term "sampling period" in Table 2.6 reflects the step size of the time of the available averaged power values of the measurement data. The specific properties for each data set are listed in Table 2.6. These include, the number of investigated households, the measurement intervals as well as the availability of reactive power values. Beyond this, for most data sets the specific appliances within the data set are sub-metered by a separate power measurement device. Additionally, the aggregated load curve from the mains power supply is usually available for being able to perform the disaggregation process.

Table 2.6: Overview of energy disaggregation data sets

Data Set	No. of Households	Sampling Period	Measurement Interval	Sub-metered/ Aggregated Data	Reactive Power	Data Gaps	Usage in Thesis
		seconds	days				Chapter
REDD [67]	6	3-4	3-19	Yes/Yes	No	Yes	-
Smart* [68]	3	1	~90	Yes/Yes	Yes	Yes	-
ADRES [66]	40	1	2x14	Yes/Yes	Yes	Yes	3
Tracebase [69]	~10	1-8	n.a.	Yes/No	No	No	-
TUG [72]	5	1	14	Yes/Yes	Yes	No	5&6
Pecan [70]	10	60	7	Yes/Yes	n.a.	n.a.	-
AMPds [71]	1	60	365	Yes/Yes	n.a.	n.a.	-
BLUED [44]	1	0,016	8	No/Yes	Yes	No	4

Above all, an analysis of the data quality of the public data sets has been carried out. The results show that most available data sets often have measurement errors in form of missing or repeated power data values. This is often caused by radio issues according to the usage of wireless sensors to collect the measurement data. Some examples of these errors are shown in Figure 2.11 and Figure 2.12. In comparison to that, a typical load curve of a fridge can for example be seen in Figure 3.2.
Since some parts of the available load curves from the public data sets include these errors, an evaluation of the actual performance of an energy disaggregation algorithm can lead to misleading results. Beyond this, a direct comparison of the performance of different energy disaggregation algorithms can provide incorrect results. However, a commonly used data set for testing the performance of different algorithms is indeed necessary to boost further research on this topic and make the individual results comparable.

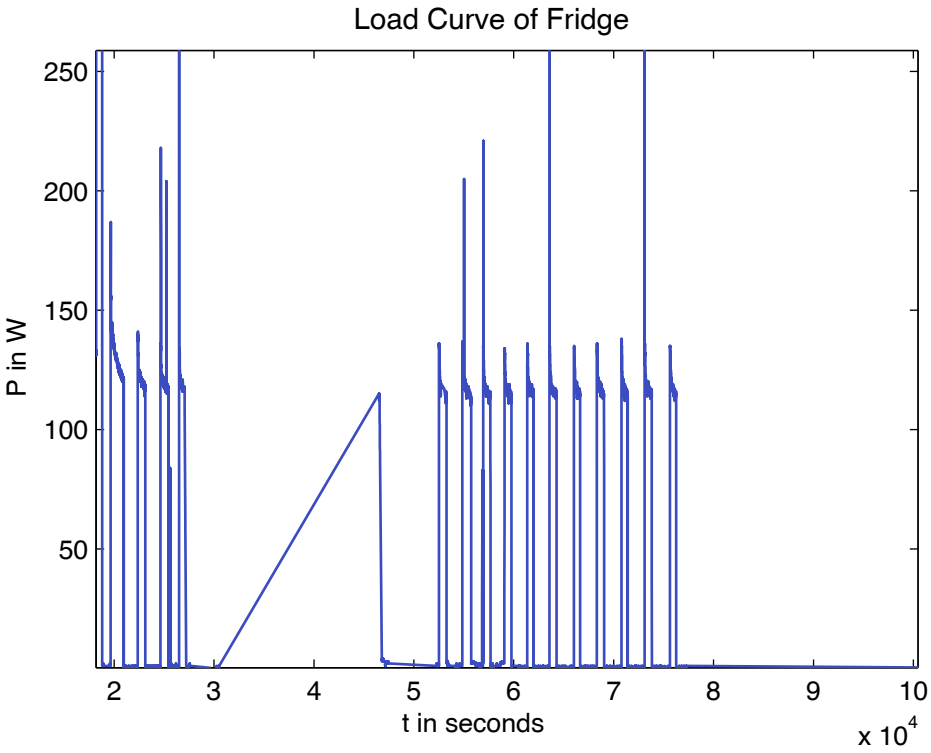

Figure 2.11: Load curve of a fridge from the REDD data set (home 3)

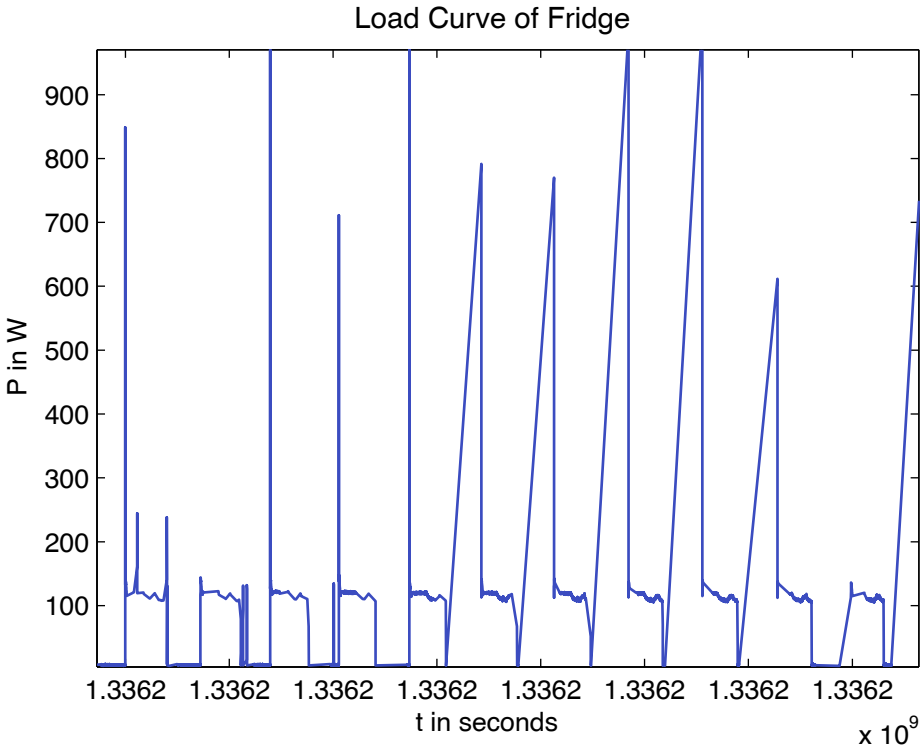

Figure 2.12: Load curve of a fridge from the Smart* data set (home A)

To fit the specific needs of an unsupervised energy disaggregation a data set which doesn't include measurement errors or data gaps is necessary. Furthermore, additional features such as reactive power values allow a further analysis of the influence on the overall performance of algorithms. Due to these reasons, in this thesis a private data set for the energy disaggregation process is chosen which was collected at Graz University of Technology (TUG) [72]. These data are used in Chapter 5 and 6 for conducting the performance analysis.

Since the prerequisites for the in this thesis presented energy disaggregation algorithm are the characteristics of electrical appliances as well as event detection two further data sets are used to investigate in these tasks.
For gathering the characteristics of electrical appliances a broad range of different appliances for the same type with a measurement period of several days is necessary. The "Autonomic Decentralized Regenerative Energy System" (ADRES) data set [66] which measurement campaign took place in about 40 households in Upper Austria for two periods of about 14 days perfectly fits these needs and is therefore used in Chapter 3. Since there are also some measurement errors available within the data set, these data is not used for evaluating the performance of the proposed energy disaggregation algorithm.

In contrast to all other available data sets the focus of the "Building-Level fUlly-labeled dataset for Electricity Disaggregation" (BLUED) [44] is to provide timestamped events of state changes from appliances measured in one single home. It forms a basis for evaluating the accuracy of event detectors. Therefore this data set is used in Chapter 4 for comparing different event detection algorithms. However, the BLUED data set doesn't include sub-metered power data and therefore cannot be used for evaluating the performance of energy disaggregation algorithms.

3 Characteristics of Domestic Appliances

This chapter provides an overview of the characteristics of the most common and energy-consuming electrical appliances which are used in the residential sector. Supplementary information on appliance-specific consumption is described in some detail in Chapter 2.

3.1 Load Profiles

In this section characteristics as well as load profiles of selected electrical appliances are presented.

3.1.1 General Description

A load profile shows how the power changes over time. This can be best described by the following notation: An active power can be defined by $P(t)$ and a reactive power reading by $Q(t)$ where a time step is denoted by $t \in \{1,...,T\}$. The sum of all power values is described with $\boldsymbol{P} = \{P(1),..., P(T)\}$ and $\boldsymbol{Q} = \{Q(1),..., Q(T)\}$ which is used for representing the load profiles in this thesis.

3.1.1.1 Appliance Models

Each electrical appliance exhibits different power consumption patterns which are contributed to the functionality of the device. The simplest functionality is given when a device can be turned on and off. In the on-state a device consumes constant power and after the turn-off the consumption reverts back to zero. Among these are devices such as water heaters, light bulbs and electrical radiators. These devices show only two different states – on and off. Moreover, their load curve is relatively simple since it changes from zero to a constant power level and goes back again to zero when a device is switched off.

Appliances which can consume different power levels, such as washing machines, have more than two states. This is typically caused by the integrated water heater which consumes a certain amount of power when activated and the electrical motor which also consumes a certain amount of power. Since the motor and the water heater turn-on and -off at specific times, the total power which is consumed by washing machines vary. Such behaviour can be best described with different states.

In general, there can be distinguished between four different appliances types. Devices with two states belong to the Type-I appliances and the latter mentioned multi-state appliances are found under Type-II appliances. A description of all types of appliances can be found in Section 2.3.

3.1.1.2 Influence of Voltage

According to EN 50160 [88] the voltage in the distribution network has to be in the range of ±10% of its nominal value (230 V). Since the power consumption of resistive loads such as heaters can be calculated with U²/R, the power consumption value can vary up to approximately ±20% of the nominal value. When the power values are compared, a correction of the measured power value P_m and the voltage U_m to the normalized power P_{230} and normalized voltage U_{230} = 230V is necessary [27], [28]:

$$P_{230} = \left(\frac{U_{230}}{U_m} \right)^{\kappa} \cdot P_m \tag{3.1}$$

The factor κ is within the interval [1...2] and depends on the actual device. Resistive loads, for example, have a factor of κ = 2. Electrical devices with a switching power supply have a factor close to 1.

3.1.1.3 Turn-On Transients

Electrical devices which are equipped with an electrical motor or with capacitive or inductive loads have a high turn-on-current. These turn-on-currents lead in a row to high turn-on power transients. In the residential sector, for example, refrigerators can be counted among them. The duration of turn-on transients is typically in the range of about one to two seconds. For measuring turn-on transients, sampling periods in the range of milliseconds (ms) are necessary.

Figure 3.1 shows turn-on transients of a fridge which last for about 1 second and have a maximum within the first 200 ms. In this special case, the maxima are between 15 and 20 times greater than their steady state power values.

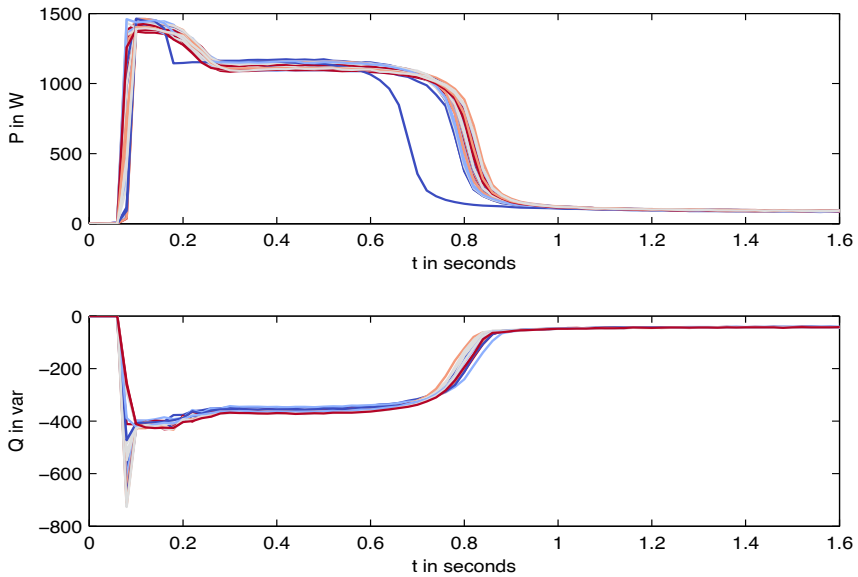

Figure 3.1: Turn-on transients of a fridge

As can be seen in Figure 3.1, the curves of the transients are very similar of the refrigerator under investigation. However, the curves of the transients can vary between different appliances which have the same functionality. Besides, some electrical appliances do not have such constant curves which makes it more difficult in terms of using the turn-on transients as a distinctive feature. Especially when low sampling periods in the range of about 1 Hz or below are selected, the total curve is averaged to approximately two or three samples. As a consequence the transient as a feature gets more difficult or even impossible to differentiate between similar devices.

3.1.2 Residential Appliances

This section represents load profiles from selected electrical appliances in the residential sector. The sampling period of the load profiles in this section is one second. The data measurement was carried out in five selected homes in Austria.

3.1.2.1 Refrigerators and Freezers

The refrigerator as well as the freezer belong to the Type-I appliances. When the temperature in the inside of the fridge or freezer exceeds a certain limit, the cooling compressor is switched on and remains switched on until the inside temperature reaches the predefined lower limit. Usually the refrigerator's turn-on and turn-off durations vary within a certain constant range, see Figure 3.2.

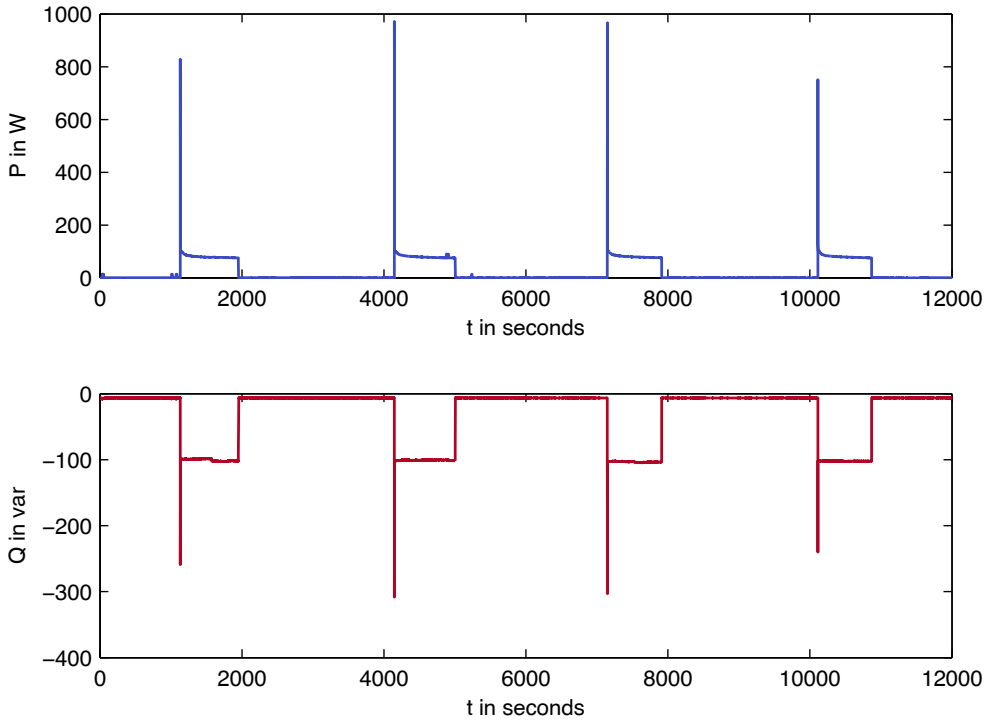

Figure 3.2: Load profile of a fridge

The load profile of a freezer looks very similar to the load profile of a refrigerator, they both have similar power consumption values. Characteristically, refrigerators and freezers have turn-on transients which can spike up to approximately 600 W to 1500 W which can be seen in Figure 3.2. The above mentioned sampling period of the load profile which is one second is similarly to the duration of the turn-on transients. Due to this reason the load curve of the turn-on transient gets mostly recorded by two samples since the starting point of the transient and the sampling period vary. This can also be seen in Figure 3.2 through the different maximums of the turn-on transient. The reactive power values of the investigated fridge are negative. Reason for this is that the device is overcompensated.

Nowadays, some refrigerators are equipped with a defrost system which turns on a few times per week and last for approximately 10 to 30 minutes. Typically the defrost cycle consumes a more or less constant power level.

3.1.2.2 Dishwashers

Dishwashers belong to the Type-II appliances since they have multiple states and different power consumption levels. Typically, there are two intervals where the heating system of dishwashers is turned as is evident in Figure 3.3. The first heating cycle increases the water temperature because of the cleaning process and the second cycle heats the water so that the dishes can be rinsed and finally drained. The power consumption of the heating element is around 2000 to 3000 Watts. By comparison, the water pump which propels the water to the spray arms has only a consumption of approximately 50-150 W. Figure 3.3 clearly shows that the water pump is switched on at the beginning of the washing process. After a certain period of time the heating element is also turned on to heat the soapy water according to the preset temperature. The exact on-duration of the heating cycle is mainly dependent on the predefined settings of the user. When the water inside the dishwasher is changed and fresh water is filled during the rinsing process the water pump stops for a while. In the last draining cycle the water is heated and the pump stops, so that the dishes can dry.

3.1.2.3 Washing Machines

The washing machine is an example of a Type-II appliance with striking load profile. Prominent is the power consumption of the motor which rotates the drum. This is illustrated by the power fluctuations of about 100 W, see Figure 3.4. The drum usually rotates counter-clockwise as well as clockwise with a short time span where it stops. Similarly to dishwashers, washing machines have a heating element for water. Depending on the selected program and the fill quantity, the heating duration varies. In specific programs the water is reheated several times. Every time before the water is changed the rotational speed of the drum is increased, power fluctuations rise to about 400 - 500 W. At the end of the program usually final spinning cycles remove water from the clothes.

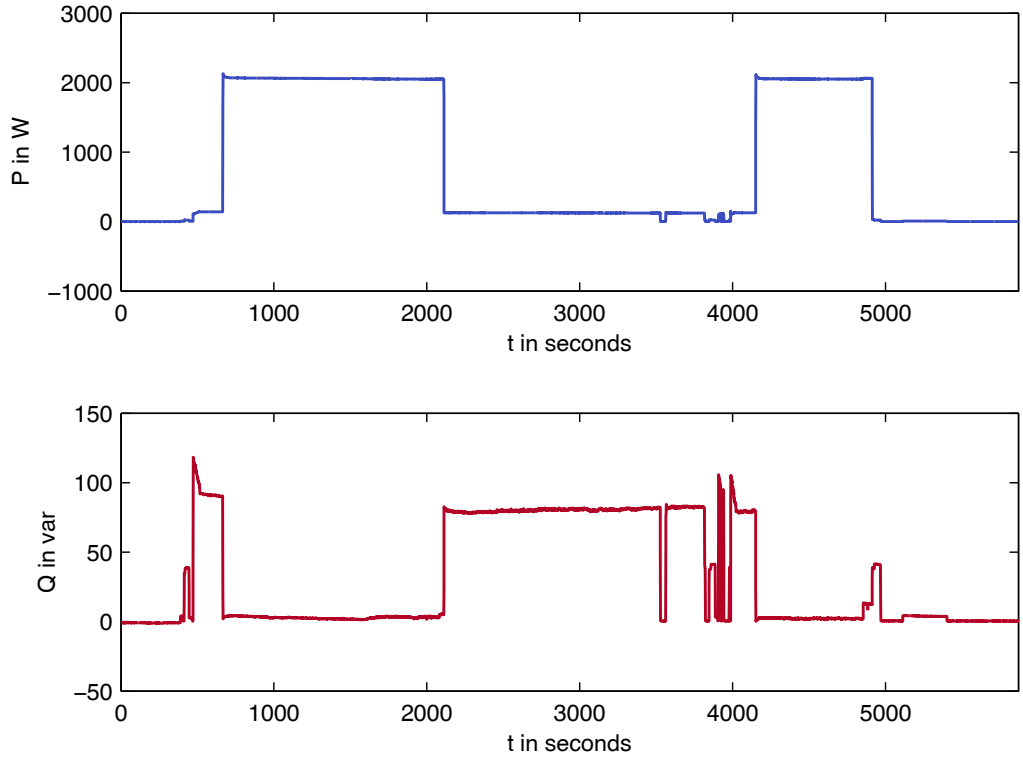

Figure 3.3: Load profile of a dishwasher

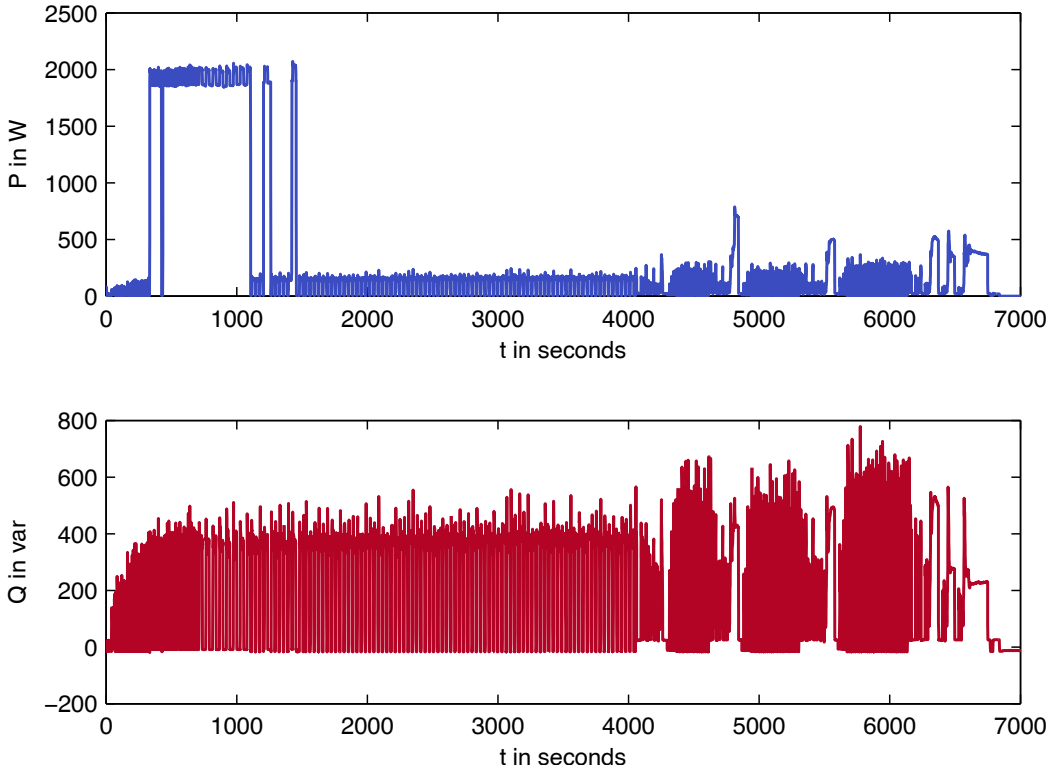

Figure 3.4: Load profile of a washing machine

3.2 Appliance-Specific Usage Patterns

In this section the typical ranges such as the on-duration or active power consumption of selected electrical appliances (including fridges, freezers, dish washers and washing machines) are presented. It can be seen that within a certain appliance group different variations such as the specific on-duration occur.

Appliance-specific usage patterns facilitate a method for assigning power events to correlated electrical devices which can be used for unsupervised energy disaggregation [73], [85]. However, one single usage pattern such as on-duration is not enough for distinguishing between different devices. The combination of on-, off-duration(s), time(s) of use as well as active and reactive power values deploy a great potential for a successful differentiation between major electrical devices.

3.2.1 Test setup

Detailed measurements of active power values in 40 households have been carried out by the project ADRES [66] (see also Section 2.4). The measurement campaign was carried out in two different periods, each of it with a time interval of two weeks in every household. The selected sampling period is one second. Typical usage patterns of each electrical device have been calculated and the mean of those values is presented in the figures. Overall, the load profiles of 23 fridges, 25 freezers, 25 dish washers and 36 washing machines have been collected, and the data of these measurements are presented in this section.

3.2.2 Power Consumption

For the purpose of calculating the power consumption, the average of the total power consumption values of a certain device when it is turned on are used. In the case of washing machines and dishwashers, the average of the power consumption of the heating element is used for illustration.
When comparing the active on-power power values in Figure 3.5 it is evident that fridges and freezers have nearly the same histograms and consume mostly approximately 100 W. In contrast the dishwashers' heaters, separated into two groups consuming about 2000 W and 2500 W respectively. The investigated washing machines in the sample are well approximated by a Gaussian-distributed power consumption with a mean of about 2000 W for the heating elements. Different types of a specific electrical appliance group have different levels of power consumption. However, the power consumption of each appliance group remains within certain predetermined ranges.

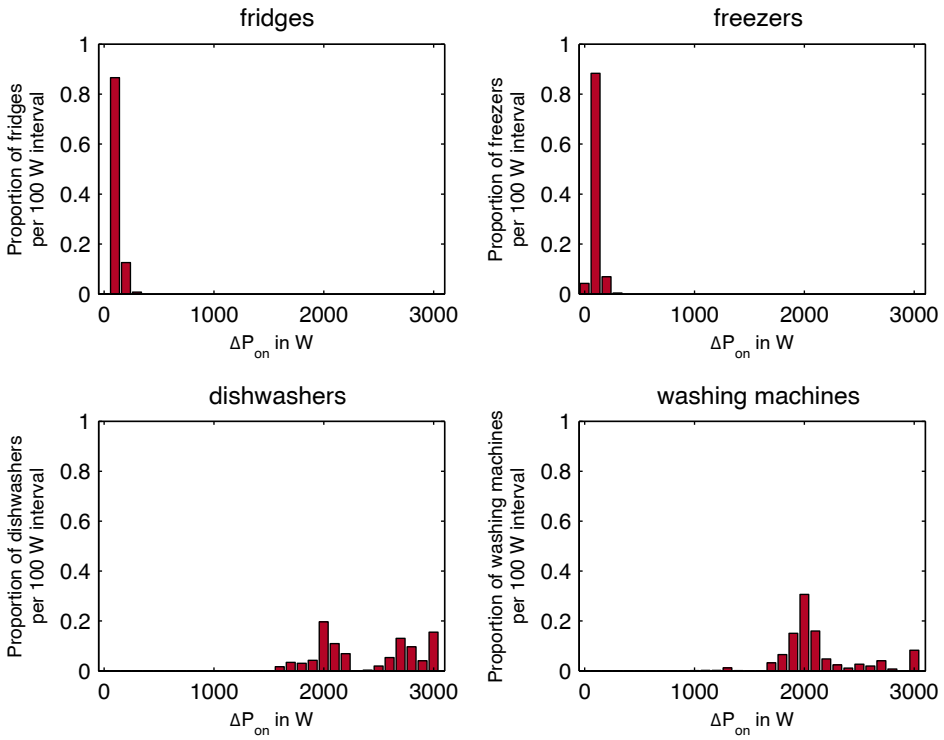

Figure 3.5: Histograms of average on-power consumption of proportion of selected electrical appliances

3.2.3 On-Duration

All investigated appliances have a typical average on-duration of about 10-20 minutes, see Figure 3.6. Especially the histograms of the fridges and the freezers look quite similar which is not all that surprising, since both electrical appliances has the same mode of operation. Apart from this, the on-time duration of fridges and freezers can vary up to several hours. This can be caused by lowering the predefined temperature in those devices or by adding large amounts of food at room temperature for cooling purposes.
The typical on-duration of heaters of dishwashers and washing machines ranges from several minutes to half an hour. The water temperature and the selected program constitute a primary influence.

The variation of the on-duration time in proportion to the median values of the on-duration of investigated devices is illustrated in Figure 3.7. An interesting fact about the on-duration of fridges and freezers is that the time interval usually remains relatively constant. On average, more than 90% of all intervals are in the range of ±10% of the median value of the on-durations. Dishwashers have relatively frequent median on-durations in approximately 40% of all heating phases. In comparison to the histograms of fridges and freezers the dishwashers' on-duration distribution is flatter. The evident distribution is predominantly caused by the second heating cycle during the cleaning process.

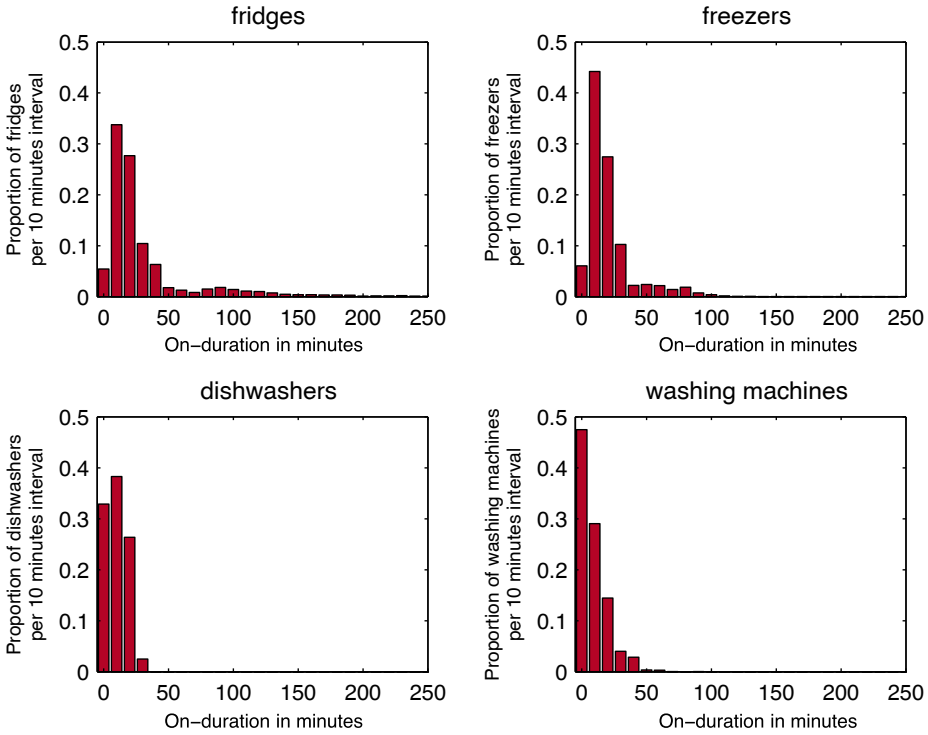

Figure 3.6: Histograms of average on-durations of proportion of selected electrical appliances

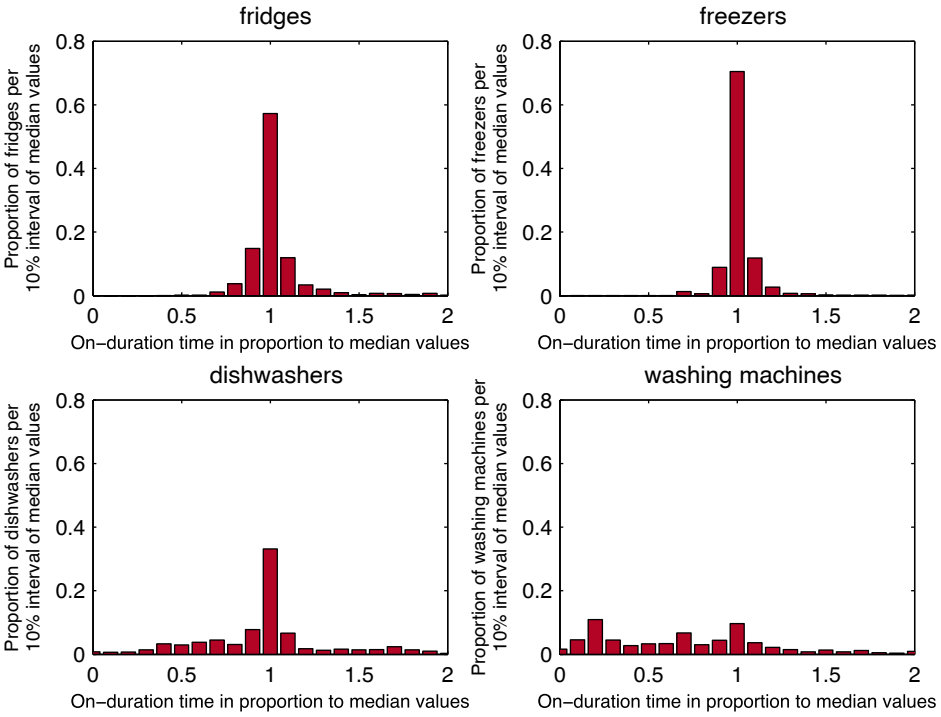

Figure 3.7: Histograms of variation of the on-duration time in proportion to the median values of selected electrical appliances

A completely different picture is illustrated by the variation of the on-duration of the washing machines. This is caused by the usage of different washing programmes which accompany variable predefined heating temperatures. Besides, the reheating process of the water plays another major role with the effect that the distribution remains rather uniform.

3.2.4 Off-Duration

The off-durations of the sample devices present a similar data in Figure 3.8. However, it can be seen that freezers tend to switch on more frequently than fridges. This can be explained particularly by the fact that those devices have a lower predefined cooling temperature, mostly a greater total volume and consequently greater heat losses than the fridges.

The off-durations of dishwashers and washing machines have two main contributors. Due to the fact that during a washing cycle the water is reheated, there is a peak in the histograms between 0 to 10 minutes. Since dishwashers dry the dishes at the end of the program, there is a particular share of off-durations in the interval of 30 to about 60 minutes. According to the histogram of the washing machines it can be seen that a proportion of washing machines has off-durations of about 100 to 200 minutes which is caused by the fact that washing machines are often used several times in a row.

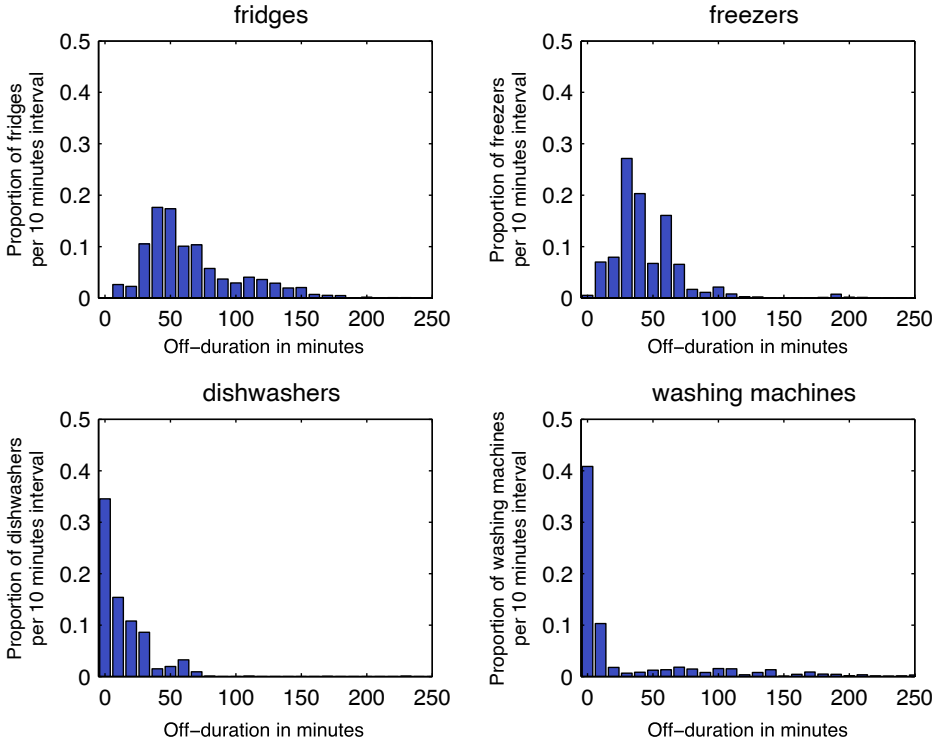

Figure 3.8: Histograms of average off-durations of proportion of selected electrical appliances

The variation of the on-duration compared with the variation of the off-duration are relatively similar, see Figure 3.9. The most apparent difference can be seen in the distribution of the variation of the off-duration of fridges which is a bit flatter. This is caused by the fact that fridges are opened more frequently than freezers. In addition to this, food which needs to be cooled is put more frequently into fridges. As a consequence the inside temperature of fridges rises faster and the upper cooling temperature level is reached in a shorter period, this causes the cooling unit to be switched on.

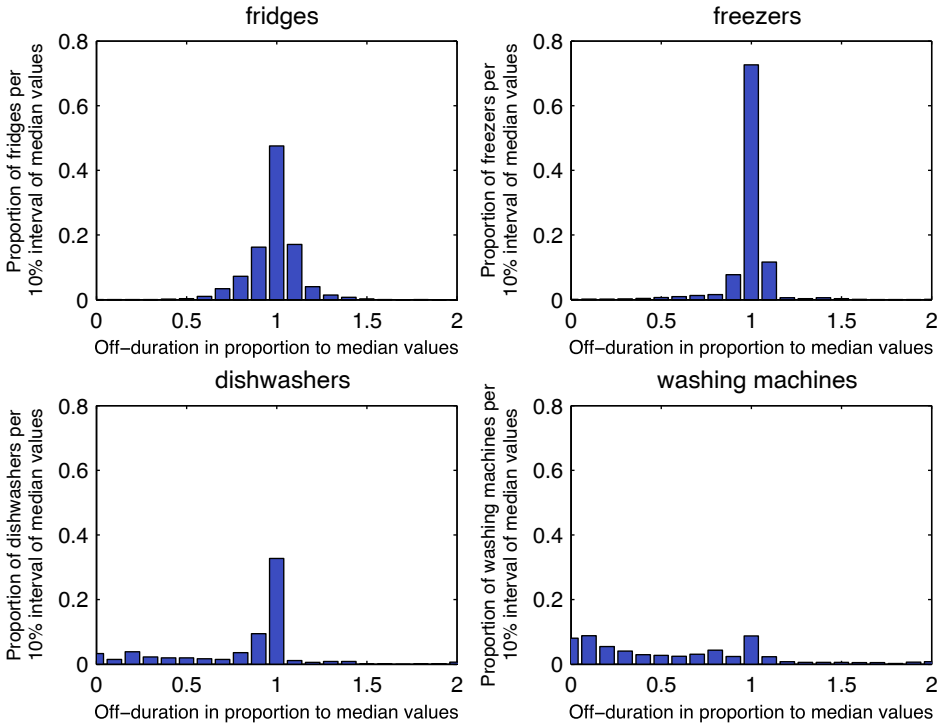

Figure 3.9: Histograms of variations of the off-duration time in proportion to the median values of selected electrical appliances

3.2.5 Time of Use

Whereas the probability distribution for the time of use for fridges and freezers is constant over the day, it varies for devices which are manually switched on. In Figure 3.10 it can be seen that dishwashers are typically turned on after lunch between 1 and 2 p.m.[1]. Besides this, between 8 and 9 a.m. as well as 8 to 10 p.m. it also has a high switch-on probability. In contrast to this, the probability curve for the switch-on of the washing machine has a peak around 10 a.m.

Due to the small sample size of the measured data a division by weekdays is hardly representative. However, if there were sufficient statistics available it would be meaningful for further appliance-specific usage patterns.

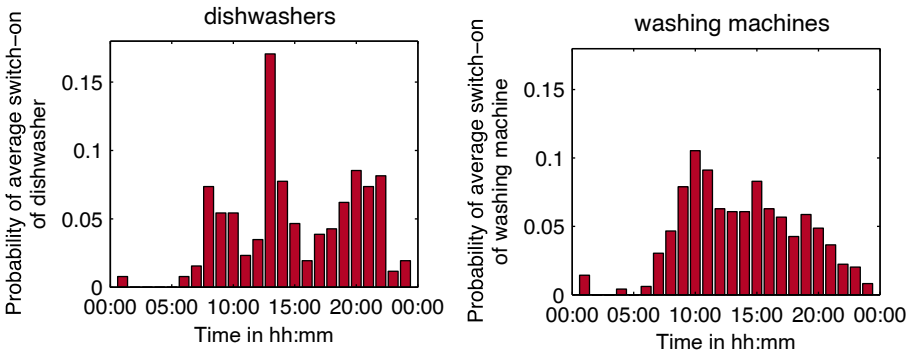

Figure 3.10: Histograms of average time of use of proportion of selected electrical appliances

[1] Austrian data – as lunch is the biggest meal

4 Feature Extraction

Different features such as current harmonics or turn-on transients were previously described in Chapter 2. In this Chapter the focus is on the accuracy of event detection algorithms.
Not all of the features which are used in energy disaggregation algorithms are adequate for analyzing low resolution power measurement data: for example, some use harmonics of the active or reactive power to detect events and monitoring this, requires high sampling frequencies. Current smart meters available on the market have only the capability to provide measurement power data of a sampling period of one second or above (see Section 2.2.2.2) of active (P) and reactive power (Q). In this Chapter only feature extraction methods which are capable of dealing with this low resolution power measurement data, with a sampling period of one second or above, are investigated.

4.1 Event Detection

This section gives some background related to event detection algorithms. Different event detection algorithms which are capable of dealing with sampling periods of about one second are described and compared against each other with different metrics. Beyond this, the influence of the sampling period on the accuracy of the event detectors as well as the incorporation of reactive power values is investigated.

4.1.1 General Overview (State of the Art)

The detection of turn-on or -off events (step changes) of electrical appliances was firstly described by Hart [28] in 1992 and is one of the commonest features in non-intrusive appliance load monitoring (NIALM) [30]. An event can be described as a certain signature which is generated when an electrical appliance is turned on or off or switches its state. An electrical appliance such as a refrigerator consumes about 100 W and 100 var for a period of several minutes. The turn-on and the turn-off results in a change in the power demand of a certain device.
When the total power consumption of an individual consumer is monitored, turning on e.g. the oven boosts the total power consumption and can be seen as a step increase. Figure 4.1 shows a total load curve of a residential consumer with events from different electrical appliances.
However, step changes are not always a consequence of turn-on or -off events. Electrical appliances with time-varying power consumption such as computers comprise of various step changes in their

power consumption. Frequent changes in power consumption and the fact that a number of electrical appliance events can occur simultaneously within the time resolution step (e.g. 1 second) are considered complicating factors when matching events to particular appliances. Importantly therefore, beyond event detection additional features are necessary for NIALM to obtain high recognition accuracy.

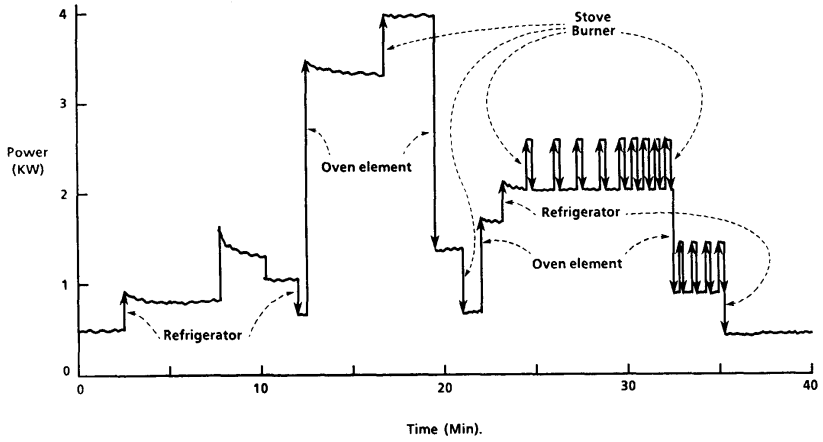

Figure 4.1: Events of different electrical appliances, source: [28]

A good overview of event detection algorithms published so far, can be found in Anderson et al. [45]. These approaches can be classified into three different categories: I) expert heuristics, II) probabilistic models and III) matched filters; relevant event detection algorithms will be described in detail below. Furthermore the following definitions will be used: an active power reading is defined by $P(t)$ and a reactive power reading is $Q(t)$ where a time step is denoted by $t \in \{1,2,...,T\}$.

4.1.1.1 Expert Heuristics

There exist various expert heuristics for event detection, most notably by Hart in the first instance. In 1992 Hart proposed an approach to non-intrusive appliance load monitoring [28]. For event detection, as a first step the power data which is computed with the voltage was normalized (see Section 3.1.1). In a second step the power data is then segmented into "*periods in which the power is steady and periods in which it is changing*" [28]. Steady periods are defined when the power does not vary within a certain tolerance of several watts and when the periods exceed a minimum time duration. The remaining periods are classified as "changing". The step sizes are calculated by subtracting the average values of steady periods which occur before and after periods of change. In comparison to Hart, Farinaccio [41] is evaluating the change of power $\Delta P(t)$ for event detection:

$$\Delta P(t) = P(t) - P(t-1) \tag{4.1}$$

Consecutive power changes are added up and compared to predefined power ranges of electrical appliances. These predefined levels are recorded with several measurements in a training phase. If the computed power changes range within the minimum and maximum values they are then assigned to a specific electrical device. A quite similar approach is proposed by Bijker [58] who also calculates more than two consecutive power readings.

Baranski (Expert Heuristic)

Another approach which is based on Farinaccio [41] is proposed by Baranski [42]. A power change $\Delta P(t)$ is assigned to an event E_i when its sign and the sign of the preceding power change $\Delta P(t\text{-}1)$ coincide. If the signs of the power changes diverge and the power change is greater than a predefined minimum power threshold P_{min} the power change is assigned to the next event E_{i+1}:

$$\Delta P(t) \mapsto \begin{cases} E_i & \text{if } |\Delta P(t)| > P_{min} \wedge \operatorname{sgn}(\Delta P(t)) = \operatorname{sgn}(\Delta P(t-1)) \\ E_{i+1} & \text{if } |\Delta P(t)| > P_{min} \wedge \operatorname{sgn}(\Delta P(t)) \neq \operatorname{sgn}(\Delta P(t-1)) \end{cases} \tag{4.2}$$

In contrary to the approach of [41] which only uses two consequent power changes, in (4.2) the number of power changes can be arbitrary with a length of N_i. The steady state power change of an event $\Delta P(E_i)$ can be calculated by the sum of corresponding changes of power $\Delta P(j)$.

$$\Delta P(E_i) = \sum_{j=1}^{N_i} \Delta P(j) \tag{4.3}$$

Turn-on transients, e.g. starting of an inductive motor, are usually detected by two events due to their power change. For avoiding this, events with positive power changes which are followed by an event with a negative power change within a predefined duration threshold τ are merged to a single event E_i':

$$E_i' = E_i + E_{i+1} \forall E_i \left| \Delta P(E_i) > 0, \Delta P(E_{i+1}) < 0 \wedge t(E_{i+1}) - t(E_i) < \tau \right. \tag{4.4}$$

Bergman (Expert Heuristic)

A relatively recent approach from Bergman et al. [43] uses a running average of power values $P_{Avg}(t)$ with a window of width w for event detection. Through the window w a smoothing of the power values is realized:

$$P_{Avg}(t) = \frac{1}{w} \sum_{j=t}^{t+w-1} P(j) \quad 1 \leq t \leq T - w + 1 \tag{4.5}$$

An event is detected when the difference of two consecutive average power values exceeds the threshold P_{thr}. The parameter of the threshold P_{thr} mainly depends on the selected width w.

For eliminating detection errors caused by transients as already mentioned above, the quantity of δ power averages has to be below the limit P_δ before an event is said to have occurred. Otherwise, the event is ignored:

$$E_i \quad if \left|P_{Avg}(t) - P_{Avg}(t+1)\right| > P_{thr} \wedge \left\{\left|P_{Avg}(t-\delta+1)\right|,...,\left|P_{Avg}(t-1)\right|\right\} < P_\delta \quad \text{where } \delta+1 \le t \le T-w+1 \tag{4.6}$$

4.1.1.2 Probabilistic Approaches

Especially with high sampling rates in the area of several kHz, the mentioned expert heuristic approaches previously discussed are not applicable due to the high-resolution load curves. An opportunity for the detection of events in these high-resolution data can be carried out with statistical analysis. The detection of *abrupt changes* in time series is referred to as *change detection* in the in statistical literature [46]. The term "abrupt changes" can be best described as "*changes in characteristics that occur very fast with respect to the sampling period of the measurements, if not instantaneously*" [46]. There have been several statistical methods applied to event detection in NIALM.

Generalized Likelihood Ratio (Probabilistic Approach)

By Dong Luo et al. [47] the generalized likelihood ratio (GLR) is applied for event detection. "*The GLR detection algorithm calculates a decision statistic from the natural log* [note: ln] *of a ratio of probability distributions before and after a potential change in mean*" [47]. However, it is necessary to train the parameters of the detector before it can be applied to real data in order to obtain results with sufficient reliability and that is a considerable drawback.

A modified GLR detector which is based on [47] is introduced by Berges et al. [48]. Through the simplification of the necessary input parameters that need to be set, this detector can be more easily applied for event detection. In this event detection algorithm two sliding windows (pre-event and post-event window) with the same length w_l are used for expert calculating the mean value μ as well as the standard deviation σ:

$$\mu(\mathbf{P}_{\mathrm{Pr}e}), \sigma(\mathbf{P}_{\mathrm{Pr}e}) \quad \text{where } \mathbf{P}_{\mathrm{Pr}e} = \{P(t-w_l),...,P(t-1)\} \tag{4.7}$$

$$\mu(\mathbf{P}_{Post}), \sigma(\mathbf{P}_{Post}) \quad \text{where } \mathbf{P}_{Post} = \{P(\mathrm{t}+1),...,P(t+w_l)\} \tag{4.8}$$

The two windows are separated by a single sample which is called the point of interest. The likelihood ratio l_t for every point of interest can be calculated with:

$$l_t = \ln \frac{P\left(P(t) \mid \mu\left(\mathbf{P}_{Post}\right), \sigma\left(\mathbf{P}_{Post}\right)\right)}{P\left(P(t) \mid \mu\left(\mathbf{P}_{\mathrm{Pr}e}\right), \sigma\left(\mathbf{P}_{\mathrm{Pr}e}\right)\right)} \quad w_l + 1 \le t \le T - w_l + 1 \tag{4.9}$$

When the likelihood ratio for every single point is computed a voting procedure is started. To carry out this task as a first step a voting window w_{vote} is used for computing the test statistics which then determines which point receives a vote. The size of the voting window w_{vote} is $2 \cdot w_l + 1$ and its starting point is equal to the starting point of the point of interest t. The test statistics s_t are calculated by the sum of the likelihood ratios from the point of interest to the last point in the event detection window for each point t:

$$s_t = \sum_{j=t}^{last(w_{vote})} l_j \tag{4.10}$$

After having calculated the test statistics s_t, in a second step the voting window is used to assign votes. To put this into practice the voting window starts at the first value of the test statistics s_t and a vote for a single point is assigned when it satisfies the maximum value in the vote window:

$$vote_{index} = \arg\max_{t \in w_{vote}} s_t \tag{4.11}$$

The point with the maximum statistics receives a vote, the vote window slides for one sample and another vote is carried out. Each point which receives more than the predetermined threshold v_{thr} votes is labeled as an event. An illustration of the functional principal can be seen in Figure 4.2.

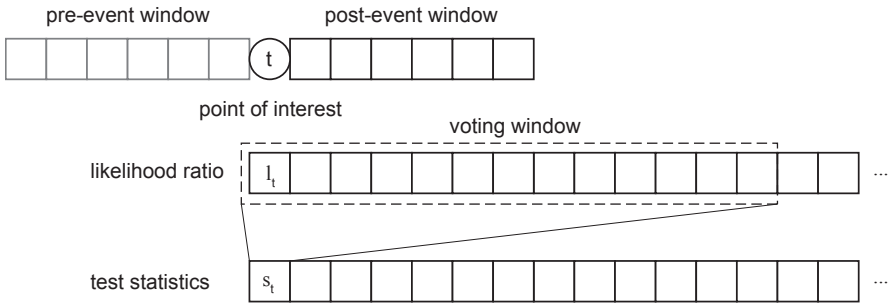

Figure 4.2: Illustration of the functional principal of the GLR detector

If there are several events within the pre- and post- event window the standard deviation σ of both windows rises. As a consequence the likelihood ratio in (4.9) decreases for this specific events and consequently not all of those events receive so many votes greater than the threshold v_{thr}. These events are not detected by the GLR detector. For increasing the detection accuracy the number of events within the pre- and post-event window have to be a minimum. This can be realized by a proper selection of the window widths w_l. If the sampling period is in the range of one second or above, an engineering approach has shown that the window widths have to be around three samples to successfully detect events.

However, to make the algorithm applicable to power measurement data with a sampling period of one second or above, and window sizes around three samples, a minimum value for the standard normal

distribution of σ=1 for the pre- and post-event window has to be established. It has been shown that the (average) power samples values in a window can be in such a small range that the standard normal distribution σ is far below 1. This leads to very unlikely probabilities of the point of interest even when its value is very close to the values in the pre-event window. However, by limiting the standard normal distribution, the probability also gets limited to more appropriate values.
It is worth to mention that the limitation and the preset of the parameters are based on engineering experience. Sample sizes of a around three elements are typically not used to compute statistical values such as the standard normal distribution.

Goodness of Fit (Probabilistic Approach)

Another relevant statistical approach for event detection in NIALM is introduced by Jin et al. [49] and is compared with GLR in [59]. "*The goodness-of-fit* [note: GOF] *test seeks to determine whether a set of data could reasonably have originated from some given probability distribution*" [49]. A known distribution function $F(x)$ is compared against an *a priori* unknown distribution function $G(x)$ which is drawn from n independent and identically distributed random samples x_i, $i = 1,...,n$. If the two distribution functions are identical the null Hypothesis H_0 of the binary hypothesis testing problem of the GOF test is fulfilled:

$$\begin{aligned} \mathrm{H}_0 &: \quad G(x) = F(x) \\ \mathrm{H}_1 &: \quad G(x) \neq F(x) \end{aligned} \tag{4.12}$$

Therefore with the GOF test the discrepancy between two distribution functions can be measured. In event detection the samples of a pre-event window are used for computing $G(x)$ and the samples of the post-event window for computing $F(x)$. If the two distribution functions do not match each other (H_1) an event is detected.
The data model for a certain power change is put together with an appliance-specific transition $e(t)$ and a disturbance $w(t)$ which is a white Gaussian process [49]:

$$P(t) = e(t) + w(t) \quad 1 \leq t \leq n \tag{4.13}$$

where n is the observation window size.
To measure if a given distribution function matches a known data set there are several tests, but the χ^2 test has been widely used in statistical literature [49]. However, the decision threshold $\chi^2_{\alpha,n-1}$ (of the χ^2 test) for distinguishing between the two hypotheses depends on the $100(1-\alpha)\%$ confidence interval and $n-1$ degrees of freedom. Hence it depends on the window size n of pre-event and post-event windows as well as on the detection confidence level α.
For calculating the minimum sample size n_0 the data model in (4.13), which is Gaussian distributed and has a standard deviation of σ_w and a mean μ_w, plays a big role. To estimate μ_w in a certain

confidence interval in the pre-event window the sample mean $\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$ can be used. It can be *100*(*1*-α)*%* confident that the error $|\mu_w - \overline{w}| \leq L$ does not exceed an maximum amount *L* when the minimum sample size is [49]:

$$n_0 = \left(\frac{z_{\alpha/2}\sigma_w}{L} \right)^2 \tag{4.14}$$

where $z_{\alpha/2}$ is the upper *100α/2* percentage point of the standard normal distribution and *L* is the minimum amount of an event to be able to detect it. If for example in a 90% confidence interval $\alpha = 0.1$ and $z_{0.05} = 1.645$ and a standard deviation of $\sigma_w = 30$ and a minimum power change of *L*=30 W is assumed, the minimum sample size $n_0 = 2.7$. So, a minimum window size of $n_0 = 3$ samples for a pre- and post-event window is necessary to be able to detect events. Due to the reason that at least 6 samples are required to detect an event which corresponds to a time interval of at least 6 seconds, the method does not work when there are several events within this time interval. The discrepancy between two distribution functions cannot be measured correctly when there are two or more events in one of the two windows. Therefore sampling periods of one second or above are not applicable for GOF.

4.1.1.3 Matched Filters

Events of electrical appliances can also be detected by matched filters. In a training phase typical transients of electrical devices such as turn-on transients can be detected and manually classified in a database. An unknown signal in the load curve is correlated with the signals in the database. When the two signals match the event can be classified which is presented in [51]. Further work based on [51], introduced a shift in time, an offset as well as a gain of the incoming signal to better match the transients in the database [52], [53].

Detailed analysis of the low resolution load profiles has been shown that “Matched Filters” are not very applicable for event detection with a measuring period of one second or above. Due to the varying offset between the sampling period and the events from a single electrical appliance, different averaged load curves such as turn-on transients are generated which cannot be easily matched against each other since they have different maximum values (see Section 3.1.1).

4.1.2 Metrics for Event Detection

This section gives an overview of how to evaluate event detection algorithms and presents some background information as well as different metrics and a score function.

4.1.2.1 Fundamentals

When events or samples are detected it can be distinguished between four different outcomes in a binary classification task[1]:

- true positive (TP, correct detection),
- true negative (TN, no detection),
- false positive (FP, incorrect detection), and
- false negative (FN, missed detection).

Figure 4.3 illustrates the different outcomes as mentioned. This notation is used in the next sections.

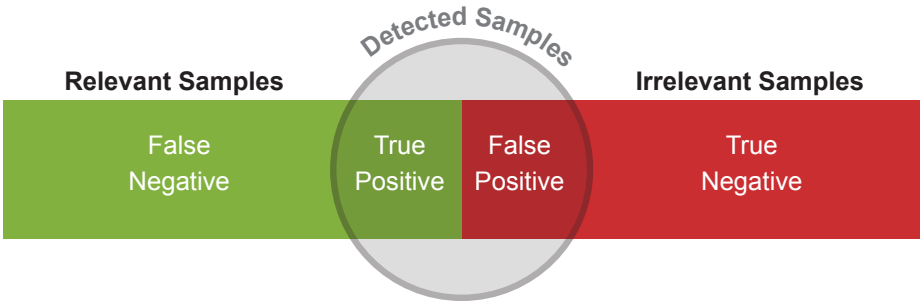

Figure 4.3: Illustration of the outcomes of a binary classification

For comparison of event detectors Anderson et al. [45] proposed four different metrics as well as a score function. Each metric has a specific goal and results in a different performance. However, the specific needs of an individual NIALM algorithm have to be taken into account by the selection of a single metric. The advantages of the individual metrics are described below.

4.1.2.2 Rate Metric

The simplest metric for evaluating an event detector is just to consider the trade-off between the true positive rate (TPR) and false positive rate (FPR) which shows the performance of a binary classifier. When the input parameters are varied and the fractions of TPR and FPR are plotted against each other in a diagram, you get the receiver operating characteristics which can illustrate the performance graphically. TPR and FPR can be computed with the number of true positives (TP), true negatives (TN), false negatives (FN) and false positives (FP):

$$TPR = \frac{TP}{TP + FN} \tag{4.15}$$

[1] For illustration, a simple example from the field of protection engineering in the case of detecting network failures: TP (correct pickup and trip), TN (correct "non-reaction"), FP (overreaction), and FN (underreaction).

$$FPR = \frac{FP}{FP + TN} \tag{4.16}$$

The perfect detector would have a TPR of 1 and a FPR of 0. When a detector is denoted with ψ and its parameter set with Ψ, the best detector $\hat{\psi}_{Rate}$ is that which is closest to the point (0,1):

$$\hat{\psi}_{Rate} = \arg\min_{\psi \in \Psi} \left\| (1,0) - (TPR, FPR) \right\|_2^2 \tag{4.17}$$

Since the number of TN >> FP, the term FPR in (4.16) goes to 0. This leads the best detector $\hat{\psi}_{Rate}$ to be the one which has the maximum number of true positives.

4.1.2.3 Percentage Metric

In the true positive percentage metric the ratio of true positives as well as the number of false positives to the overall number of events is compared. The true positive percentage (TPP) and the false positive percentage (FPP) can be calculated according to the ratio of the number of overall events n_E = TP + FN:

$$TPP = \frac{TP}{n_E} \tag{4.18}$$

$$FPP = \frac{FP}{n_E} \tag{4.19}$$

The perfect detector would have a TPP of 1 and a FPP of 0, while the optimal detector $\hat{\psi}_{Perc}$ is the one which is closest to this point:

$$\hat{\psi}_{Perc} = \arg\min_{\psi \in \Psi} \left\| (1,0) - (TPP, FPP) \right\|_2^2 \tag{4.20}$$

4.1.2.4 Total Power Metric

In the total power metric the size of the power change of the false negative and false positive events are reflected in the distance function.

So far the proposed metrics do not incorporate the size of power changes of events. When it is assumed that power changes of electrical devices with a large power change are more relevant compared to those with small changes, the next two metrics can be taken into account.

The total power change of an event $\Delta P(E_i)$ can be calculated as already described in (4.3) and (4.5). However, when the sampling period is very low (one second or below), it is necessary to mask areas such as transients to compute the correct steady state power change. A pre-event window w_1, a mask window w_2 and a post-event window w_3 are necessary to calculate the means and mask a certain area:

$$\Delta P(E_i) = \frac{1}{w_3} \sum_{t=t(E_i)+w_2+1}^{t(E_i)+w_2+w_3} P(t) - \frac{1}{w_1} \sum_{t=t(E_i)-w_1}^{t(E_i)-1} P(t) \tag{4.21}$$

To incorporate the power change of an event $\Delta P(E_i)$ to the metrics, the false negative (missing) events are denoted with $\mathcal{M}$ and the false positive events with $\mathcal{F}$. The total power change of the misses ΔP_{M} and false positives ΔP_{FP} can be calculated with:

$$\Delta P_M = \sum_{i \in \mathcal{M}} \left| \Delta P(E_i) \right| \tag{4.22}$$

$$\Delta P_{FP} = \sum_{i \in \mathcal{F}} \left| \Delta P(E_i) \right| \tag{4.23}$$

The best detector is the one which is closest to the point (0,0):

$$\hat{\psi}_{\Delta P} = \arg\min_{\psi \in \Psi} \left\| (\Delta P_{FP}, \Delta P_M) \right\|_2^2 \tag{4.24}$$

4.1.2.5 Average Power Metric

In the average power metric the trade-off between the average power of misses and the average power of false positives is balanced out. Instead of evaluating the total power changes, in this metric the trade-off of the average power changes is balanced out:

$$\overline{\Delta P}_M = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \left| \Delta P(E_i) \right| \tag{4.25}$$

$$\overline{\Delta P}_{FP} = \frac{1}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} \left| \Delta P(E_i) \right| \tag{4.26}$$

In this metric the best detector is closest to the point (0,0):

$$\hat{\psi}_{\overline{\Delta P}} = \arg\min_{\psi \in \Psi} \left\| (\overline{\Delta P}_{FP}, \overline{\Delta P}_M) \right\|_2^2 \tag{4.27}$$

4.1.2.6 Score function

In order to combine the above metrics *M,* a score function $S(\psi)$ is defined which rates the overall performance of an individual detector ψ for all four metrics:

$$S(\psi) = \frac{1}{4} \sum_{\mu \in M} \frac{D_\mu(\hat{\psi}_\mu, \theta_\mu)}{D_\mu(\psi, \theta_\mu)} \tag{4.28}$$

whereas D_μ is the Euclidean distance of the metric μ and θ_μ represents the point of the perfect detector in this space. Lets consider for example μ = rate metric: $\hat{\psi}_{Rate}$ is the best overall detector and the point of the perfect detectors is $\theta_{\mathrm{Rate}} = (0,1)$. The average of the ratios of the best detectors regarding the detector under investigation ψ is the score function.

4.1.3 Comparison of Event Detection Algorithms

Not all event detection algorithms are applicable for measurement data with a sampling period of one second or above due to the reason of the required minimum window lengths. Therefore the most relevant and promising are analyzed in detail in this section. For further investigation the algorithms of Baranski, Bergman and Modified Generalized Likelihood Ratio (GLR) are selected and compared. Hence two event detection algorithms from the category "Expert Heuristic" and one of the category "Probabilistic Models" have been selected. Matched filters are not tested for the reasons described above.

To determine the best parameter set for each event detection algorithm the metrics described in Section 4.1.2 are used. As already explained, each metric has a different objective function and as a consequence the results vary. This means that the amount of true positive, false negative and false positive events varies under each metric.

For the sake of comparison the results of the event detection algorithms with a sampling period of one second were used as the best individual scores $\hat{\psi}_{\mu}$ for each metric.

4.1.3.1 Test Setup

As a ground truth the publicly available fully labeled dataset of a single household from the BLUED data set [44] is used (see also Section 2.4). The measurement of this data set was carried out in a home in the US for two mains supplies (line A and B). The dataset has a measurement interval of one week and a measurement frequency of 60 Hz. The data values are derived by a load curve which was recorded with a 12 kHz sampling frequency. All events with a power change greater than 30 W have a timestamp and are classified into the specific electrical devices. For the comparison of the different algorithms phase line A of the dataset is used and is averaged to three different sampling periods (1 second, 2 seconds and 5 seconds). Line B is not used due to the reason that it contained several events which are not timestamped and therefore could lead to misleading results.

All detected events which are within a distance of 2.5 times of the sampling period regarding a true event are handled as a true positive.

The input parameters of the event detection algorithms under investigation vary between specific ranges and can be seen in Table 4.1. The standard step size for active power values used in the parameter variation is 2 W. For all other parameters a step size of one single sample or one single second is used.

In the original data set only events with an active power change above 30 W are classified. Due to the reason that the sampling period of the original data set is averaged, the minimum power threshold for all event detection algorithms was set to 10 W instead of 30 W.

Table 4.1: Parameter ranges used for testing different event detectors

Parameter	Definition	Baranski Range	Step size
P_{min}	Minimum power change for event detection	10-60 Watts	2
τ	Time for merging events caused by transients	1-6 seconds	1

Parameter	Definition	GLR Range	Step size
w_{vote}	Width of voting window	$2 \cdot w_l + 1$ samples	1
w_l	Width of vector $\mathbf{P}_{Pre}$ as well as $\mathbf{P}_{Post}$	1-6 samples	1
v_{thr}	Minimum threshold for labelling events	1-w_{vote} samples	1

Parameter	Definition	Bergman Range	Step size
w	Width of window	2-10 samples	1
δ	Quantity of power averages below limit P_δ	1-6 samples	1
P_{thr}	Active power threshold for event detection	10-40 Watts	2
P_δ	Active power threshold for eliminating transients	3-30 Watts	2

4.1.3.2 Results

This section provides the best performances of all event detectors under all investigated metrics for phase line A of the BLUED data set [44].

Rate Metric

The rate metric results in a maximum number of true positives, however this is accompanied by a great number of false positives which can be seen in Figure 4.4. This can be explained, on one hand, by parameter sets which lead to a high number of true positives as these also possess very low active power thresholds. These low active power thresholds allow the event detection algorithms interpret active power disturbances with low amplitudes as false positive events. On the other hand the time duration τ of the Baranski detector as well as the voting threshold v_{thr} of the GLR detector are set to one to rise the number of true positive events. As a consequence, this allows the algorithms interpret turn-on transients, for example, as two different events instead of one single event. The only exception is the Bergman event detection algorithm, which has an additional active power threshold P_δ which limits the active power changes within the time duration δ before an event is said to have occurred. The algorithm of Bergman leads to a small number of false positives but also to a greater number of false negative events compared to Baranski and GLR. The best overall score is 0.54 from the GLR detector.

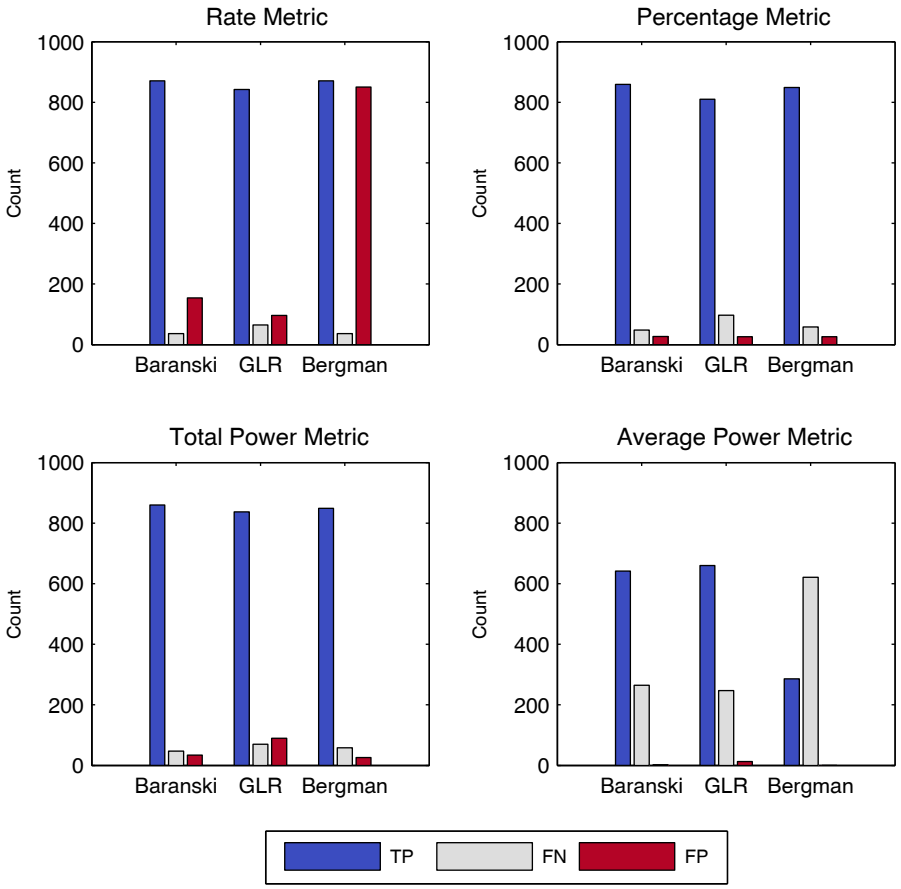

Figure 4.4: Comparison of best performances of event detectors under all metrics, sampling period: 1 second

Percentage Metric

The percentage metric balances the ratio of false positives and true positives regarding to the total events. Due to the reason that the distance function also incorporates the ratio between false positive to the total events, the number of false positives, compared to the rate metric, is reduced in this metric. The parameter sets which minimize this distance function have a greater active power threshold compared to the parameter set in the rate metric. This also leads to a slightly smaller number of true positives which is especially reflected by the event detection algorithms of Baranski and GLR (see Figure 4.4). Moreover, Bergman event's detection algorithm illustrates the same results which are caused by the equal parameter set compared to the rate metric.

In the percentage metric the number of false positives and false negative events reaches its minimum. The event detector of Baranski performs best with a score of 0.76.

Total Power Metric

In the total power metric the size of the power change of the false negative and false positive events are reflected in the distance function. This implies the assumption that greater power changes are more important than smaller ones. Hence, the optimal parameter set for this detector maximizes the number of detected events of events with greater power changes. For the load curve under investigation the parameter sets of the event detection algorithms are quite similar to the rate metric, see Figure 4.4. However, the results show that the absolute size of false negative and false positive events is best minimized with this parameter sets. Also in this metric Baranskis' event detector performs best with a score of 0.76.

Average Power Metric

In the average power metric the trade-off between the average power of misses and the average power of false positives is balanced out. The optimal detector has the minimum Euclidean distance between the average false positive and average false negative power changes and the point (0,0). As can be seen in Figure 4.4, a high number of misses best meet this objective function for the load curve under investigation. The reason for this is that most events of the load curve under investigations have a power change below 150 W. In contrast to this most of the power changes of the false positives are around several hundred Watts and are caused by turn-on transients. Thus a high number of false negative events leads to a closer distance to point (0,0):

Table 4.2 gives an overview of the best performances of the results of the score function (4.28) with the respective parameters (see also Appendix). The best results for the score function are achieved with Baranski's algorithm in the percentage and total power metric.
Different electrical devices have a different share to the total power consumption of an individual household. Moreover the simultaneous usage of electrical devices as well as the electrical equipment itself varies from household to household and plays a key role in parameter selection. The needs for detecting a high number of frequently used electrical devices which have small power changes (a few watts) differ from electrical appliances with major power changes. Furthermore the different NIALM algorithms have specific needs which also has to be taken into account for the selection of an appropriate metric and in series the parameter set.

Table 4.2: Score function and respective parameters for each metric under the different event detection algorithms for a sampling period of one second

	Metric	Parameters				Score
Baranski		P_{min}	τ			
	Rate	14	1			0.49
	Percentage	16	2			**0.76**
	Total Power	14	2			**0.76**
	Average Power	58	6			**0.40**
GLR		w_{vote}	w_l	v_{thr}		
	Rate	7	3	1		**0.54**
	Percentage	5	2	2		0.38
	Total Power	5	2	1		0.54
	Average Power	13	6	12		0.28
Bergman		w	P_{thr}	P_{δ}	δ	
	Rate	3	10	29	1	0.41
	Percentage	2	16	17	1	0.61
	Total Power	2	16	17	1	0.61
	Average Power	4	32	29	2	0.35

Variation of Time Point of Detected Event

Due to the reason that each event detector uses a different method to detect events, the indicated time points for the same event can vary under the different event detectors, especially in the case of a turn-on transient. As can be seen in Figure 4.5, the time points of events of the different event detectors differ from each other.

Baranski's event detector uses the first power value $P(t)$ where the power change $\Delta P(t)$ is bigger than a certain threshold as starting point, see (4.2) thus it is always in the forefront. In the case of the Bergman detector, changes over a certain threshold in the running average define an event and as a consequence the time point, see (4.6). Since the GLR algorithm uses two sliding windows which are separated by a single sample, its likelihood function (4.9) has its maximum when the probability of a power value $P(t)$ has its maximum probability in the post-event window. Due to the small window sizes of two and above, and sampling periods of several seconds and beyond, the probability of the post-event reaches its maximum after the power change has occurred; the time point of the GLR detector is usually in the area where the total power value has its steady state or more precisely right after the power fluctuation ends.

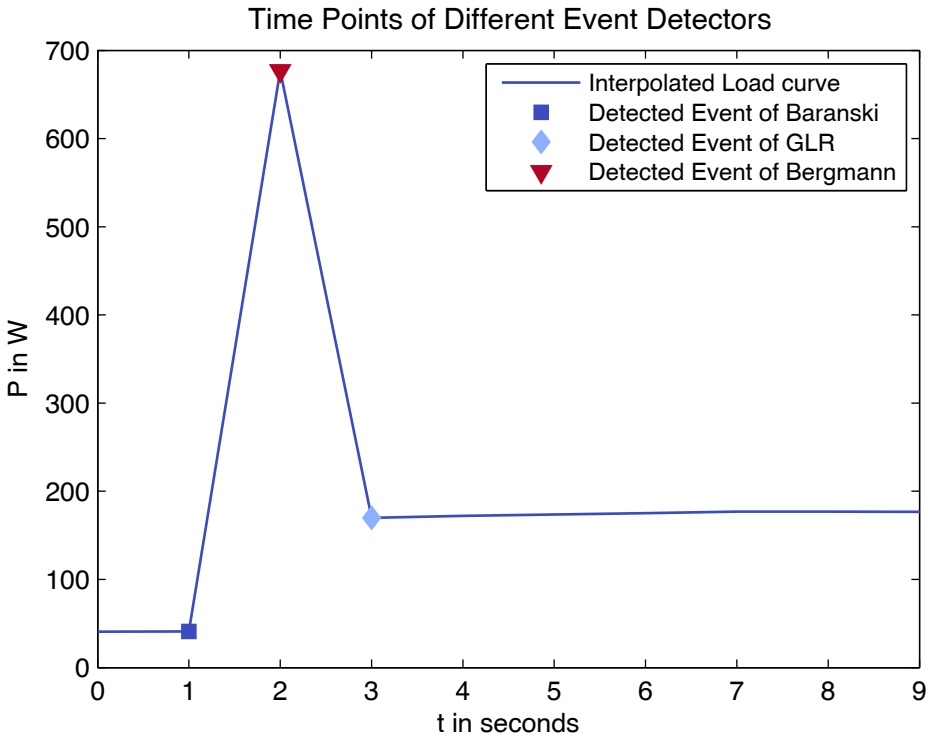

Figure 4.5: Variation of starting points from different event detectors, sampling period: 1 second, percentage metric, turn-on transient of a fridge

4.1.4 Influence of Sampling Period on Event Detection

In recent algorithms for NIALM, different sampling periods are used. Apart from measuring frequencies within the range of several kHz, low resolution power measurement data with a sampling period between one second and up to several minutes is mostly used for energy disaggregation. However, the greater the sampling periods, the more power events or state changes of electrical appliances overlap and as a consequence the single power changes are added up and can not be easily distinguished from each other. The different results of the investigated event detection algorithms according to variable sampling periods are hereby compared.

Looking at the performance of the detectors with a sampling period of five seconds in Figure 4.6 shows that a greater sampling rate raises the number of false negative and false positives events compared to a sampling period of one second (see Figure 4.4). These results are in line with the assumption that a greater sampling period reduces the number of detected true positive events.

Bergman's event detection algorithm has the highest number of false negative events in the average power metric. This is caused by the window width of 6 samples to satisfy the objective function.

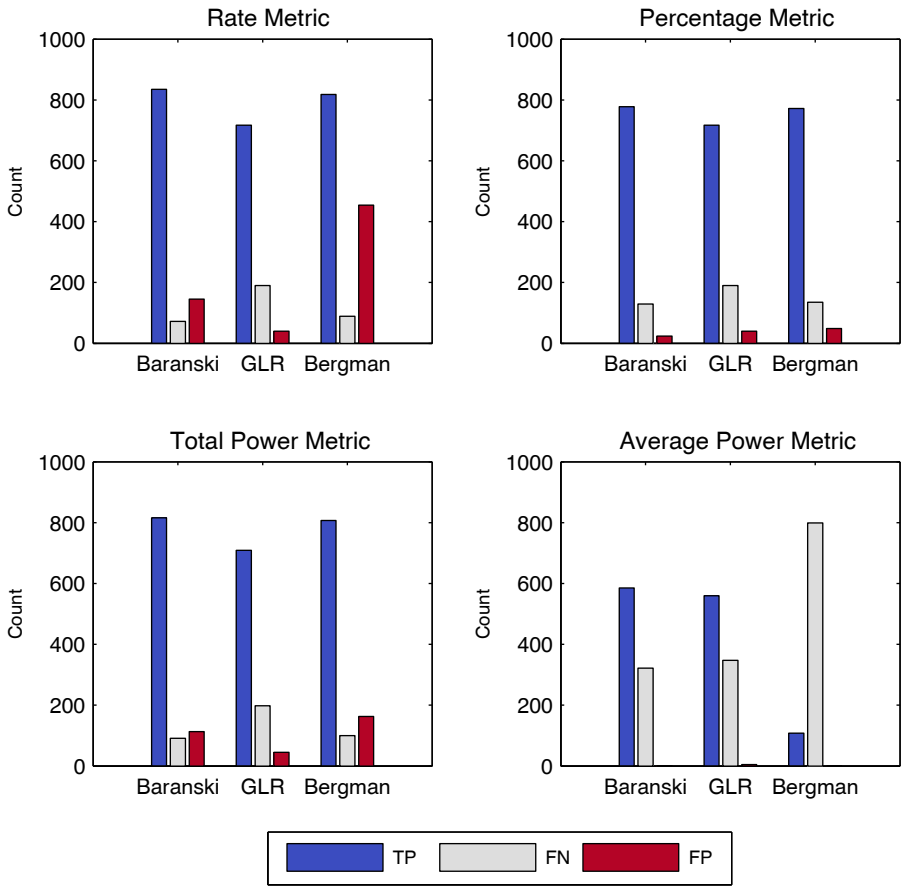

Figure 4.6: Comparison of best performances of event detectors under all metrics, sampling period: 5 seconds

Figure 4.7 shows the best total scores of the three investigated different event detection algorithms, with three distinct sampling periods. In order to be able to compare the results the best achieved overall score is taken.

As expected, the results are dependent on the sampling period and show that fewer events can be detected with greater sampling periods. The Baranski event detection algorithm has the best overall total score with a sampling period of one second. The same trend can be seen when the sampling period is reduced to two seconds. A reduction of the sampling period to five seconds brings the scores of all detectors in a range of 0.4 to 0.5 with the same positioning as described previously.

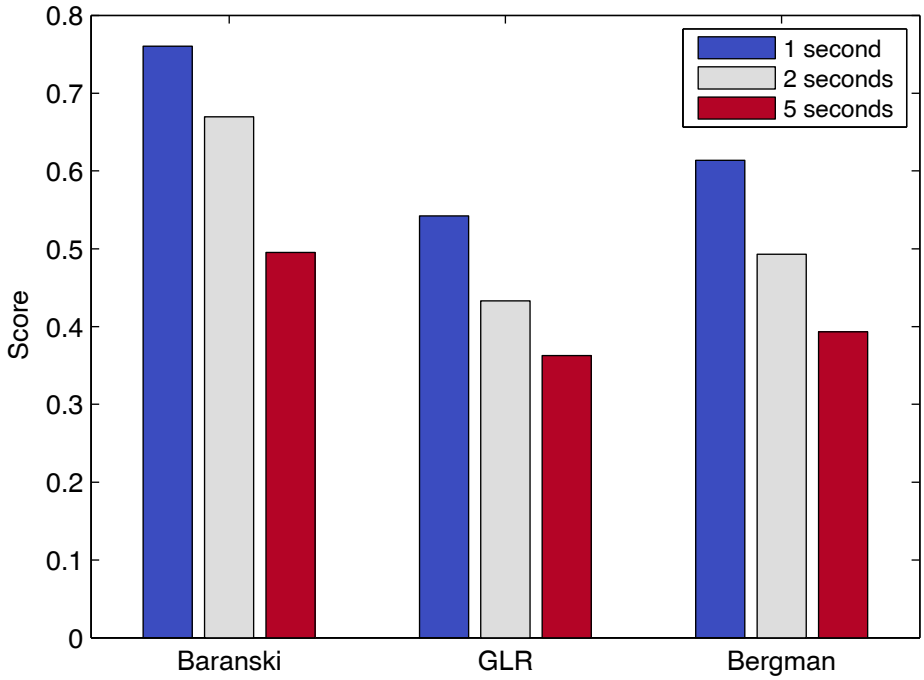

Figure 4.7: Best total score of different event detection algorithms according to varying sampling periods (1, 2 and 5 seconds)

The best overall score is achieved by the percentage and the total power metric detector (see Table 4.3). The trend shows that with greater sampling periods the differences in the scores decrease, see Figure 4.7.

Table 4.3: Score function for each metric under the different event detection algorithms for a sampling period of five seconds

	Baranski	GLR	Bergman
Rate	**0.45**	0.36	0.30
Percentage	**0.43**	0.36	0.37
Total Power	**0.50**	0.34	0.39
Average Power	**0.39**	0.27	0.31

4.1.5 Incorporation of Reactive Power Values

Besides active power values, reactive power values are used in event detection algorithms (see Section 2.3.3). A combination of active and reactive power values should lead to more accurate results since there are two different sources to separate events from the load curve. However, due to the fact that not all electrical devices have a significant reactive power consumption, the overall event detection performance can be raised only at certain devices.

Performing event detections on both data sets (active and reactive power values) with different parameters results in to two different sets of events. These two sets of events have to be merged to a single data set of events.
If for example the objective function is to maximize the number of true positives, both event data sets should be merged in such a way that all unique events from each event data set as well as events that occur in both data sets are combined (logic OR function). Thus the additional information from the reactive power source could reveal events that overlap in the active power values. But even if the number of true positives can be raised it also leads to a significant increase of false positives.

Another approach for the objective function is to minimize false positive events. This can be achieved by combining both event data sets so that just events that exist in both event data sets (doubles) remain in the overall event data set while all other unique events are removed (logic AND function). In the test setup this approach has been selected in order to illustrate how many events can be distinguished by both, reactive and active power. The input data of the active and reactive power load curves has a sampling period of one second.
To find the optimal parameter set for the event detectors in the active and reactive load curve for each load curve, k parameter sets for an event detection algorithm are defined according to Table 4.1. This results in k different event data sets. All the computed sets of events k of the reactive power values have to be combined with all the computed sets of events k of the reactive power values to find the optimal parameter sets which satisfy the objective function. Due to the reason that a combination of all these events leads to k^2 computations and the number of sets of events are above 1.000, a Monte Carlo simulation with uniformed input parameter distribution was carried out to calculate the metrics to reduce computing time and for memory saving issues. As already described for each computation of the metrics, the two different sets of events of the active and reactive power values are merged in such a way that just the events remain which are available in both (active and reactive power values) sets of events. All other events are removed from the computed data set.

As can be seen in Figure 4.8, the number of false positives is as expected to be rather low when compared to the event detection algorithms which do not incorporate reactive power values (see Figure 4.4). However, the cost of the reduced false positives are based on a high number of false negative events. Baranski's event detection algorithm has the highest number of true positives and the lowest number of false negative events in all tested metrics. The result was achieved with two different parameter sets. In comparison to the parameters in the rate metric of Baranski's detector in Section 4.1.3, the parameters for detection of the active power values such as P_{min} almost doubled with the use of this approach (see Appendix), while P_{min} has the smallest possible threshold for detection of the events of the reactive power values (see Appendix).

This underlines the assumption that due to the minor reactive power consumption of certain electrical devices a lower threshold is required to be able to detect events. This trend can also be observed for the Bergman detector.

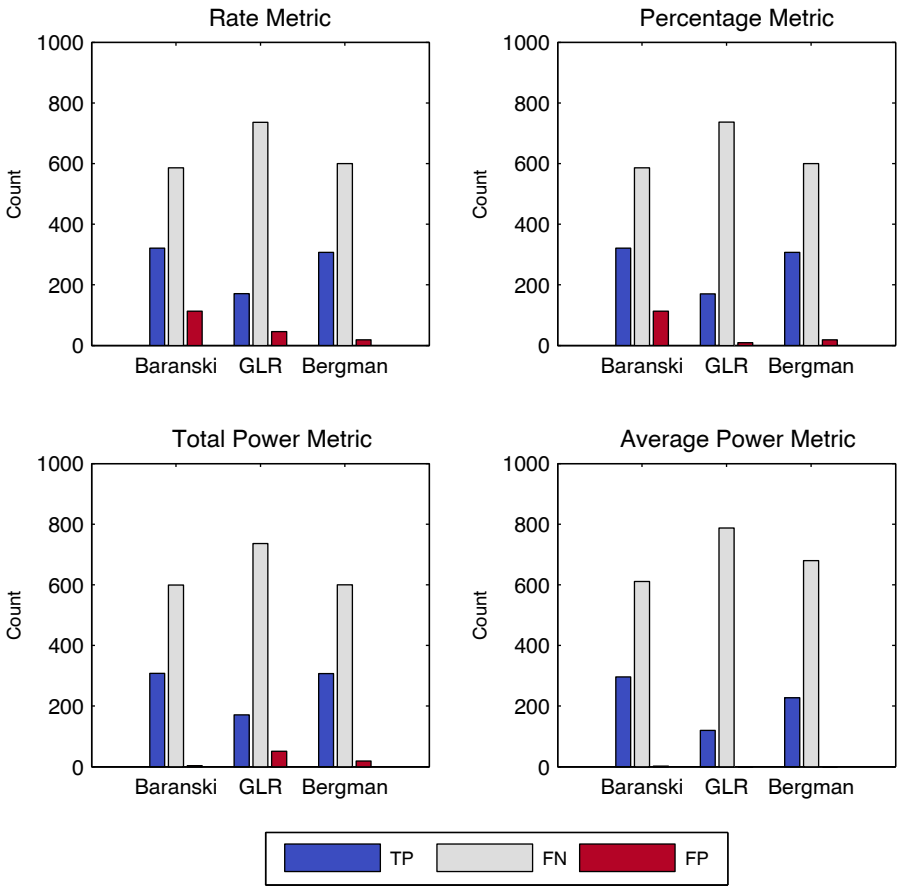

Figure 4.8: Comparison of best performances of event detectors under all metrics when active and reactive power events (doubles) are combined, sampling period: 1 second

Nonetheless, even when the overall performance is lower compared to the event detection with active power values, the false positives can be reduced. Of vital importance is the fact that in approximately one third of all events the use of the reactive power values allows for an improved assignment of events to their corresponding electrical devices.

4.2 Steady State Power Changes

This section describes different methods for the computation of steady state power changes. Beyond this an advanced version for realising this task is proposed.

4.2.1 General Overview

After the detection of events, the correlated steady state power changes of the real and reactive power are often used to distinguish between events of different electrical appliances [28]. By the term "steady state power changes" the power difference between the steady operating states of a device is defined. This means that, for example, turn-on transients of a refrigerator are masked for computing the power change from the steady off-state to the steady on-state.
Due to power fluctuations of certain electrical devices as well as the likelihood that on- and off-events from other devices overlap, the task of computing the actual steady state power change of a certain event caused by one device is difficult to ascertain. Furthermore an increase in the sampling period causes overlaps of events and inaccurate results for the actual steady state power changes. Moreover the power changes caused by turn-on transients (see Section 3.1.1) have to be taken into account for the computation of the steady state power changes. The electrical energy of the turn-on transients can also be used as additional feature when it is extracted from the power data. An accurate computation of the steady state power changes leads to an improved distinction between events caused by different electrical appliances.

4.2.1.1 Computation via Pre-Event and Post-Event Window

Standard Version (State of the Art)

One of the most common methods to compute the steady state power changes of events is to compute the average of a pre- and post-event window and to subtract the mean values [45]. This procedure can be further improved by introduction of a third window which is used to mask transients and improves the overall accuracy, see (4.21). Since the parameters for the window lengths are fixed, the individual durations of, for example, turn-on transients cannot be considered.
While this approach works well with sampling periods of several milliseconds, it is only partially applicable to sampling periods of one second or above. The reason for this is the overlap of events, the overall fluctuations of power changes of turned on devices as well as the variation of the time point of a transient where the actual event is assigned. The sampling period of one second or more leads to a variation of the offset between the start time of the sampling window and the turn-on time of a certain electrical appliance. As a consequence the exact time point, where an event detection algorithm assigns an event to a power value, varies. When, for example, a turn-on transient lasts several samples (see also Figure 4.9), the detected event could be assigned between the very first

power value of the turn-on transient and the very last value of the transient. To achieve more accurate results a new method with variable window length is introduced.

Advanced Version

In this advanced version a novel method for calculating the steady state power changes is introduced. A pre-event and a post-event window, both with variable length and position, are used to calculate the mean value before and after the detected event to provide the steady state power change.
A demonstration of the results of the advanced version can be seen in Figure 4.9 where a turn-off of a device is followed by a turn-on. As illustrated, the pre-event and post-event windows which initially have a length of 6 samples are truncated to a length of 5 and 3 samples. The remaining power values within the windows provide a more accurate estimation of the steady state power change.

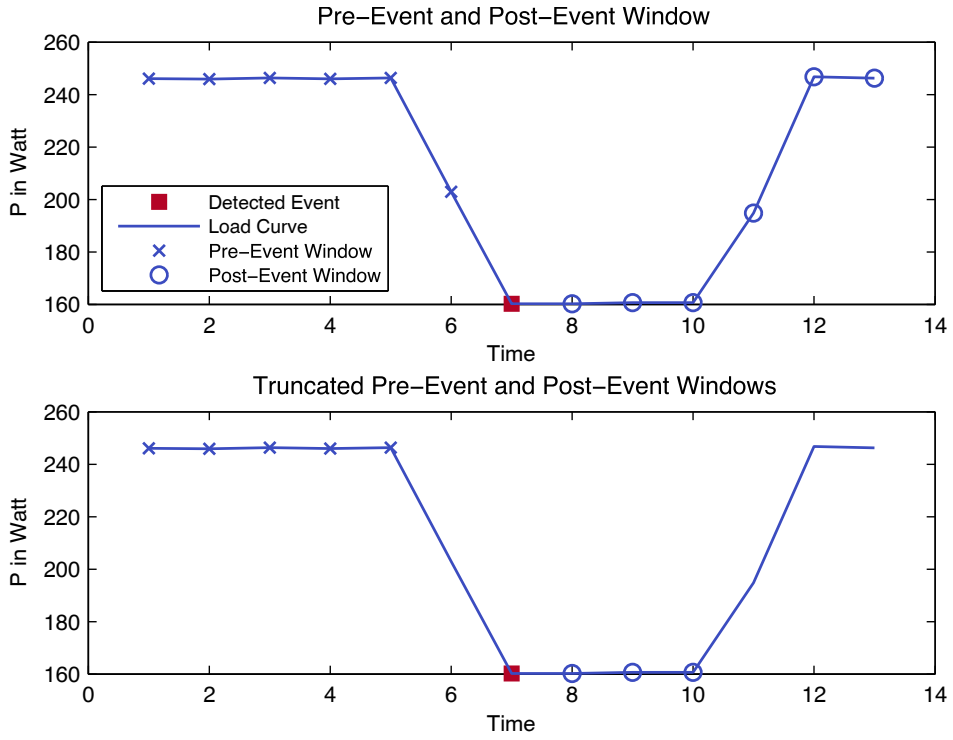

Figure 4.9: Pre- and post-event windows before (above) and after truncation of irrelevant power values, event detected with GLR detector

The novel method for computing the steady state power changes works as follows:
In a first step a pre-event window with a maximum size w_{Pre} and a post-event window with the maximum size w_{Post} are aligned before and after the detected event E_i which is gathered by Baranski's algorithm. The power values within the pre-event window are denoted with $P_{Pre}(t)$ and in the post-event window with $P_{Post}(t)$:

$$\mathbf{P}_{Pre} = \{\mathrm{P}\left(t(E_i) - w_{Pre} - 1\right), ..., \mathrm{P}\left(t(E_i) - 1\right)\} \tag{4.29}$$

$$\mathbf{P}_{Post} = \{\mathrm{P}\left(t(E_i) + 1\right), ..., \mathrm{P}\left(t(E_i) + w_{Post}\right)\} \tag{4.30}$$

$$\mathbf{P}_{Total} = \{\mathrm{P}\left(t(E_i) - w_{Pre} - 1\right), ..., \mathrm{P}\left(t(E_i) + w_{Post}\right)\} \tag{4.31}$$

To take account of the variable run length of turn-on transients as well as power fluctuations which occur within the pre- and post-event window, a power change detector is applied in the two mentioned windows. Power changes $\Delta P(t)$ which are within the pre- or post-event window and above a certain threshold ΔP_{Ch} are used as a flag to show that certain areas of the windows have to be filtered out due to factors such as irrelevant power fluctuations of other devices and turn-on transients.
For calculating the maximum threshold ΔP_{Ch} which is used for the detection of areas with relevant power changes the difference values $\Delta P_{Total}(t)$ of $\mathbf{P}_{Total}$ are used:

$$\Delta P_{Total}(t) = P_{Total}(t) - P_{Total}(t-1) \tag{4.32}$$

The maximum threshold ΔP_{Ch} can be calculated with the factor c where tests have shown that $c = 0.2$ results in a good performance:

$$\Delta P_{Ch} = c \cdot \max\left|\Delta P_{Total}(t)\right| \tag{4.33}$$

For calculating the mean value[2] of the pre-event window $\mu(\mathbf{P}_{Pre})$ one can distinguish between a complete pre-event window $\mathbf{P}_{Pre} = \{P_{Pre}(1),..., P_{Pre}(n_{Pre})\}$ and a truncated window $\mathbf{P}'_{Pre} = \{P_{Pre}(t),..., P_{Pre}(t+n'_{Pre})\}$ with $n_{Pre} > n'_{Pre}$:

$$\mu\left(\mathbf{P}_{Pre}\right) = \begin{cases} 1/n_{Pre} \cdot \sum_{t=1}^{n_{Pre}} P_{Pre}(t) & if \max\left|\Delta P_{Pre}(t)\right| <= \Delta P_{Ch} \\ 1/n'_{Pre} \cdot \sum_{t}^{t+n'_{Pre}} P'_{Pre}(t) & if \max\left|\Delta P_{Pre}(t)\right| > \Delta P_{Ch} \end{cases} \tag{4.34}$$

If a truncated pre-event window $\mathbf{P}'_{Pre}$ has to be used for calculating the mean value $\mu(\mathbf{P}_{Pre})$ in (4.34), a truncation of the complete pre-event window $\mathbf{P}_{Pre}$ has to be carried out. For this task each element $\Delta P_{Pre}(t)$ within the mentioned window which is greater than the threshold ΔP_{Rel} is considered to be irrelevant. For computing the threshold the factor r = 1.2 is used:

$$\Delta P_{\mathrm{Rel}} = r \min\left|\Delta P_{Pre}(t)\right| \tag{4.35}$$

---

[2] Note: For easier notation in the continuing paragraphs the pre-event window is used for a demonstration of the algorithm. However, the same mathematical formulations are also applied in the post-event window.

For each element within the pre-event window a logic value $P_{Pre,i}(t)$ is calculated which indicates – when it is true – that the power change is irrelevant for the truncated window $\mathbf{P}'_{Pre}$:

$$P_{Pre,i}(t) = \begin{cases} true & if\ \Delta P_{Pre}(t) > \Delta P_{Rel} \\ false & if\ \Delta P_{Pre}(t) <= \Delta P_{Rel} \end{cases} \tag{4.36}$$

If several irrelevant power changes are detected, there can also be several consecutive relevant power changes. These consecutive relevant power changes can be clustered in different sets $\mathbf{S}_j$, where $j = 1,...,n_s$ denotes the index of the individual sets $\mathbf{S}_j$. The formulation for assigning a (relevant) power value from the pre-event window $P_{Pre}(t)$ to an individual set $\mathbf{S}_j$ is as follows:

$$P_{Pre}(1) \mapsto \mathbf{S}_j \quad if \quad P_{Pre,i}(1) = false \tag{4.37}$$

$$P_{Pre}(t) \mapsto \begin{cases} \mathbf{S}_j & if\ \left(P_{Pre,i}(t) = false \wedge P_{Pre,i}(t-1) = false\right) \vee \\ & \quad \left(P_{Pre,i}(t) = true \wedge P_{Pre,i}(t-1) = false\right) \qquad t = 2,...,\mathrm{n}_{Pre}-1 \\ \mathbf{S}_{j=j+1} & if\ \ P_{Pre,i}(t) = false \wedge P_{Pre,i}(t-1) = true \end{cases} \tag{4.38}$$

$$P_{Pre}(n_{Pre}) \mapsto \begin{cases} \mathbf{S}_j & if\ P_{Pre,i}(t-1) = false \wedge P_{Pre,i}(t-2) = false \\ \mathbf{S}_{j=j+1} & if\ P_{Pre,i}(t-1) = false \wedge P_{Pre,i}(t-2) = true \end{cases} \tag{4.39}$$

Consequently n_s sets are obtained, each containing an individual number k_j of power values from the pre-event window $\mathbf{S}_j = \{P_{Pre}(t_j),...,P_{Pre}(t_j+k_j)\}$. From all of the obtained sets, just one is selected for calculating the mean value of the pre-event window in (4.34).In other words: one single set $\mathbf{S}_z$ of power values is selected and assigned to be the truncated window $\mathbf{P}'_{Pre}$.
For selection of the appropriate set, the ratio between the mean value $\mu(\mathbf{P}_{Post})$ of the complete post-event window and the mean value $\mu(\mathbf{S}_j)$ of the sets $\mathbf{S}_j$ is maximized:

$$\mu\left(\mathbf{S}_j\right) = 1/k_j \sum_{t=t_j}^{t_j+k_j} P_{Pre}(t) \tag{4.40}$$

$$z = \arg\max_{1 \le j \le n_s} \left[\max\left(\mu(\mathbf{S}_j), \mu(\mathbf{P}_{Post})\right) / \min\left(\mu(\mathbf{S}_j), \mu(\mathbf{P}_{Post})\right) \right] \tag{4.41}$$

The truncated window $\mathbf{P}'_{Pre}$ in (4.34), for calculation of the mean value of the pre-event window $\mu(\mathbf{P}_{Pre})$, equals $\mathbf{S}_z$:

$$\mathbf{P}'_{Pre} = \mathbf{S}_z \tag{4.42}$$

The steady state power change of an event $\Delta P(E)$ can be calculated by subtracting the mean value of the post-event $\mu(\mathbf{P}_{Post})$ value and the pre-event value $\mu(\mathbf{P}_{Pre})$:

$$\Delta P(E) = \mu\left(\mathbf{P}_{Post}\right) - \mu\left(\mathbf{P}_{Pre}\right) \tag{4.43}$$

4.2.1.2 Filtering of the Load Curve

Another commonly used method for computing the steady state values of events with a sampling period of one second or above, is to apply a simple median filter to the input power data as a first step [84]. As a second step a pre- and a post-event window are used for the computation of the steady state power changes.

4.2.1.3 Expert Heuristics

An expert heuristic is introduced by Baranski [42] for fulfilling the task of steady state computation (Section 4.1.1). The main advantage of this heuristic is that it merges consecutive power changes which can be caused by a turn-on transient. However, other event detection algorithms interpret such turn-on transients as two different events and in a row the steady state power changes do not reflect the actual consumption of the device.

4.2.2 Metric for Steady State Power Change

In order to compare the performance of the different steady state power change algorithms a simple metric is introduced which describes the percentage deviation from the true steady state power change $\Delta P_{True}(E_i)$. A steady state power change algorithm is denoted with λ and its parameter set with Λ. The metric M is computed by the average deviation of the estimated steady state power change $\Delta P_{Est}(E_i)$ to the true steady state power change $\Delta P_{True}(E_i)$ over the total number of events n_E:

$$M = \min_{\lambda \in \Lambda} \frac{1}{n_E} \sum_{i=1}^{i=n_E} \left| 1 - \frac{\Delta P_{Est}(E_i)}{\Delta P_{True}(E_i)} \right| \cdot 100\% \tag{4.44}$$

The parameter set of the best algorithm $\hat{\lambda}$ can be denoted with:

$$\hat{\lambda} = \arg\min_{\lambda \in \Lambda} \frac{1}{n_E} \sum_{i=1}^{i=n_E} \left| 1 - \frac{\Delta P_{Est}(E_i)}{\Delta P_{True}(E_i)} \right| \tag{4.45}$$

4.2.3 Comparison of Steady State Power Change Algorithms

Since the accuracy of the steady state power changes influences the performance of the energy disaggregation, an analysis of the different methods is carried out in this section.

4.2.3.1 Test Setup

For conducting the comparative analysis a set of events (gathered by the expert heuristic of Baranski) with the optimal parameters for the total power metric as per Section 4.1.3, is used for all tested algorithms. The dataset which was used for the event-detection was the same mentioned in Section 4.1.3.1 which comprised of a sampling period of one second.

Three different algorithm standards are compared against each other: the standard and advanced versions of the pre- and post-event window algorithm and also Baranski's expert heuristic.
As an input for the two pre- and post-event window algorithms the same set of events (gathered by the expert heuristic of Baranski) is used but there are two different versions of the load curve provided for the computation of the steady state power changes. Whereas the “raw load curve” equals the raw data set, the “filtered load curve” is a median filtered version. The reason for testing two different load curves is that in some NIALM approaches filtered load curves are used for computing steady state power changes and therefore a comparison is realized. The filtering is realized by a median filter with a window width of 5 samples, the turn-on transients as well as minor power fluctuations are removed from the original load curve.
Besides this the lengths of the windows of the different algorithms vary between specific ranges: refer to Table 4.4 for more detail. The standard step size for all parameters is one single sample.
For the evaluation of the results the metric introduced in Section 4.2.2 is used.

Table 4.4: Parameter ranges used for testing different steady state power change algorithms

Parameter	Standard Version	Step size
	Range	
w_1	1-6 samples	1
w_2	1-4 samples	1
w_3	1-6 samples	1

Parameter	Advanced Version	Step size
	Range	
w_{Pre}	1-6 samples	1
w_{Post}	1-6 samples	1

Parameter	Baranski	Step size
	Range	
P_{min}	14 Watts	-
τ	2 samples	-

4.2.3.2 Results

Due to the individual truncation of the pre- and post-event windows in the advanced version, a more accurate computation of the steady state power change is realized, as per Figure 4.10. The advanced version of the algorithm reduces the average deviation of the steady state power changes to the true values of about 1% compared to the standard version from Anderson et al. [45]. By using the filtered input data, both mentioned algorithms improve the performance of up to 0.5%, overall a minimum of 4% can be achieved.

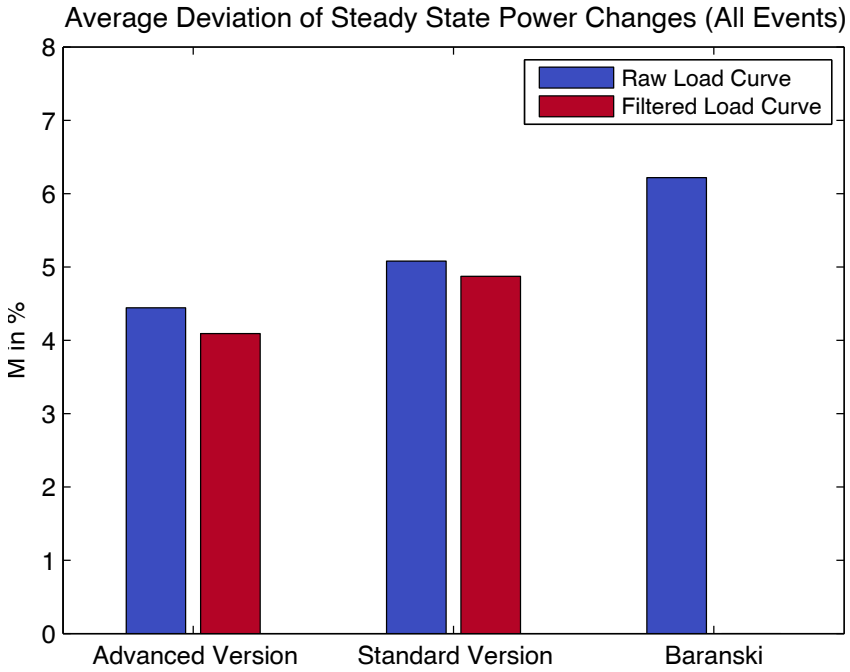

Figure 4.10: Comparison of the performance by the computation of steady state power changes with different algorithms, sampling period: 1 second

The optimal parameters of the advanced and the standard version algorithm can be seen in Table 4.5. Whereas the optimal parameters for the standard version are in between 1 and 3 samples, the advanced version has window lengths of around 4 to 6 samples. Due to the reason that in the extended version the windows can be individually truncated to a minimum size of 2 samples, longer window lengths are required. Interestingly, the optimal post-event window w_3 of the standard version has just a minimum length of 1 sample. This shows that the first value in the post-event window often reflects a good starting point to compute the steady state power change out from events detected by the Baranski algorithm. By using the filtered load curve as an input, a reduction of the window lengths can be achieved in both algorithms. Also the eliminating window w_2 is reduced to a minimum of 1 sample due to the removal of transients.

Table 4.5: Optimal parameter results for standard and advanced version algorithm

Standard Version	Optimal Parameters	
	Raw Load Curve	Filtered Load Curve
w_1	3 samples	3 samples
w_2	2 samples	1 samples
w_3	1 sample	1 sample

Advanced Version	Optimal Parameters	
	Raw Load Curve	Filtered Load Curve
w_{Pre}	4 samples	4 samples
w_{Post}	6 samples	4 samples

Baranski's algorithm has an average deviation of about 6% according to the true values of the steady state power changes. However, the filtered load curve has not been tested with this algorithm because this would also result in a changed set of events which would no longer be able to provide comparative data.

As energy disaggregation can only be carried out for a single electrical device, the performance of the results for the computation of the steady state power changes of an fridge are compared against each other in Figure 4.11.

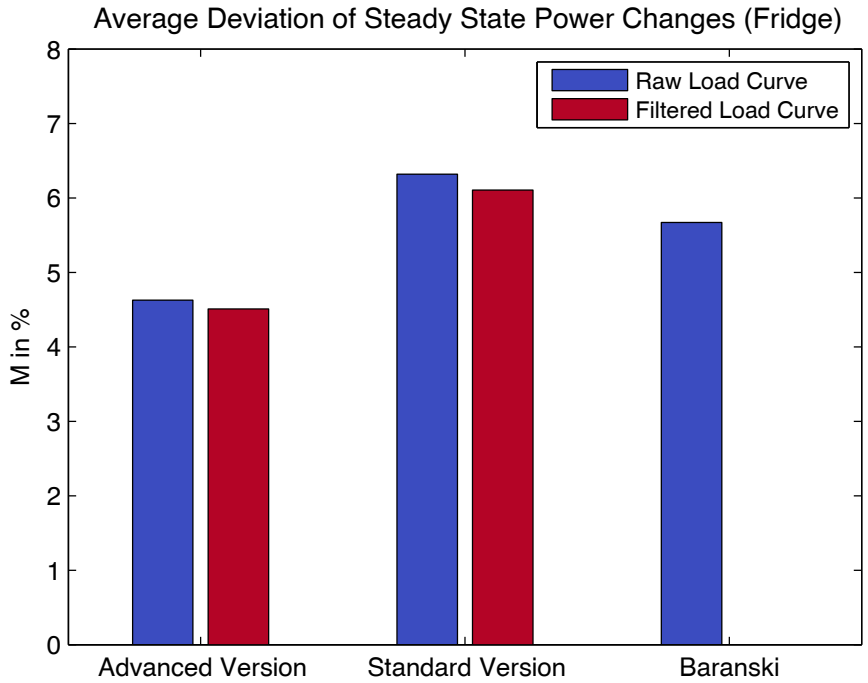

Figure 4.11: Comparison of the performance by the computation of steady state power changes with different algorithms and just a single event category (fridge), sampling period: 1 second

While the algorithms of the advanced version and of Baranski remain in a similar range, it can be seen that the average deviation of the standard version reaches more than 6%. This shows that the average deviation of the standard version can vary in a greater range in special cases, when for example just the events of a fridge are investigated. Especially the fridge has a turn-on transient which lasts for a few seconds only and can vary due to the varying offset between the sampling period and the events. Therefore the advanced version can better compensate this time variation and leads in a row to more accurate results.

4.2.4 Influence of Sampling Period on Advanced Algorithm

The sampling period represents another worthwhile parameter which influences the steady state power change computation. While the sampling period also influences the total of amount of events which can be detected, it also plays a significant role in the steady state power change calculation. The greater the sampling period the greater the probability that more overlaps occur within a certain sample step.
Whereas the influence of the different event detectors on the number of true or false positive has already been investigated in Section 4.1, this section illustrates the resulted events from each detector which are used as an input for calculating the steady state power changes. This goal is achieved by using the advanced version. Since a filtering of the input data increases results, the filtered load curve was used for steady state computation. Besides this the parameters of the advanced algorithm (w_{Pre} and w_{Post}) are varied in the same ranges as described in Section 4.2.3.1.

The results of the analysis are shown in Figure 4.12. As can be seen, for each set of events the average deviation of the steady state power change increases with each sampling period. Due to the reason that each detector uses different sets of events for computation, the results are not directly comparable to each other. However, an increase of the sampling period is accompanied by a bigger computational error in the steady state power change. This has a remarkable influence on the assignment of events of electrical appliances as it increases the error.

Table 4.6 shows the optimal parameters of the advanced version algorithm which lead to a minimum deviation M under different sampling periods. The window length remains predominantly in the same ranges for each detector.
The reason that the post-event window size w_{Post} for the GLR or Bergman algorithm equals one is the fact that a filtered load curve is used for computing the steady state power changes next to the fact that the time point of the events from the GLR and Bergman algorithm is close to the end or at the end of the power change or power fluctuation (see Section 4.1.3.2).

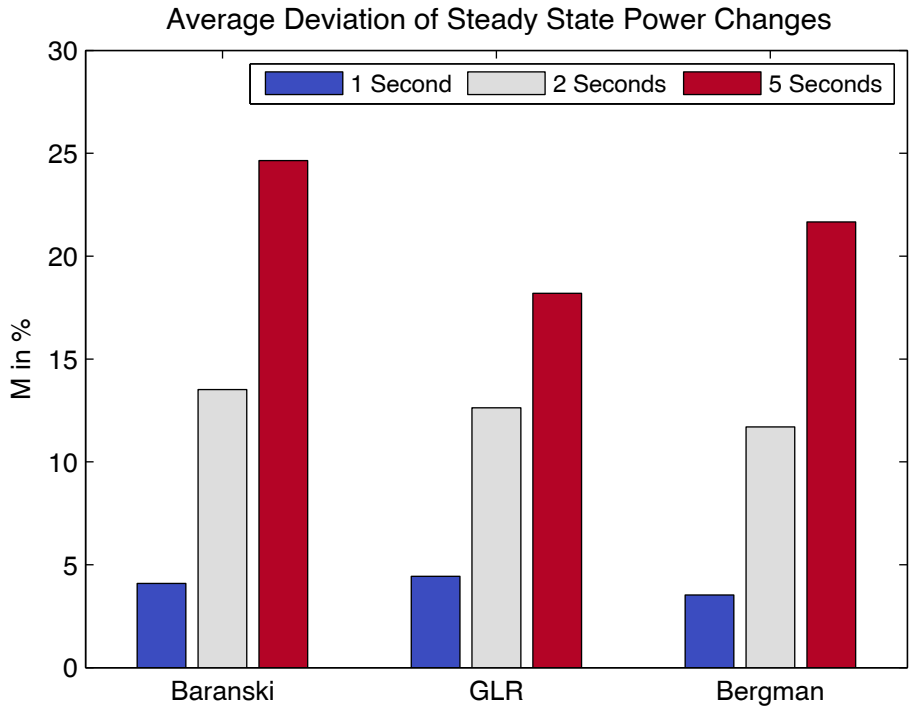

Figure 4.12: Average difference of steady state power changes with different sets of events computed with advanced version compared to true values

Table 4.6: Optimal length of pre- and post-event windows under different sampling periods and event detectors

Detector	1 Second		2 Seconds		5 Seconds	
	w_{Pre}	w_{Post}	w_{Pre}	w_{Post}	w_{Pre}	w_{Post}
Baranski	4 samples	4 samples	3 samples	4 samples	2 sample	4 samples
GLR	6 samples	1 sample	5 samples	1 sample	6 samples	1 sample
Bergman	5 samples	1 samples	5 samples	1 sample	5 samples	1 sample

5 Novel Method for Energy Disaggregation

This chapter describes the proposed novel method for Residential Energy Disaggregation Based on Appliance-Specific Characteristics (fEEDBACk). First of all, fundamentals to hidden Markov models are presented. Secondly, a general description to the algorithm and the modelling of electrical appliances on hidden semi-Markov models is given. Thirdly, a more detailed description of the single steps of the proposed algorithm is shown in the subsequent section of this chapter.

5.1 General Description

This section provides an overview of the fundamentals to hidden Markov models and a comparison of already proposed NIALM algorithms based on hidden Markov models. Beyond this, the general concept of the novel energy disaggregation method is provided.

5.1.1 Fundamentals

In order to provide some background information, this section presents the main fundamentals needed in the in the following sections.

5.1.1.1 Markov Model

A Markov model is generally based on a Markov chain, which is a stochastic process. The Markov chain is defined by the property that the future process just depends on the present state and not on the total history of the states of the process. As can be seen in Figure 5.1, a Markov chain has a finite state space and describes a sequence through which a process moves. The consecutive states of the Markov chain can be denoted with $\{S_t\}$ and the state space with $S = \{1,...,J\}$. This chain is described by:

- Initial probabilities $\pi_j = P(S_0 = j)$ where $\Sigma_j \pi_j = 1$
- State transition probabilities $p_{ij} = P(S_{t+1} = j \mid S_{t+1} \neq i, S_t = i)$ where $\Sigma_j p_{ij} = 1$ and $i \neq j$.

The initial distribution π_j specifies the start probability for each state. The single transition probabilities p_{ij} characterize the probability for the transition from one state to another. It can be distinguished between different orders of the Markov process where a first-order model has a memory size of one, in other words the likelihood of the current state just relies on the previous state of the model. [89]-[92]

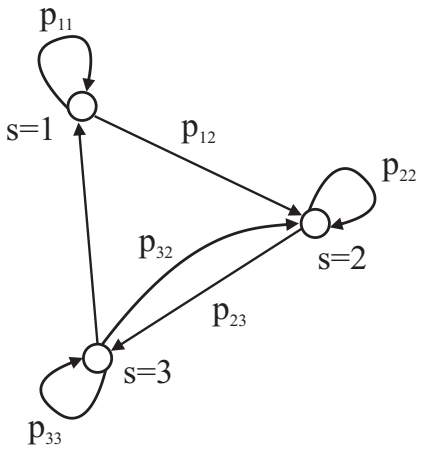

Figure 5.1: A Markov chain with three states *s* and selected state transitions

For a Markov model each state of the Markov chain emits an observation $X_t = y$ for all time steps t and is denoted by

- Emission probability $b_j(y) = P(X_t = y \mid S_t = j)$ where $\Sigma_j\, b_j(y) = 1$.

The observations y from the output process can either have discrete-valued symbols or a continuous observation density. Furthermore, univariate or multivariate distributions such as Gaussian or Gamma distribution can be modelled. Usually a Gaussian distribution is used for the observable process.

5.1.1.2 Hidden Markov Models

In a hidden Markov model, it is assumed that the latent variables $\{S_t\}$ from the stochastic 'state' process are related to the stochastic observable output signal $\{X_t\}$, as per Figure 5.2. The HMM is described by the already mentioned initial probabilities π_j, the state transition probabilities p_{ij} as well as the emission probability $b_j(y)$.

Finding the best model parameters for a HMM is usually referred as *learning*. There exist several methods for learning the model parameters such as the Baum-Welch method (expectation maximization) which is the state of the art method in terms of maximum likelihood estimation [89].

For finding the most likely sequence of latent states in the hidden Markov model different *inference* methods exist. This sequence can, for example, be obtained by the Viterbi algorithm which is a dynamic programming method. For an exact inference of the HMM by the Viterbi algorithm the complexity is $O(J^2T)$ where T denotes the number of total samples [89]. There also exist several approximate inference methods such as Gibbs sampling and variational Bayes. These methods have the advantage that they reduce the computational complexity in comparison to the exact inference. Gibbs sampling, for example, is also very simple to implement but in contrast often requires a large number of iterations for achieving the actual state sequence. [91]

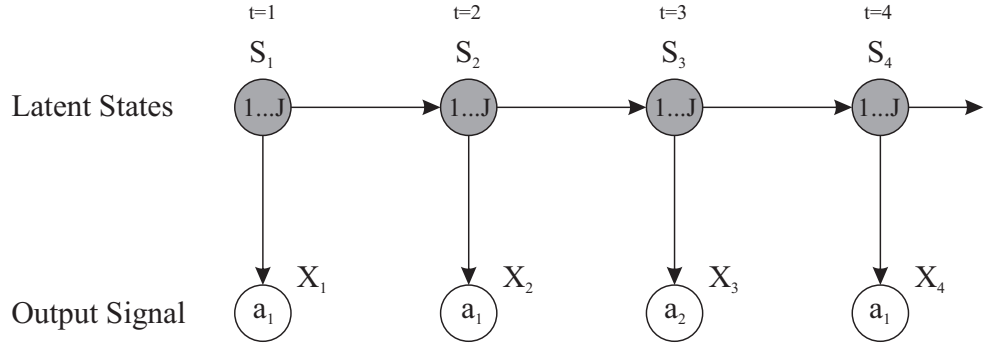

Figure 5.2: First-order hidden Markov model with latent states S_t and observable variables X_t which emit two different symbols a_1 and a_2

A major drawback of standard (hidden) Markov models is that the state durations cannot be modelled. The time duration is geometrically distributed and the durational probability $d_i(u)$ for the state i, a time duration of u (integer-valued) and the transition probability p_{ii} to remain in the state i can be written as [89]:

$$d_i(u) = \left(p_{ii}\right)^{u-1}\left(1-p_{ii}\right) \tag{5.1}$$

Hidden semi-Markov models (HSMM) are introduced by [90] to incorporate more realistic and natural time durational distributions .

5.1.1.3 Hidden Semi-Markov Models

The general description of HSMM in this section is given in [92] and [93].
As already mentioned, the main advantage of HSMMs is an explicit modelling of the duration, also called sojourn time, for each state. While in the hidden Markov model a single observation is related to a specific state, in the HSMM a sequence of observations can be related to one state, compare Figure 5.2 and Figure 5.3.

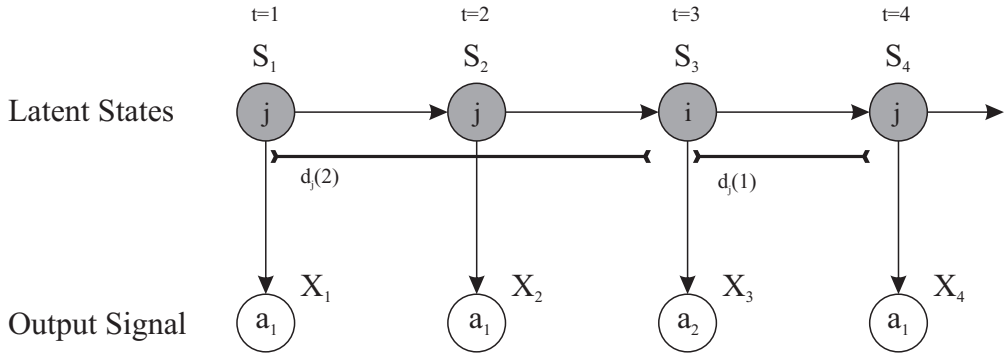

Figure 5.3: First-order hidden semi-Markov model with latent states S_t, observable variables X_t which emit two different symbols a_1 and a_2 and state durations $d_j(u)$

The state duration of a specific state j is an integer-valued, random variable $u_j = \{1,2,...,M_j\}$. This J-state hidden Markov model can be defined similarly to the HMM by the following parameters:

- Initial probabilities $\pi_j = P(S_0 = j)$ where $\Sigma_j \pi_j = 1$
- State transition probabilities $p_{ij} = P(S_{t+1} = j \mid S_{t+1} \neq i, S_t = i)$ where $\Sigma_{j \neq i} p_{ij} = 1$, $p_{ii} = 0$ and i≠j
- State duration distribution $d_j(u) = P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = 0,..., u - 1 \mid S_{t+1} = j, S_t = i)$ where $u = 1...M_j$ and i:={1...J} \ j
- Emission probability $b_j(y) = P(X_t = y \mid S_t = j)$ where $\Sigma_j b_j(y) = 1$

At the start of the Markov chain, the initial probability π_j describes the probability to start with state j. After entering a specific state i, the transition probability p_{ij} represents how likely it is that the next state is going to be state j. As opposed to regular hidden Markov models, the state transition to the state itself p_{ii} is set to zero in HSMM. Instead of using p_{ii} the state duration distribution $d_j(u)$ facilitate the likelihood for spending a certain time duration u in the specific state j. The state duration is independent of the previous state i but can also be modelled as a conditional distribution which depends on the previous state.

As already mentioned, the emission probability $b_j(y)$ represents how the output process $\{X_t\}$ is related to the semi-Markov chain $\{S_t\}$.

The state duration distribution can be modelled as a non-parametric or parametric distribution function such as the exponential family. Similarly, the observation distribution can be modelled parametrically or non-parametrically, continuously or discretely as well as dependently or independently of the state durations.

For hidden semi-Markov models the same principal methods for learning the parameters as well as inference can be applied. The complexity O for the exact inference with a Viterbi algorithm for a HSMM is $O(J M_j (J + M_j))$ in the worst case [93].

5.1.1.4 Factorial Hidden Markov Models

Factorial hidden Markov models (FHMM) are an extension to standard HMM. Instead of one single Markov chain multiple independent Markov chains of latent state variables describe a single time step t. Figure 5.4 shows a graphical representation of the model.

By a factorial hidden Markov model the number of information for a single time step t can be increased through additional Markov chains. However, this could also be realized by the increase of states J of a standard HMM but, for example, to represent Y bit a standard HMM would need 2^Y states. In contrast to this, the same information can be incorporated by Y latent chains each with 2 binary states of a FHMM. The main disadvantage of FHMM is the additional complexity $O(J^{2Y}T)$ when training them [91].

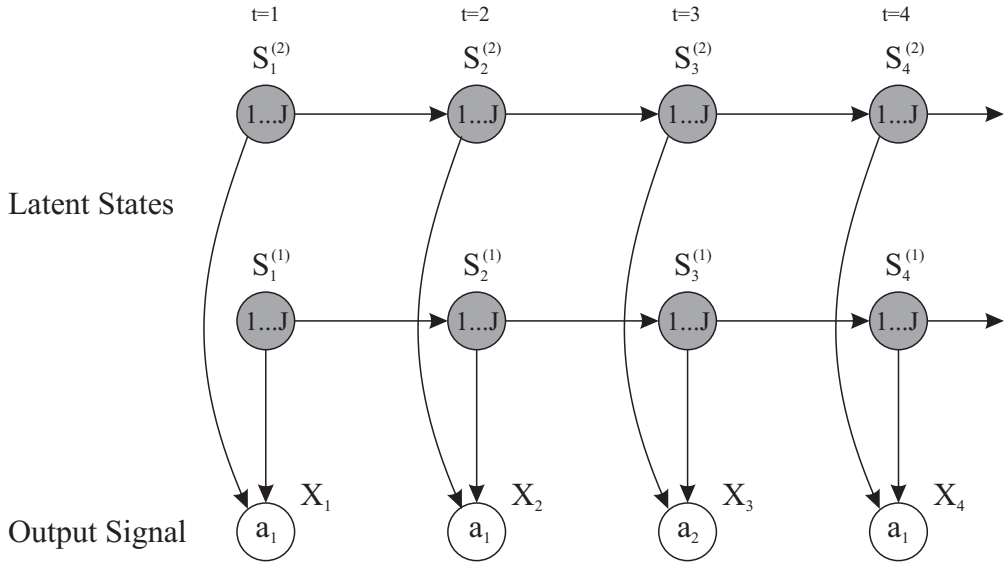

Figure 5.4: Factorial hidden Markov model comprising two Markov chains of latent variables with the same maximum number of states *J* which emit two different symbols a_1 and a_2

5.1.1.5 Viterbi Algorithm

Finding the most probable sequence of latent states in a hidden Markov model (also referred to inference) can be realized by the Viterbi algorithm. This problem can be exactly solved by the max-sum algorithm (Viterbi) due to the reason that the graph of a hidden Markov model is a directed tree [91]. The Viterbi algorithm computes the most probable state S_t for each time step *t*. At each time step *t* the most probable state is one out of *J* total different states. For illustrational reasons, two selected probable paths which connect the states of the single time steps of a HMM with *J*=3 can be seen in Figure 5.5.

The total number of possible paths for all time steps *t* grows exponentially with the total time interval *T* of the chain. By the Viterbi algorithm all these possible paths can be searched efficiently to get the most probable path for the given observation. The computational cost grows only linearly with this algorithm.

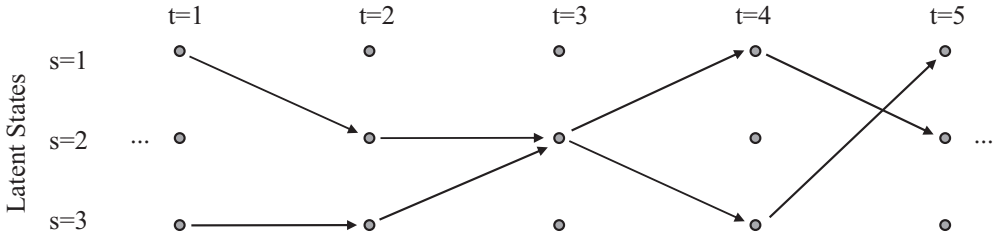

Figure 5.5: A fragment showing two possible paths for the time steps *t* of a HMM with *J*=3

For each path the probability can be evaluated by summing up the products of transition (p_{ij}) and emission probability (b_j) for each single time step t. If, for example, the time step t at a certain latent state j is considered there will be a lot of paths which go through this specific mentioned point (e.g. s=2, t=3). However, from all these paths only the path which has so far the highest probability has to be retained. Due to the reason that for each time step J different states are possible, also J different paths have to be considered for each time step t. At the next time step t+1 there have to be considered in total J^2 possible paths, but again only the most probable J paths have to be retained. This goes on until the last time step T is reached. From the last time step T, the most probable path can be revealed by tracking back to the most probable state of T-1 and so on until t=1 is reached.
For hidden semi-Markov models this procedure gets a bit more complex due to the reason that a specific state j occurs for a certain time duration $d_j(u)$. Instead of only considering the next time step as it was the case in HMM, for HSMM the joint probability of all time steps within the maximal duration M_j has to be considered for each time step t. This increases the computational complexity to $O(J\,M_j\,(J + M_j))$.

5.1.2 Comparison to Other Work

Several approaches have already been proposed to separate the electrical energy consumption of single appliances by hidden Markov models (HMM) [73]-[76]. Parson et al. [74] use a HMM in such a way that through an iterative process, the load curves of the devices are detected one after another. More complex factorial hidden Markov models (FHMM) based on multiple hidden Markov models are used by Kim et. al [73] and Kolter and Jakkola [75] to disaggregate multiple electrical appliances with their model.
One of the biggest drawbacks of FHMM is the increasing complexity that goes along with the number of electrical appliances. The complexity of the inference of FHMM increases exponentially with the number of electrical appliances, since the possible combinations of state changes rises [89].
Beyond this, electrical appliances, which are unknown and not modelled with the FHMM, are causing a residual load curve. This residual load has to be handled by the FHMM, which also increases the complexity. Due to the change of the electrical equipment by replacement or the introduction of new devices, the remaining unknown load curve also varies and as a consequence, influences the detection rate. The algorithms proposed by [73] and [75] additionally have the drawback that manual labelling of the disaggregated data is necessary, since there is no prior knowledge of the electrical appliances incorporated.
By contrast, Parson et al. [74] introduced a HMM for energy disaggregation that relies on prior appliance models. These models are tuned by the aggregated data using Expectation Maximization. The accuracy results are mainly dependent on the training data used for the prior. When sub-metered data is used for the model parameter estimation over fitting can be achieved which causes poor disaggre-

gation results. A drawback of this model is the lack of integration of reactive power values or additional features such as time of the day, which, for example, is incorporated by Kim et al. [73]. Beyond this, it is not possible to integrate state durational probabilities in the proposed method. Above all, the proposed algorithm [74] relies on power changes within the load curve which reflect steady state power changes of devices. Since this algorithm [74] is applied on a downsampled load curve with a sampling interval of one minute the power changes within the load curve from one sample to another are in the range of the actual steady state power changes of devices. When the sampling interval is decreased to a range of about several seconds these steady state power changes occur within several samples and the actual state changes can not be correctly detected anymore by using the difference values of the load curve.
To summarize, the above mentioned models provide a certain promising potential to the energy disaggregation of residential loads. However, for the unsupervised energy disaggregation the above mentioned drawbacks such as labelling of the data has to be avoided.

5.1.3 Model Selection

In general, residential appliances can be differentiated by certain appliance-specific characteristic (see Chapter 3). Especially, the time durational components such as on-duration combined with active and reactive power values are important features which improve the energy disaggregation process [73]. Statistical models such as HMM or FHMM have already been applied for energy disaggregation. To compensate the above mentioned drawbacks a model is necessary which is able to detect a certain device D when the other devices in the aggregated loads are not known. Beyond this, the model has to be able to incorporate time durational distribution for the states of a device.
The hidden semi-Markov model can be used as a base to meet the above mentioned conditions. Due to the reason that only certain sections of the load curve are generated by the device to be disaggregated the standard learning and inference procedures have to be adapted. The method developed in this thesis reduces above mentioned drawbacks by a heuristic procedure for learning the model parameters and a modified Viterbi algorithm for the approximate inference of the HSMM.

5.1.4 Basic Concept of the Novel Method for Energy Disaggregation

The fEEDBACk algorithm is based on prior models of appliances, similarly to Parson et al. [74]. The load curve of a single device can be estimated without the knowledge of other devices in the aggregated load. The aggregated load curve can be disaggregated iteratively, device by device. For the incorporation of temporal parameters duration hidden semi-Markov chains are used. The information gathered in Chapter 3 is used in an appliance-specific database that holds typical appliance parameters.

The basic concept of the represented approach is partly similar to Parson et al. [74], which uses prior models of general appliances types. However, the key differences in the fEEDBACk method are:

- the incorporation of typical estimated state durations of electrical appliances
- the integration of conditional features such as weekday, season or time
- the integration of reactive power values
- the estimation of model parameters based on filtering and clustering techniques
- an especially adapted Viterbi algorithm for hidden semi-Markov models.

The fEEDBACk algorithm works as follows (see also Figure 5.6):

Feature Extraction

In the first step, the features of the raw load curve are extracted. The events are extracted through Baranski's algorithm and the steady state power changes of the events are computed with the advanced version of the algorithm, which uses a pre- and post-event window proposed in Section 4.2.1.

Least Power Block Search

Feasible events are combined and the minimum power levels within the blocks are generated in a second step. Each block consists of an event with a positive steady state power change at the beginning and a negative at the end. Beyond this, the time duration between beginning and end gives information about the on-duration.

Appliance-Specific Filtering

The extracted power blocks reveal where a certain amount of active power is turned on within the total load curve. The statistical data of electrical appliances, which is investigated in Chapter 3, is used to build up the appliance-specific data base. Through this database, which includes typical on- and off-times as well as power consumptions of several electrical appliances, relevant power blocks and as a consequence sections of the aggregated load curve are filtered out where it is feasible that a specific electrical appliance is turned on.

Model Parameter Estimation

The filtered power blocks allow to estimate the parameters such as on- and off-duration or steady state power changes of the specific electrical device. The data of the power blocks is used for a clustering procedure where the active and reactive power changes of the device within the relevant scope are detected. This also allows an estimation of the distribution of the on- and off-durations using the clustered events.

Determination of Best Path

After extracting all relevant model parameters, the most probable path for the state changes of a certain electrical appliance is computed by a novel modified Viterbi algorithm. Only relevant events are used by the Viterbi algorithm. Additionally, data of the appliance database is used for a better performance of the modified Viterbi algorithm.

Reconstruction of Load Curve

The computed best path with the maximum likelihood is used to reconstruct the load curve for a specific device. Also an estimate of the load profile of the electrical device is extracted from the aggregated load curve.

Evaluation

The predicted load curve is compared to the ground truth data and different evaluation metrics are carried out to evaluate the performance.

After the predicted load curve of a certain device has been successfully disaggregated, it can be subtracted from the aggregated load curve. Due to the reduction of events this can make the disaggregation process of the next device simpler if the detection is correct. However, it is not a necessity to subtract the disaggregated load curve because the proposed algorithm also works with the measured aggregated load curve for other devices.

5.1.5 Model Definition for Electrical Appliances

This section provides detailed information about the model of an electrical appliance used in the fEEDBACk algorithm.
The load curve of a certain electrical appliance is modelled as a first-order hidden semi-Markov model in this new energy disaggregation approach. Due to the fact that only particular observations within the load curve are generated by the appliance to be disaggregated, some modifications to the standard algorithms for HSMM have to be done to incorporate these changes. While a modified version of the Viterbi algorithm is used to find the best path of the model, the parameter estimation methods are based primarily on expert heuristic approaches.

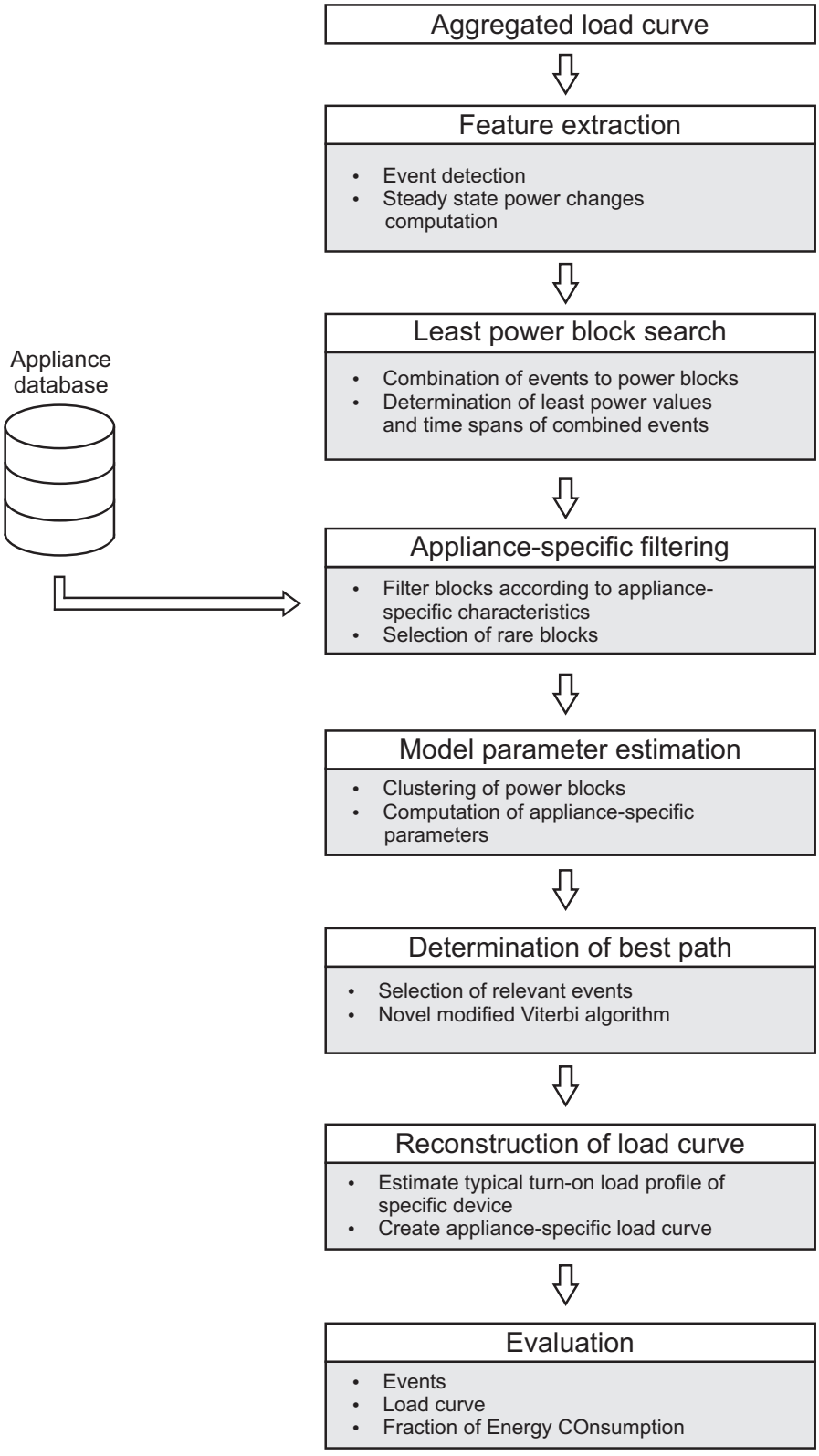

Figure 5.6: Block diagram of novel method for energy disaggregation

5.1.5.1 Specific Model Characteristics

Similarly to Parson et al. [74] the differential values of the power changes, to be exact, the total power changes of the detected events $\Delta P(E)$, are used as inputs for the HSMM. It is assumed that a state change of a device goes along with a change in power demand, which can be detected as an event. Therefore, a state change is just allowed at the time point of a certain detected event and not at time points in between.

The Markov chain of a refrigerator can be seen in Figure 5.7. There is an on- and off-state coupled with a certain power change value, which is necessary to switch from one state to another. Since the power consumption of refrigerators decreases in time to a certain limit, the power change values are not identical. Instead of using the total power spike caused by a turn-on, the steady state power changes are computed by the advanced version, which uses pre- and post-event windows (see Chapter 4). This leads to a greater accuracy for the disaggregation process.

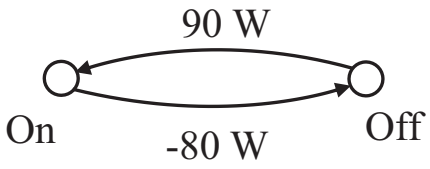

Figure 5.7: Markov chain of a refrigerator

A graphical representation including the durational components is given in Figure 5.8. The observable total power changes of the events $\Delta P(E)$ are related to the states S_t. Only events that belong to the device to be disaggregated are related to the HSMM, all other events are filtered out (dotted lines). A detailed description of the filtering process within the Viterbi algorithm (see Section 5.1.1.5) accomplished by a maximum likelihood computation can be found in the following sections.

A Gamma distribution is used for the state duration $d_j(u)$, which has already been investigated by Kim et al. [73]. The emission probabilities are modelled with the Gaussian distribution.

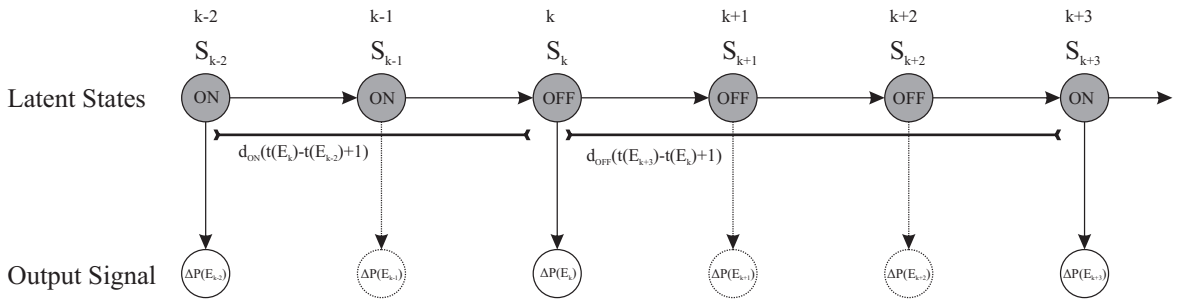

Figure 5.8: The graphical representation of a HSMM for a fridge

5.2 Steps of the Novel Method for Energy Disaggregation

5.2.1 Feature Extraction

Event Detection

In the very first step, the features (events) of the load curve are extracted. Using the results of Chapter 4 the algorithm of Baranski is applied to detect all events of the aggregated load curve. The detected points of time of the events are used in the subsequent step for the computation of the steady state power changes.

Steady State Power Changes Computation

The steady state power changes of the events are determined in this step.
For realizing this, the aggregated load curve used as input is smoothed with a simple median filter with a window size of 5 samples. The steady state power changes are computed by the "advanced algorithm" version as described in Section 4.2 and with a window length of 6 samples using the time points of the detected events.

5.2.2 Least Power Block Search

In order to find sections within the aggregated load curve where a certain amount of minimum power is used for a specific amount of time, the least power block search is used. By the term "least" the minimum power between a positive and negative power change of an event is meant.
The steady state power changes of events cannot be used as an exclusive feature for energy disaggregation, since these changes are influenced by power fluctuations caused by electrical appliances as well as probable simultaneous state changes of other appliances. For these reasons, another features are necessary to successfully detect electrical devices.
In the fEEDBACk algorithm, the minimum power value between two events as well as the duration of this power value are introduced as a feature. Combining the minimum power value and the duration leads to a power block. The combination of feasible events to the mentioned power blocks is called the least power block search. The basic principle of the method is based on the fact that all electrical appliances typically have a minimum power consumption as well as standard state duration (e.g. on- and off-time). The power blocks represent sections in the load curve where it is feasible that a certain electrical appliance is in its on-state.

Combination of Events to Power Blocks

A simple example for the basic principle of the least power block search can be seen in Figure 5.9. The method starts with the first positive power change and combines events with all negative pow-

er changes where the minimum power values $P_{k,l}^{\text{cont}}$ have to be above a certain limit P_{thr} between the events (see (5.4) and (5.5)). This results in four different blocks (B_1 - B_4), in this example. All detected power blocks are used to reveal sections where a certain device D could be in its on-state. Events with a positive total power change are denoted with E_k^+ and all events with a negative total power change with E_l^-:

$$E_k^+ := \Delta P(E_i) > 0$$
$$E_l^- := \Delta P(E_i) < 0 \tag{5.2}$$

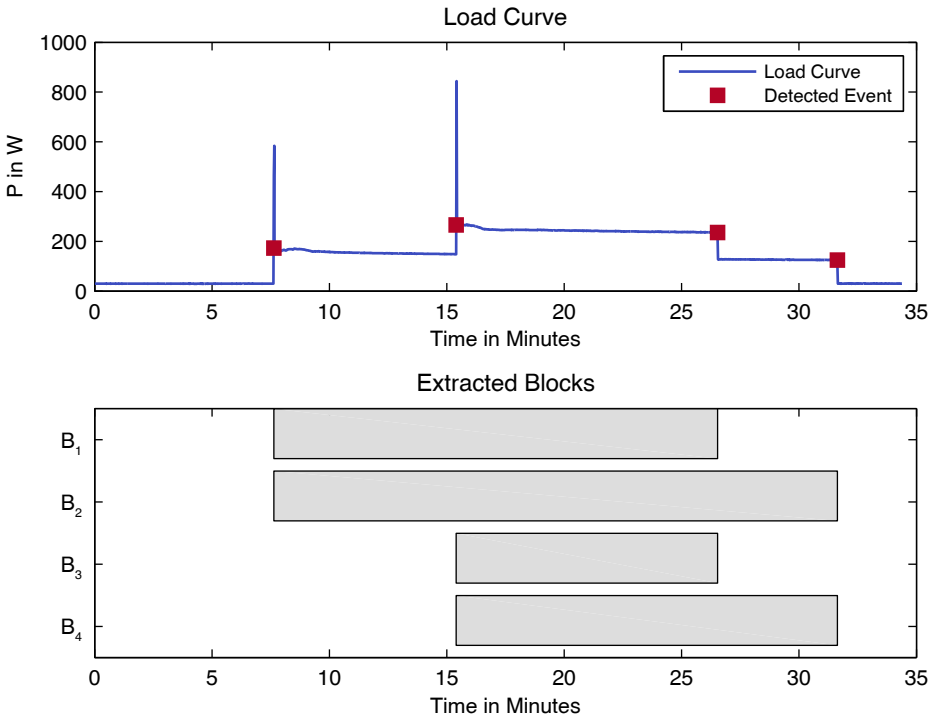

Figure 5.9: Durations of extracted least power blocks (B_1 - B_4)

To be able to compute the minimum power value of a block, the residual power value $P_{R,k}$ caused by other turned on electrical appliances, needs to be known. In Figure 5.9 the residual power value is about 25 W and it has to be subtracted from the minimum power value between the two events of a power block (E_k^+ and E_l^-). The power offset can change during the observation time and has therefore to be computed for each event and a certain time duration with length $M_{\text{on}}(D)$, which is the maximum possible on-duration of a certain electrical device D. This ensures that the minimum power value $P_{k,l}^{\text{cont}}$ continuously occurring within a block can be computed. To account for the power changes of the start event, a window w_R, which enlarges the considered section of the load $M_{\text{on}}(D)$ curve, is appended. The residual power value $P_{R,k}$ for a certain event E_k^+ can be computed with:

$$P_{R,k} = \min\left(P\left(t(E_k^+) - w_R\right), ..., P\left(t(E_k^+) + \mathrm{M}_{on}(D) + w_R\right)\right) \tag{5.3}$$

$P_{R,k}$ reflects the residual power value which is, for example, caused by devices which are constantly switched on, such as the standby losses from a TV-set or off-mode losses of a transformer. Electrical appliances which are always on cannot be detected with the proposed energy disaggregation method. The minimal continuous power value $P_{k,l}^{\mathrm{cont}}$ is the active power value that continuously arises within a detected block, and can be computed as follows:

$$P_{k,l}^{cont} = \min\left(P\left(t(E_k^+) + w_c\right), ..., P\left(t(E_l^-) - w_c\right)\right) - P_{R,k} \tag{5.4}$$

The window w_c is used to estimate the point of time where the device is in its steady state, in other words after a turn-on is completed or before a turn-off starts. This is necessary since the exact starting point of time of an event does not have to be the end of change of an event (see Section 4.1.3.2).

Determination of Least Power Values and Time Spans

To establish a power block B_m, only those events are combined that have a greater minimal continuous power value $P_{k,l}^{\mathrm{cont}}$ than a certain threshold P_{thr}. Furthermore, a positive power change E_k^+ is combined only with negative power changes E_l^- arising after positive power changes:

$$\begin{aligned} &B_m = \left\{E_k^+, E_l^-\right\} \quad if \quad P_{k,l}^{cont} > P_{thr} \\ &\text{where} \quad t(E_k^+) < \mathrm{t}(E_l^-) \end{aligned} \tag{5.5}$$

The succeeding power blocks can be computed with:

$$\begin{aligned} &B_{m+1} = \begin{cases} \left\{E_k^+, E_{l+1}^-\right\} & if \quad P_{k,l+1}^{cont} > P_{thr} \\ \left\{E_{k+1}^+, E_{l'}^-\right\} & if \quad P_{k,l+1}^{cont} \le P_{thr} \wedge P_{k+1,l'}^{cont} > P_{thr} \end{cases} \\ &\text{where} \quad t(E_{k+1}^+) < \mathrm{t}(E_{l'}^-) \end{aligned} \tag{5.6}$$

After all power blocks are found within the load curve, additional parameters such as starting point of time $\mathrm{t}_{\mathrm{start}}(B_m)$, end point of time $\mathrm{t}_{\mathrm{end}}(B_m)$ and on-duration $t_{\mathrm{on}}(B_m)$ are computed. Additionally, also the hour of the day or the weekday of the block B_m can be easily extracted.

There are always two power values $P(B_m)$ and $P_{\mathrm{E}}(B_m)$ associated to one single block. The former equals the minimum continuous power and the latter is limited to the minimal size of the associated events and the continuous power. These values are used in the next steps of the algorithm:

$$t_{start}(B_m) = t(\mathrm{E}_k^+) \tag{5.7}$$

$$t_{end}(B_m) = t(\mathrm{E}_l^-) \tag{5.8}$$

$$t_{on}(B_m) = t_{end}(B_m) - t_{start}(B_m) \tag{5.9}$$

$$P(B_m) = \mathrm{P}_{k,l}^{cont} \tag{5.10}$$

$$P_E(B_m) = \min\left(\Delta P(E_k^+), \left|\Delta P(E_l^-)\right|, \mathrm{P}_{k,l}^{cont}\right) \tag{5.11}$$

where $\mathrm{E}_k^+, \mathrm{E}_l^- \in \mathrm{B}_m$

5.2.3 Appliance-Specific Filtering

In this step, of the novel energy disaggregation algorithm all relevant power blocks, where it could be feasible that a certain electrical appliance is in its on-state, are filtered out. This can be realized due to the fact that different types of appliances have different appliance-specific characteristics. The typical parameters for active and reactive power consumption, turn-on and turn-off durations as well as time of use, has already been shown in Chapter 3. The parameter sets aggregated in Chapter 3 are used for accomplishing the filtering. The full parameter set of a certain electrical device *D* stored in the appliance database is as follows:

$P_{min}(D)$	Minimal continuous active power consumption of device *D*
$P_{max}(D)$	Maximal continuous active power consumption of device *D*
$\mathrm{t}_{\mathrm{on,min}}(D)$	Minimal on-duration time of device *D*
$\mathrm{t}_{\mathrm{on,max}}(D)$	Maximal on-duration time of device *D*
$\upsilon_{\mathrm{on}}(D)$	On-duration distribution of device *D*
$\mathrm{t}_{\mathrm{off,min}}(D)$	Minimal off-duration time of device *D*
$\mathrm{t}_{\mathrm{off,max}}(D)$	Maximal off-duration time of device *D*
$\upsilon_{\mathrm{off}}(D)$	Off-duration distribution of device *D*

Filter Blocks According to Appliance-Specific Characteristics

A simple but also effective filtering can be realized by using the active steady state power changes as well as the time spans of a certain type of device. But there could also simply added the time of day for the filtering process when it is practical. The Blocks $\mathbf{B}'(D) = \left\{B_1'(D), ..., B'_{n_{B'}}(D)\right\}$, where it is feasible that a certain appliance *D* is in its on-state, are defined as:

$$\begin{aligned}\mathbf{B}'(D) = \{B_m \,|\, & P_{min}(D) \le P(B_m) \le P_{\max}(D) \wedge \\ & t_{on,\min}(D) \le t_{on}(B_m) \le t_{on,\max}(D)\} \quad 1 \le m \le n_{B'}\end{aligned} \tag{5.12}$$

Selection of Rare Blocks

Events E_{k}^+ represented in just a few blocks of the set $\mathbf{B}'(D)$ are of special interest. The reason for this is that the lower the number of a certain event within the set $\mathbf{B}'(D)$, the higher the probability that those blocks belong to the device *D* under investigation. In other words, when an event within

the blocks **B**'(D) rarely occurs (below a number of s samples), there are just a few feasible combinations with other events. This also means that the probability that those blocks **B**''(D) are coincidentally resultant from other electrical devices is at a minimum. Typically, in such areas the load curve is not superimposed by the load profiles of other electrical devices, and as a consequence, the appliance-specific values from a certain device can be detected more easily. The blocks **B**''(D) can be computed with:

$$\mathbf{B}''(D) = \left\{ \mathrm{B}'_m(D) \middle| E_k^+ \in \mathrm{B}'_m(D) \wedge \left(1 \le \left| \mathbf{B}'(D) \in E_k^+ \right| \le s \right) \right\} \quad 1 \le m \le n_{B'} \tag{5.13}$$

The number of occurrences s depends on the device under investigation as well as the total measurement interval T. The longer the measurement interval, the more likely it is that there are periods where just a few electrical devices influence the device under investigation D.

5.2.4 Model Parameter Estimation

In this section the model parameters which are used for the disaggregation process are determined. The standard procedure which is used for learning the parameters for a HSMM cannot be applied to the fEEDBACk algorithm. Therefore a model parameter learning procedure has to be developed. The blocks **B**''(D) are used for learning the model parameters for the HSMM.

Clustering of Power Blocks

To accomplish the model parameter estimation, first of all a clustering procedure has to be carried out, to find the most probable area within the power blocks **B**''(D) of the device under investigation. Since the procedure has to be unsupervised the well-known density based clustering procedure DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [95] is used for realizing this. Clusters are considered as dense when the input data points in a n-dimensional space are close to each other. There are two parameters which have to be set for the DBSCAN algorithm: the radius ε and the number of points within the radius N_ε. For a randomly selected data point x the number of data points within the radius ε are computed. If the number of elements within the distance ε is above N_ε, the data point x is marked as dense and that it belongs to a cluster. All other points which surround the data point x within the distance ε are directly density-reachable and also belong to the same cluster. Other data points which have at least one of the before mentioned directly density-reachable data points within their radius ε are also added to the same cluster. Data points which have no one of the before mentioned data points within the distance ε are treated as noise.

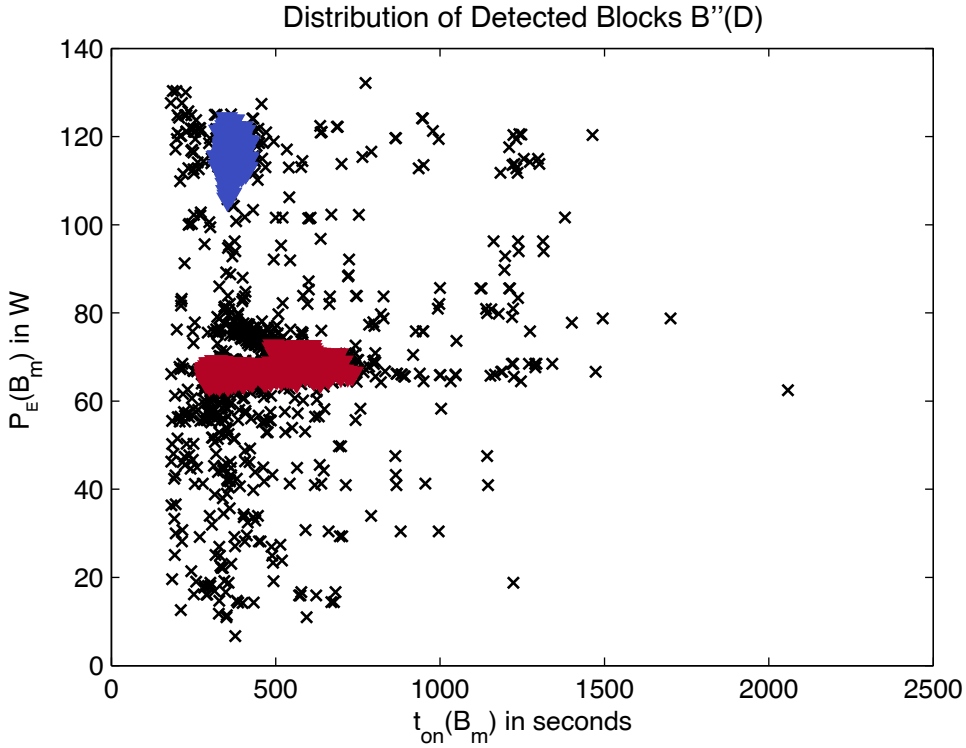

Figure 5.10: Distribution of detected blocks **B**''(D) of an aggregated load curve (black) which contains two different refrigerators (clusters are marked in blue and red)

In the fEEDBACk algorithm, as an input for the clustering procedure the continuous power values which incorporate the events $P_E(B_m)$ and their associated durations $t_{on}(B_m)$ are used. Through the usage of $P_E(B_m)$ another filtering process can be realized for power changes of events which are superimposed by events of other electrical appliances. Figure 5.10 shows an example which is generated from a load curve where two different refrigerators are found. The two clusters found are marked with blue and red data points. Data points which are not density-reachable are black, as per Figure 5.10.

Typically, the distance ε is calculated as an Euclidean distance between two data points. As can be seen in Figure 5.10, the ranges of the amplitudes of the power axes (0-140) and the time axes (0-2500) strongly differ from each other. To find dense clusters in such a case, a normalization of the parameter ranges is necessary.

For realizing this, the distance function is modified in such a way, that it just evaluates the relative distance from the data point under investigation ($t(B_x)$ and $P(B_x)$) to another data point ($t(B_d)$ and $P(B_d)$). Through the relative distance also the relative variation of the two components can be incorporated. Through this computation also the facts that the power values as well as the on-time durations vary around ±10% of their median values, in the case of the refrigerator which is figured out in Chapter 3, can be incorporated.

Additionally, a factor c_P can be added when the relative variation from the two input components vary from each other. For example, when the active power has a standard deviation of 3% and the standard deviation of the on-duration is 10% then c_P =3, otherwise it remains c_P =1. For reasons of completeness also a factor c_t =1 introduced.

$$d = \sqrt{\left(c_P \frac{P(B_d) - P(B_x)}{P(B_x)} \right)^2 + \left(c_t \frac{t(B_d) - t(B_x)}{t(B_x)} \right)^2} \tag{5.14}$$

For computing the two necessary parameters ε and N_ε for the DBSCAN algorithm a simple approach is used. First of all the number of close data points for each block in $\mathbf{B}''(D)$ which are within a relative distance of d < 3% are calculated. The value for N_ε is selected from that number which have a probability of 50 %. This procedure has shown good results.

If there are two devices of the same type within the Pt-plane for the power values, a difference of at least 15% from the greater power value and for the on-duration, a difference of more than 50% of the median value of the greater on-duration is necessary to avoid overlaps.

The power blocks which are found in a cluster can be denoted with $\mathbf{B}_C(D)$. If there are more than two appropriate cluster centres found it indicates that there are more electrical devices of the same type found. Only the power values of one single device is extracted and used for the next step.

Computation of Appliance-Specific Parameters

The parameters of all blocks from $\mathbf{B}_C(D)$ are used to learn the model parameters from the device under investigation D. To get more accurate results for the on-power changes the power values which are beyond $P_{min}(D)$ and $P_{max}(D)$ are filtered out. The median and standard deviation values are computed from the remaining power changes and are used for the estimated total on-power changes as well as for the estimated variations of those values:

$\Delta P'_{on}(D)$	Estimated total active on-power change of device D
$\Delta Q'_{on}(D)$	Estimated total reactive on-power change of device D
$\sigma'_{on}(D)$	Estimated variation of total on-power change of device D
$\Delta P'_{off}(D)$	Estimated total active off-power change of device D
$\Delta Q'_{off}(D)$	Estimated total reactive off-power change of device D
$\sigma'_{off}(D)$	Estimated variation of total off-power change of device D

For computing the time durational distributions also a filtering procedure is used. First of all the median values of the on- and off durations within the clustered blocks $\mathbf{B}_C(D)$ are computed. All durations which are around ±20% of the median values are considered for the time distribution. The parameter sets for the assigned distributions ($dist_{On}(D)$, $dist_{Off}(D)$) which are usually Gamma distributed [73] are computed with the well-known maximum likelihood estimation. In MATLAB® coded max-

imum likelihood methods for a Gamma distribution (gamfit) exist. The computed parameter set for the particular distributions can be denoted with $\zeta'_{\text{on}}(D)$ and $\zeta'_{\text{off}}(D)$.

$\zeta'_{on}(D)$ Estimated parameters for the on-duration distribution of device D

$\zeta'_{off}(D)$ Estimated parameters for the off-duration distribution of device D

The computed median values for each state duration are in a row used for the maximum duration length $M_j(D)$ of a state j. Depending on a the device D, a maximum duration between 2 and 3 times of the median value is used as the maximum run length $M_j(D)$ (see Chapter 3). Beyond this, also the state transitions have to be estimated. As already described above, in the case of cooling devices only two states are used and hence no estimation has to be carried out. For the initial probabilities π_j a factor of 0.5 is used in the case of cooling devices since a turn-on and a turn-off have the same probability.

5.2.5 Determination of Best Path

The best path, which is also called the most probable path, denotes those events in the aggregated load curve that have the maximum probability to have been generated by the electrical device D. In order to find the most probable path, all relevant events are needed which could be caused by a certain electrical device D. In this section, the procedure of finding all relevant events as well as the modified Viterbi algorithm for finding the best path are described and determined in this section, which is also described as finding the best path. To accomplish this task, the estimated real power changes ($\Delta P(D)$, $\Delta Q(D)$) of the device as well as on- and off-duration distribution extracted in the preceding section are used as input in this step. The Viterbi algorithm is very well-known for finding the correct state sequences in a hidden Markov model as already described in Section 5.1.5. In the proposed algorithm, a modified version is introduced to disaggregate a certain electrical device from the total load curve.

5.2.5.1 Relevant Events

Events within the load curve exceeding the minimum continuous power of the device D under investigation, are treated as relevant events. This means, that all events within the blocks $\mathbf{B}'(D)$ are considered as relevant. All events $\mathbf{E}'_{all}(D) = \left\{E'_{all,1}(\mathrm{D}),...,E'_{all,n_{E'a}}(D)\right\}$ within the set of power blocks $\mathbf{B}'(D)$ are determined by:

$$\mathbf{E}'_{all}(D) = \left\{E_i \,\middle|\, \mathrm{t}_{start}(B'_m(D)) \le t(E_i) \le \mathrm{t}_{end}(B'_m(D))\right\} \quad 1 \le i \le n_E, 1 \le m \le n_{B'} \tag{5.15}$$

Through the filtering of events with in the scope of a certain device D, the disaggregation process improves. Above all, areas of the load curve, where it is not feasible that a certain device is in its on-state, are eliminated.

5.2.5.2 Novel Modified Viterbi Algorithm

Since the standard Viterbi algorithm assumes that each observation belongs to a state of the HSMM, a modification to the standard procedure has to be introduced. Beyond this, not the power values $P(t)$ and $Q(t)$, but the steady state power changes of relevant events $\Delta P(E'_{all,i}(D))$ are used as observations for the modified Viterbi algorithm.
Similar to Guédon [93], a forward $\alpha_j(k,D)$ and a backward probability $\beta_j(k,D)$ for each relevant event $E'_{all}(D)$ is computed. In the standard HSMM, the observations of a certain state frequently occur in a series, which is incorporated by product sums of the probabilities of the observations in the Viterbi algorithm. A repeated observation over a certain time span increases the probability for a certain state j.

Due to the fact that in the presented model, only power changes of events are used as observations of the HSMM, and not the power level themselves, the Viterbi algorithm has to be modified. In a first step, the forward and backward probabilities for each event $E'_{all,k}(D)$ are computed. So, the forward probability $\alpha_j(k,D)$ for a certain Device D can be calculated for all events $n_{E'a}$ with:

$$\begin{aligned}
\alpha_j(k,D) = \max_{i \neq j} \Bigg[\max_{1 \le u \le k-1} \Big[& b_i\left(\Delta P\left(E'_{all,k-u}(D)\right)\right) \cdot \\
& \cdot d_i\left(t\left(E'_{all,k}(D)\right) - t\left(E'_{all,k-u}(D)\right) + 1\right)\Big] p_{ij}, \\
& b_j\left(\Delta P\left(E'_{all,k}(D)\right)\right) d_i\left(t\left(E'_{all,k}(D)\right)\right)\pi_i \Bigg] \\
\textit{where} \quad & \left(t(E'_{all,k}(D)) - t(E'_{all,k-u}(D)) + 1\right) \le M_j(D) \wedge \\
\Bigg(\textit{if } i = on: \quad & \left\{t(E'_{all,k-u}(D)), \ldots, t(E'_{all,k}(D))\right\} \in \left\{ t_{end}\left(B'_s(D)\right) \le t_{start}\left(B'_{s+1}(D)\right) \right\}_{1 \le s \le n_{B'}} \Bigg)
\end{aligned} \tag{5.16}$$

By the constraints in (5.16) the maximum time distance between the events $E'_{all,k}(D)$ and $E'_{all,k-u}(D)$ is limited to maximum run length $M_j(D)$. When the former state i was "on" all the time points of the considered events have to be within overlapping blocks B'_s where the continuous power exceeds the minimum power of the device D. Also, the constraints in (5.17) are similar to (5.16) but instead of the on-states there are only off-states considered.
The backward probability $\beta_j(k,D)$ can be written as:

$$\beta_j(k,D) = \max_{i \neq j}\left[\max_{1 \le u \le n_{E'_{all}} - k} \left[b_i\left(\Delta P\left(E'_{all,k+u}(D)\right)\right) \cdot \right.\right.$$
$$\left.\left. \cdot d_j\left(t\left(E'_{all,k}(D)\right) - t\left(E'_{all,k+u}(D)\right) + 1\right) p_{ji} \right]\right] \tag{5.17}$$
$$where \ \left(t(E'_{all,k+u}(D)) - t(E'_{all,k}(D)) + 1\right) \le M_j(D) \wedge$$
$$\left(if\ i = off: \quad \left\{t(E'_{all,k}(D)), \ldots, t(E'_{all,k+u}(D))\right\} \in \left\{ \mathrm{t}_{end}\left(B'_s(D)\right) \le \mathrm{t}_{start}\left(B'_{s+1}(D)\right) \right\} \right)$$

The total probability $\gamma_j(k,D)$ is computed with:

$$\gamma_j(k,D) = \alpha_j(k,D)\beta_j(k,D) b_j\left(\Delta P\left(E'_{all,k}(D)\right)\right) \tag{5.18}$$

Besides the forward and backward probabilities, (5.18) also incorporates the probability of the observed power change k from the event $E'_{all,k}(D)$ for the state j. Through the forward probabilities the maximum probability of the previous events belonging to a state unequal to j and being within the time range $M_j(D)$ are computed. Similarly, the backward probability is computed. An illustration of these procedures can be seen in Figure 5.11. The total probability $\gamma_j(k,D)$ shows how likely it is that the event belongs to state j.

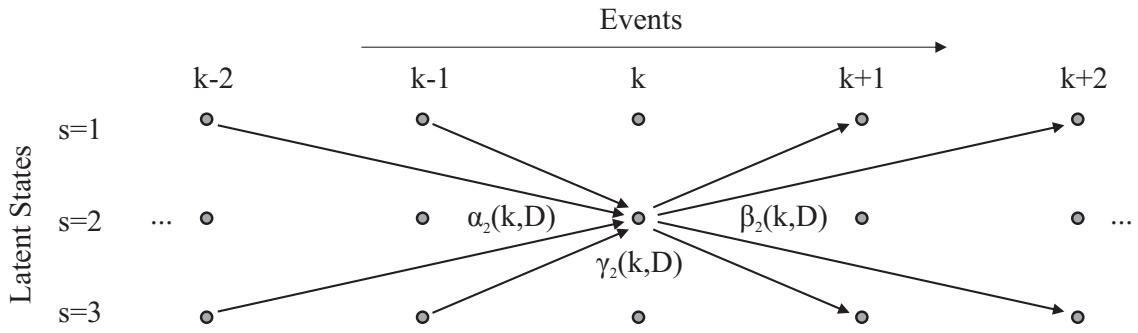

Figure 5.11: Illustration of the first paths that are incorporated for the computation of $\alpha_2(k,D)$ and $\beta_2(k,D)$ for the state j=2, an example with 3 states

Since not all observed events belong to the particular electrical device D to be disaggregated, these events have to be filtered out by the Viterbi algorithm. To realize this, a sliding window with a maximum run length $M_j(D)$ is used to find the most probable event within the window. The time step of $\gamma_j(k,D)$ within the sliding window with the highest probability is assumed to be an event from device D. This event serves as the next starting point for the sliding window with size $M_j(D)$ to extract the next most probable total probability, which is assumed to be the next state change of the particular device.

For the first estimated state $\hat{s}_1(D)$ and its assigned total probability $\gamma_i(\hat{z}_1(D),D)$ of a certain device D, which maximizes, the following formula can be written:

$$\begin{aligned}\left(\hat{s}_1(D),\hat{z}_1(D)\right) = \arg\max_{j,u}\Bigg[& \max_{2\le u\le n_{E'_{all}},\, j\ne i}\Big[\gamma_i(u,D)\cdot \\ & \cdot d_j\left(t\left(E'_{all,u}(D)\right)-t\left(E'_{all,1}(D)\right)+1\right)p_{ji}\Big]\Bigg] \\ \textit{where}\quad & \left(t(E'_{all,u}(D))-t(E'_{all,1}(D))+1\right)\le M_j(D)\end{aligned} \tag{5.19}$$

A state of the device D is denoted with the state variable $\hat{s}_m(D)$, and its duration with $t\left(E'_{all,\hat{z}_{m-1}(D)}(D)\right)-1$ to $t\left(E'_{all,\hat{z}_m(D)}(D)\right)$. The index $\hat{z}_m(D)$ corresponds to the event within the set of all relevant events $\mathbf{E'}_{all}$ in (5.15). The m-th state and its related event $E'_{all,\hat{z}_m(D)}(D)$ can be computed with:

$$\begin{aligned}\left(\hat{s}_m(D),\hat{z}_m(D)\right) = \arg\max_{i,u}\Bigg[& \max_{\hat{z}_{m-1}(D)<u\le n_{E'_{all}},\, j=\hat{s}_{m-1},\, i\ne j}\Big[\gamma_i(u,D)\cdot \\ & \cdot d_j\left(t\left(E'_{all,u}(D)\right)-t\left(E'_{all,\hat{z}_{m-1}(D)}(D)\right)+1\right)p_{ji}\Big]\Bigg] \\ \textit{where}\quad & \left(t(E'_{all,u}(D))-t(E'_{all,\hat{z}_{m-1}(D)}(D))+1\right)\le M_j(D)\wedge m>1\wedge \\ & \left(\textit{if } i=\textit{off}: \quad \left\{t(E'_{all,u}(D)),\ldots,t(E'_{all,\hat{z}_{m-1}(D)}(D))\right\}\in\left\{t_{end}\left(B'_s(D)\right)\le t_{start}\left(B'_{s+1}(D)\right)\right\}_{1\le s\le n_{B'}}\right)\end{aligned} \tag{5.20}$$

The constraints in (5.20) only allow events $E'_{all,\hat{z}_m(D)}(D)$ to be considered as a next state change of device D which are within the maximum run length $M_j(D)$ from the last detected event $E'_{all,\hat{z}_m(D)-1}(D)$. Beyond this, in the case of an off-state the events are further limited to areas within the load curve where it is feasible that the device D can be in its on-state. This is realized by just considering events for the off-state which are within the same power block as the last event $E'_{all,\hat{z}_m(D)-1}(D)$ or within overlapping power blocks which ensures that the minimum power requirement of the device D is always met.

The estimation procedure of the states $\hat{s}_m(D)$ and index of events $\hat{z}_m(D)$ gets simplified when the events from the clustered power blocks $\mathbf{B}_C(D)$ are used as detected and known on- and off-events of the specific device D. These off-events can act as a starting point for the Viterbi algorithm to find the next probable on-state.

5.2.6 Reconstruction of Load Curve

Several electrical devices such as electric heaters only have a constant power level that is dependent on the voltage. Other devices such as refrigerators or freezers typically do not have a constant power value over time (see Section 3.1). To disaggregate the load curve of the device under investigation

the load profile for each of its states has to be known. Since the focus of the fEEDBACk algorithm is to be unsupervised, a procedure for extracting the load profile of a certain state of a device *D* from the aggregated load is introduced.

The prerequisites for the extraction of the load profile for an estimated state $\hat{s}_m(D)=j$ are its appropriate time spans from $t\left(E'_{all,\hat{z}_m(D)}(D)\right)$ to $t\left(E'_{all,\hat{z}_{m+1}(D)}(D)\right)-1$ within the aggregated load curve. The load profile of a specific state *j* can be extracted from areas of the aggregated load curve where just a couple of other events from different electrical appliances can be found. The power blocks within the set **B**''(*D*) from (5.13) meet this requirement. All estimated states $\hat{\mathbf{s}}_j'(D)$ of a certain state *j* which fulfil both criteria can be denoted with:

$$\hat{\mathbf{s}}'_j(D)=\left\{\hat{\mathbf{s}}(D)\middle|\hat{s}_m(D)=j\wedge\hat{z}_m(D)\in\mathbf{B}''(D)\right\} \tag{5.21}$$

It is assumed that the total time interval *T* of the aggregated load curve long enough and contains several periods which can be used for the described procedure. A filtering procedure is carried out due to the fact that apart from different events, there are also minor power fluctuations, which are not detected as events, disturbing the aggregated load curve as well as power (state) changes from other devices. This is realised by using the difference values (4.1) of the active power values in (5.22) and applying the median function for each time step of all extracted differential power values. The estimated differential power value $\Delta\hat{P}_j(t,D)$ for the time step *t* of a certain state *j* can be computed from the aggregated load curve with:

$$\begin{aligned}&\Delta\hat{P}_j(t,D)=median\left\{\Delta P\left(t(\mathrm{E}_{all,\hat{z}_m(D)})+t-1\right)\right\}\\&where\quad m\in\hat{\mathbf{s}}'_j(D)\wedge\left(1\le t\le t_{j,min}(D)\right)\end{aligned} \tag{5.22}$$

The estimated power $\hat{P}_j(t,D)$ value of a specific state *j* for the time step *t* can be computed by the accumulated sum of (5.22) and the estimated steady state power change for the specific state *j*:

$$\hat{P}_j(t,D)=\sum_{i=1}^{t}\Delta\hat{P}_j(i,D)+\Delta P'_j(D) \tag{5.23}$$

An example for this procedure can be seen in Figure 6.5. With the knowledge of the power values $\hat{P}_j(t,D)$ from (5.23) of the specific states as well as $\hat{s}_m(D)$ and $\hat{z}_m(D)$ from (5.20) the estimated total active load power values $\hat{P}(t,D)$ for the time interval $1\le t\le T$ can be reconstructed. For the off-state, a power value of close to zero is usually used. Most of the electrical appliances have a very low standby power consumption of up to several Watts, which is just a fraction of the on-power consumption and therefore plays a minor role compared to the total energy consumption.

5.2.7 Evaluation

The most frequent evaluation algorithms for NIALM are used for computing the performance accuracy of the fEEDBACk algorithm. A detailed overview of further evaluation methods is given in Section 2.3.5.
The evaluation of the proposed energy disaggregation algorithm (fEEDBACk) is split in three parts. First of all, the estimated events where a state change occur are compared to the ground truth data. Secondly, it is checked whether the estimated load curve $P(t,D)$ is within a certain range of the ground truth. This has the additional advantage that disaggregation errors caused by a wrongly reconstructed load curve, but a correct most probable path (states), can be easily detected as well.
Finally, the estimated total energy of the device to be disaggregated is compared to the total energy of the ground truth data. However, the last metric is just provided for comparability reasons.

5.2.7.1 Events

Different metrics for event detection have already been presented in Section 4.1.2. Since the true positive and false positive events are mostly relevant for the energy disaggregation process, the true positive percentage (TPP) and false positive percentage (FPP) from (4.18) and (4.19) are used for evaluation purposes.

5.2.7.2 Load Curve

The single time points of the reconstructed load curve $\hat{P}(t,D)$ are matched with the ground truth $P^0(t,D)$ by an evaluation metric which is used in the information retrieval domain. It can be distinguished between four different outcomes in a binary classification task (see Section 4.1.2.1): true positive (TP), true negative (TN), false positive (FP), and false negative (FN). To be able to differentiate between wrong and correctly reconstructed on-power values, Kim et al. [73] proposed to distinguish between accurate true positive (ATP) and inaccurate true positive (ITP) instead of the usual single true positive (TP) classification. Let $\hat{P}(t,D)$ be the predicted value at time step t and $P^0(t,D)$ be the ground truth value:

- When $\hat{P}(t,D) = 0$ and $P^0(t,D) = 0$, the prediction is true negative (TN).
- When $\hat{P}(t,D) = 0$ and $P^0(t,D) > 0$, the prediction is false negative (FN).
- When $\hat{P}(t,D) > 0$ and $P^0(t,D) = 0$, the prediction is false positive (FP)

To incorporate the accuracy of the estimated power values the outcome of TP is further split into accurate and inaccurate true positive (TP = ATP + ITP):

- When $\hat{P}(t,D) > 0$, $P^0(t,D) > 0$, and $|\hat{P}(t,D) - P^0(t,D)| / P^0(t,D) \leq \kappa$, the prediction is an accurate true positive (ATP).

- When $\hat{P}(t,D) > 0$, $P^0(t,D) > 0$, and $|\hat{P}(t,D) - P^0(t,D)| / P^0(t,D) > \kappa$, the prediction is an inaccurate true positive (ITP).

The variable κ is a threshold due to the reason that the power values of electrical devices vary. Measurements have shown that power consumptions vary up to 20 %. Precision and recall can be calculated with [73]:

$$\text{Precison} = \frac{ATP}{ATP + ITP + FP} \tag{5.24}$$

$$\text{Recall} = \frac{ATP}{ATP + ITP + FN} \tag{5.25}$$

By the term *precision,* which is also called positive predictive value, the fraction of the number of on-power values that are relevant (*ATP*) to all estimated on-power values (*ATP+ITP+FP*) is meant. In contrast to this, *recall* which is also known as sensitivity describes the fraction of the number of on-power values that are relevant (*ATP*) to all ground truth on-power values within the load curve of a certain device (*ATP+ITP+FN*).
The F-Measure is the harmonic mean of precision and recall:

$$F\text{-}Measure = \frac{2 \cdot \text{Precison} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.26}$$

Energy disaggregation results of different electrical appliances are compared to each other in Chapter 6. As described above, the F-Measure considers both, the precision and the recall for computing a score. When both load curves (the predicted and the ground truth) match each other within the range κ, precision and recall gets one and as a consequence also the F-Measure gets one.

5.2.7.3 Fraction of Energy Consumption

Since the comparison of the total energy consumption of a certain electrical device has been used frequently as a metric for energy disaggregation, it has been included for the sake of completeness. However, a lot of papers have already mentioned the drawbacks of this metric, because a similar energy consumption can also be caused through significantly different load curves. The difference in energy $W_{\text{diff}}(D)$ can be expressed as:

$$W_{diff}(D) = \frac{\sum_t P(t,D)}{\sum_t P^0(t,D)} \tag{5.27}$$

However, through the fact that correlated power changes are randomly missing, also coincidental good results can be achieved by comparing the fraction of energy consumption.

6 Results of the Novel Energy Disaggregation Method

In this chapter, the performance of the fEEDBACk energy disaggregation algorithm is studied by analysing specific steps of the proposed algorithm as well as providing the evaluation results for the disaggregated load curve.
A case study for cooling devices (fridges an freezers) is used to evaluate the single steps of the algorithm. Several sensitivity analyses such as optimal parameter selection, the integration of the reactive power values as a feature, influence of the sampling period on the accuracy as well as the maximum sampling period for detecting appliances are carried out. A discussion about the universality of the method to other devices is given in the last section of this chapter.

6.1 Motivation

In order to evaluate the performance of energy disaggregation algorithms sub-metered data sets with a sampling period of a few seconds are necessary as a starting point. As already described in Section 2.4 there exist a couple of public data sets but most of them include measurement errors in form of repeated or missing power data values within specific sections of the load curve. For being able to correctly evaluate the performance of an energy disaggregation algorithm a data set is necessary which doesn't have missing or incorrect power values. Therefore the TUG data set [72] is used for evaluation purposes. This data set also includes reactive power values which can be used for further sensitivity analyses.
Criteria for eligibility for the device to be disaggregated are the availability of the sub-metered load curves of the device in each of the measured households, the consumption of active and reactive power, and a frequent usage of the device. Only cooling devices meet the above criteria in the available data set. To ensure the broadest possible range of different devices of a similar type, fridges and freezers are selected for the evaluation of the proposed algorithm.

6.2 Test setup

The algorithm was applied to the power measurement data TUG [72] (see also Section 2.4) with a sampling period of one second up to 20 seconds. Measurements were carried out for an interval of

two weeks in five households in Austria. The measurement devices ELOG 550 and ELOG 570 by DEWETRON® were used for gathering the test data. Current clamps which were installed in the distribution box were used for the power measurement. Originally, the measurement raw data have a sample rate of 10 kHz but these values are averaged to get the specific sampling periods.

If not otherwise stated, the event detection is realized by the algorithm of Baranski with a time duration for merging events of τ =2 samples and the minimum power for detecting a power change of an event P_{min}=10 W. The steady state power changes are computed with the proposed "advanced version" of the algorithm in Section 4.2. The window lengths used for the pre- (w_{Pre}) and post-event window (w_{Post}) have the same length of 6 samples.

6.3 Performance of the Estimation of the Appliance-Specific Parameters

This section provides information about the performance of estimating the appliance-specific parameters. The performance is demonstrated on the disaggregation of cooling devices (refrigerators and freezers).

6.3.1 Appliance-Specific Filtering

The appliance-specific parameters analyzed in Chapter 3 allow a filtering of the detected power blocks B_m from (5.5) and (5.6). The main goal of the filtering procedure is a separation of all detected power blocks **B** in such a way that the device under investigation D forms a dense and unique area in the Pt-plane of **B**''(D) (5.13). In the case of refrigerators and freezers, the filtering parameters always lead to such dense clusters in the Pt-plane of **B**''(D), which could be easily detected by the DBSCAN algorithm. Even if more than one cooling device is installed in a single household, no overlaps in the Pt-plane occurred in the investigated households, which is a prerequisite for being able to cluster the data (see Figure 5.10).
The DBSCAN algorithm allows to find dense areas in the input data set. For choosing the filter parameters ε and N_ε for the DBSCAN, an approach is used as described in detail in Section 5.2.4. The results have shown that this procedure works very well for electrical devices that are frequently turned on and off. For eliminating falsely detected clusters by the DBSCAN algorithm the number of elements within the detected cluster is a very meaningful information to perform a sanity check. If the detected number of elements in a detected cluster is very small (in the case of cooling devices: below 10 samples for a measurement interval of 5 days or above), it can be treated as noise. The exact number of data points within a cluster depends mainly on the total measurement interval T which is used for the disaggregation algorithm and the on-duration $t_{on}(B_m)$. An example for this is shown in Figure 6.1.

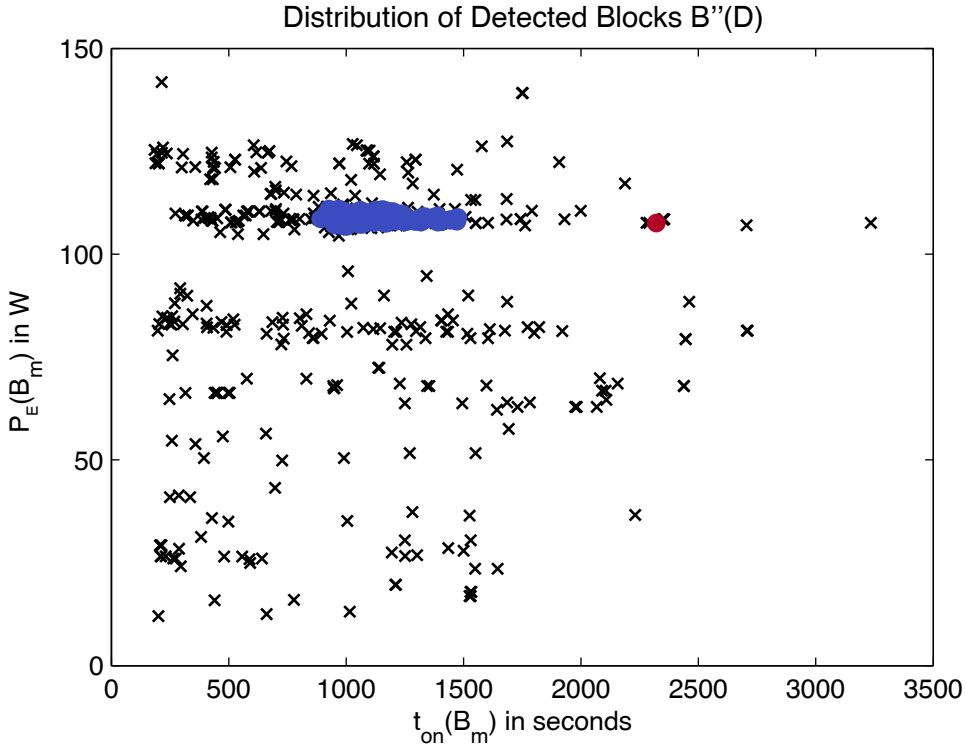

Figure 6.1: Distribution of detected blocks **B**''(*D*) of an aggregated load curve which contains a single refrigerator (blue cluster) and some data points which are treated as noise (red cluster)

In this example, the DBSCAN algorithm resulted in two different clusters since a couple of data points are close to each other (clusters are coloured in blue and red). By treating the cluster coloured in red as noise, a correct detection of the clusters of the refrigerator (marked in blue) can be performed.

Anyhow, one of the main drawbacks of the DBSCAN algorithm is that it can only find different clusters if these clusters have a similar density in their data points. If, for example, two refrigerators in one household have on-durations which differ widely, a separation of the formed clusters within the Pt-plane could perhaps not be accomplished by the DBSCAN algorithm. The reason for this is that the refrigerator with the longer on-duration leads to fewer power blocks and as a consequence to a less dense area in the Pt-plane. However, since this is a very rare condition, it is not further investigated in this thesis. A solution for this problem can be that the time durational input $t_{on}(B_m)$ is split in two or three parts, which allows different densities in each data set. Another solution for this would be the usage of another clustering algorithm.

6.3.2 Steady State Power Changes

The steady state power changes of a certain device *D* are computed via the detected cluster described in the preceding section. The mean values of the power changes within the cluster are used for computing the predicted power changes of a certain device. Table 6.1 shows a comparison of select-

ed steady state power change values. As can be seen, the maximum deviation between the average ground truth and the predicted steady state power change values is approximately 8%. The estimated values are used as mean values for the single states of the HSMM.

Table 6.1: Comparison of true steady state power changes (ΔP, ΔQ) and predicted values (ΔP', ΔQ')

j	Cooling Device	$\Delta P_j(D)$	$\Delta P'_j(D)$	-	$\Delta Q_j(D)$	$\Delta Q'_j(D)$	-
		W	**W**	**%**	**var**	**var**	**%**
on	1	81.6	76.6	-6.0	-64.7	-64.1	-1.1
	2	133.0	130.3	-2.1	-125.7	-126.0	0.2
	3	136.5	135.1	-1.0	-147.8	-147.4	-0.2
	4	121.2	118.2	-2.5	-121.8	-131.2	7.7
	5	107.3	99.7	-7.2	-104.6	-107.0	2.3
	6	118.5	118.6	0.0	-47.0	-46.9	-0.0
off	1	-65.7	-66.5	1.2	64.9	63.8	-1.7
	2	-108.2	-109.0	0.7	124.6	125.3	0.5
	3	-109.8	-109.4	-0.3	148.5	147.9	-0.4
	4	-95.9	-95.4	-0.5	122.8	132.5	7.9
	5	-87.7	-86.1	-1.8	106.4	108.0	1.5
	6	-95.6	-95.4	-0.2	46.2	46.5	0.7

As already mentioned above, in the aggregated load curve there are power fluctuations from other devices which influences the steady state power change of the device *D* to disaggregate. Also in the Viterbi algorithm these power changes are used as inputs for the disaggregation process. For computing the probability of the steady state power change of an event, the predicted power changes of the device *D* are treated as Gaussian distribution. To compute this probability a standard deviation value is also necessary. The computation of such a standard deviation value from the detected cluster, similarly to the mean value, very often leads to standard deviation values that are around a few Watts. Such small standard deviations strongly influence the modified Viterbi algorithm due to the reason that superimposed events from device *D* have a very low probability according to the normal distribution. For this reason, the minimum value of the standard deviation is set to 10% of the respective steady state power change, which leads as a consequence to a better performance of the fEEDBACk algorithm. True and predicted steady state power change values of a fridge with its respective Gaussian distribution functions can be seen in Figure 6.2.

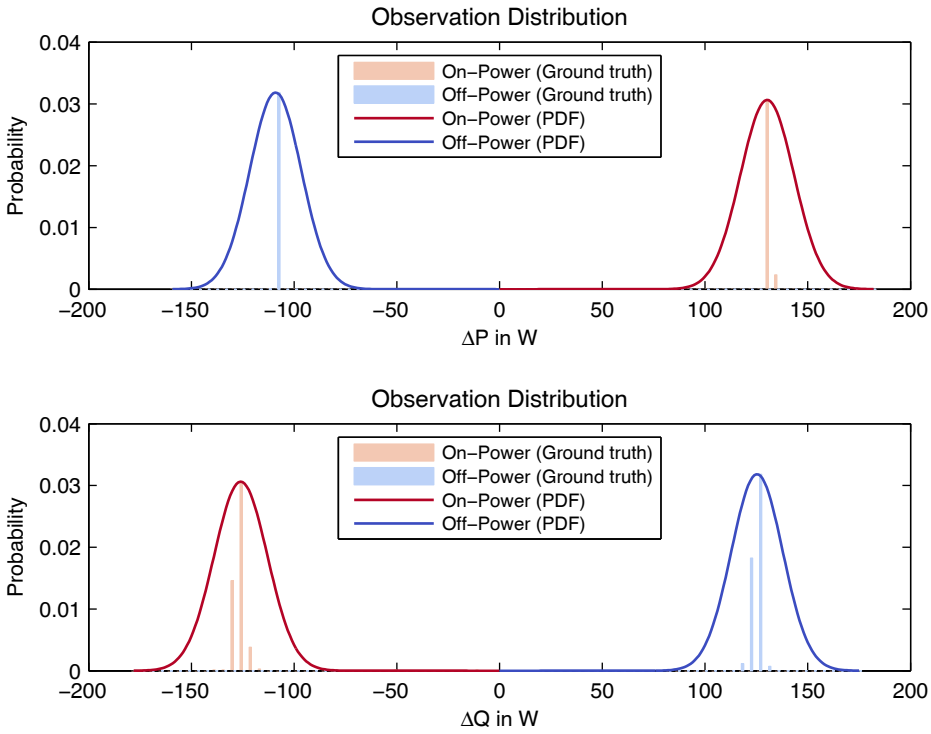

Figure 6.2: Estimated parameters for the observation distributions (active and reactive steady state power changes) for a fridge, estimated distributions compared to ground truth values.

6.3.3 Time Durational Distributions

The time durational distributions are computed from the clustered events $\mathbf{B}_C(D)$ of the precedent section. Also here, the median time duration for each state is computed first. In the case of refrigerator and freezer, the results from Chapter 3 have shown that the durational components of a state vary approximately between ±10% of the median value. To include those findings, durations outside of this range are not part of the time durational distribution computation. The parameters of a Gamma function are estimated via a maximum likelihood computation. For evaluation purposes of the described methodology, the best parameter set $\zeta_j(D)$ computed via the ground truth data is compared to the parameter set $\zeta'_j(D)$ estimated from the clustered data set in the preceding section. The comparison is realized by the likelihood function $\mathcal{L}$ for the gamma distribution where the parameters α and β are represented by the estimated parameter sets $\zeta'_{on}(D)$ and $\zeta'_{off}(D)$ and the quantity x represents the ground truth for the on-and off-duration values:

$$\mathcal{L}(\alpha,\beta|x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{6.1}$$

By comparison, the negative natural log-likelihood of the data set with the predicted parameters $-\ln(\mathcal{L}(\zeta'_j(D) \mid \mathbf{t}_j)$ and the parameters computed from the ground truth data $-\ln(\mathcal{L}(\zeta_j(D) \mid \mathbf{t}_j)$ are comput-

ed as can be seen in Table 6.2. The smaller the number, the higher the probability that the parameter set fits the ground truth data. All estimated parameter sets $\zeta'_j(D)$ for the state duration distributions are less probable than the parameter sets estimated from the ground truth data.

Table 6.2: Comparison of natural log-likelihoods of different parameter sets for the Gamma distribution

Cooling Device	$-\ln(\mathcal{L}(\zeta_{on}(D)) \mid t_{on})$	$-\ln(\mathcal{L}(\zeta'_{on}(D))\mid t_{on})$	$-\ln(\mathcal{L}(\zeta_{off}(D))\mid t_{off})$	$-\ln(\mathcal{L}(\zeta'_{off}(D))\mid t_{off})$
1	3654.1	3995.9	4550.9	4688.5
2	2526.4	3292.3	2588.9	2696.9
3	1843.7	2175.8	2014.0	2059.7
4	1797.8	4907.6	1809.3	5473.4
5	2161.8	2163.4	2790.5	2839.1
6	966.9	972.7	1816.8	4480.5

In order to explain these partly substantial differences in the natural log-likelihood, the power blocks within the extracted cluster have to be considered. As already explained in Section 5.2.2 all events with positive steady state power changes are combined with all feasible events with negative steady state power changes. Even if only power blocks $\mathbf{B}''(D)$ are selected which have a minimum number of s occurrences of events within the start and the end of each block, besides the actual blocks of the device D to be disaggregated there also exist blocks from other devices in $\mathbf{B}''(D)$. If some power blocks caused by events from other devices which have similar durations as well as continuous power levels $P_E(B_m)$ from (5.11), the parameter estimation for the Gamma function is influenced. However, the results show that even in the case when there are two cooling devices within the aggregated load curve a good model parameter estimation can be achieved.

An example of the estimated time duration distributions for a fridge can be seen in Figure 6.3. Both distribution functions covering the main part of the state durations have their peak very close to the peak of the ground truth data. Despite the differences in the natural log-likelihood function, a good approximation of the ground truth data can be realized. The estimated parameter sets $\zeta'_{on}(D)$ and $\zeta'_{off}(D)$ describe the gamma distribution γ.

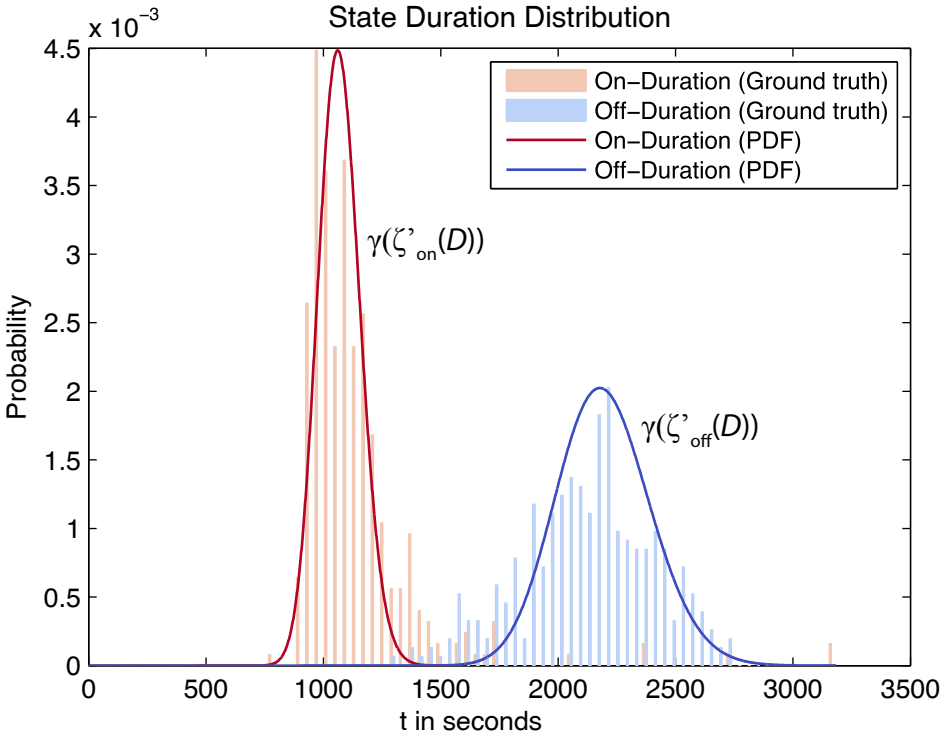

Figure 6.3: Estimated parameters for the duration distributions for a fridge, estimated distributions compared to ground truth values

6.3.4 Reconstruction of the Load Profile

In this section it is investigated in the accuracy for estimating the load profile of a certain device D for a specific state j by the fEEDBACk algorithm. For realising this, the estimated active power values $\hat{P}_j(t, D)$ from (5.23) are compared with the ground truth power values from the device D in a state j at each the time step t. As a standard test case the ground truth (measured) data of a freezer for its on-states is used. For a better comparison of the achieved accuracy, in addition to the estimated load profile of the on-state also one randomly selected ground truth set of on-power values is used in the same manner for computing the distance function in (6.2). This also shows how different the ground truth on-power values are from each other.

For being able to evaluate the accuracy of the estimated load power values, first of all, the ground truth power values for the on-cycles have to be extracted from the measured data. This can be realised, for example, by conducting an event detection on the ground truth data and extracting the power values within the events by simply filtering such sections where the power values are in the range of the on-state. In the specific test case of the freezer, the total number of $a_j(D)$ sets of power values $\mathbf{P}_{i,j}^{\,0}$ for each on-cycle i which are in the range of 115 W to 150 W are extracted.

Due to the reason that only the run of the load profile wants to be compared, the differential power values from (4.1) are used within the distance function d_P. The mean value of all squared distances for each time step t of a single load profile is used for computing the distance function. This is done for each measured set of on-power values $\mathbf{P}_{i,j}^{0}$ and the mean of all values gives d_P.

$$d_P = \frac{1}{a_j(D)} \sum_{i=1}^{a_j(D)} \left[\frac{1}{m_{i,j}} \sum_{t=1}^{m_{i,j}} \left(\Delta\hat{P}_j(t) - \Delta P_{i,j}^{0}(t) \right)^2 \right]$$
$$where\ m_{i,j} = \min\left(\left| \hat{\mathbf{P}}_j(t) \right|, \left| \mathbf{P}_{i,j}^{0} \right| \right) - 1 \tag{6.2}$$

The histogram of the computed distances d_P of a freezer are shown in Figure 6.4. As can be seen, the predicted load profile results in nearly the same distribution of distances as a randomly selected load profile of a freezer. This means that both, the randomly selected as well as the predicted load profile of the on-state have similar distances to the remaining measured load profiles. The estimated load profile may even be a bit more similar to the rest of the measured load profiles since it has a slightly higher proportion in the distance of 0.05. This can be explained by the usage of the median values of several load profiles for each time step. As an example, Figure 6.5 shows a predicted and a randomly selected measured power load profile of the on-state of a freezer. As can be seen, the estimated load profile is smoothed in comparison to a randomly selected load profile of an on-state of a freezer.

Other reconstructed load profiles also have a similar distance measure d_P, as per Figure 6.6. The median value of all distances is typically around 0.05 and 0.25. The boxes in blue show the 25th and 75th percentiles of all computed distances between the reconstructed load profile and all measured load profiles. The whiskers show the most extreme data points. Only cooling device number 2 and number 6 have a greater distance compared to all the other devices, because the measured load profile itself varied within a broader range. Anyhow, the reconstructed load profile of cooling device 2 (see Figure 6.5) shows a great similarity to the measured load profile. This also applies to cooling device 6. Even the greatest distance measurement has a median power difference of about 0.2 W^2 or according to (6.2) 0.4 W per set and sample, what is very low. These results show that through the proposed procedure an accurate estimation of the load profile can be computed.

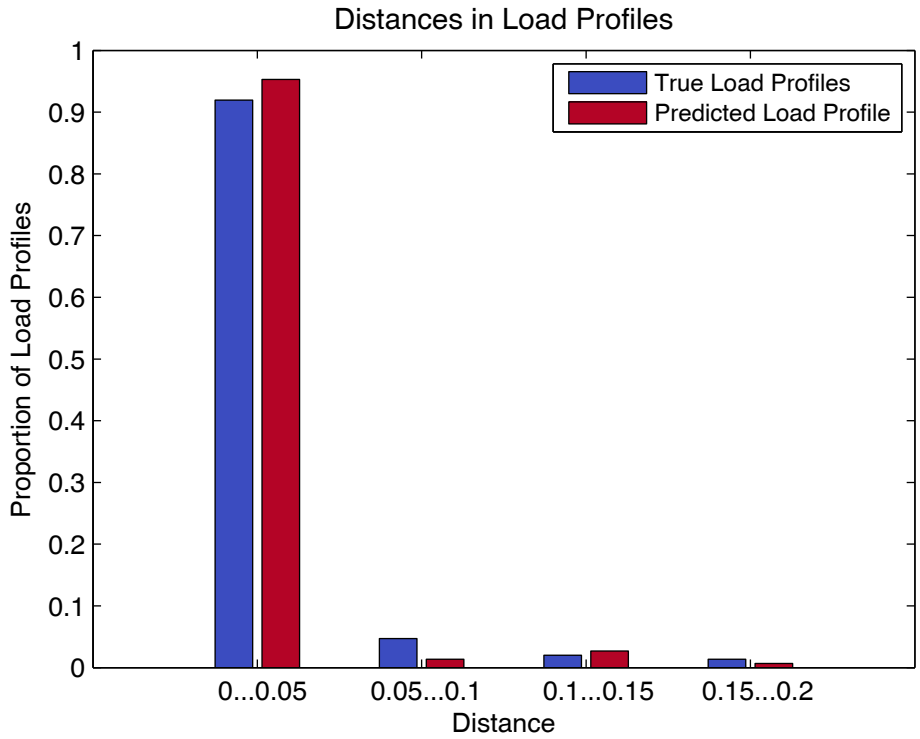

Figure 6.4: Distances between one random selected measured load profile of a freezer as well as the predicted load profile to all measured load profile

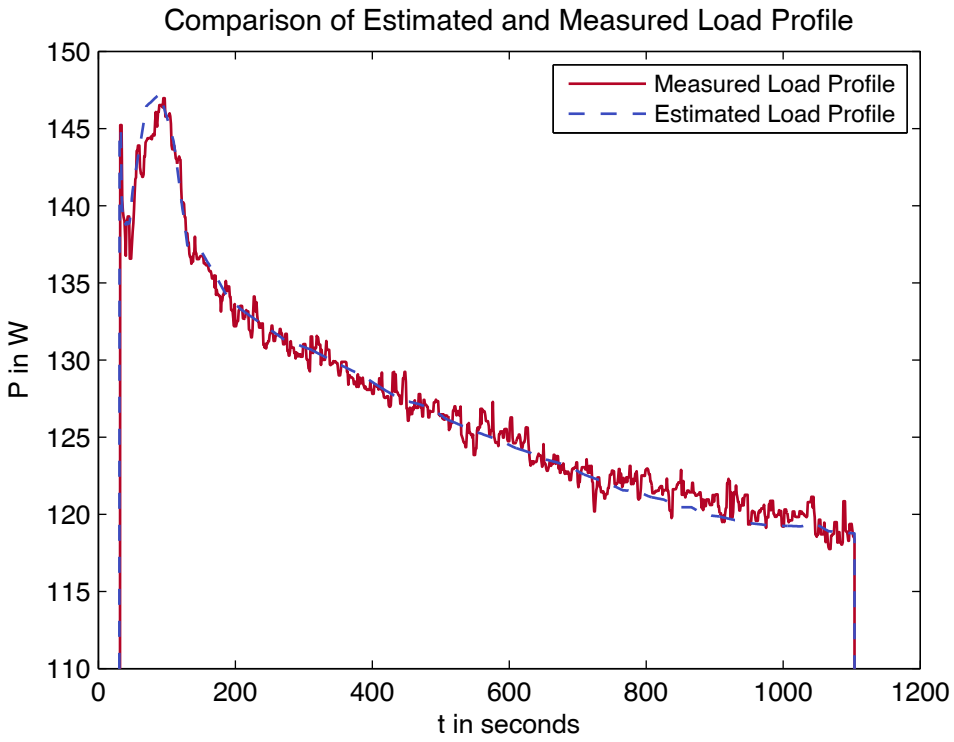

Figure 6.5: Comparison of the estimated and reconstructed load profile of cooling device 2, sampling period:1 second

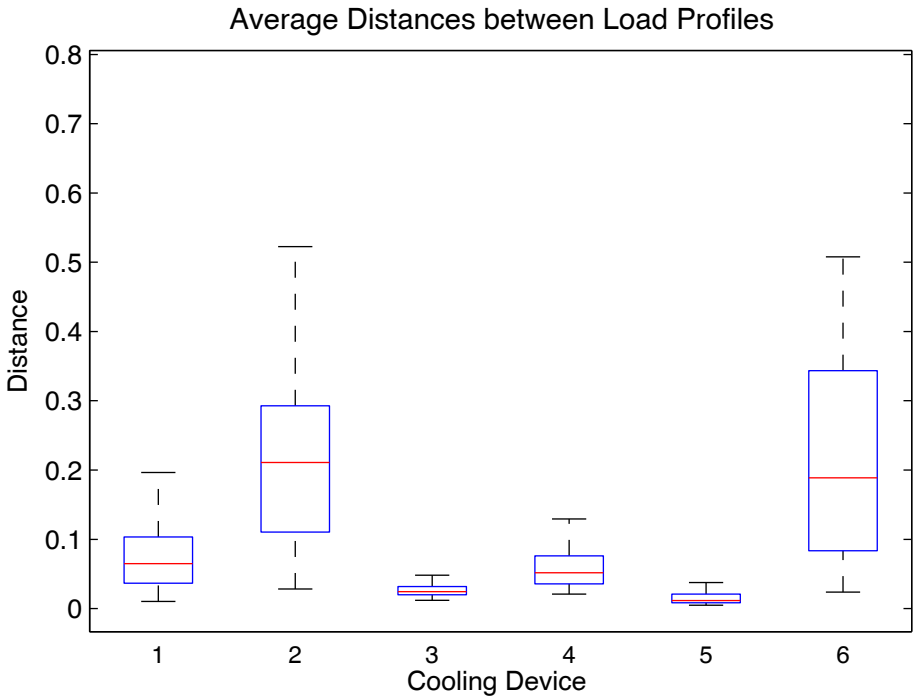

Figure 6.6: Average distances between predicted load profiles of cooling devices to true load profiles, median value (red), 25th and 75th percentile (blue) and most extreme data points (black)

6.4 Performance of the Energy Disaggregation

As described above in this section, the overall performance of the fEEDBACk algorithm is presented.

To demonstrate the performance of the energy disaggregation the load curves of cooling devices are used (see Section 6.1). Beyond this, the most important parameters of the fEEDBACk algorithm are varied in four sensitivity analyses.

6.4.1 Overall Performance

For computing the event detection results, the events within the disaggregated load curve are compared with the ground truth data. Figure 6.7 shows the performance of the event detection through the fEEDBACk algorithm. The true positive ratio (TPR) from (4.15) shows that at least 85% of the events of a cooling device are detected by the algorithm. The number of false positives to the total number of events, the false positive percentage (FPP) from (4.19), varies from approximately from 3% to 20%. The proportion of numbers of events of the cooling device related to the total number of events within the load curve does not correlate. In the case of cooling device 1, the highest proportion of events from the cooling device related to the total events is found, but in contrast the test case has the worst TPR and highest FPP.

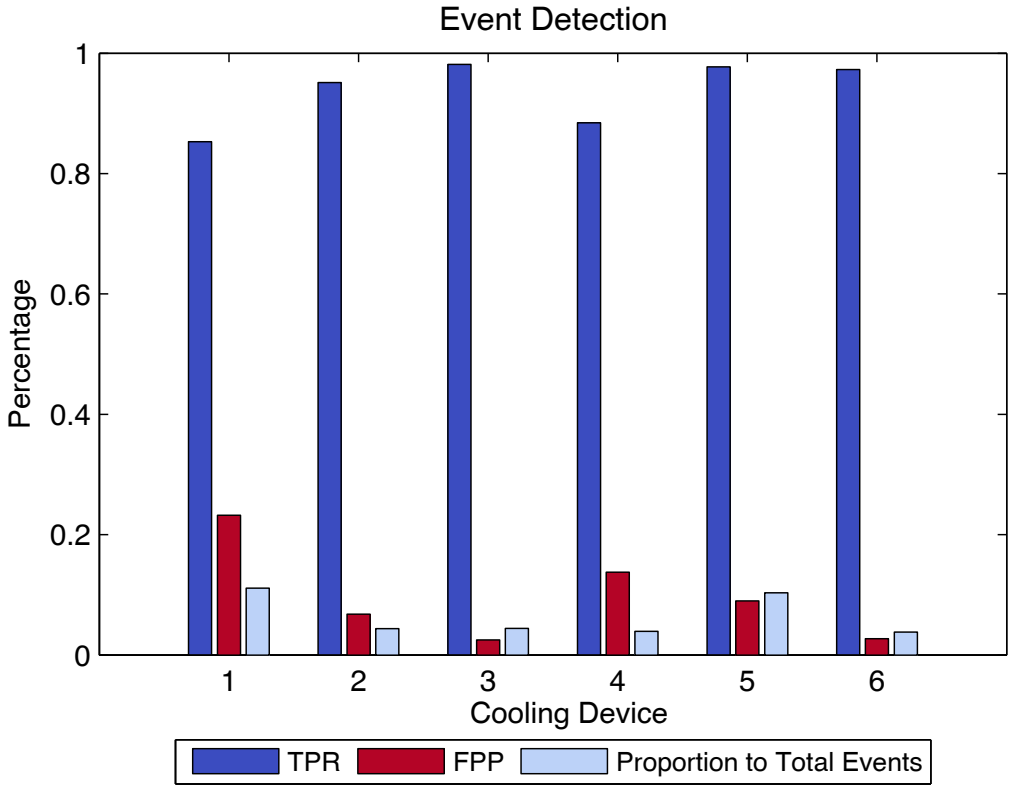

Figure 6.7: Event detection via the energy disaggregation process, true positive ratio and false positive percentage, sampling period:1 second

Metrics for the energy disaggregation process are presented in Section 5.2.7. Figure 6.8 clearly shows that the different metrics provide not always the same results. Especially the metric that compares the fraction of energy to the total amount of energy consumed by a certain device leads to misleading results. As can be seen in Figure 6.8, the fraction of energy is very well estimated, but in contrast the F-Measure varies in a wider amplitude. Reason for this is that the amount of energy can be correctly estimated through random fault detection of power changes. Especially, when the extracted on- and off-durations match the ground truth values, a load curve for the device D with a similar on/off ratio can be estimated which has nearly the same energy consumption as the ground truth. However, even when the amount of energy matches huge differences in both load curves can exist.

The results for F-Measure lie between 0.80 and 0.99 and the values strongly correlate with the results of the event detection of Figure 6.7. Table 6.3 shows the detailed energy disaggregation metrics. In contrary to the F-Measure value which has a maximum score of one, the estimated amount of energy can theoretically become a multiple of the amount of the energy of the ground truth data. The F-Measure value only represents differences in the load curve and not if the disaggregation process over or under estimates the total turn-on time.

The results in Figure 6.8 clearly show that cooling devices can be detected by the fEEDBACk algorithm with a high accuracy.

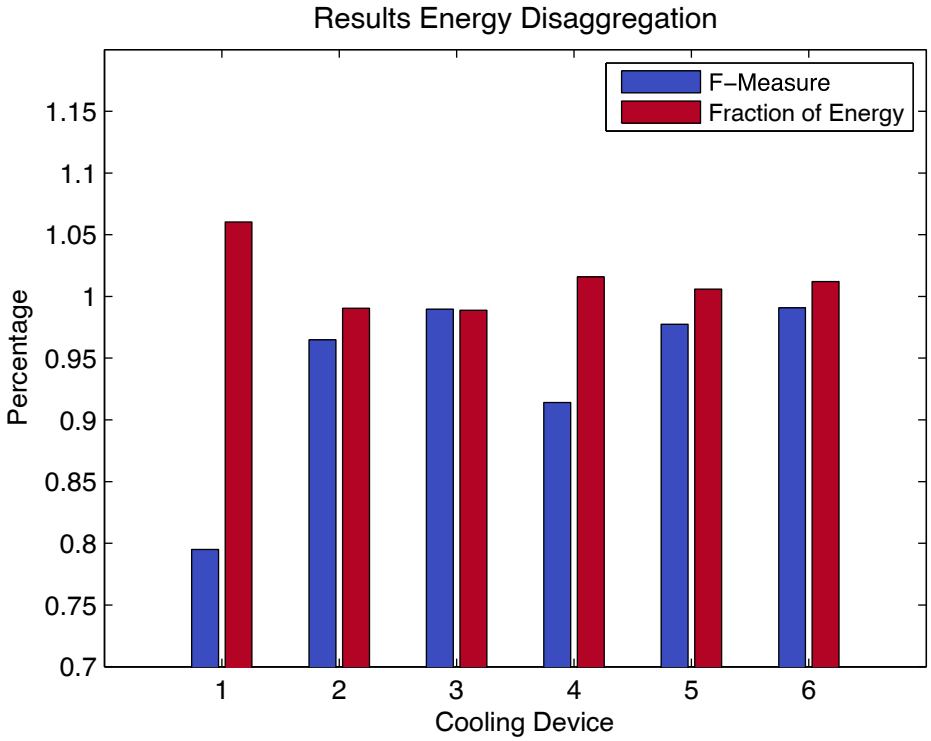

Figure 6.8: Energy disaggregation results for fEEDBACk algorithm under different metrics for several cooling devices, sampling period:1 second

Table 6.3: Overview of energy disaggregation results for cooling devices, sampling period: 1 second

Cooling Device	F-Measure	Precision	Recall	Fraction of Energy
1	0.80	0.81	0.78	1.06
2	0.96	0.98	0.95	0.99
3	0.99	0.99	0.99	0.99
4	0.91	0.91	0.92	1.02
5	0.98	0.98	0.98	1.01
6	0.99	0.99	0.99	1.01

Two samples of the results of the disaggregation process of cooling device 1 and 6 can be seen in Figure 6.9 and Figure 6.10. In both cases short samples of the total aggregated load curves are shown in blue where the sections in red indicate that the cooling device to be disaggregated is on. The green lines show the estimated load curves which are generated by the fEEDBACk algorithm.

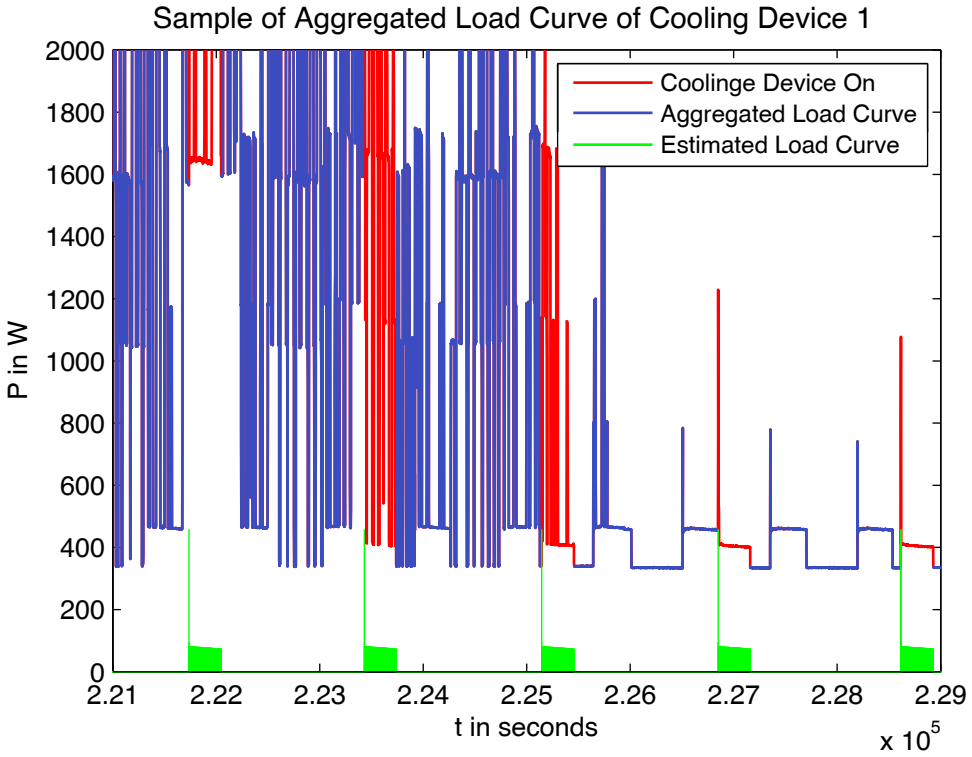

Figure 6.9: Sample of aggregated and estimated load curve of cooling device 1

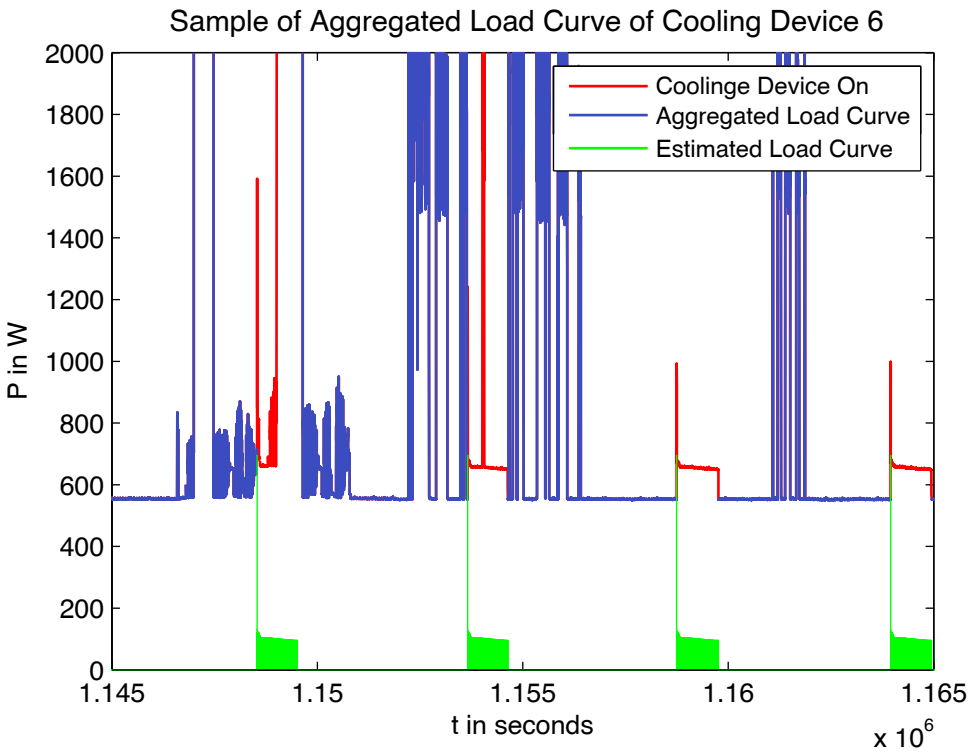

Figure 6.10: Sample of aggregated and estimated load curve of cooling device 6

6.4.2 Sensitivity Analysis 1: Effect of Different Parameter Sets in Event Detection on Energy Disaggregation

As already mentioned, the detection of events has a big influence on the performance of the fEEDBACk algorithm. The reason for this is that the time points of events are used as input for the modified Viterbi algorithm. In Chapter 4, the influence of the selection of parameters for the event detection algorithm is investigated. The findings from these investigations are used for the parameter selection for the event detection in the fEEDBACk algorithm. To verify these results, a variation of the input parameters for the Baranski algorithm is performed in this section. The time duration for merging events τ and the minimum power for detecting a power change of an event P_{min} of (4.2) are varied. The range of parameters used can be seen in Table 6.4.

Table 6.4: Parameter ranges used for verifying the fEEDBACk algorithm

Parameter	Baranski	Step size
	Range	
P_{min}	4-24 W	2
τ	2-4 seconds	1

A total of 33 combinations is used for finding the optimal parameter set. Table 6.5 shows the best results for F-Measure of the investigated cooling devices with different time duration lengths and minimum power values (further details see Appendix). In some cases the F-Measure value can be improved for a maximum increase of about 0.07 through a variation of the parameter set. In most of the cases there is nearly no difference in the F-Measure value regarding the values in Table 6.3. This shows that the results from Section 4.1 are also applicable to other measurement data.
However, the variation of the event detection parameters also shows that the parameter set has a big influence on the total performance of the algorithm. Especially in the test case cooling device 1, a major improvement of the overall performance could be achieved. The reason for this is that the turn-on transient of the cooling device 1 lasts between two and three seconds and the standard parameter of $\tau = 2$ seconds causes wrong detections. For this reason it is more practical to use a duration for $\tau = 3$ seconds for cooling devices.

For a chosen constant duration of 3 seconds a variation of the minimum power value within the range of $P_{min} = 4$ W and 24 W has been carried out. Figure 6.11 shows the average difference of the F-Measure value regarding the best detected F-Measure value for each test case by a constant duration of $\tau = 3$ seconds. Between 10 and 24 W the F-Measure value just barely decreases by approximately 0.005. This shows that the influence of the selection of the power value P_{min} is just minor when it is greater than 8 Watts. The results are in line with the optimal range of parameters from Chapter 4 (see Appendix).

Above all, it has to be considered further that for other electrical devices, which for example have greater steady state power changes, another parameter set can lead to a better performance. But also for these devices it can be assumed that the power value P_{min} will not have a very strong influence on the performance of the energy disaggregation process if it is varied between certain ranges.

Table 6.5: Overview of energy disaggregation results for all cooling devices with different window lengths for event detection, differences to best F-Measure values, sampling period: 1 second

Cooling Device	τ	P_{min}	**F-Measure**	
-	**Seconds**	**Watt**	**Result**	**Difference**
1	2	22	0.87	0.07
1	3	10	0.87	0.07
1	4	10	0.87	0.07
2	2	4	0.97	0.01
2	3	4	0.97	0.01
2	4	10	0.97	0.01
3	2	24	0.99	0.00
3	3	24	0.99	0.00
3	4	24	0.99	0.00
4	2	14	0.92	0.01
4	3	18	0.92	0.01
4	4	18	0.92	0.01
5	2	14	0.98	0.01
5	3	14	0.98	0.00
5	4	14	0.98	0.00
6	2	14	0.99	0.00
6	3	14	0.99	0.00
6	4	16	0.99	0.00

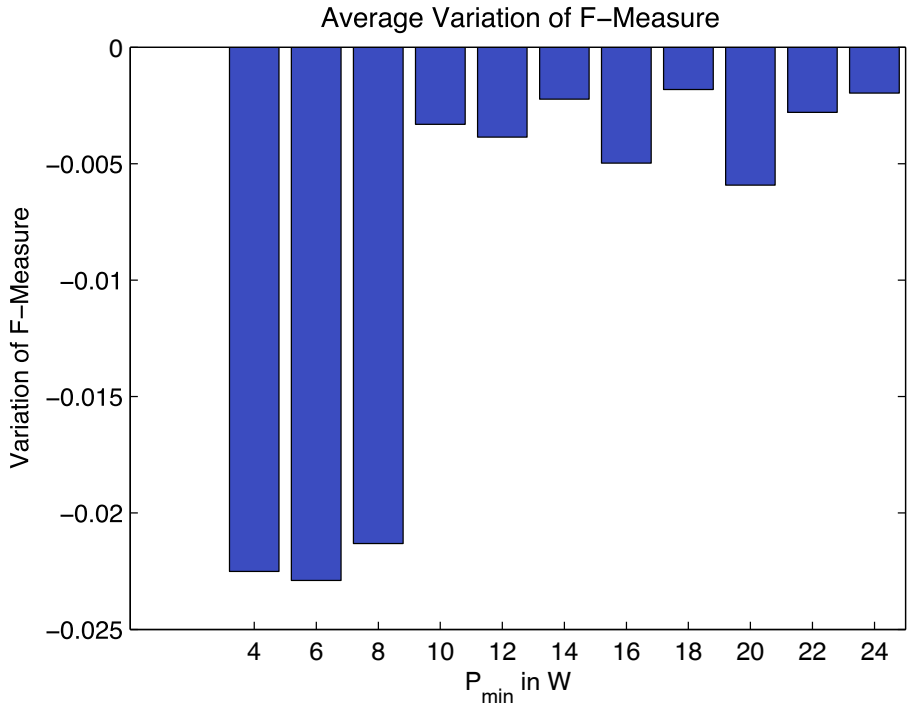

Figure 6.11: Average variation for all cooling devices to the best F-Measure value for a constant $\tau = 3$ seconds, sampling period: 1 second

6.4.3 Sensitivity Analysis 2: Effect of Incorporation of Reactive Power Values on Energy Disaggregation

The incorporation of reactive power values allows the usage of an additional feature for the energy disaggregation process. Electrical devices which consume reactive power will especially benefit from the incorporation of reactive power values. The reason for this is that if the active steady state power changes are superimposed by other electrical devices, the reactive power steady state power change remains at the previous level. This is only true if the device causing the superimposition has a low or ideally no reactive power consumption.

Figure 6.12 shows the results when the reactive power values are incorporated in the fEEDBACk algorithm. Most notably, the accuracy of the algorithm increases for those test cases where a relatively low F-Measure value is the starting point.

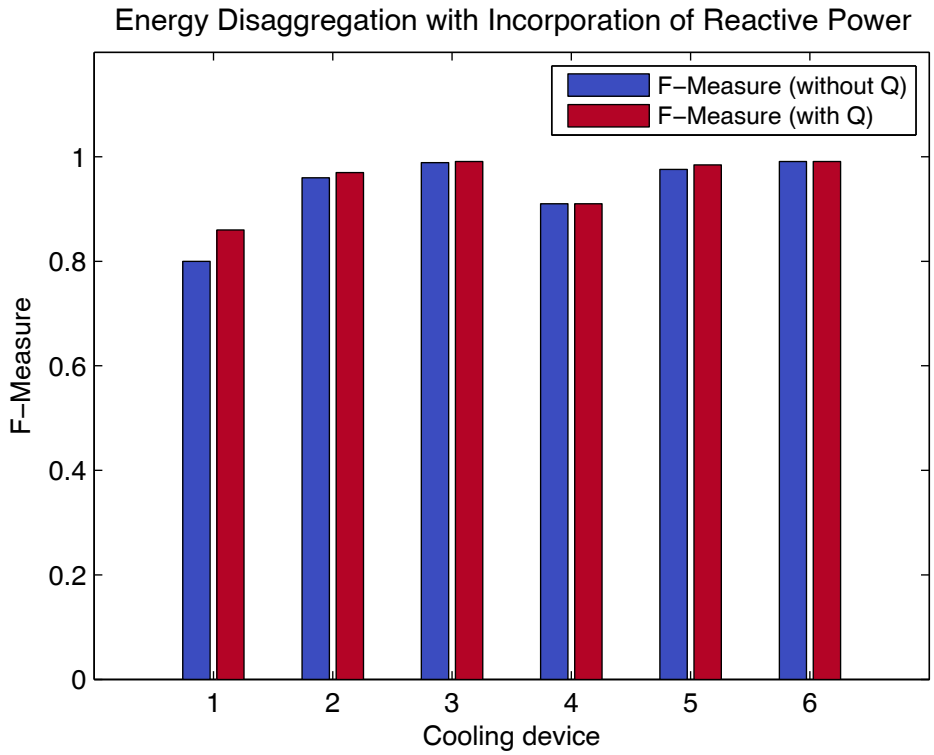

Figure 6.12: Energy disaggregation results for fEEDBACk algorithm with incorporation of reactive power values, sampling period:1 second

Table 6.6 provides detailed information about the results. As can be seen, the F-Measure value increases by 0.06 in the best case. For test cases that already have the F-Measure value above 0.90, just minor improvements of approximately 0.01 can be achieved.

Table 6.6: Overview of energy disaggregation results for cooling devices with and without incorporation of reactive power, sampling period: 1 second

Cooling Device	Without Q	With incorporation of Q
	F-Measure	F-Measure
1	0.80	0.86
2	0.96	0.97
3	0.99	0.99
4	0.91	0.91
5	0.98	0.98
6	0.99	0.99

6.4.4 Sensitivity Analysis 3: Effect of Different Sampling Periods on Energy Disaggregation

One of the greatest influences on the performance of energy disaggregation algorithms is the sampling period of the measurement data to be disaggregated. The lower the sampling period, the fewer features can be extracted from the load curve. To show the effect of the sampling period on the energy disaggregation process, four further sampling periods (2 seconds, 5 seconds, 10 seconds and 20 seconds) are compared to each other. To gather these sampling periods, the original power measurement data with a sampling period of one second is averaged to the corresponding ones.
Figure 6.13 shows the F-Measure values for the single sampling periods for different cooling devices. In all test cases, a decrease in performance can be seen when the sampling period rises. In the test case cooling device 1, which already has the worst F-Measure value, the biggest decrease in F-Measure from 0.80 (1 second) to 0.65 (20 seconds) occurs. Table 6.7 shows the detailed F-Measure values.

Interestingly, all test cases which already have a good performance result in just a minor decrease by a maximum of 0.05. One of the main reasons for this is that in the test case cooling device 1, a fridge and a freezer can be seen in the measured load curve. Since cooling devices have similarly on-durations as well as steady state power changes in the disaggregation process, some errors occur. Additionally, because an increase in the sampling period leads to a decrease of detectable events (see also Section 4.1.4), there are more superimposed power changes. Beyond this, it is unfeasible to estimate the true steady state power changes since the number of samples between the events also decreases. Therefore, the computed steady state power changes do not allow such a good differentiation between the events from different devices. This also makes it impractical to find the correct time points where a state change of a certain device occurred. As a consequence, the time points of the predicted state changes vary from the true time points by a couple of samples.
In test cases where the F-Measure value remains relatively stable, the other devices which were measured within the load curve differentiate more strongly from the behaviour of a cooling device. This leads to a smaller decrease in the performance as has already been shown.

Table 6.7: Overview of energy disaggregation results for cooling devices for different sampling periods

Cooling Device	F-Measure				
	1 second	2 seconds	5 seconds	10 seconds	20 seconds
1	0.80	0.81	0.75	0.75	0.65
2	0.96	0.95	0.95	0.93	0.92
3	0.99	0.99	0.99	0.98	0.97
4	0.91	0.89	0.89	0.90	0.86
5	0.98	0.97	0.96	0.94	0.93
6	0.99	0.98	0.99	0.98	0.95

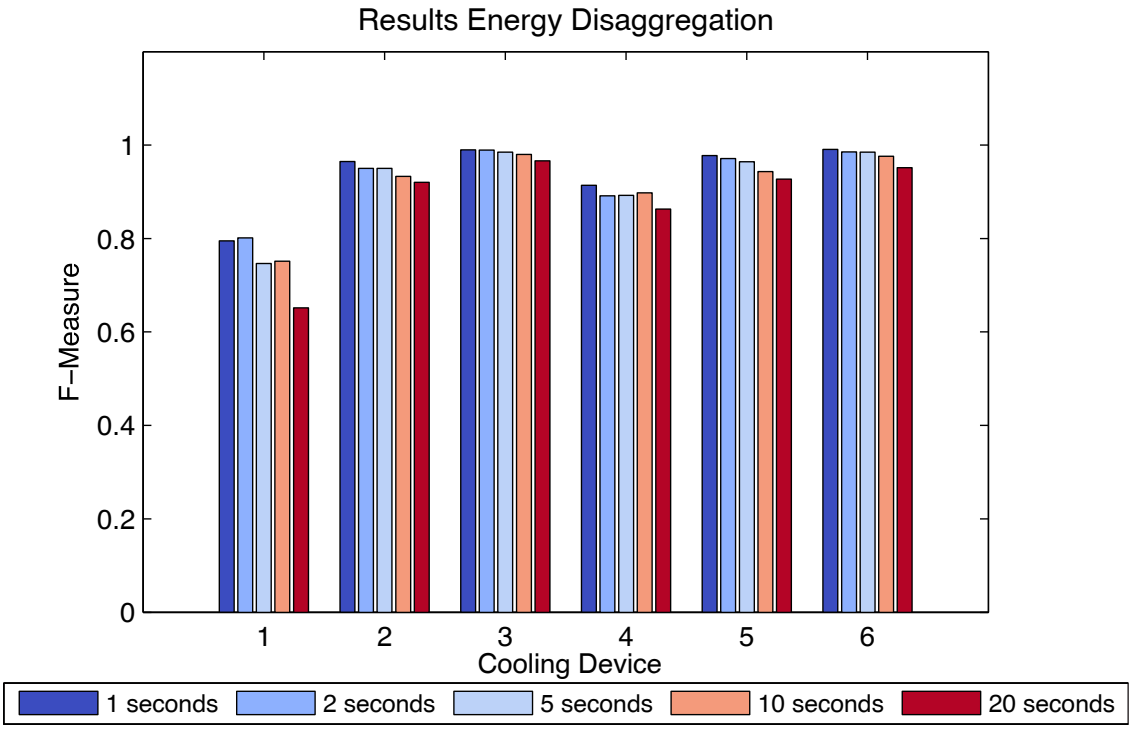

Figure 6.13: Performance of the fEEDBACk algorithm under different sampling periods

Another important fact in increasing the sampling period is that the parameters for computing the steady state power change, and computing the clusters via the DBSCAN algorithm have to be adapted. For the computation of the steady state power changes, smaller window lengths are necessary due to the decreased number of samples between the events. If the window size is smaller, a shorter time span is evaluated for getting the pre- and post-event mean values. For the sampling periods above 5 seconds, a window length of $w_{Pre} = w_{Post} = 2$ samples is used.

But the distance measurement (5.14) used for the DBSCAN algorithm also has to be adapted to be applicable for such long sampling periods. As opposed to the steady state power change values, the sampling period is a discrete variable. This leads to a fixed minimum distance between two samples with the same power value within the Pt-plane of $\mathbf{B}''(D)$. To compensate this fact, similarly to c_P, a factor c_t that depends on the time is necessary for the second term in (5.14). The factor c_t can be computed with:

$$c_t = \begin{cases} \dfrac{0.85 \cdot \varepsilon}{t} & if \dfrac{1}{t} \geq 0.85 \cdot \varepsilon \\ 1 & otherwise \end{cases} \tag{6.3}$$

Through this procedure the samples within the Pt-plane of $\mathbf{B}''(D)$ can be formed by the DBSCAN algorithm to the correspondent clusters of the certain devices to be disaggregated. The factor c_t is also a necessity for state durations of only a few samples.

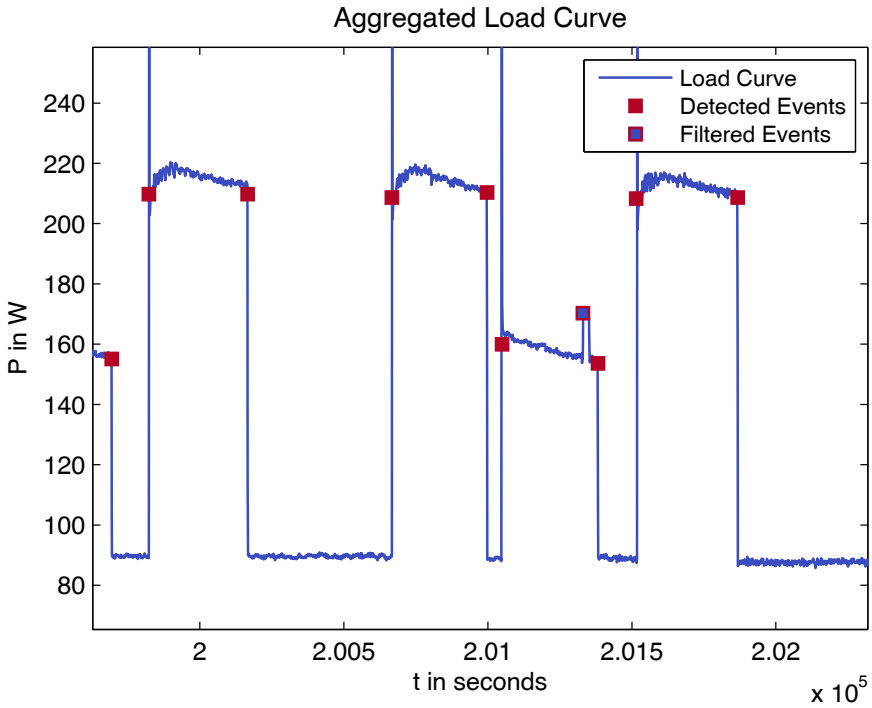

Figure 6.14: Part of the aggregated load curve of test case cooling device 1 with detected and filtered events, sampling period: 1 second, compare with Figure 6.15

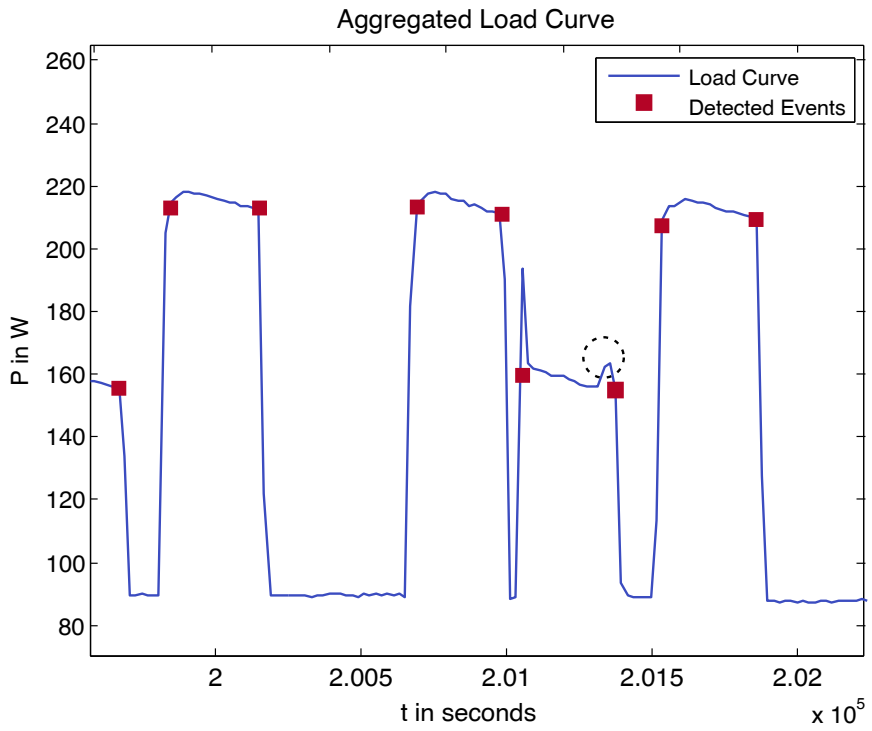

Figure 6.15: Part of the aggregated load curve of test case cooling device 1 with detected and filtered events, sampling period: 20 seconds, compare with Figure 6.14

Moreover, the needs as well as the opportunities for energy disaggregation with a different sampling period strongly vary. For example, in the case of a sampling period of one second there are a lot of events caused by other devices (in comparison to the device D to be disaggregated) which need to be filtered out to get a better performance by the Viterbi algorithm. Such power blocks typically have an on-duration of up to approximately 30 seconds and have similar values for the on and off steady state power changes. Additionally the power values before and after such power blocks usually have the same level. These events can be easily filtered out by checking the values of the steady state power changes, the maximum duration of the correlated power block as well as the similarity between the mean values before and after the power block (see Figure 6.14).
If the sampling period increases, these power blocks can no longer be seen within the load curve, compare Figure 6.14 and Figure 6.15. The filtered event from Figure 6.14 is no longer detected within the sampling period of 20 seconds (see dashed circle in Figure 6.15). The reason for this is that the required power change does not reach the minimum value P_{min} from (4.2), which is a prerequisite for the detection of events with the Baranski detector. As a consequence, no filtering procedure is necessary in those cases.

But also aspects of the load curve, such as the individual load profile of a certain device, get averaged and the specific details are reduced. Figure 6.15 clearly shows that, for example, the turn-on transients of the load profile of the cooling device 1 (which has the greater steady state power changes) are eliminated through the averaging process to get greater sampling periods. But also the number of samples between the events makes it more difficult to compute the correct steady state power changes.

6.5 Maximum Sampling Period for Detecting Appliances

This section analyses the maximum sampling period where it is feasible to detect cooling devices in aggregated load curves.
For demonstration purposes a load curve of a fridge is used because it is a electrical device which is turned on regularly and often has relatively constant on- and off-durations. These more or less fixed durations make the effect of varying steady state power changes (see below) more obvious. It will be shown that even these constant durations cannot be detected anymore when the sampling period is above a certain limit.

As already be shown in the preceding section, the performance of the energy disaggregation algorithm decreases with increasing sampling periods. By averaging the aggregated load curve, fewer samples contain the same amount of information about the power changes of the comprised electrical

devices. This information loss goes along with a loss of features such as detectable events within the aggregated load curve. But also the number of electrical appliances which are simultaneously used contribute to the steady state power changes of the events. This also means that besides the loss of the quantity of the features also the quality of the features for the disaggregation process decreases. These facts makes an energy disaggregation process more complicated or even infeasible and impossible. However, at a certain sampling period the disaggregation process ends up with the impracticality for being able to detect the individual devices in the load curve at all.
This time point where it is impractical to detect electrical appliances depends on several factors. For example, appliance-specific characteristics such as on-durations or on-power values strongly influence the mentioned time point where an energy disaggregation is impractical anymore.

In Chapter 3 the detailed analysis of specific appliances shows that, for example, the on-durations of the investigated electrical appliances have their maximum probability in a range of about 10-20 minutes. Averaged power values with a step size of 15 minutes, which are typically recorded by smart meters in Austria, only allow to accurately detect on-durations with a minimum duration of 15 minutes if the turn-on of the device match the start of the 15-minute average window. Only in this special case, the on-duration is reflected by one 15-minute power sample. In all other cases, the on-power values are averaged to two power samples. Therefore the exact time duration of a specific device is not traceable. Even if the on-duration doubles to 30 minutes, up to 3 power samples are usually used to represent the power changes of the device.

Above all, the steady state power changes of a certain device are also averaged, and can not be extracted from the 15-minute load curve. The exact time point of the window for the averaging process influences the exact power value readings for the certain device which can be measured. It is obvious that as a consequence the measurable power values from the same type of device strongly vary due to the averaging process of the active power values.
An example for the variation of the varying power changes of a fridge can be seen in Figure 6.16. It can clearly be seen that besides the varying power change values, also the on- and off time durations strongly differs between the load curves with a sampling period of one second and 15 minutes.

These facts makes it impossible to extract the true appliance-specific characteristics such as steady state power changes.

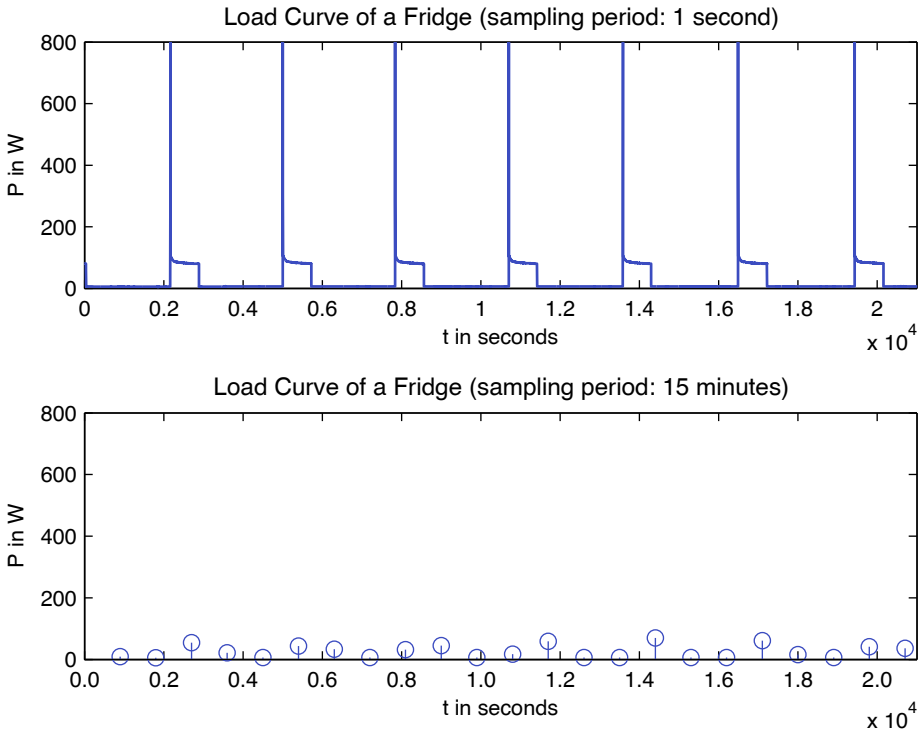

Figure 6.16: Load curve of a fridge with a sampling period of 1 second and 15 minutes

Furthermore a detection of two distinct types of devices within the two above mentioned samples can be infeasible. If, for example, a load with a great on-power value such as a water kettle with a power rating of 2 kW is switched on for one minute, it consumes an energy of about 120 kWs. Nearly the same amount of energy can be consumed by a freezer with an average on-power of 130 W with an on-duration of approximately 920 seconds (~15,3 minutes). Both examples can result in a similar power change of about 130 W within the load curve for one single sample. But also other devices which have different power values and on-durations affect the load curve. The aggregated load curve of the load curve of the fridge which can be seen in Figure 6.16 is shown in Figure 6.17. It is obvious that the load curve of the fridge which can be easily seen in the upper part of Figure 6.17 can no more be detected accurately as compared to the window with a sampling period of 1 second in the lower part of Figure 6.17. However, due to certain characteristics of cooling devices such as the periodically turn-on, in special cases where no other electrical appliances are switched on, the periodically power spikes within the load curve can hint to the fact that there might be a fridge or freezer installed in this specific home. These power spikes can be seen at the right side of the lower part of Figure 6.17.

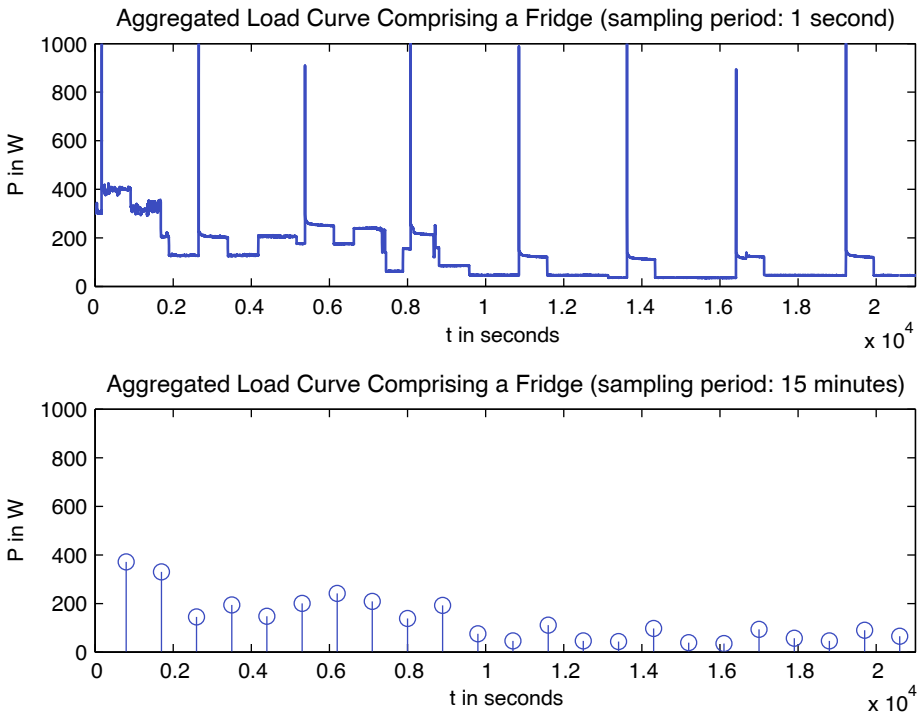

Figure 6.17: Aggregated load curve comprising a fridge with a sampling period of 1 second and 15 minutes

Other electrical devices are not switched on with such a regular pattern in comparison to cooling devices (see Chapter 3 and [73]). Depending on the specific type of device also the on-time is much shorter (e.g. water kettle, microwave) or much longer (e.g. personal computer, TV-set).
Above all, a lot of devices are used simultaneously by the consumers. Amongst other things, this includes devices installed in the kitchen or in the living room. This means that generally a couple of electrical devices are switched on simultaneously or subsequently for a certain period of time when a consumer is at home. An example where several appliances in the kitchen are used by a consumer can be seen in Figure 6.18. The figure shows that most of the used devices have a relatively short on-duration of a few minutes. By increasing the sampling period to a value of 15 minutes only a rise in the power consumption can be detected.
From this, it is apparent that one can not distinguish anymore between the power changes of the corresponding devices when the sampling period is 15 minutes. This statement is validated by the works of many different authors who proposed various energy disaggregation algorithms typically using sampling periods in the range of several kHz up to about a few minutes [31]-[41], [73]-[86]. Kolter et al. [87] use an sparse coding approach with sampling periods of one hour to disaggregate load curves of specifically selected appliances. Even when all loads within the load curve are known the average accuracy of the predicted energy for the devices to be disaggregated is below 50 % [87]. This further emphasis the fact that such sampling periods are not applicable for correctly detecting appliances.

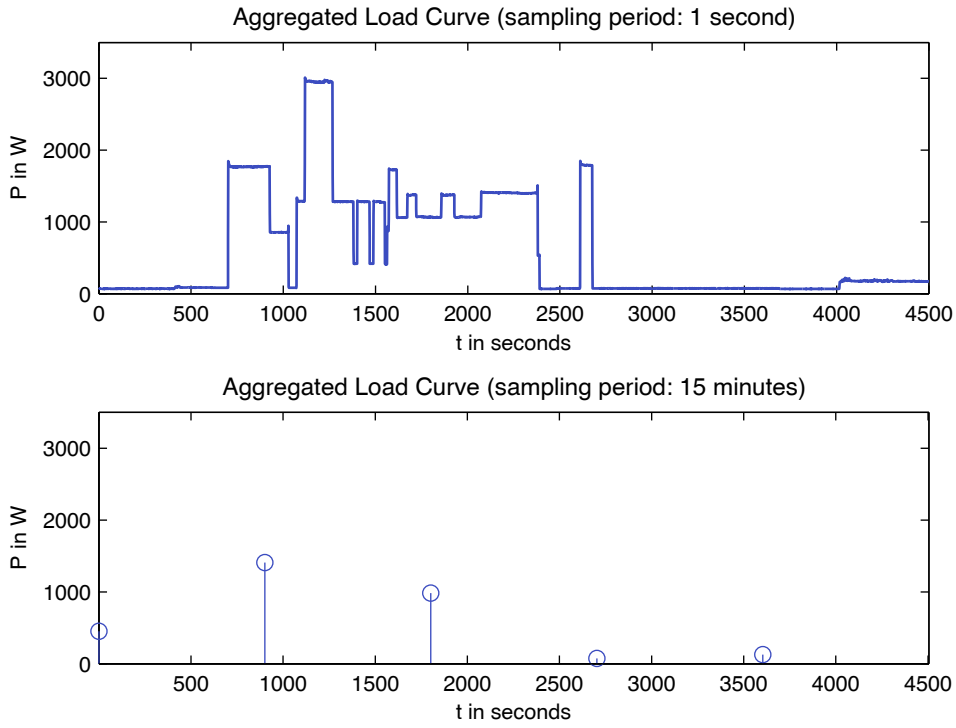

Figure 6.18: Aggregated load curve comprising several electrical appliances from the kitchen with a sampling period of 1 second and 15 minutes

However, one possible exception is a device with regular or semi-regular time of use and a more or less constant power consumption above several hundreds Watt which is switched on at time points where a constant power is drawn by the aggregated load curve (e.g. after midnight). The on-duration of this device has to be at least approximately 30 minutes or 2 sample. As a further restriction a more or less regular power pattern has to be drawn at regular or semi-regular intervals. At present from the point of view of the author, there exist only extremely few classes of electrical devices which have the above mentioned characteristics. These include, among others, fixed installed electrical heating systems such as night-storage heaters, or electric hot water boilers. These devices usually draw a constant power of several kilowatts for a period of up to several hours. Traditionally such devices are controlled by the network system operator via power line communication (PLC).

6.6 Overview of the Performance, Applicability and Limitations

6.6.1 General Issues

There are various methods to save energy in a household, to reduce the total power consumption, and as a consequence, decrease the carbon footprint. The persistence of energy saving actions plays a major role in achieving such results. Two significant prerequisites can be achieved simultaneously by monitoring the appliance-specific power consumption in a household. On the one hand, the biggest influencing factors for saving energy due to the share of the total power consumption can be identified. On the other hand, automatic energy saving advices as well as changes in the behaviour for realizing energy savings can be detected and a consumer can be informed automatically. Up to now there are no automated appliance-specific measurement devices available on the market that are affordable, easy to install and as efficient as possible in terms of time. Through the usage of smart meters in individual homes in the European Union, a measurement device will be available which provides load curves where the method of energy disaggregation can be generally applied. The output data of the smart meters, especially the sampling period, plays a major role in the performance of energy disaggregation algorithms.
The integration of new functionalities into electrical devices such as network support usually goes along with a more complex load curve. This also makes the energy disaggregation process more challenging. But also the integration of renewable energy sources in an individual home such as a photovoltaic panels typically alters the total load curve through the standard installation, and so smart meters can only detect the difference between the fed power values and the consumed power values from the grid as an output.
Above all, specific attention should be paid to data protection and privacy concerns that have arisen in the last couple of years. This is an important issue that contributes to the success of an appliance-specific load monitoring system.

6.6.2 Performance of the Method

A prerequisite for applying the fEEDBACk algorithm on a certain device which has recurring turn-on and off-events is that the measurement period T is great enough to enable the Pt-plane of $\mathbf{B}''(D)$ to form dense clusters which can be identified by the DBSCAN algorithm. This depends on the amount of filtered power blocks $\mathbf{B}''(D)$ as well as on the variation of the corresponding power values $P_{\mathrm{E}}(B''_m)$ extracted from the aggregated load curve. Results have shown that, for example for a fridge, a time duration of approximately ten days is more than enough to extract the relevant appliance-specific characteristics. But also smaller time spans such as two days can be used to successfully extract appliance-specific parameters. These newly detected events and power blocks could, for example, easily be used to relearn the parameters of a device, especially when the state durations of a certain

device change. For example, if the cooling temperature of a device is modified, a relearning of the model parameters is necessary. This can be realized by applying the learning procedure on several parts of the aggregated load curve to be disaggregated.
For applying the DBSCAN algorithm, dense clusters are necessary (see Section 5.2.4). If from the measurement interval, for example, just a few relevant blocks can be extracted, and if these blocks belong to a device with a variation of significantly more than ±10% in the on-state duration, the parameter ε of the DBSCAN algorithm needs to be increased to successfully detect such a cluster. But by the increase of ε also close blocks from other devices can be assigned to the cluster of the device to be disaggregated. The easiest solution for this issue is to increase the measurement interval T. If this is not applicable, another possibility is to use a different clustering algorithm, e.g. hierarchical clustering. Since the operating time of the energy disaggregation is assumed to be several months or even longer, the described DBSCAN usually meet its objectives.

Another influencing factor on the clustering procedure is the overlap in the Pt-plane with a device that has the same behaviour as the device to be disaggregated. In order to form overlapping areas, the two devices must have similar power on-changes as well as on-durations. Usually this is a very rare condition but can occur when there are two devices of the same type within the aggregated load curve. For example, in the special case when there is more than one fridge or freezer in the aggregated load curve, a clustering procedure can only be realized if at least one of the parameters (on-power value or on-duration) strongly differs from the other. For the power values, a difference of at least 15% from the bigger power value and for the on-duration, a difference of more than 50% of the median value of the greater on-duration is necessary. Alternatively, a third feature such as the reactive steady state power change can be integrated to form clusters which are not density-reachable and are possible to be clustered separately.

Above all, the sampling period also influences the learning of the parameters for the model. For measurement periods of up to 20 seconds, an energy disaggregation can be realized as has been shown in Chapter 6. If the state durations of certain devices are shorter than or equal to the sampling period, a disaggregation process is not possible. The reason for this is that in the said case, no exact state duration distributions can be extracted due to the concentration of the blocks in such a small area.
An effect on the performance of the algorithm can be noticed at sampling periods greater than 1 second. The reason for this is that the correct detection of steady state power changes by the event detection algorithm cannot be realized in all cases. Especially when the sampling period exceeds 10 seconds, more and more events occur within a few samples and it is very hard or sometimes nearly impracticable to gather the true value of the steady state power change. This also effects the least power block search, since this method is based on combining positive and negative steady state pow-

er changes. In special cases, greater sampling periods can result in opposing signs of the steady state power change compared to the true value. This happens when a turn-off of the device *D* is superimposed by a big positive power change of another device. Since only positive power changes are combined with negative power changes in the least power block search, this strongly influences the results. A solution for this would be the combination of all feasible events for greater sampling periods, which would increase the computational effort. Anyhow, this is a relatively rare occurrence if the sampling period is approximately below 10 seconds, but it rises with increasing sampling periods.

Results have shown that the model parameters successfully extracted by the DBSCAN algorithm are very close to the real values (see Section 6.3). The parameters such as steady state power change as well as the estimated state durational distribution are necessary to achieve a good performance by the algorithm. But even if the parameters slightly deviate from the true values, a good overall performance can be achieved. This is realized by the combination of the modified Viterbi algorithm and the selection of relevant events within all detected events. Especially through the filtering of events, which are improbable for a certain device *D*, the accuracy can be raised significantly.

The results gathered from the modified Viterbi algorithm are used to reconstruct the load profile of a specific state as a first step. The described reconstruction works well for sections within the load curve where the power fluctuations of other devices are limited. If there are a lot of power fluctuations in the section of the load profile of a specific state to be extracted, the results can get strongly influenced. This procedure is limited to devices which have a relatively constant power value within a specific state. For electrical appliances with a strongly varying power consumption within a state, such as TV-sets, the extracted load profile can vary within a larger range compared to the true values.

6.6.3 Applicability and Limitations

The fEEDBACk algorithm is based on the detection of events in an aggregated load curve. These events are used to find power blocks coming from the devices to be disaggregated. The proposed fEEDBACk algorithm works well for electrical appliances with a recurring pattern in their load profile.
The basic principles of this algorithm can be applied to other electrical devices such as dishwashers or water kettles which are regularly turned on and off. For example, in the case of a washing machine a filtering procedure can be realized to find the recurring events of the power changes of the motor for rotating the drum. The state durations of the motor are relatively constant and can be easily found within all detected power blocks. The described learning procedure that includes the DBSCAN algorithm can be used to extract the specific model parameters of the power changes from the motor of the rotating drum. After the filtering of these specific power blocks within the aggregated load curve,

other present states such as the water heating process or the spinning can be found by limiting the search range to a certain time span before and after the recurring power cycles of the drum. When all these relevant sections within the load curve are found, the parameter learning can be realized by evaluating the detected power blocks and utilizing the appliance-specific unified model. But the described method can also be applied to devices which are not frequently turned on and off. For better clustering and detection of such devices, the weekday as well as the time of day can also be incorporated as additional features to the extracted power blocks.

However, the proposed algorithm is not applicable for all types of appliances. For example, devices which are constantly switched on and consume a constant active power cannot be disaggregated with the proposed algorithm. The reason for this is that the proposed method is predominantly based on power changes that do not occur or are just within a few Watts and therefore are not traceable.

All unsupervised algorithms have the limitation that these algorithms can just extract electrical devices from the aggregated load curve if the load profile and the general behaviour of the type of appliance is known in advance. Therefore it is necessary to use unified appliance models to be able to disaggregate the load curve to its origins.

7 Conclusions

7.1 General Conclusion

This thesis presents a novel method (fEEDBACk) for unsupervised energy disaggregation for domestic consumer households. For this purpose, the measurement data from 40 households (ADRES) are analyzed to conduct the characteristics of appliance-specific usage patterns. Furthermore, different event detection algorithms are compared with each other and investigations into finding the parameter sets lead to the most accurate results. Above all, an algorithm for computing the steady state power changes is proposed, which improves the computation accuracy. These principals form the basic framework for the fEEDBACk algorithm.

7.1.1 Appliance-Specific Characteristics

Four typical electrical appliances (fridges, freezers, dish washers and washing machines) used in the residential sector (see Chapter 3) are investigated. The results of the analysis show that each individual category of electrical devices goes along with specific power consumptions, on- and off-state durations as well as time of use.

Due to the fact that fridges and freezers have a similar mode of operation, the detected characteristics are also analogical. Interestingly, the on- and off-durations of an individual fridge or freezer vary just within a certain range of their median value. As opposed to this, the specific on-durations of a single washing machine strongly vary in its usage cycles. This results from the different washing programs used within the households and the fact that the absorbed energy depends on the used capacity.

In the case of the dishwasher, one frequently used period for the heating cycle can usually be found for a single device. Above all, the time of use during, for example, a day is another feature that can be utilized for unsupervised energy disaggregation if the load profiles of certain devices are relatively similar.

7.1.2 Event Detection

The accuracy of selected event detection algorithms applicable for sampling periods between 1 and 5 seconds are investigated in this thesis (see Chapter 4). For the analysis, a publicly available labelled data set was used. Results show that the most accurate event detection is realized by the algorithm

of Baranski. This approach leads to the most accurate results for the investigated range of sampling periods. If the time duration of the sampling period rises, more and more events are superimposed, which leads to fewer detectable events within the load curve. The total score for the event detection algorithm of Baranski varies within 0.76 and 0.5 when the sampling period is increased from 1 to 5 seconds for the best metrics. The influence of the event detection accuracy is also investigated when reactive power values are incorporated to the event detection process. Through this integration, the number of false positive detections is significantly reduced. However, this goes along with a rapidly increased number of missing events. The reason for this is that just a couple of events also go along with the reactive power value required in the selected detection process.
Beyond this, in the computation of the steady state power changes of the detected events is investigated. An algorithm using variable pre- and post window lengths for the computation of the steady state power changes is introduced. The comparison to other state of the art algorithms shows that the proposed algorithm results in a more accurate computation regarding the true power change values.

7.1.3 fEEDBACk Algorithm

This unsupervised energy disaggregation method is based on using hidden semi-Markov models which are parameterised by a heuristic procedure. The parameter learning process is utilized by appliance-specific characteristics which form unified models of the specific appliances. Within the aggregated load curve, events are detected. These events are formed to power blocks which can be filtered by the parameters of unified models of the appliances. Through a clustering procedure, the relevant appliance-specific power blocks are detected, which are further used for the estimation of the model parameters. The parameterised modified Viterbi algorithm is able to find the state changes of the device to be disaggregated.
The performance analysis of the fEEDBACk algorithm is conducted by applying the method on a power measurement data set from different households with a sampling period from 1 second to 20 seconds.
The results of this novel method show that for the different test cases of cooling devices, F-Measure values between 0.80 and 0.99 can be achieved. If the parameters for the event detection are varied it has been shown that the range of F-Measure, which is the harmonic mean of precision and recall, is increased to values between 0.87 and 0.99. The selection of the durational components of the used event detection algorithm (Baranski) is sensitive to the performance in the disaggregation process of one of the investigated cooling devices. It has been shown that the selection of the parameters for the minimum power value used for the event detection only plays a subsidiary role when it is above a certain limit.

By the incorporation of reactive power values to the algorithm, a further increase of the F-Measure values can be realized. However, in test cases where the F-Measure value is above 0.90, just minor increases of about 0.01 can be achieved.
If the sampling period of the load curve to be disaggregated is increased, it has been shown that the F-Measure values remain stable for some of the test cases. If there are, for example, a fridge and a freezer within the aggregated load curve, an increase of the sampling periods to 20 seconds has a big influence on the accuracy (F-Measure: -0.15). To apply the algorithm on other sampling periods, the parameters of the event detection as well as for the computation of the steady state power changes have to be adapted. Beyond this, the distance function for the DBSCAN algorithm also needs to be altered to meet the specific needs when the sampling period is increased.

7.2 Future Work

The area for future work in the scope of unsupervised energy disaggregation algorithms lies, in general terms, in the further improvement of the accuracy as well as the reliability of the existing methods.
In the fields of automated energy analysis, it is important to introduce to the market an applicable energy disaggregation system that utilizes the smart meter data. A low-cost and easy to install system would give customers the opportunity to perform an energy analysis within the household. This would offer consumers the possibility to identify electrical devices with a high energy saving potential and beyond this, provide the possibility to trace energy saving actions.

In the field of enhancement of the proposed method, the integration of other features from the aggregated load curve such as turn-on transients or the incorporation of further sensors within the household could be investigated. Also, appliance-specific adoptions of the proposed algorithm can lead to a better performance for a specific type of device. Beyond this, the influence of a certain state duration of a specific-appliance to the next state duration could be investigated. Fridges and freezers, for example, often have a shorter off-duration after a long on-duration has occurred. This could further improve energy disaggregation.

For terms of comparability of the different available energy disaggregation methods, reviewed standardised data sets with different sampling periods and an exact description of which aggregated load curve has to be used for the disaggregation process are needed. Each algorithm should use exactly the same load curve for the disaggregation process. It is also meaningful to investigate standardized tests that have to be performed by each algorithm. For example, using different sampling periods and features, and providing the results in certain selected metrics for comparability.

8 References

[1] European Commission, "*Energy efficiency: delivering the 20% target,*" Communication from the Commission, COM (2008), vol. 772, pp. 5–6, 2008.

[2] European Commission, "*Directive 2006/32/EC of the European Parliament and of the Council of 5 April 2006 on energy end-use efficiency and energy services: Article 13,*" Official Journal of the European Union, vol. 27, 2006.

[3] European Commission, "*Directive 2009/72/EC of the European Parliament and of the Council of 13 July 2009 - concerning common rules for the internal market in electricity and repealing Directive 2003/54/EC: Annex I,*" Official Journal of the European Union, vol. 14, 2009.

[4] B. Paolo, H. Bettina, and L. Nicola, "*Energy Efficiency Status Report 2012: Electricity Consumption and Efficiency Trends in the EU-27,*" European Commission - Institute for Energy and Transport, vol. 2012.

[5] D. Bosseboeuf et al, "*Energy Efficiency Trends in Buildings in the EU: Lessons from the ODYSSEE MURE project,*" Project report. p. 42–42, 2012.

[6] M. Strasser, "*Strom-und Gastagebuch 2012: Strom-und Gaseinsatz sowie Energieeffizienz österreichischer Haushalte. Auswertung Gerätebestand und Einsatz,*" Project report. Published by Statistik Austria. Vienna, 2012.

[7] Aníbal de Almeida et al.: "*Residential Monitoring to Decrease Energy Use and Carbon Emissions in Europe* (*REMODECE*)," - Publishable Report 2008. Online available: http://remodece.isr.uc.pt/.

[8] K. Ehrhardt-Martinez, K. A. Donnelly, and J. A. Laitner, "*Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities,*" American Council for an Energy-Efficient Economy, Washington, DC, pp. 11–12, 40, 59, 2010.

[9] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "*Is Disaggregation The Holy Grail of Energy Efficiency? The Case of Electricity,*" Technical Paper Series: PTP-2012-05-1, Precourt Energy Efficiency Center, Stanford, p. 6–6, 2012.

[10] S. Darby, "*The Effectiveness of Feedback on Energy Consumption: A Review for DEFRA of the Literature on Metering. Billing and Direct Displays,*" Technical Report, Environmental Change Institute, University of Oxford, Oxford, England, pp. 8–9, 2006.

[11] B. Neenan and J. Robinson, "*Residential Electricity Use Feedback: A Research Synthesis and Economic Framework,*" EPRI Technical Update 1016844, Electric Power Research Institute, Palo Alto, California, no. 1016844, pp. 2-4 – 2-7, 2009.

[12] J. Ham and C. Midden, "*Ambient Persuasive Technology Needs Little Cognitive Effort: The Differential Effects of Cognitive Load on Lighting Feedback versus Factual Feedback,*" in Lecture Notes in Computer Science, Persuasive Technology, T. Ploug, P. Hasle, and H. Oinas-Kukkonen, Eds.: Springer Berlin Heidelberg, 2010, pp. 132-142.

[13] W. Abrahamse, L. Steg, C. Vlek, and T. Rothengatter, "*A review of intervention studies aimed at household energy conservation,*" Journal of Environmental Psychology, vol. 25, no. 3, pp. 273–291, 2005.

[14] D. Vine, "*The Effectiveness of Energy Feedback for Conservation and Peak Demand: A Literature Review,*" OJEE, vol. 02, no. 01, pp. 7–15, 2013.

[15] C. Fischer, "*Feedback on household electricity consumption: a tool for saving energy?,*" Energy Efficiency, pp. 79–104, 2008.

[16] A. Faruqui, S. Sergici, and A. Sharif, "*The impact of informational feedback on energy consumption—A survey of the experimental evidence,*" Energy, vol. 35, no. 4, pp. 1598–1608, 2010.

[17] S. Houde, A. Todd, A. Sudarshan, J. A. Flora, and K. C. Armel, "*Real-time Feedback and Electricity Consumption: A Field Experiment Assessing the Potential for Savings and Persistence,*" The Energy Journal, International Association for Energy Economics, vol. 34, no. 1, 2013..

[18] J. Schleich, M. Klobasa, S. Gölz, and M. Brunner, "*Effects of feedback on residential electricity demand—Findings from a field trial in Austria,*" Energy Policy, vol. 61, pp. 1097–1106, 2013.

[19] M. Gleerup, A. Larsen, S. Leth-Petersen, and M. Togeby, "*The Effect of Feedback by Text Message* (*SMS*) *and Email on Household Electricity Consumption: Experimental Evidence,*" EJ, vol. 31, no. 3, 2010.

[20] W. Gans, A. Alberini, and A. Longo, "*Smart meter devices and the effect of feedback on residential electricity consumption: Evidence from a natural experiment in Northern Ireland,*" Energy Economics, vol. 36, pp. 729–743, 2013.

[21] I. Vassileva, M. Odlare, F. Wallin, and E. Dahlquist, "*The impact of consumers' feedback preferences on domestic electricity consumption,*" Applied Energy, vol. 93, pp. 575–582, 2012.

[22] C. Grimm, P. Neumann, and S. Mahlknecht, "*Embedded systems for smart appliances and energy management.*" New York, NY: Springer, 2013.

[23] D. Parker, D. Hoak, A. Meier, and R. Brown, "*How much Energy Are We Using? Potential of Residential Energy Demand Feedback Devices,*" Proceedings of the 2006 Summer Study on Energy Efficiency in Buildings, American Council for an Energy Efficient Economy, Asilomar, CA, 2006.

[24] M. Fuller, C. Kunkel, M. Zimring, I. Hoffman, K.L. Soroye, and C. Goldman, "*Driving Demand for Home Energy Improvements,*" Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory.

[25] C. Elbe and E. Schmautzer, "*Automated Electrical Energy Analysis for Domestic Consumers Based on Smart Meters,*" paper no. 0408, 21st International Conference on Electricity Distribution (CIRED), Frankfurt, Germany, 2011.

[26] S. Karjalainen, "*Consumer preferences for feedback on household electricity consumption,*" Energy and Buildings, vol. 43, no. 2-3, pp. 458–467, 2011.

[27] P. Kundur, N. J. Balu, and M. G. Lauby, Power system stability and control. New York: McGraw-Hill, 1994.

[28] G. W. Hart, "*Nonintrusive appliance load monitoring,*" Proceedings of the IEEE, vol. 80, no. 12, pp. 1870–1891, 1992.

[29] M. Zeifman and K. Roth, "*Nonintrusive appliance load monitoring: Review and outlook,*" Consumer Electronics, IEEE Transactions on, vol. 57, no. 1, pp. 76–84, 2011.

[30] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "*Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey,*" Sensors, vol. 12, no. 12, pp. 16838–16866, 2012.

[31] A. Dobnikar, U. Lotrič, and B. Šter, Eds, Adaptive and Natural Computing Algorithms: Springer Berlin Heidelberg, 2011.

[32] M. Figueiredo, A. Almeida, and B. Ribeiro, "*An Experimental Study on Electrical Signature Identification of Non-Intrusive Load Monitoring Systems,*" in Lecture Notes in Computer Science, Adaptive and Natural Computing Algorithms, A. Dobnikar, U. Lotrič, and B. Šter, Eds.: Springer Berlin Heidelberg, 2011, pp. 31-40.

[33] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, Eds, Nonintrusive appliance load monitoring based on integer programming. SICE Annual Conference, 2008, 2008.

[34] A. Cole and A. Albicki, Eds, Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content. Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE, 2000.

[35] J. Li, S. West, and G. Platt, Eds, Power decomposition based on SVM regression. Modelling, Identification & Control (ICMIC), 2012 Proceedings of International Conference on, 2012.

[36] A.G. Ruzzelli, C. Nicolas, A. Schoofs, and O'Hare, G. M P, Eds, Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor. Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference on, 2010.

[37] H. Najmeddine, El Khamlichi Drissi, K, C. Pasquier, C. Faure, K. Kerroum, A. Diop, T. Jouannet, and M. Michou, "*State of art on load monitoring methods,*" 2nd International Power and Energy Conference (PECon), 2008.

[38] A. J. Bijker, Xiaohua Xia, and Jiangfeng Zhang, "*Active power residential non-intrusive appliance load monitoring system,*" AFRICON, 2009. AFRICON ,09, pp. 1–6.

[39] L. K. Norford and S. B. Leeb, "*Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms,*" Energy and Buildings, vol. 24, no. 1, pp. 51–64, 1996.

[40] M. Marceau and R. Zmeureanu, "*Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings,*" Energy Conversion and Management, vol. 41, no. 13, pp. 1389–1403, 2000.

[41] L. Farinaccio and R. Zmeureanu, "*Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses,*" Energy and Buildings, vol. 30, no. 3, pp. 245–259, 1999.

[42] M. Baranski,"*Energie-Monitoring im privaten Haushalt,*" 1st ed. Göttingen: Cuvillier, 2006.

[43] D. C. Bergman, Dong Jin, J. P. Juen, N. Tanaka, C. A. Gunter, and A. K. Wright, "*Distributed non-intrusive load monitoring,*" ISGT - Innovative Smart Grid Technologies, pp. 1–8, 2011.

[44] K. D. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe, and M. Berges, "*BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research,*" 2012.

[45] K. D. Anderson, M. E. Berges, A. Ocneanu, D. Benitez, and J. M. F. Moura, "*Event detection for Non Intrusive load monitoring,*" IECON - International Conference on: Industrial Electronics, Control, and Instrumentation, pp. 3312–3317, 2012.

[46] M. Basseville, I. V. Nikiforov, and others, Detection of abrupt changes: theory and application: Prentice Hall Englewood Cliffs, 1993.

[47] Dong Luo and Leslie K. Norford, "*Monitoring HVAC Equipment Monitoring HVAC Equipment Electrical Loads from a Centralized Location - Methods and Field Test Results,*" ASHRAE Transactions, no. 108, pp. 841–857, 2002.

[48] M. Berges, E. Goldman, H. S. Matthews, L. Soibelman, and K. Anderson, "*User-Centered Nonintrusive Electricity Load Monitoring for Residential Buildings,*" J. Comput. Civ. Eng, vol. 25, no. 6, pp. 471–480, 2011.

[49] Y. Jin, E. Tebekaemi, M. Berges, and L. Soibelman, "*Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities,*" ICASPP - International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 4340–4343, 2011.

[50] R. V. Hogg, J. W. McKean, and A. T. Craig, Introduction to mathematical statistics: Pearson, 2013.

[51] S. B. Leeb, S. R. Shaw, and J. L. Kirtley, JR, "*Transient event detection in spectral envelope estimates for nonintrusive load monitoring,*" Power Delivery, IEEE Transactions on, vol. 10, no. 3, pp. 1200–1210, 1995.

[52] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "*Power signature analysis,*" Power and Energy Magazine, IEEE, vol. 1, no. 2, pp. 56–63, 2003.

[53] S. R. Shaw, S. B. Leeb, L. K. Norford, and R. W. Cox, "*Nonintrusive Load Monitoring and Diagnostics in Power Systems,*" Transactions on Instrumentation and Measurement, IEEE, vol. 57, no. 7, pp. 1445–1454, 2008.

[54] H.-H. Chang, H.-T. Yang, and C.-L. Lin, "*Load Identification in Neural Networks for a Non-intrusive Monitoring of Industrial Electrical Loads,*" in Lecture Notes in Computer Science, Computer Supported Cooperative Work in Design IV, W. Shen, J. Yong, Y. Yang, J.-P. Barthès, and J. Luo, Eds.: Springer Berlin Heidelberg, 2008, pp. 664-674.

[55] H.-H. Chang, "*Non-Intrusive Demand Monitoring and Load Identification for Energy Management Systems Based on Transient Feature Analyzes,*" Energies, vol. 5, no. 12, pp. 4569–4589, 2012.

[56] S. Patel, T. Robertson, J. Kientz, M. Reynolds, and G. Abowd, "*At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line* (*Nominated for the Best Paper Award*)*,*" in Lecture Notes in Computer Science, UbiComp 2007: Ubiquitous Computing, J. Krumm, G. Abowd, A. Seneviratne, and T. Strang, Eds.: Springer Berlin Heidelberg, 2007, pp. 271-288.

[57] M. Hazas, A. Friday, and J. Scott, "*Look Back before Leaping Forward: Four Decades of Domestic Energy Inquiry,*" Pervasive Computing, IEEE, vol. 10, no. 1, pp. 13–19, 2011.

[58] A. J. Bijker, Xiaohua Xia, and Jiangfeng Zhang, "*Active power residential non-intrusive appliance load monitoring system,*" AFRICON, 2009, pp. 1–6.

[59] A. R. Rababaah and E. Tebekaemi, "*Electric load monitoring of residential buildings using goodness of fit and multi-layer perceptron neural networks,*" Computer Science and Automation Engineering (CSAE), 2012, vol. 2, pp. 733–737.

[60] Z. Wang and G. Zheng, "*Residential Appliances Identification and Monitoring by a Nonintrusive Method,*" Smart Grid, IEEE Transactions on, vol. 3, no. 1, pp. 80–92, 2012.

[61] J. Liang, K. K. Simon, G. Kendall, J. W. M Cheng, (2010): Load Signature Study—Part I: Basic Concept, Structure, and Methodology. In: Power Delivery, IEEE Transactions on 25 (2), S. 551–560.

[62] F. Jazizadeh and B. Becerik-Gerber, "*A Novel Method for Non Intrusive Load Monitoring of Lighting Systems in Commercial Buildings,*" Computing in Civil Engineering (2012), pp. 523–530.

[63] A. Schoofs, A. Guerrieri, Delaney, D. T, O'Hare, G. M P, and A.G. Ruzzelli, Eds, "*ANNOT: Automated Electricity Data Annotation Using Wireless Sensor Networks,*" 7th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010.

[64] Z. C. Taysi, M. A. Guvensan, and T. Melodia, "*TinyEARS: Spying on House Appliances with Audio Sensor Nodes,*" 2nd ACM Workshop, Zurich, Switzerland, pp. 31–36, 2010.

[65] A. Zoha, A. Gluhak, M. Nati, M.A. Imran, and S. Rajasegarar, Eds, "*Acoustic and device feature fusion for load recognition,* " 6th IEEE International Conference on Intelligent Systems (IS), 2012.

[66] A. Einfalt, A. Schuster, C. Leitinger, D. Tiefgraber, M. Litzlbauer, and S. Ghaemi, "*Konzeptentwicklung für ADRES - Autonome Dezentrale Regenerative Energie Systeme,*" Energie der Zukunft, Technical Report, 2012.

[67] J. Z. Kolter and M. J. Johnson, "*REDD: A public data set for energy disaggregation research,*" In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.

[68] S. Barker, A. Mishra, D. Irwin, E. Cecchet, and P. Shenoy, "*Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes,*" ACM SustKDD'12, 2012.

[69] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. o. Chonov, M. Werner, and R. Steinmetz, "*On the Accuracy of Appliance Identification Based on Distributed Load Metering Data,*" in Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT), 2012, pp. 1–9.

[70] P. Street, "*The Pecan Street project,*" Austin, TX: Working Group Report, 2013.

[71] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic, "*AMPds: A Public Dataset for Load Disaggregation and Eco-Feedback Research,*" Electrical Power and Energy Conference (EPEC), 2013 IEEE, pp. 1–6.

[72] T. Mallits, "*Analyse des Stromverbrauches in Haushalten mit Mustererkennung: Eine mögliche Anwendung für Smart Meter,*" Diploma thesis, Graz University of Technology, 2012.

[73] H. S. Kim, "*Unsupervised disaggregation of low frequency power measurements,*" Thesis, University of Illinois at Urbana-Champaign, 2011.

[74] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "*Non-Intrusive Load Monitoring Using Prior Models of General Appliance Types,*" Twenty-Sixth Conference on Artificial Intelligence (AAAI-12), 2012.

[75] J. Z. Kolter and T. Jaakkola, "*Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation,*" Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012, La Palma, Canary Islands, April 21-23, 2012, no. 22, pp. 1472–1482, 2012.

[76] M. J. Johnson and A. S. Willsky, "*Bayesian Nonparametric Hidden Semi-Markov Models,*" ArXiv e-prints, arXiv:1203.1365 [stat.ME], 2012.

[77] M. Zeifman, "*Disaggregation of home energy display data using probabilistic approach,*" Consumer Electronics, IEEE Transactions on, vol. 58, no. 1, pp. 23–31, 2012.

[78] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "*Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring,*" Journal of Industrial Ecology, vol. 14, no. 5, pp. 844–858, 2010.

[79] D. Srinivasan, Ng, W. S, and A. C. Liew, "*Neural-network-based signature recognition for harmonic source identification,*" IEEE Transactions on Power Delivery, vol. 21, no. 1, pp. 398–405, 2006.

[80] A. Marchiori, D. Hakkarinen, Qi Han, and L. Earle, "*Circuit-Level Load Monitoring for Household Energy Management,*" Pervasive Computing, IEEE, vol. 10, no. 1, pp. 40–48, 2011.

[81] Gu-yuan Lin, Shih-chiang Lee, Hsu, J. Y. -J, and Wan-rong Jih, Eds, "*Applying power meters for appliance recognition on the electric panel,*" 5th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2010.

[82] T. Zia, D. Bruckner, and A. Zaidi, Eds, "*A hidden Markov model based procedure for identifying household electric loads.*" IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society, 2011.

[83] H. Goncalves, A. Ocneanu, and M. Berges, "*Unsupervised disaggregation of appliances using aggregated consumption data,*" 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD), 2011.

[84] H. Shao, M. Marwah, and N. Ramakrishnan, "*A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation: Applications to Residential and Commercial Buildings,*" Twenty-Seventh AAAI Conference on Artificial Intelligence, Bellevue, Washington, USA, 2013.

[85] C. Elbe and E. Schmautzer, "*Appliance-Specific Energy Consumption Feedback for Domestic Consumers Using Load Disaggregation Methods,*" 22nd International Conference on Electricity Distribution (CIRED), Stockholm, Sweden, 2013.

[86] A.I. Cole and A. Albicki, Eds, Algorithm for nonintrusive identification of residential appliances. Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on, 1998.

[87] J. Z. Kolter, S. Batra, and A. Y. Ng, "*Energy disaggregation via discriminative sparse coding,*" NIPS, pp. 1153–1161, 2010.

[88] European Commission, "*European Standard EN 50160: Voltage characteristics of electricity supplied by public distribution networks,*" Brussels, 2011.

[89] L. R. Rabiner, (1989): A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE 77 (2), pp. 257–286.

[90] J. D. Ferguson: Variable duration models for speech. In: Proc. Symposium on the application of hidden Markov models to text and speech 1980, pp. 143–179.

[91] C. M. Bishop, (2006): Pattern recognition and machine learning. New York NY: Springer (Information science and statistics).

[92] Shun-Zheng Yu (2010): Hidden semi-Markov models. In: Artificial Intelligence 174 (2), S. 215–243.

[93] Y. Guédon, (2003): Estimating hidden semi-Markov chains from discrete sequences. In: Journal of Computational and Graphical Statistics (Volume 12, Issue 3).

[94] R. C. Moyer, Contemporary financial management, 12th ed. Mason, OH: South-Western, Cengage Learning, pp. 147-498, 2012.

[95] S. Theodoridis and K. Koutroumbas, Pattern recognition, 3rd ed. Amsterdam, Boston: Elsevier/Academic Press, pp. 695-699, 2006.

A Appendix

A.1 Parameters and Results of Event Detectors with Different Sampling Periods

A.1.1 Sampling Period: 1 Second

Table A.1: Overview of best parameters and results for a sampling period of 1 second

Metric	Parameters				Results			
Baranski								
	P_{min}	τ			TP	M	FP	score
Rate	14	1			871	36	154	0.49
Percentage	16	2			859	48	27	0.76
Total Power	14	2			860	47	34	0.76
Average Power	58	6			642	265	2	0.40
GLR								
	w_{vote}	w_l	v_{thr}		TP	M	FP	score
Rate	7	3	1		842	65	96	0.54
Percentage	5	2	2		810	97	26	0.38
Total Power	5	2	1		837	70	90	0.54
Average Power	13	6	12		660	247	13	0.28
Bergman								
	w	P_{thr}	P_{δ}	δ	TP	M	FP	score
Rate	3	10	29	1	871	36	850	0.41
Percentage	2	16	17	1	849	58	26	0.61
Total Power	2	16	17	1	849	58	26	0.61
Average Power	4	32	29	2	286	621	0	0.35

A.1.2 Sampling Period: 2 Seconds

Table A.2: Overview of best parameters and results for a sampling period of 2 seconds

Metric	Parameters				Results			
Baranski								
	P_{min}	τ			TP	M	FP	score
Rate	14	1			871	36	205	0.52
Percentage	12	3			849	58	36	0.66
Total Power	10	3			850	57	42	0.67
Average Power	60	5			617	290	1	0.38
GLR								
	w_{vote}	w_l	v_{thr}		TP	M	FP	score
Rate	9	4	1		777	130	40	0.39
Percentage	9	4	1		777	130	40	0.39
Total Power	7	3	1		767	140	54	0.43
Average Power	13	6	8		665	242	12	0.34
Bergman								
	w	P_{thr}	P_{δ}	δ	TP	M	FP	score
Rate	3	10	29	1	877	30	708	0.47
Percentage	3	10	11	1	809	98	29	0.49
Total Power	3	10	14	1	832	75	127	0.49
Average Power	6	22	26	2	291	616	652	0.34

A.1.3 Sampling Period: 5 Seconds

Table A.3: Overview of best parameters and results for a sampling period of 5 second

Metric	Parameters				Results			
Baranski								
	P_{min}	τ			TP	M	FP	score
Rate	10	1			835	72	145	0.45
Percentage	6	6			778	129	24	0.43
Total Power	24	1			816	91	113	0.50
Average Power	60	6			585	322	0	0.39
GLR								
	w_{vote}	w_l	v_{thr}		TP	M	FP	score
Rate	5	2	1		717	190	40	0.36
Percentage	5	2	1		717	190	40	0.36
Total Power	7	3	1		709	198	45	0.34
Average Power	13	6	12		560	347	5	0.27
Bergman								
	w	P_{thr}	P_{δ}	δ	TP	M	FP	score
Rate	3	10	29	1	818	89	454	0.30
Percentage	2	16	20	1	772	135	49	0.37
Total Power	2	14	29	1	807	100	163	0.39
Average Power	6	32	29	1	108	799	0	0.31

A.2 Parameters and Results of Event Detectors with Incorporation of Reactive Power Values

Table A.4: Overview of best parameters and results for a sampling period of 1 second

Metric	Parameters				Results			
Baranski								
	P_{min}	τ			TP	M	FP	score
Rate (P)	28	1			321	586	113	0.17
Rate (Q)	10	1						
Percentage (P)	28	1			321	586	113	0.17
Percentage (Q)	10	1						
Total Power (P)	28	3			308	599	3	0.32
Total Power (Q)	10	1						
Average Power (P)	38	3			296	611	2	0.32
Average Power (Q)	10	4						
GLR								
	w_{vote}	w_l	v_{thr}		TP	M	FP	score
Rate (P)	5	2	1		171	736	46	0.23
Rate (Q)	13	6	1					
Percentage (P)	11	5	2		170	737	9	0.25
Percentage (Q)	11	5	1					
Total Power (P)	11	5	1		171	736	51	0.25
Total Power (Q)	11	5	1					
Average Power (P)	13	6	8		120	787	0	0.26
Average Power (Q)	5	2	2					
Bergman								
	w	P_{thr}	P_{δ}	δ	TP	M	FP	score
Rate (P)	3	12	29	2	307	600	19	0.28
Rate (Q)	2	10	17	2				
Percentage (P)	3	12	29	2	307	600	19	0.28
Percentage (Q)	2	10	17	2				
Total Power (P)	3	12	29	2	307	600	19	0.28
Total Power (Q)	2	10	17	2				
Average Power (P)	9	14	29	5	227	680	0	0.32
Average Power (Q)	2	10	8	5				

A.3 Parameters and Results of fEEDBACk Algorithm

A.3.1 Parameters for Detecting Cooling Devices

$P_{min}(D)$	60 W
$P_{max}(D)$	150 W
$t_{on,min}(D)$	3 minutes
$t_{on,max}(D)$	65 minutes
$\upsilon_{on}(D)$	Gamma distribution
$t_{off,min}(D)$	7 minutes
$t_{off,max}(D)$	83 minutes
$\upsilon_{off}(D)$	Gamma distribution
J	2
A	[0,1;1,0]
π	[0.5 0.5]
$M_{on}(D)$	65 minutes

A.3.2 Sensitivity Analysis 1 - Variation of F-Measure

Table A.5: Average variation of F-Measure values to best values of each investigated cooling devices when the parameters for the Baranski event detector are varied

		τ in seconds		
		2	3	4
P_{min} in W	2	-0.150	-0.023	-0.014
	4	-0.008	-0.023	-0.020
	6	-0.016	-0.021	-0.024
	8	-0.015	-0.003	-0.004
	10	-0.019	-0.004	-0.004
	12	-0.002	-0.002	-0.002
	14	-0.004	-0.005	-0.005
	16	-0.003	-0.002	-0.002
	18	-0.003	-0.006	-0.005
	20	-0.002	-0.003	-0.003
	22	-0.001	-0.002	-0.002

A.3.3 Sensitivity Analysis 2 - Results

Table A.6: Overview of energy disaggregation results for cooling devices with and without incorporation of reactive power, sampling period: 1 second

Cooling Device	Without Q			With incorporation of Q		
	F-Measure	Precision	Recall	F-Measure	Precision	Recall
1	0.80	0.81	0.78	0.86	0.87	0.85
2	0.96	0.98	0.95	0.97	0.98	0.96
3	0.99	0.99	0.99	0.99	1.00	0.99
4	0.91	0.91	0.92	0.91	0.92	0.91
5	0.98	0.98	0.98	0.98	0.98	0.99
6	0.99	0.99	0.99	0.99	0.99	0.99