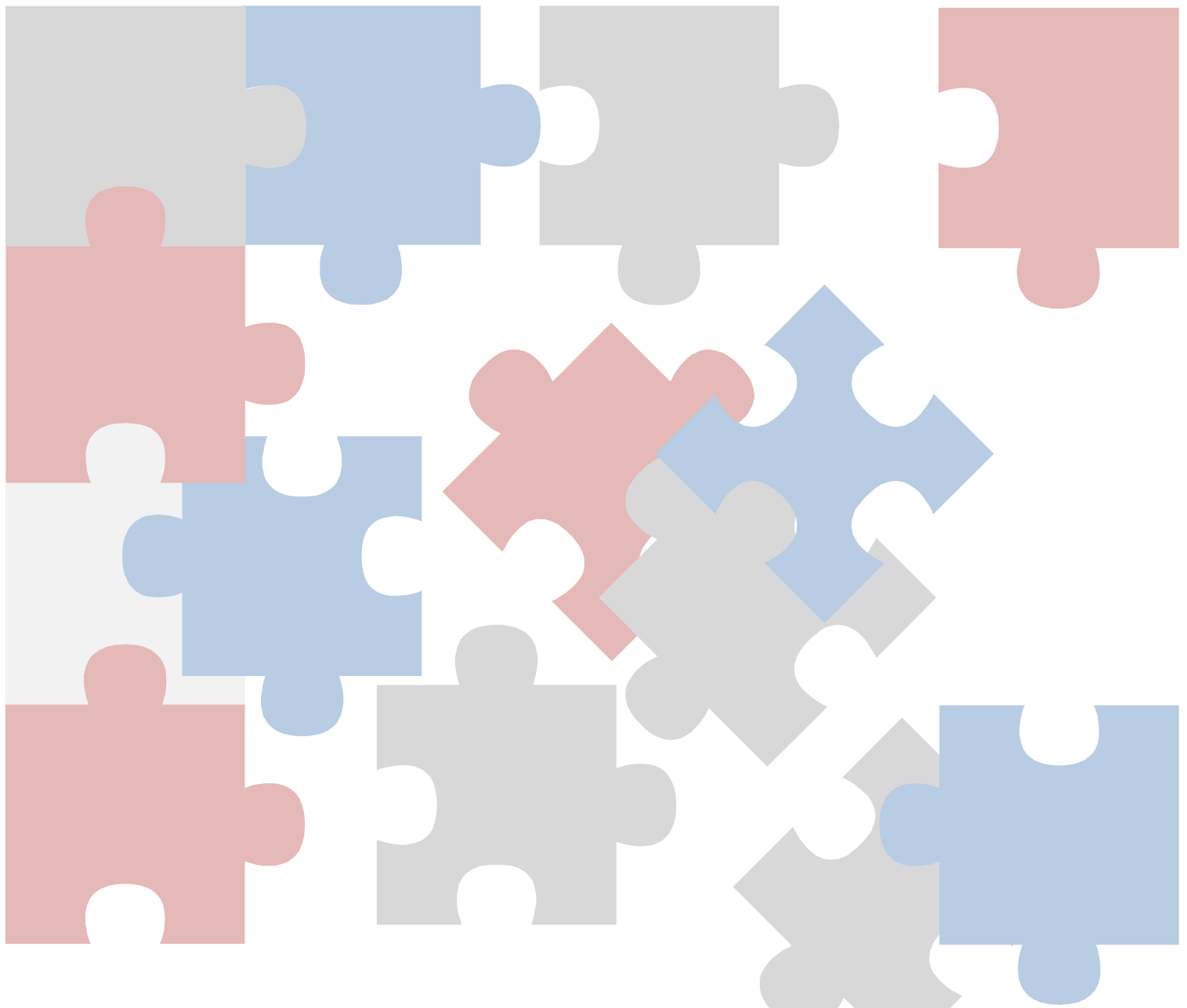
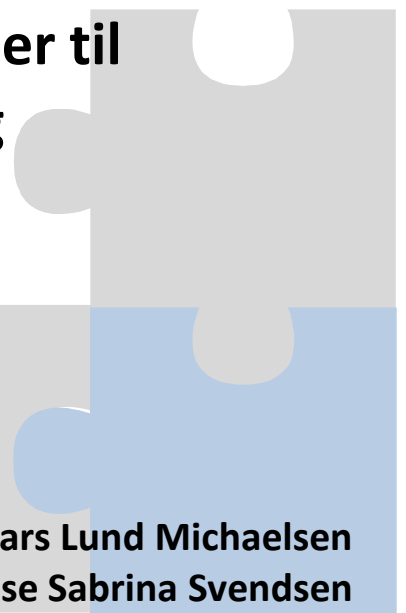
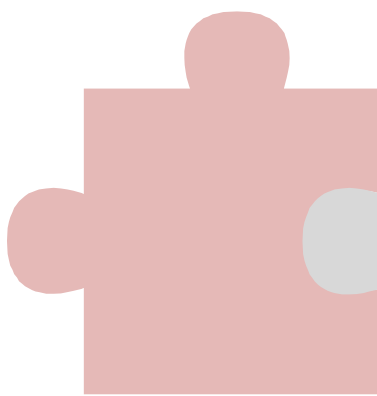
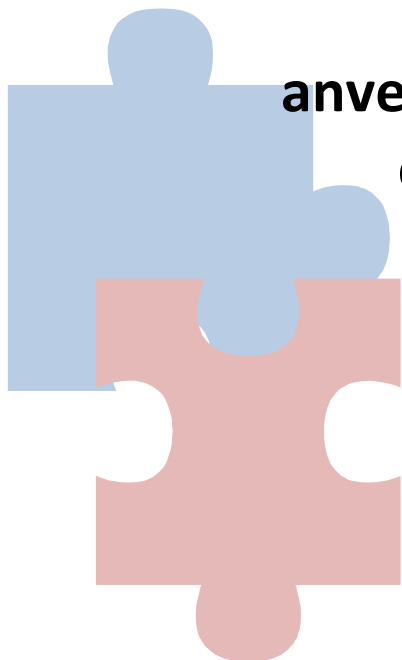




Brugeradfærd i boliger

**anvendelse af klassificeringstræer til
estimering af energiforbrug**



**Lars Lund Michaelsen
Louise Sabrina Svendsen**

Titel:

Brugeradfærd i boliger - anvendelse
af klassificeringstræer til estimering
af energiforbrug

Tema:

Brugeradfærd

Projektperiode:

Kandidatprojekt,
efterårssemester 2010 /
forårssemester 2011

Projektforfattere:

Lars Lund Michaelsen

Louise Sabrina Svendsen

Vejledere:

Henrik Brohus

Tine Steen Larsen

Oplagstal: 5

Sidetal: 188

Bilagsantal og -art:

1 appendiks og 1 dvd

Afsluttet: 9. juni 2011

Synopsis:

Denne rapport består overordnet set af tre dele. Den første del af rapporten indeholder en problemanalyse, med en beskrivelse af hvor stor en del af det samlede danske energiforbrug, der gennemsnitligt anvendes i husholdninger. I denne beskrives ligeledes hvor store variationer af energiforbruget, der kan observeres imellem boliger, blandt andet som følge af beboernes adfærd.

Den efterfølgende del af rapporten indeholder et litteraturstudie, der er skrevet med henblik på at klarlægge erfaringer og viden om beboeres påvirkning på energiforbruget i boliger. Fokus rettes mod at identificere hvilke parametre, der kan forklare dele af forbrugsvariationerne boliger imellem. Tilsvarende klarlægges nogle metoder, der tidligere er anvendt til at beskrive eller estimere beboernes påvirkning på energiforbruget.

I den sidste del af rapporten videreføres erfaringer fra litteraturstudiet til en analyse af, hvorledes disse kan inddrages i udviklingen af klassificeringstræer til estimering af energiforbrug i boliger. Der foretages en eksemplificering af udviklingen af et klassificeringstræ for elforbruget på baggrund af fiktivt data. Der redegøres endvidere for håndtering af problemstillinger, der kan opstå i forbindelse med udviklingen af et klassificeringstræ på baggrund af virkeligt data. Slutteligt konkluderes anvendelse af metoden til estimering af energiforbrug under hensyntagen til beboernes påvirkning.

Forord

Denne rapport er produktet af kandidatopgaven på Aalborg Universitet inden for retningen Indeklima og Energi ved det Teknisk-Naturvidenskabelige Fakultet.

Formål og målgruppe

Formålet med nærværende rapport er at klarlægge erfaringer og viden om beboeres påvirkning på energiforbruget i boliger, og videreføre denne til udvikling af et klassificeringstræ, der kan anvendes til estimering af energiforbrug i parcel-/rækkehuse.

Rapporten henvender sig til personer, der besidder faglig viden svarende til ingeniørvidenskabeligt niveau inden for indeklima og energi. Derudover er grundlæggende kendskab til data mining fordelagtigt.

Særlig tak til

Der skal rettes en særlig tak til Tine Steen Larsen, Aalborg Universitet og Henrik Brohus, Aalborg Universitet for konstruktiv vejledning og inspiration igennem hele projektperioden.

Der rettes tilsvarende en særlig tak til Henrik N. Knudsen, Statens Byggeforskningsinstitut og Lars Kristensen, Det Grønne Hus/Energitjenesten for udlevering af spørgeskemabesvarelser og datamateriale, samt udredning af spørgsmål i forbindelse med dette.

Endelig skal der rettes en tak til Vibeke Hansen Kjærbye, NIRAS for at vi må gengive en figur fra hendes rapport.

Læsevejledning

Rapportens indhold

Denne rapport består af tre overordnede dele. Kapitel 1 og 2 udgør den indledende problem-analyse, hvor der fokuseres på, hvor stor en andel af Danmarks samlede energiforbrug der anvendes i husholdninger. I denne del indgår ligeledes en beskrivelse af, hvor store variationer af energiforbruget der kan observeres husstande imellem, blandt andet som følge af beboernes adfærd.

Kapitel 3 udgør den anden hoveddel af rapporten, hvilket er litteraturstudiet. Dette kapitel er skrevet med henblik på at klarlægge erfaringer, viden og relevante resultater for beboernes indflydelse på energiforbruget i boliger, med udgangspunkt i den tilgængelige litteratur inden for området.

Den sidste del af rapporten indeholder en analyse af, hvorledes viden fra litteraturstudiet kan videreføres til udviklingen af et klassificeringstræ, som kan anvendes til estimering af energiforbrug i boliger. Der udføres en eksemplificering af udviklingen af et klassificeringstræ for elforbruget, på baggrund af fiktivt data. Efterfølgende beskrives metodens anvendelighed, set i et større perspektiv, med henblik på at klarlægge, hvorledes denne kan implementeres i praksis. Slutteligt diskuteres og konkluderes på erfaringer og resultater fra projektet.

Dokumentation

Denne rapport indeholder den væsentligste dokumentation for de ovenstående emner. Til rapporten hører et appendiks, der indeholder yderligere information om litteraturstudie og grundlæggende teori for anvendte metoder. Tre af de udviklede klassificeringstræer er vedlagt som separate tegninger til rapporten. For forklaring af de anvendte parametre og spørgsmål henvises til kapitel 6. Ydermere er der vedlagt en dvd, der indeholder dokumentation i form af anvendte regneark, illustrationer fra rapporten og installationsfiler til det i projektet anvendte data minings software Weka.

Kildehenvisninger og ordforklaring

I hovedrapporten og appendiks er kildehenvisningerne angivet ved nummerering. Nærmere information om kilderne, i form af titel, forfatter mv., fremgår af litteraturlisten, som er placeret bagerst i rapporten i henholdsvis hovedrapporten og appendiks.

Der er til hovedrapporten tilknyttet en ordliste, der indeholder en beskrivelse af nogle af de fagudtryk og begreber, der anvendes. Ordlisten er at finde på siderne før litteraturlisten.

Summary

Energy consumption in general is a subject discussed throughout the world, due to its effect on global warming. The climate changes are evolving consequently affecting many people around the world. In order to try to stop the process of global warming, several countries set a goal of reducing their total energy consumption within the year 2020. The total amount of energy spend in Danish dwellings account for 31 % of the total energy consumption. In order to try to meet the international agreement and reducing the energy consumption, the government in Denmark set up goals, resulting, among other things, in the legislation on buildings being edited. Usually only 1 % of the housing stock is replaced each year, and since the existing housing stock responsible for consuming the most, there is a need for focus on these. Since the government cannot force people to renovate their houses, they can only encourage them to save energy by optimizing the building through renovation or by focusing on the behaviour of the occupants. In this project the main goal is to examine how the occupant behaviour affects the energy consumption in dwellings, and hereby try to find a method that can describe this connection. This is performed in order to develop a model that can be used as a tool for informing the occupants on how to save energy based on their behaviour.

This project is divided into three main sections. The project takes its starting point in a preliminary study where the development of energy consumption primarily in Danish dwellings is evaluated. This includes an evaluation on how the energy consumption varies within the same types of buildings due to, among other things, the occupant behaviour. The following section contains a literature survey where knowledge of how the occupant behaviour affects the energy consumption in dwellings is gathered and evaluated. The knowledge from the literature survey is employed in an exemplification of how to develop and use a classification tree in order to estimate the energy consumption of a dwelling, when taking, among other things, occupant behaviour into account. In the first part of the report the main focus is on examining the development of the energy consumption in dwellings. Here it is observed that the total energy consumption for domestic space heating and heating the domestic hot water has been reduced by 11 % since 1980. The usage for utility water is reduced from $70 \text{ m}^3/\text{person per year}$ in 1987 to $45 \text{ m}^3/\text{person per year}$ in 2007. Whereas the usage of domestic hot water has increased from $10 \text{ m}^3/\text{person per year}$ to $15 \text{ m}^3/\text{person per year}$ in the same period. The electricity consumption has been increasing in the period from 1987 to 1992 but has not since then been varying remarkably. On the basis of the literature survey a study upon which air temperatures the occupants prefer in their dwellings is conducted. In that study it is found, that the preferred air temperatures varies between $19 - 24 \text{ }^\circ\text{C}$ with approximately 50 % preferring air temperatures between $21 - 23 \text{ }^\circ\text{C}$.

Within the subject of varying consumption buildings in between, both the preliminary study and the literature survey shows that the total water consumption varies with a factor of 7 within similar types of buildings. Whereas the electricity consumption is seen varying with a factor of 7 – 10. Occupant behaviour and its influence on energy consumption is evaluated throughout the literature study. Here it is ascertained that the following parameters are of

importance for describing the connection between occupant behaviour and domestic heat consumption: air temperature, habits of occupants when airing the dwelling, living area and the year of construction. For water consumption the parameters are: number of occupants, number and duration of showers, number of laundering, the presence of kids and teenagers and living area. Lastly the size of electricity consumption is affected by the parameters: number of occupants, number of electrical appliances and their service life, household income, the presence of kids and teenagers, standby power and the knowledge level of the occupants regarding size of electricity consumption.

In order to be able to estimate the energy consumption in dwellings, with regard for occupant behaviour, it is necessary to choose a method that can be used to describe the connection. To ensure that the model derived from this method can be used widely, it is necessary that the model is easy to understand and use. In the last part of the report an evaluation of five data mining methods is conducted. The result of this evaluation shows that the outcome of the classifications tree method is the easiest to understand and use. Through the last part of this report the knowledge from the preliminary study and literature survey is used for producing a fictitious dataset. This dataset is utilised in the exemplification of how to develop and apply the classification tree.

As mentioned earlier the classification tree is simple to use and easy to understand, which is why it is chosen as a tool for estimating the energy consumption with regard for occupant behaviour. The target groups for usage are: occupants in dwellings, energy advisors and real estate agents. The same classification tree can be developed for all three groups, but there is an opportunity for developing a more technical classification tree for the energy advisors. The result of using the classification tree is an estimation of energy consumption for the given household that is being tested. Based on the classification tree it is possible to evaluate which parameters affected the result and hence which parameters to focus on in order to lower the energy consumption. The developed tree indicated that there is a possibility that usage of the classification tree with regard for energy consumption and occupant behaviour will be a success. It cannot be expected that the usage of the classification tree alone will result in reduced energy consumption of the households. As a result it is recommended that the usage of the tree is followed up by additional guidance on a web page or by energy advisors etc.

Indholdsfortegnelse

Kapitel 1	Indledning	13
Kapitel 2	Problemanalyse	14
2.1	Energiforbrug i husholdninger	14
2.1.1	Energiforbrugets udvikling	15
2.1.2	Energiforbrugets variation mellem husstande	18
2.1.3	Godkendelse af boligers energibehov	22
2.2	Adfærdens påvirkning på energiforbruget i boliger	24
2.2.1	Væsentlige forhold for det estimerede energiforbrug	26
2.3	Metoder til estimering af adfærdens påvirkning på energiforbruget	29
2.3.1	Be10	29
2.3.2	BSim	30
2.3.3	Internetbaserede energiberegningsprogrammer	31
2.3.4	Udvikling af nyt beregningsprogram	32
2.4	Interesseparter	33
2.5	Problemformulering	34
2.5.1	Projektafgrænsning	35
Kapitel 3	Litteraturstudie	36
3.1	Metoder til analysering af data	36
3.1.1	Theory of Planned Behaviour	37
3.1.2	Faktor- og komponentanalyse	38
3.1.3	Klyngeanalyse	39
3.1.4	Beslutningstræer	41
3.1.5	Diskussion	44
3.2	Rumopvarmning	45
3.2.1	Lufttemperaturer i boliger	45
3.2.2	Ventilation og udluftning i boliger	52
3.2.3	Diskussion	59
3.3	Elforbrug	64
3.3.1	Elforbrugets slutanvendelser	65
3.3.2	Beboernes betydning for elforbruget	68
3.3.3	Væsentlige parametre for elforbruget	71
3.3.4	Diskussion	74
3.4	Vandforbrug	77
3.4.1	Samlet vandforbrug	77
3.4.2	Varmt brugsvand	80
3.4.3	Diskussion	84
Kapitel 4	Metoder til analysering af data	86
4.1	Introduktion af metoder	87

4.2	Diskussion af metodernes anvendelighed	89
4.3	Beslutningstræ	92
4.3.1	Klassificeringstræ	94
Kapitel 5	Beskrivelse af mulige klassificeringstræer og data til udvikling af disse	98
5.1	Beskrivelse af hvilke klassificeringstræer der kan udvikles	98
5.1.1	Beskrivelse af modeltype 1	99
5.1.2	Beskrivelse af modeltype 2	100
5.1.3	Udvælgelse af modeltype	101
5.2	Definering af klassificeringstræernes klasser	101
5.2.1	Vurdering af mulige klasser	101
5.2.2	Inddeling af de udvalgte klasser	103
5.3	Definering af klassificeringstræernes attributter	112
5.3.1	Vurdering af nye parametre på baggrund af Fremtidens Parcelhuse . .	112
5.3.2	Beskrivelse af attributter til klassificeringstræerne	121
Kapitel 6	Generering af data	124
6.1	Beskrivelse af datagenerering for klassificeringstræets klasse	125
6.1.1	Beskrivelse af datagenerering for boligareal	125
6.1.2	Beskrivelse af datagenerering for husstandens indkomst	126
6.1.3	Beskrivelse af datagenerering for beboernes alder	127
6.1.4	Beskrivelse af datagenerering for beboernes forventede elforbrug . . .	129
6.1.5	Beskrivelse af datagenerering for beboernes opmærksomhed på elforbruget	130
6.1.6	Beskrivelse af datagenerering for mængden af underholdningsapparater	131
6.1.7	Beskrivelse af datagenerering for brugstid af underholdningsapparater	132
6.1.8	Beskrivelse af datagenerering for brug af husholdningsapparater	133
6.1.9	Beskrivelse af datagenerering for antallet af elapparater med standbyforbrug	134
Kapitel 7	Udvikling af klassificeringstræ	136
7.1	Algoritme til udvikling af klassificeringstræ	136
7.2	Håndtering af manglende data	144
7.2.1	Manglende data ved udvikling af klassificeringstræ	145
7.2.2	Manglende data ved anvendelse af klassificeringstræ	150
7.3	Begrænsning af klassificeringstræets størrelse	152
7.4	Evaluerings af et klassificeringstræets nøjagtighed	155
7.4.1	Confusion matrix	156
7.4.2	Andelen af korrekt klassificerede elementer i klassificeringstræet . . .	157
7.4.3	True positive rate og false positive rate	157
7.4.4	Precision	158
7.4.5	F-measure	158
Kapitel 8	Anvendelse af klassificeringstræer	160
8.1	Klassificeringstræernes målgruppe	160
8.2	Betydningen af datagrundlagets størrelse	162
Kapitel 9	Diskussion	167

Kapitel 10 Konklusion	173
Kapitel 11 Perspektivering	178
Ordliste	180
Litteratur	183

Indledning

Der er gennem de senere år for alvor kommet fokus på den globale opvarmning og dens konsekvenser for befolkningen og miljøet, blandt andet på grund af, at der er observeret ændringer af klimaet på Jorden. En ændring af klimaet har direkte indflydelse på befolkningens levemåde, da ændringen kan føre til ekstreme vejsituationer, herunder hedeølger, tørke og oversvømmelser. Af denne årsag arbejder forskere over hele verden på at klarlægge, hvilke forhold der har betydning for den globale opvarmning. En lang række forskere er enige om, at en del af klimaproblematikken skyldes den menneskeskabte udledning af drivhusgasser. Denne udledning er en konsekvens af blandt andet det globale behov for elektricitets- og varmeproduktion, transport og lignende, da hovedparten af disse behov er baseret på anvendelsen af fossile brændsler. [59]

Håndteringen af klimaproblematikken foregår både på globalt og nationalt plan. På globalt plan sker dette ved, at der i mange lande indgås internationale forpligtigelser til at reducere den samlede udledning af drivhusgasser. Disse forpligtelser videreføres til den nationale håndtering af klimaproblematikken, hvor det i Danmark blandt andet er udtrykt gennem regeringens energipolitiske målsætning. Denne målsætning for energibesparelser er at nedbringe bruttoenergiforbruget, der er en opgørelse over det samlede indenlandske energiforbrug, med 4 % frem til 2020 i forhold til 2006 [81]. For at nå denne målsætning har regeringen blandt andet fokuseret på energiforbruget i bygninger, da disse udgør en betragtelig del af det samlede energiforbrug. Dette fokus er blandt andet kommet til udtryk gennem skærpede krav til bygningers energibehov. Da den danske boligmasse udgør en væsentlig del af den samlede bygningsmasse, er det naturligt at fokusere på energiforbruget i disse. De skærpede krav til bygningers, og derunder også boligers, energibehov kommer til udtryk gennem *Bygningsreglementet* [68]. Da disse krav hovedsagligt omhandler bygningskonstruktioners varmetekniske egenskaber, har det medført, at energiforbruget til rumopvarmning er reduceret i nye bygninger. Derved udgør energiforbruget til opvarmning af varmt brugsvand og elforbruget en større del af boligers samlede energiforbrug, end det typisk er tilfældet i ældre boliger. Da samtlige energiforbrug i boligerne er under indflydelse af beboerne, har deres adfærd og vaner væsentlig betydning for energiforbrugets størrelse. Der er af denne årsag behov for at klarlægge, hvordan og i hvor stort et omfang beboernes adfærd har indflydelse på energiforbruget, for at udnytte en del af det store besparelspotentiale der er i forbindelse med energiforbruget i boliger.

Da klimaproblematikken er en samfundsmæssig problemstilling, er der behov for, at der bidrages til løsning af denne fra flere sider. Fokusområdet, omhandlende beboernes påvirkning af energiforbruget i boliger, er en lille, men særdeles relevant, del af den samfundsmæssige problemstilling.

Problemanalyse

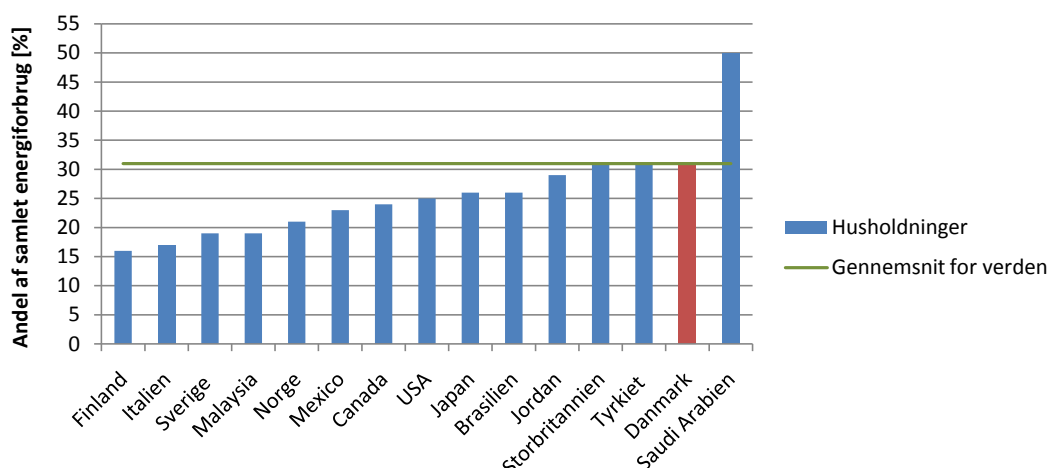
Med udgangspunkt i indledningen fremgår det at boligmassen har potentiale i forbindelse med at opnå nationale energibesparelser. Derfor indledes problemanalysen med en opgørelse over hvor meget energi, der anvendes i danske husholdninger generelt, samt en beskrivelse af hvilken udvikling der er sket inden for dette område fra 1980'erne og frem til i dag. På baggrund af den generelle udvikling af energiforbruget i husholdninger beskrives hvorledes dette forbrug varierer boliger imellem. Denne variation giver anledning til en række problemstillinger, som opstår i forbindelse med estimering af energiforbruget i boliger. På denne baggrund introduceres nogle af de forhold, der har stor betydning for energiforbruget. Analyserne danner grundlaget for problemformuleringen til dette projekt, der beskrives sidst i dette kapitel.

2.1 Energiforbrug i husholdninger

Der har gennem en længere årrække været øget fokus på det samlede energiforbrug generelt, og på hvorledes dette forbrug kan reduceres. Årsagen til den øgede opmærksomhed på området skyldes, set i et større perspektiv, blandt andet problemstillingerne omkring global opvarmning og den usikre adgang til fossile brændsler. Hvad enten der er tale om økonomiske, forsyningsmæssige eller miljømæssige hensyn, har en øget bevågenhed blandt politikere, organisationer, virksomheder, medierne og befolkningen øget interessen for at udføre en generel energibesparende indsats. Et af de områder der er i fokus, er energiforbruget til husholdninger [23, 39]. Det drejer sig om energiforbruget til eksempelvis rumopvarmning, produktion og distribution af varmt brugsvand, ventilation, husholdningsapparater m.fl.

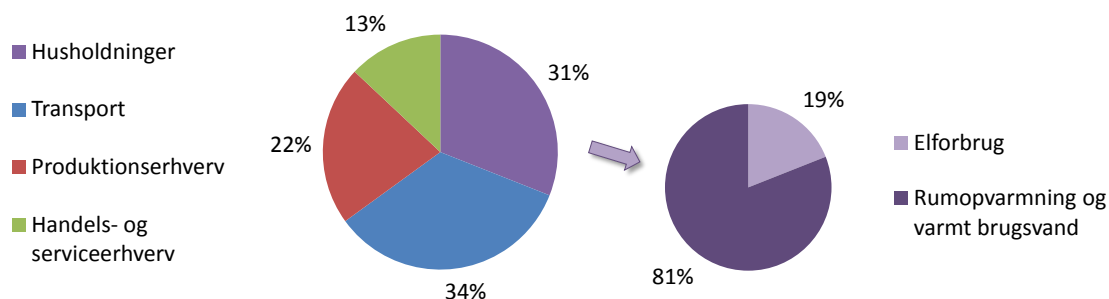
Energiforbruget i husholdninger udgør i flere lande en stor andel af det totale energiforbrug. Af figur 2.1 fremgår det, at 16 - 50 % af det totale energiforbrug i de enkelte lande anvendes i husholdningerne, samt at der gennemsnitligt i verden anvendes omkring 31 % i denne sektor [27, 86]. På baggrund af denne figur fremgår det, at andelen af det samlede danske energiforbrug, der anvendes i husholdninger, svarer til den gennemsnitlige andel i verden. Det bør i denne sammenhæng bemærkes, at energiforbruget til husholdninger er under 26 % for størstedelen af de illustrerede lande, hvilket indikerer, at der, sammenlignet med de øvrige lande, er et oplagt potentiale for at reducere energiforbruget i danske husholdninger.

2.1.1. Energiforbrugets udvikling



Figur 2.1: Husholdningers andel af samlet energiforbrug i forskellige lande [27, 86].

Rettes fokus mod energiforbruget i Danmark, blev der i 2009 anvendt energi, for hvad der svarer til 631 PJ fordelt på områderne transport, produktionserhverv, handels- og serviceerhverv samt husholdninger [27]. Fordelingen af energiforbruget mellem de førnævnte områder fremgår af figur 2.2, hvor det i denne sammenhæng kan bemærkes, at energiforbruget i husholdninger udgjorde 192 PJ i 2009 [27]. Dette forbrug er gennemsnitligt fordelt således, at omkring 81 % anvendes til rumopvarmning samt opvarmning af varmt brugsvand, og de resterende 19 % anvendes til elforbrugende apparater i husholdningen [27].



Figur 2.2: Fordeling af anvendt energi i Danmark i 2009 [27].

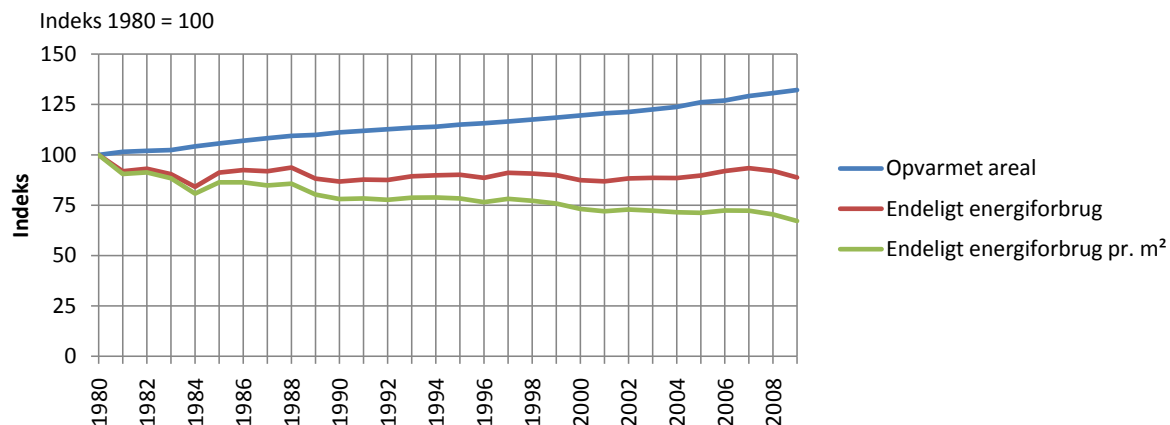
2.1.1 Energiforbrugets udvikling

Energiforbruget i husholdningerne, der fremgår af figur 2.1 og 2.2, angiver andelen af det totale energiforbrug i 2009. Denne andel dækker imidlertid over energiforbrug til rumopvarmning, opvarmning af varmt brugsvand og elforbrug. I dette afsnit beskrives hvilken udvikling der er sket inden for de enkelte forbrugsområder siden 1980'erne.

Energiforbrug til rumopvarmning og opvarmning af varmt brugsvand

Informationer om udviklingen af energiforbruget til rumopvarmning samt opvarmning af varmt brugsvand er oplyst af *Energistyrelsen* [27]. Af disse oplysninger er det ikke muligt at adskille de to forbrugsområder, da disse opgøres som ét samlet energiforbrug. Udviklingen af dette forbrug er illustreret på figur 2.3, hvor det fremgår, at der fra 1980 til 2009 er sket en

reduktion af det endelige energiforbrug, opgjort per arealenhed, svarende til 33 % [27]. Det endelige energiforbrug dækker over den nyttiggjorte energi (nettoenergiforbruget) og eventuelle lokale energitab i husstandene. Såfremt der differentieres mellem nettoenergiforbrug og energitab i husstandene, fremgår det, at nettoenergiforbruget er stigende og energitabene er reduceret [27]. Reduktionen af det endelige energiforbrug, opgjort per arealenhed, kan blandt andet forklares ud fra en energioptimering af ældre boliger, udskiftning til mere energieffektive varmeanlæg samt et lavere energiforbrug i nye boliger [27]. Idet 81 % af energiforbruget i husholdningerne anvendes til rumopvarmning og opvarmning af varmt brugsvand, jævnfør figur 2.2, kan det sammenholdes med det faktum, at det endelige energiforbrug, der fremgår af figur 2.3, ikke er markant faldende. Det betyder, at der er et besparelspotentiale i varmemeforbruget inden for husholdningssektoren.



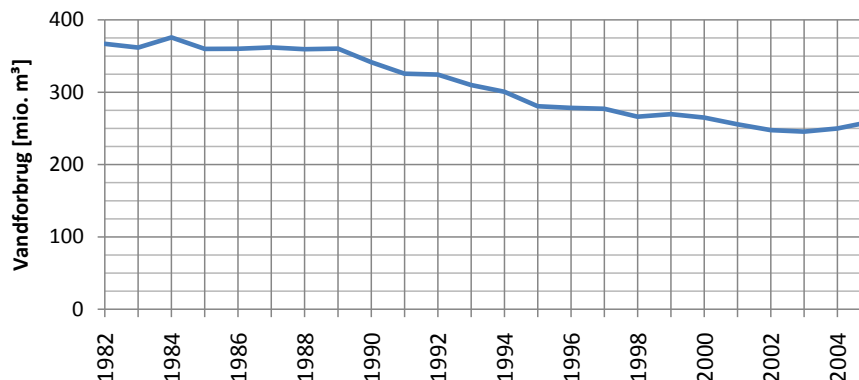
Figur 2.3: Udviklingen af energiforbruget til rumopvarmning og opvarmning af varmt brugsvand i danske boliger [27].

Vand- og varmtvandsforbrug

Det fremgår af den foregående beskrivelse, at energiforbruget til rumopvarmning og opvarmning af varmt brugsvand ikke kan adskilles i oplysningerne fra den aktuelle kilde. På denne baggrund er det relevant at anskue vand- og varmtvandsforbruget som et ressourceforbrug samt at vurdere udviklingen af dette. Forbruget af vandværksvand i husholdninger har været faldende fra 1989 og frem til 2003, hvorefter der er sket en stigning i forbruget [94]. I 2005 var der et samlet vandforbrug i husholdningerne på omkring 260 mio. m³, hvilket fremgår af figur 2.4 [94]. Hvad angår forbruget af varmt brugsvand, er der ikke i større udstrækning målt på dette. Erfaringer viser imidlertid, at omkring en tredjedel af det kolde brugsvand opvarmes til varmt brugsvand [8]. Tilsvarende viser erfaringer, at der i 1987 blev anvendt omkring 70 m³ per person, hvoraf det varme brugsvand udgjorde omkring 10 m³ per person [8]. Det samlede årlige vandforbrug var i 2007 reduceret til omkring 45 m³ per person, hvoraf forbruget af varmt brugsvand udgjorde omkring 15 m³ per person [8].

Til trods for at det gennemsnitlige årlige vandforbrug har været faldende siden 1980'erne, er det vigtigt at have yderligere besparelsmuligheder i fokus i fremtiden. Specielt når det af ovenstående fremgår, at det samlede vandforbrug er faldende, men forbruget af varmt brugsvand viser en stigende tendens, idet det er dette forbrug der har energimæssige omkostninger. Denne anskuelse skyldes det faktum, at når energiforbruget til rumopvarmning

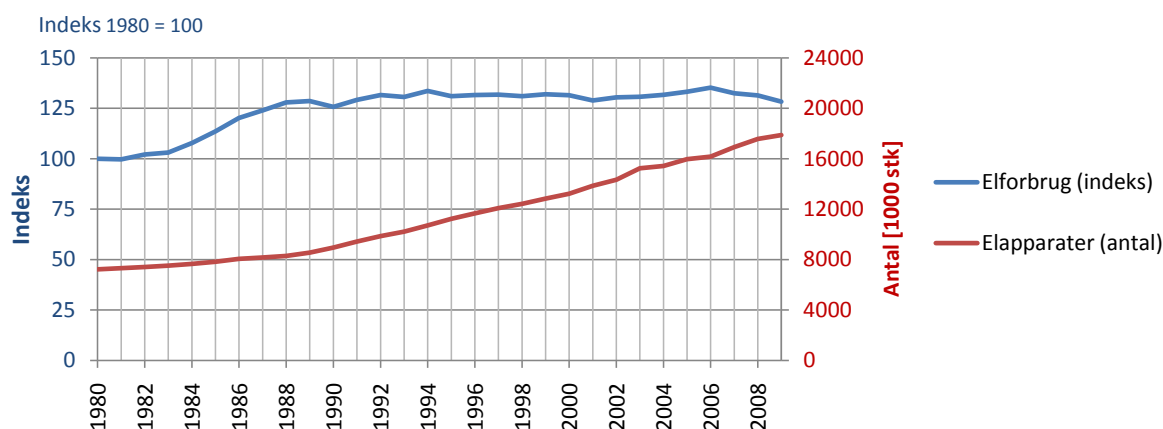
reduceres i fremtidige boliger, vil energiforbruget til opvarmning af varmt brugsvand udgøre en større andel af det samlede energiforbrug. Dermed vil vand- og varmtvandsforbruget have en større betydning i det samlede regnskab, end det tidligere har haft.



Figur 2.4: Udviklingen af forbruget af vandværksvand i danske husholdninger [94].

Elforbrug

Udviklingen af elforbruget i danske husholdninger er illustreret på figur 2.5. Af denne fremgår det, at der fra 1980 til omkring 1992 er sket en markant stigning. Siden 1992 har det gennemsnitlige elforbrug i danske husholdninger været relativt konstant, med enkelte mindre stigninger og fald [27]. Ikke desto mindre er det samlede elforbrug i husholdninger steget over 28 % fra 1980 til 2009 [27]. Dette svarer til en stigning fra 28,4 PJ i 1980 til 36,4 PJ i 2009 [27]. Det bemærkes, at elforbruget er relativt konstant til trods for en markant forøgelse af bestanden af elapparater i husholdninger¹. Stigningen i antallet af elapparater omfatter både tv, frydere, vaskemaskiner m.fl., men apparaterne er gennem årene blevet mere energieffektive, hvilket er en medvirkende årsag til det relativt konstante elforbrug [27].



Figur 2.5: Udviklingen af elforbruget og bestanden af elapparater i danske husholdninger [27].

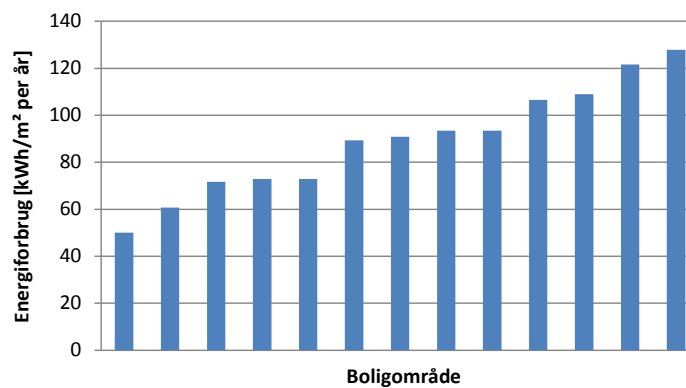
¹Bestanden er opgjort efter følgende elapparater: tv, køleskabe, frydere, mikrobølgeovne, opvaskemaskiner, vaskemaskiner samt tørretumbler [27].

2.1.2 Energiforbrugets variation mellem husstande

Den foregående beskrivelse har fokuseret på den gennemsnitlige udvikling af de enkelte energi- og ressourceforbrug i husholdninger. Undersøgelser viser imidlertid, at der er store variationer af energi- og ressourceforbrug husstande imellem. Med henblik på at klarlægge hvor store variationer der forekommer i praksis, beskrives enkelte af disse undersøgelser i det efterfølgende.

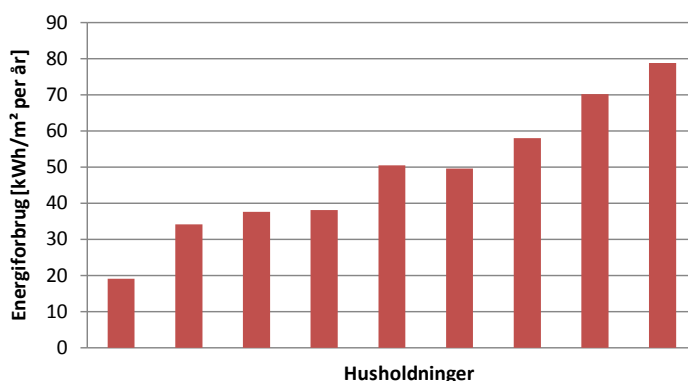
Energiforbrug til rumopvarmning

Der er, til illustrering af hvorledes energiforbruget til rumopvarmning varierer husstande imellem, inddraget resultater fra et dansk projekt af *Kirsten Gram-Hanssen* [40]. Disse resultater klarlægger variationen i det gennemsnitlige varmekonsum for 13 forskellige boligområder, som illustreret på figur 2.6. Det bør i denne sammenhæng bemærkes, at varmekonsumet dækker over energiforbruget til rumopvarmning og opvarmning af varmt brugsvand, da det ikke er muligt at adskille disse på baggrund af de tilgængelige oplysninger.



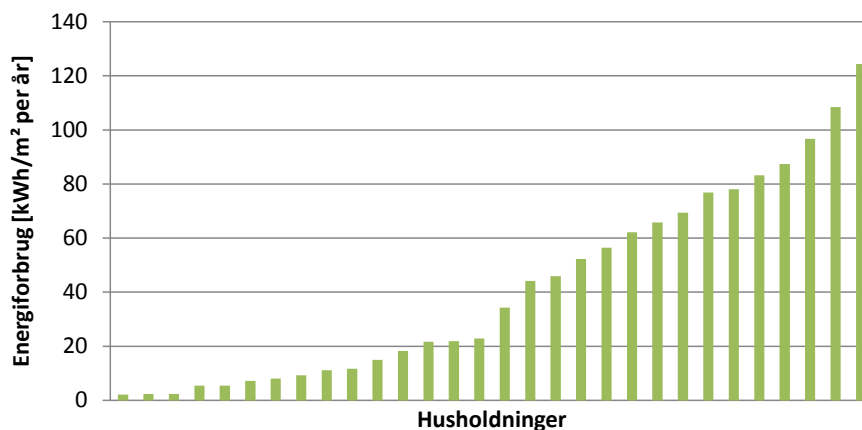
Figur 2.6: Gennemsnitligt energiforbrug til rumopvarmning og opvarmning af varmt brugsvand i 13 boligområder [40].

Den foregående beskrivelse tager udgangspunkt i den gennemsnitlige variation af varmekonsumet mellem boligområder. Til illustrering af hvorledes energiforbruget kan variere mellem enkelte boliger, er der inddraget målinger fra et projekt af *Knudsen et al.* [52]. Disse målinger stammer fra ni parcelhuse opført i 2006, som alle er klassificeret som lavenergiklasse 2 i henhold til det daværende bygningsreglement, og er illustreret på figur 2.7.



Figur 2.7: Energiforbrug til rumopvarmning og opvarmning af varmt brugsvand i ni parcelhuse [52].

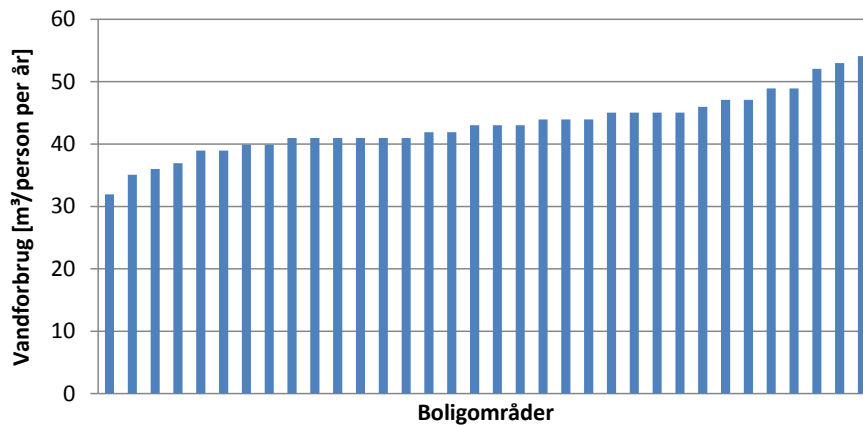
Med udgangspunkt i den foregående problemstilling med adskillelse af energiforbrug til rumopvarmning og opvarmning af varmt brugsvand, er der inddraget målinger fra et svensk projekt af *Jean Paul Zimmermann* [109]. Disse er anvendt, da der er foretaget separate målinger af energiforbruget til rumopvarmning, hvormed det er muligt at adskille disse. Målingerne er foretaget i parcelhuse, og er illustreret på figur 2.8. Af denne figur fremgår en betydelig variation af energiforbruget til trods for, at forbruget er opgjort ud fra det opvarmede etageareal.



Figur 2.8: Årligt energiforbrug til rumopvarmning i 30 svenske parcelhuse [109].

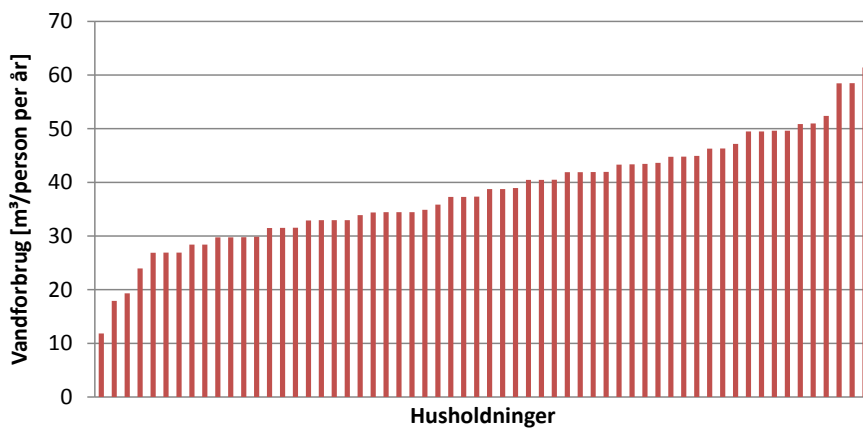
Vand- og varmtvandsforbrug

Der er generelt mangel på viden om forbruget af varmt brugsvand i boliger. Af denne årsag er det valgt ligeledes at undersøge størrelsen og variationen af det samlede vandforbruget. Et eksempel på dette er fra et dansk projekt af *Jesper Ole Jensen* [48], der omhandler 35 boligområder i og omkring Aarhus. Der har som minimum været 200 beboere i hvert af de boligområder, der er inddraget i projektet. På figur 2.9 er der illustreret en variation i det gennemsnitlige samlede vandforbrug i hvert enkelt boligområde. Det fremgår af figuren, at vandforbruget varierer fra omkring $32 \text{ m}^3/\text{person per år}$ til $55 \text{ m}^3/\text{person per år}$, hvor middelforbruget for samtlige boligområder er omkring $43 \text{ m}^3/\text{person per år}$.



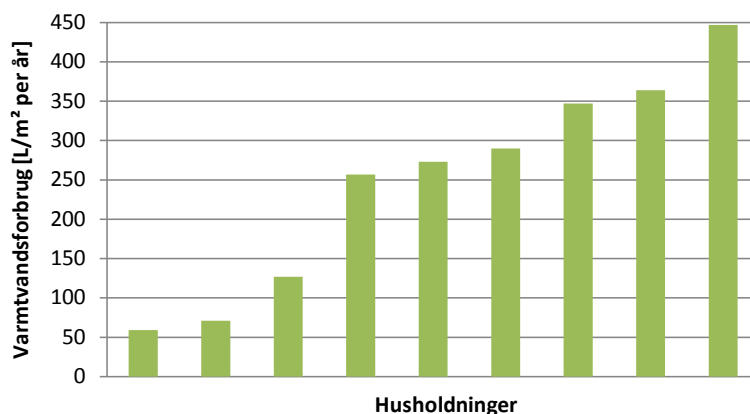
Figur 2.9: Gennemsnitligt årligt vandforbrug i 35 boligområder i Aarhus [48].

Såfremt der fokuseres på ét af boligområderne, der er inddraget i førnævnte projekt fremkommer en variation af de enkelte husstandes samlede vandforbruget som illustreret på figur 2.10. Af figuren fremgår det, at det gennemsnitlige vandforbrug i husstandene inden for ét boligområde varierer fra omkring $12 \text{ m}^3/\text{person per år}$ til $70 \text{ m}^3/\text{person per år}$.



Figur 2.10: Gennemsnitligt årligt vandforbrug husstande i ét boligområde i Aarhus [48].

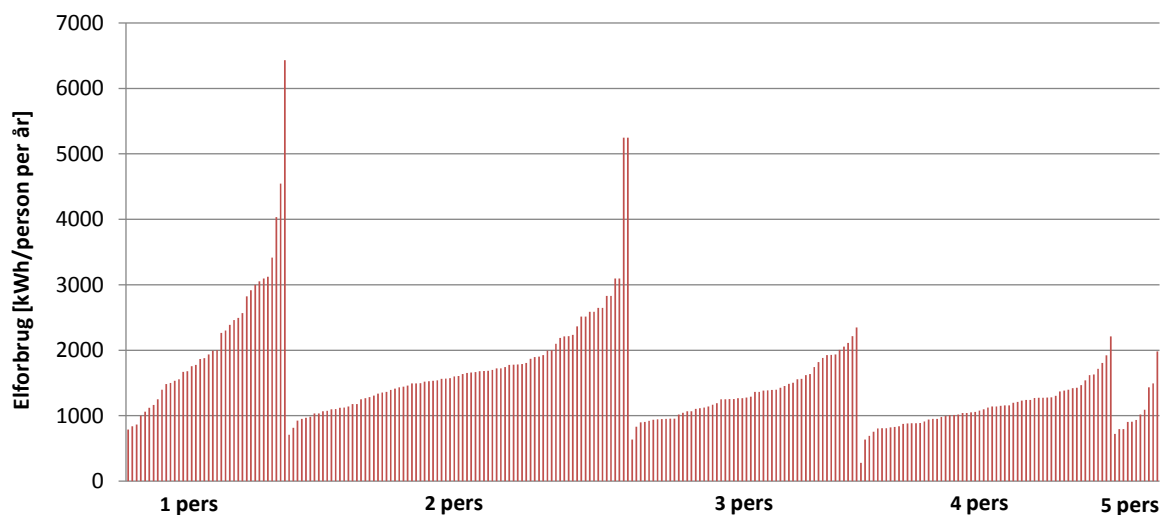
I det tidligere omtalte litteratur fra *Knudsen et al.* [52] er der målt på forbruget af varmt brugsvand i de ni parcelhuse. Af figur 2.11 fremgår en tydelig variation af forbrugsmængden, der anvendes i boligerne.



Figur 2.11: Varmtvandsforbrug i ni parcelhuse opgjort per arealenhed [52].

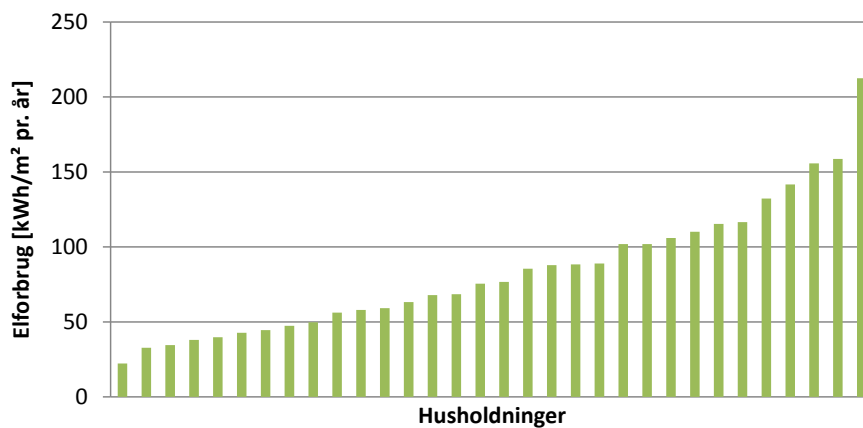
Elforbrug

I praksis forekommer der store variationer af elforbruget fra en bolig til en anden. Det fremgår eksempelvis af figur 2.12, der illustrerer det gennemsnitlige årlige elforbrug i husstande opdelt efter husstandsstørrelse. Af figur 2.12 fremgår der udelukkende målinger af elforbruget i ét boligområde.



Figur 2.12: Gennemsnitligt årligt elforbrug opgjort per person for forskellige husstandsstørrelser fra ét boligområde i Aarhus. [48].

Det ovenstående elforbrug er opgjort per person, men en tilsvarende variation fremkommer, såfremt elforbruget opgøres per arealenhed. Af figur 2.13 fremgår målinger for elforbruget i 32 svenske parcelhuse. På denne figur er der illustreret en betydelig variation af elforbruget blandt husstandene.

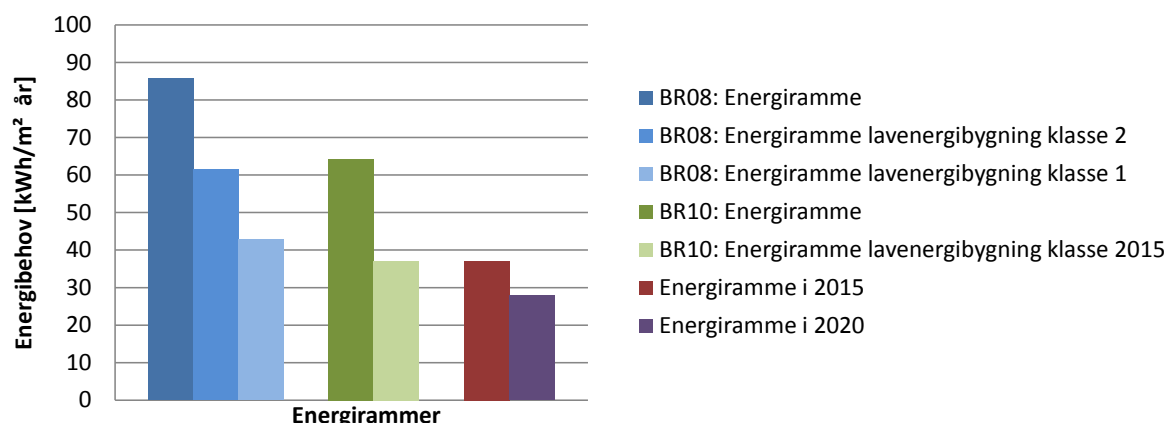


Figur 2.13: Årligt elforbrug opgjort per arealenhed fra 32 svenske parcelhuse [109].

Den foregående beskrivelse illustrerer, at der forekommer betydelige variationer mellem samtlige energiforbrug, når husstande sammenlignes. På baggrund af disse iagttagelser fremgår der i det efterfølgende afsnit en diskussion af, hvorvidt variationerne af energiforbrugene skyldes manglende restriktioner for boligernes energimæssige ydeevne.

2.1.3 Godkendelse af boligers energibehov

I de foregående afsnit er det illustreret hvor stor en andel af Danmarks samlede energiforbrug, der anvendes i husholdninger, samt hvor store variationer af energiforbruget, der kan registreres boliger imellem. Den observerede variation af energiforbruget til rumopvarmning er imidlertid ikke direkte et udtryk for manglende restriktioner om boligers energibehov. Af nuværende restriktioner er *Bygningsreglement 2010* [68], der er et normativt virkemiddel, som har til formål, at sikre en nedre grænse for kvaliteten af byggeriet i Danmark. Af denne fremgår der blandt andet mindstekrav til indeklima samt begrænsninger for størrelsen af energibehovet i boliger [21]. Energikravene til nybyggeri er skærpet gennem tiderne, specielt siden indførelsen af *Bygningsreglement 1998* inkl. tillæg [65], hvor energirammer blev introduceret i 2006. Af figur 2.14 fremgår udviklingen af energikravene til boliger siden 2006 og frem til de kommende krav i 2020.



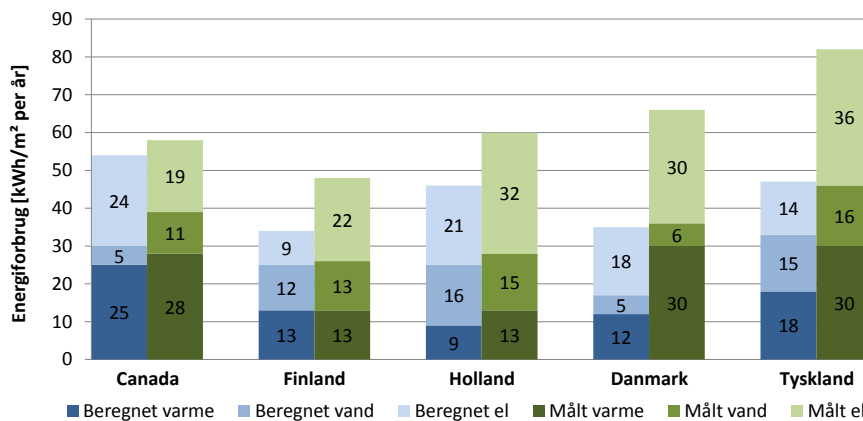
Figur 2.14: Udviklingen af krav til energibehov i det danske bygningsreglement fra 2006 og frem til de fremtidige krav i 2020 [66, 68].

Boligers energibehov skal godkendes af myndighederne på baggrund af en energirammeberegning udført i energiberegningsprogrammet Be10 [14]. I energirammeberegninger inddrages energibehov til drift af boliger så som rumopvarmning, produktion og distribution af varmt brugsvand, ventilation m.fl. De førnævnte forhold medtages på baggrund af en række standardiserede beregningsforhold. Hvad angår energibehovet til rumopvarmning, er der fastlagt hvilke inde- og udetemperaturer, der skal beregnes og dimensioneres efter. Til beregning af energibehovet anvendes der almindeligvis et setpunkt for rumopvarmning på $20\text{ }^{\circ}\text{C}$, hvilket betyder, at samtlige rum antages opvarmet til en månedlig gennemsnitstemperatur på $20\text{ }^{\circ}\text{C}$ i alle årets måneder [19]. Med hensyn til vandforbruget medtages der i energirammeberegninger udelukkende den del, der opvarmes til varmt brugsvand. I disse beregninger antages et årligt forbrug af varmt brugsvand $250\text{ L/m}^2\text{ opvarmet etageareal}$, hvor det dog gælder, at forbruget som minimum skal være $15\text{ m}^3/\text{år}$ og maksimalt $60\text{ m}^3/\text{år}$ [1]. Foruden energibehov til rumopvarmning og produktion af varmt brugsvand inddrages nogle enkelte elforbrug. Blandt disse er eksempelvis elforbrug til ventilation, pumper anvendt i vand- og varmekredse, opvarmning af brugsvand samt ved eventuel mekanisk køling. Derved undlades alt øvrigt elforbrug fra belysning og interne elapparater, der forekommer i boligen.

Ud fra den ovenstående beskrivelse fremgår det, at boligers energibehov godkendes på baggrund af en række standardiserede beregningsforhold. Disse forhold er fordelagtige at anvende i godkendelsessammenhænge, da det medfører, at forskellige boliger godkendes på baggrund af de samme forudsætninger. Det er ikke ensbetydende med, at forudsætninger i godkendelsesforløbet stemmer overens med hvorledes forholdene er i praksis. Det skyldes, at beboerne har betydelig indflydelse på hvorledes boligen og dens installationer anvendes. Da disse ikke medtages i godkendelsesforløbet, vil der naturligvis forekomme afvigelser mellem de to situationer. Afvigelserne er ikke et udtryk for, at beboernes adfærd bør medtages i godkendelsesforløbet, da det er bygnings ydeevne, der skal godkendes, og ikke beboernes anvendelse af denne. Af denne årsag bør der fokuseres på hvorledes beboerne kan rådgives, således de har den nødvendige viden for hvorledes energiforbruget i boligen optimeres.

2.2 Adfærdens påvirkning på energiforbruget i boliger

For at have mulighed for at rådgive beboere om hvilken indflydelse deres adfærd har på energiforbruget, er det i den sammenhæng nødvendigt at foretage beregninger eller simuleringer. De virkelige forhold er imidlertid komplekse at vurdere på forhånd, hvilket kan eksemplificeres ud fra tidligere undersøgelser. Et eksempel på det førnævnte kan illustreres ud fra et internationalt projekt, hvor fem forskellige parcelhuse er opført i fem forskellige lande. Inden boligerne blev opført, blev der foretaget simuleringer for at vurdere, hvilket energiforbrug der kunne forventes i boligerne. De beregnede og målte energiforbrug for de fem boliger fremgår af figur 2.15. Energiforbrugene er beregnet ved hjælp af flere forskellige simuleringsprogrammer², hvor der generelt fremkommer afvigelser mellem de beregnede og målte energiforbrug. Hvad angår afvigelsen for energiforbruget til rumopvarmning for projektet i Canada, er årsagen angivet til at være problemer med effektiviteten på varmeanlægget [96]. Derimod er der relativt store afvigelser på energiforbruget til rumopvarmning i projekterne fra Danmark og Tyskland. I begge tilfælde angives beboernes adfærd som den primære årsag til afvigelserne, da lufttemperaturen er 2 - 3 °C højere end forventet [96]. Der har dog yderligere været enkelte problemer med ventilationsanlæggene, der har medvirket til afvigelserne [96]. Elforbruget er underestimeret i samtlige tilfælde med undtagelse af det canadiske projekt, hvor det er overestimeret. Gennemgående for alle de underestimerede tilfælde er, at det øgede elforbrug primært skyldes beboernes adfærd, men med en mindre indflydelse fra effektiviteten på de tekniske installationer [96]. Effektivitet på de tekniske installationer er imidlertid årsagen til det lavere forbrug end forventet i boligen i Canada, idet den anvendte ventilator er mere energieffektiv end angivet i beregningerne [96].

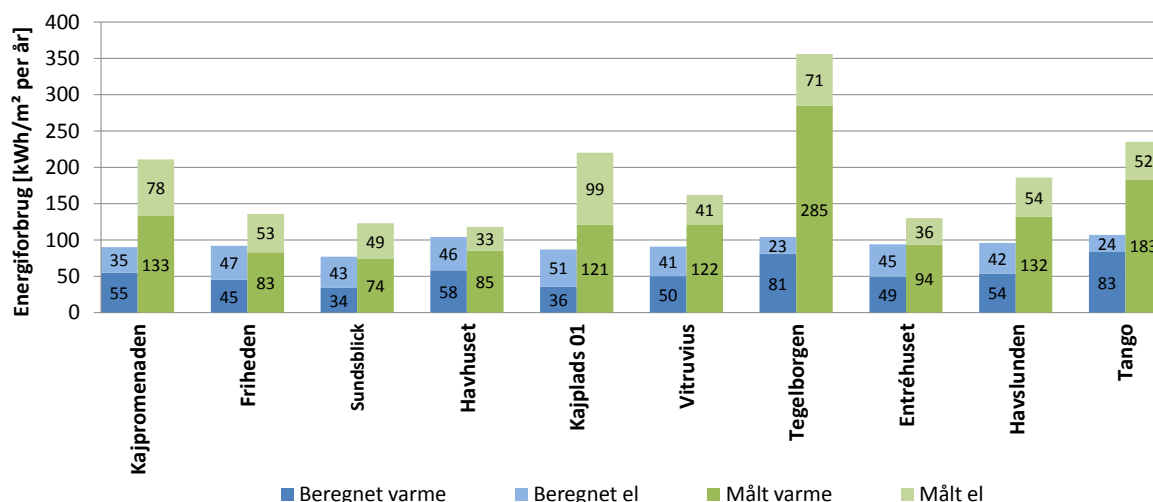


Figur 2.15: Beregnet og målt energiforbrug i fem forskellige boliger opført i fem forskellige lande [96].

Et andet eksempel, som illustrerer den vanskelige forudsigelse af energiforbruget i boliger, fremgår af figur 2.16. Disse oplysninger stammer fra et projekt i Malmö, hvor energiforbruget er beregnet ved hjælp af energiberegningsprogrammet Enorm 2004 [30]. Dette program er en svensk pendant til det danske energiberegningsprogram Be10. Af den førnævnte figur fremgår en, i flere tilfælde tydelig, afvigelse mellem de estimerede og de målte energiforbrug [63]. Afvigelserne mellem de forventede og de aktuelle energiforbrug skyldes flere forskellige

²Enerpass, TRNSYS, ESPr, Suncode og tsbi3.

forhold. Energiberegningsprogrammet Enorm 2004 tager eksempelvis ikke højde for varmeakkumulering og temperatursvingninger gennem døgnet, samtidig med at bidrag fra solstråling overvurderes [30]. Beboernes adfærd er ligledes en medvirkende årsag til afvigelserne, da lufttemperaturen i flere tilfælde er højere end forventet [63]. En anden forklaring er, at der i flere af boligerne har været problemer med de tekniske installationer og ikke mindst uopmærksomhed i udførelsen af bygningerne [63]



Figur 2.16: Beregnet og målt energiforbrug i ti forskellige boligprojekter opført i den bæredygtige bydel Bo01 i Malmö [63].

De ovenstående eksempler illustrerer, at det i flere tilfælde er vanskeligt at estimere hvilket energiforbrug, der kan forventes i boliger. I nogle af de foregående tilfælde angives beboernes ønskede lufttemperatur at være årsagen til afvigelsen mellem de forventede og de faktiske forbrug. Hvilken lufttemperatur beboerne foretrækker i deres hjem kan variere betragteligt. Ud fra grundlæggende studier om beklædningens varmeledningsmodstand, aktivitetsniveau og termisk komfort, kan vejledende værdier fastlægges. Et eksempel herpå er en stillesiddende person med et aktivitetsniveau på 1,0 *met* og beklædning svarende til 1,0 *clo*. Den optimale operative indetemperatur for denne person vil være omkring 23 °C [18].

Undersøgelser viser imidlertid, at der kan være store variationer af lufttemperaturen i et rum, afhængig af hvilken funktion de enkelte rum i boligen har. Der er en tendens til, at beboerne ofte foretrækker lufttemperaturer, der er højere end 20 °C, som almindeligvis anvendes i energiberegninger [13, 34, 97, 98, 99, 100]. Det medfører, at energibehovet til rumopvarmning undervurderes. Betydningen af dette er forskellig fra bolig til bolig, men i flere projekter er der vurderet på dette. Vurderingerne lyder på, at energiforbruget til rumopvarmning vil stige fra 7 % [70], 8 % [97, 98], 12 % [100] og op til 20 % [34] per grad højere lufttemperatur. De store forskelle i hvor meget ekstra energi, der skal anvendes ved at hæve lufttemperaturen, skyldes forskelle i boligens tekniske egenskaber. Eksempelvis omhandler projektet, der angiver en stigning på 20 %, passivhuse, hvilket vil sige, at der i forvejen anvendes en beskedne mængde energi til rumopvarmning, hvormed stigningen procentmæssigt vil være betragtelig. Beregningsforudsætningen for indetemperaturen, der anvendes i de nuværende Be10-beregninger, er god i forhold til opnåelse af godkendelse til

boligen, men ikke optimal i forhold til målet med skærperne i *Bygningsreglementet*. Som det fremgår af den foregående beskrivelse, anvendes der ofte en højere lufttemperatur end den der er angivet i beregningsforudsætningerne. Ved at anvende en forhøjet indetemperatur i beregningerne imødegås målet om at reducere energiforbruget i boliger, idet der tages højde for en fejlkilde, der introduceres af beboerne, som har stor betydning for energiforbruget til rumopvarmning. Herigennem vil boligerne blive godkendt på en mere reel baggrund, hvor der tages højde for at de praktiske forhold ikke altid stemmer overens med de teoretiske forudsætninger.

Beboerne kan påvirke lufttemperaturen i boligen på flere forskellige måder. Det kan være ved at ændre setpunktet for varmeanlægget, ændre solafskærmningen, åbne og lukke vinduer etc. Alle disse ændringer vil have indflydelse på hvor meget energi, der anvendes til rumopvarmning de pågældende steder.

Den foregående beskrivelse indikerer, at det er vanskeligt at forudsige energiforbruget i boliger. Én af årsagerne til afvigelserne i energiforbruget til rumopvarmning kan være en fejl-estimering af lufttemperaturen. Der er imidlertid andre forhold, der tilsvarende kan påvirke energiforbruget, og det er derfor relevant at identificere hvilke forhold, der har størst indflydelse på det estimerede energiforbrug.

2.2.1 Væsentlige forhold for det estimerede energiforbrug

Med henblik på at identificere de forhold, der har den største betydning for variationen af energiforbruget i boliger, er der foretaget et studie af tidligere undersøgelser omhandlende dette emne. I appendiks A gennemgås nogle af de projekter, der fremhæver disse forhold, hvilket efterfølgende skal danne grundlaget for en udvælgelse, af hvilke parametre hovedfokus skal hvile på i dette projekt. Der undersøges både for indflydelsen af konstruktionsmæssige parametre, så som U-værdier for konstruktionsdele, infiltration og lignende, men også parametre, der er under indflydelse af beboerne, så som ønsket lufttemperatur, anvendelse af elektriske apparater mv. Derved kan der skabes et overblik over hvilke faktorer, der har størst indflydelse på energiforbruget i forskellige boliger.

I flere af projekterne, der er fremhævet i appendiks A, er der ved brug af simuleringer foretaget en følsomhedsanalyse for at identificere de mest indflydelsesrige parametre for energiforbruget. Resultaterne af hvilke parametre der har indflydelse på variationen af energiforbruget, samt hvor stor betydning de enkelte parametre har på denne variation, afhænger i høj grad af hvilke af disse, der er behandlet stokastisk. Egenskaberne, i form af eksempelvis den valgte fordelingsfunktion, middelværdi og standardafvigelse for de stokastiske parametre, har ligeledes indflydelse på hvor stor betydning de enkelte parametre tillægges. I forsøg på at tage højde for dette, anvendes der efterfølgende en *rangeringsfaktor* til at vurdere parametrenes betydning for variationen af energiforbruget på tværs af forskellige projekter. Rangeringsfaktoren beregnes, som angivet i formel 2.1, for hvert enkelt projekt. Derved fremkommer en rangering af parametrene i de enkelte projekter, hvor den parameter med størst indflydelse på variationen i energiforbruget får en værdi på 1,00. Den næstvigtigste parameter får en lavere værdi, som afhænger af antallet af stokastiske parametre, og så fremdeles. Rangeringsfaktoren beregnes for samtlige parametre, hvorefter middelværdien for hver enkelt parameter beregnes. Det vil sige, at indgår en bestemt parameter i to projekter, beregnes rangeringsfaktoren i disse to projekter, og middelværdien af de to rangeringsfaktorer fastlægges efterfølgende.

$$RANG_i = \frac{N + 1 - p_i}{N} \quad (2.1)$$

Hvor:

$RANG_i$	Rangeringsfaktor for parameter i [-]
N	Antal parametre i litteratur [-]
p_i	Parameter i 's rangering i litteratur [-]

Et eksempel på hvorledes rangeringsfaktoren beregnes fremgår af det efterfølgende eksempel. Parameteren *varmebelastning fra apparatur* (omtalt som VFA i formel 2.2) fremgår i tre forskellige projekter, hvor disse projekter er præsenteret nærmere i appendiks A. Denne parameter er i de tre projekter rangeret som henholdsvis 4 ud af 5, 5 ud af 10 og 10 ud af 10, hvilket medfører, at rangeringsfaktoren beregnes på følgende måde, jf. formel 2.1:

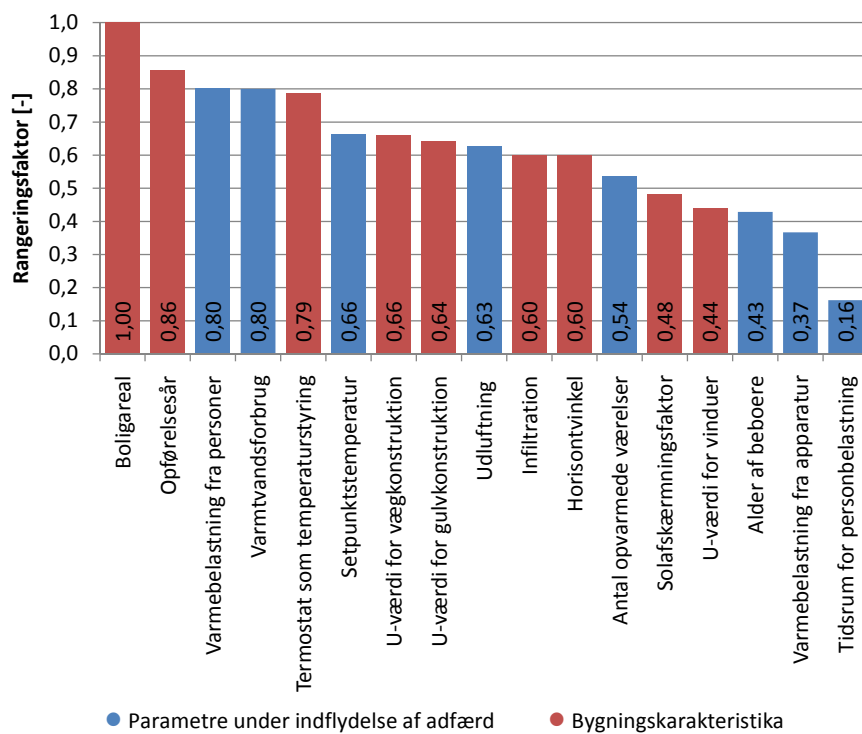
$$RANG_{VFA} = \frac{1}{3} \left(\frac{5 + 1 - 4}{5} + \frac{10 + 1 - 5}{10} + \frac{10 + 1 - 10}{10} \right)$$

$$RANG_{VFA} = 0,37 \quad (2.2)$$

Det fremgår af ovenstående eksempel, at rangeringsfaktoren tager højde for hvilken indflydelse den pågældende parameter har sammenholdt med de øvrige parametre. Den tager imidlertid ikke direkte højde for, hvor stor betydning den enkelte parameter har på variationen af energiforbruget. Det vil sige, at en parameter, som har en lille indflydelse på variationen i et projekt, kan få numerisk samme rangeringsfaktor, som en parameter i et andet projekt, der har en markant større indflydelse på variationen af energiforbruget i det aktuelle projekt. Derved kan anvendelsen af den valgte rangeringsfaktor diskuteres, men ikke desto mindre, anvendes den i dette projekt som indikation på hvilke parametre, der har den største betydning for variationen af energiforbruget i boliger.

Resultatet af dette studie er præsenteret grafisk på figur 2.17, hvor det fremgår, at det er parametrene *boligareal* og *opførelsesår*, der har den højeste rangering³. Såfremt der fokuseres på parametre, der er under indflydelse af beboernes adfærd, er det *varmebelastning fra personer* og *varmtvandsforbrug*, der er rangeret højest. Derimod er parameteren *tidsrum for personbelastning* den lavest rangeret, hvilket dog diskuteres senere i dette afsnit.

³Beregningerne fremgår af regnearket *Problemanalyse\Følsomhedsanalyse - rangering af parametre.xlsx* på den medfølgende DVD.



Figur 2.17: Rangering af parametre med indflydelse på bygningers energiforbrug.

Af resultaterne fra figur 2.17 fremgår det, at boligarealet er den parameter, der har størst betydning for variationen af energiforbruget. Dette dækker over, at jo større areal beboerne ønsker at anvende, jo større areal er der, som kræver opvarmning og belysning. Dette kan kædes sammen med parameteren *antal opvarmede værelser*, der ligeledes fremgår af figuren. Boligens opførelsesår er jævnfør figuren en væsentlig parameter, hvilket dog skal sammenholdes med forhold som U-værdier for konstruktionsdele og infiltration. Det skyldes, at ældre boliger typisk er mere utætte, og der er ofte anvendt dårligere isolerende konstruktioner sammenlignet med nyere boliger.

Solafskærpningsfaktoren kan, på baggrund af figur 2.17, vurderes at have indflydelse på energiforbruget til rumopvarmning. På figuren er denne parameter skitseret som en byggeteknisk parameter, men det bør i den sammenhæng bemærkes, at beboerne ligeledes har mulighed for at påvirke størrelsen af solafskærpningsfaktoren. Det skyldes, at beboerne kan justere denne og derved påvirke solafskærpningsfaktoren.

Foruden anvendelsen af solafskærmning har beboerne og deres brug af boligen en betragtelig indflydelse på energiforbruget. Varmebelastningen fra personer er en direkte konsekvens af beboernes tilstedeværelse. Dette forhold dækker over hvor ofte, hvor lang tid og hvor mange beboere, der er tilstede i boligen. Ydermere træffer beboerne ligeledes beslutninger om hvor ofte, og i hvor lang tid ad gangen, der skal luftens ud i boligen. Beboernes valg af hvor mange rum der skal opvarmes, samt den ønskede lufttemperatur i disse, har ligeledes en betydelig indflydelse på energiforbruget til rumopvarmning.

Udover energiforbruget til rumopvarmning, fremgår det af figur 2.17, at beboernes forbrug af varmt brugsvand har en betragtelig indflydelse af energiforbruget i boliger.

I den foregående beskrivelse er forhold som ejerformen af boligen ikke medtaget. Det bør vurderes, hvorvidt denne kan forklare dele af de forbrugsvariationer, der kan observeres for

energiforbruget i boliger. Da det ofte er tilfældet, at ejerboliger har et større boligareal end lejerboliger, kan dette forhold have indflydelse på eksempelvis antallet af hårde hvidevarer, bestanden af elapparater, behov for belysning og rumopvarmning mv. En undersøgelse indikerer, at boligernes byggetekniske stand kan være forskellig for ejer- og lejerboliger, hvor ejerboliger typisk er i en bedre energimæssig byggeteknisk stand [40]. Såfremt det er tilfældet, vil dette forhold have indflydelse på energiforbruget til rumopvarmning. Det kan imidlertid ikke forventes, at ejerformen har en direkte indflydelse på beboernes adfærd og entydigt kan forklare forbrugsforskelle mellem boligerne. Ikke desto mindre er det et aspekt, hvis påvirkning på energiforbruget bør undersøges.

2.3 Metoder til estimering af adfærdens påvirkning på energiforbruget

Med udgangspunkt i den foregående beskrivelse står det klart, at for at beboere kan rådgives om hvordan og i hvilken grad deres adfærd påvirker energiforbruget, er det nødvendigt, at der kan foretages beregninger eller simuleringer af dette. Der er flere metoder, der kan medvirke til at synliggøre beboernes påvirkning på energiforbruget. I de efterfølgende underafsnit beskrives nogle metoder, der kan anvendes til at opnå en estimering af energiforbruget i boliger. Eksemplerne er begrænset til metoder der enten anvendes eller kan anvendes i Danmark.

2.3.1 Be10

Energiberegningsprogrammet Be10 [14] anvendes på nuværende tidspunkt til beregning af energibehovet i bygninger og som godkendelsesværktøj ved ny- eller ombygning. Da programmet anvendes til myndighedsgodkendelse af bygninger, benyttes der, som tidligere beskrevet, en række standardiserede beregningsforhold. Det rejser imidlertid en række problemstillinger, såfremt programmet skal anvendes til at vurdere beboernes indflydelse på energiforbruget. Der er eksempelvis kun en begrænset mulighed for at angive realistiske forhold i energiberegningsprogrammet. Dette skal forstås således, at mængden af varmt brugsvand i boliger kan varieres i Be10-beregningerne, hvorimod bestand og anvendelse af elapparater ikke direkte kan medtages i programmet. Da forholdene omkring elapparater har stor betydning for det samlede energiforbrug i boliger, og vil udgøre en større andel i fremtiden, er det problematisk, at de ikke kan medregnes.

Energiberegningsprogrammet Be10 anvender månedsmiddeltemperaturer i beregningerne, hvilket betyder, at variationer af udetemperaturen reduceres betragteligt. I samme forbindelse anvendes ofte et gennemsnitligt internt varmetilskud fra elapparater, hvilket betyder at dette på nogle tidspunkter underestimeres. Samtidig anvendes programmet hovedsagligt til estimering af energiforbruget, hvorfor der kun tages højde for indeklimaet i bygningerne i meget begrænset omfang, i form af estimering af overtemperaturer. Det kan have den uheldige konsekvens, at der i nogle tilfælde fokuseres mest på at reducere energibehovet, hvormed der kun tages højde for indeklimaforholdene i begrænset omfang.

På baggrund af de ovenstående betragtninger, må det anses som en nødvendighed, at energiberegningsprogrammet udvikles yderligere, såfremt dette skal anvendes med henblik på at vurdere beboernes indflydelse på energiforbruget. Det er imidlertid en klar fordel, at pro-

grammet i forvejen er udbredt, hvis det kan udvikles yderligere, således en bedre estimering af energibehovet kan opnås. Ydermere kan det have den fordel, at når der alligevel skal udføres en energirammeberegning, kan der på samme tidspunkt opnås indsigt i hvordan beboere og deres adfærd påvirker det samlede energiforbrug. Det kræver imidlertid, at der er opmærksomhed på, at boligerne ikke optimeres til nogle bestemte beboeres adfærd. Hvis det sker, kan det resultere i boliger, der er mindre robust over for ændringer i adfærd, hvormed det kan medføre et større energiforbrug, når nogle andre beboere overtager boligen.

En anden fordel ved energiberegningsprogrammet er, at det er relativt simpelt og hurtigt at anvende. Inputparametrene for de enkelte systemer er begrænset til en overskuelig mængde, hvor der i flere tilfælde kan anvendes standardværdier. Ligeledes kan simplificeringerne i energiberegningsprogrammet være en fordel i forbindelse med simuleringstiden, når energibehovet skal beregnes. Simplificeringerne betyder, at effekten af en parameterændring hurtigt fremgår af resultaterne, hvilket kan være fordelagtigt gennem en optimeringsproces. Ulemper ved simplificeringerne er, at nøjagtigheden af resultaterne reduceres, hvormed brugerne af programmet skal have større opmærksomhed på, i hvilken sammenhæng resultaterne anvendes.

De ovenstående betragtninger fremhæver nogle fordele og ulemper ved energiberegningsprogrammet. Disse betragtninger bør tages til overvejelse, hvis programmet skal videreudvikles til at danne grundlag for estimering af beboernes påvirkning på energiforbruget. Dette skal blandt andet ses i sammenhæng med den politiske målsætning om at reducere energiforbruget i bygninger, men uden at gå på kompromis med indeklimaet.

2.3.2 BSim

Simuleringsprogrammet BSim [15] anvendes på nuværende tidspunkt hovedsagligt i forbindelse med større eller komplekse projekter, for at vurdere indeklimaet i bygninger. I modsætning til Be10 stilles der ingen krav om en BSim-simulering ved ny- eller ombygning, hvilket kan være en medvirkende årsag til, sammen med det faktum at programmet er tidskrævende at anvende, at programmet anvendes i mindre omfang. BSim har en række fordele og ulemper, der beskrives nærmere i det efterfølgende.

Sammenholdt med Be10 er der i BSim flere muligheder med hensyn til variation og mængde af inputdata. I BSim er der mulighed for at anvende en større mængde inputdata, hvilket resulterer i, at de aktuelle systemer i bygningen kan modelleres mere virkelighedstro. Et større og mere specifikt datainput betyder, at der kan foretages realistiske beregninger af energiforbrug og indeklima, der kan være i bedre overensstemmelse med virkeligheden sammenholdt med en Be10-beregning.

Det høje detaljeringsniveau kommer blandt andet til udtryk ved, at flere inputdata og resultater opgøres på timebasis. Dette kan danne et solidt grundlag for en efterfølgende analyse og optimering af bygningen. Det betyder, at indeklimaet kan analyseres mere detaljeret end i Be10, og det fremgår på timebasis, hvorledes indeklimaet ændres afhængigt af eksempelvis udeforhold og interne belastninger.

I BSim er der mulighed for at inddrage beboernes adfærd, hvilket ikke direkte er muligt i Be10. Det kommer eksempelvis til udtryk ved, at der kan anvendes bestemte dagsprofiler for de interne belastninger, setpunktstemperaturen for rumopvarmning kan varieres, og parametrene for hvor ofte og i hvor høj grad solafskærmningen kan justeres.

På trods af at BSim har en høj detaljeringsgrad på nogle områder, kan eksempelvis vandforbrug og energiforbrug til opvarmning af varmt brugsvand ikke medregnes i programmet. Såfremt simuleringsprogrammet anvendes til beregning af energiforbruget i boliger, skal der, foruden BSim-simuleringen, udføres en beregning af energiforbruget til opvarmning af varmt brugsvand.

Ud fra de foregående beskrivelser er der fremhævet nogle fordele og ulemper. Fordelene ved BSim er, at der er mulighed for at inddrage en større mængde information om brugeradfærd, specielt i forhold til rumopvarmning end i Be10. Det betyder, at der er mulighed for i samme program at opnå viden om bygningens tekniske egenskaber i forhold til rumopvarmning, både med og uden hensyntagen til beboernes adfærd samt opnå resultater om boligens indeklimamæssige forhold. Den største ulempe ved BSim er, at det er betydeligt mere tidskrævende at anvende sammenholdt med Be10. Det gælder både ved opsætningen af simuleringen, hvor der indledningsvis skal indtastes betydelige mængder data for konstruktioner og systemer. På tilsvarende vis er tidsforbruget for simuleringerne større som funktion af kompleksiteten af modellen. Ud over det tidsmæssige aspekt opnås der ingen resultater omkring forbruget til opvarmning af varmt brugsvand ej heller information om elforbruget i boligen. Dette vil kræve en videreudvikling af programmet.

De ovenstående betragtninger bør tages til overvejelse, hvis BSim skal anvendes til at vurdere beboernes indflydelse på energiforbrug og indeklima.

2.3.3 Internetbaserede energiberegningsprogrammer

Der eksisterer på nuværende tidspunkt talrige programmer på internettet, der er frit tilgængelige, og som kan anvendes til estimering af de forskellige energiforbrug i boligen [28, 46, 67, 69, 83, 87]. Fælles for hovedparten af denne type beregningsprogrammer er, at der ikke umiddelbart er tilgængelige beskrivelser for datagrundlaget og beregningsproceduren i disse. Derved er det vanskeligt at vurdere nøjagtigheden af beregningerne og troværdigheden af beregningsresultaterne. Flere af de internetbaserede energiberegningsprogrammer illustrerer hvor store energibesparelser, der kan opnås ved eksempelvis energirenoveringer eller anvendelse af energibesparende elapparater. Ved anvendelse af denne type energiberegningsprogrammer bør det tages op til overvejelse, hvorvidt udgiverne af programmet har interesse i beregningsresultatet. Det kan eksempelvis diskuteres om isoleringsproducenter har interesse i, at overslagsberegninger for energibesparelser udtrykker større besparelser, end hvad der kan forventes i praksis. Det skal imidlertid ikke tolkes således, at beregningsprogrammerne nødvendigvis fejkalkulerer energiforbruget, men der bør eksempelvis rettes opmærksomhed på, om forudsætninger før og efter en energirenovering er de samme, eller om disse favoriserer det sidstnævnte tilfælde. Samtidig skal det vurderes, hvor omfattende og virkelighedstro datagrundlaget for beregningerne er, og om det er brugerne af programmet, der skal oplyse flere væsentlige forhold. Såfremt brugerne af energiberegningsprogrammet skal indtaste oplysninger om centrale forhold for beregningen, og denne ikke har fuldt kendskab til forholdene, eller har viden om betydningen af oplysningernes korrekthed, kan disse påvirke resultatet betragteligt.

Blandt fordelene ved internetbaserede energiberegningsprogrammer er, at disse i flere tilfælde er udviklet med henblik på, at almene boligejere kan anvende dem. Det betyder, at programmerne ofte har en simpel brugerflade samtidig med at datagrundlaget, som brugerne af programmet skal indtaste, er begrænset til en overskuelig mængde. Flere af programmerne

angiver i resultaterne hvor store besparelser, oplyst enten i en energimængde eller økonomisk størrelse, der kan forventes ved en energirenovring, udskiftning af elapparater, adfærdsændring mv. Hvis beboernes selv skal anvende programmerne, er det en klar fordel, at de er frit tilgængelige for alle, men det kræver, at beboerne er opsøgende for at finde disse.

Der er, som beskrevet i det foregående, store fordele ved anvendelse af denne type energiberegningsprogrammer. Såfremt denne type programmer anvendes til rådgivning af beboere, er der imidlertid flere forhold, som må tages op til overvejelse for at sikre den bedste rådgivning for de enkelte beboere. Blandt disse forhold er, at programmet skal være let anvendeligt, men stadig inddrage væsentlige byggetekniske og adfærdsmæssige forhold. Ydermere bør det udvikles på baggrund af en kombination af teoretisk viden samt erfaringer/viden fra virkelige forhold, for at de bedste resultater fremkommer.

2.3.4 Udvikling af nyt beregningsprogram

Det kan være en mulighed at udvikle et nyt beregningsprogram, der har de funktioner, der ønskes i forbindelse med de ovenstående problemstillinger og betragtninger. Et beregningsprogram, der tager højde for energiforbrug til rumopvarmning, produktion af varmt brugsvand samt det interne elforbrug samtidig med at det kan estimere indeklimaet i bygningen, vil være optimalt. Dette program kan eksempelvis være en videreudvikling af BSim, idet dette program allerede kan beregne to ud af de fire nævnte egenskaber. Opmærksomheden skal rettes mod brugervenligheden af programmet, men det skal gøres uden at gå på kompromis med nøjagtigheden af beregningerne.

I forbindelse med brugervenligheden af programmet, kan det være en mulighed, at anvende en ekstern fil, til at opbevare standardinput for diverse produkter, eksempelvis ventilatorer, varmtvandsbeholdere eller lignende. Derved har de enkelte brugere af programmet mulighed for at indtaste data, der passer til de produkter de ofte anvender i simuleringerne, da der er en tendens til, at samme type produkter anvendes til flere forskellige projekter.

Der kan være både fordele og ulemper ved at inddrage samtlige energiforbrug i beregningerne. Fordelen ved at inddrage samtlige energiforbrug er, at der kan skabes overblik over størrelsen og betydningen af de enkelte forbrug set i sammenhæng med det samlede energiforbrug. Der er en vis usikkerhed forbundet med hver parameter, der anvendes i simuleringerne, og jo flere inputdata der anvendes, jo større usikkerhed kan der være forbundet med resultaterne. Dog bliver de nuværende usikkerheder bragt frem i lyset og behandlet for i fremtiden at kunne opnå bedre resultater. En af metoderne til at afhjælpe denne problemstilling er, at angive flere parametre ud fra sandsynlighedsfordelinger frem for enkelte diskrete værdier. Det vil resultere i, at der kan tages højde for flere usikkerheder i beregninger, og resultaterne kan angives ud fra, hvor stor sandsynlighed der er for at det aktuelle forbrug forekommer. Dette kan imidlertid skabe nogle problemstillinger i forbindelse med vurderingen af hvilken sandsynlighedsfordeling, og hvilke sandsynlighedsparametre de enkelte forhold skal defineres ved. Det kan imødekomes ved, at der udvikles nogle sandsynlighedsfordelinger, som knytter sig til nogle bestemte beboer- og adfærdstyper. En uddybning af hvorledes dette kan udføres fremgår af appendiks B.

Hvis det nye beregningsprogram kan inddrage samtlige energiforbrug, samtidig med at indeklimaet simuleres, dannes grundlaget for en bedre helhedsforståelse for bygningen og betydningen af de enkelte påvirkninger. Samtidig kan én beregningsmodel give oplysninger om

flere forskellige forhold, som kan danne grundlaget for en mere hensigtsmæssig optimering af bygningen. Således kan der fokuseres på både bygningens energimæssige ydeevne samt indeklimaet. Er der samtidig mulighed for at inddrage brugeradfærd i et større omfang, end det er tilfældet i de to tidligere nævnte beregningsprogrammer, er der mulighed for at opnå en bedre estimering af bygningers samlede energiforbrug, samtidig med at indeklimaet tages i betragtning.

2.4 Interesseparter

Før udvikling af en metode eller et program, bør det tages op til overvejelse, hvilken målgruppe metoden til estimering af adfærdens påvirkning på energiforbruget skal henvende sig til. Med udgangspunkt i den foregående beskrivelse er der flere forskellige interesseparter, for hvem det kan være relevant at synliggøre beboernes påvirkning på energiforbrug og indeklima i boliger. En beskrivelse af de væsentligste interesseparter udføres i det efterfølgende.

Boligejere

Boligejere er den gruppe, som kan drage direkte nytte af, at der foretages undersøgelser om samspillet mellem beboere, energiforbrug og indeklima. Ved at bidrage med informationer om dem selv, kan de inddrages i estimeringer af et forventet energiforbrug. Dette kan eksempelvis være oplysninger om tidligere energiforbrug, holdninger til miljø-/energibesparelser og beskrivelse af deres adfærd. Disse oplysninger kan anvendes i estimeringerne, for at skabe indsigt i hvilke konsekvenser beboernes handlinger har på energiforbruget, samt hvordan disse kan optimeres. Benyttes samme lejlighed til at tilbyde beboerne rådgivning om indkøb af elbesparende apparater og optimal anvendelse af boligen, kan der muligvis opnås en holdnings- eller adfærdsændring hos beboerne.

Rådgivende ingeniører og arkitekter

En anden gruppe, der kan drage nytte af en større viden inden for området brugeradfærd i boliger, er rådgivende ingeniører og arkitekter. På nuværende tidspunkt anvendes der inden for disse fagområder flere forskellige metoder til estimering af energiforbruget, som eksempelvis Be10 og BSim. Med en større viden om beboernes påvirkning på energiforbruget og indeklimaet i boliger, kan rådgivende ingeniører og arkitekter samarbejde om, at finde de mest fordelagtige tekniske løsninger, så der anvendes mindst muligt energi, under hensyntagen til hvordan denne anvendes i praksis. Det kan overvejes, om en beskrivelse af optimalt anvendelse af boligen skal udarbejdes, for at bistå beboernes eventuelle spørgsmål, og for at videregive informationen til de efterfølgende beboere.

Energirådgivere

Denne faggruppe har til formål at rådgive blandt andet boligejere om, hvordan energiforbruget i boligen kan reduceres. På denne baggrund vil en større viden om beboeres påvirkning på energiforbruget i boliger forbedre rådgivningen fra energirådgivere. Indføres energirådgivning til beboere ved enten nybyggeri eller renoveringsopgaver, kan beboerne derigennem få fremlagt hvorledes bygningen benyttes optimalt, og hvordan de største besparelser opnås.

Ejendomsmæglere

Såfremt der udarbejdes nogle beskrivelser af beboernes påvirkning på energiforbruget, eller nogle relativt simple beregningsprogrammer, er der mulighed for, at ejendomsmæglere kan vejlede købere om, hvilket energiforbrug de kan forvente at få. Det kan, sammen med ener-

gimærkningen som blev indført i 2006, gøre købere mere opmærksomme på energiforbruget i boligen.

2.5 Problemformulering

Der har gennem en længere periode været fokus på energibesparende tiltag i forbindelse med boliger. Dette fokus kommer blandt andet til udtryk gennem kampagner, der er iværksat med henblik på at reducere eksempelvis varme-, vand- og elforbruget, eller gennem den politiske aktivitet på området. Den danske regering har en målsætning om at reducere bruttoenergiforbruget med 4 % frem til 2020 [81]. For at nå denne målsætning må der blandt andet fokuseres på energiforbruget i husholdninger, da disse står for omkring en tredjedel af det samlede energiforbrug i Danmark. En del af målsætningen kan nås gennem de stramninger af blandt andet komponent- og energikrav for bygninger, som løbende indføres i bygningsreglementet. Stramningerne er en metode til at reducere energiforbruget i hovedsagligt nye bygninger samt ved større renoveringer. Disse stramninger vil blandt andet medvirke til, at energiforbruget til rumopvarmning reduceres i fremtidige boliger. På trods af at energiforbruget til rumopvarmning reduceres, vil der stadig være en variation heraf, som følge af beboeres forskellige ønsker og krav til indeklima. For at opnå regeringens målsætning er det væsentligt, at der fokuseres på samtlige energiforbrug for at udnytte besparelspotential bedst muligt. I den forbindelse bør det bemærkes, at elforbruget samt energiforbruget til opvarmning af varmt brugsvand vil antage en større del af det samlede energiforbrug i fremtidige boliger, som følge af det faldende energiforbrug til rumopvarmning.

Da det er under 1 % af bygningsmassen, der udskiftes per år, er det ikke nok udelukkende at indføre stramninger i bygningsreglementet [81]. Det betyder, at energiforbruget i eksisterende bygninger er et væsentligt forhold, når det sammenholdes med regeringens centrale målsætninger om at reducere energiforbruget. Energibesparelser i den eksisterende bygningsmasse kan opnås ved en energirenovering af bygningerne eller ved rådgivning af beboerne. I denne sammenhæng er det nødvendigt, at beboerne kan rådgives om hvilke foretag, der vil medføre de største energibesparelser - byggetekniske forbedringer såvel som adfærdsmæssige forhold. Denne rådgivning må imidlertid ikke medføre, at indeklimaet i boligerne forringes, såfremt der skal opnås holdbare løsninger. Da beboernes adfærd i boligen har stor betydning for energiforbruget, kan mindre ændringer af adfærden medføre relativt store ændringer af dette. For at beboerne kan modtage en optimal rådgivning, kræver det, at energiforbruget, indeklimaet og beboernes påvirkning på disse kan estimeres.

På baggrund af den foregående problemstilling er målet for dette projekt, at klarlægge hvilke erfaringer og undersøgelser der i forvejen er udført omhandlende beboeres påvirkning på energiforbruget i boliger. Denne viden skal inddrages i en diskussion omhandlende hvilke metoder, der kan anvendes til at estimere energiforbruget i boliger under hensyntagen til beboernes adfærd. Med udgangspunkt i denne diskussion udvælges én metode, hvorefter der udføres en eksemplificering af udviklingen og anvendelsen af denne.

Ved at undersøge erfaringer fra flere forskellige fagområder og synsvinkler opnås en bredere baggrundsviden, der kan danne et solidt grundlag for den efterfølgende diskussion af, hvorledes beboernes adfærd kan inddrages i estimeringen af energiforbruget. Nøglen til at opnå

den størst mulige succes med en bygning, set i relation til både indeklima og energi, er at se bygningen, beboerne og samspillet mellem disse som en helhed, således der kan udføres en optimering hvor robustheden af denne, med hensyn til energiforbrug, opretholdes.

2.5.1 Projektafgrænsning

En del af den danske regerings målsætninger er, som tidligere beskrevet, at reducere energiforbruget i Danmark. I projektet foretages en undersøgelse af hvorledes beboernes adfærd påvirker energiforbruget i boliger. Det betyder, at hovedfokus i projektet vil være på samspillet mellem beboernes adfærd og energiforbrugets størrelse. Med udgangspunkt i dette vil der i nærværende projekt primært fokuseres på ejerboliger i form af parcel-/rækkehuse. Da det er et særdeles omfattende fokusområde, afgrænses der i dette projekt fra direkte at foretage vurderinger af indeklimaet og beboernes indflydelse på dette. I projektet inddrages der trods alt forhold, som har betydning for indeklimaet i boliger, eksempelvis beboernes ønskede lufttemperatur og udluftningsvaner. Disse forhold behandles imidlertid med henblik på at vurdere indflydelsen på energiforbruget. Det må derimod ikke fremstå således, at indeklimaet i boliger må negligeres. Det skyldes, at undersøgelser har påvist, at der i flere lavenergi- og passivhuse er problemer med opretholdelsen af et tilfredsstillende indeklima [2].

Litteraturstudie

I problemformuleringen i afsnit 2.5 er det primære fokus lagt på at finde frem til en metode, der kan beskrive sammenhængen mellem energiforbruget i boliger og beboernes adfærd. For at imødegå denne problemstilling undersøges det hvilke analysemetoder, der med succes er anvendt til at løse lignende problemstillinger omhandlende energiforbrug og brugeradfærd i boliger.

Da formålet med dette projekt er, at anvende en metode, der kan estimere energiforbruget i boliger under hensyntagen til beboernes adfærd er det nødvendigt, at opnå indsigt i hvilke forhold, der har betydning for energiforbruget. Fra afsnit 2.2.1 står det, på baggrund af simuleringer og følsomhedsanalyser, klart, at forhold som varmtvandsforbrug, setpunkts-temperatur for rumopvarmning, udluftning, infiltration, solafskærmning og varmebelastning fra apparater er blandt de mest indflydelsesrige forhold i forbindelse med energiforbruget i boliger og variationen af dette. Da beboerne har, mere eller mindre, direkte indflydelse på alle disse forhold, er der grund til at undersøge hvilke analyser og målinger, der er udført for beboernes adfærd i forhold til de førnævnte forhold.

Dette kapitel er opbygget således, at der er et afsnit for henholdsvis energiforbruget til rumopvarmning, koldt-/varmtvandsforbrug og elforbruget. I hvert afsnit gennemgås relevante resultater, fremgangsmåder og iagttagelser fra andet litteratur, og de anvendte kilder sammenlignes, i det omfang det er muligt. De tre ovenstående afsnit afsluttes alle tre med en diskussion af den vigtigste information.

3.1 Metoder til analysering af data

Med udgangspunkt i ovenstående introduktion gennemgås der i dette afsnit flere rapporter og artikler, hvori der omtales forskellige metoder, der alle anvendes til analysering af datamængder. Foruden en beskrivelse af metodernes anvendelse gennemgås ligeledes hovedresultaterne fra de respektive undersøgelser, hvori det er forsøgt, at opdele eller klassificere beboerne i beskrivende brugertyper. Opdelingen af beboerne sker blandt andet ved at undersøge beboernes holdning til miljømæssige aspekter, deres adfærd i forbindelse med huslige aktiviteter, tekniske forhold om boligerne samt ved indsamling af informationer om demografiske forhold. Afsnittet indledes med en kortfattet gennemgang af adfærdsmodellen *Theory of Planned Behaviour*, der anvendes til at beskrive sammenhængen mellem holdninger og adfærd.

3.1.1 Theory of Planned Behaviour

Igennem flere analyser, hvor der anvendes spørgeskemaanalyser, er der ved udvikling af spørgeskemaerne anvendt en adfærdsmodel, der betegnes *Theory of Planned Behaviour* [101]. Adfærdsmodellen anvendes til at beskrive sammenhængen mellem menneskets holdninger og adfærd. Modellen anvendes primært inden for reklamebranchen og sundhedssektoren, men har i flere studier, af blandt andet *Faiers et al.* [32], *Mari Martiskainen* [57] og *Wilson et al.* [107], vist sig at være anvendelig til at forudsige forbrugeradfærden inden for husholdningssektoren. Desuden anvendes metoden i projekter fra *Gill et al.* [38], der gennemgås senere i dette afsnit, og *Barr et al.* [4] der gennemgås i afsnit 3.1.3. Yderligere information om metoden fremgår af afsnit C.1.1 i appendiks C.

Adfærdsmodellen *Theory of Planned Behaviour* blev anvendt i en analyse udført af *Gill et al.* [38]. Formålet med analysen var, at undersøge hvor stor en indflydelse beboernes adfærd havde på boligens energiforbrug. Denne undersøgelse udført på baggrund af målinger, spørgeskemaundersøgelse og interviews. Boligerne var opført som lavenergihuse og anvendte gennemsnitligt 46 % mindre energi til rumopvarmning, og henholdsvis 39 % og 34 % mindre vand og el, i forhold til regionens daværende gennemsnit. Selvom forbruget var reduceret i boligerne, kunne der observeres store forbrugsvariationer boligerne imellem [38]. Spørgeskemaundersøgelsen blev udviklet ved hjælp af vejledningen, der bygger på adfærdsmodellen *Theory of Planned Behaviour*¹. Spørgeskemaet var udformet ved hjælp af nogle udsagn, som beboerne skulle tage stilling til ud fra deres holdning til spørgsmålene. Besvarelsenerne kunne gives ud fra en femtrinsskala med svarmuligheder fra *helt uenig* til *helt enig*, hvor der blev tildelt point fra -2 til +2 for hver besvarelse. På baggrund af de tildelte point blev der beregnet tre scores (*total behaviour scores*), hvor disse blev sammenholdt med henholdsvis varmemeforbruget, den samlede mængde anvendt brugsvand og elforbruget i de enkelte boliger.

Spørgeskemaundersøgelsen blev besvaret af 18 beboere fra 15 husstande. Idet der var stor forskel på de tre tidligere omtalte forbrug boligerne imellem, blev det undersøgt, hvilke forhold der havde betydning for varmemeforbruget, den samlede vandmængde og elforbruget. Disse forhold fremgår af tabel 3.1.

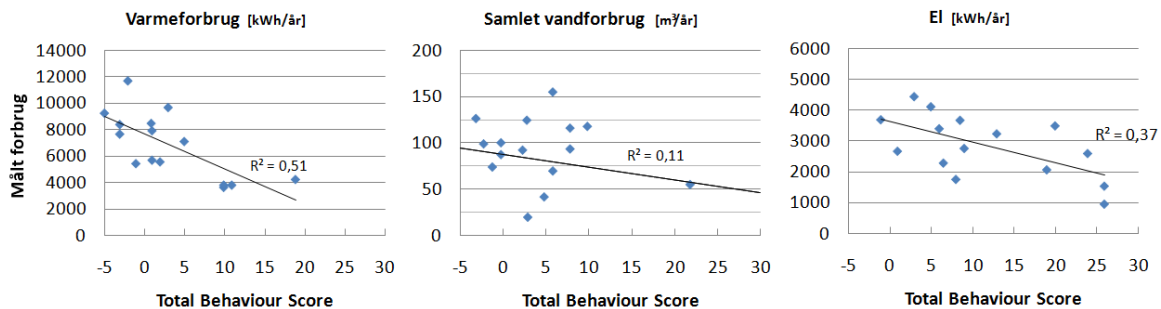
Forbrugsområde	Forhold med betydning for forbrug
Rumopvarmning	Ønsket lufttemperatur samt information om styring af varme- og ventilationsanlæg.
Brugsvand	Antallet af brusebade eller karbade per uge samt vaner i forbindelse med tøjvask og rengøring.
El	Brugstid for belysning samt antallet af elektriske apparater og brugstiden af disse.

Tabel 3.1: Vigtige parametre inden for forbrugsområderne varmemeforbrug, den samlede vandmængde og elforbruget [38].

Beboernes *total behaviour scores* blev sammenholdt med målingerne fra hvert af de tre forbrugsområder. I figur 3.1 fremgår det, at beboernes *total behaviour scores* beskriver henholds-

¹Det anvendte spørgeskema fremgår af artiklen fra *Gill et al.* [38, s. 508]. Den anvendte vejledning er udformet af *Francis et al.* [36].

vis 51 %, 11 % og 37 % af variationen for varmemeforbruget, den samlede mængde anvendt brugsvand og elforbruget.



Figur 3.1: Beboernes indflydelse på varmemeforbruget, det samlede vandforbrug og elforbruget [38].

Beboerne blev under interviewet bedt om, på et søjlediagram hvor alle 15 husstandes forbrug blev illustreret, at udpege den søjle de mente repræsenterede deres forbrugsstørrelse for varmemeforbruget, det samlede vandforbrug og elforbruget bedst. Ved sammenligning af beboernes estimat på eget forbrug og det målte forbrug i boligen, fremkommer korrelationskoefficienter på 0,69, 0,39 og 0,29 for henholdsvis varmemeforbruget, det samlede vandforbrug og elforbruget.

Det skal bemærkes, at den foregående analyse blev udført på baggrund af 18 spørgeskema-besvarelser, hvorfor den ikke bør anses som værende repræsentativ for hele Storbritannien. Det skal tilkendegives, at brugeradfærden har en relativ høj forklaringsgrad for variationen af specielt energiforbruget til rumopvarmning. Det bør derfor undersøges, hvorvidt dette forhold kan genfindes blandt et større antal husstande, da det kan give bedre indsigt i, hvor stor indflydelse brugeradfærden har på energiforbruget i boligen.

3.1.2 Faktor- og komponentanalyse

Faktoranalyse anvendes til at forenkle fortolkningen af en datamængde, der indeholder en stor mængde observationer eller variable. Denne analysemetode er anvendt af *Barr et al.* [4] i forbindelse med behandling af resultaterne fra en spørgeskemaundersøgelse. Formålet med analysen var, at undersøge hvorvidt der kunne identificeres nogle bestemte adfærdsmønstre, og derigennem klarlægge hvilke kendetegn forskellige typer af beboere havde. Analysen blev udført på baggrund af resultaterne fra en omfattende spørgeskemaundersøgelse udført i sommeren 2002 i Storbritannien. 1600 spørgeskemaer blev sendt ud til tilfældigt udvalgte husstande og 1256 besvarede skemaer blev sendt retur. I denne undersøgelse blev faktoranalysen anvendt med henblik på, at analysere hvorvidt energibesparende adfærd hos beboere kunne grupperes sammen med andre miljømæssige handlinger hos disse.

Spørgeskemaet var udformet således, at 40 forskellige spørgsmål kunne kobles sammen med et tilsvarende antal adfærd og måder at agere på hos respondenterne. Resultatet af faktoranalysen var, at de forskellige adfærd kunne grupperes i tre overordnede faktorer. Den ene faktor relaterede sig til købsbeslutninger hos respondenterne, hvilket eksempelvis drejede sig om at

købe FairTrade-mærkede produkter, købe energieffektive pærer og apparater m.fl. Den anden faktor var en gruppering af forskellige adfærd i hjemmet, hvilket eksempelvis kunne dreje sig om at reducere antallet af bade, slukke for vandet mens der børstes tænder, slukke lys i ubenyttede rum, anvende mere tøj frem for at skrue op for varmen m.fl. Den sidste faktor grupperede flere typer af adfærd inden for genbrug, som for eksempel at genbruge glas, aviser, plastikflasker, donere møbler og tøj til velgørenhed m.fl. Det vil sige, at *Barr et al.* anvendte faktoranalysen til at gruppere forskellige former for adfærd hos respondenter, for at finde nogle „grundlæggende“ adfærd og måder at agere på i besvarelsene.

Foruden den førnævnte artikel er der opnået kendskab til en undersøgelse udført af *Raaij et al.* [103], hvor formålet med anvendelsen af komponentanalyse i undersøgelsen var at forenkle den efterfølgende analyse af spørgeskemabesvarelsene. I undersøgelsen blev der analyseret spørgeskemabesvarelses fra 145 respondenter omhandlende forskellige energirelaterede adfærd og demografiske forhold. I forbindelse med de energirelaterede adfærd blev der foretaget en komponentanalyse, hvor resultatet var, at 16 forskellige adfærd kunne samles i seks komponenter. Det betyder, at *Raaij et al.* grupperede spørgsmålene efter nogle „grundlæggende“ adfærd omhandlende eksempelvis temperatur i hjemmet ved tilstedeværelse og fravær, brug af gardiner, udluftning m.fl. For yderligere beskrivelse af analysens formål og resultater, se afsnit 3.2.2.2.

3.1.3 Klyngeanalyse

I den omtalte undersøgelse fra *Barr et al.* [4] blev der yderligere udført en klyngeanalyse på data omhandlende beboernes indkøbsrelaterede adfærd, vaner i hjemmet samt beboernes holdning til miljøet. I spørgeskemaet blev der stillet 40 spørgsmål, hvoraf de mest relevante for dette litteraturstudie er oplistet i tabel 3.2. Det bør bemærkes, at nummereringen, der er anvendt i tabel 3.2, tilsvarende anvendes på figur 3.2 - 3.5.

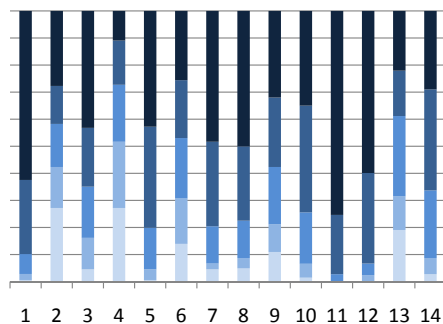
Vaner i hjemmet	Indkøbsrelateret adfærd
(1) Sluk for vandhanen når der vaskes op.	(13) Køber I Lavenergipærer?
(2) Sluk for bruseren når du sæber dig ind.	(14) Køber I energieffektive apparater?
(3) Sluk for vandhanen når der børstes tænder.	
(4) Reducer antallet af brusebade/bad i badekar.	
(5) Tag et brusebad frem for bad i badekar.	
(6) Reducer antallet af toiletskyl.	
(7) Reducer temperaturen på det varme brugsvand.	
(8) Reducer temperaturen i ubenyttede rum.	
(9) Hold temperaturen nede for at spare på energien.	
(10) Tag mere tøj på inden du skruer op for varmen.	
(11) Sluk for lyset i ubenyttede rum.	
(12) Vent med at vaske tøj før der er en fyldt maskine.	

Tabel 3.2: Relevante parametre undersøgt af *Barr et al.* inden for emnerne *vaner i hjemmet* og *indkøbsrelateret adfærd* [4].

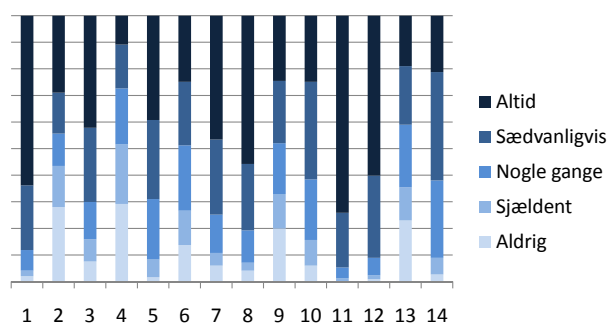
Resultatet af klyngeanalysen blev en beskrivelse af fire forskellige adfærdstyper: *den engagerede type*, *den generelle type*, *den periodiske type* og *den ikke-energibesparende type*². Forskellene mellem typerne er illustreret på figur 3.2 - 3.5. De fire adfærdstyper blev beskrevet ud fra forskellige forhold, som for eksempel husstandsstørrelse, boligtype, boligens ejerforhold, og typiske demografiske kendetegn som eksempelvis indkomst, uddannelsesniveau m.fl.

Den engagerede type blev kendetegnet ved, at der var en høj sandsynlighed for at de købte lavenergipærer og energieffektive elapparater. Samtidig var deres adfærd således, at de ikke anvendte energi til rumopvarmning før det var nødvendigt, de slukkede lys i ubenyttede rum og vaskede ikke tøj før der var til en fuld maskine, hvilket blev afspejlet i elforbruget. *Den generelle type* blev kendetegnet ved, at de havde den samme adfærd med hensyn til rumopvarmning og elforbruget som *den engagerede type*. De adskilte sig ved, at de ikke havde samme tendens til at købe energibesparende apparater og lavenergipærer. Ved *den periodiske type* skete der en stor reduktion af beboernes energibesparende adfærd og tendens til indkøb af energieffektive apparater samt lavenergipærer. *Den ikke-energibesparende type* blev kendetegnet ved, at de ikke købte energibesparende apparater og lavenergipærer. Samtidig var deres energibesparende adfærd kraftigt reduceret i forhold til *den periodiske type*, og de udviste kun en energibesparende adfærd, såfremt den ikke gik ud over deres komfort.

En dybdegående beskrivelse af de fire typer fremgår af appendiks C.2. Det er vigtigt at understrege, at der i forbindelse med denne undersøgelse ikke blev udført målinger på boligernes energiforbrug. Det betyder, at beboerne i denne undersøgelse udelukkende inddeles efter deres besvarelse af spørgsmålene i spørgeskemaet. Der er dermed ingen sikkerhed for, at de angivne adfærdsmønstre reelt set afspejles i husstandenes energiforbrug, samt at beboerne ikke agerer i overensstemmelse med de svar, de har angivet. Denne tendens, til at svare anderledes end hvad ens vaner reelt set er, er observeret i en undersøgelse fra *Kirsten Gram-Hanssen* [40], hvor den adfærd beboerne tilkendegiver ikke er genkendelig i forbrugets størrelse.

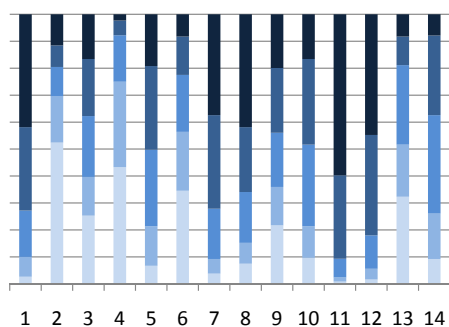


Figur 3.2: *Den engagerede types* handling i forhold til vaner i hjemmet og indkøbsrelateret adfærd [4].

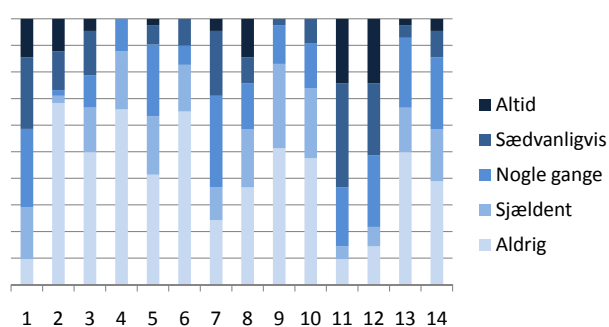


Figur 3.3: *Den generelle type* handling i forhold til vaner i hjemmet og indkøbsrelateret adfærd [4].

²Antallet af personer i hver type er: den engagerede type (294 personer), den generelle type (412 personer), den periodiske type (505 personer) og den ikke-energibesparende type (43 personer).



Figur 3.4: *Den periodiske type* handling i forhold til vaner i hjemmet og indkøbsrelateret adfærd [4].



Figur 3.5: *Den ikke-energibesparende type* handling i forhold til vaner i hjemmet og indkøbsrelateret adfærd [4].

Foruden ovenstående undersøgelse er der opnået kendskab til en artikel fra *Raaij et al.* [103], der behandles mere dybdegående i afsnit 3.2.2.2. *Raaij et al.* har forsøgt at anvende en klyngeanalyse på besvarelser fra en spørgeskemaundersøgelse. Dette skete med henblik på at identificere forskellige adfærdsmønstre blandt respondenterne, i forhold til temperaturniveau og behov for udluftning i boligen. Uden en nærmere uddybning angav *Raaij et al.*, at klyngeanalysen i dette tilfælde ikke gav nogle meningsfulde resultater, hvor der kunne identificeres nogle klare adfærdsmæssige forskelle blandt respondenterne. Artiklen bevidner, at klyngeanalyse ikke i alle tilfælde er anvendelig, hvorfor resultaterne af sådanne undersøgelser må analyseres nøje, hvis der skal identificeres meningsfulde resultater.

3.1.4 Beslutningstræer

Beslutningstræer anvendes overordnet til at søge efter mønstre og sammenhænge i større datamængder for herefter at visualisere sammenhængen grafisk. Et eksempel på hvorledes et beslutningstræ kan anvendes fremgår af en fra *Zhun et al.* [108]. Målet i denne artikel var at inddele beboerne i to prædefinerede grupper, som tog udgangspunkt i beboernes samlede energiforbrug. Inddelingen i de prædefinerede grupper skete på baggrund af viden om udeklima, boligtype, boligareal, husstandsstørrelse, konstruktionernes U-værdi, areal af utætheder, viden omkring boligens varme- og varmtvandssystem samt om køkkenet var tilsluttet gas eller elektricitet. Beslutningstræet blev opbygget på baggrund af en datamængde, med målinger fra 55 japanske boliger, og blev efterfølgende testet på målinger fra yderligere 12 boliger. Det endelige beslutningstræ, der blev udviklet, fremgår af figur 3.6.

Definitioner:

TEMP: Udetemperatur [°C]

U-værdi: Konstruktionernes U-værdi [W/m²K]

UTH: Utætheder i konstruktionerne [cm²/m²]

OPV: System til opvarmning af varmt brugsvand

KØK: Energikilde i køkken

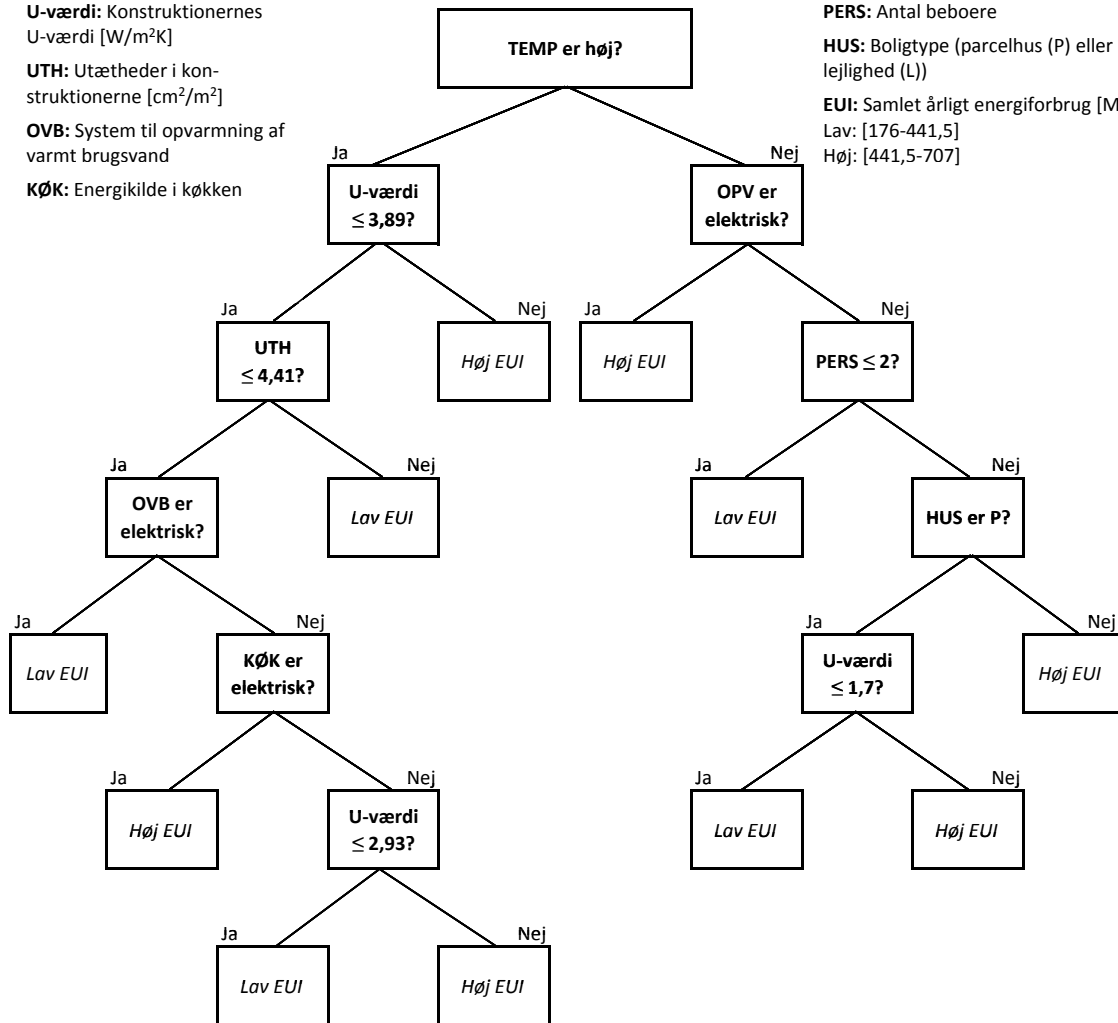
Definitioner:

OPV: Opvarmningssystem i boligen

PERS: Antal beboere

HUS: Boligtype (parcelhus (P) eller lejlighed (L))

EUI: Samlet årligt energiforbrug [MJ/m²]
Lav: [176-441,5]
Høj: [441,5-707]



Figur 3.6: Beslutningstræ til vurdering af energiforbruget til rumopvarmning, vand- og elforbrug [108].

På trods af at den anvendte datamængde havde en beskeden størrelse, gav modellen gode resultater, idet 93 % og 92 % af boligerne blev klassificeret korrekt, for henholdsvis det træningsdata modellen blev udviklet på baggrund af, og testdataet som modellens nøjagtighed blev vurderet på. Resultaterne fra denne undersøgelse kunne se anderledes ud, såfremt den anvendte datamængde var større. Det vil øge risikoen for fejlklassificering af boligerne i grupperne, og yderligere kan det udviklede beslutningstræ blive større og være mere uoverskueligt. Beslutningstræets egenskaber i artiklen fra *Zhun et al.* viser, på trods af den beskudne datamængde, gode indikationer for at metoden er anvendelig inden for estimering af energiforbrug.

Foruden ovenstående eksempel blev metoden med beslutningstræer anvendt i et projekt af *Tso et al.* [102]. I analysen blev der undersøgt tre forskellige metoder til analysering af data. Dette havde til formål at klarlægge anvendeligheden af disse, samt at undersøge hvilke

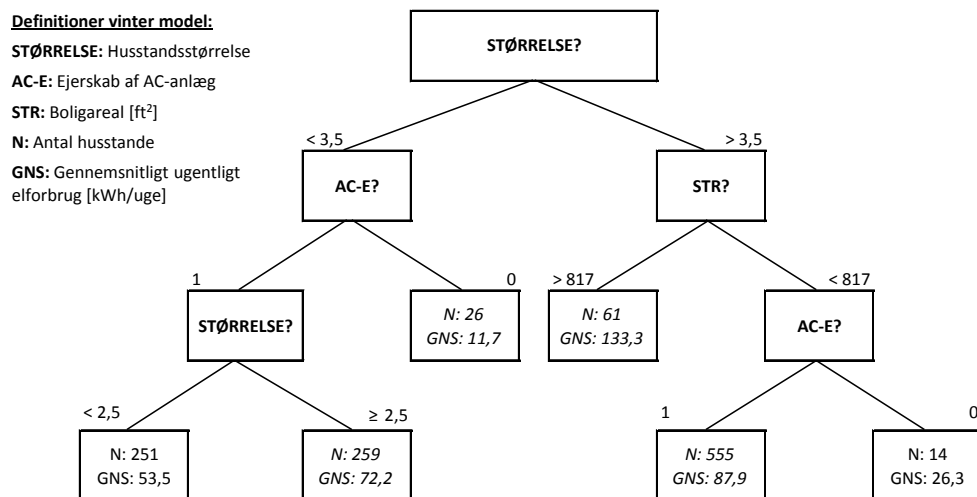
3.1.4. Beslutningstræer

metoder, der gav de bedste resultater. Metoderne, der blev undersøgt, var regressionsanalyse, beslutningstræ og neurale netværk. Alle tre metoder anvendte information fra 1166 og 1000 spørgeskemaundersøgelser udført henholdsvis om sommeren og om vinteren i Hong Kong. I spørgeskemaundersøgelserne blev der indsamlet information om boligtype, karakteristika for boligen, antallet af elektriske apparater samt brugstid af disse m.fl. Målet for inddelingen var at finde frem til hvilke parametre, der havde indflydelse på elforbrugets størrelse, og derigennem indirekte estimere elforbruget størrelse i boligerne opgjort på ugebasis.

Resultatet af undersøgelsen blev to beslutningstræer; ét udviklet for elforbruget på baggrund af besvarelserne fra sommerundersøgelsen og ét fra vinterundersøgelsen. De vigtigste variable fra de to spørgeskemaundersøgelser, fremgår af tabel 3.3, og det endelige beslutningstræ, udviklet på baggrund af vinterundersøgelsen, er illustreret på figur 3.7.

Beslutningstræ - sommer	Beslutningstræ - vinter
Boligtype	Boligens størrelse
Boligens størrelse	Husstandsstørrelse
Husstandens indkomst	Ejerskab af aircondition
Ejerskab af elektrisk vandvarmer	
Ejerskab af emhætte	

Tabel 3.3: De vigtigste parametre i beslutningstræet for henholdsvis sommer- og vinterundersøgelsen [102].



Figur 3.7: Det endelige beslutningstræ for elforbruget i vinterperioden [102].

I analysen sammenlignede *Tso et al.* de tre tidligere omtalte analysemetoder med udgangspunkt i data fra vinterperioden. Gennem denne analyse fandt *Tso et al.* frem til, at det udviklede beslutningstræ var den metode, der estimerede energiforbruget bedst. Det betyder, at metoden med beslutningstræer var anvendelig til at beskrive sammenhængen mellem elforbruget og de introducerede parametre i den omtalte analyse.

3.1.5 Diskussion

Sammenhængen mellem energiforbruget i boligen, beboernes adfærd og andre relevante parametre er i de foregående afsnit beskrevet ved hjælp af flere forskellige metoder. De fire analysemetoder, der er introduceret i de foregående afsnit viser, at alle metoder er anvendelige til at finde og beskrive en sammenhæng mellem beboernes adfærd og energiforbruget i boligen. Nogle af modellerne anvender sociologiske og demografiske variable andre anvender tekniske variable til at beskrive sammenhængen.

I analysen udført af *Gill et al.* [38] og indirekte i analysen fra *Barr et al.* [4] anvendes en adfærdsmodel. I undersøgelsen fra *Gill et al.* inddrages der sociologiske og demografiske forhold samt målinger, hvorimod der i analysen fra *Barr et al.* kun inddrages sociologiske og demografiske forhold. Det er indledningsvis i kapitel 3 fremhævet, at det er vigtigt, at inddrage flere forskellige forhold, når energiforbruget skal vurderes. Det gælder både sociologiske parametre, tekniske aspekter og demografiske forhold. Adfærdsmodellen kan anses som et hjælpemiddel til at udvikle og bedømme et spørgeskema, for herefter at anvende de fremkomne resultater i udviklingen af en model der beskriver sammenhængen mellem energiforbruget i boligen og beboernes adfærd.

I undersøgelserne fra *Raaij et al.* [103] og *Barr et al.* benyttes der henholdsvis faktor- og klyngeanalyse og komponent- og klyngeanalyse. For begge undersøgelser er der forskel på, hvorledes metoderne er anvendt. Gennemgående for begge undersøgelser er, at faktor- og komponentanalyserne som hovedpunkt anvendes til at reducere/sammenlægge antallet af variable i en undersøgelse, for derigennem at skabe et bedre overblik over hvilke variable der overordnet set kan beskrive de samme forhold. Klyngeanalyserne er ligeledes anvendt i begge undersøgelser. I analysen fra *Raaij et al.* fremkom der ingen brugbare resultater, hvorimod metoden gav brugbare resultater, i form af en beskrivelse af fire forskellige brugertyper i undersøgelsen af *Barr et al.* Faktor- og komponentanalysen anses derfor som anvendelig i den indledende behandling af indsamlet data, hvor den samlede mængde variable kan reduceres eller beskrives ved hjælp af et overordnet emne. Klyngeanalysen gav gode resultater for den ene af analyserne, hvor resultaterne af disse er fremstillet grafisk, og forskellene imellem de fremkomne brugertyper er tydeliggjort igennem en beskrivelse.

Metoden om beslutningstræer er anvendt i to forskellige analyser af henholdsvis *Tso et al.* [102] og *Zhun et al.* [108]. I analysen fra *Zhun et al.* er målet, at klassificere beboere til enten at tilhøre et lavt/højt forbrug ud fra viden om forskellige variable. I analysen fra *Tso et al.* var hovedformålet at undersøge hvilke variable, der har betydning for energiforbruget i boliger ud fra viden om forskellige variable og det ugentlige energiforbrug. Det fremgår af den foregående beskrivelse, at anvendelsen af metoden er meget forskellig i de to analyser. Fælles for begge analyser er, at resultaterne er fremstillet grafisk og er dermed lette at overskue og forklare.

Som det fremgår af nærværende afsnit, kan alle de beskrevne metoder anvendes til at beskrive en sammenhæng mellem energiforbruget i boliger og beboernes adfærd. Der er stor forskel på hvordan metoderne analyserer data, og flere af metoderne kan kombineres. Det kan eksempelvis ske ved at reducere antallet af variable ved hjælp af en faktor-/klyngeanalyse, for herefter at anvende data til udvikling af et beslutningstræ. Fire af de fem nævnte metoder beskrives nærmere i kapitel 4, hvor den grundlæggende teori for disse introduceres. På baggrund af den praktiske og teoretiske indsigt i metoderne udvælges der i nærværende rapport

én metode, der vil blive anvendt til udvikling af en model til estimering af energiforbrug i boliger.

3.2 Rumopvarmning

Energiforbruget til rumopvarmning står, sammen med opvarmningen af varmt brugsvand, for 81 % af det samlede energiforbrug i boliger, jf. figur 2.2 på side 15. Foruden de tekniske forhold for boligen er der nogle adfærdsrelaterede forhold, der har vist sig at have stor indflydelse på energiforbruget til rumopvarmning. Ifølge afsnit 2.2.1 er de væsentligste forhold, de aktuelle setpunkttemperaturer for varmeanlægget, udluftning, infiltration og brugen af solafskærmning. For at undersøge disse forhold nærmere er relevante emner i den gældende lovgivning, med hensyn til krav til operative temperaturer, indledningsvis studeret. Derefter er der udarbejdet en sammenfatning af hvilke setpunkttemperaturer, der typisk anvendes i boliger i forskellige europæiske lande. Anvendelsen af ventilation i boliger er undersøgt gennem lovgivning og forskellige artikler på området, hvor beboernes adfærd er i højsæde. Samtidig er boligens infiltration og beboernes adfærd i forbindelse med udluftning ligeledes undersøgt.

Flere af de oplistede forhold er enten ikke behandlet, eller kun behandlet i et begrænset omfang, i undersøgelser i boliger. Flere af disse forhold er undersøgt og beskrevet i forbindelse med kontorbyggerier, og resultaterne fra analyserne fremgår af afsnit C.4, C.4.1 og C.4.3 i appendiks C.

3.2.1 Lufttemperaturer i boliger

I *Bygningsreglementet* [68] stilles der krav til, at der i rum hvor der opholder sig personer i længere tid, kan opretholdes et sundheds- og sikkerhedsmæssigt tilfredsstillende indeklima samt sundhedsmæssigt tilfredsstillende temperaturer, når der tages hensyn til den menneskelige aktivitet i rummene. For funktionskrav og metoder til specifikation, verifikation og kontrol af termisk indeklima henvises der til *DS 474* [18] samt *DS/EN/ISO 7730* [89]³. Ud over kravet i bygningsreglementet og vejledningen i standarderne har beboernes komforttemperatur og holdning til forbrug også indflydelse på størrelsen af energiforbruget til rumopvarmning.

I det følgende afsnit introduceres det væsentligste indhold i de to normer, hvad angår krav til den operative temperatur, med henblik på at anvende denne viden i den fremtidige udvikling af en model. Yderligere beskrivelse af krav til indeklima fra normerne fremgår af afsnit C.3 - C.3.3 i appendiks C.

3.2.1.1 DS 474

Denne norm indeholder metoder, der skal hjælpe med at specificere, projekttere og kontrollere det termiske indeklima i bygninger, hvor personer opholder sig. Ved specifikation af krav til

³I den forbindelse bør det bemærkes, at kravene fra *DS 474* er gældende for de fleste rum i boliger. Derimod er *DS/EN/ISO 7730* gældende til vurdering af indeklima i kontorer, men den er ligeledes anvendelig for andre typer af rum og bygninger.

indeklimaet skal der tages højde for brugernes aktivitetsniveau og beklædning. I boliger kan beboernes aktivitetsniveau variere fra sovende tilstand til hård fysisk aktivitet, som eksempelvis rengøring. Samtidig er beboerne, inden for hjemmets fire vægge, yderst adaptive i form af, at de er i stand til at regulere deres beklædningsniveau. Kravene, til den operative temperatur i *DS 474* [18], der fremgår af tabel 3.4, er baseret på, at mindre end 10 % er utilfredse med det generelle termiske indeklima.

Generelle påvirkninger	Beklædning	Temperatur
	[met]	[°C]
Operativ temperatur i opholdszonen (vinter), t_o	1,0	$20 < t_o < 24$
Operativ temperatur i opholdszonen (sommer), t_o	0,5	$23 < t_o < 26$

Tabel 3.4: Krav til den operative temperatur fra *DS 474* under forudsætning af et aktivitetsniveau på 1,2 *met* [18].

Indeklimaet i forskellige boliger kan blive vurderet og sammenlignet ud fra de angivne grænser, men kun med hensyn til om boligens indeklima er godkendt eller falder uden for de angivne acceptable grænser. Grænserne fra denne norm sammenlignes i det efterfølgende med kravene i *DS/EN/ISO 7730*, for at klarlægge hvilken kategori de angivne grænser i *DS 474* dækker over.

3.2.1.2 DS/EN/ISO 7730

Den europæiske standard *DS/EN/ISO 7730* [89] indeholder metoder til vurdering af termisk indeklima, komfort, lokal diskomfort, og til at vurdere i hvilken grad brugerne er tilfredse med indeklimaet, gennem brugen af PMV og PPD. Tilsvarende *DS 474* lægger *DS/EN/ISO 7730* stor vægt på, at der skal differentieres mellem anvendelsen af forskellige rum, brugernes påklædning og deres aktivitetsniveau, for at opnå de bedste forhold for brugerne. De generelle påvirkninger i tabel 3.5 er sammenlignet med grænserne, der er angivet for kontorer, klasseværelser og kantiner, idet der i *DS/EN/ISO 7730* ikke er angivet grænser for boliger. De givne designkriterier i *DS/EN/ISO 7730* er gældende for de omtalte rum, men kan overføres til andre typer af rum, der anvendes på tilsvarende måde [89]. Kriterierne for de anvendte grænser er beregnet på grundlag af følgende forudsætninger:

- Aktivitetsniveau på 70 W/m^2 svarende til 1,2 *met*.
- Beklædningsniveau sommer: 0,5 *clo*.⁴
- Beklædningsniveau vinter: 1,0 *clo*.⁵

I tabel 3.5 fremgår det, at grænserne for den operative temperatur, fra *DS 474*, svarer til en kategori B i *DS/EN/ISO 7730*.

⁴Svarer til: underbukser, skjorte m. korte ærmer, lette bukser, sokker og sko.

⁵Svarer til: underbukser, undertrøje m. korte ærmer, skjorte, bukser, v-sweater, sokker og sko.

3.2.1. Lufttemperaturer i boliger

Generelle påvirkninger	Kategori
Operativ temperatur i opholdszonen (vinter), t_o	B
Operativ temperatur i opholdszonen (sommer), t_o	B

Tabel 3.5: Kategori for indeklimaparametrene i *DS 474* [18] når de sammenlignes med grænserne i *DS/EN/ISO 7730* [89].

Peeters et al. [75] har fremlagt et forslag til en mulig metode til beregning af lufttemperaturen i tre forskellige rumtyper i boligen⁶, på baggrund af standarden *DS/EN/ISO 7730*, rapporten fra *P.O. Fanger* [33] og flere andre artikler. Denne metode fremgår i afsnit C.4.2 i appendiks C.

3.2.1.3 Målte lufttemperaturer i boliger samt beboernes holdning til varmemeforbruget

Der er i flere undersøgelser udført målinger på lufttemperaturen for at skabe et overblik over, hvilke forhold beboerne har i deres boliger, samt hvor meget lufttemperaturerne varierer imellem disse. Undersøgelserne er udført i flere forskellige lande og deres omfang, med hensyn til måleperiodens varighed og antallet af boliger der undersøges, er meget varierende. Nogle af disse undersøgelser introduceres i dette afsnit. Efterfølgende gennemgås to undersøgelser fra henholdsvis *Kirsten Gram-Hanssen* [40, 41] og *Jesper Ole Jensen* [48], hvor hovedformålet i disse er, at klarlægge hvilke parametre der har indflydelse på varmemeforbrugets størrelse, samt at undersøge beboernes adfærd på området.

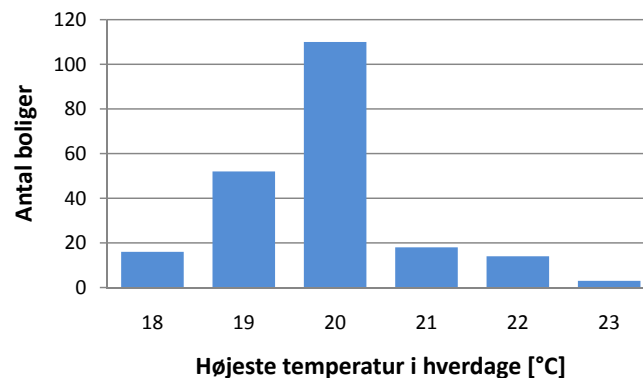
Målte lufttemperaturer

Thomsen et al. [96] fandt ved undersøgelse af boligers energiforbrug i et internationalt demonstrationsprojekt, fra 1989-1994, frem til, at lufttemperaturen i boligerne har stor betydning for det endelige varmemeforbrug. I en af de danske boliger fra undersøgelsen blev energiforbruget til rumopvarmning beregnet til $12 \text{ kWh/m}^2 \text{ per år}$ ved brug af BSim. Til sammenligning fremgik det af målinger, at det reelle forbrug var $30 \text{ kWh/m}^2 \text{ per år}$, hvilket svarede til et underestimeret forbrug på 150 %. Årsagen til afvigelsen skyldtes primært, at beboerne havde en lufttemperatur på $22\text{-}23 \text{ }^\circ\text{C}$. Dette var højere, end der blev antaget i beregningerne, hvor der blev anvendt en standard rumtemperatur på $20 \text{ }^\circ\text{C}$, hvilket almindeligvis anvendes til energiberegning af boliger i Danmark [96].

I et andet projekt fandt *Brohus et al.* [13] frem til, gennem målinger, spørgeundersøgelser og simuleringer, hvilke parametre der havde størst betydning for det samlede energiforbrug til boliger. Undersøgelsen beskriver hvorledes der i 2005 blev udført målinger og analyser af otte tilnærmelsesvis identiske rækkehuse. Der blev blandt andet målt på infiltration, setpunkttemperatur for rumopvarmning, varmebelastning og udluftning. Der blev endvidere udført en spørgeundersøgelse for at indsamle viden om brugernes alder og forskellige vaner. Målingerne i rækkehusene viste, at setpunktstemperaturen for rumopvarmningen havde en middelværdi på $22,19 \text{ }^\circ\text{C}$ med en standardafvigelse på $1,03 \text{ }^\circ\text{C}$.

⁶Rumtyperne omfatter badeværelse, soveværelse og andre rum (stue, køkken og kontorer etc.).

Lignende undersøgelser af temperaturfordelinger i boliger blev undersøgt af *Santin et al.* [44] i 2008. Analysen blev udført på baggrund af en spørgeundersøgelse i to boligområder i Holland. Beboerne i undersøgelsen var inddelt i tre grupper, ud fra hvilken type styring de havde på varmeanlægget. I nogle boliger var der manuelle radiatorventiler, i andre manuelle termostater og den tredje type var der programmerbare termostater. Af denne undersøgelse fremkom temperaturfordelingen for setpunkter i 220 boliger. Den primært anvendte setpunktstemperatur var på $19 - 20\text{ }^{\circ}\text{C}$, jf. figur 3.8. Resultaterne fra spørgeundersøgelsen viste, at omkring 75 % af husstandene foretrak $19 - 20\text{ }^{\circ}\text{C}$ som den højeste, foretrukne setpunktstemperatur for varmeanlægget. Ved en statistiske analyse af resultaterne fremkom en middelværdi for den højeste, foretrukne setpunktstemperatur er $20\text{ }^{\circ}\text{C}$, med en standardafvigelse svarende til $7\text{ }^{\circ}\text{C}$. *Santin et al.* henviser til en lignende hollandsk undersøgelse udført i 2001 af *Jeeninga et al.* [47], der fandt frem til at 73 % af de undersøgte husholdninger, hvilket svarer til omkring 130 husholdninger, havde en lufttemperatur mellem $18 - 20\text{ }^{\circ}\text{C}$.



Figur 3.8: Temperaturfordeling i ca. 220 boliger i Holland. Der er i figuren frasorteret enkelte oplyste setpunktstemperaturer, hvor disse lå imellem $14 - 17\text{ }^{\circ}\text{C}$ og enkelte ved henholdsvis 25 og $30\text{ }^{\circ}\text{C}$ [44].

Der er gennem nærværende litteraturstudie indsamlet viden om lufttemperaturen fra 59 danske boliger. Noget data er fra større projekter og andet er enkeltstående målinger. I figur 3.9 fremgår det, at lufttemperaturen i boligerne varierer mellem $19 - 24\text{ }^{\circ}\text{C}$, men med en større del af boligerne mellem $21 - 23\text{ }^{\circ}\text{C}$. Disse temperaturer stemmer godt overens med indeklimamålingerne fra *Kirsten Gram-Hanssen* [40], der gav en klar indikation af, at beboerne foretrak forskellige temperaturer i hjemmet. Temperaturerne i stuen varierede hovedsagligt mellem $18 - 23,5\text{ }^{\circ}\text{C}$ og temperaturerne i soveværelset mellem ca. $15 - 22\text{ }^{\circ}\text{C}$. Der var dog i enkelte tilfælde målt temperaturer ned til omkring $12\text{ }^{\circ}\text{C}$ og op til $23\text{ }^{\circ}\text{C}$ ⁷.

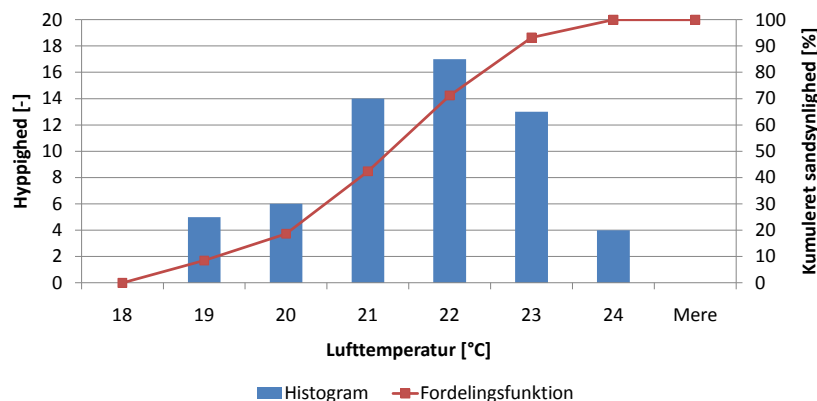
I en anden undersøgelse fra *Jesper Ole Jensen* [48] blev der ikke målt temperaturer i boligerne. Derimod gav viceværten for et boligområde udtryk for, at de fleste beboere har valgt setpunktstemperaturer på $19\text{ }^{\circ}\text{C}$ om natten og $21\text{ }^{\circ}\text{C}$ om dagen⁸. Der blev tilsvarende givet udtryk for, at der dog var nogle personer, der havde brug for højere temperaturer i dagtimerne, hvilket skyldtes stillesiddende arbejde, som det er kendetegnende for eksempelvis nogle studerende, pensionister og arbejdsløse. I forbindelse med undersøgelsen har *Jesper Ole*

⁷Information om undersøgelsen fremgår i det efterfølgende afsnit.

⁸Information om undersøgelsen fremgår i det efterfølgende afsnit.

3.2.1. Lufttemperaturer i boliger

Jensen fortaget interviews med nogle beboere. Disse interviews indikerer ligeledes, at nogle personer oplever en lav temperatur i boligen som naturligt, hvor der i stuen vurderes at være $18\text{ }^{\circ}\text{C}$ og omkring $20\text{ }^{\circ}\text{C}$ i arbejdsværelset. Ved anvendelse af oplyste lufttemperaturer fra beboerne skal der tages højde for, at beboerne ikke altid angiver den korrekte lufttemperatur for boligen. Dette fremgår bl.a. af projekterne fra *Kirsten Gram-Hanssen* [40] og *Knudsen et al.* [52], hvor der er observeret flere graders forskelle mellem oplyst og målt lufttemperatur.



Figur 3.9: Histogram over målte lufttemperaturer i 59 danske boliger. De anvendte kilder fremgår af appendiks D.

Beboernes holdning til varmekonsumet og deres indflydelse på størrelsen af forbruget

I de efterfølgende afsnit tages der udgangspunkt i to forskellige danske undersøgelser udført af henholdsvis *Kirsten Gram-Hanssen* [40] og *Jesper Ole Jensen* [48] til beskrivelse af beboernes holdninger og adfærd. Undersøgelserne introduceres ganske kort i det efterfølgende, hvorefter resultaterne fra undersøgelserne præsenteres og sammenlignes.

Jesper Ole Jensen har undersøgt hvilke årsager, der kan være til forskelle i varmekonsumet mellem husstande i samme boligområde, og blandt forskellige boligområder. Udgangspunktet for det pågældende projekt er forbrugsmålinger fra 1998 omhandlende 5631 husstande fordelt på 35 boligområder i og omkring Aarhus. Hvad angår varmekonsumet, har der kun været adgang til forbrugsmålinger i 33 boligområder.

I undersøgelsen fra *Kirsten Gram-Hanssen*, var målet at afdække sociale og tekniske forklaringer på forskelle i energiforbrug i ensartede boliger. Undersøgelsen blev udført i fire dele: teoretiske beregninger, spørgeskemaundersøgelser, interviews og indeklimamålinger. Spørgeskemaundersøgelsen blev udført i 1000 husstande med en svarrate på 50 %, og ud fra alle disse svar, og de udførte målinger, blev ti husstande udvalgt til et uddybende interview.

Som fortsættelse af afsnittet ovenfor, omhandlende lufttemperaturen i boliger, er der i undersøgelsen fra *Jesper Ole Jensen* [48] registreret markante forskelle beboerne imellem med hensyn til hvordan de regulerer varmen i boligen.

Hvad angår vaner for justering af setpunktstemperaturer i boligerne, gav projektet udtryk for forskellige vaner i denne forbindelse. Nogle personer gav udtryk for, at de kun justerede setpunktstemperaturen, såfremt det var yderst nødvendigt, hvilket underbygges af følgende udtalelse fra en beboer i en af de interviewede husstande:

„Jeg har det sådan, at når jeg først slukker for varmen, så er der bare slukket, indtil jeg simpelthen ikke kan holde det ud. Og så tænder jeg. Og så står den på 1-2 stykker, det gør den så altid, hele døgnet rundt.“ [48, s. 131]

En anden beboer gav udtryk for, at årsagen til at setpunktstemperaturen sjældent blev justeret i boligen skyldtes, at noget af vasketøjet blev tørret indendørs. I og med at samme beboer gav udtryk for, at der typisk blev vasket tøj 12 gange om ugen, blev der sjældent skruet ned for varmen, når boligen blev forladt.

I et af de omtalte boligområder i projektet blev der i 1997 – 1998 indført et nyt varmestyringssystem i boligerne. Med dette system havde beboerne mulighed for at indstille systemet med en valgt temperatur om dagen og en anden temperatur om natten. I den forbindelse udtalte en af beboerne at:

„Før vi fik systemet har vi ikke skruet ned om natten, der stod temperaturen fast om vi var hjemme eller ej. (...) Selvom man forsøger at være ansvarlig og økonomisk overfor det, er det alt for besværligt at begynde at stille på de termostater.“ [48, s. 128]

Efter varmesystemet blev indført, og det var muligt at justere setpunktstemperaturen for rumopvarmning i hele huset ved et tryk på én knap, gav beboerne udtryk for, at de altid anvendte denne funktion, når huset blev forladt. Ud over denne funktion beskrev beboerne, at de ydermere anvendte automatisk natsænkning af setpunktstemperaturen. Dette kan indikere, at det faldt beboerne mere naturligt at ændre eksempelvis setpunktstemperaturer for rumopvarmning, hvis det var teknisk nemt, hvilket *Jesper Ole Jensen* beskriver på følgende måde i rapporten:

„De tekniske systemer kan opleves som noget uigennemskueligt, der er teknisk svært at håndtere, som man derfor lader passe sig selv, hvis det fungerer.“ [48, s. 128]

Et af de forhold der har indflydelse på varmekonsumet i boliger, er beboernes adfærd for, hvorvidt de justerer temperaturen og lukker af til ubenyttede rum. Denne adfærd kan både være et spørgsmål om dagligdagsvaner og holdninger omkring hygge og komfort, hvilket en af de interviewede beboere gav udtryk for på følgende måde:

„(...) vi kan godt lide, at der er åbent i hele huset, at man har en fornemmelse af, at der er liv over det hele, at man ikke er isoleret inde i et bestemt rum.“ [48, s. 129]

Modsat er der i undersøgelsen fra *Kirsten Gram-Hanssen* [41] et eksempel på, at en familie lukker dørene til første salen i hjemmet, når der anvendes varme i stueetagen, for at spare på energien. Med hensyn til hygge og komfort udtrykte en af familierne i den pågældende undersøgelse deres holdning til varmekonsumet på følgende måde:

„Varme, det er noget vi bruger. Det går jeg ikke og tænker over.“ [40, s. 77]

og denne holdning stammer bl.a. fra deres behov for hygge:

„Du kan ikke hygge dig, hvis det er sådan, at du sidder og kravler sammen for at holde varmen.“ [40, s. 75]

3.2.1.4 Væsentlige parametre med betydning for varmemeforbruget

I et forsøg på at analysere varmemeforbruget indikerede resultaterne fra undersøgelsen udført af *Jesper Ole Jensen* [48], at boligforbruget, der er opgjort som boligarealet delt med antal personer i husstanden, har indflydelse på varmemeforbruget på en sådan måde, at et stort boligforbrug typisk medfører et stort varmemeforbrug. Denne sammenhæng er fundet i tidligere undersøgelser, der introduceres af *Jesper Ole Jensen*, hvorfor resultaterne fra denne undersøgelse underbygger de tidligere. I nogle boliger viser det sig, at der er en sammenhæng mellem et stort boligforbrug, lavt antal beboere og en høj alder for beboerne. *Jesper Ole Jensen* vurderer, at det tyder på, at ægtepar vælger at blive boende, selvom børnene er flyttet hjemmefra, hvilket betyder, at boligen måske er større, end hvad beboernes behov reelt er. Det kan endvidere tyde på, at andre værdier så som tilknytning til boligen eller området, i form af sociale kontakter og netværk, i nogle tilfælde vægter højere, end de økonomiske og miljømæssige fordele der kan være i at flytte til en mindre bolig.

Gram-Hanssen et al. [42] har ikke, som *Jesper Ole Jensen*, undersøgt sammenhængen mellem boligforbruget og varmemeforbruget. Derimod er sammenhængen mellem boligarealet og varmemeforbruget undersøgt, hvor der blev registreret en tendens til, at varmemeforbruget stiger med størrelsen af boligen.

Foruden boligarealet har resultaterne fra en undersøgelse af *Kirsten Gram-Hanssen* [40] indikeret, at boligens opførelsesår har en vis betydning for varmemeforbruget, hvor dette har tendens til at være højest i de ældste boliger. Dette forhold kan være forventeligt, eftersom blandt andet isoleringsstandarder har ændret sig gennem årene, således der er krævet bedre varmeisolerende konstruktioner.

Det fremgår ligeledes af undersøgelsen fra *Jesper Ole Jensen*, at varmemeforbruget typisk er højere i husstande, hvor beboerne har en høj indtægt sammenholdt med husstande, hvor beboerne har en lavere indtægt. I disse tilfælde bliver der imidlertid argumenteret for, at dele af forklaringen på sammenhængen mellem varmemeforbrug og indtægt kan findes i boligen, og den status, som området har på det lokale boligmarked. Med dette menes, at de aktuelle områder og boliger var attraktive for familier med børn i teenagealderen, hvilket har indflydelse på energiforbruget på den måde, at teenagere, ifølge projektet, har nogle vaner, der medfører et større energiforbrug. [48]

Foruden ovenstående konkluderer *Kirsten Gram-Hanssen* [40] yderligere, at beboernes alder har betydning for varmemeforbruget, jf. tabel 3.6. Af tabellen fremgår det, at der for beboere over 66 år er en sammenhæng med deres alder og varmemeforbruget, idet 44 % af beboerne over 66 år tilhører gruppen storforbrugere. Derudover fremgår det af den førnævnte tabel, at der i aldersgruppen 18 – 66 år er en stor variation, i hvor meget varme der anvendes, da hver af forbrugstyperne består af ca. 1/3 af alle adspurgte i den pågældende aldersgruppe⁹. Denne tendens kan underbygges yderligere af resultaterne fra *Santin et al.* [44], hvor det

⁹De anvendte grupper er defineret som *småforbrugere*: den tredjedel med det laveste varmemeforbrug, *storforbrugere*: den tredjedel med det højeste varmemeforbrug

fremgår, at det højeste antal timer hvor radiatorerne var tændt, var i husstande, hvor der boede personer over 65 år.

Ældste persons alder	Småforbruger	Normalforbruger	Storforbruger
18 – 66 år	34 %	33 %	32 %
over 66 år	20 %	36 %	44 %

Tabel 3.6: Sammenhæng mellem varmekonsum og alderen på den ældste beboer i husstanden [40].

3.2.2 Ventilation og udluftning i boliger

Danskerne opholder sig i gennemsnit 80–90 % af livet indendørs [60]. Det er derfor afgørende, at der er adgang til frisk luft i de bygninger, som de opholder sig i, så indeklimaet ikke skaber gener eller påvirker deres helbred [60]. Den friske luft kan opnås ved naturlig ventilation, ved et udsugningsanlæg hvor udeluften ledes ind gennem friskluftsventiler eller ved balanceret ventilation i form af et ventilationsanlæg installeret i boligen.

I dette afsnit introduceres den gældende lovgivning for indeklimaet fra *Bygningsreglementet* [68]. Derudover fokuseres der på beboernes adfærd, nærmere betegnet hvordan beboerne anvender de forskellige typer af ventilation i deres boliger, og der vurderes på hvilke parametre, der har indflydelse på brugen af ventilation. Anvendelsen af mekanisk ventilation i boliger er vurderet i afsnit C.4.6 i appendiks C.

3.2.2.1 Lovgivning om luftmængder og infiltration

I dette afsnit introduceres de krav der stilles i *Bygningsreglementet* [68] til luftmængderne og infiltrationen i boligen. Forhold som: atmosfærisk indeklima, træk og støj beskrives i afsnit C.3.3 appendiks C. *Bygningsreglementet* [68] foreskriver for ventilation af bygninger generelt at:

„Bygninger skal ventileres. Ventilationssystemer skal projekteres, udføres, drives og vedligeholdes, så de i benyttelsestiden mindst yder de tilsigtede ydelser“. [68, kap. 6.3.1.1 stk. 1]

hvor *ventilationssystemer* refererer både til naturlig, hybrid og mekanisk ventilation. Kravene der er stillet specifikt for boliger i kapitel 6.3.1.2 i *Bygningsreglementet* er angivet i tabel 3.7. Krav til den tekniske del af ventilationsanlæg, er givet i kapitel 8.3 i *Bygningsreglementet*. Der er angivet krav til følgende emner: temperaturvirkningsgrad for varmegenvinding både for anlæg der forsyner én bolig og flere boliger, krav til SEL-værdien, krav til at anlæggene udføres, indreguleres og afleveres som foreskrevet i *DS 447* [18] og som i *DS 428* [20] mht. brandsikkerhed og krav til vedligehold af anlæg.

3.2.2. Ventilation og udluftning i boliger

Område	Krav
Luftskifte, generelt	Luftskiftet i boliger skal som minimum være på $0,3 \text{ l/s m}^2 \text{ opvarmet etageareal}$, samt der skal være en emhætte i køkkenet over komfuret. [68, stk. 1]
Naturlig ventilation, enfamiliehuse*	Ved naturlig ventilation tilføres luften gennem ventiler i ydervæggen, vinduer eller ventiler heri og ledes bort vha. den naturlige opdrift gennem aftrækskanaler i wc/badværelse og køkken. Der stilles krav til aftrækskanalernes størrelse ved både tilførsel og fjernelse af luften. [68, stk. 2]
Mekanisk ventilation, enfamiliehuse og andre beboelsesbygninger	Systemet skal imødekomme minimumskravet fra stk. 1 og skal indeholde varmegenvinding. Desuden skal luften indblæses i beboelsesrum og fjernes fra bad/wc, køkken og bryggers. Om sommeren kan indblæsning erstattes af udelufttilførsel gennem vinduer, udeluftventiler og lignende [68, stk. 2]. Der kan anvendes behovsstyrede anlæg men minimums luftmængden må ikke blive mindre end kravet angivet i stk. 1. [68, stk. 4]
Behov for øget luftmængde	Luftskiftet i køkken, baderum, wc-rum, bryggers og lignende rum skal som minimum kunne forøges til følgende: udsugning køkken 20 l/s , udsugning baderum og wc-rum 15 l/s . I særskilt wc-rum, bryggers og kælderrum 10 l/s . [68, stk. 5]

* Huse med én bolig, herunder sommerhuse, samt dobbelthuse, rækkehuse, gruppehuse, kædehuse og lignende, hvor boligerne ikke er adskilt ved et vandret lejlighedsskel.

Tabel 3.7: Krav til ventilation i boliger fra *Bygningsreglementet* [68].

Af figur 2.17 på side 28 fremgår det, at infiltrationen har stor betydning for energiforbruget, hvorfor kravene til dette beskrives i nærværende afsnit.

Kravene til konstruktionernes tæthed fra *Bygningsreglementet* [68] fremgår af tabel 3.8. Disse krav er minimumskrav, hvilket betyder, at bygherren kan vælge at stille skærpede krav til dennes projekt [68, kap. 7.2.1]. Boligernes tæthed afprøves ved hjælp af en blowerdoor-test, når klimaskærmen er lukket. Kommunalbestyrelserne i landet skal i mindst 5 pct. af byggesagerne stille krav til at målingerne foretages. Resultaterne af testen angives som et gennemsnit, hvor der måles både ved 50 Pa over- og undertryk i boligen [68, kap. 7.2.1]. Der er i nærværende projekt undersøgt infiltrationsmængder i boliger afhængig af deres opførelsesår. Dette fremgår i afsnit C.4.4 i appendiks C.

Energiramme	Infiltration
	$[\text{l/s pr. m}^2 \text{ opvarmet etageareal}]$
Nybyggede boliger	1,5
Nybyggede lavenergi boliger	1,0
Nybyggede bygninger med høje rum*	0,5
Nybyggede lavenergi bygninger med høje rum*	0,3

* Gældende når klimaskærmens overflade divideret med etagearealet er større end 3.

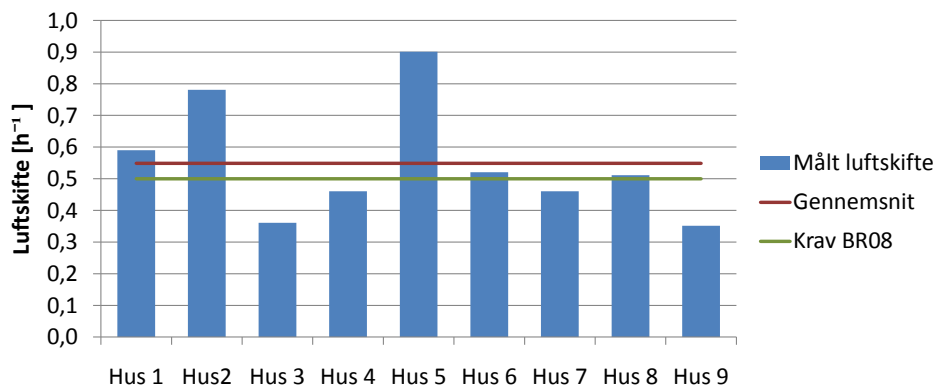
Tabel 3.8: Krav til konstruktionernes tæthed fra *Bygningsreglementet* [68].

3.2.2.2 Målt luftskifte i danske boliger samt beboernes vaner for udluftning

Adgangen til frisk luft i danske boliger skal som minimum, for nyopførte boliger, opfylde de krav der stilles i tabel 3.7 [68]. Der har siden 1982 været krav til adgangen af frisk luft i boliger, hvor kravet dengang var et minimum på $0,5 \text{ h}^{-1}$ [16]. I det efterfølgende introduceres litteratur omhandlende målte luftskifte for boliger, opført såvel før som efter 1982, hvor der beskrives om årsagerne til de forskelle, der kan observeres mellem boliger. Efterfølgende flyttes fokus til hvilken betydning beboerne har på udluftningen af boligen, samt hvilke konsekvenser dette kan have for varmemeforbrugets størrelse.

Målt luftskifte

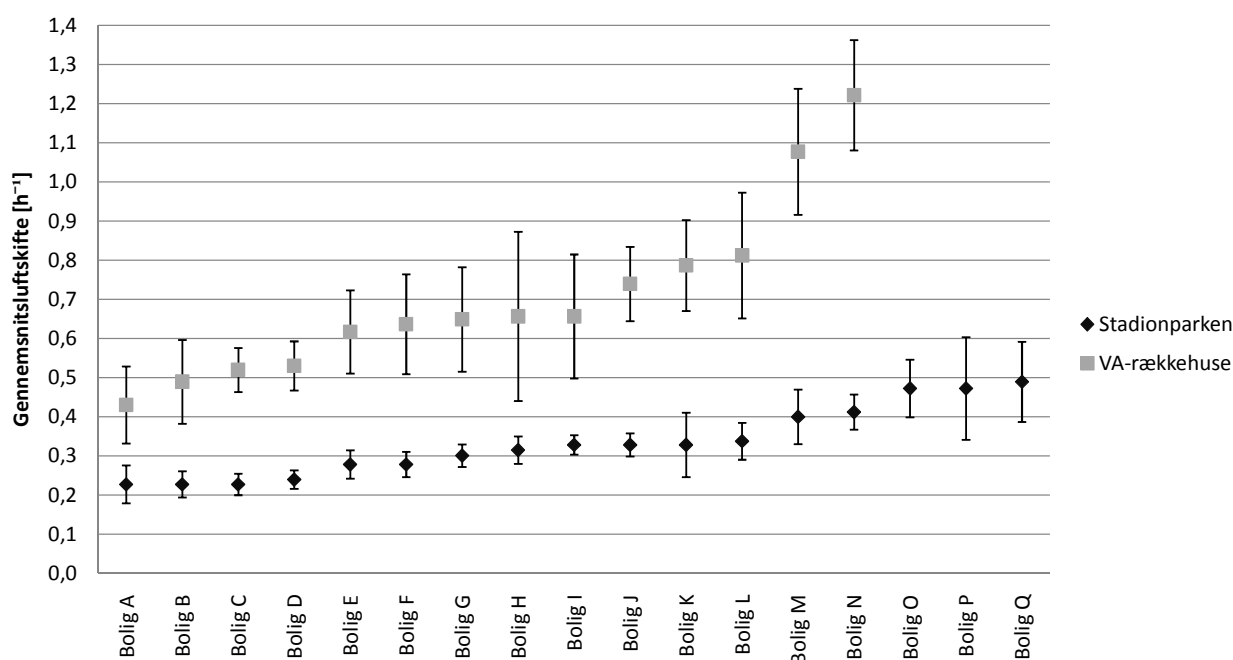
Der er i en undersøgelse udført af *Knudsen et al.* [52] foretaget målinger af luftskiftet i ni forskellige boliger, der alle er opført i perioden omkring 2006/2007. I figur 3.10 fremgår det, at luftskiftet i de ni boliger, varierer imellem $0,35 - 0,9 \text{ h}^{-1}$



Figur 3.10: Målt luftskifte i ni forskellige boliger opført i 2006/2007 i Herfølge, Danmark [52].

Som det fremgår fra figurerne 3.10 og 3.11 er der stor forskel på de målte luftskifte i boligerne. Luftskiftet for *hus 2* og *hus 5* er over kravene i *Bygningsreglementet* [68] og for *hus 3*, *4*, *7* og *9* er luftskiftet under kravene givet i *Bygningsreglementet*. Den øgede mængde i *hus 2* og *5* kan ikke umiddelbart forklares, dog er der mistanke om at anlægget i *hus 2* muligvis ikke har kørt optimalt i målingsperioden [52]. For boligerne med for lavt luftskifte kan det for *hus 3* og *hus 4* skyldes, at det er boliger med naturlig ventilation, og disse ifølge erfaring ofte er en smule under kravene i *Bygningsreglementet* [52]. For *hus 7* og *hus 9* kan det skyldes ventilationsanlæggets indstilling [52].

For boliger der blev opført i henholdsvis starten af 1960'erne og 1970'erne, er luftskiftet målt i 31 boliger i en 14 dages periode, af *Kirsten Gram-Hanssen* [40]. Disse boliger blev opført, før kravene fra 1982 blev indført, og der observeres en stor variation i de målte luftskifte. Boligerne blev udvalgt på baggrund af, at de enten havde et lavt eller højt varmemeforbrug. Af figur 3.11 fremgår de målte gennemsnit og standardafvigelse for det totale luftskifte per bolig. Det fremgår på den førnævnte figur, at luftskiftet varierer mellem $0,2 - 1,3 \text{ h}^{-1}$. Det gennemsnitlige luftskifte for Stadionparken er $0,32 \pm 0,09 \text{ h}^{-1}$ og for VA Rækkehusene er det $0,70 \pm 0,21 \text{ h}^{-1}$.



Figur 3.11: Målt luftskifte i 31 boliger fordelt på to forskellige byggerier [40].

Flere af boligerne i Stadionparken imødekommer ikke de nutidige krav, og flere af boligerne i VA Rækkehusene har meget højere luftskifte, end hvad kravene er i dag. Der fremgår ikke oplysninger om hvilken type ventilation, der anvendes i begge de omtalte boligområder. Tilsvarende haves der ikke kendskab til, om der er forskel boligerne imellem pga. renoveringsarbejde. Blandt årsagerne til de store forskelle af luftskiftet i boligerne kan bl.a. være bygningernes tilstand med hensyn til tæthed.

Beboernes vaner for af udluftning

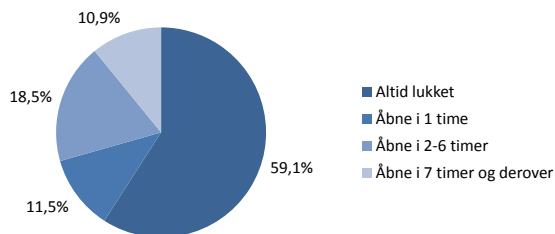
Det er på baggrund af afsnit 2.2.1 erfaret, at udluftning af boliger har stor indflydelse på energiforbruget, herunder hovedsagligt varmemeforbruget. Dette afsnit omhandler en gennemgang af beboernes anvendelse af vinduerne i boligen, og hvilke vaner de har herom. Derudover gennemgås ligeledes nogle karakteristika for beboere, der har indflydelse på hyppigheden af udluftning i boligen. Boligtypens betydning for mængden af udluftning og en gennemgang af to modeller, der beskriver sandsynligheden for brugen af vinduer, fremgår af afsnit C.4.5 i appendiks C.

Vaner i forbindelse med udluftning drejer sig i de fleste tilfælde om at få tilgang til frisk luft, forebygge allergi eller lukke tobaksrøg ud. Det er i en undersøgelse udført af *Kirsten Gram-Hanssen* [40], som omtales senere i dette afsnit, erfaret, at der var stor forskel på hvilke vaner beboerne har med hensyn til udluftning. Nogle luftede stort set ikke ud, andre foretrak en masse frisk luft grundet en stor forkærlighed til friluftslivet eller forebyggelse af allergi. Når beboerne blev adspurgt hvorfra deres vaner omkring udluftning stammer, oplyste flere af dem, at det var noget, de havde lært fra barns ben, eller noget de var blevet oplyst omkring gennem medierne. Vigtigheden af dette emne blev påvist gennem et studie udført af *Brohus et al.* [13], tidligere nævnt i afsnit 3.2 i nærværende rapport. Det fremgik af en

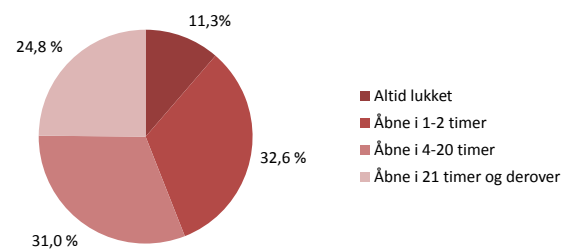
følsomhedsanalyse, at udluftning og infiltration tilsammen stod for 20 % af indflydelsen på usikkerheden i energiforbruget, jf. figur A.2 i appendiks A.

I samme den førømtalte undersøgelse blev brugen af vinduer ligeledes undersøgt. Resultatet af denne undersøgelse var en middelværdi og standardafvigelse for antallet af udluftninger per dag, hvor disse er henholdsvis 1,21 og 0,50. Varigheden af udluftningerne viste sig at have en middelværdi på 10,88 minutter per udluftning med en standardafvigelse på 8,72 minutter per udluftning.

Santin et al. [44] har ligeledes undersøgt brugen af vinduer i boliger. Undersøgelsen bygger på en spørgeskemaundersøgelse udført i 2008 for 313 boliger. Brugen af vinduer er hovedsagligt undersøgt for dagligstuen og soveværelset, hvor resultaterne er skitseret på henholdsvis figur 3.12 og 3.13. Det er værd at bemærke, at omkring 59 % af de adspurgte aldrig lufter ud i dagligstuen ved brug af vinduer. I modsætning hertil svarer næsten 25 % af de adspurgte, at de har vinduer åbne i soveværelset i 21 timer eller derover.



Figur 3.12: Anvendelse af vinduer til udluftning i dagligstue ifølge spørgeundersøgelse [44].



Figur 3.13: Anvendelse af vinduer til udluftning i soveværelse ifølge spørgeundersøgelse [44].

I en tidligere omtalt undersøgelse fra *Kirsten Gram-Hanssen* [40], blev der spurgt ind til beboernes vaner omkring udluftning. Der blev observeret en stor variationen boligerne imellem, og årsagerne var yderst forskellige. Igennem analysen viste det sig, at udluftningen af boligen havde stor betydning for beboernes varmekonsum [41]. Spørgeundersøgelserne gav indsigt i flere familiers daglige rutiner omkring deres brug af varme, samt vaner for udluftning af boligen. Fem familier blev udvalgt som typiske familier, hvor deres kendetegn for vaner og forbrug fremgår af tabel 3.9.

Familie	Kendetegn	Varmeforbrug [kWh/år]
1	Varmt, hyggeligt og nemt Lufter ofte ud, men slukker ikke for varmen imens. Termostater står altid på samme indstilling.	14600
2	Frisk luft - nyd livet Elsker frisk luft og udeliv. De har svært ved at finde sig tilrette indenfor. Lufter ofte ud, men slukker ikke for varmen imens. Har det ifølge venner koldt.	14000
3	Bekymret - men det er ikke nemt Tænker meget over at spare pga. miljøet, men kan ikke finde ud af teknikken.	10300
4	Ønsker at spare, og ved hvordan Sparer pga. økonomien. Lufter ofte ud. Skruer ned for varmen om natten og om dagen når de ikke er hjemme.	4900
5	Vaner og en smule omtanke Elsker frisk luft og kan ikke lide at have det for varmt. Regulerer varmen hver dag og slukker når de lufter ud.	4000

Tabel 3.9: Variation mellem vaner for varmekonsum og udluftning i fem familier [41].

Kendetegnende for alle familier var, at de alle luftede meget ud. Der var imidlertid stor forskel på hvordan beboerne bar sig ad med dette. Familie 1 og 2 slukkede ikke for varmen, imens der blev luftet ud. I familie 1 skyldtes det en holdning til, at der ikke blev opnået en besparelse ved at slukke for varmen, hvorimod det i familie 2 skyldtes, at de ganske simpelt ikke viste, om det kunne betale sig at slukke for varmen.

Familie 4 ønskede primært at spare på grund af økonomien, men miljøet var bestemt ikke ubetydeligt for dem. I denne familie var begge parter vokset op med fokus på besparelse, modsat den vestlige „brug og smid væk“-kultur, og de ville hellere bruge pengene på ferier end på spild af energi. Ved familie 5 skyldtes sparementaliteten og vanerne en ungdom med energikrisen i 1970'erne:

„Man havde de der kriser i 70'erne ikke, og så tænkte vi, nå, men så må vi jo også gøre det, ikke, og så er det blevet ved en“. [40, s. 76]

En af de øvrige familier der deltog i undersøgelsen havde en trang opvækst eller en hård periode i livet betydet, at de nu, hvor de havde råd, var gode ved dem selv. Konsekvensen var, at de havde et øget forbrug, som de ikke tænkte specielt meget over, og de nød den ekstra komfort. Udtrykt af en af familierne:

„Vi ved udmærket godt, at vi har en lidt højere temperatur end 21 (...) det tager vi helt cool, for vi betaler selv for det“. [40, s. 76]

Familie 1 og 5 noterede deres forbrug ned, og begge familier sammenlignede forbruget med kommunens grønne regnskaber. Det blev dog ikke gjort for optimere, og det gjorde, uanset forbrugets størrelse, heller ikke det store indtryk på de to familier. For familie 1 gjaldt, at de viste, at de brugte mere end gennemsnittet, men til det svarede manden:

„We like the way we live, and then some heat is needed. I do not want to economize on it, because what we can save in money is not very much (...), of course we do not use just for using, but we use what we think is needed to feel nice and comfortable (...) there is always cost involved in living, so we just pay the cost.“
[41, s. 180]

Tabel 3.9 og de ovenstående udtalelser viser, at der kan være stor forskel på varmekonsumet i nærmest ens boliger. Årsagerne til det meget varierende forbrug skyldes forskellige holdninger blandt beboerne overfor miljøet, forskellige vaner for udluftning af boligen, økonomiske årsager, vaner fra opvæksten og forskel i ønsket lufttemperatur.

Karakteristika for beboerne har ligeledes indflydelse på, hvordan udluftningen finder sted i boligen. Dette blev undersøgt af *Raaij et al.* [103] ved at opdele husstandene i forskellige typer, ud fra varmekonsumet sammenholdt med ønskede lufttemperaturer og ventilationsrate i boligen. Undersøgelsen blev udført ved at indsamle data gennem en selvrappporterende spørgeundersøgelse på 145 husholdninger i Vlaardingen i Holland gennem en periode fra 1976 – 1977. I undersøgelsen blev husholdningernes varmekonsum undersøgt, samtidig med at dette blev sammenholdt med 17 forskellige aktiviteter, der blev udført i hjemmet. De fire mest signifikante parametre blev udvalgt til videre undersøgelse igennem en faktor- og klyngeanalyse. De fire udvalgte variable fremgår af tabel 3.10.

Energirelateret adfærd

Temperatur i hjemmet, når der er personer til stede.

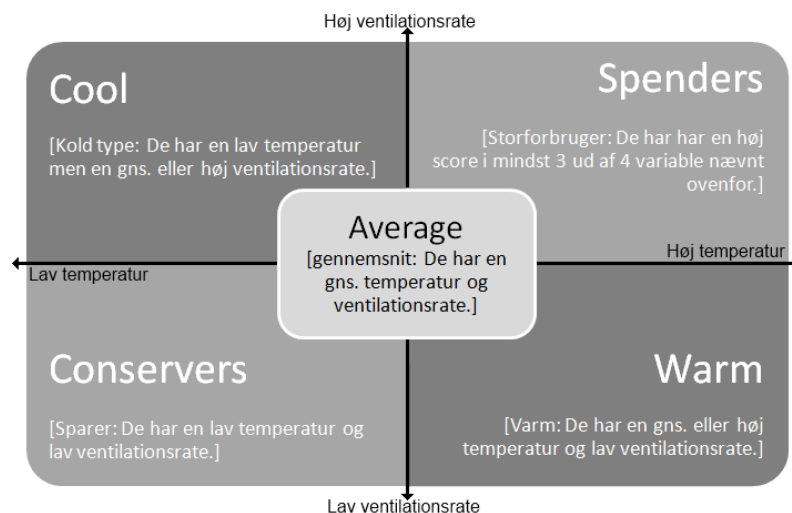
Temperatur i hjemmet, når der ikke er nogen til stede.

Mønster for udluftning i forskellige rum, brugen af vinduer, udluftventiler samt ventilationssystemer. Forhold i soveværelse om natten hvad angår brugen af radiator og vinduer til udluftning.

Anvendelse af yderdøren til udluftning.

Tabel 3.10: Relevante parametre undersøgt af *Raaij et al.* [103].

Igennem klyngeanalysen blev der defineret fem forskellige brugertyper, hvor hver type er tilknyttet en adfærd i forhold til temperatur og brugen af udluftning, jf. figur 3.14.



Figur 3.14: De fem typer beboere præsenteret af *Raaij et al.* [103].

Det blev i et hollandsk projekt fra 2007 af *Soldaat et al.* [88] undersøgt hvilken indflydelse beboerne havde på udluftning og balanceret ventilation, og derigennem også deres indflydelse på energiforbruget. Undersøgelsen blev foretaget som en kombination af interviews og simulering. Igennem et litteraturstudie fandt *Soldaat et al.* frem til, at fire faktorer spiller en stor rolle i adfærden omkring brugen af ventilation og udluftning. Disse fire faktorer er; karakteristika for beboerne, karakteristika for boligen, karakteristika for klimaet og karakteristika for ventilationssystemet, som er nærmere uddybet i appendiks C.4.5 og C.4.6. Beboernes karakteristika var kendetegnet ved, at det var husstandsstørrelsen, tilstedeværelsen af beboerne, samt deres aktivitetsniveau, der var afgørende for varigheden af udluftningen. Aktiviteter som rengøring, madlavning, vask og tørring af tøj og rygning kunne hver især bidrage til at gøre luften tung eller lugte, og beboernes opfattelse af indeluften var derfor også afgørende for hvor meget de luftede ud. Derudover drejede det sig også om beboernes komforttemperaturer, hvor disse varierede meget fra person til person. Generelt gjaldt der, at jo højere temperaturer der var ønsket af beboerne, des mindre blev der luftet ud. Dette understøtter resultaterne fra *Raaij et al.* [103], der fremgår af figur 3.14, ganske godt. Lufttemperaturerne varierede mellem de forskellige rum, hvor der var én ønsket temperatur for stuen og øvrige opholdsrum, mens der ofte forekom en lavere temperatur for soveværelset.

3.2.3 Diskussion

Indledningsvis diskuteres indholdet fra afsnit 3.2.1 omhandlende målte lufttemperaturer i boliger, samt hvilke parametre der har den største indflydelse på varmekonsumets størrelse. Efterfølgende diskuteres indholdet fra afsnit 3.2.2 omhandlende anvendelsen af ventilation og udluftning i boliger med fokus på beboernes påvirkning på disse.

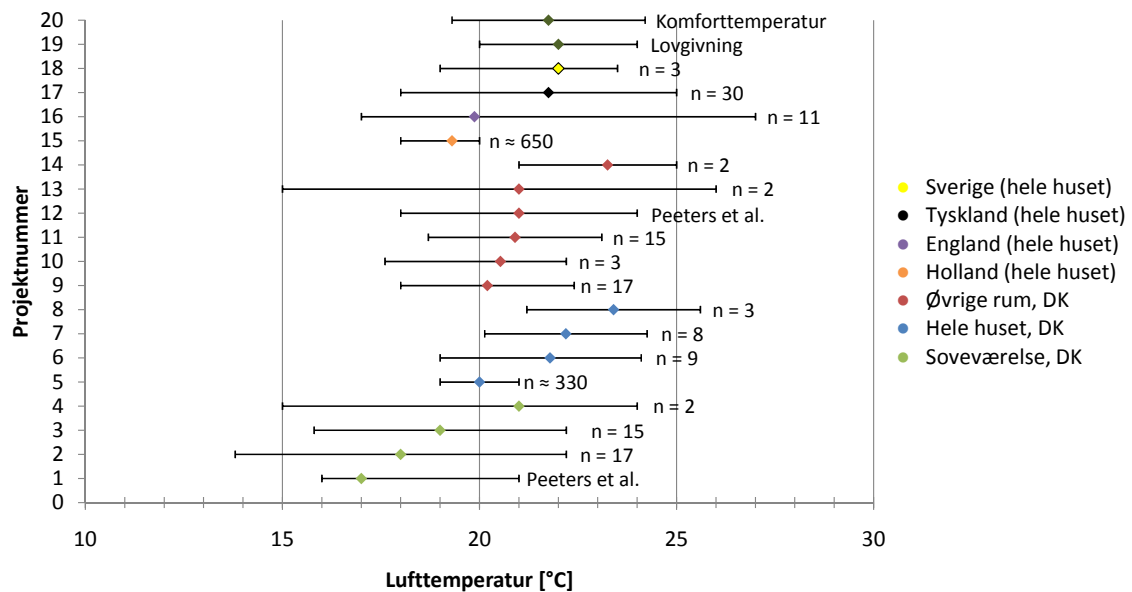
Lufttemperatur i boliger

Igennem de foregående afsnit, er der opnået et større kendskab til brugen af rumopvarmning i boliger. Dette er opnået gennem et studie af beboernes adfærd i forbindelse med valg og ændring af lufttemperaturer. Rumopvarmning anvendes typisk i de koldeste perioder af året

jf. figur C.2 på side 21 i appendiks, hvorfor udgangspunktet for den efterfølgende diskussion vil være denne periode, set i forhold til det danske udeklima.

Lovgivningen omkring operative temperaturer og anvisningen til beregning af menneskets komforttemperatur bygger på anselige mængder forsøg og erfaringer inden for hvilke forhold mennesket befinder sig i termisk komfort. Ifølge lovgivningen, beskrevet i afsnit 3.2.1.1, skal der i vintermånederne kunne opnås en operativ temperatur på mellem 20°C og 24°C i rum, hvor der opholder sig personer i længere tid. Det er meget forskelligt fra bolig til bolig, i hvilke rum beboerne ofte opholder sig, afhængig af deres arbejdsforhold og deres hverdagsmønster. Ikke desto mindre kan dette sammenholdes med menneskets komforttemperatur, der ved en almindelig indendørs vinterbeklædning på 1 *clo* typiske ligger i et interval mellem omkring 19°C til 24°C , jf. tabel C.11 på side 18 i appendiks. Det illustrerer, at de i lovgivningen angivne temperaturintervaller stemmer godt overens med de komforttemperaturer, der kan estimeres.

Det er i dette litteraturstudie valgt at gå et skridt videre og foretage et studie af hvilket forhold, der ifølge andet litteratur, videnskabelige undersøgelser og målinger, typisk gør sig gældende i boliger. Nogle af resultaterne af dette studie er skitseret på figur 3.15 for bedre at kunne danne et overblik over disse. De på figuren angivne lufttemperaturer stammer fra forskellige projekter, yderligere oplysninger om disse kan findes i afsnit 3.2.1.3 i denne rapport og i appendiks D. På figur 3.15 er der angivet en middelværdi og et interval for lufttemperaturen ud fra hvert projektnummer. Det bør dog bemærkes, at nogle intervaller er optegnet på baggrund af målte minimum og maksimum temperaturer og andre for øvre og nedre middeltemperaturer. Dette kan give et misvisende billede af intervallerne for lufttemperaturerne, når de enkelte projekter sammenholdes. Ydermere bør der sættes fokus på det antal af boliger, der er angivet for de forskellige projekter. Nogle af de angivne lufttemperaturer stammer fra målinger i enkelte boliger, andre på baggrund af målinger i et større antal boliger, mens nogle lufttemperaturer er vurderet på baggrund af udtalelser eller spørgeundersøgelser. Alle disse forhold kan give store usikkerheder på de angivne temperaturer, men ikke desto mindre vil der foretages nogle vurderinger af de aktuelle forhold i boliger.



Figur 3.15: Indetemperaturer i boliger fundet på baggrund af litteraturstudiet. Kildehenvisning til de enkelte projektnumre kan ses i appendiks D.

Figur 3.15 indikerer, at der kan være en tendens til, at der haves lavere temperaturer i soveværelset sammenholdt med øvrige rum og for boligen generelt. Hvor der i soveværelser observeres lufttemperaturer omkring $16\text{ }^{\circ}\text{C}$ til $21\text{ }^{\circ}\text{C}$, kan det forventes, at lufttemperaturer i øvrige rum ligger omkring $19\text{ }^{\circ}\text{C}$ til $24\text{ }^{\circ}\text{C}$, hvilket stemmer udmærket overens med de i lovgivningen angivne operative temperaturer og de beregnede komforttemperaturer.

I forbindelse med dimensionering af varmeanlæg i boliger og beregning af boligens energibehov anvendes der typisk en lufttemperatur på $20\text{ }^{\circ}\text{C}$. Hvis dette sammenholdes med de ovenstående betragtninger, giver det en indikation på, at de anvendte dimensioneringsforhold, med en lufttemperatur på $20\text{ }^{\circ}\text{C}$, ligger i den nedre del af intervallet for, hvilke gennemsnitlige lufttemperaturer litteraturstudiet indikerer, at der typisk anvendes i boliger. Derfor kan der sættes spørgsmålstegn ved godkendelsen af boligerne, ud fra det teoretisk beregnede energiforbrug, idet der fremgår en tendens til, at der haves højere lufttemperaturer i boligerne og energiforbruget til rumopvarmning dermed oftest underestimeres.

Gennem dette litteraturstudie er der opnået viden om hvor ofte og inden for hvilket tidsrum setpunkttemperaturen for rumopvarmningen typisk ændres i boliger. Undersøgelserne indikerer, at beboernes holdninger og vaner har indflydelse på hvor ofte de justerer setpunkttemperaturen, men det kan også være et spørgsmål om, hvor besværligt det er for beboerne at justere det. Vigtigheden af det sidstnævnte er blandt andet vist gennem et renoveringsprojekt i en boligforening, hvor varmestyringssystemerne i boligerne blev udskiftet. Denne udskiftning førte til betydelige besparelser for det samlede boligområde, og et interview med en beboer viser tilsvarende, at der før det nye varmestyringssystem blev indført, meget sjældent blev justeret på radiatortermostaterne i husstanden.

Hvad angår vaner, er nogle af disse eksemplificeret gennem en beskrivelse af fem forskellige familiers kendetegn. I den forbindelse er det beskrevet, at der er en familie, som, tilsvarende

det tidligere eksempel, ikke justerer radiatortermostaterne, til trods for der ofte luftes ud i boligen. I kontrast til dette står en anden familie, der regulerer for varmen hver dag, og slukker mens der luftes ud. Forskellighederne blandt familierne og beboerne kan både være et spørgsmål om vaner, men personernes holdninger har også en betydelig indflydelse på, hvordan de agerer. Der er vist eksempler på personer, der sparer, ofte justerer på varmen og tænker over deres forbrug på baggrund af økonomiske årsager, mens andre angiver miljøet, som årsagen på deres spareadfærd. Andre personer giver udtryk for, at de er opvokset med en bestemt adfærd, som stadig, på den ene eller anden måde, påvirker deres måde at agere på. Disse eksempler giver alle en indikation på hvor forskellige personers adfærd og holdninger er, og på hvordan det kan komme til udtryk i forbindelse med brugen af deres bolig.

Det er i flere projekter undersøgt hvilke parametre, der kan have en forklarende effekt i forbindelse med varmemeforbruget, og for hvornår der i det hele taget anvendes rumopvarmning. I et projekt er det særligt tre parametre, der indikerer at have en sammenhæng med energiforbruget til rumopvarmning. *Jesper Ole Jensen* [48] beskriver, at den ene parameter er boligforbruget, der er opgjort som boligareal per person, hvor der fremgår en tendens til, at et stigende boligforbrug medfører et stigende varmemeforbrug. I nogle tilfælde er det husstande med et stort boligforbrug, et lavt antal beboere og en høj alder for beboerne, der har et højt varmemeforbrug. *Kirsten Gram-Hanssen* [40] har ikke, på samme måde som *Jesper Ole Jensen*, sammenholdt boligforbruget med varmemeforbruget, men i denne undersøgelse blev der observeret en sammenhæng mellem boligarealet og varmemeforbruget. I begge undersøgelser blev der indikeret, at ældre beboere i nogle tilfælde havde et højere varmemeforbrug. Dette kan skyldes flere forskellige forhold, for eksempel at beboerne har en større bolig, end hvad deres behov reelt er, og at rum der ikke anvendes bliver opvarmet. Derudover kan deres holdninger, vaner og adfærd, som for alle typer af beboere, påvirke varmemeforbruget. Det stigende varmemeforbrug hos ældre beboere kan også grunde i, at beboerne opholder sig længere tid i boligen, fordi de ikke har samme tilknytning til arbejdsmarkedet som tidligere. Samtidig kan deres komforttemperatur muligvis være lidt højere, når de opholder sig i boligen på grund af et lavere aktivitetsniveau. De førnævnte forhold bliver bekræftet i et projekt af *Santin et al.* [44], hvor det er påvist, at beboere over 65 år typisk har radiatorerne stående på en høj temperaturindstilling i længere tid end yngre personer. Beboernes vaner for justering af radiatortermostaterne i forbindelse med udluftning har *Kirsten Gram-Hanssen* endvidere påvist at have en effekt på varmemeforbruget, hvor beboere, der ikke skruer ned for termostaterne, typiske har et højere varmemeforbrug.

Foruden de ovennævnte forhold har *Jesper Ole Jensen* endvidere påvist en sammenhæng mellem beboernes indtægt og varmemeforbruget, hvor tendensen er, at husstande med høje indtægter typisk har et højere varmemeforbrug, sammenholdt med husstande der har en lavere indtægt. Dette kan grunde i forhold, som er påvist i samme projekt, hvor der er en tendens til, at boligforbruget typisk er stigende i takt med husstandens indkomst. Forklaringerne til et højere varmemeforbrug kan, som beskrevet tidligere, være, at boligen reelt set er større end beboernes behov, og at en stor bolig alt andet lige vil kræve mere energi til rumopvarmning end en mindre bolig.

Foruden boligforbruget, alder, vaner for udluftning og indkomst er det påvist, at boligens opførelsesår har en vis sammenhæng med varmemeforbruget. I dette tilfælde kommer det til udtryk ved, at der typisk er et større varmemeforbrug i ældre boliger. Dette er dog også at forvente, eftersom der gennem de senere år er sket en markant forbedring af boligkonstruk-

tionernes varmeledningsevne generelt.

Ventilation og udluftning i boliger

Igennem de ovenstående afsnit, er der ligeledes opnået et større kendskab til brugen af ventilation i boliger. Dette er opnået gennem et studie af beboernes adfærd, karakteristika for beboerne samt vigtigheden af information om brugen/styring af ventilationsanlæg. Denne del af diskussionen knytter sig både til de foregående afsnit og de dele af appendiks, der fremgår af afsnit C.4.4-C.4.6.2. Der er stor forskel på typen af ventilation i de danske boliger, hvor det kan spænde lige fra naturlig ventilation, mekanisk udsugning til balanceret ventilation. Uanset hvilken type der anvendes, kan brugen af disse have stor indflydelse på det samlede energiforbrug, enten igennem varmemeforbruget eller elforbruget.

I lovgivningen er der klare retningslinjer for hvilke luftmængder, der kræves i danske boliger. Som gennemgået tidligere i dette afsnit kræves der et luftskifte på minimum $0,3 \text{ l/s per m}^2 \text{ opvarmet etageareal}$. Luftskiftet i danske boliger er meget varierende, og igennem litteraturstudiet er det erfaret, at der i nogle tilfælde forekommer et luftskifte mellem $0,2 - 1,3 \text{ h}^{-1}$ i boliger fra 1960'erne og 1970'erne, og i nyere boliger et luftskifte mellem $0,35 - 0,9 \text{ h}^{-1}$. Af dette foregår det, at luftskiftet er meget varierende, dog vil der for nyopførte boliger være krav til et minimumsluftskifte, der er foreskrevet af lovgivningen. Der er fra *Bygningsreglement* 2010 indført mulighed for at anvende behovsstyret ventilation, hvilket betyder, at der for beboernes side er mulighed for at regulere på luftmængden, så længe denne ikke reduceres under den krævede mængde.

Mængden af information er et område med stigende vigtighed, idet det ved opførelse af bedre boliger med flere tekniske installationer er vigtigt også at informere beboerne om hvorledes den optimale drift og den største energibesparelse opnås. Ved tekniske problemer med anlæggene og manglende information hos beboerne stiger frustrationen, og de reagerer enten ved vrede eller giver op og slukker teknikken. Hvis denne type boliger tildeles dette ry, er der risiko for, at lysten til at bo i denne type boliger falder, og dermed udbredelsen af disse huse. Et eksempel, der illustrerer vigtigheden af information til beboerne, er omhandlende deres vaner i forbindelse med udluftning. Selvom der i projektet fra Holland var et ventilationsanlæg med balanceret ventilation, så fortsatte beboerne med at lufte ud vha. vinduerne, på trods af at dette ikke burde være nødvendigt. Årsagen til denne handling skyldes enten at ventilationsanlægget ikke fungerede optimalt, beboerne manglede information om hvordan anlægget fungerede mest optimalt, samt at beboernes handlinger ofte er præget af vaner.

Igennem litteraturstudiet har det vist sig, at muligheden for at anvende forskellige luftmængder i boliger har været anvendt i Holland. I nogle boliger fra var der tre indstillingsmuligheder for ventilationen, hvor indstilling tre, den højeste indstilling, gav hvad der svarede til den krævede minimumsluftmængden. De tre indstillinger betød først og fremmest, at de fleste af beboerne valgte indstilling to, da denne indstilling, ifølge beboernes intuition, måtte yde den krævede mængde, og at der derudover var muligt at skrue op eller ned på luftmængden efter ønske. Det viste sig derfor allerede fra begyndelsen, at beboerne fik mindre luft ind i boligerne, end der reelt var tiltænkt blot pga. manglende information. Desuden viser dette, at det udover information også er vigtigt at udforme styringspanelerne og indstillingerne, så de er intuitive for beboerne. Konsekvensen af det manglende luftskifte blev derfor, at mange af beboerne begyndte at lufte ud, og dermed forsvandt pointen med et balanceret ventilationsanlæg. Samtidig forsvandt de tiltænkte besparelser, der kan opnås ved brug af varmegenvinding. En anden gennemgående problemstilling, der blev beskrevet tidligere ved

brugen af balanceret ventilation i boliger, er at beboerne generes af støj og træk fra anlæggene. Problemet er set i projekter fra Holland, Danmark og Sverige og beboerne reagerer ved at skærme indblæsningsarmaturerne af, skrue ned for luftmængden eller slukke anlæggene. Det er derfor vigtigt for succesraten for det nye tiltag i *Bygningsreglementet* med behovsstyring, at designet og udførelsen af anlæggene er gennemtænkte, hvor der skal anvendes nogle intuitive løsninger. Tilsvarende skal beboerne udrustes med den nødvendige information, der gør løsningerne let tilgængelige og forståelige.

Det kan både være godt og skidt at give beboerne en mulighed for at have medbestemmelse i styringen. Det indbyder, som illustreret ovenfor, til større fejlkilder, da beboernes indflydelse kan have betydning for om anlægget bruges optimalt. Der kan dog ved optimalt brug samtidig også skabes større tilfredshed for beboerne. Med dette i mente er det derfor en vigtig pointe, at beboerne skal introduceres til alt teknik i deres bolig, både gældende for varme- og ventilationsanlæggene.

Med hensyn til vedligeholdelse af anlæggene er information om teknikken ligeledes vigtig for driften. I flere af projekterne fra Holland blev filtrene i ventilationsanlæggene ikke skiftet, hvilket medførte at driften forringes, både med hensyn til ydelsen i luftmængder, men også et stigende elforbruget på grund af et højere tryktab gennem anlægget. Den faldende ydelse har yderligere indflydelse på beboernes vaner.

Gennemgående for flere af de gennemførte undersøgelser gælder det, at beboernes opfattelse af indeklimaet og deres termiske komfort spiller en stor rolle for deres vaner. Årsagerne til udluftning kan være: „tung“ varm luft, dårlig lugt, allergier eller røg, men hvornår folk reagerer på påvirkningerne er meget forskelligt. Generelt afhænger beboernes vaner af deres ønskede lufttemperatur, og hvor meget frisk luft de ønsker i boligen. Sammenspillet mellem disse to parametre blev beskrevet af *Raiij et al.* [103], der kategoriserer beboerne ud fra deres behov for frisk luft og deres ønskede niveau for lufttemperaturen i boligen, jf. figur 3.14. Ud fra de øvrige informationer fra litteraturstudiet virker betragtningen fra *Raiij et al.* som et godt værktøj til overordnet at beskrive beboernes forhold til indeklimaet i deres hjem. Andre parametre, der har betydning er beboernes alder, idet det har vist sig, at ældre mennesker og familier med små børn lufter mindre ud. Det viser sig også at allergikere og rygere, af forskellige årsager, lufter mere ud end andre. Derudover har beboernes forhold til naturen også betydning, idet der er en svag tendens til, at beboere der dyrker flere aktiviteter i det fri lufter meget ud.

De vurderinger, der er beskrevet gennem denne diskussion, bygger på oplysninger fra litteraturstudiet. Flere af vurderingerne bygger dog kun på et beskedent antal undersøgelser eller målinger, hvorfor det skal bemærkes, at der i nogle tilfælde bør indgå flere oplysninger, for at der kan dannes et mere nuanceret billede af, hvorledes de aktuelle forhold i boliger er, og for hvor forskellig beboeres adfærd i virkeligheden er.

3.3 Elforbrug

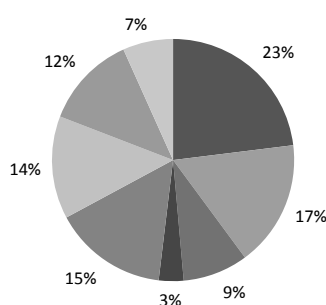
Elforbruget udgør ifølge figur 2.2 på side 15 omkring 17 % af det samlede energiforbrug i boliger. Størrelsen af dette forbrug ser, i forhold til den samlede energimængde, ikke specielt stort ud, men efter indførelse af de nye krav til boligens tekniske egenskaber og de fremtidige stramninger i *Bygningsreglementet* [68], er der risiko for, at elforbruget i fremtiden vil udgøre

3.3.1. Elforbrugets slutanvendelser

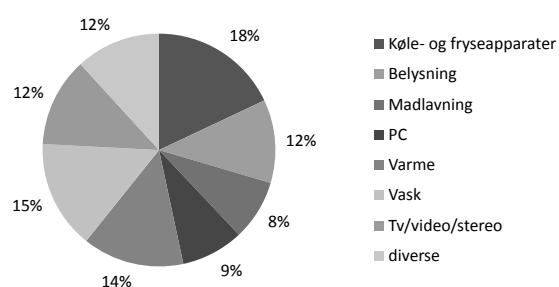
en større procentdel af det samlede energiforbrug. For at opnå indsigt i hvordan elforbruget fordeler sig på de forskellige apparater, og for at undersøge hvilke forhold, der gør, at elforbruget varierer betragteligt mellem enkelte husstande, som afsnit 2.1.2 indikerer, bliver disse områder studeret nærmere gennem de efterfølgende afsnit.

3.3.1 Elforbrugets slutanvendelser

Ifølge figur 2.5 på side 17 har elforbruget været relativt konstant siden 1992, men slutanvendelsen af elforbruget har været meget varierende, jf. figurerne 3.16 og 3.17. Af disse fremgår det, at specielt elforbruget til køle- og fryseapparater, belysning og pc'er har ændret sig markant indenfor perioden 1998 – 2008.



Figur 3.16: Elforbrugets fordeling mellem otte slutforbrug i 1999 [25]. Standbyforbruget er inkluderet i de enkelte forbrug.



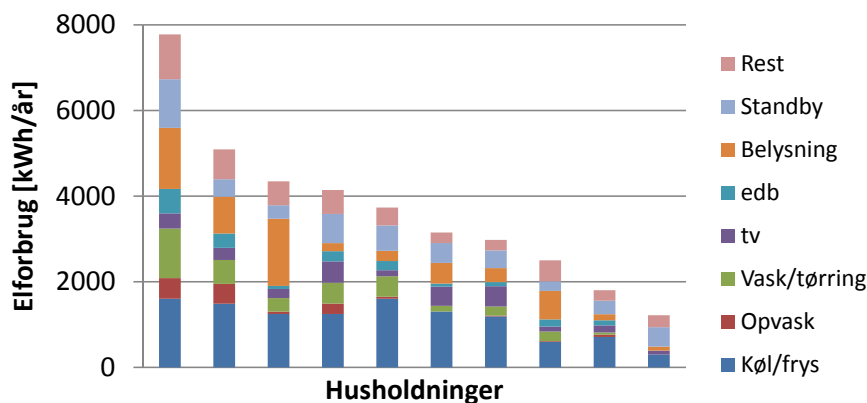
Figur 3.17: Elforbrugets fordeling mellem otte slutforbrug i 2008 [25]. Standbyforbruget er inkluderet i de enkelte forbrug.

I tabel 3.11 fremgår minimum-, maksimum- samt gennemsnitsforbruget for de mest almindelige husholdningsapparater i boliger. Forbrugene stammer fra en undersøgelse af *Casper Kofod* [53]. En mere udførlig beskrivelse af undersøgelsen fra *Casper Kofod* fremgår i appendiks C.5.

Apparattype	Minimum [kWh/år]	Maksimum [kWh/år]	Typisk forbrug [kWh/år]	Procentvis fordeling [%]
Køle/frys	250	2000	500 – 1000	25, 1
Opvask	50	700		3, 9
Tøjvask	23	739		10, 6
Tørring af tøj	20	1200		
Tv/video	4	844	211	6, 3
EDB	0	800	162	3, 8
Belysning	48	1670	452	12, 9
Standby	0	1300	308	8, 8

Tabel 3.11: Minimum- og maksimumforbrug, det typiske forbrug i parcelhuse og den procentmæssige fordeling opgjort i 100 boliger i 1999/2000 [53].

Ud over den store variation for elforbruget imellem forskellige husstande er der stor forskel på hvad elforbruget i de forskellige boliger anvendes til. Dette fremgår af figur 3.18 der præsenterer variationen på det samlede elforbrug og fordelingen af elforbruget apparaterne imellem i 10 ud af 100 boliger i Odense. De 100 boliger, der blev undersøgt, var 85 parcelhuse og 15 lejligheder, hvor der blev foretaget målinger i én måned i hver bolig, ti boliger af gangen med en samlet måleperiode på ti måneder. På grund af målemetoden kan der dog forekomme nogle usikkerheder i resultaterne. For det første er januar og februar måned ikke medtaget i måleperioden, hvilket betyder, at forbruget til belysning kan blive undervurderet, idet januar og februar er præget af lange perioder uden dagslys. Desuden er antallet af målinger i undersøgelsen umiddelbart for beskedent til at kunne anvende resultaterne, som et generelt billede af elforbruget i danske boliger. Det viser sig dog, at resultaterne for det samlede elforbrug, for både parcelhuse og lejligheder i undersøgelsen, er tæt på landsgennemsnittet, og ved en sammenligning af bestanden af elapparater med ELMODEL-bolig ses samme tendens. Derfor betragtes resultaterne fra denne undersøgelse som gældende for et repræsentativt udsnit af den danske befolkning. [53]

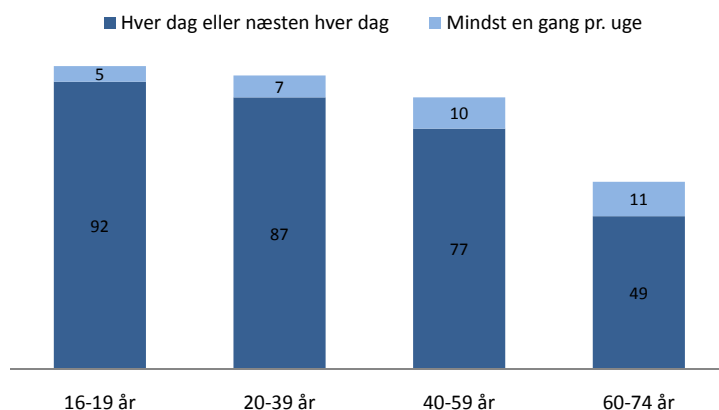


Figur 3.18: Elforbruget fordelt på otte slutanvendelser i 10 udvalgte boliger [53].

Undersøgelsen af *Casper Kofod* blev udført i 1999/2000 hvorfor der i forhold til i dag burde være en anderledes indbyrdes fordeling imellem slutanvendelserne, jf. figur 3.16 og 3.17. Specielt inden for forbruget til computere er forbruget støt stigende. I 2008 havde 82 % af den danske befolkning adgang til internettet i hjemmet, i 2009 var dette steget til 83 % og i 2010 var det oppe på 86 % [92]. Brugen af computere er dog, ligesom for andre apparater, meget forskellig fra familie til familie. Nogle anvender adgangen til internettet til betaling af regninger og enkelte mails, hvor andre anvender den til betaling af regninger, mails, daglig surfing og som underholdning i form af spil.

Det viser sig, at specielt alderen har betydning for udbredelsen og anvendelsen af EDB-udstyr. I 2009 havde ni ud af ti under 60 år adgang til en computer, hvorimod mindre end syv ud af ti over 60 år havde adgang til en [90]. Rapporten „*Befolkningens brug af internet*“ fra 2009 [90] viser, at 75 % af befolkningen anvender internettet hver dag eller næsten hver dag, og yderligere 9 % anvender det mindst én gang om ugen. Figur 3.19 illustrerer sammenhængen mellem regelmæssig brug af computeren og befolkningens alder, hvor det tydeligt fremgår, at det er standard for den yngre generation at have adgang til computere, og indenfor nogle årtier kan det forventes, at anvendelsen vil stige yderligere blandt de ældre generationer.

Forbruget til edb varierer mellem 0 – 800 kWh/år for de 100 husstande i Odense, med et gennemsnitligt forbrug på 162 kWh/år. Den afgørende faktor for størrelsen af forbruget er, ligesom for tv/video, varigheden af forbruget og typen af apparat.



Figur 3.19: Brugen af computere i den danske befolkning anno 2009 i forhold til beboerens alder [90].

Som det fremgår af tabel 3.11, udgør forbruget til belysning i boliger 12,9 % af det samlede elforbrug. Ligesom for mange øvrige elforbrug varierer elforbruget til belysning også betragteligt fra husstand til husstand, som det er illustreret i figur 3.18. Generelt fremkom det af undersøgelsen af *Casper Kofod*, at hvis der i husstanden er en tendens til et højt elforbrug, er forbruget til belysning ofte også højt, hvilket bekræftes i figur 3.18. For yderligere studie af brugen af kunstig belysning i boliger, se appendiks C.5.1.

Standbyforbruget i boliger udgjorde i gennemsnit 308 kWh/år i undersøgelsen fra *Casper Kofod*, hvilket svarer til 8,8 % af det samlede elforbrug. I undersøgelsen fra *Dansk Energi* [25] for år 2008, der er illustreret i figur 3.17, udgør standbyforbruget 12 %. Standbyforbruget er gennemsnitligt fordelt på apparaterne, som det er vist i tabel 3.12. Det fremgår af

tabel 3.11, at standbyforbruget udgør stort set samme andel af det samlede elforbrug, som andelen til vask/tørring, eller som den samlede andel af underholdningsapparaterne i form af tv/video/HIFI samt edb. Standbyforbruget er, modsat de øvrige elforbrug et forbrug, der intet eller ingen gavner.

Anvendelse	Apparater	Elforbrugsandel [%]
Madlavning	Display for mikroovn, ovn, kaffemaskine	5
Teleudstyr	Trådløs tlf. (oplader), vis nr., tlf.svarer	4
Opladere	Mobiltlf., tandbørste, ministøvsuger mv.	3
Audio og visuelle	Radio, tv, video, ghettoblaster, mv.	63
Kontorudstyr	pc, printer, fax, mv.	24

Tabel 3.12: Standbyforbruget fordelt på fem anvendelser [53].

Af tabel 3.12 fremgår det, at størstedelen af standbyforbruget går til tv, pc'er og lignende apparater. Det er derfor inden for disse forbrug, at der specielt skal fokuseres med henblik på at reducere standbyforbruget. Beboerne behandler dette område meget forskelligt, hvor nogle er yderst opmærksomme og slukker hvor slukkes kan, er der derimod andre, der er ligeglade og betragter ikke den økonomiske vinding som tilstrækkelig i forhold til arbejdet, der skal lægges i at reducere det.

En undersøgelse fra *Kirsten Gram-Hanssen* [40], der tidligere er nævnt i afsnit 3.2, har vist, at der er en klar sammenhæng mellem husstandens samlede elforbrug og antallet af apparater med et standbyforbrug. *Kirsten Gram-Hanssen* har opgjort standbyforbruget ud fra de tre grupper, der er beskrevet tidligere, hvor der i gruppen af småforbrugere er der 51 %, der kun har ét apparat med standbyforbrug, hvorimod der for storforbrugere er 48 %, der fem eller flere apparater med standbyforbrug.

Den samlede størrelse af standbyforbruget kan både indikere et højt standbyforbrug for nogle enkelte apparater eller en stor mængde af apparater med et lavere standbyforbrug, hvilket dog afhænger af apparatets alder. Standbyforbruget har vist sig, ifølge *Kirsten Gram-Hanssens* undersøgelse, at være blandt de områder hvor mange beboere er inkonsekvente. Størstedelen af beboerne i undersøgelsen har ytret, at de synes, de skal spare af hensyn til miljøet, og beboerne vil gerne spare, men kun et fåtal af dem tænker over standbyforbruget i hverdagen.

3.3.2 Beboernes betydning for elforbruget

I undersøgelser fra *Jesper Ole Jensen* [48] og *Kirsten Gram-Hanssen* [40] er det erfaret, at elforbruget i boliger varierer meget husstande imellem. Igennem en undersøgelse, udført af *Jesper Ole Jensen*, der tidligere er omtalt i afsnit 3.2, havde han adgang til målinger af elforbruget i 16 ud af 35 boligområder, der var inddraget i projektet. Forbrugsmålingerne for det samlede elforbrug, delt ud per beboer i husstandene, viste, at de varierede med en faktor syv mellem de laveste og højeste elforbrug. I samme undersøgelse erfares det, at bestanden af elapparater varierede betydeligt mellem de enkelte husstande. Eksempelvis havde nogle husstande fem fjernsyn, mens andre havde et enkelt. For pc'ere var situationen den samme, hvor nogle husstande rådede over flere computere mens andre ikke ejede nogle. Det bør

dog bemærkes, at forbrugsmålingerne og interviewene, som *Jesper Ole Jensen* analyserede i projektet, er fra 1997/98. Jævnfør afsnittet ovenfor og figurerne 3.16 - 3.17, hvoraf det fremgår, at elforbruget til pc'ere siden 1999 til 2008 er steget fra at udgøre 3 % til 9 % af det samlede energiforbrug¹⁰ må det antages, at der siden denne undersøgelse, er sket en stigning inden for de enkelte husstande i forhold til mængden af pc'ere. En stigning i mængden af elapprater i boliger er dog også observeret, men eftersom elapparaterne bliver mere energieffektive medfører denne stigning i mængde ikke nødvendigvis et øget elforbrug, jf. figur 2.5 på side 17. I samme projekt viste det sig, at ud over fjernsyn, computere og de øvrige elapparaters indflydelse på elforbruget, er det specielt tilstedeværelsen af vaskemaskine og tørretumbler i boligen, og ikke mindst brugen af disse, der kan forklare dele af forskellen i elforbruget mellem husstandene.

Denne sammenhæng med mængden af elapparater og brugen af disse havde *Kirsten Gram-Hanssen* [40] også observeret. De elapparater der i denne undersøgelse viste sig, at have en stor betydning for elforbrugets størrelse var: opvaskemaskine, vaskemaskine, tørretumbler, tv/video, mikrobølgeovn og ovn. Analysen viste specielt en stærk sammenhæng mellem elforbruget og brugen af tørretumbler og tv/video, hvilket fremgår i tabel 3.13 og 3.14.

Brug af tørretumbler	Småforbruger [%]	Normalforbruger [%]	Storforbruger [%]
1 gang pr. uge	28	33	38
2 gange pr. uge	13	39	48
3 gange pr. uge	14	28	58
4 gange pr. uge	8	28	64
5 eller flere gange pr. uge	9	21	70

Tabel 3.13: Sammenhæng mellem typen af forbruger og brugen af tørretumbler, samlet for husstanden [40].

Antal af Tv/video	Småforbruger [%]	Normalforbruger [%]	Storforbruger [%]
1 Tv/video	50	30	20
2 Tv/video	31	40	29
3 Tv/video	22	32	46
4 Tv/video	16	36	48
5 eller flere tv/video	7	13	80

Tabel 3.14: Sammenhæng mellem typen af forbruger og det samlede antal af tv-/ videoapparater for husstanden [40].

Tv og videoudstyr er i Danmark standardapparater i boligen, hvilket *Kirsten Gram-Hanssens* undersøgelse også indikerede, eftersom alle adspurgte familier ejer som minimum ét tv. Dele

¹⁰Der jævnfør figur 2.5 på side 17 har været stort set konstant siden 1992

af forskellene i elforbruget til tv og video lå i hvor mange af disse apparater familierne besad. Et mønster, som udtalt af en af familierne, viste sig at være gældende for flere af familierne:

„Ja, det i soveværelset, det er jo, fordi det er meget hyggeligt, synes jeg. Og så det på arbejdsværelset, det var et vi havde, da knægten flyttede hjemmefra, så røg det derind. (...) Og det i kælderen, det ved jeg, der har vi altid bare haft et dernede og så har vi et lille gammelt sort/hvid, der bare står.“ [40, s. 85]

Dette citat viser, at flere familier anskaffer nye apparater men skaffer sig ikke af med de gamle, ej heller når børnene flytter hjemmefra. Hvis disse apparater ikke anvendes men standbyforbruget ikke afbrydes vedbliver et unødvendigt forbrug, hvor der reelt set var en mulighed for at reducere det.

Generelt har analysen fra *Kirsten Gram-Hanssen* vist, at der var stor sammenhæng mellem beboernes elforbrug, og deres vaner omkring udførelse af husligt arbejde samt brugen af apparater til underholdning. Dette blev blandt andet bekræftet igennem interviews, der blev gennemført samtidig med undersøgelsen. I forbindelse med brug af vaskemaskine og tørretumbler viste undersøgelsen, at de blev anvendt meget forskelligt blandt husstandene. Brugen af vaskemaskiner afhang meget af familiens størrelse og mængden af vasketøj, der blev genereret af familien. Ved gennemgang af dette område med familierne kom det frem, at flere af dem ikke følte, at deres tøj var beskidt, men at det derimod var samfundets normer, der var den bestemmende faktor for, hvor længe de gik i samme stykke tøj. Alle husstande anvendte en vaskemaskine, hvorimod anvendelsen af tørretumbleren delte beboerne i to grupper. Nogle familier kunne slet ikke drømme om at eje en tørretumbler, mens andre ikke kunne leve uden. Som luksusbrug af tørretumbleren blev det erfaret det, at en familie anvendte den til at opvarme håndklæder, så de var lune, til når de skulle tørre sig efter badet.

Husholdningsmaskinerne betragtes af flere beboere som et overskueligt område at være energigivende på. Udskiftningen af et apparat til en nyere og energieffektiv model, betragtes af beboerne som en nem energigivende handling, idet når udskiftningen er gjort så er det overstået og kræver dermed ikke en ændring af vaner. Et eksempel herpå er en familie der, efter eget udsagn, udskifter apparater af hensyn til miljøet og ikke af hensyn til deres økonomi.

„Det er ligesom overstået nu, ikke. Det skal du ikke tænke så meget mere over.“ [40, s. 87]

Foruden beboere, der udskiftede apparater af hensyn til miljøet og ikke af hensyn til deres økonomi, viste undersøgelsen også, at andre beboere først udskiftede apparater, når det gik i stykker eller for eksempel når hele køkkenet skulle renoveres. I forbindelse med renovering var det dog ikke altid energiforbruget, men derimod designet, der kom i første række. Specielt denne overvejelse omkring design havde også stor betydning for andre elforbrugende apparater til hjemmet, specielt tv/dvd/spillemaskiner og musikanlæg. Igennem de seneste år har udviklingen på markedet dog styret i en retning, hvor mange af producenterne gør meget ud af at opnå en høj energieffektivitet.

Elforbruget til belysning er blevet undersøgt af *Jesper Ole Jensen* [48] gennem interviews undersøgt forskellen på anvendelse af belysning i boliger. Interviewene blev udført i 1997/98 og det fremgik, at omkring en tredjedel af de adspurgte husstande ikke anvendte lavenergipærer hvilket står i kontrast til andre af de adspurgte husstande, hvor nogle udelukkende

3.3.3. Væsentlige parametre for elforbruget

anvendte lavenergipærer i boligen. Dette forhold vurderes dog at have ændret sig betydeligt, eftersom der i 2009 blev indført forbud mod distribution af matterede glødepærer, og dette forbud medfører ligeledes, at samtlige glødepærer skal udfases i tiden frem mod 2012 [31]. Hvad angår vaner for anvendelse i af belysning i boliger, indikerede *Jesper Ole Jensens* undersøgelse en betydelig forskel mellem de enkelte husstande. I nogle husstande blev der ikke anvendt unødigt belysning, hvilket en beboer beskriver på følgende måde:

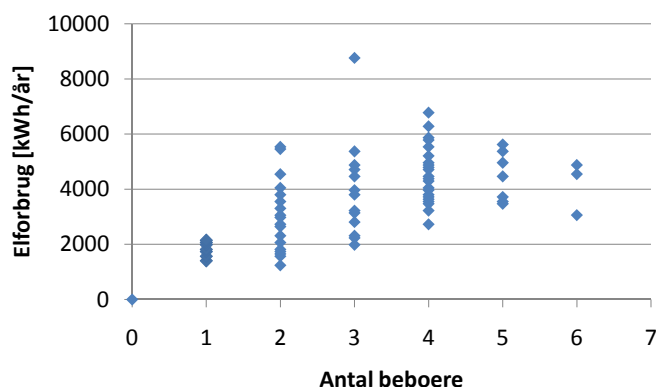
„Så snart vi begynder at se fjernsyn, så har vi kun lige den der tændt (lampen over fjernsynet), og den i reolen. Vi synes ikke det er nødvendigt med projektører på når vi ser fjernsyn.“ [48, s. 130]

I andre husstande blev belysningen anvendt på en helt anden måde, hvor det ikke altid var en vane hos beboerne at slukke for lyset i ubenyttede rum. Et eksempel på dette beskrev en beboer gennem interviewene på denne måde:

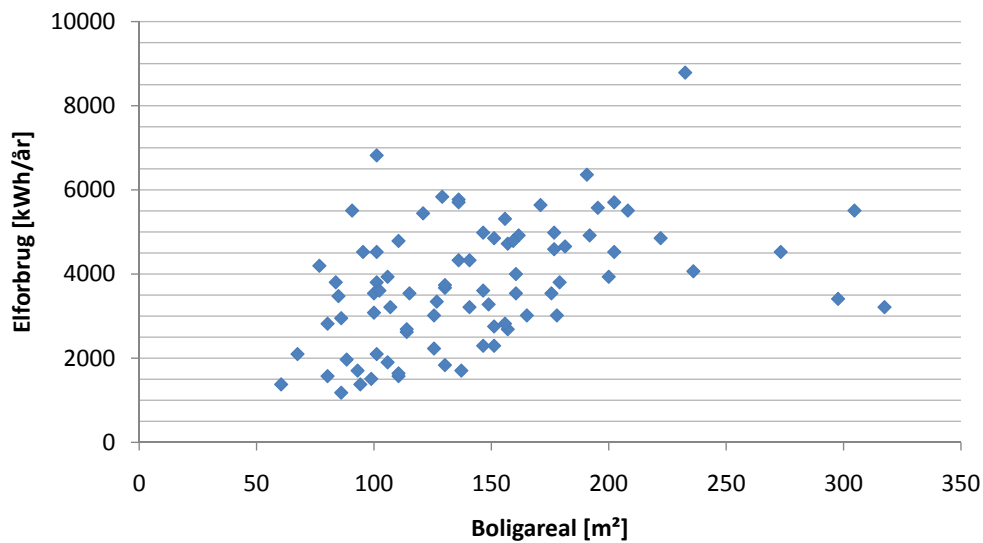
„Der er nogen der kan lide at sidde meget i mørke, og have et stearinlys tændt, men sådan har vi det ikke. Så jeg tror tit vi har mere lys tændt end mange andre har.“ [48, s. 129]

3.3.3 Væsentlige parametre for elforbruget

I forbindelse med målingerne fra Odense, der er udført af *Casper Kofod*, har de vist en tendens til, at antallet af beboere, samt størrelsen af boligen, har indflydelse på det samlede elforbrug. Dette er illustreret på figur 3.20 og 3.21, hvor det kan ses, at der ved henholdsvis et stigende antal beboere og et stigende boligareal er en tendens til, at elforbruget er stigende. Samme tendens er set i et projekt fra *Gram-Hanssen et al.* [42].



Figur 3.20: Elforbruget i 85 parcelhuse sammenholdt med antallet af beboere [53].



Figur 3.21: Elforbruget i 85 parcelhuse sammenholdt med boligarealet [53].

På tilsvarende måde har *Jesper Ole Jensen*, der er omtalt tidligere i dette afsnit, forsøgt at identificere nogle parametre, der kan forklare dele af den variation af elforbruget, der er mellem husstandene. I projektet er boligforbruget, der er opgjort som boligarealet delt med antal personer i husstanden, vurderet således, at et stigende boligforbrug typisk medfører et stigende elforbrug. På tilsvarende måde konkluderes det, at der er en tendens til, at elforbruget stiger med stigende husstandsstørrelser, hvilket beskriver antal personer per bolig.

I en anden undersøgelse af *Gram-Hanssen et al.* [42] blev det undersøgt, hvilken forklaringsgrad beboernes alder har i forhold til elforbruget i boliger. Elforbrugets størrelse er i analysen vurderet i forhold til et livscyklusforløb, hvor beboerne har forskellig adfærd i forskellige stadier i deres liv, der dermed forårsager forskellige størrelser på elforbruget. *Kirsten Gram-Hanssen* betragter sammenhængen mellem elforbrug og ældste beboers alder således, at forbruget er stigende når alderen er mellem 20 – 50 år, hvorefter den er faldende. I denne sammenhæng er der taget højde for at antallet af beboere i husstanden er varierende over tid, dvs. et stigende forbrug i perioden hvor der er børn/teenagere i husstanden, og et faldende forbrug når børnene flytter hjemmefra. Overordnet set blev husstandene i vurderet ud fra den ældste persons alder, da dette er valgt som indikator for husstandens placering i livscyklusforløbet, der er beskrevet kort ovenfor. Sammenhængen mellem alderen for husstandens ældste beboer og elforbruget beskrives i alle typer af boliger som værende kurvelineært, denne tendens er dog stærkest i parcelhuse, række-, kæde- og dobbelthuse, hvor den for lejlighed kun er svagt varierende i hele livscyklussen. *Kirsten Gram-Hanssen* [40] erfarede yderligere, at beboernes køn har betydning for elforbruget. Resultatet fra undersøgelsen viste, at enlige kvinder har tendens til at have et højere elforbrug end enlige mænd.

I undersøgelsen fra *Gram-Hanssen et al.* [42] blev der fundet frem til, at tilstedeværelsen af børn og teenagere har betydning for størrelsen af elforbruget i boliger. Resultaterne viste, at børn har en positiv effekt på forbruget, idet de ikke forbruger ligeså store mængder som en voksen. Teenagere derimod har en negativ effekt, da en teenagers elforbrug er højere end en voksens elforbrug. For yderligere beskrivelse af sammenhængen mellem børn og teenagers tilstedeværelse se afsnit C.5.2 i appendiks C.

I et forskningsprojekt har *Pedersen et al.* [74] foretaget en analyse af elforbruget i husholdninger, hvor baggrundsmaterialet dækker over forbrugsopgørelser og spørgeskemaundersøgelser, blandt 811 danske husstande, foretaget i 1996.

Gennem projektet er der lagt fokus på at undersøge tre forskellige hovedområder indenfor beboernes levemåde og disse områders forklaring af elforbruget. De tre hovedområder er beboernes levevilkår, livsorientering samt sociale relationer. Indenfor området levevilkår undersøges faktorer som indtægt, uddannelse, alder mv. Hvad angår livsorientering undersøges det, hvorvidt beboerne er bevidste om deres elforbrug i dagligdagen, i hvor høj grad de føler, at de selv har mulighed for at påvirke miljøforholdene i deres omgivelser, samt undersøgelser af hvor stressede beboerne føler sig i hverdagen mv. I forbindelse med de sociale relationer er det styrken af beboernes sociale netværk, samt hvilke normer med relevans for el- og miljøområdet, der deles inden for deres omgangskreds mv.

For at danne et indtryk af hvilke faktorer der kan forklare variationen i elforbruget, er der i projektet foretaget en regressionsanalyse, hvor hovedresultaterne er gengivet i tabel 3.15. Den første søjle i tabellen beskriver de enkelte faktorerers signifikansniveau, og den anden søjle beskriver de enkelte faktorerers relative betydning for elforbruget, således at deres betydning kan sammenlignes på trods af de forskellige enheder, som faktorerne opgøres i.

Sig.	Beta	Variabler (R = 0,42)
* 0,00	0,30	Viden om elregningens størrelse
* 0,00	0,29	Indtægt
0,21	−0,06	Sociale relationer
0,82	0,01	Uddannelse
0,90	0,00	Handlingsparathed

* Værdierne er signifikante.

Tabel 3.15: Levemådens indflydelse på elforbruget [74].

Overordnet set forklarer de fem faktorer, der er angivet i tabel 3.15, 42 % af variationen i elforbruget. Det bør dog bemærkes, at kun to af de angivne faktorer har et lavt signifikansniveau, og at disse faktorer har omtrent lige stor indflydelse på elforbruget.

Viden om elregningens størrelse er bestemt på baggrund af beboernes besvarelse i spørgeskemaet, om hvor meget husstanden kvartalsmæssigt betaler for el angivet i intervaller. Beboernes besvarelser er derefter sammenholdt med forbrugsopgørelser fra den aktuelle elforsyning, og derved kan denne faktor inddrages i regressionsanalysen. Undersøgelsen fra *Gill et al.* [38], der blev gennemgået i afsnit 3.1.1, viste samme tendens, som resultaterne fra *Pedersen et al.* Resultaterne skal forstås således, at jo mindre viden beboerne har om størrelsen af deres elregning, jo større elforbrug kan husstanden forventes at have. På tilsvarende måde skal indtægten, i denne sammenhæng, forstås således, at jo højere indtægt beboerne har, jo større elforbrug kan husstanden forventes at have.

I et forsøg på at nuancere analyseresultaterne er der i projektet også udført en multipel regressionsanalyse, hvor resultaterne er gengivet i tabel 3.16. B-koefficienten, der er angivet i den første søjle i tabellen betyder, at eksempelvis et ekstra barn i husholdningen vil medføre en stigning i elforbruget på 420 kWh/år. Beta-koefficienten, der er angivet i den anden søjle

i tabellen, beskriver de enkelte faktoreres relative betydning for elforbruget, således at deres betydning kan sammenlignes på trods af de forskellige enheder, som de opgøres i.

B	Beta	Variabler (R = 0,72)
7	0,25	Brutto kvadratmeter i boligen
420	0,20	Børn
503	0,16	Boligtype: Parcelhus *
380	0,15	Voksne
349	0,15	Viden om elregningens størrelse
96	0,11	Hårde hvidevarer **
37	0,10	Elektriske apparater ***
444	0,10	Boligtype: Rækkehus *

* Der er regnet med en lejlighed som basisniveau.

** Antal hårde hvidevarer angives i det præcise antal.

*** Antal elektriske apparater angives i intervaller og ikke ved det præcise antal.

Tabel 3.16: Signifikante variablers indflydelse på elforbruget. Alle variable er signifikante ved 1 %-niveauet [74].

Det er værd at bemærke, at samtlige faktorer i tabel 3.16 har et signifikansniveau på omkring 1 %, samtidig med, at disse faktorer tilsammen beskriver 72 % af variationen i elforbruget mellem husstandene. I projektet konkluderes det, at det er disse faktorer, der har størst påvirkning på elforbruget. Samtidig konkluderes det, at beboernes kendskab til elregningens størrelse har en signifikant betydning for husstandens elforbrug, hvilket blev vist i tabel 3.15, og at det indikerer, at det konkrete fokus på elforbruget påvirker adfærden.

Igennem projektet er der også opnået kendskab til, i hvilket omfang beboerne lægger vægt på elforbruget i det daglige. Her viser det sig, at 42 % ikke, eller kun i begrænset omfang, lægger vægt på elforbruget i det daglige, 35 % at de sommetider lægger vægt på elforbruget, og 23 % at de lægger stor vægt på elforbruget. Sammenholdes dette med, at analysen også viser, at de personer, der føler sig stresset/tidspresset i hverdagen, også er de folk, der har det højeste elforbrug, samt at 66 % svarer, at de ikke interesserer sig for elforbruget, da det blot anses som værende en del af de faste udgifter, indikerer det, at elforbruget kun i begrænset omfang er i fokus i det daglige.

3.3.4 Diskussion

Igennem de foregående afsnit, er der opnået et større kendskab til elforbruget i boliger, med hensyn til størrelsen af dette samt fordelingen af forbruget blandt de enkelte slutanvendelser. Ydermere er studier af personers holdninger i forbindelse med valg, udskiftning og brug af elapparater studeret, og det er forsøgt at klarlægge hvilke forhold, der kan beskrive de, i nogle tilfælde betydelige, forbrugsvariationer, der kan ses mellem enkelte husstande.

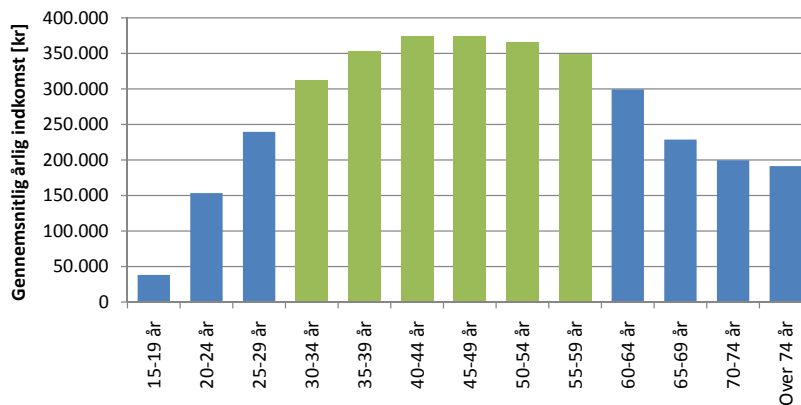
Dette litteraturstudie påviser, at der er betragtelige variationer af elforbruget blandt danske boliger. Undersøgelser af *Jesper Ole Jensen* og *Casper Kofod* viser, at variationen mellem husstanden med det laveste og det højeste elforbrug svarer til henholdsvis til en faktor syv

og en faktor ti i husstandens totale elforbrug. Årsagen til disse betydelige forskelle skal findes i beboernes forskellige behov for elforbrugende apparater og udstyr i boligen, men også beboernes vaner for brug af apparaterne har en stor indflydelse på forbruget.

At beboere har forskellige behov for køleskabe og frydere, kommer ifølge *Casper Kofods* undersøgelse ikke til udtryk blandt bestanden af disse, eftersom det påvises, at disse hårde hvidevarer er standardudstyr i boliger, og der ikke ses nogen stigning som følge af indkomst. Da der alligevel fremgår en betydelig variation i forbruget på dette område, kan det tyde på, at det er størrelsen af de hårde hvidevarer og deres energieffektivitet, der er afgørende for elforbruget til disse apparater. Størrelsen af apparaterne har betydning for, hvor stort et volumen apparaterne skal nedkøle, og energieffektiviteten afhænger typisk af apparaternes alder og kvalitet, hvor beboernes opmærksomhed og handlekraft på det område er meget varierende. Det er blandt andet kommet til udtryk gennem interviews med personer, der, efter eget udsagn, udskifter elforbrugende apparater under hensyntagen til miljøet, mens andre personer går mere op i apparaternes design end deres energieffektivitet.

Opvaskemaskiner er, sammenholdt med køleskabe og frydere, mindre udbredt, idet undersøgelser tyder på, at det er omkring halvdelen af boligerne, der ejer en opvaskemaskine. Ikke desto mindre viser en analyse på 100 husstande i Odense, at elforbruget til dette apparat er særdeles varierende, hvor der kan være op til en faktor 14 imellem husstande med lavt og højt forbrug. Årsagerne til de store forskelle mellem husstandenes elforbrug til opvaskemaskiner kan primært findes i forskellige energieffektiviteter samt brugen af disse. I forbindelse med energieffektiviteten af opvaskemaskiner hænger dette sammen med apparatets kvalitet og alder. Gennem litteraturstudiet er der, ligesom det er nævnt tidligere, fundet forskelle mellem beboernes handlekraft i forbindelse med udskiftning af ineffektive elapparater, hvor nogle beboere ikke udskifter disse, før de enten går i stykker eller i forbindelse med en større udskiftning eller renovering af boligen. Undersøgelser viser, at levetiden på kvalitetsopvaskemaskiner kan være op mod 15 til 20 år [10], at der er sket over en tredobling af bestanden af opvaskemaskiner i den danske husholdning over en tyveårig periode fra 1989 til 2009 [27], samt at opvaskemaskiners gennemsnitlige, specifikke elforbrug i samme periode er faldet med omkring 37 % [27]. Det tyder på, at det er blevet mere almindeligt med opvaskemaskiner i boligen, men også at den eksisterende bestand af opvaskemaskiner kan være af betydelig varierende alder og energieffektivitet, hvilket kan være dele af årsagen til det varierende elforbrug til opvaskemaskiner.

Casper Kofod har gennem undersøgelsen fra Odense fundet indikationer på, at beboere i aldersgruppen 30 til 60 år anvender opvaskemaskinen oftere end personer over 60 år. Årsagen til dette kan være, at der i husstande med beboere i den førnævnte aldersgruppe, kan være flere beboere tilstede i boligen på grund af hjemmeboende børn. Dette vil naturligvis medføre større mængder opvask og derved en hyppigere anvendelse af opvaskemaskinen. Foruden ovennævnte forhold har *Casper Kofods* undersøgelse indikeret, at der er et stigende brug af opvaskemaskiner som funktion af en stigende husstandsindkomst. Det kan dog hænge sammen med førnævnte begrundelse, hvor personer i aldersgruppen 30 til 60 år oftere anvender opvaskemaskiner. Dette vurderes på grundlag af, at det netop er denne aldersgruppe, der gennemsnitligt har den højeste indkomst, hvilket er vist på figur 3.22, hvor aldersgruppen 30 til 60 år er markeret med grøn.



Figur 3.22: Gennemsnitlig årlig indkomst i Danmark i 2009 for forskellige aldersgrupper [95].

Hvad angår vaskemaskiner og tørretumblere viser undersøgelsen af *Casper Kofod*, at vaskemaskiner er særdeles udbredt i danske boliger, specielt i parcelhuse, mens tørretumblere er mindre udbredte. Det er endvidere beskrevet, at elforbruget til vaskemaskiner varierer med omkring en faktor 30 mellem husstanden med det laveste og det højeste forbrug i undersøgelsen med de 100 husstande i Odense. For tørretumblere kan der ses en endnu større forskel mellem husstandene, hvor det er vist, at der kan være op til en faktor 60 i forskel. Den store variation blandt husstandene viser, at brugen af disse to hvidevarer kan være vidt forskellig, hvor et tidligere nævnt interview tillige viser, at tørretumbleren i nogle tilfælde anvendes til, hvad der kan betegnes som luksusbrug. *Jesper Ole Jensen* har beskrevet, at brugen af vaskemaskine og tørretumbler har stor indflydelse på boligens samlede elforbrug. Samme forhold gengives i en undersøgelse af *Kirsten Gram-Hanssen*, hvor der tydeligt kan ses en opdeling af brugen af tørretumblere afhængigt, af om husstanden er blandt den laveste eller højeste tredjedel, når husstandes totale elforbrug sammenholdes.

Af øvrige væsentlige elforbrug kan nævnes brugen af apparater som for eksempel tv, video, dvd, computere og HIFI. *Casper Kofods* undersøgelse viser, at der i 1999/2000 gennemsnitligt blev anvendt omkring 160 kWh/år på EDB-udstyr, hvor det dog i dette projekt vurderes, at dette forbrug vil være højere i dag, da der som vist tidligere, er sket en stor udvikling i hvor mange personer, der dagligt anvender computere. Hvad angår tv, video, dvd og HIFI var der 1999/2000 et gennemsnitligt forbrug på omkring 210 kWh/år i husstandene. Brugen af sidstnævnte apparater har *Kirsten Gram-Hanssen* vist, at der har en stor indflydelse på husholdningens elforbrug, hvor der også kan ses en betydelig forskel på, hvor mange tv-apparater der er til rådighed, afhængig af om husstanden er blandt den laveste eller højeste tredjedel, når husstandes totale elforbrug sammenholdes.

Der er i flere projekter undersøgt hvilke parametre, der kan have en forklarende effekt i forbindelse med elforbruget i boliger. Analyser fra *Jesper Ole Jensen*, *Kirsten Gram-Hanssen* samt *Pedersen et al.* indikerer alle, at antallet af beboere er af afgørende betydning for elforbruget, hvor alle undersøgelserne peger i retning af, at elforbruget er stigende som følge af et stigende antal beboere. Ifølge *Jesper Ole Jensen* og *Pedersen et al.* er det specielt tilstedeværelsen af børn og teenagere i husstanden, der har betydning for elforbrugets størrelse. I mindre husstande har *Kirsten Gram-Hanssen* endvidere vist indikationer på, at enlige kvinder typisk har et højere elforbrug end enlige mænd.

Foruden antallet af beboere, tyder undersøgelser også på, at boligarealet kan have en for-

klarende effekt i forbindelse med elforbruget, hvor et stigende boligareal typisk medfører et stigende elforbrug.

Pedersen et al. har på baggrund af elforbrugsmålinger i 1996 opstillet et matematisk udtryk, der ved at inddrage parametre som boligareal, antal børn og voksne, boligtype, antallet af hårde hvidevarer og andre elektriske apparater samt en vurdering af beboernes viden omkring elforbruget, kan forklare 72 % af den variation af elforbruget, der var blandt de omkring 800 husstande, der deltog i undersøgelsen.

3.4 Vandforbrug

Indledende undersøgelser viser, at varmtvandsforbruget har en betydelig indflydelse på energiforbruget i boliger, jævnfør figur 2.17 på side 28. Til trods for dette er det gennem litteraturstudiet erfaret, at der kun findes en beskedent mængde litteratur på dette område. Der er fundet flere undersøgelser på det samlede vandforbrug, men mindre omkring mængden af varmtvandsforbruget i boliger eller energiforbruget til opvarmning af dette. Af denne årsag er det i de efterfølgende afsnit valgt, at opdele litteraturstudiet i henholdsvis et samlet vandforbrug og varmt brugsvand, for at opnå så meget kendskab til områderne som muligt.

3.4.1 Samlet vandforbrug

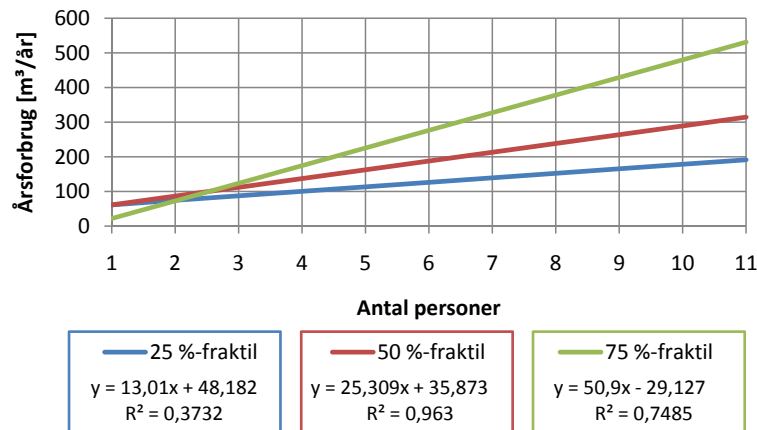
Det samlede vandforbrug der blev anvendt i danske husholdninger i 2007 svarede til et årligt forbrug på $45 \text{ m}^3 \text{ per person}$, hvilket fremgår af figur 2.4 på side 17 [8]. I en opgørelse fra *Dansk Vand- og Spildevandsforening (DANVA)* [71] blev det årlige vandforbrug opgjort til $41,3 \text{ m}^3 \text{ per person}$. Forskellen mellem de to opgørelser er, måden hvorpå opgørelserne er udformet. Oplysningerne fra *DANVA* bygger på data fra over 2.575.000 personer, hvilket svarer til omkring 47 % af den danske befolkning [93]. Oplysningerne om vandforbruget fra *Danmarks Statistik* bygger blandt andet på oplysninger fra *DANVA*, men de er korrigeret af *Danmarks Statistik* efter deres oplysninger om blandt andet husholdninger uden tilknytning til almene vandforsyninger [94]. Denne forskel i måden hvorpå opgørelserne udføres er skyld i differensen mellem forbrugets størrelse.

Miljøstyrelsen et al. beskrev, at der ikke var tilstrækkeligt grundlag for entydigt at afklare, om anvendelse af boligenhed som normeringsenhed var bedre end andre enheder. På baggrund af disse iagttagelser anbefales det i projektet, at der anvendes antal personer som normeringsenhed. Denne vurdering er sket på baggrund af det i tabel 3.17 angivne antal forbrugsdata, der var indsamlet over en femårig periode.

Figur 3.23 viser en grafisk præsentation af nogle resultater fra projektet, hvor disse kan anvendes til estimering af det årlige vandforbrug, som funktion af antal personer i parcel- og rækkehuse. I forbindelse med databehandlingen i projektet, og derved også på figur 3.23, fremgår der at være over ti personer i nogle parcel- og rækkehuse. Dette vurderes, at have samme årsag som ved lejlighederne, hvor nogle vandmålere dækker over flere husstande. Da det for parcel- og rækkehuse kun er 2 % af husstandene, hvor der var over fem beboere, vurderes analyserne af *Miljøstyrelsen et al.* at være brugbare.

Forbrugskategori	Data			
	Time	Døgn	Måned	År
Parcel-/rækkehuse	27.994	4.018	51	36.889

Tabel 3.17: Antal forbrugsdata ud fra målingstype [61].



Figur 3.23: Årligt vandforbrug som funktion af antal personer, gældende for parcel- og rækkehuse [61].

3.4.1.1 Beboernes indflydelse på vandforbrugets størrelse

Undersøgelsen af *Jesper Ole Jensen* [48], som tidligere er omtalt i afsnit 3.2, har foretaget en analyse af energiforbrug, derunder også vandforbrug, i over 5500 danske boliger. Det bør dog bemærkes, at der ikke var målt på energiforbruget til opvarmning af varmt brugsvand, men derimod på mængden af det samlede vandforbrug i boligerne. Analysen viste, at forhold som boligforbrug, bebyggelsens alder eller beboernes indtægt ikke kunne forklare forskelle i vandforbrug mellem enkelte husstande. Derimod indikerede undersøgelsen, at forhold som ejerform af husstanden, tilstedeværelsen af teenagere i boligen og daglige vaner og rutiner kunne forklare dele af den variation, der var i vandforbruget mellem husstandene. Undersøgelsen viste, at det gennemsnitlige vandforbrug per person i hvert boligområde varierede med over en faktor to mellem de forskellige områder; fra et gennemsnitligt forbrug på $32 \text{ m}^3/\text{person per år}$ i et område til $69 \text{ m}^3/\text{person per år}$ i et andet. Tilsvarende viste projektet, at internt i et af boligområderne varierede vandforbruget med en faktor syv mellem det laveste og det højeste forbrug i området.

I en undersøgelse fra *Gram-Hanssen et al.* [42], hvor der blev anvendt regressionsanalyse, fremgik det, at de parametre der havde størst indflydelse på vandforbrugets størrelse var husstandsstørrelse, boligarealet, husstandens samlede indkomst samt tilstedeværelsen af småbørn¹¹, hvor den vigtigste parameter klart var antallet af beboere i husstanden.

I en anden undersøgelse fra *Kirsten Gram-Hanssen* [40] hvor fokus igennem en spørgeundersøgelse var orienteret mere imod brugernes adfærd omkring forskellige vaner i hjemmet. I denne analyse blev det igen erfaret, at husstandsstørrelsen har stor indflydelse på det sam-

¹¹Definition småbørn for undersøgelsen er 0 – 6 år.

lede vandforbrug. Ydermere blev det erfaret, at beboernes vaner omkring tøjvask og bad var meget varierende og der blev registreret en sammenhæng mellem beboernes badevaner og det samlede varmeforbrug.

I undersøgelsen af *Jesper Ole Jensen* blev der foretaget interviews med forskellige beboere for blandt andet at klarlægge hvilken adfærd beboerne havde omkring vask af tøj. Nogle beboere havde ikke vaskemaskine i hjemmet, men anvendte et fællesvaskeri, hvor de vaskede fire til fem maskiner hver 10. – 12. dag. Andre beboere beskrev, at de anvendte vaskemaskinen fem til syv gange om ugen og andre igen op til 16 gange om ugen, hvilket blandt andet skyldtes behov for vask af sportstøj, arbejdstøj mv.

En af årsagerne til variationen af vandforbruget mellem de enkelte husstande, kunne forklares ud fra forholdene mellem hvor ofte der blev vasket tøj og hvor tøjet blev vasket henne. Det betyder, at der kan være betydelig forskel i det årlige vandforbrug i boligen, afhængig af om eksempelvis beboere der ofte vasker tøj har privat vaskemaskine, eller om de benytter et fællesvaskeri, hvilket kommer til udtryk ved, at vandforbruget i boligen vil syne lavere hvis fællesvaskeriet benyttes.

I forbindelse med vaner for tøjvask fandt *Kirsten Gram-Hanssen KGH029* ligeledes frem til, at flere af beboerne vasker tøjet før det når at blive beskidt. Årsagen til denne tendens bliver af flere af familierne beskrevet deres vaskevaner i forhold til hvad der er normalen i samfundet i dag. Dette er udtrykt på følgende måde for en af familierne:

„Men idag kunne jeg da ikke tænke mig at have de samme par sokker på to dage i træk. Men det gjorde man jo dengang, ikke. Fordi, altså, sådan var tiden jo. Det er da tiden, der har ændret sig. [40, s. 73]

I forbindelse med personlig hygiejne var der i *Jesper Ole Jensens* projekt, nogle beboere der skønnede, at varigheden af deres bade varede tre til fem minutter, og at der ikke var forskel på om det var de voksne eller børnene, der badede. En anden beboer beskrev, at hun kunne tage karbad, der varede i et par timer, hvor der blev skiftet vand ad flere omgange. Når beboerne blev spurgt omkring badevaner, nævnte flere, at teenagerne tog lange bade - flere op til 30 minutter. I undersøgelsen var vandforbruget sammenholdt med beboernes alder i husstandene, hvilket viste en tendens til, at familier med små børn brugte mindre vand per person end familier uden børn. *Jesper Ole Jensen* vurderede, at det kunne forklares ved, at små børn ikke optrådte som selvstændige forbrugere i boligerne, hvorved et lidt større forbrug kunne deles ud på flere personer, når forbruget blev opgjort på basis af antallet af beboere. Tilsvarende indikerede undersøgelsen, at familier med teenagebørn kunne have et betydeligt højere forbrug per person end andre familier.

Kirsten Gram-Hanssen har i en lignende undersøgelse undersøgt det samlede vandforbrug i boliger i forbindelse med et projekt fra Albertslund [40]. Her blev det konkluderet, at der er en sammenhæng mellem forbruget af vandforbrugets størrelse og antallet af bade per person per uge. Det fremgår tydeligt, fra tabel 3.18, at andelen af småforbrugere, der tager bad mindre end tre gange om ugen, er 53 %, hvorimod der ved storforbrugeren er hele 42 % der tager bad hver dag. Det bør dog bemærkes, at nogle badeaktiviteter kan foretages andre steder end i boligen, eksempelvis i sportsklubber eller lignende. Det betyder, at en familie med dette mønster kan tilhøre gruppen småforbrugere, idet mængden af bade i hjemmet er lav, men i disse tilfælde er tilknytningen til en sådan gruppe ikke et udtryk for miljøvenlig adfærd i familien.

Badevaner	Småforbrugere [%]	Normalforbrugere [%]	Storforbruger [%]
Bader hver dag	25	33	42
2 – 3 gange per uge	40	39	21
Mindre end 3 gange per uge	53	19	28

Tabel 3.18: Sammenhæng mellem badevaner og vandforbruget per person [40].

Vurderes årsagerne til mængden af bade ud fra interviewmaterialet, fremgår det, at der er stor forskel iblandt familierne. Flere af dem mener dog, ligesom det også blev nævnt for tøjvask i afsnittet ovenfor og afsnit 3.3, at det har noget med samfundets normer at gøre, hvor ofte de går i bad. Nogle beboere mener, at renlighed er lig med social status i samfundet i dag, som udtrykt ved følgende udtalelse:

„Det er rart, og så føler man, at man er ren. Og så er man velsoigneret, når man møder på arbejde. Ja, det giver en hvis selvtillid, at man bare er i orden. Og det er lige meget, hvilket job man har. Sådan har jeg det i hvert fald.“ [41, s. 79]

En anden familie mener, at det skyldes samfundets normer i dag og at årsagen til at tage bad ikke skyldes, at de er så beskidte hver dag at det er nødvendigt:

„... Nej, ikke for at blive ren, for så beskidte tror jeg ikke folk er nu om stunder. Jeg tror, det er noget, en livsstil vi har fået, ikke. Jeg tror ikke på.“ [41, s. 78]

Andre eksempler findes også blandt andre familier, hvor der er familiemedlemmer, der kun bader få gange om ugen. Andre familier argumenterer for at det slet ikke er sundt for kroppens balance at bade så mange gange om ugen. Mange af familierne bader dog ofte, så denne problemstilling dukker mere op som et dilemma mellem samfundets normer og hvad der ifølge dem selv, er sundt for kroppen.

3.4.2 Varmt brugsvand

Badevanerne der er beskrevet ovenfor og sammenhængen mellem det samlede vandforbrug og antallet af bade må at der er en sammenhæng I forbindelse med energimærkning af en boligblok i Nørresundby, blev der målt på det totale varmtvandsforbrug for denne. Boligblokken er opført i 1973 og ombygget i 1993, og den består af alt i alt 128 lejligheder. I 1996 var der et samlet varmtvandsforbrug for boligblokken på 3148 m^3 , hvilket svarede til 38 % af det totale vandforbrug eller gennemsnitligt $24,6 \text{ m}^3$ per lejlighed. [45]

I Albertslund Kommune blev der i 2007 udfærdiget et grønt regnskab for kommunen, hvor boligernes vandforbrug blev opgjort og sammenholdt med forbruget fra tidligere år. I det grønne regnskab, var der skønnet hvor stor en del af varmekonsumet, der går til opvarmning af varmt brugsvand afhængig af boligtype og opførelsesår. De anvendte værdier fra det omtalte grønne regnskab fremgår i tabel 3.19.

Boligtype	Opførelsesår	Andel
		[%]
Parcelhus	Før 1979	25
	Efter 1979	35
Tæt/lav ejerbolig	Før 1979	30
	Efter 1979	40
Tæt/lav lejebolig	Før 1979	30
	Efter 1979	40
Etagebolig	Før 1979	35
	Efter 1979	45

Tabel 3.19: Andel af det samlede varmekonsum der anvendes til opvarmning af det varme brugsvand [72].

Bøhm et al. [8] udgav i 2009 en rapport omhandlende et større måleprojekt af blandt andet varmtvandsforbruget i danske boliger. I projektet er der gennem flere år målt på varmtvandsforbrugets størrelse hovedsagligt i større boligblokke men også i tæt/lav-bebyggelse og enkelte enfamiliehuse. Det primære formål med målingerne var, at danne grundlag for at vurdere størrelsen af energiforbruget til opvarmning af varmt brugsvand, eftersom det i projektet oplyses, at der ikke hidtil har været udført større registreringer og undersøgelser af dette i Danmark. Dette er på trods af, at energiforbruget til opvarmning af varmt brugsvand indgår som en del af energiberegninger på nuværende tidspunkt.

I mange projekter var der ikke muligt, at adskille energiforbruget til rumopvarmning og energiforbruget til opvarmning af varmt brugsvand, eftersom disse to forbrug ofte opgøres samlet. En gennemgang af resultaterne fra *Bøhm et al.* viser, ifølge tabel 3.20, at omkring 30 % af det samlede energiforbrug til opvarmning anvendes til opvarmning af varmt brugsvand. Dette er i overensstemmelse med, hvad *Bøhm et al.* oplyser i rapporten, hvor der gives udtryk for, at vurderinger blandt flere energikonsulenter lyder på, at det er omkring en tredjedel af energiforbruget til opvarmning, der anvendes til produktion af varmt brugsvand. I de 30 % samt den tredjedel, der er beskrevet af *Bøhm et al.*, kommer derudover energiforbruget til den opvarmning af brugsvandet, der sker under eksempelvis madlavning, opvask og tøjvask. På tilsvarende måde kan mængden af det varme brugsvand i nogle af målingerne sammenholdes med det totale vandforbrug i bygningerne, og som det er vist i tabel 3.20, udgør det varme brugsvand 31 % af dette forbrug.

Boligtype	Andel af totalt varmeforbrug	Andel af totalt vandforbrug
	[%]	[%]
Enfamiliehus, 1 husstand *	20,5	31,0
Boligblok, 214 husstande	28,0	-
Tæt/lavt, 185 husstande	33,0	-
Boligblok, 71 husstande	34,0	-
Boligblok, 148 husstande	36,5	28,4
Enfamiliehus, 1 husstand	-	33,7
Middelværdi	30,4	31,0

* Der fremgår kun målinger for 6 måneder af rapporten [8].

Tabel 3.20: Varmtvandsforbrugets andel i forhold til henholdsvis det totale varmeforbrug, der dækker over energiforbrug til rumopvarmning og produktion af varmt brugsvand, og andelen i forhold til det totale vandforbrug i boligtyperne.

Nogle af resultaterne fra projektet af *Bøhm et al.* er oplistet i tabel 3.21, hvor der kort er nævnt hvilken bygningstype målingerne er udført i, samt antallet af husstande i hver type. Af rapporten fremgår det ikke hvor mange personer, der boede i hver husstand, og det er ikke muligt at skelne de enkelte husstandes varmtvandsforbrug fra hinanden, da varmtvandsforbruget i rapporten er opgjort, som et samlet forbrug for hele den pågældende bygningstype. Det har af denne grund kun været muligt, at fordele varmtvandsforbruget ligeligt mellem husstandene i de aktuelle boligtyper, da det også kun var for tre af bygningerne at boligarealet var opgivet. Med mindre andet er anført i tabel 3.21, er der taget en middelværdi af 12 målinger, svarende til én opgørelse af varmtvandsforbruget på månedsbasis, og angivet under betegnelsen *gennemsnit* i tabellen. Betegnelserne *laveste forbrug* og *højeste forbrug* er opgjort som henholdsvis det laveste og det højeste månedsforbrug af de 12 målinger ligeligt fordelt på husstandene i boligtypen.

3.4.2. Varmt brugsvand

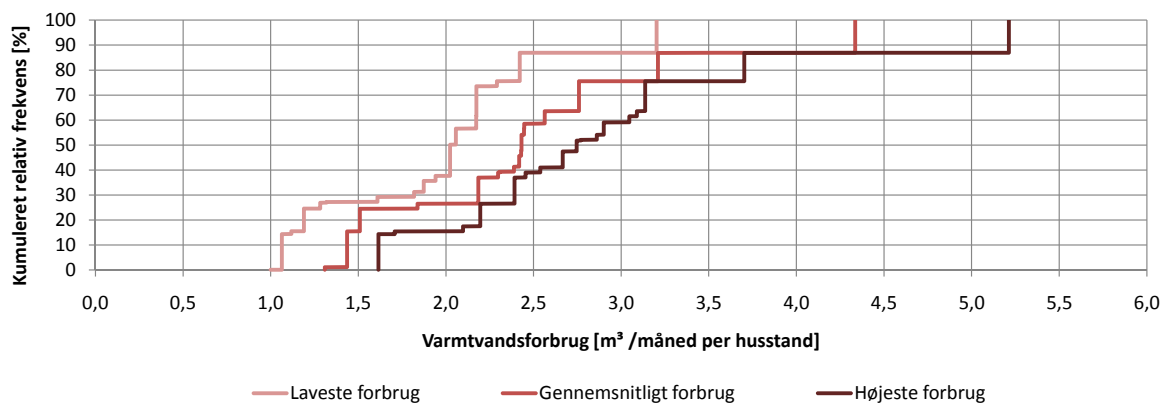
Boligtype	Gennemsnit	Laveste forbrug	Højeste forbrug
	$[m^3/måned/husstand]$	$[m^3/måned/husstand]$	$[m^3/måned/husstand]$
Boligblok, 18 husstande *	1,31	1,12	1,71
Boligblok, 234 husstande	1,44	1,06	1,62
Boligblok, 148 husstande	1,51	1,19	2,20
Boligblok, 33 husstande **	1,84	1,61	2,10
Boligblok, 171 husstande	2,18	2,02	2,39
Boligblok, 33 husstande	2,30	1,94	2,54
Tæt/lavt, 5 husstande	2,31	1,32	2,77
Boligblok, 33 husstande *	2,39	2,29	2,45
Boligblok, 71 husstande	2,42	1,87	2,75
Boligblok, 33 husstande	2,43	1,82	2,86
Tæt/lavt, 105 husstande	2,43	2,06	2,67
Boligblok, 39 husstande	2,45	1,28	3,05
Boligblok, 33 husstande	2,45	2,02	3,09
Enfamiliehus, 1 husstand *	2,53	2,25	3,05
Boligblok, 81 husstande	2,56	2,17	2,90
Boligblok, 196 husstande *	2,76	2,17	3,14
Tæt/lavt, 185 husstande	3,21	2,42	3,70
Enfamiliehus, 1 husstand	3,54	1,00	4,50
Boligblok, 214 husstande	4,34	3,20	5,21

* Der fremgår kun målinger for 6 måneder af rapporten [8].

** Der fremgår kun målinger for 11 måneder af rapporten [8].

Tabel 3.21: Varmtvandsforbrug for forskellige boligtyper, hvor det totale månedsforbrug er delt ud på antallet af husstande.

Ud fra dataet, der er angivet i tabellen ovenfor, er der optegnet en kumulativ fordeling af det månedlige varmtvandsforbrug fordelt ligeligt mellem husstandene i de aktuelle boligtyper. Optegnelsen er foretaget ud fra de samme tre betegnelser, der er anvendt i tabellen og kan ses på figur 3.24. Det er værd at bemærke, at de resultater, der er fremkommet på baggrund af en sammenskrivning af målingerne fra *Bøhm et al.*, ikke gengiver forskellene mellem enkelte husstande, men det er en middelbetragtning af forbruget i boligtyperne. Med dette menes, at den største forskel i varmtvandsforbruget, der fremgår af både tabel 3.21 og figur 3.24 er med omkring en faktor 5,2, men hvis eksakte varmtvandsforbrug for de enkelte lejligheder kunne sammenholdes, ville forskellen mellem det laveste og det højeste forbrug kunne være betragteligt større.



Figur 3.24: Kumulativ fordeling for det månedlige varmtvandsforbrug.

3.4.3 Diskussion

Igennem studiet af tilgængeligt litteratur i forbindelse med det samlede vandforbrug og forbruget af varmt brugsvand, er der opnået kendskab til størrelse og variation af begge vandforbrug. Ydermere er det studeret hvilke forhold, der kan have betydning for størrelsen af vandforbruget, samt hvilken adfærd der er i boliger omkring brugen af vand.

Ved en gennemgang af en større mængde forbrugsdata har *Miljøstyrelsen et al.* vist, at der er store variationer mellem det samlede vandforbrug i parcel- og rækkehuse sammenholdt med lejligheder, og de anbefalede derfor at behandle de to forbrug separat. Mængden af målinger og undersøgelser omhandlende størrelsen på varmtvandsforbruget er yderst begrænset og de fleste undersøgelser på området vurderer hvor meget energi, der anvendes til at opvarme det varme brugsvand. Igennem flere undersøgelser blev det erfaret at varmtvandsforbruget udgør omkring 1/3 af det samlede vandforbrug, og mængden af energi, der anvendes til opvarmning af det varme brugsvand, udgør i undersøgelser imellem 25 – 45 % af det samlede varmeforbrug med et gennemsnit omkring 30 %.

Gennem behandling af forbrugsdata fra en stor mængde husstande har *Jesper Ole Jensen* påvist, at der imellem nogle husstande er en faktor syv til forskel på det samlede vandforbrug. Nogle af årsagerne til forskellene kan findes blandt beboernes vaner i forbindelse med tøjvask, madlavning og badning mv. I forbindelse med tøjvask bevidner *Jesper Ole Jensens* undersøgelse, at behovet blandt husstandene er vidt forskellige, hvor nogle vasker et par enkelte maskiner med over en uges mellemrum, er der eksempler på andre, der vasker op mod 16 gange ugentligt. Ligeledes fandt *Kirsten Gram-Hanssen* inden for samme emne frem til, at det ifølge flere beboere er udviklingen i samfundets normer med hensyn til hygiejne og status der er skyld i at mængden af vasketøj er øget.

Hvad angår beboernes personlige hygiejne i form af badning, viser undersøgelser ligeledes forskelle blandt husstandene, hvor det er antallet af bade og varigheden af disse, der varierer. En undersøgelse fra *Kirsten Gram-Hanssen* viste, at der var en tendens til hvor mange ugentlige bade beboerne tog, afhængig af om husstanden var blandt den laveste eller højeste tredjedel. Hvad angår varigheden af badene, viser *Jesper Ole Jensens* undersøgelse, at der i nogle husstande blev taget bade med nogle få minutters varighed, mens badenes varighed i

andre husstande kunne være op mod en halv time eller længere. Det tyder på, at beboernes holdning til ressourceforbrug kan være meget forskellig blandt enkelte husstande. Som det viste sig gældende for tøjvask, er der indikationer på at samfundets normer er skyld i at hyppigheden for beboernes bade er øget i forhold til tidligere. Påvirkning fra samfundet med hensyn til hygiejne og status virker meget sandsynlig idet der for de fleste vedkommendes tilfælde ikke er et reelt behov for denne type adfærd, så længe ens job ikke kræver hårdt fysisk arbejde.

Metoder til analysering af data

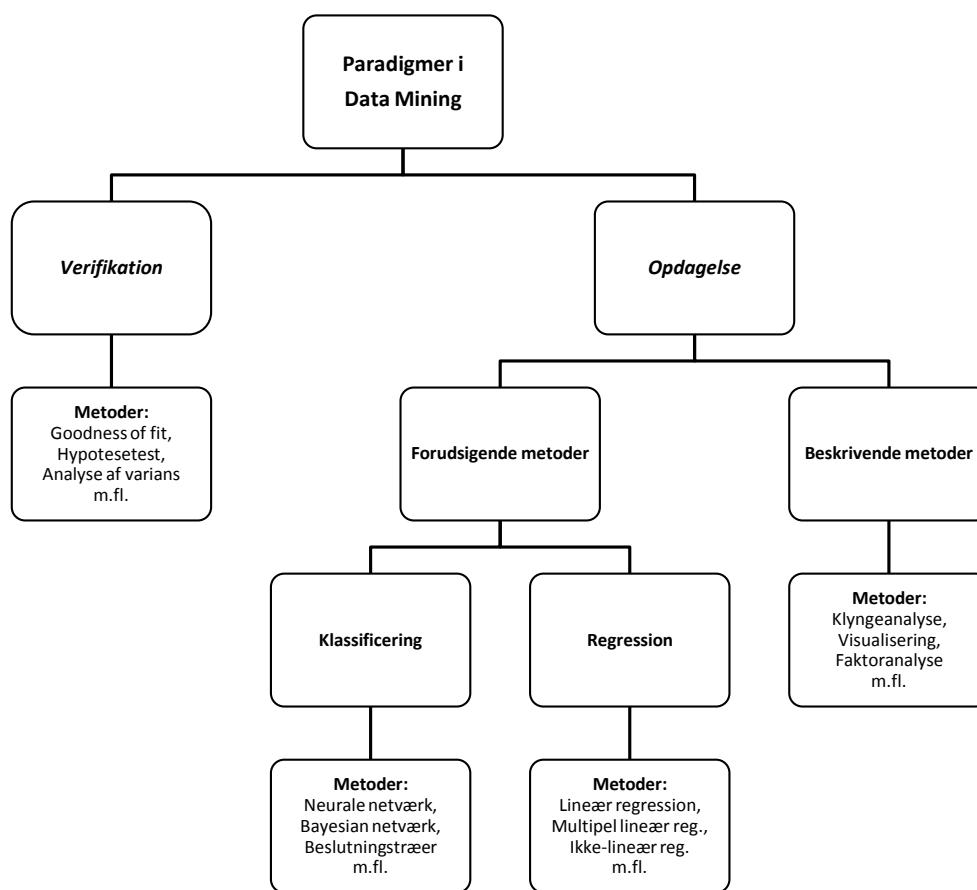
I afsnit 3.1 i litteraturstudiet er der gennemgået en række artikler, hvori der er anvendt forskellige metoder til analysering af data. Blandt disse metoder kan nævnes faktor-, komponent-, klyngeanalyse samt klassificeringstræ, som alle er en del af metoden kaldet *data mining* [84]. Data mining er overordnet set en metode, hvor der søges efter nye brugbare strukturer, mønstre eller ukendte sammenhænge ved at sortere i store mængder af data. Denne metode anvendes blandt andet til markedsanalyser, inden for det finansielle område og videnskaben. Data mining opstod som en konsekvens af, at data blev målt og opbevaret med så stor hast, at ingen havde mulighed for at analysere disse store mængder af data.

Der kan overordnet set skelnes mellem to typer data mining; *verifikation* og *opdagelse*, hvilket er skitseret på figur 4.1. Metoderne, der hører under *verifikation*, har til formål at verificere data ud fra en opstillet hypotese. Det sker på baggrund af nogle af statistiske metoder som eksempelvis *hypotesetest*, *goodness of fit*, *test af varians* m.fl. Hvad angår metoderne under *opdagelse*, er målet at opdage nye sammenhænge i data. De fleste metoder i typen *opdagelse* i data mining er baseret på induktiv læring¹, hvor en model ofte opbygges på baggrund af en stor datamængde og derefter valideres af testdata. Der er flere mulige fremgangsmåder inden for denne del af data mining, hvor der ofte skelnes mellem to undergrupper; *forudsigende* og *beskrivende metoder*, jf. figur 4.1.

De *forudsigende metoder* forsøger at finde frem til en model, hvorpå en ny og ukendt sammenhæng kan beskrives i en eksisterende datamængde, samt at kunne estimere en given variabels størrelse. Ofte opdeles disse metoder i yderligere to undergrupper bestående af metoder inden for *klassificering* og *regression*. Klassificeringsmetoderne anvendes til at forudsige, inden for hvilken kategori eller klasse en given variabel kan placeres. Blandt disse metoder hører eksempelvis *neurale netværk*, *beslutningstræer*, *bayesiansk netværk* m.fl. Regressionsmetoderne anvendes til at forudsige størrelsen af en given variabel angivet ved en værdi og ikke en klasse. Der skelnes ofte mellem *lineær regression*, *multipel lineær regression*, *ikke-lineær regression* m.fl. [84, 73]

De *beskrivende metoder* har til formål at analysere data, og finde en forståelse i hvordan det underliggende data fungerer. Eksempler på denne type metoder kan være *faktor-*, *klyngeanalyse*, *visualisering* m.fl. [84]

¹Læring hvor sammenhænge erkendes ud fra eksperimenter og forsøg.



Figur 4.1: Taksonomi for data mining [84].

4.1 Introduktion af metoder

Anvendelse af metoder som faktor-, komponent-, klyngeanalyse samt beslutningstræ er relativt nyt inden for området omhandlende brugeradfærd i boliger, men ikke desto mindre er de særdeles relevante i henhold til det nærværende projekts overordnede mål. På baggrund af dette vil der foretages en sammenligning af de førnævnte analysemetoder i det efterfølgende. Metoderne sammenholdes med fordele og ulemper, således det kan klarlægges, hvilken metode der er mest anvendelig i henhold til projektets mål.

Faktor- og komponentanalyse

I afsnit 3.1 i litteraturstudiet er der omtalt nogle artikler fra *Barr et al.* [4] samt *Raaij et al.* [103] omhandlende brugeradfærd i boliger. I disse artikler er der anvendt henholdsvis faktor- og komponentanalyse. Faktoranalyse er en generel betegnelse for en række forskellige analysemetoder, der, sammen med komponentanalyse, befinder sig blandt de *beskrivende metoder* inden for data mining, jf. figur 4.1. I appendiks F fremgår en uddybende teoretisk beskrivelse af de nævnte analysemetoder, men disse introduceres kort i det efterfølgende.

Analysemetoderne, der går under betegnelsen faktoranalyse, har til formål at klarlægge uobserverede strukturer mellem nogle observerede variable [22]. Ofte anvendes disse analysemetoder i forbindelse med spørgeskemaundersøgelser, hvor det undersøges, hvorvidt der kan

identificeres en struktur blandt besvarelsene af spørgsmålene, der ikke direkte er registreret, men muligvis eksisterer [9]. Metoden til undersøgelse af de uobserverede strukturer gør, at faktoranalyse generelt kan opdeles i to underkategorier; eksplorativ og konfirmativ faktoranalyse, hvor de til trods for opdelingen har det samme overordnede mål.

Eksplorativ faktoranalyse dækker over en række metoder, der har til formål at klarlægge antallet af underliggende faktorer i et datasæt, samt at estimere i hvor høj grad disse faktorer påvirker besvarelsene af en spørgeskemaundersøgelse, såfremt dette er datagrundlaget. Et kendetegn ved de eksplorative faktoranalyser er, at det er det observerede data, der danner grundlaget for analysen af den underliggende struktur. Det vil sige, at der i eksplorative faktoranalyser arbejdes ud fra nogle data og frem mod en model, hvilket er modsat de konfirmative faktoranalyser. [17]

Konfirmativ faktoranalyse tager udgangspunkt i en hypotese, der opstilles på forhånd, om antallet af faktorer, og ofte også hvilke besvarelser der påvirkes af hvilke faktorer [37]. Modsat eksplorativ faktoranalyse så arbejdes der ved konfirmativ faktoranalyse fra en model mod nogle data. Derved kan data analyseres, for at be-/afkræfte hvorvidt en given model/hypotese kan identificeres i de aktuelle data.

Hovedformålet med komponentanalyser er eksempelvis at samle spørgsmålene i spørgeskemabesvarelsene til et færre antal komponenter, der tager højde for hovedparten af den variation, der er mellem besvarelsene af disse spørgsmål [22]. Ofte forveksles komponentanalyse med eksplorativ faktoranalyse grundet en lignende fremgangsmåde [22]. Det bør imidlertid bemærkes, at de to analysemetoder har forskellige hovedformål; eksplorativ faktoranalyse har til formål at „*udforske*“ data, mens komponentanalyse har til formål at *reducere* data.

Klyngeanalyse

I afsnit 3.1 i litteraturstudiet er der omtalt nogle artikler fra *Barr et al.* [4] samt *Raaij et al.* [103] omhandlende brugeradfærd i boliger, hvor der i disse artikler er anvendt klyngeanalyse. Klyngeanalyse er en generel betegnelse for en række forskellige analysemetoder, der befinder sig blandt de *beskrivende metoder* inden for data mining, jf. figur 4.1 [85]. De forskellige analysemetoder har et fælles mål; at identificere hvilke elementer i et datasæt (eksempelvis personer fra en spørgeskemaundersøgelse), der har nogle fælles karakteristika for efterfølgende at gruppere elementerne i klynger. Dette sker således, at elementerne inden for hver enkelt klynge har flest mulige fællestræk, og således at forskellene mellem klyngerne er størst mulig [58]. Grupperingen sker gennem en række matematiske procedurer, hvor der i litteraturen ofte skelnes mellem tre forskellige fremgangsmåder, og derfor tre forskellige typer klyngeanalyse; hierarkisk, ikke-hierarkisk og two-step klyngeanalyse [58, 85].

Hierarkisk klyngeanalyse er en betegnelse for de metoder, hvor grupperingen af elementer foregår skridt for skridt, og hvor grupperede elementer ikke kan skifte til en anden klynge i et senere skridt, selvom dette muligvis ville være hensigtsmæssigt [58].

Ikke-hierarkisk klyngeanalyse forløber således, at det på forhånd er valgt hvor mange klynger analysen skal ende op med. Det første skridt i analysen går ud på, at vurdere hvor „tyngdepunktet“ for de enkelte klynger er placeret, hvilket ofte sker ud fra et subjektivt skøn. [58]

I en two-step klyngeanalyse anvendes der, som navnet antyder, både en hierarkisk og en ikke-hierarkisk klyngeanalyse. Indledningsvis udføres en hierarkisk klyngeanalyse, hvilket

danner grundlaget til vurdering af hvor mange klynger, det vil være meningsfuldt at arbejde videre med. Når det er fastlagt, anvendes en ikke-hierarkisk klyngeanalyse til at gruppere elementerne i det aktuelle antal klynger. [58]

I appendiks G fremgår en uddybende teoretisk beskrivelse af klyngeanalyserne.

Beslutningstræer

I afsnit 3.1 i litteraturstudiet er der omtalt artikler fra *Zhun et al.* [108] og *Tso et al.* [102], hvor der med succes er anvendt beslutningstræer inden for emnet energiforbrug i boliger. Beslutningstræer betegner en samling af metoder der tilhører de *forudsigende metoder* inden for data mining, jf. figur 4.1. Disse metoder anvendes til at søge efter mønstre og sammenhænge i både små og store datamængder, for efterfølgende at visualisere sammenhængen grafisk. De forskellige analysemetoder har et fælles mål; at udvikle et beslutningstræ ud fra en identificering af hvilke variable i et datasæt, der kan adskille elementerne i separate grupper, som karakteriserer ét bestemt forhold for disse. Et eksempel på det foregående kan være at identificere hvilke spørgsmål i en spørgeskemaundersøgelse, der kan adskille respondenterne i separate grupper, som er karakteriseret ved respondenterne elforbrug. Grupperingen sker gennem en række matematiske procedurer, der medfører, at overordnet set skelnes mellem to typer beslutningstræer; regressionstræer og klassificeringstræer [84].

Regressionstræer er udviklet på baggrund af en række metoder, der kan anvendes til estimering af hvilket forhold, der kan forventes for et element. Sammenholdt med det foregående eksempel kan et regressionstræ estimere en numerisk værdi for hvilket elforbrug, en respondent kan forventes at have, på baggrund af de besvarelser denne har angivet i spørgeskemaet.

Klassificeringstræer kan anvendes til at estimere hvilken kategori et element kan forventes at tilhøre, hvor de mulige kategorier er fastlagt inden udviklingen af klassificeringstræet. I det tidligere omtalte eksempel kan det illustreres ved, at klassificeringstræet estimere hvilken kategori for elforbruget en respondent kan forventes at have, på baggrund af de besvarelser denne har angivet i spørgeskemaet.

4.2 Diskussion af metodernes anvendelighed

De foregående beskrivelser introducerer hovedformålene med de udvalgte metoder til analysering af data. Metoderne har hver især en række fordele og ulemper, som sætter begrænsninger for hvilke områder de er anvendelige indenfor. I dette afsnit sammenholdes metodernes anvendelighed med det primære mål for dette projekt, der er beskrevet i afsnit 2.5.

Beskrivelsen af henholdsvis faktor-, komponent- og klyngeanalyse angiver, at disse metoder tilhører de *beskrivende analysemetoder* inden for data mining. De beskrivende analysemetoder kan anvendes, såfremt der skal foretages nogle uddybende undersøgelser af beboere og deres adfærd. Artiklerne, der er beskrevet i afsnit 3.1 i litteraturstudiet, er eksempler på, hvorledes disse metoder kan anvendes i praksis. En af fordelene ved faktoranalyser er, at de kan anvendes til at gruppere en række forskellige adfærd hos beboere ud fra nogle underliggende faktorer. Det resulterer i at kompleksiteten af et datasæt kan reduceres, når dette skal analyseres. På baggrund af der eksisterer forskellige typer faktoranalyser, er det muligt både at „udforske“ eksisterende data eller at afprøve en hypotese om nogle sammenhænge i et datasæt. Komponentanalyser kan anvendes til datareduktion, eksempelvis hvis

besvarelsesne af en omfattende spørgeskemaundersøgelse skal analyseres, hvilket derved kan lette det efterfølgende arbejde. Blandt ulemperne ved faktor- og komponentanalyser er, at der skal foretages flere subjektive valg af metoder undervejs. Disse metoder har, som det fremgår af appendiks F, betydelig indflydelse på resultaterne af analyserne. Desuden kan disse analysemetoder være vanskelige at anvende direkte til estimering af energiforbrug, da de, som tidligere beskrevet, tilhører de beskrivende metoder inden for data mining. På denne baggrund er analysemetodernes fordele at finde ved beskrivelsen af data. I tabel 4.1 er de nævnte fordele og ulemper ved faktor- og komponentanalyser oplistet.

Faktor- og komponentanalyse	
Fordele	Ulemper
• Kan reducere kompleksiteten af et datasæt. • Kan skabe overblik over nogle „grundlæggende“ adfærd eller sammenhænge mellem besvarelser (faktoranalyse). • Kan samle information fra flere variable under ét komponent (komponentanalyse). • Der er mulighed for at „udforske“ eksisterende data (eksplorativ faktoranalyse) eller afprøve en hypotese på eksisterende data (konfirmativ faktoranalyse).	• Der foretages flere subjektive valg af metoder under udførelsen af analysen (eksempelvis ved vurdering af antal faktorer/komponenter, estimering af faktorladninger og valg af rotationsmetoder, jf. appendiks F) • Fortolkningen af resultaterne kan blive subjektivt præget ud fra hvilken titel faktorerne/komponenterne tildeles. • Kan være vanskelige at anvende til estimering af energiforbrug.

Tabel 4.1: Tabel over nogle af de fordele og ulemper der er ved faktor- og komponentanalyser.

Ved klyngeanalyser skal der, som det ligeledes er tilfældet ved faktor- og komponentanalyse, foretages flere subjektive valg af metoder under udførelsen af analysen. Det medfører, at resultaterne kan blive påvirket af disse subjektive valg. Blandt fordelene ved klyngeanalyser er, at de i nogle tilfælde kan inddеле respondenterne i et datasæt i tydeligt adskilte klynger, som kan karakteriseres ved nogle bestemte kendetegn. Denne fordel kan imidlertid ligeledes anses som en ulempe, hvis der på baggrund af datagrundlaget ikke kan identificeres nogen tydeligt adskilte klynger. I flere tilfælde kan der udelukkende identificeres nogle klynger, hvis kendetegn er sammenfaldende og derfor vanskelige at tolke. Mulighederne for at anvende klyngeanalyse i forbindelse med estimering af energiforbrug er imidlertid større end for faktor- og komponentanalyse. I tabel 4.2 er de nævnte fordele og ulemper ved klyngeanalyser oplistet.

Klyngeanalyse	
Fordele	Ulemper
• Metoden udvikle klynger direkte på baggrund af data. • Der er mulighed for at identificere tydeligt adskilte klynger, som kan adskilles fra hinanden ved nogle bestemte kendetegn. • Metoden kan anvendes til estimering af energiforbrug.	• Der foretages subjektive valg af metoder undervejs (eksempelvis ved valg af lighedsmål og grupperingsmetode, jf. appendiks G). • Der er risiko for, at de enkelte klynger kan være sammenfaldende og derved være vanskelige at adskille på baggrund af kendetegnene. • Der skal gennemføres en længere beregningsprocedure for at estimere et energiforbrug.

Tabel 4.2: Tabel over nogle af de fordele og ulemper der er ved klyngeanalyse.

Beskrivelsen af klassificerings- og regressionstræer angiver, at disse metoder tilhører de forudsigende analysemetoder inden for data mining. Blandt fordelene ved disse metoder er, at de udvikles på baggrund af et tildelt datasæt, hvorefter klassificerings- og regressionstræet kan anvendes til estimering af energiforbrug. Afhængigt af datagrundlaget er der mulighed for at udvikle et simpelt og let overskueligt klassificerings- og regressionstræ. Modsat er der risiko for, at klassificerings- og regressionstræet i nogle tilfælde kan blive stort og komplekst. Der er oplistet fordele og ulemper ved disse metoder i tabel 4.3.

Klassificerings- og regressionstræ	
Fordele	Ulemper
• Der kan udvikles et klassificerings- eller regressionstræ direkte på baggrund af data (informationer, målinger eller spørgeskemabesvarelser). • Metoderne kan give overskuelige og let anvendelige resultater. • Metoderne kan direkte anvendes til estimering af energiforbrug i form af enten en kategori (klassificeringstræ) eller ved en numerisk størrelse (regressionstræ).	• Det foretages subjektive valg af metoder undervejs (eksempelvis ved valg af algoritme og reducere af beslutningstræets størrelse, jf. kapitel 7). • Der er risiko for, at klassificerings- og regressionstræer kan blive store og uoverskuelige.

Tabel 4.3: Tabel over nogle af de fordele og ulemper der er ved klassificerings- og regressionstræer.

Det bør bemærkes, at der generelt for samtlige metoder til analysering af data ikke er nogen garanti for, at der opnås nogle anvendelige resultater, da anvendeligheden af metoderne afhænger af datagrundlaget.

På baggrund af de foregående beskrivelser anses metoden med beslutningstræer at være den bedst anvendelige i forbindelse med at nå projektet hovedmål. Det skyldes, at denne metode tilhører den samling af metoder inden for data mining, hvis formål er, at forudsige eller estimere et forhold. I nærværende projekt er formålet, at udvikle en model på baggrund af målinger af energiforbrug i boliger samt ved kendskab til forskellige karakteristika for beboerne og deres adfærd. Metoderne med beslutningstræer åbner mulighed for, at produktet kan være anvendelig for en bred målgruppe. Det skyldes, at der ikke skal gennemføres en større matematisk procedure for at estimere et forbrug, som det er tilfældet ved klyngeanalyser. Derimod er der mulighed for at udvikle et let anvendeligt beslutningstræ. På baggrund af disse iagttagelser introduceres den grundlæggende teori om beslutningstræer i det efterfølgende.

4.3 Beslutningstræ

I afsnit 3.1 i litteraturstudiet er anvendelsen af et beslutningstræ beskrevet i to forskellige analyser. På baggrund af, at metoden er anvendt med succes inden for emnet *energiforbrug i boliger*, foretages en vurdering af denne metodes anvendelighed i forhold til det nærværende projekts overordnede mål, jf. afsnit 2.5. Idet metoden stadig er ny inden for dette fagområde, vil denne analysemetode blive beskrevet nærmere, med henblik på at beskrive den overordnede fremgangsmåde for metoden. Det er således muligt at klarlægge fordele, ulemper og anvendelsesmuligheder for beslutningstræer. Denne beskrivelse skal danne grundlag for vurdering af, hvorvidt metoden er brugbar i nærværende projekt.

Beslutningstræet betegner en metode, hvor der, som for de øvrige metoder inden for *opdagelse* i data mining, jf. figur 4.1 på side 87, søges efter mønstre og sammenhænge i både små og store datamængder. Beslutningstræet er en del af det begreb, der betegnes *superviserede metoder* inden for data mining. Det betyder, at det er en metode, hvori der findes en sammenhæng mellem flere input attributter (uafhængige variable) og en målattribut (afhængig variabel, klasse), og hvor sammenhængen præsenteres via en opstillet model; beslutningstræet. Begreberne *elementer*, *attributter* og *klasse* anvendes i den efterfølgende beskrivelse, hvorfor de overordnet defineres. Alle tre begreber knytter sig til datamængden som klassificeringstræet udvikles på baggrund af, og de to sidstnævnte begreber anvendes ligeledes i klassificeringstræet. Elementerne beskriver antallet af datamængder i datasamlingen, jf. figur 4.2. Attributter er de oplistede oplysninger, der har en sammenhæng med klassen, i en datasamling, dvs. de spørgsmål der stilles, eller variable der undersøges. Attributterne anvendes, ved udvikling af klassificeringstræet, til at inddele data i undergrupper. En af attributterne tildeles titlen *klasse*, og er den attribut, der søges estimeret. En klasse kan for eksempel være husstandens samlede energiforbrug, forbruget af varmt brugsvand eller lignende. Klassen estimeres ved hjælp af de øvrige attributter i datamængden. Fælles for attributter og klasser i beslutningstræer er, at de kan have enten numeriske og kategoriske værdier, hvor dette er defineret i tabel 4.4.

4.3. Beslutningstræ

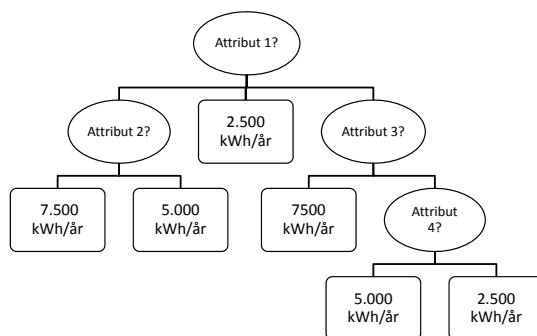
		Attributter					
		Spørgsmål 1	Spørgsmål 2	Spørgsmål 3	Spørgsmål 4	Spørgsmål n	Klasse
Elementer	Person A	...	...	...	...	...	...
	Person B	...	...	...	...	...	...
	Person C	...	...	...	...	...	...
	Person D	...	...	...	...	...	...
	...	...	...	...	...	...	...

Figur 4.2: Illustrering af elementer, attributter og klassen i en datamsamling.

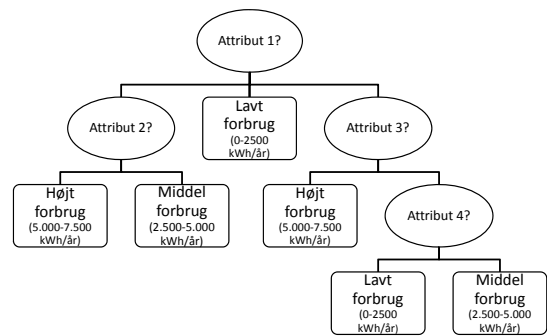
Attribut/klasse	Beskrivelse
Numerisk	Reelle talværdier, f.eks. U-værdi, temperaturer, infiltration, boligareal og lignende.
Kategorisk	Ingen talværdier, men værdier i form af begreberne høj/lav, bred/smål, parcelhuse/lejlighed, on/off osv. eller i form af nogle prædefinerede kategorier.

Tabel 4.4: Beskrivelse af forskellen mellem numeriske og kategoriske attributter, der indgår i et beslutningstræ.

Der findes to typer af beslutningstræer; regressions- og klassificeringstræer [84]. Den overordnede forskel mellem de to typer af beslutningstræer er, at der i regressionstræet estimeres en numerisk værdi for klassen, hvorimod der i klassificeringstræet estimeres en prædefineret kategoriske klasser [84]. I figur 4.3 og 4.4 fremgår der eksempler på henholdsvis et regressions-træ og et klassificeringstræ. Begge figurer illustrerer en model, hvor et energiforbrug forsøges estimeret ud fra nogle forskellige attributter, der har betydning for det pågældende forbrug. Attributterne i de to typer af beslutningstræer kan være de samme, da det der overordnet adskiller de to træer er formen på outputtet.



Figur 4.3: Visualisering af et regressionstræ.



Figur 4.4: Visualisering af et klassificeringstræ.

Begge typer af beslutningstræer er udelukkende gældende for det data, de er udviklet på baggrund af. Det betyder, at de udviklede beslutningstræer ikke kan generaliseres for et større område, eksempelvis et helt land, hvis data kun er indsamlet i et lokalt område. Hvis

beslutningstræet skal være gældende for større områder, er det nødvendigt, at indsamle tilstrækkeligt samt repræsentativt data. Selvom beslutningstræerne er udviklet på baggrund af lokalt data, kan dets egnethed godt afprøves på data fra andre områder. Det skal dog holdes for øje, afhængig af udvalget af attributter, at de givne outputs ikke nødvendigvis bliver korrekte.

Indsamlingen af data til udviklingen af et beslutningstræ kan ske på flere forskellige måder. Den indsamlede datamængde kan indeholde vidt forskellige oplysninger, afhængigt af hvilken type beslutningstræ der skal udvikles. Ved udvikling af beslutningstræer er der ingen krav til en minimums mængde af data. Det kræves dog en større datamængde for at kunne varetage en stor variation i attributter. Der opstår nogle problemstillinger i udviklingen af beslutningstræet, hvis antallet af informationer i datamængden, dvs. mængden af attributter, er for omfangsrig. Udfordringerne kan eksempelvis være, at der ved mange attributter er risiko for, at de udviklede beslutningstræer bliver store og uoverskuelige. Derudover vil indsamling og bearbejdning af en stor datamængde være tidsmæssigt og økonomisk krævende. Disse problemstillinger kan eventuelt imødegås ved at udføre en indledende analyse af de variable, eksempelvis en komponent- eller klyngeanalyse. Derigennem opnås en reducere af mulige variable og det kan dermed gøre træet mindre omstændeligt og anvendelig i et større perspektiv.

Det overordnede mål for nærværende projekt er, at inddele beboere i parcelhuse i nogle prædefinerede grupper for deres energiforbrug til rumopvarmning, varmt brugsvand og el ud fra viden omkring flere attributter. Grundet den store variation i beboernes forbrug er der dermed en risiko for, hvis klassen er numerisk, at antallet af opdelinger bliver uoverskueligt stort. Det er af denne årsag mest oplagt at anvende et klassificeringstræ, hvorfor der udelukkende fokuseres på denne type beslutningstræ i det efterfølgende.

4.3.1 Klassificeringstræ

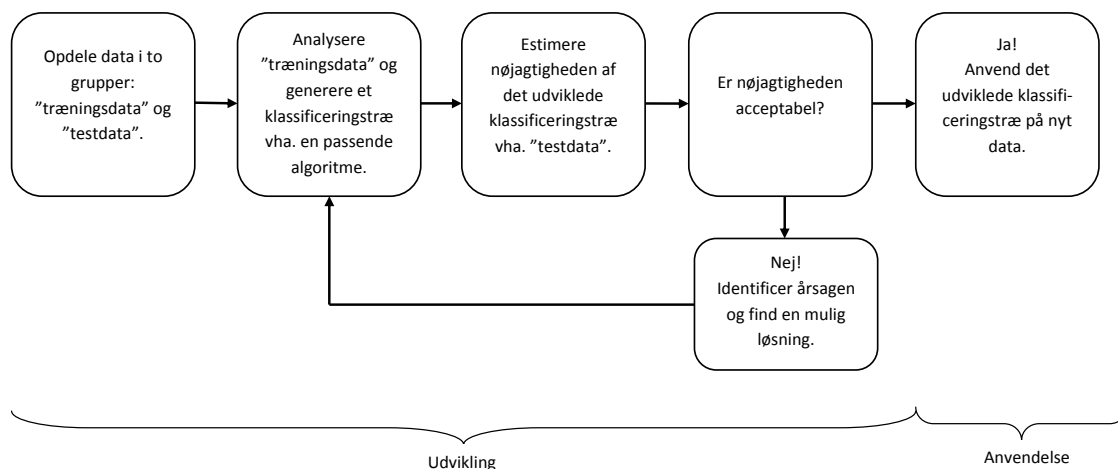
Klassificeringstræet kan anvendes inden for fagområderne: den finansielle verden, marketing, ingeniørvidenskab, botanik, datalogi og medicin [54]. Blandt årsagerne til den udbredte anvendelse af klassificeringstræer er, at det er et godt værktøj til at visualisere en løsning på et givet problem og dennes konsekvens. Eksempelvis kan metoden anvendes inden for lægevidenskab, hvor der, på basis af viden om patienten og dennes symptomer, kan skabes et overskueligt overblik over mulige tests og diagnoser.

Et klassificeringstræ fungerer som et hierarki, hvilket vil sige, at denne data mining metode inddeler data i grupper ved at vurdere én variabel ad gangen. Det betyder, at den endelige inddeling er afhængig af alle de forudgående inddelinger. Når et nyt datasæt anvendes i et klassificeringstræ, kan små ændringer i inputdata have stor betydning for, hvilken klasse beboerne tildeles, når de vurderes vha. klassificeringstræet. Foruden den hierarkiske opbygning kendetegnes klassificeringstræer ved, at de kan udføre inddelinger ud fra anvendelse af både numeriske og kategoriske attributter, jf. tabel 4.4. Efter udviklingen af klassificeringstræet kan den fundne sammenhæng, der beskrives visuelt igennem beslutningstræet, omformes til nogle regelsæt, beskrevet som tekst, for hvordan data skal inddeles.

Udviklingsprocessen for et klassificeringstræ kan opdeles i to dele. Første del er en udviklingsproces, hvori klassificeringstræet udvikles på baggrund af data og dets nøjagtighed testes og

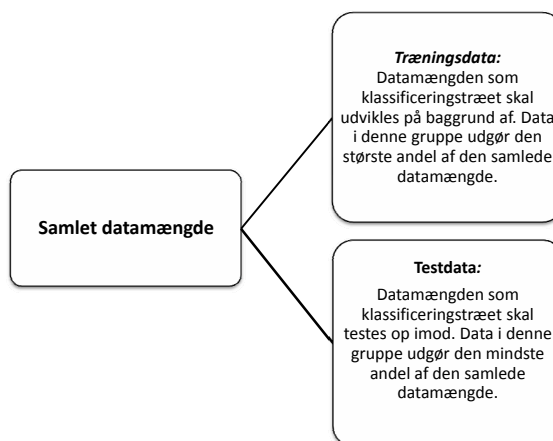
4.3.1. Klassificeringstræ

evalueres. Anden del i udviklingsprocessen er en anvendelsesdel, hvor det udviklede klassificeringstræ anvendes til et tiltænkt formål. Udviklingsprocessen er illustreret i figur 4.5.



Figur 4.5: Illustrering af udviklingsprocessen for et klassificeringstræ [108].

I den illustrerede udviklingsproces på figur 4.5 indledes delen *udvikling* med, at det indsamlede data skal gennemgås for fejl. Det kan eksempelvis være manglende besvarelser eller urealistisk målte værdier. Herefter inddeles dataet i et „træningsdata“ og et „testdata“, jf. figur 4.6. Træningsdata anvendes til at udvikle klassificeringstræet, dvs. det er den datamængde, der danner grundlaget for modellen. Det er derfor i dette data, at den største del af datamængden forekommer. Ved anvendelse af den største datamængde til at opbygge modellen, imødegås sandsynligheden for, at der tages højde for de fleste tilfælde af kombinationer i datamængden.

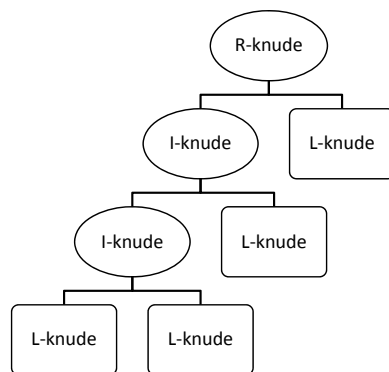


Figur 4.6: Illustrering af inddeling i „træningsdata“ og et „testdata“ [108].

Testdata udgør derfor en mindre del af den samlede datamængde. Før klassificeringstræet bliver udviklet, opstilles der et kriterium, for hvor godt klassificeringstræet skal kunne håndtere data, og placere dem i den korrekte prædefinerede klasse. Dette kriterium gælder både

for træningsdata og testdata, og anvendes som kriterium for hvornår klassificeringstræets nøjagtighed kan accepteres. Imødegår klassificeringstræet kriteriet, først ved træningsdata og efterfølgende ved anvendelse af testdata, som illustreret på figur 4.5, vil modellen blive accepteret. Efter denne inddelingen af data er fortaget, er målet, at udvikle et klassificeringstræ ved anvendelse af en algoritme.

Et klassificeringstræ er, som navnet indikerer, opbygget efter træets princip, og består af *root*, *branch* og *leaf*. Modsat træet i naturen vokser et klassificeringstræ dog i en nedadgående retning [54]. De overordnede objekter som et klassificeringstræ består af, er tre forskellige typer af knuder, og led der forbinder disse. I figur 4.7 er sammenhæng mellem knuderne illustreret, og tabel 4.5 angiver en udførlig beskrivelse af knudernes egenskaber.



Figur 4.7: Illustrering af hvilke typer af knuder der kan anvendes i et klassificeringstræ [84].

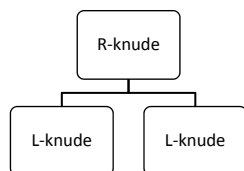
Knudetype	Engelsk navn	Beskrivelse
R-knode	Root node	R-knuden er klassificeringstræets startpunkt. I R-knuden udføres der en splittest på en attribut, hvorefter data inddeles i undergrupper. R-knuden er ikke afhængig af andre knuder, og har kun udadgående kontakt til enten en I- eller L-knode.
I-knode	Internal/test node	I-knuden figurerer, som navnet afslører, internt i træet. Ligesom for R-knuden udføres der splittests på attributter, hvorefter data inddeles i undergrupper. I-knuden er afhængig af R-knuden og eventuelt andre I-knuder. Den har udadgående kontakt med andre I- eller L-knuder.
L-knode	Leaf/decision node	L-knuden repræsenterer et output af en klassificering. Knuden tildeles en klasseinddeling, der angives på baggrund af hvilken klasseinddeling hovedparten af elementerne fra træningsdata i knuden har. Den kan ligeledes tildeles en sandsynlighedsvektor, der beskriver sandsynligheden for et udfald. Denne knude er afhængig af R- eller I-knuder og har ingen udadgående kontakt.
Led		Leddet er det, der skaber kontakten mellem de tre forskellige knuder i et klassificeringstræ.

Tabel 4.5: Beskrivelse af de knuder der indgår i et klassificeringstræ [84, 54].

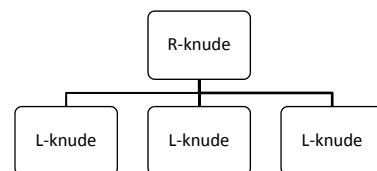
Alle knudetyper, der er beskrevet i tabel 4.5, indgår altid i et klassificeringstræ. Den visuelle fremstilling kan opnå større overskuelighed ved at tildele forskellige geometriske udtryk eller forskellige farver til hver knudetype. Det kan eksempelvis være ved at angive R-knuden og I-knuder cirkulære figurer og L-knuder i firkantede eller lignende, eller forskellige farver for de pågældende knuder.

Udviklingen af et klassificeringstræ sker ved hjælp af en algoritme. Der findes flere forskellige algoritmer, der kan anvendes. Beskrivelsen af fremgangsmåden i dette afsnit tager udgangspunkt i algoritmen C4.5 [77]. Denne algoritme anvender tre forskellige mål under udviklingen af klassificeringstræet, som anvendes til at udvælge den attribut, der er bedst til at inddеле elementerne i de aktuelle klasseinddelinger. Ved udviklingen af klassificeringstræet udføres såkaldte splittests, der anvender begrebet *gain ratio* til udvælgelsen af den attribut, der er bedst til at inddеле elementerne i de aktuelle klasseinddelinger. Denne splittest udføres i R-knuden og I-knuderne, hvor data inddeles til enten nye I-knuder eller L-knuder, jf. figur 4.7. En splittest udføres som et spørgsmål, hvorefter data inddeles. Et eksempel herpå kan være: „Antal personer i boligen?“, til dette spørgsmål svares der med et helt tal mellem eksempelvis 1 – 5 personer. Hvert svar tildeles et led med kontakt til en L-knude. Andre spørgsmål kan være orienteret mod et „ja/nej“ svar, hvor spørgsmålet eksempelvis kan være: „Er U-værdien for ydervægge $< 0,11 \leq W/m^2K$ “. En splittest visualiseres i klassificeringstræet ved de forskellige knuder, hvor hvert led, der forbinder knuderne, tildeles en beskrivelse, der betegner de svarmuligheder, der er til splittesten.

Klassificeringstræet kan udvikles som enten et binært eller ikke-binært træ. Forskellen kendetegnes ved om data i R- og I-knuderne inddeles efter henholdsvis to eller flere led, hvor en illustration heraf fremgår af figur 4.8 og 4.9. Det er i denne forbindelse vigtigt at pointere, at der altid er flere løsninger for inddeling af data for et klassificeringstræ. Alle løsninger er rigtige, men den bedste løsning afhænger af hvilke kriterier, der er opstillet for klassificeringstræet, eksempelvis: mindste antal L-knuder (mindst muligt træ), mindste fejlmargen og lignende. [62]



Figur 4.8: Eksempel på et binært klassificeringstræ, hvor R-knuden har forbindelse til to L-knuder [84].



Figur 4.9: Eksempel på et ikke-binært klassificeringstræ, hvor R-knuden har forbindelse til flere end to L-knuder [84].

Beskrivelse af mulige klassificeringstræer og data til udvikling af disse

Det er i det foregående kapitel fastlagt, at metoden omhandlende klassificeringstræer skal anvendes til opstilling af en model. Denne model skal anvendes til estimering af energiforbrug, hvor der blandt andet kan inddrages information om beboernes adfærd, jf. problemformuleringen i afsnit 2.5. Formålet med nærværende kapitel er, at skabe et overblik over hvilke typer af klassificeringstræer der kan udvikles inden for det udvalgte emne, samt beskrive hvilke data der er anvendelige i udviklingsprocessen.

5.1 Beskrivelse af hvilke klassificeringstræer der kan udvikles

Inden for metoden omhandlende klassificeringstræer er der adskillige muligheder for hvilke type emner, der kan behandles og beskrives. Der er derfor ingen overordnet opskrift på, hvilken klasse samt hvilket data der er anvendeligt. Et andet forhold der er behov for at klarlægge, før indsamlingen af data samt udviklingen af klassificeringstræer påbegyndes, er hvilket formål klassificeringstræet har, og hvordan det er tiltænkt at skulle anvendes. Inden for området energiforbrug er der mulighed for, at klassificeringstræet kan anvendes både under designfasen af en bolig, eller det kan anvendes for den eksisterende boligmasse. I forhold til projektets formål, er løsningen om et klassificeringstræ i designfasen ikke optimal, idet det, ud fra tidligere vurderinger i dette projekt, ikke findes optimalt, at inddrage beboernes adfærd i designfasen. Årsagen til dette er, at der er risiko for at boligen, ved inddragelse af beboernes adfærd, bliver optimeret specifikt til en familie, og dermed er der risiko for, at boligens robusthed med hensyn til energiforbrug reduceres.

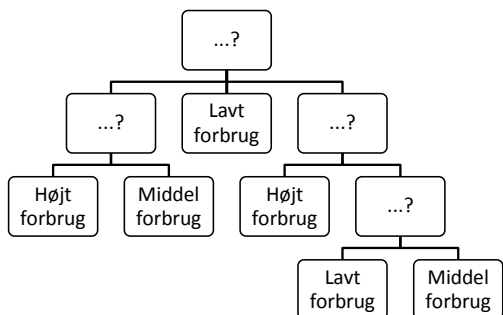
Den anden mulighed for klassificeringstræet er, at det anvendes i eksisterende boliger. Her er det tiltænkt at kunne anvendes til at estimere energiforbruget i boliger under hensyntagen til beboernes adfærd. Dette skaber en anvendelsesmulighed for klassificeringstræet i forbindelse med reducere af energiforbruget i den eksisterende boligmasse. Dermed kan det indgå som et virkemiddel til at nå regeringens mål om at reducere bruttoenergiforbruget med 4 % inden år 2020. Ud fra ovenstående vurdering er det valgt, at fokusere på at udvikle et klassificeringstræ, der kan anvendes på den eksisterende boligmasse.

Med udgangspunkt i litteraturstudiet i kapitel 3 er der, for nærværende projekt, mulighed for at udvikle to forskellige typer af klassificeringstræer, hvor de overordnede klasser er meget forskellige. I tabel 5.1 er det oplistet, hvilke typer klassificeringstræer der beskrives i nærværende afsnit.

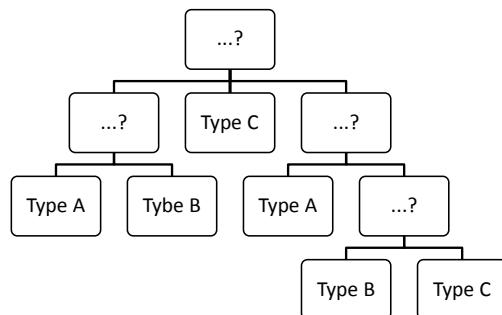
Mulige klassificeringstræer	
Modeltype 1:	Klassificeringstræ til estimering af energiforbrug – Tre separate klassificeringstræer – Ét samlet klassificeringstræ
Modeltype 2:	Klassificeringstræ til estimering af brugertype – Flere separate klassificeringstræer

Tabel 5.1: Beskrivelse af hvilke muligheder for udvikling af klassificeringstræer der gennemgås i nærværende afsnit.

Den ene modeltype klassificeringstræ kan anvendes til estimering af energiforbrug. For denne modeltype beskrives to muligheder, hvor den ene består af tre forskellige klassificeringstræer. Disse er tiltænkt at estimere henholdsvis energiforbruget til rumopvarmning, varmtvands- og elforbrug, hvor der inddrages oplysninger om blandt andet beboernes adfærd. Den anden mulighed for denne modeltype er udvikling af ét klassificeringstræ til estimering af boligens samlede energiforbrug, som illustreret på figur 5.1. I tabel 5.1 omtales en anden modeltype, der består af flere forskellige klassificeringstræer. Disse er hver tiltænkt at beskrive forskellige brugertyper med udgangspunkt i beboernes adfærd, som illustreret på figur 5.2. De kan eksempelvis beskrive beboernes vaner med hensyn til udluftning, justering af varme anlæg, solafskærmning etc. Begge typer klassificeringstræer beskrives nærmere i det efterfølgende.



Figur 5.1: Visualisering af modeltype 1 til estimering af energiforbrug på baggrund af viden om tekniske, demografisk og adfærdsrelaterede forhold.



Figur 5.2: Visualisering af modeltype 2 til estimering af prædefineret brugertype på baggrund af viden om tekniske, demografisk og adfærdsrelaterede forhold.

5.1.1 Beskrivelse af modeltype 1

Den førstnævnte type klassificeringstræ er tiltænkt, som beskrevet i det foregående, at beskrive sammenhængen mellem beboernes adfærd og energiforbruget i boliger. I dette afsnit

beskrives to forskellige muligheder for udvikling af klassificeringstræer. Den ene mulighed består af tre individuelle klassificeringstræer, hvor der udvikles et til henholdsvis energiforbrug til rumopvarmning, varmtvands- og elforbrug. Ved udvikling af tre forskellige klassificeringstræer er der forskellige faktorer, der har indflydelse på de tre enkelte energiforbrug, jf. kapitel 3. Gennemgående for alle tre klassificeringstræer er, at de attributter, der anvendes ved udviklingen af disse, ikke udelukkende knytter sig til beboernes adfærd. Det er nødvendigt, for at opnå det bedst mulige klassificeringstræ, at inddrage viden om tekniske, demografisk såvel som adfærdsrelaterede forhold for henholdsvis boligen og beboerne. Det skyldes, at det ud fra litteraturen er erfaret, at hverken de tekniske, demografiske eller adfærdsrelaterede forhold alene beskriver den fulde variation i energiforbruget i boliger.

En anden mulig løsning af den omtalte problemstilling er udvikling af ét samlet klassificeringstræ, hvor beboerne inddeles efter boligens samlede energiforbrug. Der er risiko for, ved udvikling af ét samlet klassificeringstræ, at nogle af de faktorer, der knytter sig til brugeradfærden inden for de enkelte forbrugsområder, vil forsvinde i baggrunden. Det skyldes, at disse faktorer, set i et større perspektiv, ikke nødvendigvis har lige så stor indflydelse på energiforbruget til rumopvarmning, varmtvands- og elforbruget, som eksempelvis konstruktionernes U-værdi har på energiforbruget til rumopvarmning. Det betyder, at der er risiko for, ved udvikling af ét samlet klassificeringstræ, at indsigten til sammenhængen mellem brugeradfærd og energiforbrug vil forsvinde i baggrunden. Idéen omkring ét samlet klassificeringstræ kan være en mulighed, når en større indsigt i emnet brugeradfærd er opnået.

5.1.2 Beskrivelse af modeltype 2

Hensigten med denne modeltype er at beskrive en samling af flere klassificeringstræer, der hver især beskriver nogle forskellige brugertyper, omhandlende beboernes vaner og adfærd. I forbindelse med energiforbrug til rumopvarmning kan det eksempelvis dreje sig om justering af solafskærmning, varmeanlæg, mekanisk ventilation samt beboernes udluftningsvaner. Disse enkeltstående klassificeringstræer skal udvikles på baggrund af viden om eksempelvis demografiske oplysninger om beboerne og adfærdsmæssige forhold, i form af eksempelvis reaktionsmønstre når boligen opfattes for varm eller kold og lignende. De endelige klasser, som data skal klassificeres efter, tiltænkes at blive beskrevet ved hjælp af en klyngeanalyse af det indsamlede data. Målet for klassificeringstræet er, at det ved optimal udvikling kan give indsigt i de overordnede forskelle mellem forskellige typer af beboere. Klassificeringstræerne kan anvendes ved simuleringsopgaver i eksempelvis BSim, for derigennem at tage højde for konsekvensen af varierende adfærd hos beboerne, når der simuleres med henblik på resultater for indeklima og energiforbrug til rumopvarmning. Der opstår imidlertid nogle problemstillinger, i forhold til hvordan data kan anvendes i et klassificeringstræ, og hvordan det kan anvendes i eksempelvis BSim. I BSim er præcisionen, med hensyn til angivelse af tidsskemaer og styring af anlæg, begrænset i forhold til de mange mulige variationer, der kan eksistere blandt beboere i boliger. Nogle af disse problemstillinger kan imødegås under udviklingen af brugertyperne, ved eksempelvis at tage højde for „begrænsningerne“ i programmet når brugertyperne udvikles. En anden løsning af problemstillingen kan være, at eksempelvis tidsperioderne i BSim udvikles til at kunne angives i mindre intervaller end de nuværende timeangivelser. Denne ændring er ikke nødvendig for samtlige forhold, men specielt ved angivelse af udluftning kan denne mulighed være fordelagtig i forhold til at vurdere brugeradfærdens betydning for indeklima og energiforbrug.

En anden problemstilling i forbindelse med denne model er, at programmer som BSim ikke anvendes i så stort et omfang i forbindelse med projektering af boliger. Det medfører, at klassificeringstræet til estimering af brugertype bliver mindre anvendeligt i praksis.

5.1.3 Udvælgelse af modeltype

Med udgangspunkt i den foregående beskrivelse fokuseres i det efterfølgende på modeltype 1, hvor der udvikles tre klassificeringstræer, som kan estimere en forbrugsstørrelse baggrund af blandt andet beboernes adfærd, som illustreret på figur 5.1. Under udviklingen af et klassificeringstræ viser de underliggende beregningerne, hvilke attributter der har størst betydning for energiforbruget til rumopvarmning, varmtvandsforbrug og elforbrug. Det vælges, at forberede udviklingen af et klassificeringstræ for hver af de tre udvalgte energiforbrug samt at eksemplificere udviklingen af et klassificeringstræ for ét af de omtalte forbrug. I det efterfølgende beskrives hvilken klasse data tildeles, samt hvilke attributter der kan indsamles data for i forbindelse med udviklingen af klassificeringstræerne. Attributterne, der skal anvendes til udviklingen af klassificeringstræerne, udvælges på baggrund af viden fra litteraturstudiet i kapitel 3. Derudover vurderes nogle af attributterne fra litteraturstudiet med forholdene i projektet fra *Fremtidens Parcelhuse* af Knudsen et al. [52], der samtidig danner grundlag for inspiration til nye attributter, som kan anvendes i klassificeringstræerne.

5.2 Definerer af klassificeringstræernes klasser

Ved udvikling af klassificeringstræer anvendes der, som gennemgået i afsnit 4.3, én klasse og flere forskellige attributter. Klassificeringstræer er kendetegnet ved, at deres klasse er kategorisk. Det betyder, at klassen kan angives som en skala, eksempelvis „højt forbrug“, „gennemsnitligt forbrug“ eller „lavt forbrug“, frem for at anvende den målte værdi for forbruget af varme, varmt brugsvand og el [54].

5.2.1 Vurdering af mulige klasser

Klassen i klassificeringstræet har stor betydning for hvilket data, der skal indsamles. Den har ligeledes indflydelse på, hvordan data tilhørende attributterne til klassificeringstræet skal undersøges og indsamles. Af denne årsag bør klassen for klassificeringstræer fastlægges, for at danne grundlag for vurdering af hvilke attributter der vil være optimale at anvende i udviklingsprocessen. Målet for hvad de tre klassificeringstræer skal beskrive, samt deres klasse er beskrevet nedenfor:

- Klassificeringstræ til estimering af energiforbruget til rumopvarmning i boligen (klassen) på baggrund af oplysninger om beboernes adfærd samt demografiske og tekniske oplysninger (attributterne).
- Klassificeringstræ til estimering af forbruget af varmt brugsvand (klassen) på baggrund af oplysninger om beboernes adfærd samt demografiske og tekniske oplysninger (attributterne).

- Klassificeringstræ til estimering af elforbruget (klassen) på baggrund af oplysninger om beboernes adfærd samt demografiske og tekniske oplysninger (attributterne).

Af den ovenstående beskrivelse fremgår det, at de forskellige klasser er fastlagt. Det skal imidlertid fastlægges hvilke enheder, der er mest optimale at anvende for klassen. Det skyldes, at disse enheder har indflydelse på hvorledes attributterne i klassificeringstræet skal opgøres. Beskrivelsen af anvendelige attributter og deres enhed fremgår i afsnit 5.3.2.

Energiforbruget i boliger opgøres i forskellige enheder, afhængig af hvilket forbrug der evalueres. Fælles for de fleste enheder ved opgørelsen af energi er, at de sjældent er optimale i forhold til beboernes forståelse for boligens energiforbrug. Denne problemstilling er blandt andet synliggjort i et projekt fra Storbritannien af *Gill et al.* [38]. Beboerne, i en af de involverede boliger, fik udskiftet en elmåler, der opgjorde forbruget i kWh , til en måler der angav forbruget i en økonomisk størrelse. Dette resulterede i, at beboerne fik en større forståelse for elforbruget. Denne ændring medførte, at beboerne efterfølgende var mere opmærksomme på, hvor ofte de anvendte elforbrugende apparater. Tilsvarende fik beboerne, ved observation af måleren, forståelse for hvor meget forskellige aktiviteter betød for størrelsen af deres regning. Dette aspekt, omhandlende beboernes forståelse for energiforbrug, er værd at have med i vurderingen af hvilken klasse, der skal anvendes i de tre tiltænkte klassificeringstræer.

Der er flere forskellige muligheder for at angive forbruget til rumopvarmning, varmt brugsvand og el på en mere håndgribelig måde. Anvendelige enheder for elforbruget kan være $kr./år$, $kr./måned$ eller $kr./person$ opgjort enten per år eller måned. Varmtvandsforbruget kan eksempelvis opgøres som $L/år$ eller $L/person\ per\ år$. Enheden for forbruget til rumopvarmning er mere kompliceret at beskrive ved hjælp af en let forståelig enhed. Årsagen til dette er, at forbruget beskrevet i forhold til et økonomisk aspekt er prismæssigt afhængig af hvilken type opvarmningssystem, der er tilgængeligt i boligen samt prisen for energikilden. En af fordelene, ved at angive forbruget i en enhed der er nemmere for beboerne at relatere sig til, er at modellen, ved en optimal udvikling, vil være anvendelig med henblik på at visualisere eventuelle besparelsesområder. Ulempen ved at anvende en enhed hvori der indgår økonomi er, at klassificeringstræerne ikke kan anvendes generelt, idet energipriserne varierer imellem landsdele, og afhænger af hvilket abonnement beboerne har. Det betyder, at klassificeringstræerne konstant skal opdateres. Enheder til forbruget for varmtvands- og elforbruget, opgjort ud fra antal beboere, er overvejet som en mulighed. Det skyldes, at der gennem litteraturstudiet i kapitel 3 er erfaret, at antallet af personer har stor betydning for forbrugets størrelse. Ulempen ved at opgøre forbruget per person er, at det er vanskeligt at prædefinere nogle klasser, der kan anvendes til at beskrive forbruget i alle størrelser af husstande. Denne problemstilling kan eksemplificeres ved at en forbrugsmængde, der er betragtet som „højt“ for boliger med én beboer, kan anses som værende „gennemsnitligt“ for en bolig med to beboere.

Selvom der anvendes en mere forståelig enhed for at synliggøre de enkelte forbrug, er der ingen garanti for, at beboerne, som det fremgår af analysen fra *Gill et al.* [38], vil revurdere deres forbrug. Nogle beboere er modtagelige over for ændringer. Derimod er andre beboere mere eller mindre ligeglade, og de anser deres forbrug som en nødvendighed i forhold til at opretholde den valgte livsstil. Denne forskel på beboere er illustreret i kapitel 3, hvor det fremgår, at der er stor forskel på, hvorvidt beboerne er villige til at indføre ændringer på bekostning af eksempelvis komfort, økonomisk vinding eller omlægning af vaner.

5.2.2. Inddeling af de udvalgte klasser

Med udgangspunkt i den foregående vurdering af enheder fastlægges følgende klasser for de tre klassificeringstræer:

- Klasse til klassificeringstræet for rumopvarmning: kWh/m^2 per år.
- Klasse til klassificeringstræet for varmt brugsvand: $m^3/år$.
- Klasse til klassificeringstræet for elforbrug: $kWh/år$.

5.2.1.1 Indsamling af data til klasse

Den ovenstående definering af klasserne indikerer, i hvilke enheder data til klassen skal indsamles. Data til rumopvarmning er det, der kan give anledning til de største udfordringer ved indsamling. Det skyldes, at forbruget er afhængigt af boligens varmeanlæg, og vil derved kræve en omregning til den valgte enhed. Data om beboernes varmeforbrug kan indsamles via aflæsning af målere i boliger med fjernvarme, hvorimod boliger med gas-/oliefyr kan opgøres ved hjælp af mængden af anvendt brændsel. Fælles for alle forbrug er, at energiforbruget til opvarmning af varmt brugsvand kan være inkluderet, afhængigt af systemernes opbygning. Det indsamlede data vil kræve en graddøgnsberegning. Dette skal udføres for at normalisere forbruget, for derigennem at sikre at klassificeringstræet ikke udvikles på baggrund af data fra eksempelvis en meget kold vinter, hvor forbruget til rumopvarmning ofte er forøget i forhold til normalen.

Data om beboernes forbrug af varmt brugsvand er ofte vanskeligt at fremskaffe, idet der i mange boliger udelukkende måles på forbruget af det kolde vand. Omkostningerne til opvarmning af brugsvandet er ofte en del af det samlede varmeforbrug. Derigennem er det varme brugsvand et skjult forbrug, idet forbruget ikke opgøres separat. På nuværende tidspunkt er forbruget af varmt brugsvand antaget at udgøre omkring 1/3 af det samlede vandforbrug, men ifølge afsnit 2.1.2 og 3.4 erfares det samtidig, at der inden for varmtvandsforbruget er store variationer, jf. figur 2.11 på side 21. Forbruget af det varme brugsvand kan undersøges ved at anvende skønnet om 1/3 af det samlede vandforbrug for at lette processen, men risikoen for fejl er til stede. Fordelen ved dette er, at det er en simpel beregning, samtidig med at data for det samlede vandforbrug kan indsamles gennem forsyningsselskaberne. En anden mulighed for hvorledes problemstillingen ved forbruget af varmt brugsvand kan undersøges, kan være ved en gennemgående undersøgelse, hvor både forbruget af det varme brugsvand undersøges samtidig med beboernes vaner. Denne løsning vil kræve, at der installeres separate vandmålere på det varme brugsvand i boligerne. Dette er simpelt ved nybyggeri, men ved eksisterende boliger er der risiko for, at det bliver en større opgave. Denne metode vil give de bedste resultater, men det kræver både tid og økonomi for at gennemføre dette i en større skala.

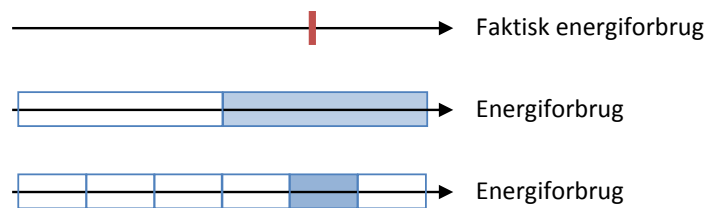
Data for elforbruget er enkelt at fremskaffe, idet det kan oplyses af forsyningsselskaberne og opgøres i forvejen i den valgte enhed for klassen til klassificeringstræet.

5.2.2 Inddeling af de udvalgte klasser

Ved anvendelse af en kategorisk klasse i et klassificeringstræ kræves det, som beskrevet tidligere, at der på forhånd defineres nogle opdelingsmuligheder af denne. I analysen fra *Zhun*

et al. [108] er klassen for klassificeringstræet angivet som et årligt energiforbrug. Klassen er inddelt i to undergrupper; „højt EUI“ eller „lavt EUI“, jf. figur 3.6 på side 42 hvor der ligeliges fremgår en definition af EUI. Beboerne inddeles i disse grupper afhængig af de øvrige attributters egenskaber. Hver af disse underopdelinger dækker over et samlet årligt energiforbrug på mellem $176 - 441,5 \text{ MJ/m}^2$ og $441,5 - 707 \text{ MJ/m}^2$ for hhv. det lave og høje årlige energiforbrug. Grænsen for opdelingen er valgt som gennemsnittet af den laveste og højeste måling.

Den endelige klasse i et klassificeringstræ kan beskrives ved hjælp af flere inddelinger. Der er fordele og ulemper ved både små og store intervaller, hvorfor opdelingen skal vurderes i forhold til det data, som klassificeringstræet skal udvikles på baggrund af. Figur 5.3 illustrerer princippet i inddelingen af klassen for energiforbruget, hvor dette kan gælde for både energiforbrug til rumopvarmning, varmtvands- og elforbrug.



Figur 5.3: Skitsering af princip for klasseinddeling.

Den endelige klasse i et klassificeringstræ kan beskrives ved hjælp af flere inddelinger. Der er fordele og ulemper ved både små og store intervaller, hvorfor opdelingen skal vurderes i forhold til det data, som klassificeringstræet skal udvikles på baggrund af. Figur 5.3 illustrerer princippet i inddelingen af klassen for energiforbruget, hvor dette kan gælde for både energiforbrug til rumopvarmning, varmtvands- og elforbrug. Fordelen ved at inddele klassen i store intervaller er, at risikoen for at klassificeringstræet fejlklassificerer elementerne i forhold til klasseinddelingerne reduceres. Ulempen i denne forbindelse er, at når et element er klassificeret, angives energiforbruget med et stort interval. Blandt fordelene ved at inddele klassen i små intervaller er, at der er mulighed for, at et element klassificeres med et energiforbrug, der er angivet inden for et mindre interval, som skitseret nederst på figur 5.3. Ulempen i denne forbindelse er, at der ved udviklingen af klassificeringstræet er behov for større datamængder, for at opretholde et rimeligt antal elementer i knuderne. Derudover er der en øget risiko for, at flere elementer misklassificeres ved brug af klassificeringstræet hvormed nøjagtigheden reduceres.

Ved at inddele klassen i små intervaller kan der, ved optimal udvikling, påvises forskelle af forbrugsstørrelsen på baggrund af beboernes adfærd. Ved at inddele klassen i mange undergrupper kan der opstå en mulighed for at illustrere, at en lille ændring i beboernes adfærd kan medføre et reduceret forbrug. Derigennem kan en inddeling af klassen i små intervaller virke som incitament til at spare på energien for beboerne i boligen.

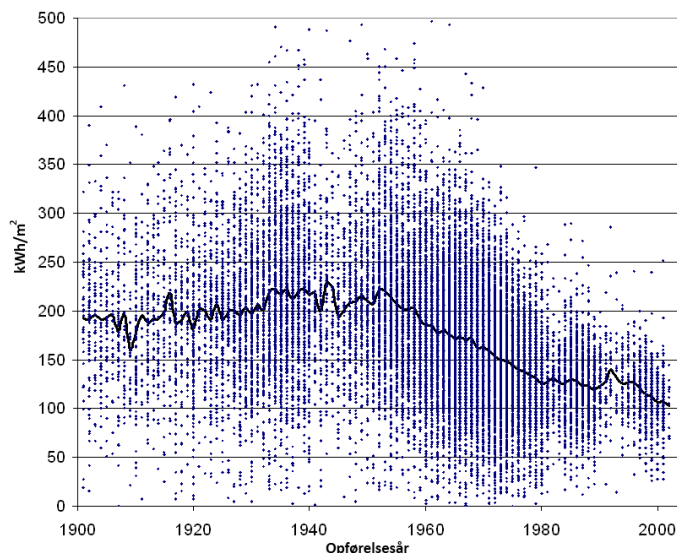
5.2.2.1 Energiforbrug til rumopvarmning

Klassens enhed for energiforbruget til rumopvarmning er i afsnit 5.2 fastlagt til $kWh/m^2 \text{ per år}$. Der er adskillige muligheder, for hvordan inddelingen af den valgte klasse kan defineres.

Der er mulighed for, som i analysen af *Zhun et al.* [108], at vælge to inddelinger, eller der kan vælges et alternativ med flere inddelinger. Antallet af inddelinger afhænger af det tilgængelige data, samt eventuelle subjektive krav til klassificeringstræet. Antallet af undergrupper, som klassen skal inddeles i, har betydning for intervallerne størrelse, og har derigennem direkte indflydelse på klassificeringstræets egenskaber.

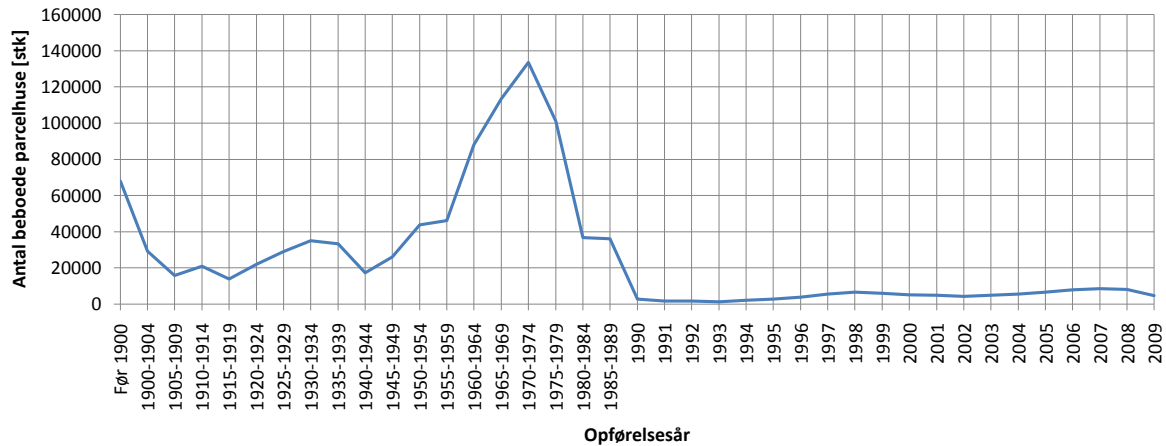
Energiforbruget til rumopvarmning påvirkes af flere forskellige faktorer, der knytter sig til én af tre følgende kategorier; tekniske, demografiske og adfærdsrelaterede forhold, jf. kapitel 3. Kombinationen af faktorerne inden for disse tre forhold er essentielle for at beskrive en sammenhæng mellem energiforbruget til rumopvarmning og brugeradfærd. Af denne årsag bør alle faktorerne medtages ved udvikling af et klassificeringstræ. Foruden de faktorer, der er beskrevet i kapitel 3, bør forhold som boligtype og boligens opførelsesår vurderes. Begge faktorer ændrer indirekte på intervallerne, der skal defineres, idet de har betydning for, hvor store variationer der kan forekomme imellem det laveste og det højeste forbrug. I nærværende projekt fokuseres på ejerboliger, jævnfør problemformuleringen i afsnit 2.5, men i appendiks H er der foretaget en uddybning af hvilken betydning boligtypen har på energiforbruget til rumopvarmning, og hvordan dette forhold kan medtages i klassificeringstræet.

Boligens opførelsesår har betydning for, hvor stor forskel der er imellem det laveste og det højeste forbrug. Derfor har denne faktor en stor betydning for, hvordan klassen for energiforbruget til rumopvarmning kan inddeles. Energiforbruget til rumopvarmning i boliger varierer betragteligt, hvilket fremgår af figur 5.4. Af denne fremgår det tydeligt, at der er en stor variation af forbruget, og dermed for intervallerne af klasseinddelingen, afhængig af om der fokuseres på boliger fra alle årtier, eller om der udelukkende fokuseres på enkelte perioder, eksempelvis efter 1980'erne.



Figur 5.4: Skitsering af årligt naturgasforbrug i forhold til boligens opførelsesår i danske parcelhuse. Gengivet med tilladelse fra *Vibeke Hansen Kjærbye* [51].

Figur 5.5 illustrerer antallet af beboede parcelhuse i 2010 under hensyntagen opførelsesåret for boligen¹. I figuren er det illustreret, at de parcelhuse, der udgør den største andel af beboede parcelhuse i 2010, er opført imellem 1950 – 1984.



Figur 5.5: Antal beboede parcelhuse i Danmark i 2010 i forhold til boligens opførelsesår [91].

Der er adskillige muligheder for, hvilke parcelhuse klassificeringstræet skal være gældende for. I det efterfølgende er der oplistet enkelte forslag til hvilke perioder, der kan anvendes. Valget, af hvilke opførelsesår klassificeringstræet skal være gældende for, skal træffes ud fra hvilket overordnede formål der er fastlagt for dette.

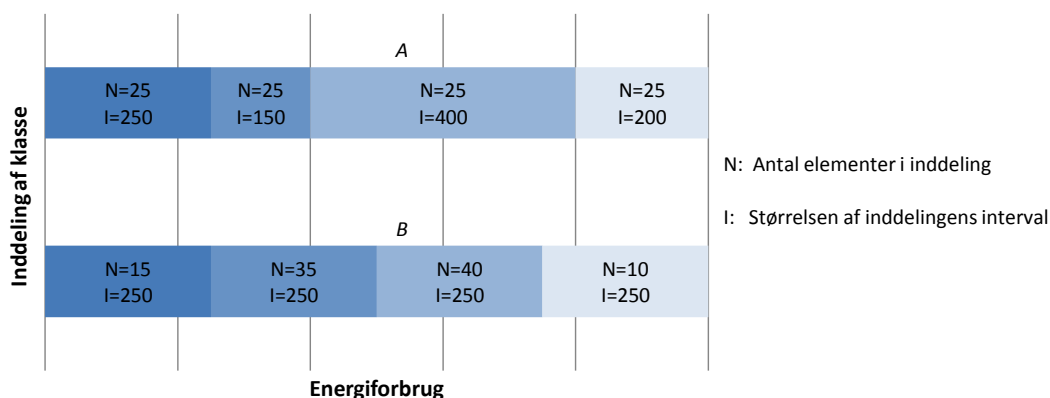
- Et klassificeringstræ der er gældende for alle opførelsesår.
- Et klassificeringstræ der er gældende for boliger opført i perioden 1950 – 1984.
- Et klassificeringstræ der er gældende for boliger, der er opført inden for de seneste ti år.

Et klassificeringstræ, der er udviklet til at være gældende alle opførelsesår, vil kræve et omfattende arbejde med at fremskaffe data fra boliger fra de forskellige årstal. Foruden indsamling af informationer om energiforbrug til rumopvarmning, er det nødvendigt, at indsamle oplysninger omhandlende i hvor høj grad boligen er blevet renoveret. Ved at fokusere på parcelhuse fra 1950 – 1984 opnås der en analyse der, ved optimal udvikling², kan bidrage til at beskrive sammenhængen mellem beboernes adfærd og energiforbruget til rumopvarmning i 52 % af parcelhusene i Danmark [91]. Det er ligeledes nødvendigt, for boligerne fra denne periode, at vurdere i hvor høj grad boligerne er renoveret. Ved at fokusere udelukkende på boliger der er opført inden for de seneste ti år, opnås der et klassificeringstræ, hvor beboernes adfærd formentlig har større indflydelse på energiforbruget. Det skyldes, at de tekniske forhold, der knytter sig til energiforbruget til rumopvarmning, er forbedret i de nye boliger.

¹Af figuren fremgår det samlede varmeforbrug i boliger, hvilket både dækker over energiforbruget til rumopvarmning og opvarmning af varmt brugsvand.

²For at opnå en optimal udvikling af klassificeringstræet kræves der tilstrækkeligt data, som er repræsentativt for den samlede mængde af husstande fra den udvalgte periode.

Der er flere forskellige muligheder for inddeling af klassen for energiforbruget til rumopvarmning. De to simpleste metoder tager begge udgangspunkt i det data, som klassificeringstræet skal udvikles på baggrund af. Den første mulighed er, at inddele klassen i et forudbestemt antal undergrupper, hvor den samlede mængde af husstande fordeles ligeligt imellem grupperne. Fordelen ved denne metode er, at intervallerne er lette at definere. Det skyldes, at det kun kræver aflæsninger af energiforbruget til rumopvarmning i de respektive undergrupper, og derigennem angive henholdsvis en minimum- og maksimumværdi for undergruppens interval. Ulempen ved denne metode er, at størrelsen på de definerede intervaller kan variere betragteligt. Dette scenarie kan blandt andet forekomme i de tilfælde, hvor en større mængde husstande har omtrent samme størrelse forbrug. Disse husstande vil skabe et interval i midten af en datamængde, der dækker over den samme mængde af husstande som de øvrige undergrupper. Intervallet for energiforbruget i denne gruppe vil forekomme beskedent, sammenholdt med de øvrige intervaller for de resterende undergrupper. Et eksempel herpå er illustreret øverst på figur 5.6. Den anden metode for inddeling af klassen kan defineres ved at betragte alle forbrugsstørrelser til rumopvarmning, og efterfølgende inddele datamængden i lige store intervaller, illustreret som del B i figur 5.6. Fordelen ved denne metode for inddeling af klassen er, at intervallerne får samme størrelse. Ulempen ved denne metode er, at der kan opstå store forskelle på antallet af husstande i de definerede inddelinger. Dette kan medføre, at klassificeringstræet bliver visuelt uens, med eksempelvis mange knuder i den ene side og et næsten tomt træ i den modsatte side.



Figur 5.6: To måder, A og B, hvorpå klassen for et klassificeringstræ kan inddeles.

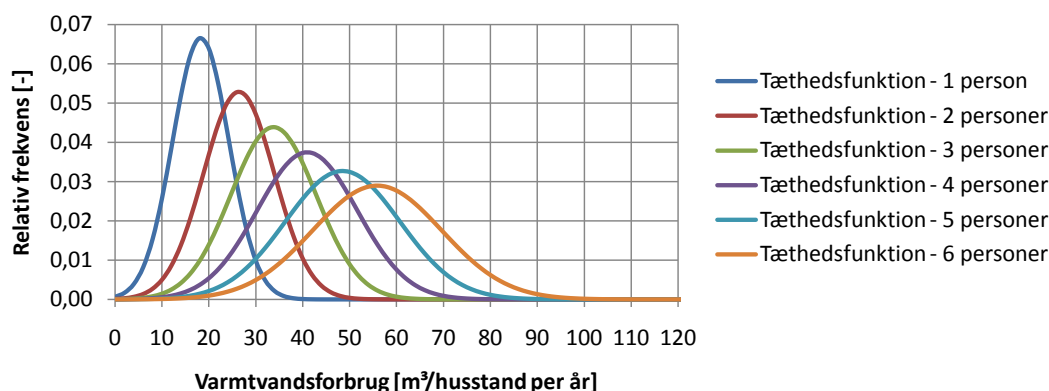
I det nærværende projekt er der ikke defineret en fast inddeling af klassen med tilhørende intervaller for energiforbruget til rumopvarmning. Det skyldes, at klassen er afhængig af målet for klassificeringstræet, og det dertilhørende indsamlede data. Det anbefales, at klassen tildeles enheden kWh/m^2 per år, idet det dermed er muligt at sammenligne forskellige boligstørrelser. Det er med ovenstående gennemgang klarlagt, hvilke faktorer der har betydning for inddelingen af klassen. Der er yderligere informeret om de adskillelige muligheder, der er for udviklingen af et klassificeringstræ for energiforbruget til rumopvarmning, afhængig af hvilket mål det forventes at beskrive.

5.2.2.2 Varmtvandsforbrug

I afsnit 5.2 fremgår det, at der er flere muligheder for valg enhed til klassen for varmtvandsforbruget. Inddelingen af klassen for varmtvandsforbruget foregår på baggrund af hvilke muligheder der er i det tilgængelige data, på baggrund af litteraturstudiet og ud fra subjektive valg. Da det fremgår af litteraturstudiet, at antallet af beboere har stor indflydelse på varmtvandsforbruget, anbefales det i nærværende rapport, at der udvikles ét klassificeringstræ for hver af de forskellige husstandsstørrelser. Dette sker for at imødegå den tidligere omtalte problemstilling omkring defineringen af hvad der er højt og lavt forbrug.

Udvikling af klassificeringstræer med hensyntagen til husstandens størrelse får konsekvens for inddelingen af intervaller, idet arbejdet med inddeling af klassen bliver mere omfattende. Det skyldes, at der skal defineres specifikke intervaller for alle de mulige husstandsstørrelses klasser. Fordelen ved at der udvikles ét klassificeringstræ for alle husstandsstørrelser er, at dette kan blive mindre, mere overskueligt, samt at inddelingen af klassen til klassificeringstræet bliver mere gennemskuelig.

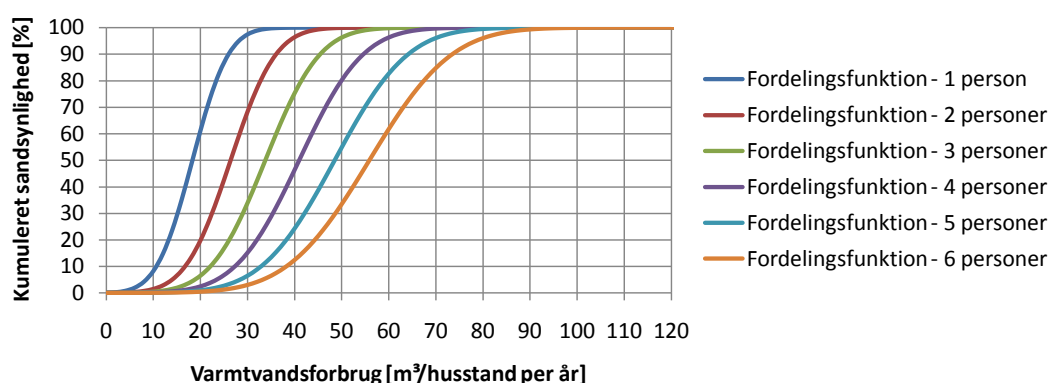
Der er adskillige muligheder for inddelingen af klassen til varmtvandsforbruget. Af figur 5.7 fremgår det, at forbruget ifølge fordelingerne er meget varierende³. Denne spredning i forbruget har betydning for antallet af inddelinger for klassen.



Figur 5.7: Tæthedsfunktioner for varmtvandsforbrug opgjort per husstand for normalfordelinger med parametrene angivet i tabel L.8 i appendiks.

Inddelingen af klassen kan foretages ved hjælp af flere forskellige metoder, hvor der i nærværende afsnit vurderes tre forskellige; inddeling af klassen på baggrund af subjektive valg, inddeling af klassen i forhold til at hver klasseinddeling skal indeholde samme antal husstande eller inddeling af klassen ved anvendelse af fraktiler i fordelingsfunktionerne, der fremgår af figur 5.8.

³Sandsynlighedsfordelingerne er udviklet i dette projekt, hvor nærmere information om disse fremgår af appendiks L



Figur 5.8: Fordelingsfunktioner for varmtvandsforbrug opgjort per husstand for normalfordelinger med parametrene angivet i tabel L.8 i appendiks.

Der er fordele og ulemper ved alle tre omtalte metoder for inddelingen af klassen for det årlige varmtvandsforbrug i boliger. I den førstnævnte mulighed bliver inddelingen af klassen baseret på baggrund af viden fra litteraturstudiet samt subjektive vurderinger af projektgruppen. Fordelen ved denne metode er, at der kan fastsættes nogle anvendelige grænser med udgangspunkt i vurdering af viden fra litteraturstudiet og fra de udviklede fordelinger. Det betyder, at grundlaget for inddelingerne er tilgængelige. Ulempen ved denne metode er, at intervallerne dannes på baggrund af et beskedent datagrundlag, hvormed det kan afvige fra de faktiske forhold. Det betyder, at der dermed er risiko for, at intervallerne ikke er dækkende, når der inddrages data fra større undersøgelser.

Metoden hvor intervallerne defineres ved at inddele den samlede mængde af husstande i et forudbestemt antal undergrupper, med lige mange beboere i hver, har den fordel, at intervallerne er simple at definere. Det skyldes, at det kun kræver aflæsninger af mængden af varmtvandsforbruget i de respektive undergrupper, og derigennem angive henholdsvis en minimum- og maksimumværdi for undergruppens interval. Ulempen ved denne metode er, at der er risiko for, at de definerede intervaller er uens, se uddybende forklaring i afsnit 5.2.2.1. Ved anvendelse af store intervaller for klassen reduceres præcisionen for vurderingen af forbruget størrelse, og ved mindre intervaller er der en større risiko for at elementerne klassificeres forkert. Denne metode er derfor kun optimal, hvis intervallerne for de definerede klasseinddelinger har tilnærmelsesvis samme intervalstørrelse.

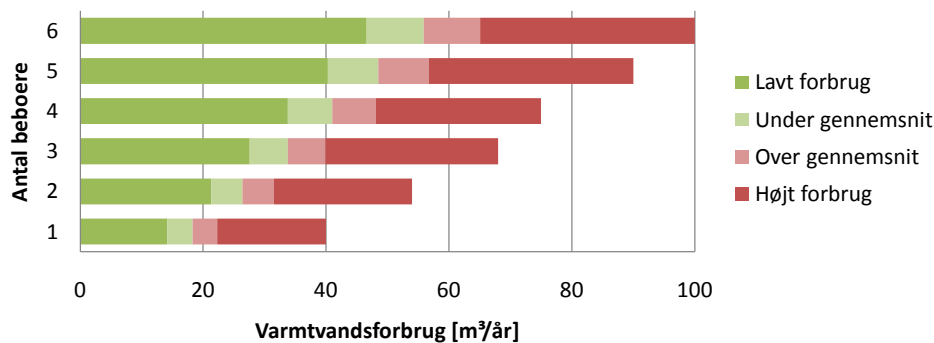
Ved definering af klassens intervaller kan der, foruden de to ovenstående metoder, anvendes nogle fastsatte fraktiler i fordelingsfunktionerne, der fremgår af figur 5.8. Fordelen ved at anvende nogle fastsatte fraktiler er, at forbrugsgrenserne for intervallerne er nemme at beskrive og anvendelig for flere husstandsstørrelses fordelingsfunktioner. Derudover er der mulighed for, at anvende fraktiler der kan beskrive eksempelvis realistiske lave, gennemsnitlige samt høje forbrug, jf. viden fra litteraturstudiet, på trods af de udførte undersøgelser begrænsede omfang inden for emnet om varmtvandsforbrug i boliger.

På baggrund af de foregående vurderinger foretages følgende forudsætninger for klassificeringstræet til varmtvandsforbruget. Inddelingen af klassen tager udgangspunkt i en inddeling, som består af fire dele, der beskriver varmtvandsforbruget: „lavt“, „under gennemsnit“, „over gennemsnit“ og „højt“. Inddeling sker på baggrund af fraktilerne: 25 %, 50 % og 75 %, hvilket betyder, at klasseinddelingen med betegnelsen „lavt“ har intervallet mellem 0 – 25 %

fraktilen, klasseinddelingen „under gennemsnit“ defineres ved intervallet mellem 25 – 50 %, „over gennemsnit“ mellem 50 – 75 % fraktilen og klasseinddelingen „højt“ er defineret til 75 – 100 % fraktilen. Intervallerne for de angivne klasseinddelinger fremgår af tabel 5.2 og er illustreret i figur 5.9.

Antal beboere	Lavt forbrug [m ³ /år]	Under gennemsnit [m ³ /år]	Over gennemsnit [m ³ /år]	Højt forbrug [m ³ /år]
1 beboer	[0 – 14, 2]	]14, 2 – 18, 3]	]18, 3 – 22, 3]	> 22, 3
2 beboere	[0 – 21, 3]	]21, 3 – 26, 4]	]26, 4 – 31, 5]	> 31, 5
3 beboere	[0 – 27, 6]	]27, 6 – 33, 8]	]33, 8 – 39, 9]	> 39, 9
4 beboere	[0 – 33, 8]	]33, 8 – 41, 0]	]41, 0 – 48, 1]	> 48, 1
5 beboere	[0 – 40, 3]	]40, 3 – 48, 5]	]48, 5 – 56, 7]	> 56, 7
6 beboere	[0 – 46, 6]	]46, 6 – 55, 9]	]55, 9 – 65, 1]	> 65, 1

Tabel 5.2: Angivelse af intervallerne for den valgte inddeling af varmtvandsforbruget, angivet for den samlede husstand.



Figur 5.9: Inddeling af klassen for varmtvandsforbruget for forskellige husstandsstørrelser på baggrund af fraktiler.

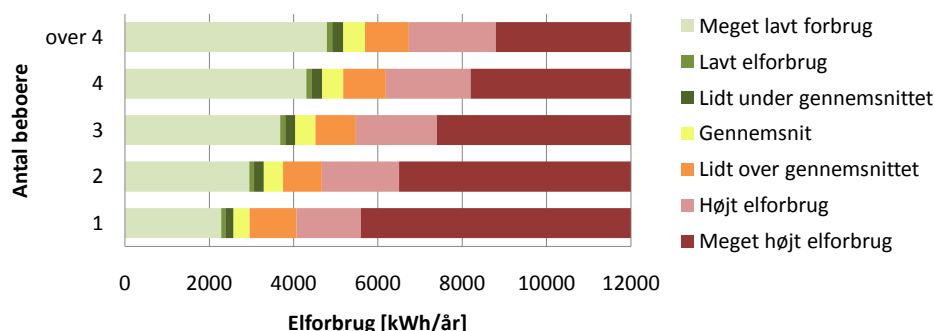
5.2.2.3 Elforbrug

Der indgår flere af de samme overvejelser for inddeling af klassen for elforbruget, som er beskrevet i det foregående afsnit med varmtvandsforbruget. Da det fremgår af litteraturstudiet, at antallet af beboere har stor indflydelse på varmtvandsforbruget, anbefales det i nærværende rapport, at der udvikles ét klassificeringstræ for hver af de forskellige husstandsstørrelser, af samme årsag som angivet i det foregående afsnit.

Inddelingen af klassen for elforbruget kan defineres som i analysen af *Zhun et al.* [108], hvor der er to inddelinger, eller den kan indeledes, som kundetyperne er defineret hos elseskaberne *Energi Nord* [64] og *DONG Energy* [29]. I disse tilfælde er forbruget inddelt i syv forskellige intervaller med forskellige grænser afhængig af husstandsstørrelsen. Som beskrevet tidligere er inddelingen af klassen afhængig af det tilgængelige data og subjektive valg. I det efterfølgende beskrives, hvorledes klasseinddelingen for elforbruget fastlægges.

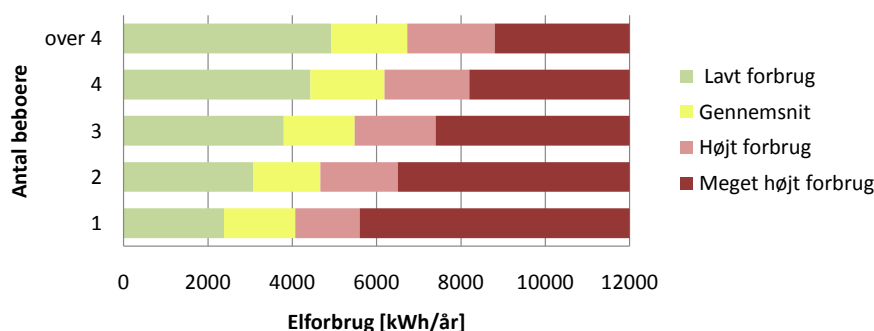
5.2.2. Inddeling af de udvalgte klasser

Inddelingen af klassen kan foretages ved hjælp af flere forskellige metoder, hvor disse er gennemgået i det foregående afsnit. Idet de tidligere omtalte forsyningsselskaber har defineret nogle forbrugskategorier på baggrund af deres kunders forbrug, vil disse intervaller beskrives med henblik på at anvende disse som grundlag for inddelingen af klassen for elforbruget. Inddelingen af forbruget fra forsyningsselskaberne fremgår af figur 5.10 og beskriver elforbruget i parcelhuse under hensyntagen til husstandsstørrelsen.



Figur 5.10: Kategorisk inddeling af elforbrug ifølge *Energi Nord* [64] og *DONG Energy* [29] under hensyntagen til husstandsstørrelsen.

Det fremgår af figur 5.10, at der er flere af inddelingerne, der er defineret ved et mindre interval. Dette drejer sig blandt andet om kategorierne „lavt forbrug“, „lidt under gennemsnit“ og „gennemsnit“, hvor kategorierne hver især spænder imellem hhv. $[96 - 170]$, $[191 - 259]$ og $[382 - 517]$ kWh/år. Det vurderes, at risikoen for fejlklassificering er for høj, såfremt de angivne intervaller anvendes i klassificeringstræet for elforbruget. På denne baggrund inddeles klassen i fire dele. De fire inddelinger bygger på en reducere af de syv kategorier fra figur 5.10. Kategorierne „meget lavt forbrug“ og „lavt forbrug“ er kombineret og tilsvarende med „lidt under gennemsnit“, „gennemsnit“ og „lidt over gennemsnit“. Resultatet af dette samt den endelige definering af klassens inddeling fremgår af figur 5.11, og de numeriske værdier for grænserne fremgår i tabel 5.3.



Figur 5.11: Den valgte inddeling af klassen for elforbruget.

Antal beboere	Lavt forbrug [kWh/år]	Gennemsnit [kWh/år]	Højt forbrug [kWh/år]	Meget højt forbrug [kWh/år]
1 beboer	[0 – 2381]	]2381 – 4070]	]4070 – 5600]	> 5600
2 beboere	[0 – 3065]	]3065 – 4668]	]4668 – 6500]	> 6500
3 beboere	[0 – 3800]	]3800 – 5480]	]5480 – 7400]	> 7400
4 beboere	[0 – 4426]	]4426 – 6187]	]6187 – 8200]	> 8200
over 4 beboere	[0 – 4919]	]4919 – 6730]	]6730 – 8800]	> 8800

Tabel 5.3: Angivelse af grænserne for den udvalgte inddeling for elforbruget.

5.3 Definerings af klassificeringstræernes attributter

I dette afsnit beskrives hvilke attributter, der kan anvendes i forbindelse med klassificeringstræer til estimering af energiforbrug til rumopvarmning, varmtvandsforbrug og elforbrug i boliger. Betegnelsen attributter dækker i denne forbindelse over forhold for boligen, beboerne og deres vaner. Afsnittet er opdelt i to underafsnit, hvor det første beskriver et casestudie af hvilke andre forhold, end de i litteraturstudiet beskrevne, der kan forklare forbrugsforskelle boliger imellem. I det efterfølgende beskrives hvilke attributter, der bør søges nærmere informationer om, når de førnævnte klassificeringstræer skal udvikles.

5.3.1 Vurdering af nye parametre på baggrund af Fremtidens Parcelhuse

I dette afsnit beskrives et casestudie, der er udført i nærværende projekt, med henblik på at identificere nogle mulige forhold i forbindelse med beboere, deres vaner og holdninger, der kan forklare dele af de forbrugsforskelle, der kan observeres mellem husstandes energiforbrug.

Undersøgelsen er udført på baggrund af data fra otte, og i nogle tilfælde ni, parcelhuse. Datagrundlaget stammer fra et tidligere omtalt projekt, benævnt *Fremtidens Parcelhuse af Knudsen et al.* [52], hvor der er foretaget detaljerede målinger af energiforbrug og indeklima. Parcelhusene er opført i Herfølge i 2006, og har været beboet siden 2007. De er opført som lavenergibygning klasse 2, men kun fire ud af de ni parcelhuse overholder disse krav i praksis. Samtidig med at de tekniske forhold er evalueret, er der udsendt et spørgeskema til beboerne, hvor der stilles spørgsmål til deres opfattelse af at leve i boligerne, deres vaner og holdninger. Otte af husstandene har besvaret spørgeskemaet, og der er i nærværende projekt opnået adgang til disse spørgeskemabesvarelser. I de efterfølgende underafsnit beskrives de væsentligste resultater fra casestudiet. I tabel 5.4 fremgår enkelte basisoplysninger for de ni boliger, hvor der kan søges yderligere information om undersøgelsen i appendiks N.

5.3.1. Vurdering af nye parametre på baggrund af Fremtidens Parcelhuse

Basisoplysninger	Hus 1	Hus 2	Hus 3	Hus 4	Hus 5	Hus 6	Hus 7	Hus 8	Hus 9
Boligareal [m^2]	167	171	143,5	154,8	147,5	167	172	172	149,7
Antal beboere	4	4	2	4	4	4	4	6	2
Antal voksne	4	2	2	2	2	2	2	—	2
Antal børn	0	2	0	1	2	2	2	—	0
Antal teenagere	0	0	0	1	0	0	0	—	0
Elforbrug* [$kWh/\text{år}$]	5638	3859	5773	4920	4791	2974	2939	9868	1966
Varmeforbrug [$kWh/\text{år}$]	8279	9920	5395	5291	10345	6367	13555	3288	7561
Vandforbrug [$L/\text{person per døgn}$]	68,0	90,6	180,0	98,9	73,3	97,0	124,6	39,0	75,1
Varmtvandsforbrug [$L/\text{person per døgn}$]	14,9	44,0	39,0	36,8	26,4	33,3	30,2	4,6	14,5
Varmtvandsforbrugets andel af det samlede vandforbrug [%]	22	49	22	37	36	34	24	12	19

* Elforbruget er opgjort uden bygningsel.

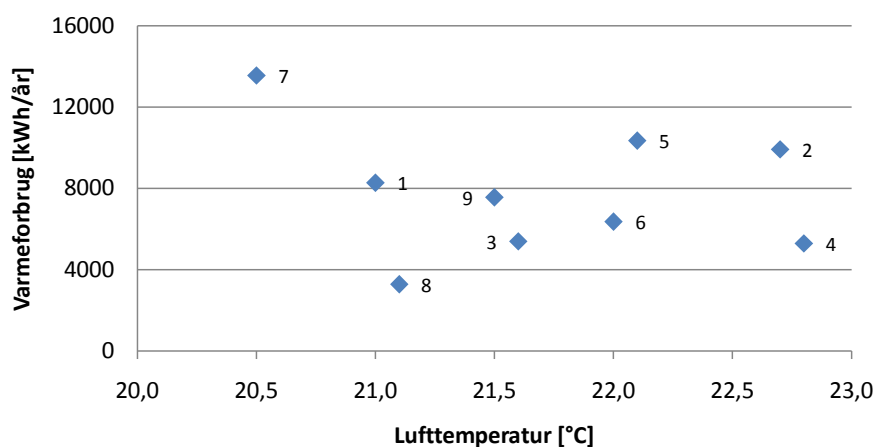
Tabel 5.4: Basisoplysninger om parcelhusene fra projektet *Fremtidens Parcelhuse* [52].

5.3.1.1 Test af parametre i forhold til det samlede varmekorbrug

På baggrund af afsnit 2.2.1 og 3.2 er der opnået viden om hvilke forhold, der har betydning for husstandes samlede varmekorbrug og energikorbruget til rumopvarmning. Det er i nogle af de artikler og rapporter, der fremgår i de førnævnte afsnit, undersøgt hvilke parametre, der kan forklare dele af den forbrugsvariation, der kan observeres mellem forskellige husstande, når det samlede varmekorbrug eller energikorbruget til rumopvarmning sammenholdes. Blandt de forhold, der i litteraturstudiet har vist sig at forklare dele af den førnævnte variation, er blandt andet lufttemperaturen, boligens størrelse, beboernes alder m.fl. Der er foretaget en analyse af, hvorvidt disse forhold viser samme tendenser i det tilgængelige data fra *Fremtidens Parcelhuse* [52], samt om der kan identificeres nogle andre parametre, der kan forklare dele af forbrugsvariationen. I den forbindelse bør det bemærkes, at der tages udgangspunkt i boligernes samlede varmekorbrug. Et detaljeret casestudie af førnævnte data fremgår af appendiks N.1, hvor der skal gøres opmærksom på, at vurderingsgrundlaget er beskedent grundet antallet af husstande i undersøgelsen. Da det i forvejen er begrænset hvor meget tilgængeligt litteratur, der er om brugeradfærd i boliger og dets betydning for varmekorbruget, vil den efterfølgende vurdering have til formål at klarlægge hvilke forhold, der kan undersøges nærmere i fremtidige projekter.

Ikke alle tendenser fremgår lige tydeligt i det tilgængelige data, men hvad angår lufttemperaturen, viser undersøgelsen, at flere forhold bør tages i betragtning, hvis denne anvendes til vurdering af boligernes varmekorbrug. Et eksempel herpå er illustreret i figur 5.12, hvor det fremgår, at husstanden med den laveste målte gennemsnitstemperatur i perioden fra okto-

ber 2009 til og med december 2009 er den husstand, der har det højeste energiforbrug til rumopvarmning.



Figur 5.12: Skitsering af den gennemsnitlige lufttemperatur i de ni husstande, målt i perioden fra oktober 2009 til og med december 2009, sammenholdt med varmekonsumet. Punkternes nummerering knytter sig til husnummeret i undersøgelsen, jf. tabel 5.4. [52]

Årsagen til det høje energiforbrug og den, i forhold til de øvrige husstande, lave lufttemperatur kan være et større varmetab fra boligen, men også forhold som udluftningsvaner har betydning for varmekonsumet. En detaljeret beskrivelse fremgår af appendiks N.1, men det viser sig blandt andet, at beboerne i *hus 7* har problemer med, at de føler træk, samt at det føles for koldt både om sommeren og om vinteren. Det fremgår ikke af spørgeskemaerne, hvad der er årsagen til, at beboerne føler træk, og at det er for koldt. Derved vides det ikke, om det er når der luftes ud, eller om det er på andre tidspunkter, at beboerne ikke er i komfort. I forsøg på at indkranse problemstillingen er boligens tæthed undersøgt ved hjælp af resultatet af en blowerdoortest, hvor disse fremgår af tabel 5.5. Resultatet fra *hus 7* er sammenlignet med resultaterne fra tre andre boliger, hvor det vurderes, at beboernes oplevelse af træk ikke skyldes fejl eller utætheder i konstruktioner.

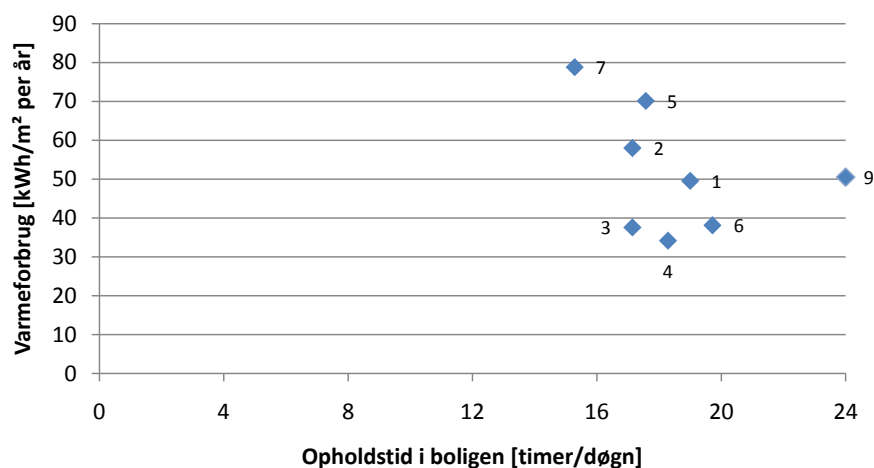
Hus	Utæthed [L/s per m ²]
1	1,00
2	0,80
5	0,52
7	1,06

Tabel 5.5: Resultater for tæthedsprøvning af fire boliger fra *Fremtidens Parcelhuse* [52].

Da resultaterne fra de udførte tæthedsprøvninger ikke indikere, at der er problemer med boligens tæthed, sammenholdes besvarelsene af spørgeskemaerne at vurdere en anden faktor, udluftning i boligen, som kan have betydning for varmekonsumet. Beboerne i *hus 7* angiver, som nogle af de eneste i undersøgelsen, at de lugter ud om dagen og om natten både sommer og vinter. Beboerne angiver endvidere, at årsagen til udluftning er, at de ønsker at

reducere temperaturen i hjemmet, samt at de ønsker frisk luft i boligen. Det fremgår ikke af spørgeskemaerne, hvordan der udluftes, eller hvor længe udluftningerne varer, hvorfor det er vanskeligt at vurdere, hvor stor betydning førnævnte har for energiforbruget til rumopvarmning. Ikke desto mindre vil en øget udluftning medføre en øget luftmængde til boligen, og dermed kræves yderligere energi til at opvarme luften i kolde perioder. Idet varmemeforbruget er opgjort som det samlede forbrug, dvs. inklusiv både energiforbruget til rumopvarmning og opvarmning af varmt brugsvand, er det vurderet om beboerne i *hus 7* har et større varmtvandsforbrug i forhold til de øvrige boliger. Dette vurderes imidlertid ikke at være årsagen til afvigelsen, idet deres varmtvandsforbrug svarer til det gennemsnitlige forbrug for alle ni husstande. Det har i nærværende casestudie ikke været muligt entydigt at identificere årsagen til det høje varmemeforbrug i *hus 7*. Det kan imidlertid være interessant at undersøge, om beboernes udluftningsvaner er årsagen til det høje forbrug.

Ovenstående bevidner, at andre forhold end lufttemperaturen må tages i betragtning, når boligers varmemeforbrug skal vurderes. Et af de forhold der kan undersøges nærmere i fremtidige projekter er beboernes opholdstid i boligen sammenholdt med varmemeforbruget. Det fremgår af figur 5.13, at der kan være en tendens til, at beboere der opholder sig mindre tid i boligen har et større varmemeforbrug. Tendensen er dog ikke entydig, eftersom beboerne i husstanden der har oplyst den længste opholdstid i boligen, ikke har det laveste varmemeforbrug. Beboerne i *hus 9* skiller sig ud fra beboerne i de øvrige huse ved at være et ældre ægtepar, hvor den ene part har trukket sig tilbage fra arbejdsmarkedet. Ifølge spørgeskemabesvarelsenerne skiller deres vaner sig ikke markant ud, men dette er det blandt de eneste forhold, der kan forårsage forskellen. Den svage tendens til at varmemeforbruget er faldende med stigende opholdstid i boligen kan blandt andet skyldes, at jo længere tid beboerne er til stede i boligen des større intern varmebelastning og dermed et større varmetilskud i boligen.



Figur 5.13: Skitsering af beboernes oplyste opholdstid i otte husstande sammenholdt med varmemeforbruget opgjort per arealenhed i de pågældende husstande.

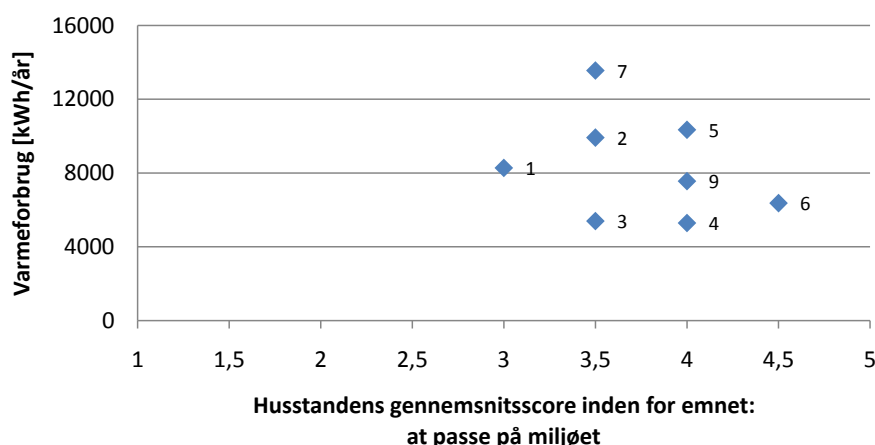
Foruden de parametre der gennem litteraturstudiet har vist sig, at kunne forklare dele af den forbrugsvariation, der kan observeres mellem husstandene, er det muligt at sammenholde varmemeforbruget med henholdsvis beboernes holdninger til miljø og energibesparelser. Spørgsmålene om beboernes holdning til miljø og energi er stillet på følgende måde i spørgeskemaet:

„Hvor meget er du optaget af at passe på miljøet?“ [52, s. 139]

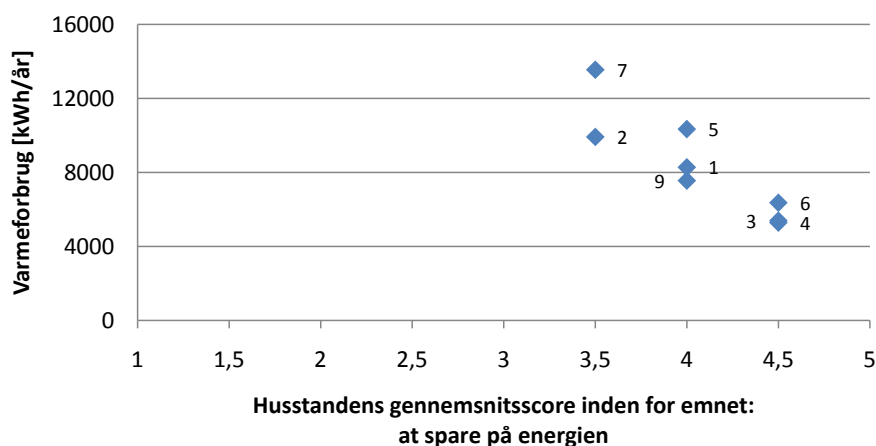
„Hvor meget er du optaget af at spare på energien?“ [52, s. 139]

hvor der svares på en femtrinsskala fra „lidt“ til „meget“. Hvad angår undersøgelsen med beboernes holdninger til miljøet, fremgår der intet tydeligt mønster mellem dette og varmemeforbruget, hvilket fremgår af figur 5.14. Anderledes ser det ud med beboernes holdning til energibesparelser, som det er illustreret på figur 5.15. Af denne figur fremgår en relativ tydelig sammenhæng mellem beboernes besvarelser og varmemeforbrug. Sammenhængen viser, at jo mere optaget beboerne er af at spare på energien, jo lavere varmemeforbrug er der i de pågældende husstande. Det skal i denne sammenhæng bemærkes, at besvarelsenerne kun er fra otte husstande, hvorfor undersøgelsen bør udføres på et større antal husstande. En lignende analyse er udført af *Kirsten Gram-Hanssen* [40], hvor det fremgår, at beboerne forbinder miljøet med vand og affald. Det samme kan være tilfældet for beboerne i *Fremtidens Parcelhuse*, hvorfor der ikke fremgår en tydelig sammenhæng i figur 5.14. Ud fra spørgeskemabesvarelsenerne er der defineret tre forskellige forbrugstyper (små-, middel- og storforbrugere) hvor deres holdning til deres forbrug er meget varierende. Der blev registreret både små- og storforbrugere, der mener, det er vigtigt at spare på energien af hensyn til miljøet og ikke økonomien. Der er andre små- og storforbrugere, der havde valgt deres adfærd på grund af muligheden for at spare penge. En tredje gruppe forbrugere bekymrede sig stort set ikke om hverken miljøet eller økonomien i den sammenhæng, og anvendte de mængder, de mente var passende for at opretholde deres levestandard. *Kirsten Gram-Hanssen* fandt en klar tendens til, at beboerne i husstande med et lavt forbrug ikke kan lide at bruge mere end højst nødvendigt, og at spare på energien er for disse beboere en bekræftelse af deres livsværdier.

Ikke desto mindre giver resultaterne af denne undersøgelse anledning til, at en tilsvarende undersøgelse blandt et større antal husstande kan være interessant, for at undersøge hvorvidt en tilsvarende sammenhæng fremgår. I den forbindelse kan det anbefales at tilknytte nogle handlinger til de enkelte svarmuligheder. Derved reduceres respondenternes forskellige tolkning af besvarelsesmulighederne, eksempelvis gennem fortolkning af udtrykkene „meget lidt“ i forhold til „lidt“. Dette bør foretages i forsøg på at opnå, at respondenterne svarer i henhold til deres reelle handlinger, og ikke ud fra hvad de mener er rigtigt at gøre



Figur 5.14: Skitsering af beboernes holdning til at passe på miljøet sammenholdt med varmemeforbruget i de pågældende husstande.



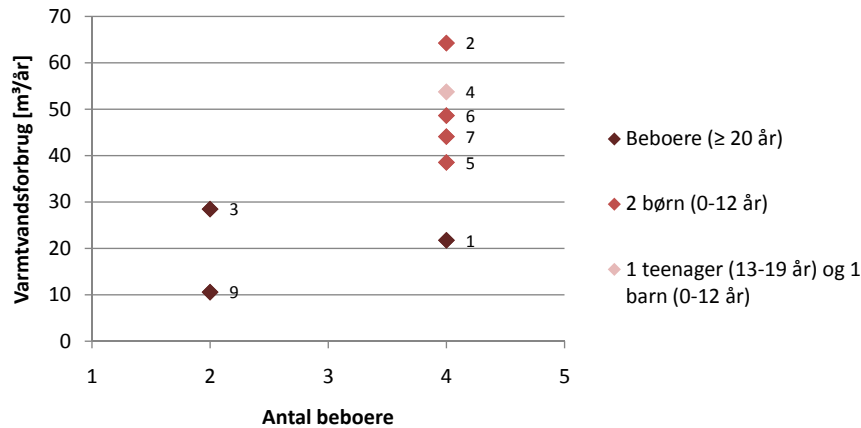
Figur 5.15: Skitsering af beboernes holdning til at spare på energien sammenholdt med varmekonsumet i de pågældende husstande.

5.3.1.2 Test af parametre i forhold til vandforbrug

Der er gennem litteraturstudiet i afsnit 3.4 opnået viden om hvilke forhold, der har betydning for henholdsvis det samlede vandforbrug og varmtvandsforbrugets størrelse. Desuden er der undersøgelser, der sammenholder forskellige forhold for boligerne, beboerne og deres holdninger med vandforbruget. Fra disse undersøgelser er der fremhævet nogle parametre, der kan forklare dele af den forbrugsvariation, der kan observeres mellem forskellige husstande. Der er gennem litteraturstudiet skelnet mellem projekter, der undersøger forbrugsvariationen for henholdsvis det samlede vandforbrug og varmtvandsforbruget. Årsagen til denne opdeling er, at det ikke nødvendigvis er de samme parametre, der kan forklare variationen for henholdsvis det samlede vandforbrug og varmtvandsforbruget. Eftersom det er begrænset hvilke undersøgelser, der er opnået kendskab til, som omhandler forbrugsvariationer for varmtvandsforbruget, er forklarende parametre i forbindelse med det samlede vandforbrug ligeledes afprøvet på varmtvandsforbruget i boligerne fra *Fremtidens Parcelhuse* [52]. Desuden undersøges det, om der kan identificeres nogle andre parametre end de i litteraturstudiet angivne, der kan forklare dele af forbrugsvariationen. En detaljeret undersøgelse af førnævnte data fremgår af appendiks N.2.

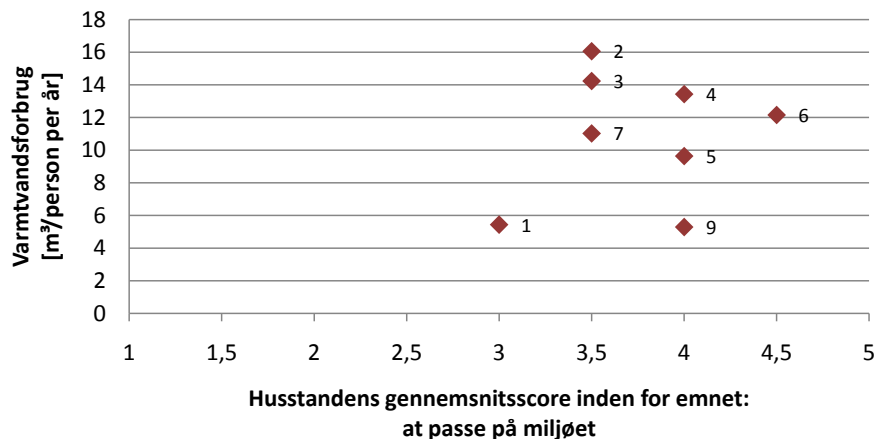
Blandt de forhold, der i litteraturstudiet har vist sig at forklare dele af variationen for husstandes samlede vandforbrug, er blandt andet antallet af beboere, bade, tøjvaske, tilstedeværelse af teenagere m.fl. Der er foretaget en undersøgelse af, hvorvidt disse forhold viser samme tendenser i det tilgængelige data, hvilket fremgår af appendiks N.2. I forhold til det tilgængelige data er der indikationer på, at der kan være en sammenhæng mellem varmtvandsforbruget og tilstedeværelsen af børn og teenagere. Som det også var tilfældet i litteraturstudiet i forhold til det samlede vandforbrug, er der indikationer på, at varmtvandsforbruget kan være højere i husstande, hvori der er børn eller teenagere, jf. figur 5.16. I litteraturstudiet blev der observeret en tendens til, ved vurdering af det samlede vandforbrug i husstande af samme størrelse, at husstande med børn tilstede havde et lavere forbrug end i husstande kun med voksne beboere hvorimod husstande med teenagere tilstede havde det højeste forbrug [42]. Dette stemmer imidlertid ikke entydigt overens med nærværende casestudie, hvor familien med fire voksne beboere har et lavere forbrug end de øvrige hus-

stande, der har samme størrelse jf. figur 5.16. Blandt årsagerne til denne afvigelse kan være, at de to voksne børns tilstedeværelse, taget deres alder i betragtning, muligvis ikke er så hyppig. Da der ikke kan konkluderes nogle sammenhænge på baggrund af otte husstande, er det et forhold, der bør analyseres nærmere i en større undersøgelse.

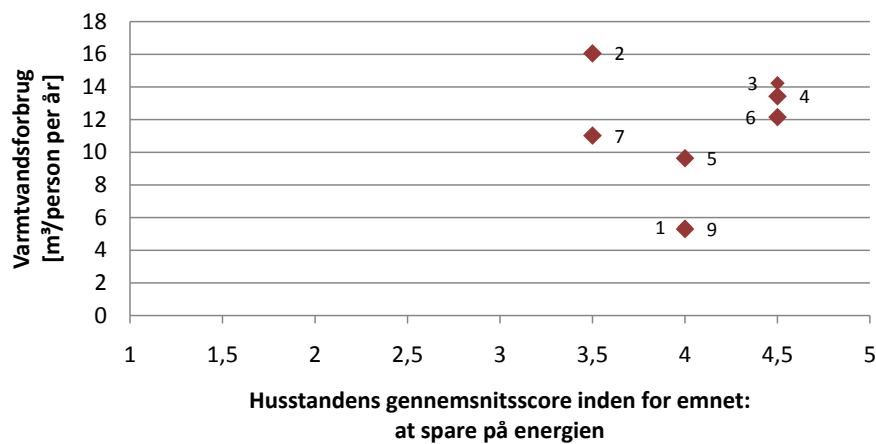


Figur 5.16: Skitsering af otte husstandes varmtvandsforbrug sammenholdt med antallet af børn og teenagere for de pågældende husstande.

Foruden førnævnte forhold er det i det tilgængelige data muligt at sammenholde vandforbruget med henholdsvis beboernes holdninger til miljø og energibesparelser, som det ligeledes var tilfældet i det foregående. Hvad angår disse undersøgelser, viser de ikke umiddelbart sammenhæng med hverken det samlede vandforbrug eller varmtvandsforbruget, hvor sidstnævnte er illustreret på figur 5.17 og 5.18.



Figur 5.17: Illustrering af husstandenes varmtvandsforbrug opgjort per person sammenholdt med beboernes gennemsnitsscore for spørgsmålet om hvor optaget de er af at passe på miljøet.



Figur 5.18: Illustrering af husstandenes varmtvandsforbrug opgjort per person sammenholdt med beboernes gennemsnitsscore for spørgsmålet om hvor optaget de er af at spare på energien.

Der kan være flere årsager til, hvorfor der ikke fremkommer en umiddelbar sammenhæng mellem de førnævnte forhold. For det første er spørgsmålet ikke direkte rettet mod hverken vandforbruget eller varmtvandsforbruget. Det betyder, at det kan være forskelligt fra person til person, hvad der forstås ved miljø. Dette forhold er blandt andet beskrevet i en tidligere undersøgelse af *Kirsten Gram-Hanssen* [40], hvor det dog har vist sig, at personer ofte forbinder vandforbrug og affald med forhold, der har med miljø at gøre. Samtidig gives der i samme undersøgelse udtryk for, at det ikke altid er miljøholdninger, der medfører en sparsommelig adfærd. Foruden førnævnte er det samtidig subjektivt, hvor meget respondenterne selv vurderer, at de er optaget af at passe på miljøet. Besvarelsen kan se anderledes ud, hvis det er respondentens bekendte, der skal vurdere hvorledes respondentens holdninger kommer til udtryk i praksis. Hvad angår besvarelsenerne om beboernes holdning til, om de er optaget af at spare på energien, fremgår der, ligesom ovenfor, ikke en direkte sammenhæng med hverken det samlede vandforbrug eller varmtvandsforbruget.

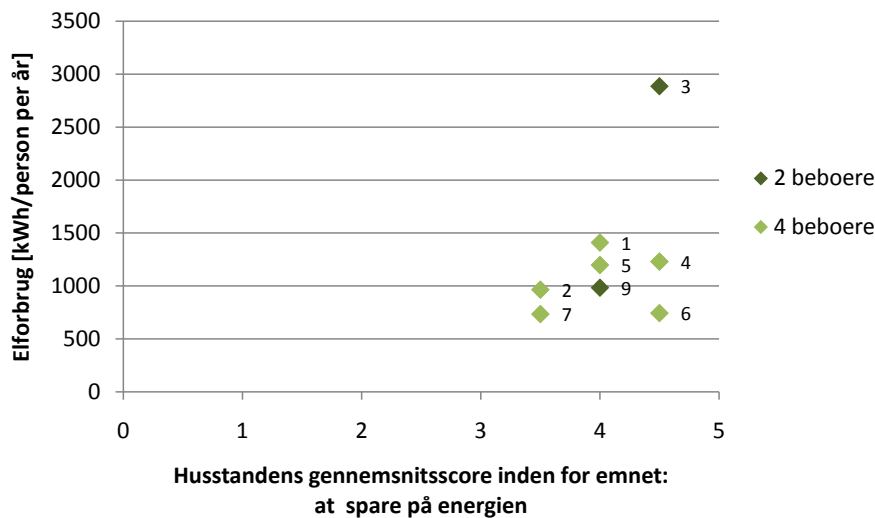
5.3.1.3 Test af parametre i forhold til elforbrug

Der er gennem litteraturstudiet i afsnit 3.3 opnået viden om hvilke forhold, der har betydning for elforbrugets størrelse. På tilsvarende måde som i de foregående afsnit er viden fra litteraturstudiet sammenholdt med forskellige forhold for boligerne, beboerne og deres holdninger med elforbruget. Fra disse undersøgelser er der fremhævet nogle parametre, der kan forklare dele af den forbrugsvariation, der kan observeres mellem forskellige husstande. Blandt de forhold, der i litteraturstudiet har vist sig at forklare dele af variationen for husstandes elforbrug, er blandt andet antallet af beboere, boligens størrelse, indkomst, beboernes alder, tilstedeværelse af teenagere, bestanden af elektriske apparater m.fl. Der er foretaget en undersøgelse af, hvorvidt disse forhold viser samme tendenser i det tilgængelige data, som det gjorde i litteraturstudiet, hvilket fremgår af appendiks N.3. Blandt de undersøgte forhold kan der i det tilgængelige data fra *Fremtidens Parcelhuse* [52] genfindes nogle tendenser mellem elforbruget og henholdsvis antal beboere samt tilstedeværelse af børn og teenagere. For de øvrige forhold, så som boligareal, beboernes alder, uddannelsesniveau, vask/tørring af tøj,

anvendelse af elbesparende foranstaltninger og opholdstid, er der ikke fundet nogle entydige sammenhænge med elforbrugets størrelse.

I nærværende afsnit gennemgås resultaterne for hvilke andre forhold, end de nævnte i litteraturstudiet, der kan indikere en sammenhæng med elforbruget. Dette udføres med henblik på at fremhæve hvilke undersøger, der kan udføres i fremtidige projekter for at opnå større viden om beboernes indflydelse på elforbrugets størrelse.

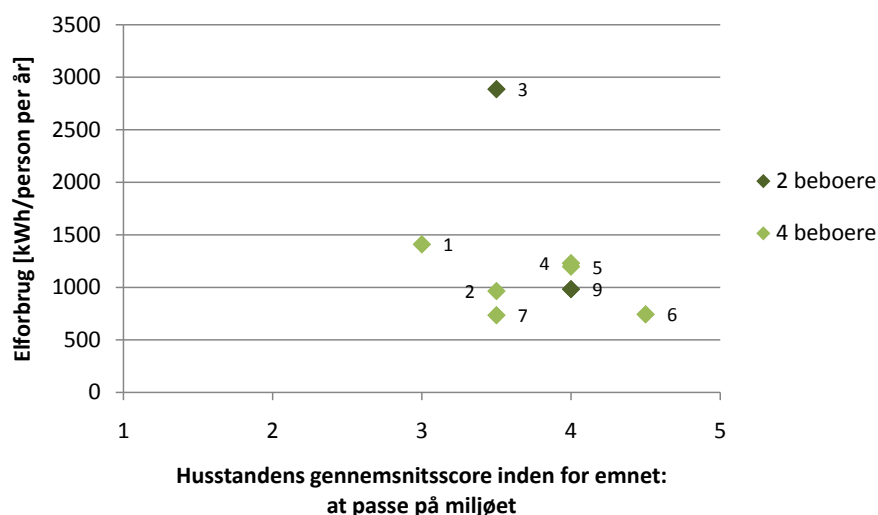
Som det er tilfældet i de foregående afsnit, er det muligt at sammenholde elforbruget med henholdsvis beboernes holdninger til miljø og energibesparelser. Analysen omhandlende beboernes holdning til energibesparelser, viser ingen umiddelbar sammenhæng til elforbrugets størrelse, hvilket fremgår af figur 5.20.



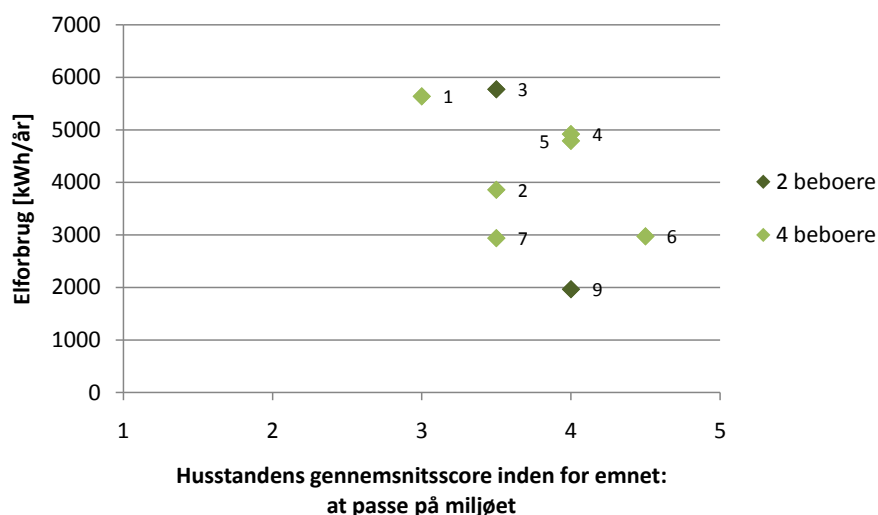
Figur 5.19: Illustrering af elforbruget sammenholdt med beboernes gennemsnitsscore for spørgsmålet om hvor optaget de er af at spare på energien.

Med hensyn til sammenhængen mellem beboernes holdninger til miljø og elforbruget er det i casestudiet erfaret, at sammenhængen kan evalueres på to forskellige måder. Forskellen på de to analyser er hvilken enhed elforbruget evalueres efter. De to enheder der er evalueret, er $kWh/person\ per\ år$ og $kWh/år$. Resultaterne af analysen fremgår af figur 5.19 og 5.21. I figur 5.19 fremgår ingen tydelig tendens, hvor specielt *hus 3* skiller sig ud med et højt forbrug⁴. Hvis *hus 3* udelukkes fremgår en svag tendens til, at elforbruget er svagt faldende ved stigende miljøscore. Idet elforbruget i *hus 3* ikke er urealistisk højt, kan intet konkluderes ud fra dette casestudie, hvorfor en nærmere analyse på større datamængder kan være interessant. I figur 5.21 er sammenhængen vurderet ud fra enheden $kWh/år$, hvor der fremgår en svag sammenhæng mellem beboernes holdning til miljøet og det samlede årlige elforbrug i husstanden. Det samme forbehold som for den foregående analyse er gældende i dette tilfælde, hvorfor en nærmere analyse på større datamængder tilsvarende kan være interessant.

⁴Elforbruget i *hus 3* er sammenholdt med grænserne fra *Energi Nord* [64], hvor det tildeles kategorien „højt forbrug“.



Figur 5.20: Illustrering af elforbruget sammenholdt med beboernes gennemsnitsscore for spørgsmålet om hvor optaget de er af at passe på miljøet, når elforbruget opgøres ud fra enheden $kWh/person\ per\ år$.



Figur 5.21: Illustrering elforbruget sammenholdt med beboernes gennemsnitsscore for spørgsmålet om hvor optaget de er af at spare på energien, når elforbruget opgøres ud fra enheden $kWh/år$.

5.3.2 Beskrivelse af attributter til klassificeringstræerne

Med henblik på at eksemplificere udviklingen og anvendelsen af klassificeringstræer inden for emnet brugeradfærd, er det nødvendigt at indsamle data, der kan anvendes til at beskrive sammenhængen mellem beboernes adfærd og de udvalgte klasser for de enkelte klassificeringstræer. Det nødvendige data afhænger af, hvad klassificeringstræet skal beskrive. Formålet med nærværende afsnit er at eksemplificere hvilket data, der vurderes at være anvendelig ved udvikling af klassificeringstræer. I nærværende projekt er målet med klassificeringstræer-

ne, som tidligere beskrevet, at forberede udviklingen af tre forskellige klassificeringstræer og eksemplificere udviklingen af ét klassificeringstræerne i kapitel 7. I tabel 5.6 til 5.8 fremgår det hvilke attributter, tilknyttet enten teknisk, demografisk eller adfærdsmæssig relevans, der ud fra litteratur- og casestudiet har indikeret at have en sammenhæng med energiforbruget til rumopvarmning, varmtvandsforbruget eller elforbruget. Disse attributter skal imidlertid ikke betragtes som en begrænsning for hvilke forhold, der er mulige at medtage i klassificeringstræerne. I appendiks O fremgår en uddybende beskrivelse af hvad de enkelte betegnelser i de førnævnte tabeller dækker over, og hvordan oplysninger om disse kan indsamles.

Tekniske attributter	Rumopvarmning	Varmtvandsforbrug	Elforbrug
Boligareal	•	•	•
Boligens varmetekniske egenskaber	•		
Boligens opførelsesår	•		
Infiltration	•		
Styring af mekanisk ventilation	•		•

Tabel 5.6: Angivelse af hvilke tekniske attributter der er relevante for de tre beskrevne klassificeringstræer på baggrund af erfaringer fra litteratur- og casestudie.

Demografiske attributter	Rumopvarmning	Varmtvandsforbrug	Elforbrug
Antal beboere		•	•
Beboernes alder	•	•	•
Køn			•
Indtægt	•		•

Tabel 5.7: Angivelse af hvilke demografiske attributter der er relevante for de tre beskrevne klassificeringstræer på baggrund af erfaringer fra litteratur- og casestudie.

5.3.2. Beskrivelse af attributter til klassificeringstræerne

Adfærdsrelaterede attributter	Rumopvarmning	Varmtvandsforbrug	Elforbrug
Beboernes opholdstid i boligen	•	•	•
Holdninger til miljø, energi og økonomi	•	•	•
Viden om størrelsen af forbrug	•	•	•
Ønsket lufttemperatur	•		
Justering af termostater	•		
Lukke af til ubenyttede rum	•		
Vaner mht. udluftning	•		
Adfærd hvis det opfattes for koldt	•		
Vaner i forbindelse med hygiejne		•	
Vaner i forbindelse med opvask		•	
Antal og brugstid af elapparater			•
Elapparaternes alder			•
Standbyforbrug			•
Slukke lys i ubenyttede rum			•

Tabel 5.8: Angivelse af hvilke adfærdsrelaterede attributter der er relevante for de tre beskrevne klassificeringstræer på baggrund af erfaringer fra litteratur- og casestudie.

Flere af forholdene, der fremgår af de foregående tabeller, kan vurderes både på baggrund af udsagn fra beboere, eller der kan foretages målinger i boligerne. Der er fordele og ulemper ved begge muligheder, som bør tages i betragtning, inden data til klassificeringstræerne indsamles. Såfremt der søges informationer om de enkelte forhold gennem en spørgeundersøgelse blandt beboerne, er besvarelsener subjektivt præget, og der kan være afvigelser mellem besvarelsener og de faktiske forhold. Omvendt kan det skabe problemer, hvis nogle af forholdene, der indgår i klassificeringstræerne, skal undersøges ved brug af måleudstyr. Problemstillingerne opstår i, at når klassificeringstræerne skal anvendes til vurdering af et energiforbrug, kan denne estimering muligvis ikke foretages, uden der foreligger måleresultater for en bolig. Det betyder, at ønsket om et simpelt, brugervenligt og hurtigt værktøj kan falde til jorden. Af denne årsag bør det overvejes, hvorvidt de enkelte forhold skal undersøges gennem besvarelsener fra beboere eller fra målinger, inden der indsamles data til opstilling af klassificeringstræer.

Forhold som boligens orientering, skyggeforhold fra omgivelser, solafskærmning, vinduers g-værdi og andre tekniske forhold har indflydelse på energiforbruget til rumopvarmning. På trods af de forskellige forholds betydning for energiforbruget, er det i det nærværende projekt valgt, at disse forhold ikke medtages i opbygningen af klassificeringstræerne. Årsagen til dette valg skal findes i det overordnede mål med opstilling af klassificeringstræerne i dette projekt. Målet er, som tidligere beskrevet, at udvikle et værktøj, der er let anvendeligt og relativt hurtigt kan estimere et forventet energiforbrug. Det betyder, at der på nogle områder må gås på kompromis med detaljeringsgraden af informationerne, som klassificeringstræerne opbygges på baggrund af. Det er en balancegang mellem at opnå nøjagtige estimeringer af energiforbruget og ende ud med et simpelt, brugervenligt og hurtigt klassificeringstræ, hvor fordele og ulemper må opvejes, når det skal vurderes, om informationer skal medtages eller ej.

Generering af data

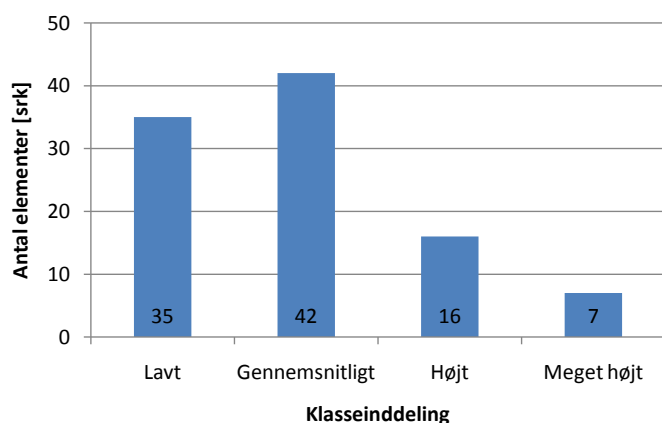
På baggrund af beskrivelserne af klasser og attributter i det foregående kapitel er det i nærværende projekt fastlagt, at der eksemplificeres udviklingen af et klassificeringstræ til estimering af elforbruget i boliger. I klassificeringstræet skal der indgå oplysninger om blandt andet demografiske forhold for beboerne samt informationer om deres adfærd. Der er to forhold, der har en medvirkende årsag til, at elforbruget er valgt til eksemplificering af udviklingsprocessen. For det første er det gennem litteraturstudiet erfaret, at oplysninger om attributter og størrelsen af elforbruget er tilgængeligt i et større omfang end for de øvrige forbrug. Den anden og vigtigste årsag til udvælgelsen af elforbruget som klasse er, at det er interessant at eksemplificere udviklingen af en relativt simpel model, der kan øge fokus på elforbruget i boliger. Ifølge *Statens Byggeforskningsinstitut* [56] er der behov for et øget fokus på elforbruget i boliger, såfremt den politiske målsætning om at reducere nye bygningers energiforbrug med 75 % skal nås. For at kunne eksemplificere udviklingen af et klassificeringstræ er der behov for data. I nærværende projekt er alt data fiktivt, men genereret på baggrund af erfaringer fra litteratur- og casestudiet. Elforbrugets størrelse er fastlagt på baggrund af sandsynlighedsfordelingen, der er udviklet og beskrevet i appendiks K. I de efterfølgende afsnit beskrives det genererede data i forhold til hvilke forudsætninger der er foretaget under datagenereringsprocessen¹.

Der genereres, i nærværende projekt, data for 100 fiktive husstande. Mængden af data er fastlagt for at illustrere, at der kan udvikles et klassificeringstræ uden behov for tusindvis af målinger. Dette åbner dermed op for, at problemstillingen omhandlende brugeradfærd i boliger, kan undersøges, uden der er behov for omfattende måleprojekter. Det er gennem litteraturstudiet erfaret, at antallet af beboere i boligen forklarer en del af variationen i elforbruget. Denne iagttagelse er anvendt i kapitel 5, hvor det er fastlagt, at der udvikles ét klassificeringstræ for hver husstandsstørrelse. Nærmere argumentation om dette fremgår af førnævnte kapitel. Idet metoden for udviklingen af klassificeringstræet er ens uanset husstandsstørrelsen, er det i nærværende projekt valgt at generere data, og efterfølgende udvikle et klassificeringstræ, for husstande med fire beboere.

¹Det genererede data, med tilhørende kildeangivelse, der anvendes til at eksemplificere udviklingen af klassificeringstræet, fremgår af følgende regneark: *Klassificeringstræ | Generering af data for elforbrug.xlsx* på den medfølgende DVD.

6.1 Beskrivelse af datagenerering for klassificeringstræets klasse

Genereringen af data til elforbruget sker, som tidligere omtalt, på baggrund af sandsynlighedsfordelingen, der er udviklet og beskrevet i appendiks K. Elforbruget er beskrevet ved en lognormalfordeling, hvor forudsætningerne for denne fremgår af tabel 6.1. Intervallet for det genererede data spænder imellem 2245–10227 $kWh/år$ med et gennemsnit på 5142 $kWh/år$. Det genererede data er inddelt efter de klasseinddelinger, der er fastlagt i afsnit 5.2.2.3, hvor intervallerne for disse er gengivet i tabel 6.1. Det genererede data inddeles i de respektive klasseinddelinger, hvor figur 6.1 illustrerer et histogram over det genererede data.



Figur 6.1: Histogram over genereret data for klassen med udgangspunkt i en lognormalfordeling.

Forudsætning	Beskrivelse [$kWh/år$]
Middelværdi	5035
Standardafvigelse	1540
Lavt	0 – 4426
Gennemsnit	4427 – 6187
Højt	6188 – 8200
Meget højt	> 8200

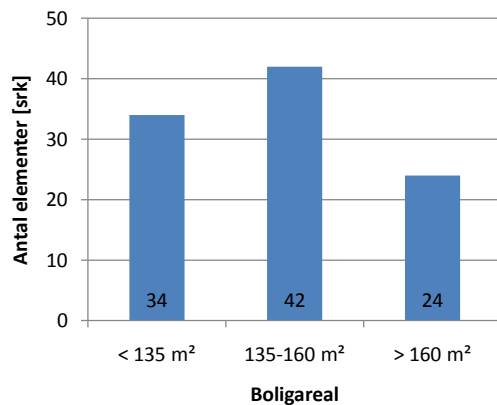
Tabel 6.1: Forudsætninger for generering af data for elforbruget og klassens inddeling.

6.1.1 Beskrivelse af datagenerering for boligareal

Det fiktive data for boligarealet er genereret på data fra *Danmarks Statistik*², med de forudsætninger der er oplistet i tabel 6.2. Informationerne fra *Danmarks Statistik* beskrives ved hjælp af en tilnærmet normalfordeling, hvor forudsætningerne for denne fremgår af tabel 6.2. Ved generering af data er boligarealet gjort afhængig af elforbruget, grundet den erfarne sammenhæng i litteraturstudiet i kapitel 3, således at stigende boligareal resulterer i et stigende elforbrug. Denne sammenhæng er beskrevet ved at anvende den opstillede normalfordeling og angive elforbrugets størrelse som betingelse for, om data skal genereres ud fra de nederste eller øverste del af normalfordelingen.

Intervallet for det genererede data spænder imellem 53 – 272 m^2 med et gennemsnit på 143,6 m^2 . Idet der er mange mulige udfald for attributten er det valgt, at anvende attributten som kategorisk med tre mulige udfald: < 135 m^2 , 135 – 160 m^2 , > 160 m^2 . På figur 6.2 fremgår et histogram over det genererede data for boligarealet.

²Følgende link: <http://www.statistikbanken.dk/BOL511>



Figur 6.2: Histogram over genereret data for boligareal med udgangspunkt i en normalfordeling.

Forudsætning	Beskrivelse
Middelværdi	147 m ²
Standardafvigelse	34 m ²
Forudsætninger til Danmarks Statistik	
Område	Hele landet
Boligtype	Parcelhuse
Boligstørrelse	Alle undtagen „uoplyst“
År	2010

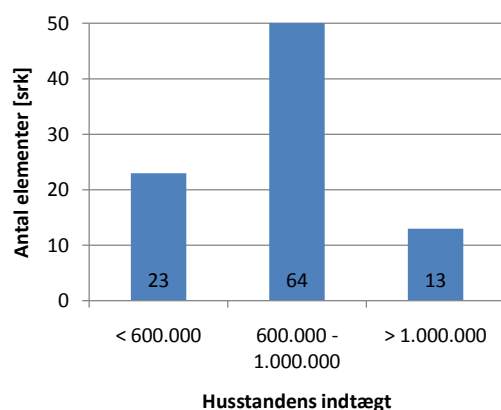
Tabel 6.2: Forudsætninger for generering af data for boligarealet.

6.1.2 Beskrivelse af datagenerering for husstandens indkomst

Det fiktive data for husstandens indkomst er genereret på baggrund af data fra *Danmarks Statistik*³, med de forudsætninger der er oplistet i tabel 6.3. Informationerne fra *Danmarks Statistik* beskrives ved hjælp af en tilnærmet normalfordeling, hvor forudsætningerne for denne fremgår af tabel 6.3. Ved genereringen af data er indkomsten gjort afhængig af elforbruget, grundet den erfarede sammenhæng i litteraturstudiet i kapitel 3, således at stigende indkomst resulterer i et stigende elforbrug. Denne sammenhæng er beskrevet ved at anvende den opstillede normalfordeling, og angive elforbrugets størrelse som betingelse for om data skal genereres ud fra de nederste eller øverste del af normalfordelingen.

Intervalleret for det genererede data spænder imellem 150.836 – 1.291.337 kr. med et gennemsnit på 737.666 kr.. Idet der er mange mulige udfald for attributten, er det valgt at anvende attributten som kategorisk med tre mulige udfald: <600.000, 600.000-1.000.000, >1.000.000. På figur 6.3 fremgår et histogram over det genererede data for husstandens indkomst.

³Følgende link: <http://www.statistikbanken.dk/INDKF2>



Figur 6.3: Histogram over genereret data for husstandens indkomst med udgangspunkt i en normalfordeling.

Forudsætning	Beskrivelse
Middelværdi	749.378 <i>kr.</i>
Standardafvigelse	206.540 <i>kr.</i>
Forudsætninger til Danmarks Statistik	
Område	Hele landet
Enhed	Antal familier i gruppen
Ejer/lejer	Beboet af ejer
Familietype	Familier med børn
Indkomstinterval	Alle intervaller
År	2009

Tabel 6.3: Forudsætninger for generering af data for husstandens indkomst.

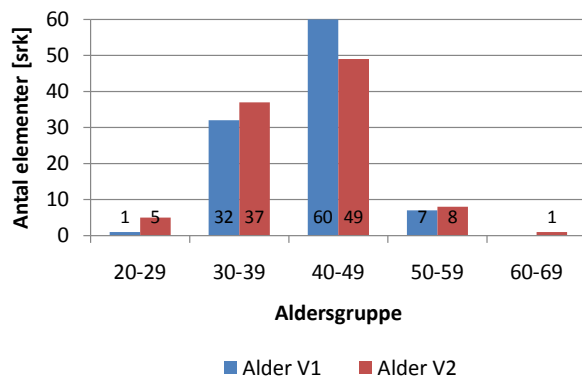
6.1.3 Beskrivelse af datagenerering for beboernes alder

Det fiktive data for beboernes alder er genereret på forskellige måder. I dette afsnit gennemgås forudsætningerne for generering af beboernes alder for både voksne og børn. Alderen på den ene voksne beboer (V1) er genereret på baggrund af data fra *Danmarks Statistik*⁴, med de forudsætninger der er oplistet i tabel 6.4⁵. Informationerne fra *Danmarks Statistik* beskrives ved hjælp af en tilnærmet normalfordeling, hvor forudsætningerne for denne fremgår af tabel 6.4. Intervallet for det genererede data spænder imellem 29 – 56 *år* med et gennemsnit på 41 *år*.

Alderen på den anden voksne (V2) genereres ud fra en normalfordeling. Middelværdien for denne er den genererede alder på voksen V1, og standardafvigelsen er fastlagt til at være tre *år*. Intervallet for det genererede data spænder imellem 22 – 61 *år* med et gennemsnit på 41 *år*. På figur 6.4 fremgår et histogram over det genererede data for alderen på de to voksne beboere. Idet der er mange mulige udfald for denne attribut, anvendes der en kategorisk inddeling med intervaller på 10 *år*.

⁴Følgende link: <http://www.statistikbanken.dk/FAM122N>

⁵De ti udvalgte kommuner, der er omtalt i tabellen, er: København, Brøndby, Faaborg-Midtfyn, Odense, Silkeborg, Aarhus, Ringkøbing-skjern, Struer, Jammerbugt og Aalborg.



Figur 6.4: Histogram over genereret data for alderen på de to voksne beboere.

Forudsætning	Beskrivelse
Middelværdi	41,34 år
Standardafvigelse	7,58 år
Forudsætninger til Danmarks Statistik	
Område	Ti udvalgte kommuner
Husstandstype	Ægtepar og par iøvrigt
Husstansstørrelse	Fire beboere
Antal børn	To
Køn	Mænd
Alder	20 – 70 år
År	2008

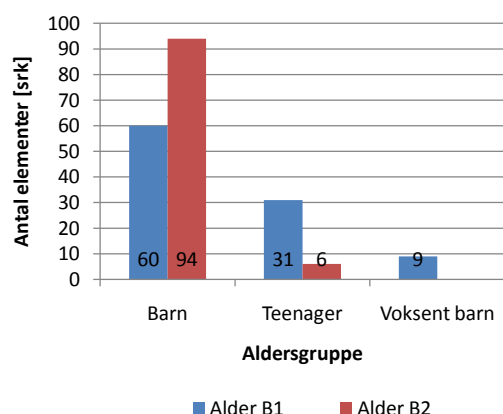
Tabel 6.4: Forudsætninger for generering af data for alderen på voksen V1.

Alderen på det ældste barn (B1) afhænger af alderen på den yngste af de voksne beboere. Det fremgår af tabel 6.5 hvilke forudsætninger der er anvendt, hvor barnet tildeles en tilfældig værdi inden for de angivne intervaller. Intervallet for det genererede data spænder imellem 1 – 22 år med et gennemsnit på 11 år. Børnenes alder er efterfølgende inddelt i følgende kategorier; *barn* (0–12 år), *teenager* (13–19 år) eller *voksant barn* (>19 år). På figur 6.5 fremgår et histogram over det genererede data for alderen på de to børn i husstande.

Alderen på det yngste barn (B2) afhænger af alderen på det ældste barn. Alderen genereres ud fra følgende forudsætninger: hvis det ældste barn er 1 år tildeles det yngste barn en alder mellem 0 – 1 år. Hvis det ældste barn er ældre end 1 år, tildeles det yngste barn en tilfældig alder mellem 1 år og det ældste barns alder fratrukket 1 år, se formel 6.1. Intervallet for det genererede data spænder imellem 0 – 18 år med et gennemsnit på 5 år.

$$1 \leq \text{alder barn } B2 \leq \text{Alder barn } B1 - 1\text{år} \quad (6.1)$$

6.1.4. Beskrivelse af datagenerering for beboernes forventede elforbrug



Figur 6.5: Histogram over genereret data for anden på de to børn.

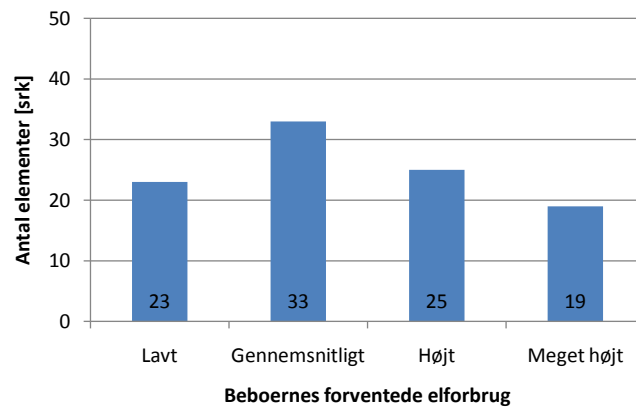
Alder, yngste voksne (YV)	Alder barn B1
[år]	[år]
$18 \leq YV < 26$	1 – 2
$26 \leq YV < 31$	1 – 8
$31 \leq YV < 36$	1 – 13
$36 \leq YV < 41$	2 – 18
$41 \leq YV < 46$	6 – 22
$46 \leq YV < 51$	10 – 22
$51 \leq YV < 56$	14 – 22
$56 \leq YV < 61$	18 – 22
$61 \leq YV < 66$	20 – 22

Tabel 6.5: Forudsætninger for generering af data alderen på barn B1.

6.1.4 Beskrivelse af datagenerering for beboernes forventede elforbrug

Denne attribut omhandler hvilket elforbrug beboerne vurderer, at de har. Informationerne beskrives ved hjælp af en tilnærmet normalfordeling, med elforbruget som middelværdi og en spredning på 2000 kWh/år . Ved at benytte elforbrugets værdi som middelværdi for normalfordelingen er de to faktorer korreleret, og der er derigennem sikret en sammenhæng imellem elforbrugets reelle størrelse, og det elforbrug beboerne vurderer, at de har. Sammenhængen er beskrevet ved at anvende den opstillede normalfordeling og angive elforbrugets størrelse som betingelse for, om data skal genereres ud fra de nederste eller øverste del af normalfordelingen.

Intervallerne for det genererede data spænder imellem $1527\text{--}14440 \text{ kWh/år}$ med et gennemsnit på 6127 kWh/år . Afvigelsen mellem beboernes bud på størrelsen af deres elforbrug og det reelle forbrug varierer mellem $1\text{--}144\%$ med et gennemsnit på omkin 28% . Idet der er mange mulige udfald for attributten, er det valgt, at angive denne som kategorisk med fire mulige udfald: „lavt“, „gennemsnit“, „højt“ og „meget højt“. Dette svarer til den inddeling, der anvendes for klassen, hvor de samme intervaller anvendes for denne attribut, se evt. tabel 6.1. På figur 6.6 fremgår et histogram over det genererede data for beboernes forventede elforbrug.



Figur 6.6: Genereret data for beboernes forventede elforbrug.

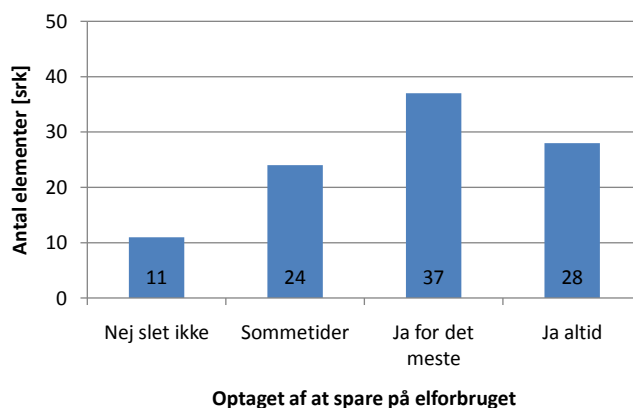
6.1.5 Beskrivelse af datagenerering for beboernes opmærksomhed på elforbruget

Denne attribut angiver hvorvidt beboerne er optaget af at spare på elforbruget, og er tildelt fire besvarelsesmuligheder: „ja altid“, „ja for det meste“, „sommetider“ og „nej slet ikke“. Hvis der skal indsamles informationer om denne attribut i praksis, er der to forskellige muligheder for, hvordan dette kan udføres. Den ene mulighed er at lade det fremgå som et holdningsspørgsmål, hvor beboerne selv skal fortolke dette, og svare ud fra deres mening hvorvidt de er optaget af at spare på elforbruget. Dette kan imidlertid give anledning til nogle problemstillinger i forbindelse med analysering af data, da spørgsmålet er underlagt en subjektiv fortolkning fra beboernes side. En anden mulighed er, at tilknytte nogle aktiviteter til de enkelte svarmuligheder. Det kan eksempelvis være, at for svarmuligheden „ja altid“ er tilknyttet aktiviteter som at beboerne altid slukker lyset i ubenyttede rum, undgår at have elapparater med standbyforbrug og altid køber energieffektive elapparater.

I forbindelse med genereringen af data er denne attribut korreleret med elforbruget, hvilket er udført for at opnå det mest optimale data at arbejde med under udviklingen af klassificeringstræet. Forudsætningerne for genereringen af data fremgår af tabel 6.6. På figur 6.7 fremgår et histogram over det genererede data for om beboerne er optaget af at spare på elforbruget.

Klassen for elforbrug	Svarmulighed for beboerne
Lavt	ja altid / ja for det meste
Gennemsnit	ja altid / ja for det meste / sommetider
Højt	ja for det meste / sommetider / nej slet ikke
Meget højt	sommetider / nej slet ikke

Tabel 6.6: Forudsætninger til generering af data for om beboerne er optaget af at spare på elforbruget. Den kategoriske værdi er genereret tilfældig inden for de viste muligheder.



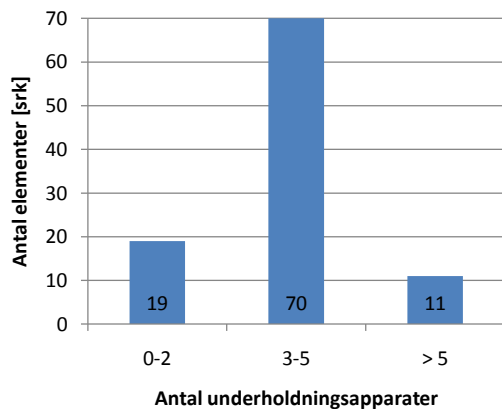
Figur 6.7: Genereret data for om beboerne er optaget af at spare på elforbruget.

6.1.6 Beskrivelse af datagenerering for mængden af underholdningsapparater

Det fiktive data for antallet af underholdningsapparater dækker over tv, computere og spillekonsoller. Data for tv og computere er genereret på baggrund af informationer fra *TNS Gallup*⁶. Disse informationer tilnærmes ved hjælp af en lognormalfordeling, der beskrives af forudsætningerne der fremgår af tabel 6.7. Ved generering af data er mængden af underholdningsapparater gjort afhængig af elforbruget, grundet den erfarede sammenhæng i litteraturstudiet i kapitel 3, således at stigende antal apparater resulterer i et stigende elforbrug.

Antallet af spillekonsoller er genereret som et tilfældigt tal mellem 0 – 2 stk. Data for alle tre typer af apparater summeres, hvor det generede data har et gennemsnit på 3,8 stk. Idet der er mange mulige udfald for attributten er det valgt, at anvende attributten som kategorisk med tre mulige udfald: 0 – 2, 3 – 5 og > 5. På figur 6.8 fremgår et histogram over det generede data for antallet af underholdningsapparater.

⁶Følgende link: <http://ida.dk/netvaerk/idaforum/u0631a/Documents/Facts%20and%20Figures,%2012.%20maj%202009/TNS%20Gallup.pdf>



Figur 6.8: Histogram over genereret data for antallet af underholdningsapparater.

Forudsætning	Beskrivelse
Middelværdi	1,43 <i>stk</i>
Standardafvigelse	1,38 <i>stk</i>

Tabel 6.7: Forudsætninger for generering af data for antallet af henholdsvis tv og computere.

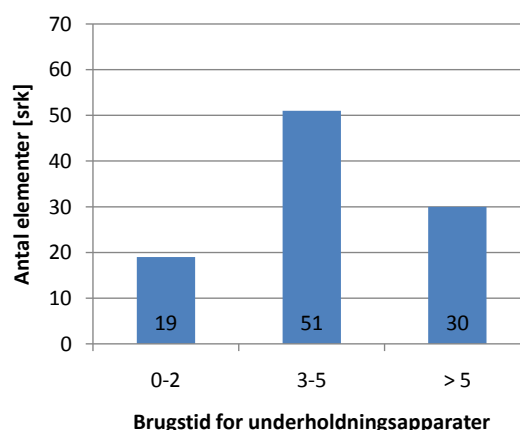
6.1.7 Beskrivelse af datagenerering for brugstid af underholdningsapparater

Det fiktive data for brugstiden af underholdningsapparater er genereret på baggrund af data fra *ELMODEL-Bolig* [80]. Oplysningerne fra *ELMODEL-Bolig* er opgivet som middelværdier og standardafvigelser for brugstiden af både tv og computere. Det antages, at brugstiden for disse apparater kan beskrives ved en normalfordeling med de oplysninger, der fremgår af tabel 6.8, samt at denne brugstid tilsvarende omfatter anvendelsen af spillekonsoller.

Ved generering af data er brugstiden af underholdningsapparater gjort afhængig med elforbruget, hvor der tages hensyn til, at hvis mængden af underholdningsapparater i boligen er lig 0, er brugstiden af disse tilsvarende nul. Sammenhængen mellem brugstiden af underholdningsapparater og elforbruget er beskrevet ved at anvende den opstillede normalfordeling, og angive elforbrugets størrelse som betingelse for om data skal genereres ud fra de nederste eller øverste del af normalfordelingen. Intervallet for det genererede data spænder imellem 2 – 19 timer med et gennemsnit på 8,6 timer. Det bør i denne sammenhæng bemærkes, at brugstiden omfatter samtlige underholdningsapparater for hele husstanden. Idet der er mange mulige udfald for attributten, er det valgt at angive attributten som kategorisk med tre mulige udfald: 0 – 5, 5 – 10, > 10. På figur 6.9 fremgår et histogram over det genererede data for brugstiden af underholdningsapparater.

Forudsætning	tv		pc	
	Hverdag	Weekend	Hverdag	Weekend
Middelværdi*	5,02	7,15	2,46	4,12
Standardafvigelse*	2,51	4,00	3,42	5,96

Tabel 6.8: Forudsætninger til generering af data for brugstiden af underholdningsapparater.



Figur 6.9: Genereret data for brugstiden af underholdningsapparater.

6.1.8 Beskrivelse af datagenerering for brug af husholdningsapparater

Det forudsættes, at alle 100 fiktive boliger er i besiddelse af husholdningsapparaterne vaskemaskine, tørretumbler og opvaskemaskine, hvorfor der kun genereres data for anvendelsen af disse. Det fiktive data for den ugentlige anvendelse af de omtalte husholdningsapparater er genereret på baggrund af data fra *ELMODEL-Bolig* [80]. Oplysningerne fra *ELMODEL-Bolig* er opgivet som middelværdier og standardafvigelser for antallet af gange husholdningsapparaterne anvendes ugentligt. Det antages, at anvendelsen af husholdningsapparater kan beskrives ved en normalfordeling med de oplysninger, der fremgår af tabel 6.9.

Forudsætning	Antal tøjvaske	Antal tørretumblinger	Antal opvaske
Middelværdi*	4, 11	3, 25	5, 18
Standardafvigelse*	2, 16	2, 27	2, 17

Tabel 6.9: Forudsætninger til generering af data for den ugentlige anvendelsen af husholdningsapparaterne.

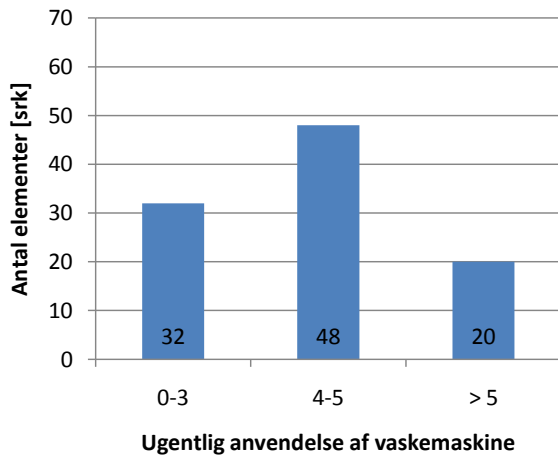
Ved generering af data er brugstiden af husholdningsapparaterne korreleret med elforbruget, således at stigende anvendelse af disse resulterer i et stigende elforbrug. Sammenhængen er beskrevet ved at anvende den opstillede normalfordeling, og angive elforbrugets størrelse som betingelse for, om data skal genereres ud fra de nederste eller øverste del af normalfordelingen.

Intervallet for det genererede data spænder for den ugentlige anvendelse af vaskemaskine imellem: 1 – 9 gange med et gennemsnit på 4, 1 gange. Den ugentlige anvendelse af tørretumbler varierer mellem: 0 – 9 gange med et gennemsnit på 3, 4 gange. Med hensyn til den ugentlige anvendelse af opvaskemaskine varierer det imellem 2 – 9 gange med 5, 3 gange som gennemsnit. Idet der er mange mulige udfald for de tre nævnte attributter er det valgt at anvende attributten som kategorisk med hver tre mulige udfald:

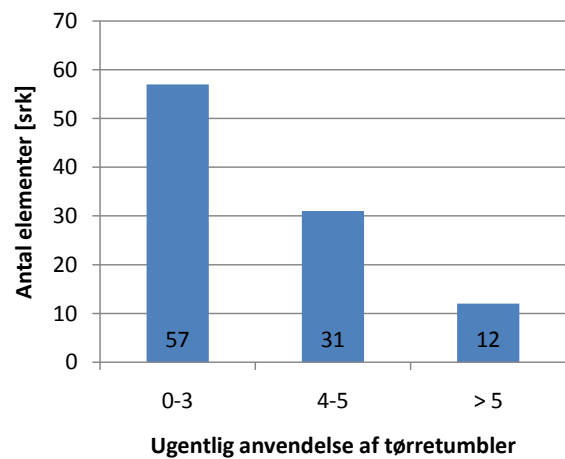
- Kategorisk inddeling for den ugentlige anvendelse af vaskemaskine: 0 – 3, 4 – 5, > 5.
- Kategorisk inddeling for den ugentlige anvendelse af tørretumbler: 0 – 3, 4 – 5, > 5.

- Kategorisk inddeling for den ugentlige anvendelse af opvaskemaskine: 0 – 4, 5 – 7, > 7.

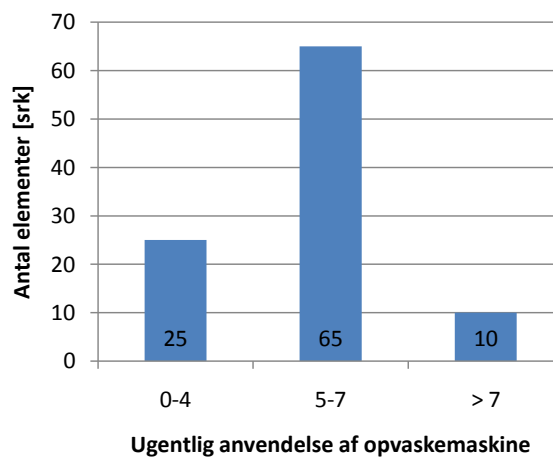
På figur 6.10 til 6.12 fremgår et histogram over det genererede data for den ugentlige anvendelse af henholdsvis vaskemaskine, tørretumbler og opvaskemaskine.



Figur 6.10: Histogram over genereret data for den ugentlige anvendelse af vaskemaskine.



Figur 6.11: Histogram over genereret data for den ugentlige anvendelse af tørretumbler.



Figur 6.12: Histogram over genereret data for den ugentlige anvendelse af opvaskemaskine.

6.1.9 Beskrivelse af datagenerering for antallet af elapparater med standbyforbrug

Det fiktive data for antallet af elapparater med et standbyforbrug er genereret på baggrund af data fra *Forsyningsvirksomhederne* [35] og *Energi Midt* [26]. Oplysningerne fra begge kilder er angivet som et typisk gennemsnit på 10 – 15 apparater, der står standby.

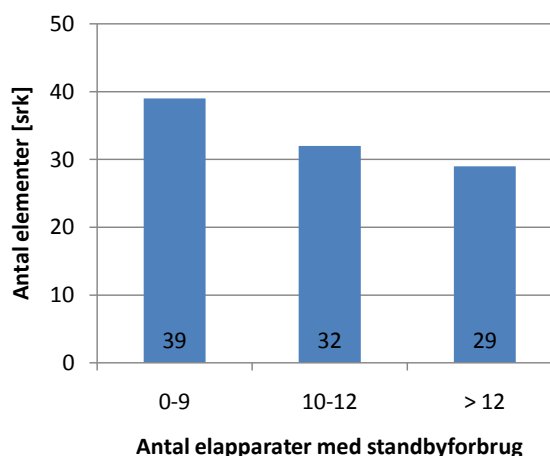
Ved generering af data er antallet af apparater med standbyforbrug korreleret med det genererede data for mængden af underholdningsapparater. Dette er foretaget således, at et

stigende antal underholdningsapparater resulterer i et stigende antal apparater med standbyforbrug. Forudsætningerne, der er anvendt ved generering af data, fremgår af tabel 6.10. Det bør i denne sammenhæng bemærkes, at antallet af elapparater med standbyforbrug ikke kun dækker over underholdningsapparater, men også apparater som opladere til mobiltelefoner, modem, radio m.fl.

Intervallerne for det genererede data spænder imellem: 5 – 18 apparater med et gennemsnit på 10,4 apparater. Idet der er mange mulige udfald for attributten er det valgt at anvende attributten som kategorisk med hver tre mulige udfald: 0 – 9, 10 – 12, > 12. På figur 6.13 fremgår et histogram over det genererede data for antallet af elapparater med standbyforbrug.

Antal underholdningsapparater	Antal apparater med standbyforbrug
≤ 3	5 – 10
≤ 6	10 – 15
> 6	15 – 20

Tabel 6.10: Forudsætninger til generering af data for antallet af apparaterne med standbyforbrug. Ved generering af data anvendes der et tilfældigt tal indenfor de oplyste intervaller.



Figur 6.13: Histogram over genereret data for antallet af elapparater med standbyforbrug.

Udvikling af klassificeringstræ

Der er i kapitel 4 beskrevet flere forskellige metoder til analysering af data, hvor disse er sammenholdt med fordele og ulemper i afsnit 4.2. På baggrund af en række iagttagelser er det i førnævnte afsnit fastlagt, at der skal anvendes et klassificeringstræ til behandling af data i dette projekt. I nærværende kapitel er der fokuseret på teorien om, hvorledes et klassificeringstræ udvikles, afprøves og anvendes i praksis. Den nærværende beskrivelse skal betragtes som en videreførelse af den overordnede teori og beskrivelse, der er præsenteret i afsnit 4.3. Til eksemplificering af teorien om klassificeringstræer, og hvorledes dette skal anvendes i praksis, vil der gennem dette kapitel anvendes det fiktive data, der er beskrevet i kapitel 6, hvor dette anses som træningsdata.

Dette kapitel er opbygget således, at der indledningsvis introduceres nogle af de adskillelige muligheder, der er for udvikling af forskellige typer klassificeringstræer. Forskellene blandt de enkelte klassificeringstræer er, at de kan udvikles på baggrund af forskellige algoritmer. Der er foretaget en undersøgelse af fordele, ulemper og anvendelsesmuligheder for forskellige algoritmer, hvor disse sammenholdes med målet for nærværende projekt. Undersøgelsen leder videre til, at der udvælges en algoritme, hvor der i det efterfølgende afsnit redegøres for hvorledes denne kan anvendes i forbindelse med udvikling og anvendelse af et klassificeringstræ. I kapitel 6, med genereringen af det fiktive data, er det fastlagt, at der tages udgangspunkt i eksemplificering af sammenhængen mellem kendetegn for beboere, deres vaner, boligens kendetegn og elforbruget. Dette valg går igen i nærværende kapitel, hvor eksemplificeringen af teorien vil ske med udgangspunkt i elforbruget som klasse. Det betyder, som det er beskrevet nærmere i afsnit 5.2, at der er mulighed for at inddele data, og dermed elforbruget, i fire forskellige klasser; „lavt“, „gennemsnitligt“, „højt“ og „meget højt“.

7.1 Algoritme til udvikling af klassificeringstræ

Fremgangsmåden for udvikling og anvendelse af et klassificeringstræ afhænger i høj grad af, hvilken algoritme der anvendes. Der findes flere forskellige algoritmer, der kan anvendes til udvikling af et klassificeringstræ. Blandt disse algoritmer er eksempelvis: ID3 [76], C4.5 [77], C5.0/See5 [78], CHAID [50], QUEST [55] og CART [12] m.fl. I appendiks M er alle de omtalte algoritmer introduceret, og der er sammenholdt fordele, ulemper og anvendelsesmuligheder for disse. Der henvises til det førnævnte appendiks for en nærmere gennemgang af algoritmerne, men på baggrund af viden fra et studie af disse, er det i det nærværende projekt valgt, at tage udgangspunkt i algoritmen C4.5, udviklet af *John Ross Quinlan* [77]. Dette valg er foretaget på baggrund af, at algoritmen kan udvikle relativt små og ukomplicerede

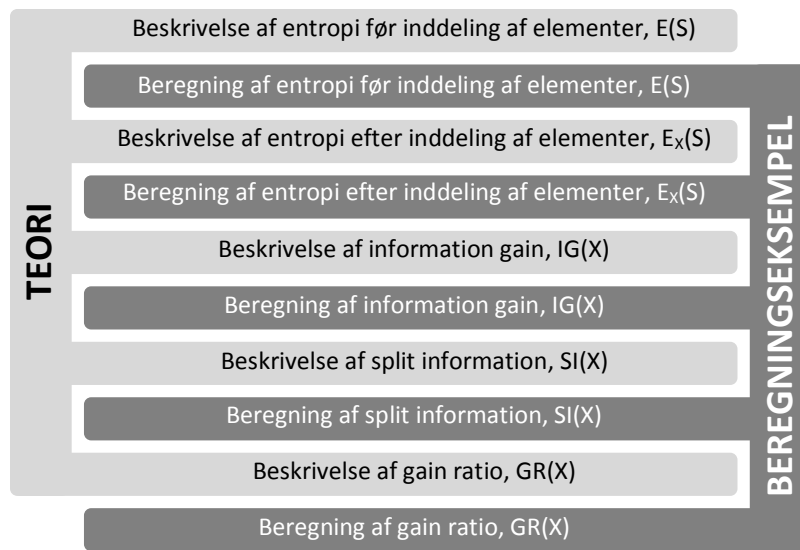
klassificeringstræer [79]. Samtidig kan algoritmen håndtere både numeriske og kategoriske attributter [77], og er frit tilgængeligt under betegnelsen J48 i data mining softwaren Weka [3]. Ydermere er C4.5 algoritmen blandt de mest anvendte algoritmer til udvikling af klassificeringstræer, hvilket betyder, at der er et udbredt kendskab til denne, samt at algoritmen er testet i mange forskellige sammenhænge og til flere forskellige formål [5, 24].

Et af de forhold, der har betydelig indflydelse på klassificeringstræets udseende, er algoritmens splitkriterium. Et splitkriterium opstiller krav til, hvornår en attribut kan anvendes i et klassificeringstræ, når der er mulighed for at vælge mellem flere forskellige attributter. I algoritmen C4.5 anvendes parameteren *gain ratio* som splitkriterium jf. appendiks M. Til beregning af *gain ratio* anvendes følgende parametre: *entropi*, *information gain* samt *split information*. En definition af de førnævnte parametre fremgår af tabel 7.1.

Parameter	Beskrivelse
Entropi	Denne parameter beregnes både før og efter en inddeling af datasættet ved en attribut. En beregning af <i>entropien</i> før en inddeling af datasættet giver et mål for graden af uorden, og kan anvendes som indikation på forudsigeligheden af informationen i et datasæt. En beregning af <i>entropien</i> efter en inddeling af datasættet angiver en vægtet sum af, hvor klar en klasseinddeling af elementerne der kan udføres med en given attribut.
Information gain	Denne parameter angiver et mål for ændringen i <i>entropi</i> mellem to forskellige tilstande, eksempelvis før og efter en inddeling af data.
Split information	Denne parameter anvendes til „normalisering“ af <i>information gain</i> , da den førnævnte parameter vil favorisere attributter med mange svarmuligheder.
Gain ratio	Denne parameter anvendes som splitkriterium i C4.5 algoritmen og beregnes på baggrund af <i>information gain</i> og <i>split information</i> . Attributten med den højeste værdi af <i>gain ratio</i> anvendes som knudepunkt i klassificeringstræet.

Tabel 7.1: Definition af nøglebegreber som anvendes til udvikling af et klassificeringstræ på baggrund af C4.5 algoritmen [77].

Den efterfølgende beskrivelse af teorien for udvikling af et klassificeringstræ følger den beregningsmæssige korrekte fremgangsmåde for C4.5 algoritmen. Det nærværende afsnit er opbygges således, at de teoretiske begreber introduceres enkeltvis med en efterfølgende eksemplificering af den praktiske anvendelse af disse. Opbygningen af afsnittet er illustreret på figur 7.1, hvor det øverste begreb introduceres først. Når de enkelte beregningstrin fra figur 7.1 er udført, kan resultaterne af beregningen illustreres grafisk i et klassificeringstræ.



Figur 7.1: Opbygningen af nærværende afsnit med hensyn til beskrivelse og eksemplificering af teorien til udvikling af et klassificeringstræ på baggrund af C4.5 algoritmen.

Entropi før inddeling af elementer

Ved udvikling af et klassificeringstræ skal der udvælges de attributter i et datasæt, der inddeler elementerne i de korrekte klasser ved brug af så få attributter som muligt. I den forbindelse er begrebet *entropi* særdeles relevant. På baggrund af definitionen af parameteren i tabel 7.1 anvendes denne til beregning af graden af uorden før en eventuel inddeling ved en attribut. Udtrykket til beregning af *entropien* i dette tilfælde fremgår af formel 7.1 [77]. Hvad angår værdien af denne parameter, kan den variere mellem nul og to, såfremt der haves fire inddelinger for klassen, som det er tilfældet for elforbruget i nærværende rapport. Betydningen af entropien i det aktuelle tilfælde fremgår af tabel 7.2¹.

$$E(S) = - \sum_{j=1}^k \left(\frac{\text{freq}(C_j, S)}{|S|} \cdot \log_2 \left(\frac{\text{freq}(C_j, S)}{|S|} \right) \right) \quad (7.1)$$

Hvor:

$E(S)$	Entropi før en eventuel inddeling af elementer fra datasæt S [<i>bit</i>].
$\text{freq}(C_j, S)$	Antal elementer i datasæt S der tilhører klassen C_j af de k mulige klasser.
$ S $	Antallet af elementer i datasæt S .

¹Enheden *bit* anvendes som enhed for en informationsmængde, og afhænger af antallet af mulige, diskrete tilstande samt sandsynligheden for hver af disse forekommer [104].

Entropi	Beskrivelse
$E(S) = 0 \text{ bit}$	En <i>entropi</i> på nul svarer til, at samtlige elementer i datasættet tilhører samme klasse.
$E(S) = 1 \text{ bit}$	En <i>entropi</i> på én svarer til, at samtlige elementer i datasættet er ligeligt fordelt på to af de fire klasseinddelinger.
$E(S) = 2 \text{ bit}$	En <i>entropi</i> på to svarer til, at samtlige elementer i datasættet er fordelt ligeligt på alle fire klasseinddelinger.

Tabel 7.2: Beskrivelse af *entropiens* betydning såfremt denne beregnes før inddeling af elementer. Ovenstående er kun gældende såfremt klassen har fire inddelinger.

Et eksempel på hvorledes ovenstående *entropi* beregnes, kan illustreres ud fra data i tabel 7.3. Af tabellen fremgår det, at datasættet indeholder 100 elementer, der er fordelt på de fire omtalte klasseinddelinger. På baggrund af disse oplysninger kan *entropien* før en eventuel inddeling af elementer beregnes, som det fremgår af formel 7.2.

	Klasse				I alt
	Lavt	Gennemsnitligt	Højt	Meget højt	
Antal elementer	35	42	16	7	100

Tabel 7.3: Klasseinddeling af elementer fra det fiktive data, der er genereret i kapitel 6.

$$\begin{aligned}
 E(S) &= - \left(\frac{35}{100} \log_2 \left(\frac{35}{100} \right) + \frac{42}{100} \log_2 \left(\frac{42}{100} \right) + \frac{16}{100} \log_2 \left(\frac{16}{100} \right) + \frac{7}{100} \log_2 \left(\frac{7}{100} \right) \right) \\
 &\approx 1,747 \text{ bit}
 \end{aligned} \tag{7.2}$$

Entropi efter inddeling af elementer

Til vurdering af hvor effektivt en attribut kan adskille elementerne i et datasæt på de forskellige klasseinddelinger, anvendes en *entropi* tilsvarende den foregående. Definitionen af *entropien* i dette tilfælde fremgår af tabel 7.1. Formeludtrykket til beregning af den førnævnte parameter fremgår af formel 7.3 [77], hvor betydningen af *entropiens* størrelse er vist i tabel 7.4. *Entropien* i den førnævnte formel angiver en vægtet sum af, hvor klar en klasseinddeling af elementerne der kan udføres med attribut X [77].

$$E_X(S) = \sum_{i=1}^n \left(\frac{|S_i|}{|S|} \cdot E(S_i) \right) \tag{7.3}$$

Hvor:

$E_X(S)$	<i>Entropi</i> efter inddeling af elementerne i datasæt S ved attribut X [<i>bit</i>].
$ S_i $	Antallet af elementer i datasæt S der for attribut X indeholder besvarelsen i af de n forskellige svarmuligheder for attributten.
$E(S_i)$	<i>Entropi</i> for elementerne i datasæt S der indeholder besvarelsen i til attribut X beregnet på baggrund af formel 7.1 [<i>bit</i>].

Entropi	Beskrivelse
$E_X(S) = 0 \text{ bit}$	En <i>entropi</i> på nul svarer til, at samtlige elementer, med samme besvarelse af den aktuelle attribut, tilhører samme klasseinddeling. Tilsvarende kan <i>entropien</i> opnå en værdi på nul, hvis alle elementerne på med svarmulighed A tilhører klasse 1, og samtlige elementer med svarmulighed B tilhører klasse 2 og så fremdeles.
$E_X(S) = 1 \text{ bit}$	En <i>entropi</i> på én kan opnås, hvis samtlige elementer, med samme besvarelse af den aktuelle attribut, er ligeligt fordelt på to af de fire klasseinddelinger for enten en eller flere af attributtens svarmuligheder.
$E_X(S) = 2 \text{ bit}$	En <i>entropi</i> på to kan opnås, hvis samtlige elementer, med samme besvarelse af den aktuelle attribut, er ligeligt fordelt på alle fire klasseinddelinger. Det gælder både hvis alle elementerne har samme besvarelse af attributten, men også hvis elementerne har forskellige besvarelser af attributten.

Tabel 7.4: Beskrivelse af *entropiens* betydning såfremt denne beregnes efter inddeling af elementer. Ovenstående er kun gældende såfremt klassen har fire inddelinger.

Til eksemplificering af hvorledes *entropien* efter en inddeling skal beregnes, tages der i det efterfølgende udgangspunkt i attributten omhandlende, hvorvidt beboerne i en husstand er optaget af at spare på elforbruget. Fordelingen af elementer, afhængigt af klasseinddelingen og besvarelsen til den pågældende attribut, fremgår af tabel 7.5. På baggrund af disse oplysninger kan *entropien* efter en inddeling ved attributten beregnes, som illustreret i formel 7.4.

Besvarelse til attribut	Klasse				I alt
	Lavt	Gennemsnitligt	Højt	Meget højt	
Ja altid	17	11	0	0	28
Ja for det meste	18	13	6	0	37
Sommetider	0	18	4	2	24
Nej slet ikke	0	0	6	5	11

Tabel 7.5: Fordelingen af elementer for det fiktive data, der er genereret i kapitel 6, opdelt mellem elementernes klasseinddeling og besvarelse af attributten omhandlende hvorvidt beboerne i en husstand er optaget af at spare på elforbruget.

$$\begin{aligned}
E_X(S) = & -\frac{28}{100} \left(\frac{17}{28} \log_2 \left(\frac{17}{28} \right) + \frac{11}{28} \log_2 \left(\frac{11}{28} \right) \right) \\
& -\frac{37}{100} \left(\frac{18}{37} \log_2 \left(\frac{18}{37} \right) + \frac{13}{37} \log_2 \left(\frac{13}{37} \right) + \frac{6}{37} \log_2 \left(\frac{6}{37} \right) \right) \\
& -\frac{24}{100} \left(\frac{18}{24} \log_2 \left(\frac{18}{24} \right) + \frac{4}{24} \log_2 \left(\frac{4}{24} \right) + \frac{2}{24} \log_2 \left(\frac{2}{24} \right) \right) \\
& -\frac{11}{100} \left(\frac{6}{11} \log_2 \left(\frac{6}{11} \right) + \frac{5}{11} \log_2 \left(\frac{5}{11} \right) \right) \\
& \approx 1,171 \text{ bit}
\end{aligned} \tag{7.4}$$

Information gain

Det er beskrevet i tabel 7.1, at de ovenstående entropiberegninger anvendes til beregning af *information gain*, hvilket fremgår af formel 7.5 [77]. Parameteren *information gain* angiver et mål for ændringen i *entropi* mellem to forskellige tilstande - før og efter en inddeling af data. På baggrund af definitionen af *entropien*, kan *information gain* potentielt opnå værdier mellem nul og to, såfremt klassen består af fire inddelinger. Eksempler på betydningen af *information gain* fremgår af tabel 7.6.

$$IG(X) = E(S) - E_X(S) \tag{7.5}$$

Hvor:

$IG(X) \mid$ *Information gain* for attribut X på baggrund af elementerne i datasæt S [bit].

Information gain	Beskrivelse
Lav værdi:	En lav værdi for <i>information gain</i> betyder, at klasseinddelingen af elementerne i et datasæt ikke vil forbedres betydeligt ved anvendelse af den aktuelle attribut.
Høj værdi:	En høj værdi for <i>information gain</i> betyder, at der kan foretages en relativ klar inddeling af elementerne i et datasæt ved anvendelse af den aktuelle attribut. Den maksimale værdi for <i>information gain</i> fås, hvis en attribut kan inddele samtlige elementer i et datasæt i de korrekte klasser, hvor elementerne forinden var ligeligt fordelt mellem de enkelte klasser.

Tabel 7.6: Beskrivelse af betydning af *information gain*.

I nogle algoritmer anvendes *information gain* som splitkriterium. Undersøgelser har imidlertid vist, at *information gain* vil favorisere de attributter, der har flest svarmuligheder [49, 77, 108]. Det betyder eksempelvis, at numeriske attributter vil favoriseres i forhold til kategoriske attributter med et beskedent antal svarmuligheder såfremt *information gain* anvendes som splitkriterium. Dette forhold er årsagen til, at *information gain* ikke anvendes som splitkriterium i C4.5 algoritmen [77].

Formel 7.5 kan anvendes til beregning af værdien af *information gain* fra det tidligere eksempel i dette afsnit. Oplysninger til beregning af *information gain* fås fra entropiberegningerne i formel 7.2 og 7.4. Resultatet af denne beregning, for attributten omhandlende hvorvidt beboerne i husstanden er optaget af at spare på elforbruget, bliver: $IG(X) = 0,576 \text{ bit}$.

Split information

På baggrund af *information gains* favorisering af attributter med talrige svarmuligheder, foretages en form for normalisering af denne ved brug af parameteren *split information*. Det matematiske udtryk til beregning af *split information* fremgår af formel 7.6 [77]. Af formelen fremgår det, at *split information* beregnes lidt tilsvarende *entropien* $E(S)$ fra formel 7.1. Det bør dog bemærkes, at *split information* beregnes på baggrund af antallet af elementer i forhold til den aktuelle attributs svarmuligheder, og ikke i forhold til klasseinddelingen som det er tilfældet ved *entropien*. Eksempler på betydningen af *split information* fremgår af tabel 7.7.

$$SI(X) = - \sum_{i=1}^n \left(\frac{|S_i|}{|S|} \cdot \log_2 \left(\frac{|S_i|}{|S|} \right) \right) \quad (7.6)$$

Hvor:

$SI(X)$ | *Split information* for attribut X med n svarmuligheder og på baggrund af elementerne i datasæt S [bit].

Split information	Beskrivelse
Lav værdi:	En lav værdi for <i>split information</i> opnås, hvis størstedelen af elementerne i datasættet har samme besvarelse til den aktuelle attribut.
Høj værdi:	En høj værdi for <i>split information</i> opnås, hvis der er adskillige svarmuligheder til en attribut samtidig med, at elementerne i datasættet har forskellige besvarelser til denne.

Tabel 7.7: Beskrivelse af betydning af *split information*.

Til eksemplificering af hvorledes *split information* beregnes i praksis anvendes det tidligere eksempel fra dette afsnit. Oplysningerne til beregning af *split information* for attributten, omhandlende hvorvidt beboerne i husstanden er optaget af at spare på elforbruget, fremgår af tabel 7.5. På denne baggrund kan *split information* beregnes som illustreret formel 7.7.

$$SI(X) = - \left(\frac{28}{100} \log_2 \left(\frac{28}{100} \right) + \frac{37}{100} \log_2 \left(\frac{37}{100} \right) + \frac{24}{100} \log_2 \left(\frac{24}{100} \right) + \frac{11}{100} \log_2 \left(\frac{11}{100} \right) \right) \approx 1,889 \text{ bit} \quad (7.7)$$

Gain ratio

Sammenhængen mellem *gain ratio*, *information gain* og *split information* fremgår af formel

7.8 [77]. Splitkriteriet for algoritmen anvendes på den måde, at attributten med den højeste *gain ratio* anvendes som knudepunkt i klassificeringstræet [77]. Der er dog det krav, at *gain ratio* kun skal beregnes for de attributter, der har en *information gain*, der er højere end gennemsnittet for samtlige attributters *information gain*. Dette krav er fastlagt, da en høj værdi for *gain ratio* kan opnås med en lav værdi af *split information*. Hvis hovedparten af elementerne i datasættet har samme besvarelse til en attribut, vil det medføre, at $|S_i|/|S|$ vil nærme sig værdien én, hvorefter $\log_2(|S_i|/|S|)$ vil nærme sig nul. Det betyder, at *split information* får en lav værdi, hvormed attributten favoriseres i beregningen af *gain ratio*, selvom attributten ikke nødvendigvis giver en klar klasseinddeling af elementerne [49, 77].

$$GR(X) = \frac{IG(X)}{SI(X)} \quad (7.8)$$

Hvor:

$GR(X)$ | *Gain ratio* for attribut X [–].

Af ovenstående formel fremgår det, at de forhold der kan reducere sandsynligheden for, at en attribut vælges som knudepunkt i klassificeringstræet er en lav *information gain* eller en høj *split information*. Med hensyn til det tidligere eksempel kan *gain ratio* for attributten, omhandlende hvorvidt beboerne i husstanden er optaget af at spare på elforbruget, beregnes til: $GR(X) = 0,305$.

Opstilling af klassificeringstræ

På baggrund af den foregående beskrivelse kan *gain ratio* beregnes for samtlige attributter i et datasæt, hvor fremgangsmåden er illustreret på figur 7.1. Der er gennem den foregående beskrivelse anvendt et beregningseksempel ud fra attributten omhandlende hvorvidt beboerne i en husstand er optaget af at spare på elforbruget. Beregninger af *gain ratio* for de øvrige attributter viser, at attributten, der er anvendt som eksempel gennem dette afsnit, har den højeste *gain ratio*. Det betyder, at denne attribut skal anvendes som R-knude i klassificeringstræet, hvor de 100 elementer inddeles efter besvarelsene til attributten.

Datasættet skal efterfølgende opdeles ud fra elementernes besvarelse til attributten, der er anvendt som R-knude, hvilket betyder, at datasættet opdeles i fire mindre sæt. For hvert af disse sæt foretages en tilsvarende beregning af *gain ratio* for samtlige attributter. Denne beregningsprocedure fortsætter, indtil samtlige elementer i de enkelte knudepunkter tilhører samme klasseinddeling, eller hvis ingen attributter kan medføre en bedre inddeling af elementerne. Når det er tilfældet omdannes knudepunktet til en L-knude. Den beskrevne beregningsproces er automatiseret i den tidligere omtalte data minings software Weka, hvorfor dette program er anvendt i forbindelse med det førømtalte datasæt. På baggrund af det fiktive data kan der udvikles et klassificeringstræ, der vil fremstå som illustreret på *tegning A*, der er vedlagt rapporten. Det bør dog i den forbindelse bemærkes, at nogle svarmuligheder ikke fremgår af den førnævnte figur. Det skyldes det aktuelle datasæts beskedne størrelse, hvor alle kombinationsmuligheder af besvarelser til attributterne ikke fremkommer. Klassificeringstræet ender med at have 19 R- og I-knuder og yderligere 34 L-knuder, hvor i alt syv ud af hundrede elementer fejlklassificeres.

Af vedlagte *tegning A* fremgår en middelværdi og standardafvigelse i hver L-knude. Disse er beregnet på baggrund af elementerne i de respektive L-knuder. Da klassificeringstræet er udviklet på baggrund af 100 elementer, resulterer det i, at der er et beskedent antal elementer i hver L-knude. Hvis der udvikles et klassificeringstræ på baggrund af et større antal elementer, kan det være en mulighed, at undersøge hvilken sandsynlighedsfordeling der kan beskrive elementerne i de enkelte L-knuder. I stedet for at beregne en middelværdi og standardafvigelse i hver L-knude, kan det være en mulighed, at beregne henholdsvis én middelværdi og standardafvigelse for hver klasse. Det vil, i det aktuelle tilfælde, resultere i fire middelværdier og standardafvigelser. Denne beregningsmetode vil fjerne muligheden for, at illustrere eventuelle forskelle mellem fordelingen af elementer i de enkelte L-knuder, der tilhører samme klasseinddeling. Af denne årsag er det i nærværende projekt valgt, at beregne en middelværdi og standardafvigelse for hver L-knude.

Beskrivelsen af udviklingen af klassificeringstræet i det nærværende kapitel tager udgangspunkt i, at samtlige attributter er kategoriske. Dette valg er foretaget, da det vil resultere i mere overskuelige håndberegninger til eksemplificering af hvorledes et klassificeringstræ udvikles. Ikke desto mindre vil der være situationer i praksis, hvor det anvendes numeriske attributter. C4.5 algoritmen kan håndtere både numeriske og kategoriske attributter, men algoritmen har en bestemt metode til håndtering af numeriske attributter. I appendiks S er det redegjort for hvorledes numeriske attributter håndteres af henholdsvis C4.5 algoritmen og af data minings softwaren Weka².

7.2 Håndtering af manglende data

I afsnit 7.1 forefindes en udførlig beskrivelse af, hvorledes et klassificeringstræ udvikles på baggrund af algoritmen C4.5. Den førnævnte beskrivelse tager udgangspunkt i et datasæt med fiktivt data, der er nærmere beskrevet i kapitel 6. Der er tale om et komplet datasæt, hvilket vil sige, at der haves besvarelser på samtlige attributter for samtlige elementer. I praksis er det almindeligt, at der kan forekomme mangler i et datasæt [43, 77]. Det kan eksempelvis dreje sig om ufuldendte spørgeskemabesvarelser, manglende målinger eller manglende viden i forhold til den aktuelle attribut. I alle tilfældene vil det resultere i, at der kan være en eller flere ukendte attributter for et eller flere elementer. Når et ufuldstændigt datasæt skal analyseres, eller eksempelvis anvendes til udvikling af et klassificeringstræ, må der tages stilling til, hvorledes ukendte attributter skal håndteres. Én mulighed kan være kun at medtage besvarelser for de elementer, hvis besvarelse indeholder svar på samtlige attributter [6]. En anden mulighed kan være, at anvende den oftest forekommende værdi af alle de kendte besvarelser for den aktuelle attribut [6]. Det kan også være en mulighed at undersøge eventuelle korrelationer mellem attributter for de kendte besvarelser, og anvende denne viden til estimering af ukendte attributter [6]. Ved anvendelse af C4.5 algoritmen til udvikling af et klassificeringstræ er der imidlertid en konkret fremgangsmåde for, hvorledes manglende data skal håndteres. På baggrund af algoritmens måde at håndtere manglende data på, må der skelnes mellem henholdsvis udvikling og anvendelse af klassificeringstræer, da der er tale om to forhold, der håndteres på hver sin måde. I de efterfølgende afsnit re-

²I det følgende regneark foretages en håndberegning af hvorledes numeriske attributter håndteres i C4.5 algoritmen: *Klassificeringstræ | Udvikling af klassificeringstræ, numerisk attribut | Udvikling af klassificeringstræ, numerisk attribut.csv* på den medfølgende DVD.

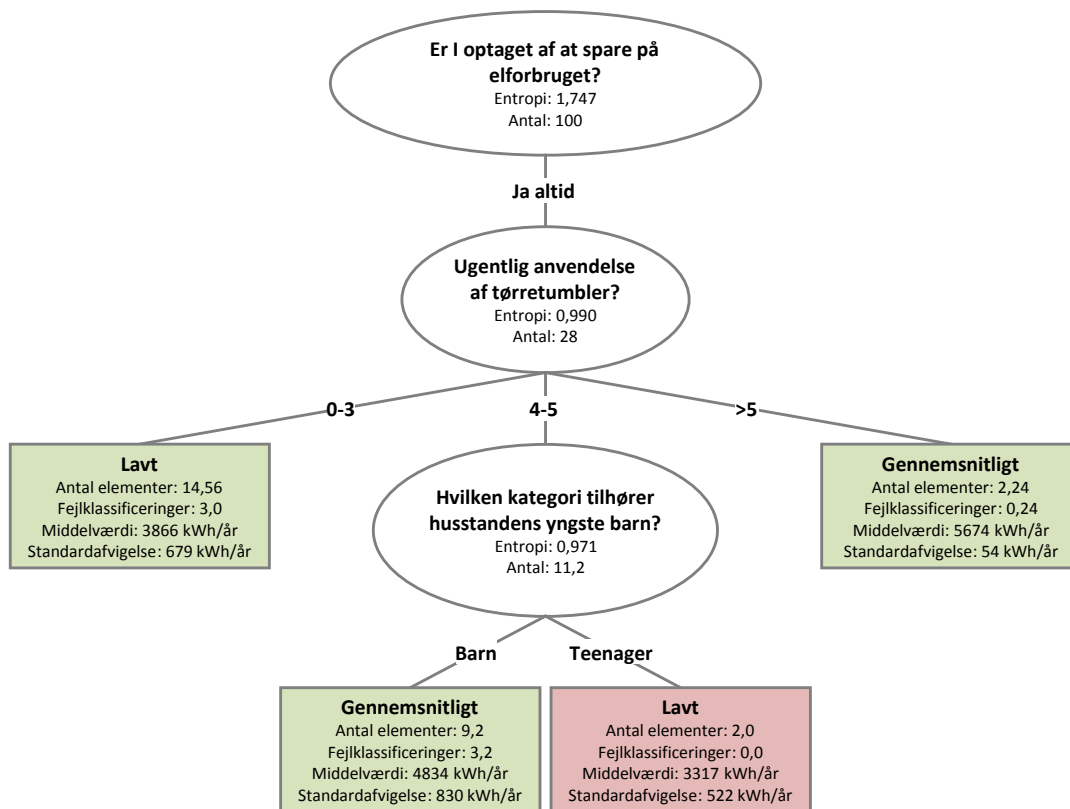
degøres for hvorledes et ufuldstændigt datasæt kan håndteres ved henholdsvis udvikling og anvendelse af et klassificeringstræ.

I de efterfølgende afsnit anvendes et eksempel gennem beskrivelsen, hvor der tages udgangspunkt i det fiktive data, der tidligere er anvendt i afsnit 7.1, men dog med enkelte modifikationer. For tre af de fiktive elementer, der har svaret „ja altid“ til spørgsmålet om, hvorvidt beboerne i husstanden er optaget af at spare på elforbruget, er den tilhørende besvarelse for den ugentlige anvendelse af tørretumbler i husstanden slettet. Alle de tre fiktive elementer havde tidligere besvarelsen „0-3“ for attributten omhandlende den ugentlige anvendelse af tørretumbler i husstanden³.

7.2.1 Manglende data ved udvikling af klassificeringstræ

Ved udvikling af et klassificeringstræ på baggrund af algoritmen C4.5 har manglende data direkte indflydelse på træets udseende. Hvis der udvikles et klassificeringstræ på baggrund af det tidligere nævnte datasæt, vil resultatet se ud som det fremgår af figur 7.2, hvor der dog kun er vist et udsnit af klassificeringstræet. Det fremgår af figur 7.2, at der er sket en ændring for hvor mange elementer, der er placeret i de enkelte L-knuder. Det kan eksempelvis ses, at der er placeret 14,56 elementer i L-knuden, hvor disse har svaret „ja altid“ til spørgsmålet om, hvorvidt de er optaget af elforbruget, samtidig med at de anvender tørretumbler „0-3“ gange ugentligt. I samme L-knude er der 3,00 ud af de 14,56 elementer, der er fejlklassificeret. Hvis dette resultat sammenholdes med det oprindelige klassificeringstræ, er der, jævnfør *tegning A*, 16,00 elementer i samme L-knude med tilsvarende 3,00 fejlklassificeringer. Det bør dog bemærkes, at algoritmen C4.5 ikke kan håndtere manglende data for den attribut, der ønskes anvendt som klasse ved udvikling af et klassificeringstræ [6].

³Datasættet med de tre manglende værdier for den ugentlige anvendelse af tørretumbler i husstanden fremgår af filen *Klassificeringstræ\ Udvikling af klassificeringstræ, manglende data\ Udvikling af klassificeringstræ, manglende data.csv* på den medfølgende DVD.



Figur 7.2: Udsnit af klassificeringstræ udviklet på baggrund af fiktivt data med tre manglende værdier for attributten omhandlende husstandes ugentlige brug af tørretumbler. Farverne på figuren anvendes kun til beskrivelse i afsnit 7.2.2.

Af figur 7.2 fremgår det, at der ikke er placeret et helt antal elementer i flere af L-knuderne. Årsagen skal findes i den måde, hvorpå C4.5 algoritmen håndterer manglende data. Ved udvikling af et klassificeringstræ på baggrund af den førnævnte algoritme har manglende data direkte indflydelse på beregning af *information gain* og *split information*, der tidligere er introduceret i afsnit 7.1 [79]. I tilfælde af manglende data for en attribut beregnes *information gain* som udtrykt i formel 7.9 [79].

$$IG_{MD}(X) = \frac{|S - S_0|}{|S|} IG(X) \quad (7.9)$$

Hvor:

$IG_{MD}(X)$	<i>Information gain</i> for attribut X ved manglende data i datasæt S [bit].
S	Det samlede antal objekter der skal fordels ved attribut X .
S_0	Antal objekter med ukendt besvarelse i datasæt S for attribut X .
$IG(X)$	<i>Information gain</i> beregnet udelukkende på baggrund af objekter med kendte værdier for attribut X . Beregnes af formel 7.5 på side 141 [bit].

Det fremgår af formel 7.9, at hvis der mangler data i et datasæt, bliver *information gain* beregnet udelukkende på baggrund af de kendte data. Derefter reduceres denne *information gain* svarende til andelen af kendte data i det samlede datasæt. Hvad angår *split information*, udvides det oprindelige formeludtryk (formel 7.6 på side 142), således at de ukendte data udtrykker et muligt „udfald“. Det betyder, at *split information* beregnes som udtryk i formel 7.10, såfremt der mangler data i datasættet [79].

$$SI_{MD}(X) = -\frac{|S_0|}{|S|} \log_2 \left(\frac{|S_0|}{|S|} \right) - SI(S - S_0, X) \quad (7.10)$$

Hvor:

$$\begin{array}{l|l} SI_{MD}(X) & \text{Split information for attribut } X \text{ ved manglende data i datasæt } S \text{ [bit].} \\ SI(S - S_0, X) & \text{Split information beregnet udelukkende på baggrund af objekter med} \\ & \text{kendte værdier for attribut } X. \text{ Beregnes af formel 7.6 på side 142 [bit].} \end{array}$$

Da det ufuldstændige datasæt har indflydelse på beregningerne af *information gain* og *split information*, skal der foretages nye beregninger for hele klassificeringstræet. Det skyldes, at manglende værdier i et datasæt i nogle tilfælde kan ændre på rækkefølgen af de enkelte attributter. Det fremgår af figur 7.2, at rækkefølgen på attributterne i dette tilfælde ikke umiddelbart er ændret, når det sammenholdes med klassificeringstræet udviklet på baggrund af det komplette datasæt, hvilket er illustreret på *tegning A*. Til trods for at den umiddelbare rækkefølge af attributterne ikke er ændret, fremgår en ændring i *information gain*, *split information*, *gain ratio* samt rangeringen af attributten med de ukendte værdier. Hovedresultaterne fremgår i tabel 7.8, hvor de viste værdier gælder for attributten omhandlende husstandens ugentlige anvendelse af tørretumbler i henholdsvis det komplette datasæt og det ufuldstændige datasæt⁴.

⁴Detaljerede håndberegninger af de enkelte faktorer med det ufuldstændige datasæt fremgår af følgende regneark: *Klassificeringstræ \ Udvikling af klassificeringstræ, manglende data \ Udvikling af klassificeringstræ, manglende data.xlsx* på den medfølgende DVD.

	Komplet datasæt	Ufuldstændigt datasæt
R-knude		
Information gain	0,240	0,170
Split information	1,353	1,521
Gain ratio	0,178	0,111
Rangering	6	—*
I-knude (kun besvarelser med „ja altid“ til R-knuden)		
Information gain	0,222	0,173
Split information	1,264	1,656
Gain ratio	0,176	0,104
Rangering	1	1

* *Information gain* for attributten omhandlende husstandens ugentlige anvendelse af tørretumbler er under middelværdien af samtlige attributters *information gain*, hvorfor den pågældende attribut ikke kan rangeres.

Tabel 7.8: Resultatskema for håndberegninger af *information gain*, *split information* og *gain ratio* for attributten omhandlende husstandens ugentlige anvendelse af tørretumbler for henholdsvis det komplette datasæt samt datasættet med tre manglende besvarelser til den pågældende attribut.

I overensstemmelse med *Quinlan et al.* [79] fremgår det af tabel 7.8, at ukendte besvarelser for en attribut vil medføre en reduceret *information gain* og en forøget *split information*. Det medfører, at *gain ratio* reduceres, hvormed attributtens rangering kan falde. Det betyder tilsvarende, at sandsynligheden for at en attribut med en eller flere ukendte besvarelser anvendes som enten R- eller I-knude reduceres. Dette fremgår af tabel 7.8, hvor attributten omhandlende husstandens ugentlige anvendelse af tørretumbler er rangeret som nummer seks ved R-knuden i det komplette datasæt. I det ufuldstændige datasæt tages attributten ikke i betragtning, da *information gain* er under grænsen for, hvornår der skal beregnes en *gain ratio* for attributten. Reduktionen af *information gain* og *gain ratio* har dog ikke reduceret rangeringen for attributten i I-knuden.

På trods af at de ukendte data udtrykkes som et muligt „udfald“ ved beregning af *split information*, betyder det ikke, at der tilføjes en ekstra gren på klassificeringstræet. De ukendte data fordeles på samtlige udfald ud fra en vægtfaktor, der beregnes, som illustreret i formel 7.11 [79].

$$w = \frac{|S_i|}{|S - S_0|} \quad (7.11)$$

Hvor:

w	Vægtfaktor til fordeling af ukendte besvarelser mellem de mulige udfald for en given attribut.
S_i	Antal objekter med kendt besvarelse i datasæt S med udfaldsmulighed i for en given attribut.

Denne vægtfaktor er årsagen til, at der ikke er placeret et helt antal elementer i flere af L-knuderne på figur 7.2. Til illustrering af hvorledes vægtfaktoren beregnes, kan der anvendes et eksempel med L-knuden, der udgrener sig ved besvarelsen „0-3“ for attributten omhandlende husstandens ugentlige anvendelse af tørretumbler. Tabel 7.9 gengiver resultaterne af en optælling af antallet af elementer i det fiktive data for hver svarmulighed. Det fremgår af tabellen, at der er tre elementer, der ikke har besvaret den aktuelle attribut, men disse elementers klasse er dog alle opgivet til et lavt elforbrug. Tilsvarende er der ti af elementerne, der har besvarelsen „0-3“ til attributten, som har et lavt elforbrug, og tre der har et gennemsnitligt elforbrug.

Besvarelse ved attribut	Antal besvarelser	Besvarelser fordelt på klasse		Vægtfaktor
		Lavt	Gennemsnitligt	
Manglende besvarelse	$S_0 = 3$	$S_{L,0} = 3$	$S_{G,0} = 0$	
0-3 gange ugentligt	$S_{0-3} = 13$	$S_{L,0-3} = 10$	$S_{G,0-3} = 3$	$w_{0-3} = 0,52$
4 – 5 gange ugentligt	$S_{4-5} = 10$	$S_{L,4-5} = 4$	$S_{G,4-5} = 6$	$w_{4-5} = 0,40$
> 5 gange ugentligt	$S_{>5} = 2$	$S_{L,>5} = 0$	$S_{G,>5} = 2$	$w_{>5} = 0,08$
Sum	$S_\Sigma = 28$			

Tabel 7.9: Resultatskema for håndberegninger af vægtfaktor ved I-knuden, hvor der inddeles efter attributten omhandlende husstandens ugentlige anvendelse af tørretumbler, som vist på figur 7.2, i det ufuldstændige datasæt.

I tabel 7.9 er de enkelte vægtfaktoreres størrelse oplyst, hvor disse er beregnet på baggrund af formel 7.11. Den praktiske anvendelse af den førnævnte formel er illustreret i formel 7.12, hvor der fremgår et eksempel på, hvorledes vægtfaktoren for besvarelsen „0-3“ beregnes.

$$w_{0-3} = \frac{|S_{0-3}|}{|S_\Sigma - S_0|} = \frac{13}{28 - 3} = 0,52 \quad (7.12)$$

På tilsvarende måde som vist ovenfor skal vægtfaktorerne for de øvrige svarmuligheder beregnes. Vægtfaktorerne indgår direkte i beregningen af antallet af elementer i de enkelte L-knuder. Dette antal kan, for besvarelsen „0-3“, beregnes som illustreret i formel 7.13. I samme formel fremgår det, hvorledes vægtfaktoren ligeledes anvendes til beregning af antallet af fejlplacerede elementer i L-knuden⁵.

$$\begin{aligned} S_{elementer,0-3} &= S_{0-3} + w_{0-3} \cdot S_0 = 13 + 0,52 \cdot 3 = 14,56 \\ S_{fejl,0-3} &= S_{G,0-3} + w_{0-3} \cdot S_{G,0} = 3 + 0,52 \cdot 0 = 3,00 \end{aligned} \quad (7.13)$$

⁵Alle håndberegninger af blandt andet vægtfaktorer og antal elementer i L-knuderne fremgår af regnearket: *Klassificeringstræ \ Udvikling af klassificeringstræ, manglende data \ Udvikling af klassificeringstræ, manglende data.xlsx* på den medfølgende DVD.

Hvor:

$S_{elementer,0-3}$	Det samlede antal elementer i L-knuden for besvarelsen „0-3“ til attributten omhandlende husstandens ugentlige anvendelse af tørretumbler.
$S_{fejl,0-3}$	Antal fejlklassificerede elementer i L-knuden for besvarelsen „0-3“ til attributten omhandlende husstandens ugentlige anvendelse af tørretumbler.

Fremgangsmåden, der er vist i formel 7.13, skal tilsvarende anvendes i de øvrige L-knuder. Derved vil der nås frem til det antal elementer, der er oplyst på figur 7.2. Der er gennem ovenstående beskrivelse redegjort for, hvorledes udviklingen af et klassificeringstræ skal foregå, såfremt der ikke haves adgang til et komplet datasæt.

7.2.2 Manglende data ved anvendelse af klassificeringstræ

I det forrige afsnit er der redegjort for, hvorledes manglende data skal håndteres ved udviklingen af et klassificeringstræ. En anden problemstilling opstår, hvis et eksisterende klassificeringstræ skal anvendes til klassificering af et element, hvor en af attributterne for dette er ukendt. I det efterfølgende anvendes et eksempel til beskrivelse af, hvorledes den førnævnte problemstilling skal håndteres.

Der skal, som tidligere beskrevet, tages udgangspunkt i et eksisterende klassificeringstræ. I det efterfølgende eksempel antages det, at klassificeringstræet, der er vist et udsnit af på figur 7.2, er afprøvet på testdata, og ikke blot udviklet på baggrund af træningsdata. Tilsvarende antages det, at elementerne i testdataet er fordelt på samme måde, som træningsdataet er fordelt på figur 7.2. På baggrund af disse antagelser kan oplysningerne fra klassificeringstræet, der fremgår af den førnævnte figur, anvendes direkte i det efterfølgende eksempel. I eksemplet skal klassificeringstræet anvendes til test af hvilken klasse for elforbruget, et fiktiv element kan forventes at tilhøre. Dette element omtales i det efterfølgende som *element A*, og hvad angår besvarelser til attributterne, er disse oplistet i tabel 7.10.

Attribut	Besvarelse
Er I optaget af at spare på elforbruget?	Ja altid
Ugentlig anvendelse af tørretumbler?	Ikke oplyst
Hvilken kategori tilhører husstandens yngste barn?	Barn

Tabel 7.10: Besvarelse fra *element A* til eksemplificering af fremgangsmåden for klassificering af denne i et eksisterende klassificeringstræ, hvor besvarelsen for én attribut er ukendt.

Det fremgår af tabel 7.10, at elementets besvarelse af husstandens ugentlige anvendelse af tørretumbler ikke er oplyst. På denne baggrund af er der sandsynlighed for, at elementet kan klassificeres med enten et lavt eller et gennemsnitligt elforbrug, jf. figur 7.2. Der er dog kun mulighed for, at elementet ender i tre af de i alt fire forskellige L-knuder, hvor disse er illustreret med grøn på figur 7.2. Årsagen til at elementet ikke kan ende i den på figur 7.2 røde L-knude er, at besvarelsen for attributten omhandlende alderen på husstandens yngste barn er oplyst.

Eftersom klassificeringstræet er udviklet på baggrund af algoritmen C4.5, er der en bestemt fremgangsmåde, der skal anvendes til klassificering af en respondent [77]. Fremgangsmåden, for vurdering af hvilken klasse respondenter tilhører, bygger på en statistisk vurdering af de mulige udfald. Dette sker på baggrund af en beregning af resultaterne for testdataets klassificering i det aktuelle klassificeringstræ. Ved klassificeringen skal der tages højde for sandsynligheden for et givent udfald ved I-knuder samt sandsynligheden for en given klassificering i relevante L-knuder.

Ovenstående resulterer i, at der i dette eksempel skal beregnes sandsynligheder for udfald ved to forskellige I-knuder; husstandens ugentlige anvendelse af tørretumbler og alderen på det yngste barn i husstanden. Hvad angår udfaldet ved attributten omhandlende den ugentlige anvendelse af tørretumbler, skal dette vurderes på baggrund af antallet af elementer, der i testdataet er inddelt i denne attribut. I dette eksempel er der tale om i alt 28 elementer, hvoraf 14,56 føres ud af grenen „0-3“, 11,2 føres ud af grenen „4-5“ og de resterende 2,24 elementer føres ud til „>5“. Ved attributten omhandlende alderen på husstandens yngste barn er sandsynligheden for udfaldet „Barn“ 100 %, da oplysningerne om denne attribut er kendt. Havde besvarelsen for denne attribut tilsvarende været ukendt, skulle sandsynligheden for udfaldet beregnes på tilsvarende måde, som det er beskrevet ovenfor.

Foruden ovenstående skal der beregnes en sandsynlighed for en given klassificering i relevante L-knuder. I dette eksempel er der to forskellige klassificeringsmuligheder i de aktuelle L-knuder; „lavt“ eller „gennemsnitligt“, hvilket fremgår af figur 7.2. Det bør i denne sammenhæng bemærkes, at der skal have kendskab til, hvilken klassificering de fejlklassificerede elementer i L-knuderne reelt har. I dette eksempel forholder det sig således, at samtlige elementer fra grenen „ja altid“ har enten et lavt eller gennemsnitligt elforbrug. Det betyder, at de elementer, der i L-knuderne fejlklassificeres til „lavt“ reelt har klassificeringen „gennemsnitligt“ og omvendt for de øvrige L-knuder.

På baggrund af ovenstående beskrivelse kan sandsynligheden for, at *element A* kan klassificeres med henholdsvis et lavt eller et gennemsnitligt elforbrug, beregnes. Fremgangsmåden for denne beregning fremgår af formel 7.14 [5].

$$\begin{aligned}P_{Lavt} &= \frac{14,56}{28} \cdot \frac{11,56}{14,56} + \frac{2,24}{28} \cdot \frac{0,24}{2,24} + \frac{11,2}{28} \cdot \frac{3,2}{9,2} \\&\approx 0,561 = 56,1 \% \\P_{Gennemsnitligt} &= \frac{14,56}{28} \cdot \frac{3}{14,56} + \frac{2,24}{28} \cdot \frac{2}{2,24} + \frac{11,2}{28} \cdot \frac{6}{9,2} \\&\approx 0,439 = 43,9 \% \end{aligned} \tag{7.14}$$

Beregningerne i formel 7.14 viser, at sandsynligheden, for at *element A* kan klassificeres med et lavt elforbrug, er højere end for en klassificering med et gennemsnitligt elforbrug. Dermed er det af ovenstående fremgangsmåde redegjort for en særdeles relevant problemstilling, da det er beskrevet, hvorledes manglende data for et element skal håndteres, når dette skal klassificeres på baggrund af et eksisterende klassificeringstræ.

7.3 Begrænsning af klassificeringstræets størrelse

Ifølge beskrivelsen i afsnit 7.1 skal udviklingen af klassificeringstræet som udgangspunkt forløbe indtil alle elementerne i L-knuderne tilhører samme klasseinddeling. Der kan dog, som beskrevet i det førnævnte afsnit, i nogle tilfælde nås et stopkriterium, selvom et eller flere elementer er fejlklassificeret. På baggrund af denne procedure er der risiko for, at klassificeringstræet bliver stort og komplekst [77]. For at undgå dette er der i C4.5 algoritmen en fast procedure for, hvorledes klassificeringstræets størrelse skal begrænses. Formålet med at begrænse klassificeringstræets størrelse er, at fjerne de dele af klassificeringstræet, der ikke bidrager betragteligt til dets nøjagtighed [7].

Udviklingen af klassificeringstræet sker efter en *top-down* fremgangsmåde, hvor træningsdata fordeles fra R-knuden og ned til der nås en L-knude. Metoden i C4.5 algoritmen, til begrænsningen af klassificeringstræets størrelse, følger en *bottom-up* fremgangsmåde. Det betyder, at begrænsningen af klassificeringstræets størrelse foregår fra L-knuderne og op efter. For at processen kan forløbe, skal det undersøges, hvorvidt der kan forventes en lavere fejlklassificering, hvis klassificeringstræets størrelse begrænses ved at udskifte I-knuder med L-knuder. I C4.5 algoritmen foretages denne vurdering på baggrund af træningsdataet. Da klassificeringstræet er udviklet direkte på baggrund af træningsdataet, har klassificeringstræet den opbygning, der inddeler dette datasæt bedst muligt. L-knuderne i klassificeringstræet består hovedsagligt af elementer med samme klasseinddeling, jf. vedlagte *tegning A*. Det betyder, at såfremt en I-knude erstattes af en L-knude vil det medføre en fejlklassificering af nogle elementer [77]. Det kan eksemplificeres ved knuderne i det røde felt på figur 7.3. Der er ingen fejlklassificerede elementer i de to L-knuder, men hvis I-knuden erstattes med en L-knude, vil dette medføre en fejlklassificering på 50 % af elementerne. Det betyder, at hvis fejlklassificeringen vurderes direkte på baggrund af træningsdataet, vil der i størstedelen af tilfældene ikke foretages en begrænsning af klassificeringstræets størrelse, da antallet af fejlklassificerede elementer vil stige. I nogle algoritmer tages der højde for denne problemstilling, ved at anvende et andet datasæt til begrænsning af klassificeringstræets størrelse. Problemstillingen håndteres på en anden måde i C4.5 algoritmen. Det skyldes, at udvikleren af denne algoritme har fastlagt, at algoritmen skal være anvendelig i de tilfælde, hvor der kun er adgang til datasæt med et beskedent antal elementer [77]. For at tage højde for den førnævnte problemstilling er det i C4.5 algoritmen fastlagt, at fejlklassificeringen fra træningsdataet skal korrigeres, når den anvendes til begrænsning af klassificeringstræets størrelse.

Metoden til begrænsning af klassificeringstræets størrelse tager udgangspunkt i en beregning af fejlraten P i formel 7.15. Fejlraten kan antage værdier mellem nul og én, og er et mål for hvor stor en andel af elementerne i knuden, der kan forventes at blive fejlklassificeret.

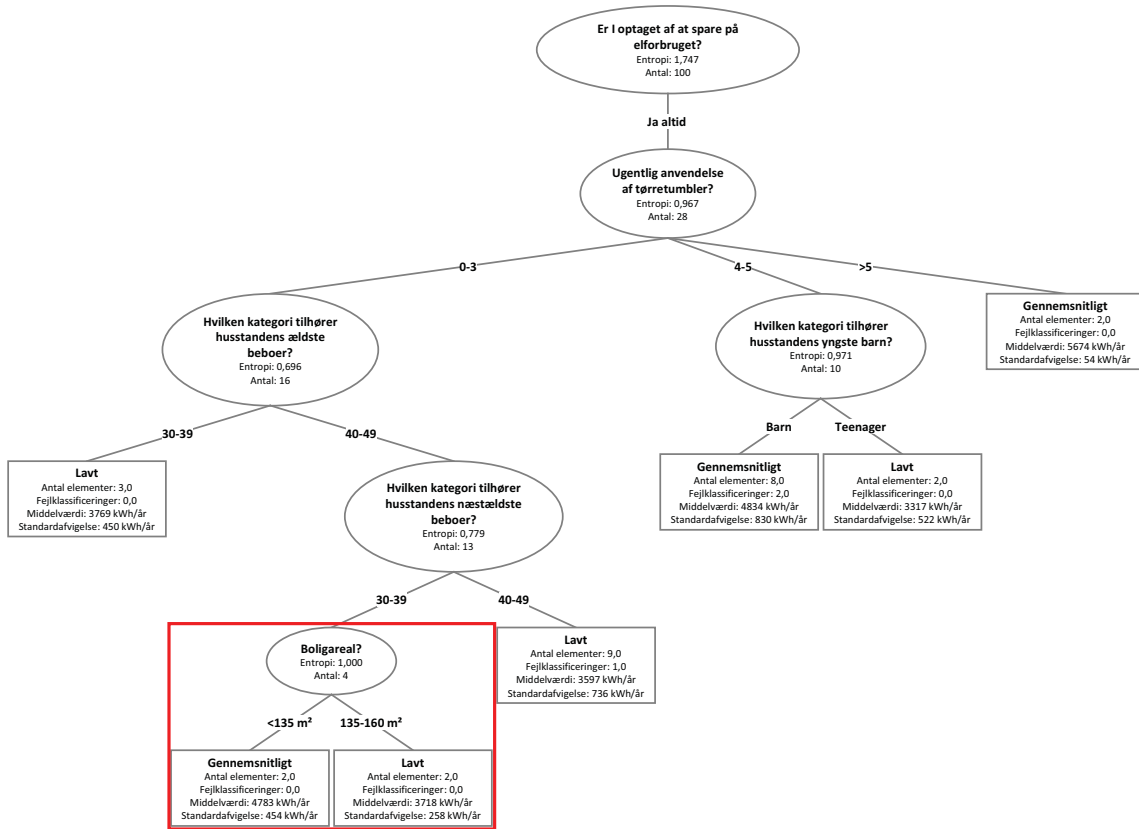
$$CF = \sum_{x=0}^E \left(\frac{N!}{(x!)(N-x)!} \cdot P^x \cdot (1-P)^{N-x} \right) \quad (7.15)$$

Hvor:

CF	Faktor, der som standard er 0,25 i C4.5 algoritmen [77].
E	Antal fejlplacerede elementer i knudepunktet.
N	Samlet antal elementer i knudepunktet.
P	Fejlrate.

Det fremgår af formel 7.15, at der indgår en CF -faktor til beregning af fejlraten. Denne faktor antager som standard en værdi på 0,25, hvilket er fastlagt af udvikleren af algoritmen [77]. En ændring af faktorens størrelse kan foretages ved udviklingen af et klassificeringstræ, men det vil ændre klassificeringstræets størrelse. En lavere CF -værdi vil resultere i, at en højere fejlrate accepteres, hvormed klassificeringstræet størrelse kan blive mindre [7]. Omvendt vil en højere CF -værdi betyde en lavere fejlrate, hvormed klassificeringstræet nødvendigvis skal være større, for at klassificere flere elementer korrekt [7]. I det nærværende projekt er der anvendt en CF -værdi på 0,25. I appendiks R fremgår en uddybende beskrivelse af CF -faktorens betydning for klassificeringstræets størrelse og nøjagtighed. Desuden fremgår der i samme appendiks en diskussion af de grundlæggende antagelser for formel 7.15 og anvendelsen af denne.

I det efterfølgende eksemplificeres anvendelsen af metoden til begrænsningen af klassificeringstræets størrelse. Der tages udgangspunkt i eksemplet på figur 7.3, der er et udsnit af det oprindelige klassificeringstræ, som er illustreret i *tegning A*. Det undersøges, hvorvidt I-knuden i det røde felt på figur 7.3 kan erstattes med en L-knude. For at undersøge dette skal fejlraten beregnes for alle knuderne i det røde felt. Begge L-knuder indeholder to elementer, hvor ingen af disse er fejlklassificeret. Såfremt I-knuden erstattes med en L-knude, vil denne indeholde fire elementer, hvoraf to af disse vil blive fejlklassificeret.



Figur 7.3: Udsnit af klassificeringstræet, der fremgår af vedlagte *tegning A* til eksemplificering af hvorledes klassificeringstræets størrelse begrænses.

Fejlraten for de to L-knuder i det røde felt kan beregnes ved anvendelse af formel 7.15. Parametrene, der skal anvendes i de to tilfælde, fremgår af formel 7.16, hvor fejlraten kan beregnes, ved isolering af $P_{L-knude,1}$. Da begge L-knuder indeholder to elementer, hvoraf ingen er fejlklassificeret, vil fejlraten have samme størrelse i begge knuder. Det medfører at $P_{L-knude,1} = P_{L-knude,2}$.

$$CF = \sum_{x=0}^E \left(\frac{N!}{(x!)(N-x)!} \cdot (P_{L-knude,1})^x \cdot (1 - P_{L-knude,1})^{N-x} \right) \Rightarrow$$

$$0,25 = \sum_{x=0}^0 \left(\frac{2!}{(x!)(2-x)!} \cdot (P_{L-knude,1})^x \cdot (1 - P_{L-knude,1})^{2-x} \right) \Rightarrow$$

$$P_{L-knude,1} = 0,500 \quad (7.16)$$

Fejlraten for I-knuden kan beregnes på tilsvarende vis, hvor der blot anvendes oplysningerne om, at knuden vil indeholde fire elementer med to fejlklassificeringer. En beregning af fejlraten for I-knuden vil give $P_{I-knude} = 0,757$ ved anvendelse af formel 7.15. De beregnede fejlratener skal multipliceres med antallet af elementer i knuderne, for at beregne det antal fejl-

klassificeringer, der skal sammenholdes ifølge C4.5 algoritmen. Resultatet af dette fremgår af formel 7.17.

$$\begin{aligned} F_{L-knode,1+2} &= N_1 \cdot P_{L-knode,1} + N_2 \cdot P_{L-knode,2} = 2 \cdot 0,500 + 2 \cdot 0,500 = 1,000 \\ F_{I-knode} &= N \cdot P_{I-knode} = 2 \cdot 0,757 = 1,514 \end{aligned} \quad (7.17)$$

Det fremgår af formel 7.17, at antallet af fejlklassificeringer er lavere for L-knuderne end for I-knuden ($F_{L-knode,1+2} < F_{I-knode}$). Det betyder, at klassificeringstræets størrelse ikke skal begrænses ved denne I-knode. På trods af dette udfald skal der, ifølge C4.5 algoritmen, foretages tilsvarende beregninger for knudepunkter højere oppe i klassificeringstræet. I eksemplet på figur 7.3 viser beregninger blandt andet, at I-knuden med attributten omhandler alderen på husstandens ældste beboer skal erstattes med en L-knode. Foretages tilsvarende beregninger for hele klassificeringstræet vil det resultere i, at det skal begrænses til den størrelse, som fremgår af vedlagte *tegning B*. Klassificeringstræet ender med at have 10 R- og I-knuder og yderligere 19 L-knuder, hvor i alt femten ud af hundrede elementer fejlklassificeres. Denne beregningsproces har begrænset det samlede antal knudepunkter i klassificeringstræet med 45 % i forhold til udgangspunktet i den vedlagte *tegning A*.

7.4 Evaluering af et klassificeringstræets nøjagtighed

Efter et klassificeringstræ er udviklet, skal dets nøjagtighed for klassificeringen af henholdsvis det anvendte trænings- og testdata vurderes. Der er flere forskellige mål, der kan anvendes i forbindelse med denne vurdering, hvor den efterfølgende beskrivelse tager udgangspunkt i de mål, som programmet Weka præsenterer i resultatfilen. Resultatfilen for klassificeringstræet, der er udviklet på baggrund af fiktivt data for 100 husstande, fremgår af appendiks Q. Denne resultatfilen er delt op i flere afsnit, hvor begreberne i afsnittet *Evaluation on training set* beskriver nøjagtigheden for et regressionstræ, hvorfor disse ikke beskrives nærmere i dette afsnit [11]. Målene for nøjagtigheden, der fokuseres på i nærværende afsnit, er at finde i under *Detailed accuracy by class* i resultatfilen. Blandt disse mål er; *confusion matrix*, *accuracy*, *TP-rate*, *FP-rate*, *precision* og *F-measure*. Betydningen af disse fremgår af tabel 7.11, hvor en uddybende beskrivelse og eksemplificering fremgår af appendiks S.1. I det efterfølgende vurderes de førnævnte mål for nøjagtigheden af det udviklede klassificeringstræ.

Parameter	Beskrivelse
Confusion matrix	Denne matrice danner grundlaget for alle de efterfølgende mål for nøjagtigheden af klassificeringstræet. Matricen indeholder oplysninger om, hvorledes de enkelte elementer er klassificeret i forhold til elementernes faktiske klasseinddeling. [82]
Accuracy	Der beregnes én værdi af <i>accuracy</i> for klassificeringstræet. Denne værdi angiver andelen af det samlede elementer, der er klassificeret korrekt, i forhold til det totale antal elementer i det aktuelle datasæt. [82]
True positive rate (TP-rate)	Der beregnes én værdi af <i>TP-rate</i> for <i>hver</i> klasseinddeling. Disse angiver et mål for andelen af korrekte klassificerede elementer i den aktuelle klasseinddeling i forhold til det totale antal elementer i datasættet, der har denne klasseinddeling. [11, 106]
False positive rate (FP-rate)	Der beregnes én værdi af <i>FP-rate</i> for <i>hver</i> klasseinddeling. Disse angiver et mål for andelen af fejlklassificerede elementer i den aktuelle klasseinddeling i forhold til det totale antal elementer i datasættet, der ikke har den aktuelle klasseinddeling. [11, 106]
Precision	Der beregnes én værdi for <i>hver</i> klasseinddeling. Disse angiver forholdet mellem antallet af elementer, der er klassificeret korrekt i den aktuelle klasseinddeling, og det samlede antal elementer, der er klassificeret med denne klasseinddeling. [105]
F-measure	Der beregnes én værdi for <i>hver</i> klasseinddeling. Disse angiver et forhold mellem antallet af elementer, der er klassificeret korrekt i den aktuelle klasseinddeling, og det antal elementer, der enten er klassificeret med denne klasseinddeling, eller har denne klasseinddeling i praksis. [105]

Tabel 7.11: Definition af nøglebegreber som anvendes til vurdering af et klassificeringstræs nøjagtighed i Weka.

7.4.1 Confusion matrix

Denne matrice indeholder oplysninger om, hvorledes de enkelte elementer er klassificeret i forhold til elementernes virkelige klasseinddeling. For det udviklede klassificeringstræ fremgår denne *confusion matrix* af tabel 7.12. Søjlerne i matricen repræsenterer de elementer, som klassificeringstræet har estimeret at tilhøre de enkelte klasseinddelinger, og rækkerne repræsenterer den virkelige klasseinddeling for elementerne. Af matricen, der er illustreret i tabel 7.12, fremgår det, at 31, 37, 13 og 4 elementer er klassificeret korrekt af klassificeringstræet, som henholdsvis *lavt*, *gennemsnitligt*, *højt* og *meget højt* forbrug, markeret med rød tekst i tabel 7.12. De elementer, der er klassificeret korrekt, fremgår altid af diagonalen, som går fra det øverste venstre hjørne mod nederste højre hjørne i matricen. Ud over at illustrere de elementer, der er klassificeret korrekt, angiver matricen ligeledes, hvilke elementer der er fejlklassificeret. Matricen illustrerer blandt andet, at fem elementer er klassificeret som *lavt*, til trods for de reelt tilhører klasseinddelingen *gennemsnitligt*. Ud fra *confusion matrixen* fremgår det yderligere, at klassificeringstræet tildeler fire elementer klasseinddelingen *meget højt*, med tre elementer fejlklassificeret som *højt*. Det vil sige, at næsten halvdelen af elementerne, der tilhører klasseinddelingen *meget højt*, bliver fejlklassificeret, hvilket medfører, at det er denne klasseinddeling, som klassificeringstræet håndterer dårligst. Denne svaghed

kan skyldes, at det er klasseinddelingen *meget højt*, der er færrest elementer af i den aktuelle datamængde. Det betyder, at når der nås en L-knude, vil sandsynligheden for, at denne knude tildeles klasseinddelingen *meget højt* være beskeden, da der ofte kan forekomme et større antal elementer med en anden klasseinddeling. [82]

Confusion matrix				
↓ virkelig klasse / → klassificeret som:	a	b	c	d
a: Lavt	31	4	0	0
b: Gennemsnitligt	5	37	0	0
c: Meget højt	0	0	4	3
d: Højt	0	3	0	13

Tabel 7.12: Resultater for det udviklede klassificeringstræ angivet i en *confusion matrix*.

7.4.2 Andelen af korrekt klassificerede elementer i klassificeringstræet

Før udviklingen af et klassificeringstræ påbegyndes er der behov for, jf. afsnit 4.3 og figur 4.5, at fastsætte en nøjagtighed som klassificeringstræet kan testes imod, før det besluttes at anvende klassificeringstræet på nyt data. Klassificeringstræets nøjagtighed vurderes på baggrund af både trænings- og testdata, idet træets evner til klassificering af elementer er bedre for det træningsdata, det er udviklet på baggrund af, frem for testdata. Hvilken nøjagtighed klassificeringstræet kan have, afhænger af hvilket data modellen skal anvendes til at vurdere, samt hvad konsekvensen for en fejlklassificering er. Ved udvikling af et klassificeringstræ, der skal anvendes inden for det medicinske område, kan der være en større menneskelig konsekvens ved en fejlklassificering, end der kan inde for dette fagområde, hvor det drejer sig om brugeradfærd og energiforbrug. [82]

Nøjagtigheden (*accuracy*) af et klassificeringstræ afhænger af hvor mange elementer, ud af den samlede datamængde, der er klassificeret korrekt. *Accuracy* for det udviklede klassificeringstræ er på 85 % for træningsdata, hvilket fremgår af appendiks Q. Resultatet for *accuracy* betyder, at klassificeringstræet klassificerer 85 ud af 100 elementer korrekt, hvilket må betragtes som værende acceptabelt, taget i betragtning af hvilket formål klassificeringstræet har.

7.4.3 True positive rate og false positive rate

TP-rate og *FP-rate* er defineret i tabel 7.11. Resultaterne af de to førnævnte mål for det udviklede klassificeringstræ fremgår af tabel 7.13. Af denne fremgår det, at *TP-rate* er relativt høj for klasseinddelingerne *lavt*, *gennemsnit* og *højt*. Dette betyder, at størstedelen af elementerne med de respektive klasseinddelinger klassificeres korrekt. Derimod er *TP-rate* for klasseinddelingen *meget højt* lig 0,57, hvilket betyder, at klassificeringstræet kun klassificerer lidt over halvdelen af elementerne i klasseinddelingen korrekt. Vurderes der derimod på klassificeringstræets værdier for *FP-rate* fremgår det, at disse er lave for alle fire klasseinddelinger. Dette betyder, at der for hver klasseinddeling ikke fejlklassificeres specielt mange elementer med en anden klasseinddeling.

Begreb	Klasseinddeling			
	Lavt	Gennemsnitligt	Højt	Meget højt
TP-rate	0,89	0,88	0,81	0,57
FP-rate	0,08	0,12	0,04	0,00

Tabel 7.13: Værdier af *TP-rate* og *FP-rate* for det udviklede klassificeringstræ.

7.4.4 Precision

Definitionen af *precision* fremgår af tabel 7.11, hvor resultaterne af denne, for det udviklede klassificeringstræ, fremgår af tabel 7.14. *Precision* kan variere mellem værdien 0 – 1, hvor en høj værdi betyder, at klassificeringstræet er god til at klassificere elementerne i den pågældende klasseinddeling korrekt [105]. Ved vurdering af resultaterne for det udviklede klassificeringstræ fremgår det, at der opnås en maksimal værdi for *precision* for klasseinddelingen *meget højt*. Det skyldes, at der ikke er nogle elementer med andre klasseinddelinger, der tildeles klasseinddelingen *meget højt*. For de øvrige klasseinddelinger, har alle en værdi over 0,8, hvilket betyder, at der for disse klasseinddelinger ikke er en særlig stor andel af elementer fra de øvrige klasseinddelinger, der fejlklassificeres som tilhørende den pågældende klasseinddeling.

Begreb	Klasseinddeling			
	Lavt	Gennemsnitligt	Højt	Meget højt
Precision	0,86	0,88	0,81	1,00

Tabel 7.14: Værdier for *precision* for det udviklede klassificeringstræ.

7.4.5 F-measure

F-measure er defineret i tabel 7.11. Værdien for *F-measure* kan variere imellem 0 – 1, hvor nøjagtigheden for klassificeringstræet er bedst ved en værdi på 1 [105]. Denne værdi opnås, når der ingen fejlklassificerede elementer er for den pågældende klasseinddeling. Resultatene af *F-measure* for det udviklede klassificeringstræ fremgår af tabel 7.15. Af den førnævnte tabel fremgår det, at værdierne for *F-measure* varierer imellem 0,73 – 0,87. Værdierne må for klasseinddelingerne *lavt*, *gennemsnitligt* og *højt* anses som værende relativt høje, hvori-mod værdien for klasseinddelingen *meget højt* er en anelse lavere. De høje værdier betyder, at størstedelen af elementerne i de respektive klasseinddelinger er klassificeret korrekt, når det sammenholdes med det antal elementer, der enten er klassificeret med den aktuelle klasseinddeling, eller har denne klasseinddeling i praksis.

Begreb	Klasseinddeling			
	Lavt	Gennemsnitligt	Højt	Meget højt
F-measure	0,87	0,86	0,81	0,73

Tabel 7.15: Værdier for *F-measure* for det udviklede klassificeringstræ.

Anvendelse af klassificeringstræer

Der er i det foregående kapitel redegjort for, hvorledes der kan udvikles et klassificeringstræ på baggrund af det data, der er beskrevet i kapitel 6. Det nærværende kapitel omhandler en beskrivelse af, hvorledes et klassificeringstræ anvendes, og hvilke målgrupper der er for de tidligere omtalte klassificeringstræer. Ydermere beskrives hvilke problemstillinger, der kan opstå, hvis klassificeringstræerne skal udvikles på en større og mere repræsentativ datamængde.

Udviklingsprocessen for klassificeringstræer omfatter to overordnede dele, jf. figur 4.5 på side 95. I det foregående kapitel er der redegjort for udviklingen af et klassificeringstræ på baggrund af træningsdata. Den efterfølgende proces er at afprøve klassificeringstræets nøjagtighed med udgangspunkt i testdata. Denne proces går i hovedtræk ud på, at testdata skal anvendes i klassificeringstræet og tildeles en klasse ud fra attributternes informationer. Den tildelte klasse sammenlignes med den faktiske klasse, hvor der efterfølgende beregnes en nøjagtighed for klassificeringstræet. Dette udføres med udgangspunkt i forholdet mellem korrekte og fejlklassificerede elementer, hvor denne nøjagtighed for klassificeringstræet angiver, hvor præcist klassificeringstræet håndterer testdata. Denne proces er imidlertid ikke udført på eksemplet fra det foregående kapitel. Det skyldes, som tidligere omtalt, at det fiktive træningsdata er genereret tilfældigt. En angivelse af klassificeringstræets nøjagtighed på baggrund af tilfældigt genereret testdata, med samme forudsætninger som træningsdata, vurderes i nærværende projekt ikke at medføre anvendelige resultater, hvor der kan konkluderes på klassificeringstræets nøjagtighed.

8.1 Klassificeringstræernes målgruppe

Der er i kapitel 5 introduceret nogle muligheder for hvilke typer klassificeringstræer, der kan udvikles, samt hvilke overvejelser der skal foretages med hensyn til klasse og attributter. På denne baggrund kan der være flere målgrupper med klassificeringstræerne, der afhænger af det aktuelle datagrundlag. I nærværende projekt fokuseres på udvikling og anvendelse af tre separate klassificeringstræer for henholdsvis energiforbrug til rumopvarmning, forbrug af varmt brugsvand samt elforbrug. Et af formålene med disse klassificeringstræer er, at sætte fokus på energiforbruget i eksisterende boliger med henblik på at opnå besparelser.

Det fremgår af tidligere afsnit i rapporten, at udviklede klassificeringstræer er simple at anvende, hvor eksempelvis elforbruget kan estimeres ved, at spørgsmålene i de enkelte knuder besvares. Den simple anvendelse er en af klassificeringstræernes store fordele, hvilket

medfører, at de kan henvende sig til en stor målgruppe. I nærværende projekt er fokus på energiforbruget i boliger, hvor det anbefales, at inddrage demografiske, tekniske såvel som adfærdsmæssige forhold i klassificeringstræerne, jf. afsnit 5.3. Omfanget af data inden for de enkelte områder kan variere afhængigt af hvilken målgruppe klassificeringstræet skal henvende sig til.

En af mulighederne er, at udvikle et klassificeringstræ, der kun inddrager tekniske attributter i et omfang, så det er muligt for almene boligejere at besvare spørgsmålene. Derved kan klassificeringstræerne anvendes af boligejerne, til vurdering af hvilke områder det bedst kan betale sig for dem at optimere. Det kan foretages ved, at de, ud fra besvarelse af spørgsmålene i de enkelte knuder, vurderer, hvor i klassificeringstræet de placeres. Ved at følge grenene op gennem klassificeringstræet, kan boligejerne vurdere hvilke forhold, i form af attributter, de har mulighed for at optimere. Derved kan de udvælge de forhold, der falder dem mest naturligt, og direkte vurdere ud fra klassificeringstræet hvilke reduktioner af energiforbruget, det typisk kan forventes at medføre. Ved at optimere på de forhold, der falder dem mest naturligt, er der størst chance for, at de energibesparende tiltag realiseres, og den ønskede effekt opnås.

For at sikre at klassificeringstræet vil blive anvendt, er der behov for, at der føres oplysningskampagner i diverse medier, og eventuelt udbreder anvendelsen gennem ejendomsmæglere. Dette er der behov for, for at vække beboernes interesse, og for at gøre opmærksom på at der er mulighed for at opnå indsigt i hvor og hvordan der i husstanden kan opnås besparelser. Hvis der ikke fokuseres på at udbrede modellen, vil det kun være den type af beboere, som normalt har interesse for dette område, der vil anvende modellen. Hvis denne type af beboere allerede har fokus på dette område, er det sandsynligvis ikke dem, der vil blive introduceret til hvordan den største besparelse kan opnås. Derfor er der behov for at udbrede modellen, for at forsøge at opnå, at klassificeringstræet vil blive anvendt af den brede del af befolkningen, og dermed sørge for at det beskrevne potentiale udnyttes mest optimalt.

En anden målgruppe, der kan anvende klassificeringstræerne, er energirådgivere. Da disse har større indsigt i de tekniske forhold i boliger, er der færre begrænsninger for hvilke forhold, der kan medtages i klassificeringstræet. Energirådgivere vurderes at kunne anvende klassificeringstræet til at rådgive boligejere om hvilke energibesparende tiltag, der vil være fordelagtige for disse at udføre. I den forbindelse kan energirådgiverne anvende klassificeringstræet til at foretage estimeringen af energiforbruget, samt i formidlingen af hvilke konsekvenser beboerens adfærd har for forbrugets størrelse.

Ligesom for beboerne skal der for denne gruppe også lægges fokus på udbredelsen af klassificeringstræets anvendelse. Dette kan enten ske igennem lovgivningen hvor det pålægges energirådgiverne at anvende metoden. Eller ved at energirådgiverne skal oplyse deres kunder om, at der mulighed for, at de selv kan anvende denne metode, hvis de ønsker at opnå et større indblik i størrelsen af deres energiforbrug.

I forbindelse med formidling af forventet energiforbrug er der et andet fagområde, som klassificeringstræerne vurderes at være anvendelige for. I forbindelse med boligkøb formidler ejendomsmæglere energimærkninger af boliger til eventuelle købere. I den forbindelse vurderes der i nærværende projekt at være potentiale i, at ejendomsmæglere kan give et estimat af hvilket energiforbrug, primært med hensyn til varmemeforbrug, købere kan forvente. Ejendomsmæglere kan i disse situationer bidrage med informationer om tekniske forhold for boligen. Fordelen ved denne anvendelse er, at køber kan få oplyst et forventet energiforbrug, som

kan anvendes i forbindelse med budgettering, og ejendomsmægleren kan udføre en mere omfattende vejledning til køber.

8.2 Betydningen af datagrundlagets størrelse

I det foregående afsnit er det klarlagt hvilke målgrupper, der kan drage nytte af klassificeringstræer til estimering af energiforbrug. For at de omtalte klassificeringstræer kan anvendes i de forskellige situationer, kræver det, at disse er udviklet på baggrund af et repræsentativt datagrundlag. Eksemplificeringen af klassificeringstræers udvikling er foretaget på baggrund af data fra 100 fiktive besvarelser. Dette har resulteret i et overskueligt klassificeringstræ, hvor egenskaben til at klassificere elementerne korrekt er på 85 %, hvilket må betragtes som værende acceptabelt. Der er, som beskrevet i det foregående kapitel, flere steder i udviklingen af klassificeringstræet, hvor det valgte data kan være med til at påvirke det endelige resultat. Disse kritiske punkter er blandt andet mængden af data, og hvilke attributter der udvælges til klassificeringstræet. Såfremt der på et senere tidspunkt skal udvikles et klassificeringstræ på baggrund af en større og mere repræsentativ datamængde, beskrives det i nærværende afsnit, hvilken indflydelse mængden af data har på udviklingen af klassificeringstræet, og hvilke konsekvenser dette har.

Til eksemplificering af ovenstående er der generet fiktivt data for 1000 boliger. Datagenereringen er foretaget på baggrund af de samme forudsætninger, der beskrevet i kapitel 6. Det udviklede klassificeringstræ er illustreret på vedlagte *tegning C*, hvor det fremgår, at dette er betragteligt større end klassificeringstræet udviklet på baggrund af data for 100 boliger, jf. vedlagte *tegning B*. Begge de førnævnte klassificeringstræer er udviklet på baggrund af en CF -faktor på 0,25 samt et krav til det mindste antal elementer i en knude før en inddeling på to. En sammenligning af antallet af knudepunkter for de to forskellige klassificeringstræer fremgår af tabel 8.1. Det er værd at bemærke, at for de to klassificeringstræer tilnærmelsesvis har samme nøjagtighed, kræver det over fire gange så mange knuder at klassificere de 1000 elementer.

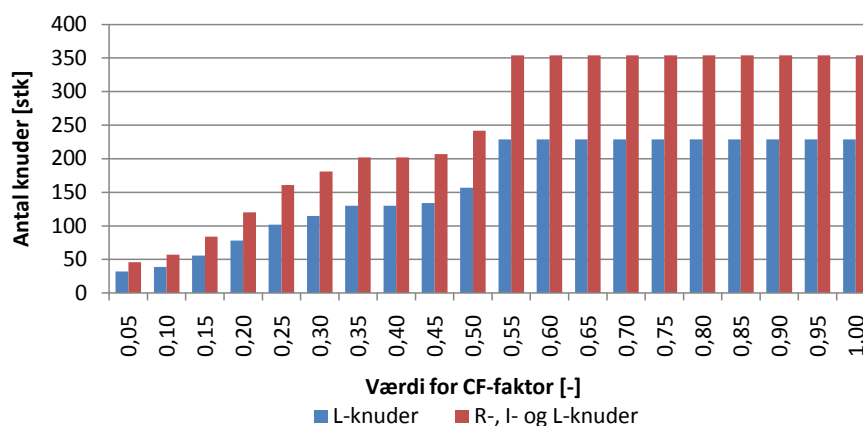
Beskrivelse	Datagrundlagets størrelse		
	100 boliger	1000 boliger	1000 boliger
Samlet antal knuder [<i>stk</i>]	29	161	66
Antal L-knuder [<i>stk</i>]	19	102	45
Andelen af fejlklassificerede elementer [%]	15	16	22
Mindste antal elementer før inddeling [<i>stk</i>]	2	2	7
CF -faktor [–]	0,25	0,25	0,35

Tabel 8.1: Antal L-knuder og det samlede antal knuder i de to forskellige klassificeringstræer.

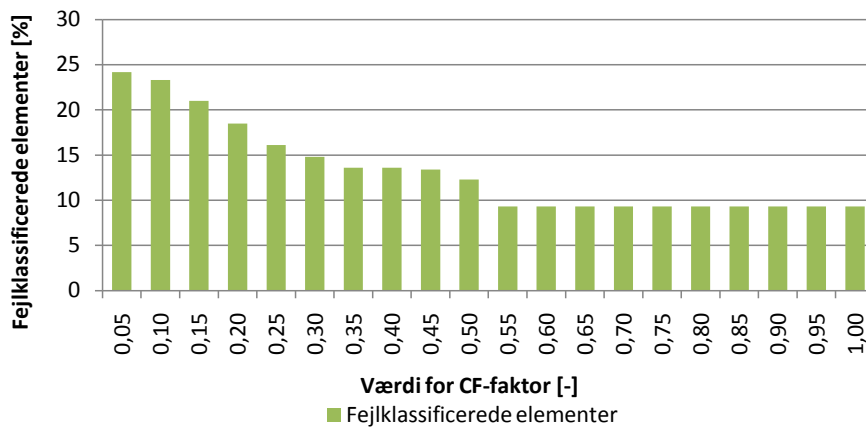
Denne størrelsesforskel kan visuelt gøre klassificeringstræet mere uoverskueligt, men til trods for dette besværliggøres anvendelsen af klassificeringstræet ikke væsentligt. Selvom klassificeringstræet er udviklet på baggrund af data fra 1000 boliger, er der risiko for at det ikke er alle mulige udfald fra en I-knude, der indeholder elementer. Hvis denne situation stadig er til stede efter udviklingsprocessen er gennemført betyder det, at elementet med knudens tiltænkte

udfald og de forudgående betingelser ikke er til stede, hvilket betyder, at denne knude ikke indgår i det samlede klassificeringstræ. Det har den konsekvens, at der kan opstå situationer, når klassificeringstræet anvendes, hvor en familie ikke kan tildeles en klasseinddeling, hvilket gør træet ubrugeligt for denne familie. Det optimale træ vil indeholde data for alle mulige udfald fra en I-knude, men idet et elements placering i en knude er betinget af de foregående udfald, er der ingen garanti for, at alle muligheder for udfald indgår i klassificeringstræet. Denne problemstilling kan imødegås, under forudsætning af anvendelsen af kategoriske attributter, ved at revurdere intervallets inddeling. Hvis dette ikke afhjælper problemet, eller gør klassificeringstræets nøjagtighed uacceptabel, er der kun mulighed for at imødegå problemstillingen, ved at indsamle viden om den pågældende familie. Når der er tilstrækkeligt med sådanne eksempler kan disse inddrages i en ny udvikling af klassificeringstræet.

Det er beskrevet i afsnit 7.3, at klassificeringstræets størrelse kan påvirkes af den omtalte CF -faktor. Nærmere information om denne fremgår af appendiks R. I klassificeringstræet, der er udviklet på baggrund af 1000 elementer, har CF -faktoren betydelig indflydelse på både dets størrelse og nøjagtighed. Dette er illustreret på figur 8.1 og 8.2. Det er tidligere omtalt, at standardværdien for CF -faktoren i C4.5 algoritmen er 0,25. Det fremgår af de førnævnte figurer, at denne værdi, i det aktuelle tilfælde, vil medføre et kompromis mellem et klassificeringstræ med få knuder og en høj nøjagtighed.

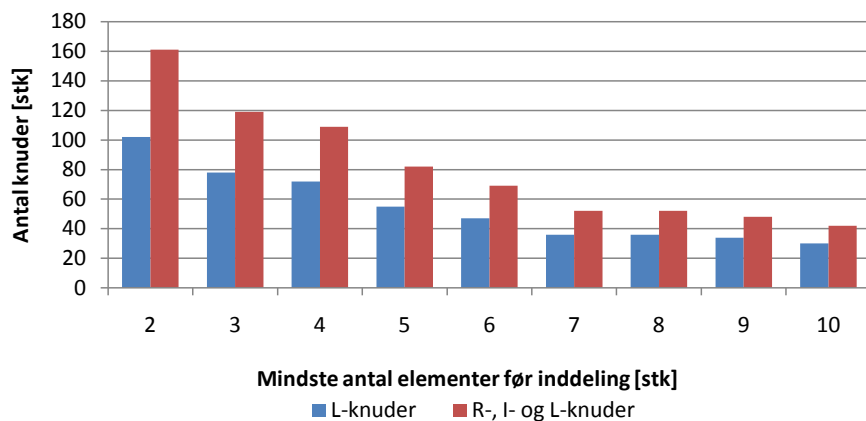


Figur 8.1: Illustrering af CF -faktorens betydning for antallet af L-knuder og det samlede antal knuder i klassificeringstræet for elforbruget. Det er forudsat, at kravet til det mindste antal elementer i en knude før en inddeling er to.

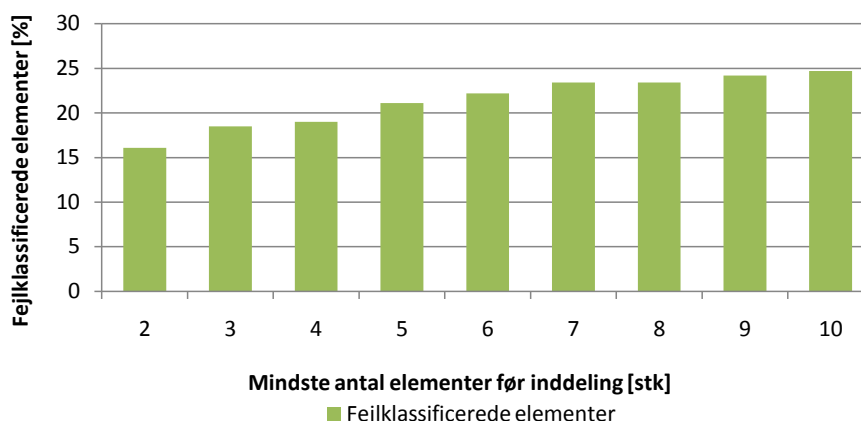


Figur 8.2: Illustrering af CF -faktorens betydning for andelen af fejlklassificerede elementer i klassificeringstræet for elforbruget. Det er forudsat, at kravet til det mindste antal elementer i en knude før en inddeling er to.

Foruden den omtalte CF -faktor er der mulighed for at påvirke klassificeringstræets størrelse, ved at ændre kravet til det mindste antal elementer der skal være i en knude, før der kan foretages en yderligere inddeling af disse. Dette krav antager som standard en værdi på to, hvilket ligeledes er anvendt i klassificeringstræerne udviklet på både 100 og 1000 elementer. En ændring af dette krav er en mulighed for at påvirke klassificeringstræets størrelse, hvis det som udgangspunkt kræver et stort klassificeringstræ at inddele det aktuelle datasæt. Betydningen af dette krav er illustreret på figur 8.3 og 8.4, der angiver henholdsvis antallet af knuder og andelen af fejlklassificerede elementer.



Figur 8.3: Betydningen af kravet til det mindste antal elementer i en knude før en inddeling sammenholdt med antallet af knuder i klassificeringstræet. Figuren er gældende for en CF -faktor på 0,25.



Figur 8.4: Betydningen af kravet til det mindste antal elementer i en knude før en inddeling sammenholdt med andelen af fejlklassificerede elementer i klassificeringstræet. Figuren er gældende for en CF -faktor på 0,25.

Da datagrundlaget for de tidligere omtalte klassificeringstræer er fiktivt, er der ikke grundlag for at vurdere, hvilken størrelse klassificeringstræet vil antage, hvis det udvikles på baggrund af et repræsentativt datasæt. Der er mulighed for, at klassificeringstræet antager en overskuelig størrelse og kan klassificere elementerne med en høj nøjagtighed. Modsat er der risiko for, at dataet enten ikke vil medføre brugbare resultater eller et stort og uoverskueligt klassificeringstræ. Såfremt det sidstnævnte er tilfældet, kan der, som illustreret i det foregående, justeres på standardværdierne for C4.5 algoritmen. Dette er udført for klassificeringstræet, der er udviklet på baggrund af 1000 besvarelser. Ved at ændre kravet til det mindste antal elementer i en knude, før en yderligere inddeling kan foretages, fra to til syv, samt ændre CF -faktoren til 0,35, vil der udvikles et klassificeringstræ med 66 knuder, jf. tabel 8.1. Af den førnævnte tabel fremgår det, at klassificeringstræets størrelse er reduceret med omkring 60 % uden dets nøjagtighed reduceres væsentligt.

Den foregående beskrivelse illustrerer, at til trods for at et klassificeringstræ, udviklet på baggrund af standardværdier, som udgangspunkt kan antage en uoverskuelig størrelse, er der muligheder for at optimere det. En reduceret størrelse af klassificeringstræet vil, som det fremgår af foregående eksempel, medføre en reduceret nøjagtighed. Såfremt dette foretages på et klassificeringstræ, udviklet på baggrund af repræsentativt data, bør det undersøges, hvor meget dette klassificerer elementerne forkert. Det vil eksempelvis sige, om elementer, hvis faktiske klasse er „lavt“, klassificeres som „gennemsnitligt“, „højt“ eller „meget højt“ og så fremdeles. Optimeringen af et klassificeringstræ er et kompromis mellem størrelse, nøjagtighed, og hvilke krav dette ønskes at opfylde.

Under udviklingen af klassificeringstræet er der risiko for, at der ikke fremkommer nogen klar sammenhæng mellem det udvalgte energiforbrug og beboernes adfærd. Når denne situation opstår, fremgår det klart, igennem visualiseringen af klassificeringstræet, elementernes placering i confusion matricen og derigennem antallet af elementer der er fejlklassificeret, at klassificeringstræet ikke kan finde en sammenhæng mellem de angivne attributter og den udvalgte klasse. For at imødegå denne problemstilling er det oplagt at vurdere inddelingen for klassen på ny og herefter udvikle træet igen for at undersøge om den udvalgte inddeling giver et bedre resultat. Derudover er der også mulighed for, ved anvendelse af kategoriske attributter, at revurdere intervallernes inddeling af disse. Hvis ingen af ovenstående mulighe-

der afhjælper problemet, må det herefter konkluderes, at metoden med klassificeringstræer ikke er anvendelig for den indsamlede datamængde. Indsamlingen af data er ikke nødvendigvis spildt, da der er mulighed for at anvende en anden analysemetode, eksempelvis en af metoderne der blev præsenteret i afsnit 4. Hvis der allerede er viden om emnet fra andre undersøgelser, kan det foreslås, at en lignende undersøgelse gennemføres, for at undersøge om de samme forhold gør sig gældende for det indsamlede data.

Diskussion

I de foregående kapitler er der gennemgået flere forskellige emner, der alle har det til fælles, at de enten er direkte eller indirekte knyttet til emnet brugeradfærd og energiforbrug. I litteraturstudiet er der behandlet både tekniske, demografiske og adfærdsrelaterede forhold i relation til hvilken påvirkning disse har på energiforbruget i boliger. Erfaringer fra litteraturstudiet er videreført til en beskrivelse af, hvorledes der kan udvikles klassificeringstræer til estimering af energiforbrug til rumopvarmning, forbrug af koldt eller varmt brugsvand og elforbrug. I løbet af arbejdsprocessen, fra udvælgelsen af en metode og frem til det færdigudviklede klassificeringstræ, er der fastlagt nogle forudsætninger, der har påvirket hele udviklingsprocessen. I nærværende kapitel diskuteres flere forskellige forhold, der har haft betydning for udviklingsprocessen og det endelige resultat.

Indeklima

Der er i nærværende projekt fokuseret på beboernes påvirkning af energiforbruget i boliger, jf. problemafgrænsningen i afsnit 2.5.1. Dette fokusområde er væsentligt, i forbindelse med indførelsen af strengere krav til hvor stort et energibehov, boliger kan godkendes på baggrund af, i *Bygningsreglementet* [66]. Til trods for at fokus er på energiforbruget i boliger, er indeklimaet i disse et forhold, der er mindst lige så væsentligt. Det skyldes, at de strengere krav til energibehovet har medført et øget fokus på optimering af dette, hvor konsekvensen heraf er, at indeklimaets betydning bliver undervurderet, hvorfor dette bliver forringet. Problemer med indeklimaet, specielt i lavenergi- og passivhuse, omfatter blandt andet for lave og høje lufttemperaturer [2]. Disse problemstillinger har medført et behov for øget fokus på indeklimaet i boliger, hvor der på intet tidspunkt under projektering eller ved renovering må gås på kompromis med eller ses bort fra dette.

Det er ikke muligt at vurdere indeklimaet i boliger ved brug af de klassificeringstræer, der er beskrevet i de foregående kapitler. At klassificeringstræerne ikke kan anvendes til vurdering af indeklimaet skyldes ikke direkte begrænsninger i metoden, men de forhold der er inddraget i denne i form af attributter. Det kan være en mulighed, at anvende et klassificeringstræ til eksempelvis at estimere hvilken kategori for indeklimaet der kan forventes i en bolig, eksempelvis ud fra definitionerne i *DS/EN/ISO 7730* [89]. Fordelen ved denne anvendelse er, at der kan udføres et estimat af indeklimaet, uden at der er behov for at udføre beregninger eller simuleringer. Der er imidlertid risiko for, at klassificeringstræet bliver stort og komplekst, såfremt dette anvendes til estimering af indeklimaet. Det skyldes, at klassificeringstræet skal estimere nogle dynamiske forhold, der kan være vidt forskellige fra bolig til bolig, som følge

af blandt andet konstruktioners varmetekniske egenskaber, vinduers orientering, varierende udeforhold etc.

En anden mulighed for at anvende et klassificeringstræ til vurdering af indeklima kan være, at dette udvikles til at beskrive nogle forskellige brugertyper, jf. afsnit 5.1. Det vil efterfølgende være muligt, at inddrage disse brugertyper i et hygrotermisk bygningssimuleringsprogram, for herigennem at vurdere hvilket indeklima der kan forventes i boligen. Ved at anvende denne løsning kan der tages højde for beboernes adfærd, og samtidig vurderes på både indeklima og energiforbrug. Det vil imidlertid kræve omfattende undersøgelser for at indsamle det nødvendige data, og der skal efterfølgende gennemføres en omfattende arbejdsproces, inden der opnås nogle simuleringsresultater.

Indsamling af data

Alle data, der er anvendt i nærværende projekt, er fiktive data, som er genereret på baggrund af oplysninger fra forskellige kilder, jf. kapitel 6. Såfremt et klassificeringstræ, på et senere tidspunkt, skal udvikles, er der behov for at indsamle en datamængde, som kan anvendes til både trænings- og testfasen. Den optimale måde hvorpå data til et klassificeringstræ kan indsamles, er ved en kombination af spørgeskemaer og forbrugsmålinger. Gennem spørgeskemaerne kan der opnås indsigt i beboernes vaner og holdninger, og oplysninger om energiforbrug kan anskaffes på baggrund af forbrugsmålinger.

Det er igennem litteraturstudiet erfaret, at der i flere af projekterne er introduceret en risiko for fejlfortolkning ved anvendelse af spørgeskemaer, fra både respondentens og interviewerens side. Risikoen ligger i, at beboerne kan tolke spørgsmålene forskelligt, når de skal besvares. Et eksempel herpå fremgår af nedenstående:

Hvor meget er du optaget af at passe på miljøet? [52, s. 47]

Spørgsmålet besvares ved, at respondenterne afkrydser én ud af fem svarmuligheder, der varierer fra „lidt“ til „meget“. Det ovenstående eksempel introducerer to problemstillinger; den første problemstilling omhandler beboernes tolkning af svarmulighederne, og den anden knytter sig hvordan besvarelsene senere hen analyseres. Problemstillingen omhandlende beboerne drejer sig om, at der kan være stor forskel på, hvordan beboerne tolker de angivne svarmuligheder. Med udgangspunkt i det ovenstående eksempel kan der for eksempel være stor forskel på, hvad beboerne betragter som værende „lidt“ og „meget“ optaget af at passe på miljøet. Derved fremgår spørgsmålet som et holdningsspørgsmål, hvor beboeren afgiver en besvarelse på baggrund af dennes tolkning af spørgsmål og svarmuligheder. Til trods for at beboerne selv skal fortolke disse, kan denne type spørgsmål godt anvendes i forbindelse med udvikling af klassificeringstræer. Denne type spørgsmål kan introducere nogle usikkerheder i forbindelse med den efterfølgende analyse af besvarelsene, da det ikke fremgår, hvorledes beboernes har fortolket spørgsmål og svarmuligheder.

Såfremt der skal foretages en analyse af besvarelsene, kan det være fordelagtigt, at reducere mængden af disse holdningsspørgsmål. Dette kan udføres ved at introducere spørgsmål, hvortil der er knyttet nogle handlinger til hver svarmulighed, som beboerne kan forholde sig til i forbindelse med besvarelsen. En anden mulighed kan være, at udfærdige nogle specifikke spørgsmål, som direkte knytter sig til nogle bestemte vaner eller aktiviteter. Derved kan beboernes vaner analyseres ud fra en række bestemte handlinger, frem for der analyseres

på nogle spørgsmål, hvor de forskellige fortolkninger af spørgsmål og svarmuligheder ikke er kendt.

I projektet fra *Knudsen et al.* [52] er der udarbejdet nogle spørgsmål omhandlende forskellige handlinger, som beboerne kan udføre i bestemte scenarier. I et af spørgsmålene opstilles et scenarie, hvor beboerne skal tage stilling til, hvilke handlinger de udfører, og ikke mindst i hvilken rækkefølge handlingerne udføres, når der er for koldt i boligen. I dette eksempel virker spørgsmålet lige til, men af besvarelsene fremgår der stadig en stor forskel på, hvorledes beboerne svarer. Nogle besvarer spørgsmålet meget overfladisk, og angiver at nogle af handlingerne udføres samtidigt, hvilket gør det vanskeligt at analysere besvarelsene efterfølgende. Andre beboere besvarer spørgsmålene meget udførligt, og beskriver yderligere hvordan den primære styring af varmeanlægget foregår. Inden for de enkelte husstande er der ligeledes observeret forskelle i hvilken rækkefølge, de enkelte beboere udfører handlingerne. Denne forskel, både internt i hjemmet og i forhold til de øvrige husstande i undersøgelsen, gør det vanskeligt at analysere de aktuelle besvarelser. Derfor er indsamlingen af data en balancegang mellem hvor meget data, der er tilstrækkeligt, og hvor udførlige besvarelser er der behov for, således det ikke ender i en uigennemskuelig og uanvendelig mængde data.

Udfordringer med udvikling af spørgeskemaer kan ligeledes opstå i forbindelse med, hvor mange svarmuligheder der skal være for de udvalgte spørgsmål. Igennem litteraturstudiet fremgår der eksempler fra to til fem svarmuligheder, hvor de to muligheder dækker over „ja“ eller „nej“ besvarelser. I andre tilfælde kan der være fem svarmuligheder, som eksempelvis fra „lidt“ til „meget“, som illustreret i det foregående, eller fra „meget uenig“ til „meget enig“. Ved anvendelse af fem svarmuligheder er der som regel mulighed for at angive et neutralt svar, hvor der ikke behøves stillingtagen til spørgsmålet. Det kan imidlertid diskuteres, om denne mulighed skal forefindes. I nogle tilfælde kan spørgsmålet omhandle et emne, der ingen interesse har for den pågældende respondent, hvorfor denne ofte ingen holdning har til problemstillingen. I dette tilfælde kan det være relevant at have en neutral svarmulighed. I andre situationer kan en neutral svarmulighed anvendes som en „smutvej“ til ikke at skulle tage stilling til et spørgsmål. Det skal derfor vurderes for hvert spørgsmål, om en neutral holdning er velkommen i undersøgelsen. Samtidig skal det bemærkes, at manglende svarmuligheder kan irritere respondenterne, som derfor, hvis der er mulighed for det, springer spørgsmålet over, hvilket i nogle tilfælde kan være værre end et neutralt svar.

Det sidste forhold, der skal tages højde for ved udformningen af spørgsmålene, er, at spørgsmålene ikke skal indikere, hvad intervieweren opfatter som en „rigtig“ eller „forkert“ besvarelse. Det kan medføre, at respondenterne giver det svar, som den tror, at intervieweren ønsker at høre. Dette forhold er blandt andet observeret i undersøgelser fra *Kirsten Gram-Hanssen* [40] og *Pedersen et al.* [74], hvor den såkaldte intervieweffekt eksempelvis fremgik af beboernes besvarelser omhandlende deres adfærd og bestemte handlinger i hjemmet. Denne intervieweffekt kan have indflydelse på de besvarelser, der indsamles gennem en spørgeskemaundersøgelse. Derved kan det medføre problemstillinger i forbindelse med analysering af data, da spørgsmål, der er under indflydelse af en intervieweffekt, ikke altid afspejler respondenterne reelle handlinger.

De ovenstående forhold har betydning, for hvorledes et spørgeskema skal udformes, hvis det skal anvendes som datagrundlag for et klassificeringstræ. Foruden de ovenstående forhold skal det tilsvarende vurderes, hvor høj detaljeringsgrad af de enkelte attributter der er nødvendig i den aktuelle situation. Da indsamling af data kan være en tidskrævende og økonomisk dyr proces, er det optimalt at få indsamlet så meget, og så detaljeret, data som

muligt - specielt hvis det er i forbindelse med nogle områder, som der kun er et beskedent kendskab til. Hvis indsamlingen af data sker i form af en spørgeskemaundersøgelse er det imidlertid en balancegang mellem at indsamle detaljeret data og udforme nogle enkle og korte spørgeskemaer, der er hurtige for respondenterne at besvare. Vurderingen af detaljeringsgraden kan eksemplificeres ved, hvis der for eksempel skal indsamles oplysninger om elforbruget i boliger. I spørgeskemaerne kan der stilles spørgsmål om hvor ofte vaskemaskinen i husstanden anvendes, da brugen af denne har indflydelse på elforbruget. Det er ikke udelukkende antallet af tøjvaske der har betydning for elforbruget, men i høj grad også ved hvilke temperaturer der vaskes. Ved udviklingen af spørgeskemaet skal det vurderes, om der kan opnås et mere nøjagtigt klassificeringstræ, hvis der indsamles oplysninger om hvor ofte og ved hvilke temperaturer vaskemaskinen anvendes, frem for udelukkende at indsamle oplysninger om hvor ofte den anvendes. Det er relativt enkelt for beboerne at besvare hvor ofte de vasker tøj, men det bliver mere omfattende, hvis de skal redegøre for, hvor ofte de vasker ved de enkelte temperaturer.

Klassen og attributternes betydning

En af de forudsætninger, der blev fastlagt i afsnit 5.2, som har haft størst indflydelse på det endelige resultat, er klassen for klassificeringstræerne. I kapitel 4 er der vurderet fordele og ulemper for flere mulige klasser til klassificeringstræerne. En af de forudsætninger, der er fastlagt af projektgruppen før udviklingen af klassificeringstræerne, er beslutningen om at udvikle klassificeringstræerne med hensyntagen til husstandsstørrelsen¹. Der er taget hensyn til husstandsstørrelsen, uden den direkte er inddraget som en attribut i klassificeringstræerne. Det kan imidlertid være en mulighed, at inddrage husstandsstørrelsen som en attribut tilsvarende de øvrige forhold. Årsagen til at dette er fravalgt, skyldes de gentagne konklusioner fra forskellige kilder i litteraturstudiet, hvoraf det fremgår, at antallet af beboere har stor indflydelse på energiforbruget. Af denne årsag er det valgt at udvikle et klassificeringstræ for samtlige husstandsstørrelser. Konsekvensen heraf bliver et relativt stort antal klassificeringstræer, hvor beboerne skal vælge det klassificeringstræ, der stemmer overens med husstandsstørrelsen, før energiforbruget kan estimeres. En medvirkende årsag til, at der udvikles et klassificeringstræ for samtlige husstandsstørrelser, er, at enheden for klassen er mere anvendelig. Dette skal forstås således, at der kan udformes klasser, som eksempelvis kan definere størrelsen af et højt forbrug inden for en specifik husstandsstørrelse. Dette er imidlertid ikke muligt, såfremt der anvendes ét samlet klassificeringstræ, hvor samtlige husstandsstørrelser indgår.

Ved udvikling af ét klassificeringstræ for samtlige husstandsstørrelser kan der tages højde for antallet af beboere, ved at inddrage dette som attribut i klassificeringstræet. Derigennem kan det undersøges, om husstandsstørrelsen reelt set har størst betydning for vand- og elforbruget, som det fremgår af litteraturen og antages i nærværende rapport. Betydningen af parameteren kan med enkelthed påvises igennem udviklingen af klassificeringstræet. Såfremt parameteren fremgår af R-knuden, eller en af de øverste I-knuder, kan antagelsen om parameterens vigtighed understøttes yderligere. En af fordelene ved et samlet klassificeringstræ er ligeledes, at der ikke er begrænsninger for hvilken husstandsstørrelse, der kan anvende.

Såfremt der udvikles et samlet klassificeringstræ for flere husstandsstørrelser, kan enheden for klassen stadig være $m^3/\text{år}$ eller $kWh/\text{år}$ for henholdsvis varmtvands- og elforbruget.

¹Den omtalte forudsætning for klasserne er gældende ved udviklingen af et klassificeringstræ for henholdsvis varmtvand og elforbruget.

Klassen kan inddeles i prædefinerede intervaller, men idet flere husstandsstørrelser kan være til stede i samme knude, kan denne imidlertid ikke tildeles benævnelser, der angiver et størrelsesmæssigt forhold til forbruget, eksempelvis et „lavt forbrug“. Ved klassificering af et element i en klasse kan der oplyses en middelværdi og en standardafvigelse. Disse oplysninger kan anvendes til sammenligning af eksempelvis de tidligere anvendte intervaller, angivet i figur 5.9 og 5.11, for derigennem at finde frem til hvordan størrelsen af husstandens forbrug er i forhold til andre husstande af samme størrelse. Der er dog risiko for, hvis den angivne standardafvigelse er høj, at det oplyste resultat fra knuden kan figurere i flere af de angivne klasser. Begge metoder er anvendelige, og det anbefales at forsøge at udvikle klassificeringstræer med begge typer af udfald, idet der kan forekomme flere forskellige sammenhænge mellem forbruget og de udvalgte attributter i de to klassificeringstræer.

Det er igennem litteraturstudiet erfaret, at der er få undersøgelser, som omhandler mængden af varmt brugsvand, der anvendes i boliger. Det overordnede skøn, der går igen i flere kilder, er, at omkring 1/3 af det samlede vandforbrug opvarmes til varmt brugsvand. Denne mangel på data medfører, at der kan overvejes en løsning, hvor der udvikles et klassificeringstræ for det samlede varmekonsum i boligen. Det betyder, at energiforbruget til rumopvarmning og opvarmning af varmt brugsvand vurderes som en helhed. Dette kan gøre metoden nemmere at udvikle, idet der dermed kun er behov for at få oplyst varmekonsumet i boligen. Ved at betragte det samlede varmekonsum er der ikke forskel på hvilke attributter, der kan anvendes i udviklingen af klassificeringstræet. Der kan imidlertid være forskel på hvilke attributter, fra henholdsvis rumopvarmning og varmt brugsvand, der inkluderes i klassificeringstræet. Samtidig er der risiko for, at attributterne fra et af forbrugene overskygger det andet. Det betyder, at der er en del information, der kan blive skjult i klassificeringstræet og dermed ikke direkte vise sammenhængen mellem beboernes adfærd og de to energiforbrug, der er inkluderet. Denne ulempe ved en samlet model gør, at der mistes et overblik over hvilke attributter, der har betydning for de to enkelte forbrug, i forhold til hvis de betragtes separat, som det er tiltænkt i nærværende rapport.

Fiktive data

Der er ved udviklingen af klassificeringstræet genereret data for 100 husstande. Dette data er genereret ud fra de oplysninger, der er beskrevet i kapitel 6. De anvendte attributter er udvalgt på baggrund af litteraturstudiet i kapitel 3 og appendiks C, men det bør ikke anses som en begrænsning for hvilke muligheder, der er for at inddrage andre parametre. Under udviklingen af klassificeringstræet er det erfaret, at der er mulighed for, at flere af de udvalgte attributter kan anvendes anderledes, end det er gjort. Det drejer sig specielt om de to nedenstående attributter:

- Alderen for de voksne beboere (forældrene).
- Tilstedeværelsen af børn, teenager og voksne børn.

Alderen på forældrene er angivet uden nærmere specifikation af hvem af dem, der er ældst, samt hvilket køn der er knyttet til hver af dem. Dette giver udfordringer i forhold til analysen af det endelige klassificeringstræ, idet der på nuværende tidspunkt ikke er defineret en reel forskel i, hvad de to enkeltstående attributter repræsenterer. Det kan samtidig vurderes, om der reelt set er behov for at få oplysninger om alderen på begge forældre. Alderen på

begge forældre kan være relevant, hvis der knyttes et køn til beboerne, ellers kan der nøjes med eksempelvis alderen på den ældste forælder.

Tilstedeværelsen af børn, teenagere og voksne børn kan ligeledes vurderes anderledes, end det er gjort i nærværende rapport. Forholdene for børnene er, modsat forholdene for forældrene, defineret således, at det på forhånd er defineret hvilket barn, der er ældst. Fordelen ved at betragte børnene separat er, at betydningen for børnenes alder kan vurderes på tværs af husstandenes størrelse. Idet der er risiko for, at alle tilstedeværende børn figurerer i klassificeringstræet, kan det derfor overvejes om børnene skal betegnes vha. mulige kombinationer i form af eksempelvis: to børn, et barn og en teenager eller to teenagere. Dette fjerner risikoen for gentagelser af børnenes alder i klassificeringstræet, men det fjerner derimod også muligheden for direkte at kunne vurdere på sammenhængen mellem energiforbruget og børnenes alder.

Konklusion

Energiforbruget i danske boliger udgør 31 % af det samlede energiforbrug i Danmark [27]. Inden for energiforbrug til rumopvarmning, mængden af varmt brugsvand og elforbruget, er der påvist en variation boliger imellem. Denne variation skyldes tekniske forhold omkring boligerne, men beboernes adfærd i hjemmet spiller ligeledes en essentiel rolle. For at kunne bidrage til regeringens mål om at reducere bruttoenergiforbruget i Danmark med 4 % inden 2020 er der behov for en fælles indsats på mange forskellige områder [81]. Her er energiforbruget i boliger er oplagt fokusområde, idet der er et stort potentiale for energibesparelser. Der er derfor behov for at øge fokus på området brugeradfærd i boliger, med henblik på at opnå viden om sammenhængen mellem beboernes adfærd og energiforbrugets størrelse. Denne viden kan herefter anvendes til at opnå energibesparelser gennem energioptimering af boligen og påvirkning af beboerne til ændring af adfærd.

Projektet er delt op i tre overordnede emner: indsamling af viden om beboernes påvirkning på energiforbruget, en vurdering og diskussion af hvilke analysemetoder der kan anvendes til beskrivelse af sammenhængen mellem brugeradfærd og energiforbrug i boliger samt eksemplificering på anvendelsen af et klassificeringstræ. Med udgangspunkt i rapportens overordnede struktur opbygges konklusionen med en tilsvarende rækkefølge, hvor hvert afsnit opsummerer de vigtigste resultater igennem projektet.

De vigtigste parametre i forbindelse med størrelsen af energiforbruget til rumopvarmning er en blanding af tekniske og adfærdsrelaterede forhold: lufttemperaturen i boligen, beboernes vaner for udluftning, boligens størrelse og opførelsesår. Der er på baggrund af litteraturstudiet udviklet en sandsynlighedsfordeling for lufttemperaturen i boliger¹. Fra denne sandsynlighedsfordeling fremgår det, at de foretrukne lufttemperaturer i boligen varierer imellem 19 – 24 °C, hvor omkring 50 % har en lufttemperatur i boligen imellem 21 – 23 °C. Det betyder, som det fremgår af litteraturstudiet, at energiberegningerne, som boligerne på nuværende tidspunkt godkendes på baggrund af, underestimerer energiforbruget til opvarmning af disse.

Vigtigheden af udluftning af boligen er erfaret igennem litteraturstudiet, hvor det fremgår, at beboere der lufte meget ud, uden at slukke for varmen, har et markant højere forbrug end beboere, der slukker for varmen imens. Ved analysering af data fra projektet *Fremtidens Parcelhuse* [52] er det ydermere erfaret, at der kan være belæg for at undersøge forholdene omkring beboernes vaner, med hensyn til udluftning af boligen, yderligere. I dette projekt skilte en af husstandene sig markant ud ved at have den laveste lufttemperatur, men det

¹Se appendiks J.1 for anvendte kilder

højeste varmekonsum. Dette er bemærkelsesværdigt idet det igennem litteraturstudiet er erfaret, at høje lufttemperaturer i boligen ofte medfører et højt forbrug. Der er behov for at undersøge forholdene nærmere, men ud fra de tilgængelige oplysninger indikeres det, at beboernes vaner omkring udluftning har stor indflydelse på varmekonsumet.

Forbruget af varmt brugsvand er vanskeligt at analysere, idet omfanget af undersøgelser på området er beskedent. Derfor indeholder litteraturstudiet både undersøgelser for det kolde og varme brugsvand. De vigtigste parametre for størrelsen af det samlede vandforbrug er: antallet af beboere, antallet og længden af bade, antallet af tøjvaske, tilstedeværelsen af børn og teenagere samt boligarealet. I de få projekter, der indeholder information om forbruget af varmt brugsvand, fremgår det, at der er store variationer i dette forbrug. Derfor er der behov for at få større indsigt i dette forbrug, specielt fordi energiforbruget til opvarmning af varmt brugsvand i den kommende årrække vil få større betydning i den samlede energimængde, der forbruges i boliger.

I de seneste år har der i medierne, gennem forskellige kampagner, været stort fokus på, at få forbrugerne til at spare på elforbruget. Valget af elforbruget som fokusområde er forståeligt, eftersom der i flere undersøgelser fremgår en stor forskel på hvor meget el, der anvendes i husstande af samme størrelse. Elforbruget størrelse har i to undersøgelser varieret med en faktor 7 og 10 mellem det laveste og højeste forbrug. De parametre, der igennem litteraturstudiet har vist sig at have stor betydning for elforbrugets størrelse er: antallet af beboere, boligens størrelse, antallet af elapparater og deres brugstid, husstandens indtægt, tilstedeværelsen af børn og teenagere, standbyforbruget samt beboernes viden om størrelsen af deres forbrug. Der er stor forskel på hvornår beboerne udskifter deres elapparater, og hvilket incitament der er for udskiftningen. Ved udskiftning af elapparater er der stor forskel på, om beboerne vægter design eller energimærkningen højest. Det er i nogle tilfælde erfaret, at anskaffelsen af et apparat med et godt energimærke indbyder til et nyt forbrugsmønster, som derigennem introducerer et merforbrug.

Beboernes viden, om tekniske installationer og størrelsen af deres forbrug, har vist sig at have betydning for forbrugets størrelse, specielt for energiforbruget til rumopvarmning og elforbruget. Information om tekniske installationer er derfor essentiel i nye bygninger, hvor teknikken er mere fremtrædende, end den har været tidligere, for at opnå optimal drift og derigennem det lavest mulige energiforbrug. Derfor er det nødvendigt både at informere beboerne om styringen af anlæggene, men også at anlæggene udformes intuitive og brugervenlige. Grundet den store variation i vanerne blandt beboerne er det vanskeligt at klarlægge og inddele beboerne i særskilte typer, idet hver person har sin egen individuelle måde at gøre tingene på. Grundet denne kompleksitet er det vanskeligt, igennem en model, at beskrive hvilken påvirkning brugeradfærden har på energiforbruget.

Med udgangspunkt i ovenstående problemstilling er det andet fokusområde for nærværende projekt, at finde frem til en metode der kan beskrive sammenhængen mellem energiforbruget og demografiske, tekniske og adfærdsrelaterede forhold. Der er opnået kendskab til fem forskellige metoder fordelt på to områder inden for data mining; beskrivende og forudsigende metoder. Metoderne faktor-, komponent-, og klyngeanalyse tilhører gruppen af beskrivende metoder.

Faktor- og komponentanalyser anvendes til at reducere kompleksiteten i et datasæt. Faktoralysen gør det ved at gruppere en række forskellige adfærd hos beboerne, for derigennem

at skabe overblik over nogle sammenhænge mellem besvarelser. Derimod kan komponentanalysen anvendes til at reducere data ved at samle information fra flere variable under et færre antal komponenter. Klyngeanalyse er en betegnelse for flere analysemetoder, der alle har det fælles mål; at identificere hvilke elementer i et datasæt, der har nogle fælles karakteristika for efterfølgende at gruppere elementerne i klynger. Blandt de beskrivende metoder er anvendeligheden af faktor-, komponent- og klyngeanalyse sammenholdt med projektets hovedmål, hvor disse er nedprioriteret i forhold til de forudsigende metoder. De beskrivende metoder er nedprioriteret grundet en vurdering af, at metodernes fordele kommer til udtryk, når der skal udføres beskrivende analyser af data. Det vil sige, når der søges efter kendte eller ukendte sammenhænge i indsamlet data, eller såfremt der skal udføres uddybende undersøgelser af beboere og deres adfærd. Disse metoder vurderes anvendelige såfremt det aktuelle data skal simplificeres, eller hvis der skal udføres beskrivelser eller undersøgelser af karakteristika for bestemte typer beboere.

Blandt de forudsigende metoder er der udført en undersøgelse af henholdsvis regressions- og klassificeringstræer. Disse metoder er vurderet som værende de bedst anvendelige i forbindelse med at nå projektets hovedmål. Det skyldes, at metodernes fordele findes i den forudsigende proces og deres simple anvendelse til estimering af eksempelvis energiforbrug. Anvendelsen og udviklingen af de to metoder har flere ligheder, men de grundlæggende forskelligheder fremgår af hvorledes klassen defineres. Risikoen for at størrelsen af et regressionstræ for energiforbruget er for kompleks, er i nærværende projekt vurderet til at være for stor. Tilsvarende vurderes en klasseinddeling af energiforbruget som værende fordelagtig i forbindelse med formidlingen af resultaterne til en bred målgruppe. En klasseinddeling, med en tilknyttet betegnelse som eksempelvis „højt forbrug“, vurderes at være mere enkel for eksempelvis beboerne at forholde sig til frem for en given numerisk værdi. Disse vurderinger er baggrunden for, at der i nærværende projekt fokuseres på udviklingen og anvendelsen af et klassificeringstræ til estimering af energiforbrug.

Det konkluderes i nærværende rapport, at metoden omhandlende klassificeringstræer har et potentiale i forbindelse med estimering af energiforbrug. Det skyldes, at der i udviklingen af et klassificeringstræ kun er få begrænsninger for hvilke forhold, der kan inddrages. Det er, på baggrund af viden fra litteraturstudiet og analysen af boligerne fra *Fremtidens Parcelhuse* [52], vurderet hvilke attributter, der kan være anvendelige i udviklingen af et klassificeringstræ. De beskrevne attributter skal ikke anses, som en begrænsning for hvilke forhold der er mulige at medtage, men som inspiration til hvilke forhold der kan inddrages. Det er eksempelvis muligt at inddrage demografiske, tekniske såvel som adfærdsrelaterede forhold, samtidig med at der kan anvendes både numeriske og kategoriske attributter i klassificeringstræet. Der vurderes dermed, på baggrund af erfaringer fra litteraturstudiet, at være et stort potentiale i anvendelsen af denne metode i forbindelse med estimering af energiforbrug. Det skyldes, at det gennem litteraturstudiet er erfaret, at de bedste løsninger fremkommer, når viden fra flere fagområder kombineres.

Ved udvikling af et klassificeringstræ kan der anvendes forskellige algoritmer, hvor der i nærværende projekt er redegjort for nogle af mulighederne. Til eksemplificering af klassificeringstræets udvikling er C4.5 algoritmen anvendt. Det skyldes, at den ikke er begrænset til udelukkende at foretage binære split, samtidig med at den kan håndtere numeriske og kategoriske attributter. Gennem eksemplificeringen er der anvendt fiktivt data med udelukkende kategoriske attributter, hvor der til trods for dette er redegjort for, hvorledes numeriske attributter håndteres af den valgte algoritme. Ydermere er der redegjort for, hvorledes manglende

data for enkelte attributter håndteres af algoritmen. Dette er foretaget på baggrund af, at der i praksis kan opstå tilfælde, hvor der er enkelte mangler i indsamlet data - hvad enten det er fra en spørgeskemaundersøgelse eller fra målinger.

Da der gennem eksemplificeringen er anvendt fiktivt data, kan der ikke konkluderes direkte på resultaterne, og sammenhængen der fremgår af det udviklede klassificeringstræ. Derimod er der mulighed for at konkludere på metodens egnethed, i forhold til at kunne beskrive sammenhængen mellem energiforbrug og beboernes adfærd. En af årsagerne til at klassificeringstræet anvendes som metode i denne rapport, skyldes metodens egenskab til at inddrage mange forskellige typer af data, samt at resultatet for metoden er simpel at anvende. Disse to forhold betyder, at metoden kan inddrage viden om både tekniske, demografiske og adfærdsrelaterede forhold, der igennem litteraturstudiet viste sig, at være optimale at anvende i kombination med hinanden, for at kunne beskrive den største variation i energiforbrugene. Den simple og lette anvendelse gør blot målgruppen for det udviklede klassificeringstræ bredere, hvilket er med til at muliggøre en større udbredelse af klassificeringstræets potentiale for anvendelse.

Klassificeringstræerne der, efter indsamling af data, kan udvikles til at beskrive henholdsvis energiforbruget til rumopvarmning, mængden af varmtvandsforbruget eller den samlede vandmængde samt elforbruget, er tiltænkt, at kunne anvendes i boliger til estimering af de pågældende energiforbrug under hensyntagen til beboernes adfærd. Anvendelsen af klassificeringstræerne er, afhængig af størrelsen af disse, relativt simpel. Som tidligere beskrevet er det i nærværende rapport valgt, med hensyn til varmtvandsforbruget og elforbruget, at der skal udvikles klassificeringstræer for de mest sandsynlige husstandsstørrelser. Praktisk anvendelse af et klassificeringstræ starter, i forhold til det udviklede eksempel, med at udvælge klassificeringstræet, der er gældende for den pågældende husstandsstørrelse. Herefter gennemgås og besvares spørgsmålene knude for knude i klassificeringstræet, indtil husstanden tildeles en klasseinddeling, der estimerer deres forbrug med eksempelvis en middelværdi og standard afvigelse. Denne viden kan anvendes på to måder. Enten accepteres resultatet som det er og kan dermed anvendes som en vejledning for hvor stort et forbrug husstanden kan forventes at have. Den anden løsning er, at beboerne kan gå baglæns i træet, på den vej i klassificeringstræet de fulgte, for dermed at indsamle informationer om hvilke parametre familien kan fokusere på, i håb om at opnå en energibesparelse. Derved kan de selv vurdere hvilke parametre, der falder dem mest naturligt at optimere, såfremt de ønsker at reducere energiforbruget i boligen. Dette er under forudsætning af, at der i klassificeringstræet anvendes nogle attributter, som almene boligejere har grundlag for at kunne besvare.

Anvendelsen af klassificeringstræet har potentiale inden for tre overordnede typer af brugere: boligejere, som gennemgås i afsnittet ovenfor, energirådgivere og ejendomsmæglere.

Klassificeringstræet, udviklet med henblik anvendelse af beboerne, kan ligeledes anvendes af energirådgiverne. Dog er der, grundet deres tekniske baggrund, mulighed for at udvikle klassificeringstræer, hvor der indgår detaljerede informationer om tekniske forhold. I disse tilfælde kommer et af klassificeringstræernes andre fordele til udtryk. Grundet den simple opbygning af klassificeringstræet kan energirådgiverne anvende det som et værktøj til at formidle hvilke betydningsfulde forhold, boligejerne kan optimere på. Tilsvarende er der et potentiale i, at klassificeringstræer kan anvendes af ejendomsmæglere til vejledning af hvilket energiforbrug, primært med hensyn til varmemeforbrug, købere kan forvente. Samtidig kan der være mulighed for inddrage de øvrige forbrug, eller gøre opmærksom på hvor information om dette kan findes. Ejendomsmægleren kan bidrage med viden om tekniske oplysninger,

eksempelvis fra energimærkerapporterne samt eventuelle udførte renoveringer. Derved kan køber på forhånd vurdere hvilket energiforbrug, der kan forventes i den kommende bolig, hvormed dele af de faste udgifter bedre kan budgetteres. Ligeledes kan ejendomsmæglerne, i forbindelse med salg af bolig, anvende klassificeringstræet sammen med energimærkningen til at informere køber om konsekvenser af eventuelle konstruktionsmæssige ændringer, som køber overvejer.

Som det fremgår af ovenstående vurderinger er der et stort potentiale for at klassificeringstræer kan anvendes med henblik på at beskrive sammenhængen mellem energiforbruget i boliger under hensyntagen til beboernes adfærd. Fordelen er som beskrevet, at klassificeringstræer er simple at anvende både for folk med og uden kendskab til brugeradfærd og energiforbrug i boliger. Den simple tilgang gør det muligt at indkredse en stor målgruppe, hvormed der er et stort potentiale for, at udbredelsen og anvendelsen af klassificeringstræet kan lykkes. Ved at udvikle og introducere klassificeringstræer, er der dannet grundlag for at øge opmærksomheden på mulige energibesparelser i boliger, og dermed kan klassificeringstræerne være medvirkende til, at energiforbruget i boliger reduceres. Udbredelsen af budskabet om klassificeringstræets anvendelighed for boligejerne kræver også stor opmærksomhed, idet der er langt imellem at introducere problemstillingen til beboerne, og herfra forvente at de straks implementerer nye vaner i hjemmet. Dette aspekt er blandt andet observeret igennem litteraturstudiet, idet der er stor forskel på hvor mange kræfter beboerne vil bruge på at implementere vanerne, samt hvor meget det går ud over deres velbefindende. Derudover er der ligeledes stor forskel på, hvilke virkemidler der påvirker forskellige beboere bedst i forhold til at opnå energibesparelser. Det kan derfor ikke forventes, at udviklingen og præsentationen af klassificeringstræerne er tilstrækkeligt arbejde. Derfor kan det anbefales, efter anvendelse af klassificeringstræer, at beboerne tilbydes en yderligere mulighed for vejledning, enten igennem en energirådgiver eller en hjemmeside med uddybende information.

Perspektivering

Af konklusionen fremgår det, at metoden omhandlende klassificeringstræer er yderst anvendelig til estimering af energiforbruget i boliger, når der skal inddrages informationer om både tekniske, demografiske og adfærdsrelaterede forhold. Udviklingen af klassificeringstræet i rapporten er udført på baggrund af fiktivt data, men for at lignende klassificeringstræer kan blive anvendelige i praksis, skal disse udvikles på baggrund af virkelige data. Dette vil kræve en del indledende arbejde, da det gennem litteraturstudiet er erfaret, at der ikke foreligger tilstrækkeligt data inden for området.

Behovet for indsamling af data er gældende for alle de tre forbrugsområder, der er behandlet gennem rapporten. Det forbrugsområde, hvor der foreligger mindst tilgængeligt litteratur, er for varmtvandsforbruget i boliger. Det er gennem litteraturstudiet erfaret, at der kun i meget begrænset omfang foreligger tilgængeligt litteratur om forbruget af varmt brugsvand i danske boliger, og på hvilke måder beboernes adfærd påvirker dette forbrug. Som det er nærmere diskuteret i rapporten, kan varmtvandsforbruget opgøres enten på baggrund af en mængdeangivelse eller som et energiforbrug til opvarmning af dette. Idet opvarmningen af varmt brugsvand ofte sker lokalt i boligerne, vil mængden af varmt brugsvand indgå som en del af det samlede vandforbrug i disse. Tilsvarende indgår energiforbruget til opvarmning af varmt brugsvand ofte som en del af det samlede varme- eller elforbrug i boligen. Det betyder, at det varme brugsvand i begge tilfælde optræder som et skjult forbrug. Såfremt der skal indsamles oplysninger om forbruget af varmt brugsvand, både med hensyn til mængder og energiforbrug, vil det kræve en del indledende arbejde, på grund af den samlede opgørelse af forbruget, som beskrevet i det foregående. Ved at undersøge forbruget af varmt vand nærmere, er der mulighed for at synliggøre forbruget, og dermed lade dette være incitament for, at kunne opnå besparelser i boligerne.

Det er i nærværende rapport beskrevet, at flere informationer omhandlende tekniske, demografiske og adfærdsrelaterede forhold kan indsamles gennem spørgeskemaundersøgelser. For at opnå så anvendelige resultater som muligt, bør der foretages dybdegående analyser af, hvorledes spørgeskemaerne kan udformes mest optimalt. Det er igennem litteraturstudiet erfaret, at der er udviklet en adfærdsmodel, der kan anvendes til at sammenholde menneskets holdninger og adfærd. I fremtidige studier kan det undersøges, hvorvidt den adfærdsmodel, eller lignende modeller, kan være fordelagtige at anvende i forbindelse med udvikling af spørgeskemaundersøgelser til uddybning af brugeradfærd.

I nærværende rapport er klassificeringstræerne beskrevet og udviklet med henblik på anvendelse for parcel-/rækkehuse. Da boligmassen består af flere typer af boliger, skal det overvejes, før der indsamles oplysninger til udvikling af klassificeringstræer, hvilke boligtyper disse skal

være gældende for. Det skyldes, at metoden lige såvel kan anvendes til estimering af energiforbrug lejligheder, såfremt datagrundlaget tillader det. Hvis der skal foretages estimeringer af energiforbruget i flere boligtyper end parcel-/rækkehuse, skal det fastlægges, om boligtyperne skal betragtes separat, eller om flere kan betragtes ved hjælp af samme klassificeringstræ. Derved er der ikke nogle begrænsninger i metoden omhandlende klassificeringstræer, der forhindrer, at der i fremtidige projekter anvendes klassificeringstræer på andre boligtyper end parcelhuse.

Der er i nærværende projekt hovedfokus på udvikling og anvendelse af klassificeringstræer i forbindelse med estimering af energiforbrug i boliger. Det er omtalt tidligere i denne rapport, at der er mulighed for, at anvende andre metoder til at foretage lignende estimeringer, men disse metoder er udelukkende vurderet på baggrund af den grundlæggende teori. For at vurdere metodernes anvendelighed i forbindelse med estimering af energiforbrug i boliger, under hensyntagen til beboernes adfærd, bør der i fremtidige studier foretages lignende undersøgelser på andre metoder. Først efter disse uddybende undersøgelser er foretaget, kan der foretages en direkte sammenligning af metodernes anvendelig inden for dette anvendelsesområde.

Såfremt der i fremtidige projekter skal anvendes et klassificeringstræ til estimering af energiforbrug, kan det anbefales, at der foretages en undersøgelse af hvilke tilgængelige softwareprogrammer, der kan anvendes i denne forbindelse. I nærværende projekt er Weka anvendt til udvikling af klassificeringstræerne. Programmet har imidlertid en række begrænsninger og områder hvor informationsmængden fra dette kan forbedres. Eksempler herpå er, at det i udviklingsprocessen ikke er muligt at sammenholde attributternes rangering. I nærværende rapport er dette udført gennem håndberegninger. Derudover informerer Weka ligeledes kun om antallet af fejlklassificeringer i de enkelte L-knuder. I forbindelse med vurdering af de udviklede klassificeringstræer kan det imidlertid være fordelagtigt, at have kendskab til hvilke klasseinddelinger de fejlklassificerede elementer har. I det nærværende projekt er dette undersøgt ved en manuel gennemgang af data, men dette vil imidlertid være et omfattende arbejde, hvis der i fremtidige projekter anvendes større datamængder.

I nærværende rapport er der foretaget vurderinger og diskussion af CF -faktorens betydning i den anvendte algoritme for klassificeringstræer. I $C4.5$ algoritmen, der er anvendt i nærværende rapport, er standardværdien for CF -faktoren fastlagt til 0,25. Denne værdi har betydning for størrelsen af de udviklede klassificeringstræer, og er fastlagt af udvikleren af algoritmen, ud fra en vurdering af at denne værdi giver fornuftige resultater i de fleste tilfælde. Det bør imidlertid undersøges, hvorvidt denne værdi ligeledes kan anses som en standardværdi i forbindelse med klassificeringstræer til estimering af energiforbrug. Denne værdi kan med fordel evalueres i et software program, hvor det er muligt, at vurdere middelværdi og standardafvigelse af energiforbruget for elementerne i de enkelte L-knuder, hvilket ikke er muligt i Weka. Ved evaluering af standardafvigelsen for energiforbruget kan det vurderes hvorvidt en ændring af CF -faktoren kan reducere standardafvigelsen for elementerne i de enkelte L-knuder.

Ordliste

Algoritme: En regneforskrift eller metode til løsning af matematiske eller logiske problemer i et endeligt antal beregningsskridt.

Attributter: Er en fællesbetegnelse for de variable der er i en indsamlet datamængde og anvendes til at inddele data efter i et beslutningstræ.

Be10: Be10 er et beregningsprogram, der anvendes til at dokumentere at energikravene i Bygningsreglementet og anden lovgivning er fulgt.

Bivariat korrelation: Er et kvantitativt mål for graden af lineær sammenhæng mellem to variable, der er uafhængig af hvilken skala, de to variable er målt på.

Bottom-up: Angiver princippet i fremgangsmåden for en agglomerativ klyngeanalyse, hvor elementerne indledningsvis inddeles i et tilsvarende antal klynger, hvorefter disse klynger grupperes skridt for skridt, således der til sidst haves én klynge indeholdende samtlige elementer.

Bruttoenergiforbrug: Bruttoenergiforbruget er det faktiske forbrug af brændsler mv. korrigeret for import/eksport af elektricitet og fjernvarme.

BSim: BSim anvendes til simulering og beregning af bl.a. termisk indeklima, energiforbrug, dagslysforhold, fugtsimulering, naturlig ventilation.

Common Factor Model: Er en model, der dels fastlægger at flere variable er påvirket af de samme fælles faktorer men i forskellig grad, og dels at de enkelte variable er påvirket individuelt af nogle unikke faktorer.

Demografiske forhold: Forhold der beskriver parametre som eksempelvis alder, indkomst, uddannelsesniveau, køn, civilstatus og lign.

edb: En fællesbetegnelse for computere, printere, scannere og lignende.

Euklidisk afstand: Angiver et mål for ligheden mellem to elementer, der er beregnet på baggrund af samtlige attributter for disse. Antager en værdi på nul såfremt de to elementer har identiske attributter, og en stigende værdi som funktion af større forskellighed mellem de to elementers attributter.

IHC-anlæg: IHC står for Intelligent House Control, og kan anvendes til at programmere lamperne, fyret eller andre elinstallationer i dit hus, så de bliver styret automatisk.

Kategorisk attribut: Ingen talværdier, men værdier i form af begreberne høj/lav, bred/smal, on/off osv. eller i form af nogle prædefinerede kategorier

Klasse: Er det navn der anvendes om den attribut der, i et beslutningstræ, beskriver det endelige mål for inddelingen af data.

Knuder: Er en fællesbetegnelse for de tre dele, der udgør hovedbestanddelen i et beslutningstræ.

Komforttemperatur: Temperaturen hvor personer føler sig i termisk komfort, dvs. personen føler sig termisk neutral for kroppen som helhed uden at ønske hverken koldere eller varmere omgivelser.

Kvalitativ: Kvalitativ analyse er en analyseform, der baserer sig på at skaffe materiale, der som udgangspunkt skal analyseres uden diverse statistiske modeller. Eksempelvis et interview.

Kvantitativ: Kvantitativ analyse er en analyseform, der baserer sig på ting der kan måles. Et eksempel herpå er en spørgeskemaundersøgelse.

Lufttemperatur: Temperaturen i opholdszonen omkring personen, men uden for grænselaget af opvarmet luft tæt ved personen.

Met: Energiomsætningen i et menneske der afhænger af menneskets aktivitet. Den kan variere imellem 0,8 *met* for et sovende menneske og op til 15 *met* for en elitesportsudøver.

Middelstrålingstemperatur: Den ensartede temperatur af de omgivende flader, der ville medføre samme varmeafgivelse ved stråling fra en person som de faktiske omgivelser.

Nettoenergiforbrug: Dette begreb anvendes i forbindelse med energiforbruget til rumopvarmning og opvarmning af varmt brugsvand. Det udtrykker den mængde energi, der er nyttiggjort i husstanden. Det betyder, at nettoenergiforbruget angiver en energimængde eksklusiv eventuelle lokale energitab hos de enkelte forbrugere.

Nominal attribut: Dette er en *kategorisk attribut* der er kendetegnet ved at man ikke kan rangere det man måler. Eksempelvis for attributterne: mand/kvinde eller personers øjenfarve osv.

Numerisk attribut: Denne attribut type er målbar dvs. objekternes interne afstand kan bestemmes og vurderes. Et eksempel herpå er: en persons højde.

Operativ temperatur: En teoretisk temperatur der kombinerer middelstrålingstemperaturen og lufttemperaturen. Det den beskriver er den ensartede temperatur af luft og omgivende flader, der vil medføre samme varmeafgivelse som de faktiske omgivelser.

Ordinal attribut: Dette er en *kategorisk attribut* der er kendetegnet ved at vi kan rangere de forskellige udfald som vi måler med vores variabel. Et eksempel herpå er f.eks. holdninger til emner - 'meget enig', 'enig', 'neutral', 'uenig' eller 'meget uenig' eller ved beskrivelse af vejret som 'koldt', 'mildt' eller 'varmt'.

PMV: står for *Predicted mean vote* og angiver den forventede middelvotering for et stort antal personers angivelse af om disse følte sig tilpas i termisk henseende.

PPD: står for *Predicted percentage of dissatisfied* og angiver det forventede antal utilfredse i større gruppe af personer i forhold til termisk henseende.

Rumopvarmning: En fællesbetegnelse for det varmekonsum der anvendes til rumopvarmning og opvarmning af varmt brugsvand. $[^{\circ}\text{C}]$

Setpunktstemperaturen: setpunktstemperaturen er indstillingen på for eksempel radiatortermostaten til den lufttemperatur i boligen som varmeanlægget ønskes at opretholde.

Splittests: En splittest anvendes i et klassificeringstræ for at inddele data i forhold til den endelige klasse. Splittests udføres ved hjælp af de udvalgte attributter og kan være i form af enten binære eller ikke binære spørgsmål.

Top-down: Angiver princippet i fremgangsmåden for en divisiv klyngeanalyse, hvor samtlige elementer indledningsvis antages at tilhøre én klynge, hvorefter klyngen opdeles skridt for skridt, således der til sidst haves et antal klynger, der svarer til antallet af elementer.

Varmekonsum: Det samlede energikonsum der anvendes til rumopvarmning og opvarmning af det varme brugsvand.

Litteratur

- [1] Søren Aggerholm og Karl Grau. *Bygningers energibehov - Beregningsvejledning*. Teknisk rapport, 2008.
- [2] Ulrik Andersen. *Lavenergihuse plages af overophedning og kolde rum*.
<http://ing.dk/artikel/111357-lavenergihuse-plages-af-overophedning-og-kolde-rum>, 2010. Ingeniøren.
- [3] Machine Learning Group at University of Waikato. *Weka 3: Data Mining Software in Java*.
<http://www.cs.waikato.ac.nz/ml/weka/>, 2011.
- [4] Stewart Barr, Andrew W. Gilg og Nicholas Ford. *The household energy gap: examining the divide between habitual- and purchase-related conservation behaviours*. Energy Policy, 33:1425–1444, 2005.
- [5] Gustavo E. A. P. A. Batista og Maria Carolina Monard. *A Study of K-Nearest Neighbour as an Imputation Method*. Hybrid Intelligent Systems, side 251–260, 2002.
- [6] Gustavo E. A. P. A. Batista og Maria Carolina Monard. *An Analysis of Four Missing Data Treatment Methods for Supervised Learning*. Applied Artificial Intelligence, 17:519–533, 2003.
- [7] Jason R. Beck, Maria E. Garcia, Mingyu Zhong, Michael Georgiopoulos og Georgios Anagnostopoulos. *A Backward Adjusting Strategy for the C4.5 Decision Tree Classifier*.
http://www.amalthea-reu.org/pubs/amalthea_tr_2007_01.pdf, 2007. Amalthea.reu - Advances of Machine Learning in Theory & Applications.
- [8] Benny Bøhm, Flemming Schrøder og Niels C. Bergsøe. *Varmt Brugsvand - Måling af forbrug og varmetab fra cirkulationsledninger*. Teknisk rapport 978-87-563-1372-8, 2009.
- [9] Niels J. Blunch. *Faktoranalyse - en metode til datakomprimering*. Ledelse og Erhvervsøkonomi/Handelsvidenskabeligt Tidsskrift/Erhvervsøkonomisk Tidsskrift, 36, 1972.
- [10] Bolius Boligejernes videntcenter. *Køb af opvaskemaskine*.
<http://www.bolius.dk/alt-om/koekken/artikel/koeb-af-opvaskemaskine/>, 2007.
- [11] Remco R. Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann og Alex Seewald David Scuse. *WEKA Manual for Version 3-6-4*. Teknisk rapport, The University og Waikato, 2010.
- [12] Leo Breiman, Jerome H. Friedman, Richard A. Olshen og Charles J. Stone. *Classification and regression trees*. isbn 0-412-04841-8. Chapman & Hall, 1984.
- [13] Henrik Brohus, Per Heiselberg, Allan Simonsen og Kim C. Sørensen. *Influence of Occupants' Behaviour on the Energy Consumption of Domestic Buildings*. Proceedings of Clima 2010 Rehva World Congress, Antalya, Turkey, 9.-12. maj 2010.
- [14] Statens Byggeforskningsinstitut. *Bygningers energibehov: Be10 beregningsprogrammet*.
<http://www.sbi.dk/miljo-og-energi/energiberegning/anvisning-213-bygningers-energibehov>, 2011.
- [15] Statens Byggeforskningsinstitut. *Simulering med BSim*.
<http://www.sbi.dk/indeklima/simulering>, 2011.
- [16] Byggestyrelsen. *Bygningsreglement 1982*. isbn 87-503-4205-3. Byggestyrelsen, 1982.
- [17] Anders Bech Christiansen, Tytte Hetmar, Jan Ivanouw og Henriette Kirkeby. *Metodologiske indblik og udsyn*. isbn 978-87-992738-0-5. Forlag for Psykologisk Forskningsmetode, 2008.

-
- [18] Dansk Standard. *DS 474 - Norm for specifikation af termisk indeklima*. isbn ICS: 91.040.01. Byggedata, 1993.
- [19] Dansk Standard. *DS 418 - Beregning af bygningers varmetab*. isbn ICS: 91.120.10. Byggedata, 2002.
- [20] Dansk Standard. *DS 428 - Norm for brandtekniske foranstaltninger ved ventilationsanlæg*. isbn ICS: 13.220.20; 91.140.30. Byggedata, 2009.
- [21] Ernst Jan de Place Hansen, Birgitte Dela Stang, Søren Ginnerup, Inge Mette Kirkeby, Thomas Cornelius Buch-Hansen, Niels-Jørgen Aagaard, Alectia A/S, Lars Schjøtt Sørensen, Niels Christian Bergsøe, Dan Hoffmeyer, Søren Aggerholm og Erik Brandt. *Anvisning om Bygningsreglement 2010 - SBI-anvisning 230*.
http://anvisninger.dk/Publikationer/Sider/Anvisning_om_Bygningsreglement_2010.aspx, 2010.
- [22] Jamie DeCoster. *Overview of Factor Analysis*.
<http://www.stat-help.com/factor.pdf>, 1998.
- [23] DI Energibranchen. *Veje til grøn vækst - 10 politiske fokuspunkter*.
http://energi.di.dk/SiteCollectionDocuments/Foreningssites/energi.di.dk/Publikationer/Veje%20til%20gr%C3%B8n%20v%C3%A6kst_web.pdf, 2011.
- [24] Tom Dietterich, Michael Kearns og Yishay Mansour. Applying the weak learning framework to understand and improve c4.5. In *In Proceedings of the Thirteenth International Conference on Machine Learning*, side 96–104. Morgan Kaufmann, 1996.
- [25] Dansk Energi. *Dansk Elforsyning Statistik 2008*. Teknisk rapport, Dansk Energi, 2009.
- [26] Energimidt.dk. *Undgå unødvendigt standbyforbrug*.
<http://www.energimidt.dk/Privat/Energi-med-omtanke/Energiraadgivning/Gode-raad-om-el/Sider/Undgaa-unoedvendigt-standbyforbrug.aspx>, 2011.
- [27] Energistyrelsen. *Energistatistik 2009*. isbn 978-87-7844-872-9 www. Energistyrelsen, 2010.
- [28] DONG Energy. *Test dit hus*.
<http://www.dongenergy.dk/PRIVAT/CLEANTECH/Pages/testdithus.aspx>, 2011.
- [29] Dong Energy. *Typisk elforbrug - hus*.
<http://www.dongenergy.dk/privat/energiforum/tjekditforbrug/typiskelforbrug/Pages/hus.aspx>, 2011.
- [30] Equa. *Enorm 2004*.
<http://www.equa.se/enorm/index.html>, 2010.
- [31] Europa-parlamentet. *Så bliver der slukket for glødepærerne*.
<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+IM-PRESS+20090904ST060248+0+DOC+XML+V0//DA>, 2009.
- [32] Adam Faiers, Matt Cook og Charles Neame. *Towards a contemporary approach for understanding consumer behaviour in the context of domestic energy use*. Energy Policy, 35:4381–4390, 2007.
- [33] P.O. Fanger. *Thermal Comfort - Analysis and Application in Environmental Engineering*. isbn 0-07-019915-9. McGraw-Hill Book Company, 1970.
- [34] Dr. Wolfgang Feist, Søren Peper og Manfred Görg. *CEPHEUS - Projectinformation No. 36*. Teknisk rapport, Passiv Haus Institut, 2001.
- [35] Forsyningsvirksomhederne.dk. *Standbyforbrug*.
<http://www.forsyning.dk/FV/Energicenter/Energibesparelser/Godeelvaner/Standbyforbrug.htm>, 2011.

-
- [36] Jillian J. Francis, Martin P. Eccles, Marie Johnston, Anne Walker, Jeremy Grimshaw, Robbie Foy, Eileen F. S. Kaner, Liz Smith og Debbie Bonetti. *Constructing Questionnaires based on the Theory of Planned Behaviour: A Manual for Health Services Researchers*. Teknisk rapport 0-9540161-5-7, Quality of Life and Management of Living Resources, Centre for Health Services Research, University of Newcastle, 2004.
- [37] Åge Andersen. *Speciale om: Latente variable, faktor analyse og LISREL modeller anvendt på kommuners tilkendelse af førtidspensioner*. Specialeafhandling, Københavns Universitet - Økonomisk institut, 2004.
- [38] Zachary M. Gill, Michael J. Tierney, Ian M. Pegg og Neil Allan. *Low-energy dwellings: the contribution of behaviours to actual performance*. Building Research & Information, 38:491–508, 2010.
- [39] Go' energi - Center for energibesparelser under Klima- og Energiministeriet. *Go' energi*. <http://www.goenergi.dk/>, 2011.
- [40] Kirsten Gram-Hanssen. *Boligers energiforbrug - sociale og tekniske forklaringer på forskelle*. Teknisk rapport 87-563-1170-2, 2003.
- [41] Kirsten Gram-Hanssen. *Residential heat comfort practices: understanding users*. Building Research and Information, 38:175–186, 2010.
- [42] Kirsten Gram-Hanssen og Kirstine Nærvig Petersen. *Husholdningers energi- og vandforbrug*. Teknisk rapport 87-563-1231-8, 2005.
- [43] Jerzy W. Grzymala-Busse og Ming Hu. A comparison of several approaches to missing attribute values in data mining. In Wojciech Ziarko og Yiyu Yao, editors, *Rough Sets and Current Trends in Computing*, volume 2005 of *Lecture Notes in Computer Science*, side 378–385. Springer Berlin / Heidelberg, 2001.
- [44] Olivia Guerra-Santin og Laure Itard. *Occupants' behaviour: determinants and effects on residential heating consumption*. (38:3, 318-338), 2010.
- [45] Grontmij Carl Bro A/S Hans Jørgen Gjerløv. *Energimærkning for Strubjerg 22, 9400 Nørresundby*. <http://www.ejerforeningen7fe.dk/Energimaerke.pdf>, 2007.
- [46] HusetsWeb. *Energiberegner*. <http://www.husetsweb.dk/>, 2011.
- [47] H. Jeeninga, M. Uytterlinde og J. Uitzinger. *Energy use of energy efficient residences*. Teknisk rapport, 2001.
- [48] Jesper Ole Jensen. *Livsstil, boform og ressourceforbrug*. Ph.d.-afhandling, By og Byg, Statens Byggeforskningsinstitut, 2002.
- [49] Earl Harris Jr. *Information Gain Versus Gain Ratio: A Study of Split Method Biases*. http://www.mitre.org/work/tech_papers/tech_papers_01/harris_biases/harris_biases.pdf, 2005. Computer Science Department William & Mary.
- [50] G. V. Kass. *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. Applied Statistics, 29:119–127, 1980.
- [51] Vibeke Hansen Kjaerbye, Anders E. Larsen og Mikael Togeby. *The effect of building regulations on energy consumption in single-family houses in Denmark*. http://www.ea-energianalyse.dk/reports/the_effect_of_building_regulations_on_energy_consumption_in_single_family_houses_in_Denmark.pdf, 2010.
- [52] Henrik N. Knudsen, Lars Kristensen, Ole Michael Jensen, Heidi Arvedsen, Niels Christian Bergsøe, Jesper Kragh, Tommy Olsen og Torben Valdbjørn Rasmussen. *Erfaringsopfølgning på lavenergibyggeri klasse 1 og 2 - med "Fremtidens Parcelhuse" som eksempel*. Teknisk rapport, 2011.
- [53] Casper Kofod. *Elforbrugets sammensætning, slutforbrugsanalyse på basis af målinger i 100 boliger i Odense*. Teknisk rapport.

-
- [54] Pawel Lewicki og Thomas Hill. *STATISTICS: Methods and Applications*. isbn 1-884233-59-7. StatSoft, 2006.
 - [55] Wei-Yin Loh og Yu-Shan Shih. *Split selection methods for classification trees*. Statistica Sinica, 7:815–840, 1997.
 - [56] Anne marie Primdal. *Galoperende elforbrug i bygninger får varmemeforbruget til at blegne*. <http://ing.dk/artikel/92643-galoperende-elforbrug-i-bygninger-faar-varmemeforbruget-til-at-blegne?highlight=Energi>, 2008. Ingeniøren.
 - [57] Mari Martiskainen. *Affecting Consumer Behaviour on Energy Demand*. Teknisk rapport, Sussex Energy Group, Brighton, 2007.
 - [58] Mario Mazzocchi. *Statistics for Marketing and Consumer Research*. isbn 978-1-4129-1122-1. SAGE Publications, 2008.
 - [59] Det Europæiske Miljøagentur. *Om klimaændringer*. <http://www.eea.europa.eu/da/themes/climate/about-climate-change>, 2011.
 - [60] Miljøstyrelsen. *Frisk luft i boligen styrker sundheden*. http://www.mst.dk/Nyheder/Pressemeddelelser/2009/Frisk_luft_styrker_sundheden.htm, 2009.
 - [61] Miljøstyrelsen og Watertech a/s. *Vandforbrug og forbrugsvariationer, Miljøprojekt Nr. 998*. Teknisk rapport, 2005.
 - [62] Michael Nashvili. *Decision Trees & Data Mining*. <http://www.decisiontrees.net/>, 2004.
 - [63] Annika Nilsson. *Energianvändning i nybyggda flerbostadshus på Bo01-området i Malmö*. Teknisk rapport 91-88722-30-9, Lunds Tekniska Högskola, Lunds universitet, 2003.
 - [64] Energi Nord. *Sammenlign dit elforbrug med gennemsnittet*. http://www.energinord.dk/dk/privat/spar_el/sammenlign_dit_elforbrug.htm, 2011.
 - [65] Erhvervs og Byggestyrelsen. *Bygningsreglement 1998 - Småhuse (inkl. tillæg 1-12) (01.01.2007 - 01.02.2008)*. <http://www.ebst.dk/bygningsreglementet.dk/brs98/0/42>, 1998-2006.
 - [66] Erhvervs og Byggestyrelsen. *Bygningsreglement 2008 (01.04.09 - 29.12.10)*. <http://www.ebst.dk/bygningsreglementet.dk/br07/0/42>, 2008.
 - [67] Erhvervs og Byggestyrelsen. *Boligejer.dk - Energiberegner*. <http://www.boligejer.dk/energiberegner>, 2011.
 - [68] Erhvervs og Byggestyrelsen. *Bygningsreglement 2010 (30.06.10 ->)*. <http://www.ebst.dk/bygningsreglementet.dk/br10/0/42>, 2011.
 - [69] Elsparefonden og Dansk Energi | Net. *Selvtjek Bolig - Hvor meget kan du spare?* <http://www.selvtjekbolig.sparel.dk/>, 2011.
 - [70] Sanne Wittrup og Erik Holm. *Varmeforbruget i nye bygninger er ude af kontrol*. <http://ing.dk/artikel/68433-varmeforbruget-i-nye-bygninger-er-ude-af-kontrol>, 2006.
 - [71] Dansk Vand og Spildevandsforening. *Vand i tal 2010 - DANVAs benchmarking og vandstatistik 2010*. Teknisk rapport, DANVA, 2010.
 - [72] Albertslund Kommune Miljø- og Teknikforvaltningen. *Grønt regnskab 2007 - Bilagsrapport, boliger og erhverv*. http://www.albertslund.dk/Borger/MiljoeOgNatur/GroentRegnskab/~media/Filer_Albertslund_dk/MiljoeOgForsyning/GroentRegnskab/Groentregnskab%202007/AK_GR07_Brap_BoligErhverv.ashx, 2008.

-
- [73] Oracle. *Oracle Data Mining Concepts, 11g Release 1 (11.1)*.
http://download.oracle.com/docs/cd/B28359_01/datamine.111/b28129.pdf, 2008.
- [74] Lene Holm Pedersen og Eva Broegaard. *Husholdningernes elforbrug - en analyse af attituder og adfærd på energi- og miljøområdet*. isbn 87-7509-532-7. Amternes og Kommunernes Forskningsinstitut, 1997.
- [75] Leen Peeters, Richard de Dear, Jan Hensen og William D'haeseleer. *Thermal comfort in residential buildings: comfort values and scales for building energy simulation*. Applied Energy, 86:772–780, 2009.
- [76] John Ross Quinlan. *Induction of Decision Trees*. Machine Learning, 1:81–106, 1986.
- [77] John Ross Quinlan. *C4.5: programs for machine learning*. isbn 1-55860-238-0. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [78] John Ross Quinlan. *Data Mining Tools See5 and C5.0*.
<http://www.rulequest.com/see5-info.html>, 2011.
- [79] John Ross Quinlan og Ron Kohavi. *Decision Tree Discovery*. Handbook of data mining and knowledge discovery, side 267–276, 2002.
- [80] Ali Rahbar, Hans Skyum Larsen og Troels Fjordbak Larsen. *Nye husholdningsapparaters elforbrug*. Teknisk rapport EB 2008/1, IT Energy, 2008.
- [81] Regeringen. *Strategi for reduktion af energiforbruget i bygninger*. Regeringen, 2009.
- [82] Andrew Roberts. *AI32 - Guide to Weka*.
<http://www.andy-roberts.net/teaching/ai32/weka.pdf>, 2005.
- [83] Rockwool. *Beregningsprogrammer*.
<http://guiden.rockwool.dk/vaerktoejer/beregningsprogrammer>, 2011.
- [84] Lior Rokach og Oded Maimon. *Data Mining with Decision Trees: Theory and Applications*. isbn 981-277-171-9. World Scientific, 2007.
- [85] H. Charles Romesburg. *Cluster Analysis for Researchers*. isbn 0-534-03248-6. Lifetime Learning Publications, 1984.
- [86] R. Saidur, H.H. Masjuki og M.Y. Jamaluddin. *An application of energy and exergy analysis in residential sector of Malaysia*. Energy Policy, 35:1050–1063, 2007.
- [87] SEAS-NVE. *Tjek dit energiforbrug*.
<http://www.seas-nve.dk/Privat/Energiraadgivning.aspx>, 2011.
- [88] Karin Soldaat og Laure Itard. *Influence of occupants on the energy use of balanced ventilation*. Proceedings of Clima 2007.
- [89] Dansk Standard. *DS/EN/ISO 7730 - Ergonomi inden for termisk miljø - Analytisk bestemmelse og fortolkning af termisk komfort ved beregning af PMV- og PPD-indekser og lokale termiske komfortkriterier*. Byggedata, 2006.
- [90] Danmarks Statistik. *Befolkningens brug af internet*.
<http://www.dst.dk/pukora/epub/upload/14039/it.pdf>, 2009.
- [91] Danmarks Statistik. *BOL22: Beboede boliger efter område, boligart, opførelsesår og installationsforhold*.
<http://www.statistikbanken.dk/BOL22>, 2010.
- [92] Danmarks Statistik. *Familiernes adgang til pc og internet i hjemmet efter type, region, indkomst, antal børn Pct af alle familier*.
<http://www.statistikbanken.dk/FABRIT01>, 2010.
- [93] Danmarks Statistik. *HISB3: Nøgletal om befolkningen*.
<http://www.statistikbanken.dk/HISB3>, 2010.

-
- [94] Danmarks Statistik. *VAND1: Forbruget af drikke- og råvand efter område og anvendelse*. <http://www.statistikbanken.dk/vand1>, 2010.
- [95] Danmarks Statistik. *INDKP5: Indkomst i alt efter område, enhed, køn, alder og indkomstinterval*. <http://www.statistikbanken.dk/INDKP5>, 2011.
- [96] K. E. Thomsen, J.M. Schultz og B. Poel. *Measured Performance of 12 demonstration projects - IEA Task 13 „advanced solar low energy buildings“*. Energy and Buildings, 37:111–119, 2005.
- [97] Henrik Tommerup. *Måling af bruttoenergiforbrug i nybyggeri svarende til BR2005 energikrav, Byggesystem: Skalmurede porebetonelementer 1*. Teknisk rapport 87-7877-162-5, BYG-DTU, 2004.
- [98] Henrik Tommerup. *Måling af bruttoenergiforbrug i nybyggeri svarende til BR2005 energikrav, Byggesystem: Skalmurede porebetonelementer 2*. Teknisk rapport 87-7877-163-3, BYG-DTU, 2004.
- [99] Henrik Tommerup. *Energirenovering af typisk parcelhus opført i perioden 1960-1980, Del 2: Målinger*. Teknisk rapport 9788778772404, BYG-DTU, 2007.
- [100] Henrik Tommerup og Jørgen Rose. *Forsøgshus med nye typer klimaskærmskonstruktioner, Byggesystem: Lette ydervægselementer i stålskelet*. Teknisk rapport 87-7877-099-8, BYG-DTU, 2002.
- [101] Toolbox. *TPB - Theory of Planned Behavior (A)*. <http://toolbox.systime.dk/index.php?id=852>, 2011.
- [102] Geoffrey K.F. Tso og Kelvin K.W. Yau. *Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks*. Energy, 32:1761–1768, 2007.
- [103] W. Fred van Raaij og Theo M.M. Verhallen. *Patterns of residential energy behavior*.
- [104] Version2. *It-leksikon*. <http://www.version2.dk/leksikon/Bit>, 2011.
- [105] Wikipedia. *Precision and recall*. http://en.wikipedia.org/wiki/Precision_and_recall, 2011.
- [106] Wikipedia. *Receiver operating characteristic*. http://en.wikipedia.org/wiki/Receiver_operating_characteristic, 2011.
- [107] Charlie Wilson og Hadi Dowlatabadi. *Models of decision making and residential energy use*. Annual Review of Environment and Resources, 32:169–203, 2007.
- [108] Zhun Yu, Fariborz Haghighat, Benjamin C.M. Fung og Hiroshi Yoshino. *A decision tree method for building energy demand modeling*. Energy and Buildings, 42:1637–1646, 2010.
- [109] Jean Paul Zimmermann. *End-use metering campaign in 400 households in Sweden - Assessment of the Potential Electricity Savings*. Teknisk rapport CONTRACT 17-05-2743, ENERTECH, 2009.