



Appendiks

Brugeradfærd i boliger

**anvendelse af klassificeringstræer til
estimering af energiforbrug**

**Lars Lund Michaelsen
Louise Sabrina Svendsen**

Indholdsfortegnelse

Appendiks A	Parametre med indflydelse på energiforbruget	4
Appendiks B	Udvikling af nyt beregningsprogram	9
Appendiks C	Litteraturstudie	11
C.1	Metoder til analysering af data	11
C.1.1	Theory of Planned Behaviour	11
C.2	Beskrivelse af brugertyper	13
C.3	Lovgivning	15
C.3.1	DS 474	15
C.3.2	DS/EN/ISO 7730	16
C.3.3	Bygningsreglementet	19
C.4	Rumopvarmning	20
C.4.1	Anvendelse og styring af rumopvarmningen	20
C.4.2	Praktisk anvendelse af vejledning fra normer	23
C.4.3	Solafskærmning	25
C.4.4	Infiltration	28
C.4.5	Udluftning	29
C.4.6	Mekanisk ventilation	32
C.4.7	Diskussion	35
C.5	Elforbrug	37
C.5.1	Slutanvendelse for elforbruget	37
C.5.2	Væsentlige parametre for elforbruget	40
C.6	Vandforbrug	41
C.6.1	Samlet vandforbrug	41
Appendiks D	Opsummering af indetemperaturer	46
Appendiks E	Udeklima	48
E.1	Dansk udeklima sammenlignet med forholdene i andre projekter	48
Appendiks F	Faktor- og komponentanalyse	51
F.1	Eksplorativ faktoranalyse	52
F.2	Konfirmativ faktoranalyse	55
F.3	Komponentanalyse	56
Appendiks G	Klyngeanalyse	58
Appendiks H	Boligtypens betydning i klassificeringstræet for rumopvarmning	64
Appendiks I	Hypotesetest	65

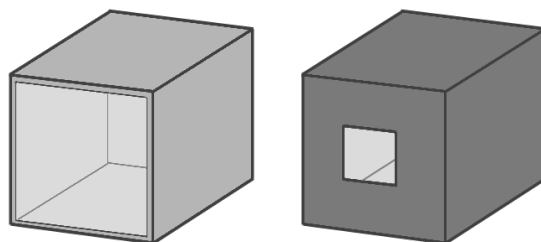
I.1	Grundlæggende teori om Chi-square test for goodness of fit	65
I.2	Grundlæggende teori om Kolmogorov-Smirnov one-sample test	67
I.3	Regneark til udførelse af Chi-square test for goodness of fit	68
I.4	Regneark til udførelse af Kolmogorov-Smirnov one-sample test	69
Appendiks J Vurdering af lufttemperatur		70
J.1	Overordnet temperaturspænd	70
Appendiks K Vurdering af elforbrug		78
K.1	Valg af normeringsenhed	78
K.2	Vurdering af elforbrugets udvikling samt betydningen heraf	79
K.3	Test af fordelingsstype	80
K.4	Estimering af fordelingsparametre	83
Appendiks L Vurdering af varmtvandsforbrug		90
L.1	Husstandes samlede vandforbrug	91
L.1.1	Valg af normeringsenhed for det samlede vandforbrug	91
L.1.2	Korrektion af oplysninger fra kilde i litteraturstudiet	92
L.1.3	Vurdering af vandforbrugets udvikling samt betydningen heraf	93
L.1.4	Korrektion for sammenhængen mellem vandforbruget per person og husstandsstørrelsen	94
L.1.5	Test af fordelingsstype	95
L.1.6	Estimering af fordelingsparametre	99
L.2	Husstandes varmtvandsforbrug	100
Appendiks M Algoritmer		103
M.1	QUEST	103
M.2	CART	104
M.3	ID3, C4.5 og See5	104
M.3.1	Fordele og ulemper ID3, C4.5 og See5	105
M.4	CHAID	106
M.4.1	Fordele og ulemper, CHAID	107
M.5	Opsummering	107
Appendiks N Undersøgelse af nye mulige attributter		108
N.1	Rumopvarmning	108
N.1.1	Lufttemperatur	108
N.1.2	Boligens størrelse og boligforbrug	111
N.1.3	Beboernes alder	112
N.1.4	Tilstedeværelse af børn og teenagere	113
N.1.5	Uddannelsesniveau	113
N.1.6	Opholdstid i boligen	114
N.1.7	Beboernes holdning til miljø og energibesparelser	115
N.2	Vandforbrug	116
N.2.1	Antal beboere	117
N.2.2	Antal bade	120
N.2.3	Antal tøjvaske	122
N.2.4	Tilstedeværelse af børn og teenagere	123
N.2.5	Beboernes alder	124

N.2.6	Uddannelsesniveau	125
N.2.7	Opholdstid i boligen	126
N.2.8	Boligareal	127
N.2.9	Beboernes holdning til miljø og energibesparelser	128
N.3	Elforbrug	130
N.3.1	Antal beboere	131
N.3.2	Boligareal	131
N.3.3	Beboernes alder	132
N.3.4	Tilstedeværelse af børn og teenagere	133
N.3.5	Uddannelsesniveau	134
N.3.6	Antal tøjvaske og tørring af tøj	134
N.3.7	Anvendelse af elbesparende foranstaltninger	135
N.3.8	Opholdstid i boligen	136
N.3.9	Beboernes holdning til miljø og energibesparelser	137
Appendiks O Beskrivelse af attributter til klassificeringstræerne		139
O.1	Tekniske attributter	139
O.2	Demografiske attributter	140
O.3	Adfærdsrelaterede attributter	141
O.4	Attributter der ikke er medtaget	143
Appendiks P Validering af brugen af programmet WEKA		145
P.1	Introduktion til eksemplet fra Ross Quinlan	145
P.2	Anvendelse af eksemplet fra Ross Quinlan	147
P.2.1	Sammenfatning	148
Appendiks Q Resultatfil fra Weka		149
Appendiks R Begrænsning af klassificeringstræets størrelse		151
Appendiks S Håndtering af numeriske attributter i klassificeringstræet		154
S.1	Evalueret af et klassificeringstræets nøjagtighed	157
S.1.1	Confusion matrix	158
S.1.2	Andelen af korrekt klassificerede elementer i klassificeringstræet	160
S.1.3	True positive rate og False positive rate	160
S.1.4	Precision	161
S.1.5	F-measure	161
Litteratur		163

Parametre med indflydelse på energiforbruget

I dette appendiks beskrives de projekter nærmere, der er anvendt i afsnit 2.2.1 i hovedrapporten, til vurdering af væsentlige parametre i forbindelse med bygningers energiforbrug.

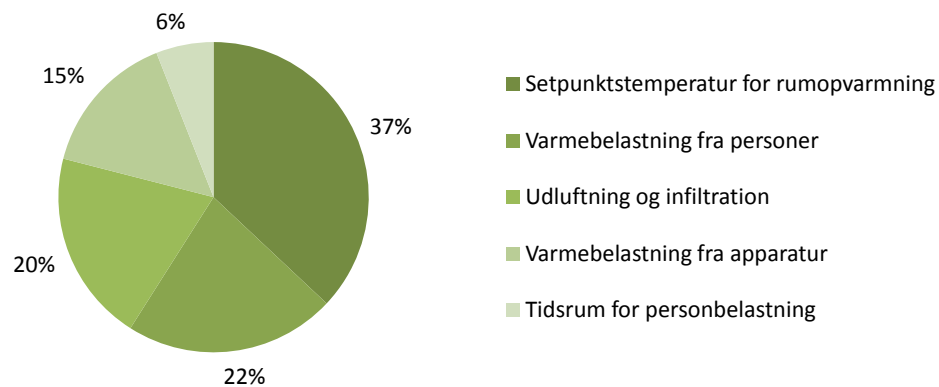
Med henblik på at identificere de parametre, der har den største betydning for variationen af energiforbruget i boliger, er der foretaget et studie af tidligere undersøgelser omhandlende dette emne. I det efterfølgende bliver der gennemgået nogle af de projekter, der har fremhævet disse parametre, hvilket skal danne grundlaget for en efterfølgende udvælgelse, af hvilke parametre hovedfokuset skal hvile på i dette projekt. Der undersøges både for indflydelsen af konstruktionsmæssige parametre, såsom transmissionskoefficienter, vinduesareal og lignende, men også parametre, der er under indflydelse af brugerne, såsom setpunktstemperatur for rumopvarmning, anvendelse af elektriske apparater mv. Derved kan der dannes et overblik over hvilke faktorer, der har størst indflydelse på energiforbruget i forskellige boliger, eftersom undersøgelser viser, at det kan være vidt forskelligt fra bolig til bolig. Et eksempel derpå kan findes i en robusthedsanalyse udført af *Hoes et al.* [29] på to vidt forskellige, tænkte lokaler som skitseret på figur A.1. Lokalet til venstre på figur A.1 har en lav termisk masse og en stort vinduesareal i modsætning til lokalet til højre på figuren, der har en høj termisk masse og et mindre vinduesareal. Lokalet med den lave termiske masse og det store vinduesareal er mere robust overfor brugeradfærd end det andet lokale, hvilket skydes at de ydre påvirkninger, såsom udetemperatur og solstråling, har større indflydelse på energiforbruget til eksempelvis opvarmning og køling end brugeradfærden har. I disse to tilfælde vil konstruktionsmæssige parametre have størst indflydelse variationen af energiforbruget i lokalet med det store vinduesareal, hvorimod parametre der er under indflydelse af brugerne vil have stort indflydelse på lokalet med det mindre vinduesareal.



Figur A.1: Skitsering af anvendte lokaler fra *Hoes et al.* [29].

I forbindelse med en række målinger, der er foretaget på otte tilnærmelsesvis identiske rækkehuse, har *Brohus et al.* [9] udarbejdet en model af boligerne i et hygrotermisk bygnings-simuleringsprogram. På denne simuleringsmodel er der indledningsvis foretaget en screeningsanalyse for at klarlægge hvilke inputparametre, der kan behandles deterministisk og for hvilke der skal indtastes stokastiske oplysninger om. Screeningsanalysen ender ud med, at der vælges at behandle fem parametre stokastisk, hvorfor der fastlægges fordelingsfunktioner, middelværdier og spredninger på baggrund af litteratur, erfaringer og skøn for disse parametre.

Efterfølgende blev der foretaget en følsomheds- og usikkerhedsanalyse, hvor resultaterne viser, at de parametre, der har størst indflydelse på usikkerheden i energiforbruget relaterer sig til brugeradfærd. Det drejer sig eksempelvis om setpunktstemperatur for rumopvarmning, varmebelastning fra personer, brugen af naturlig ventilation i boligen, varmebelastningen fra apparatur samt tidsrum for personbelastningen. Parametrenes indbyrdes indflydelse på den samlede usikkerhed for energiforbruget er skitseret på figur A.2.



Figur A.2: Parametrenes indflydelse på usikkerheden i energiforbruget [9].

I forbindelse med udvikling af en metode, der skal tage højde for nogle af de usikkerheder og spredninger der måtte være i beregninger af energibehov i bygninger, har *Brohus et al.* [8] foretaget en følsomhedsanalyse af forskellige parametres indflydelse på det beregnede energibehov. Analysen er foretaget på et 180 m^2 parcelhus hvor 75 forskellige parametre i relation til boligen er behandlet stokastisk. Der er anvendt to forskellige analysemetoder til identificering af de vigtigste parametre, hvilket er henholdsvis *Morris metoden* og *Fourier Amplitude Sensitivity Test* (FAST). Metoderne rangerer de ti vigtigste parametre forskelligt, men ikke desto mindre indikerer analyserne hvilke parametre, der i det aktuelle tilfælde er vigtigst. Disse parametre er oplistet i tabel A.1 i forhold til de to forskellige analyser, hvor rangeringen angiver faktorernes indflydelse på det beregnede energibehov, hvor 1 er den vigtigste.

Rangering	Morris	FAST
1	Setpunktstemperatur for rumopvarmning	Naturlig ventilation om vinteren i brugstiden
2	Naturlig ventilation om vinteren i brugstiden	Setpunktstemperatur for rumopvarmning
3	Varmtvandsforbrug	Varmtvandsforbrug
4	Horisontvinkel	Transmissionskoefficient for vægkonstruktion
5	Varmetilskud fra apparatur i brugstiden	Solafskærmningsfaktor
6	Naturlig ventilation om vinteren udenfor brugstiden	Horisontvinkel
7	Transmissionskoefficient for vinduer	Numerisk referenceparameter for opvarmning
8	Naturlig ventilation om sommeren	Reference tidskonstant for opvarmning
9	Numerisk referenceparameter for opvarmning	Transmissionskoefficient for vinduer
10	Transmissionskoefficient for vægkonstruktion	Varmetilskud fra apparatur i brugstiden

Tabel A.1: Rangering af vigtige parametre fra følsomhedsanalyse [8].

Verhallen et al. [87] har undersøgt hvilke parametre, der kan forklare variationer af energiforbrug til opvarmning af boliger. Undersøgelsen inddrager 145 tilnærmelsesvis identiske boliger i Holland og er udført som en kombination af forbrugsaflysninger fra november 1976 til november 1977 samt en spørgeundersøgelse blandt beboerne omhandlende holdninger og brug af boligen.

De følgende fire overordnede parametres forklaring af variationen af energiforbruget til rumopvarmning undersøges af *Verhallen et al.*; adfærdsparametre, bygningskarakteristika, specielle forhold samt sociodemografi. Adfærdsparametre dækker over de faktorer, der er angivet i tabel A.2, såsom lufttemperatur ved fravær, styring af brænder i gasfyr mv. I denne forbindelse bør det nævnes, at alle husstandene i denne undersøgelse har gasfyr i boligerne, som anvendes til opvarmning, hvilket er årsagen til, at denne faktor er medtaget i analysen. Bygningskarakteristika dækker foruden de faktorer, der er nævnt i tabel A.2 også over vindretning. Denne faktor er dog ikke medtaget i tabellen, eftersom dette forhold ikke kan ændres. Faktorer som fravær fra hjemmet i forbindelse med arbejde, weekend eller sygedage, besøg af andre personer etc. der har indflydelse på energiforbruget i boligen er samlet under den overordnede parameter specielle forhold. Sociodemografi beskriver faktorer som indkomst, uddannelse, alder, antal børn mv.

I tabel A.2 er de underordnede faktorer indfyldt på de fire overordnede parametre vist, samtidig med at det er oplyst hvor stor forklaring hver enkelt af de overordnede parametre har af variationen af energiforbruget til rumopvarmning. Nederst i tabellen er der beskrevet hvor meget de fire parametre kan forklare af variationen, når alle medtages på samme tid.

Parameter	Effekt på parameter [%]	Forklaring af variation [%]
1. Adfærdsparametre		26
<i>Lufttemperatur ved fravær</i>	33	
<i>Styring af brænder i gasfyr</i>	19	
<i>Temperatur i soveværelse om natten</i>	14	
<i>Brug af hoveddør</i>	14	
<i>Lufttemperaturer ved tilstedeværelse</i>	9	
<i>Udluftning</i>	6	
<i>Brug af gardiner</i>	3	
<i>Brug af soveværelse</i>	-	
2. Bygningskarakteristika		24
<i>Isoleringsgrad</i>	35	
<i>Naboernes energiforbrug til opvarmning</i>	22	
<i>Sammenbygget med én bygning</i>	21	
<i>Sammenbygget med to bygninger</i>	5	
3. Specielle forhold		11
4. Sociodemografi		6
De fire overordnede parametre sammen		58

Tabel A.2: De overordnede parametres indbyrdes forklaring af variationen af energiforbruget til rumopvarmning, samt de beskrivende faktorerers effekt på de overordnede parametre [87].

Santin et al. [66] har undersøgt hvilke parametre, der kan forklare variationer af energiforbrug til rumopvarmning og opvarmning af brugsvand i boliger. Undersøgelsen er foretaget på baggrund af en database indeholdende forbrugsdata fra tre år og spørgeundersøgelser fra 15000 husstande i Holland.

Analysen er delt op med to overordnede parametre; bygningskarakteristika og så boligkarakteristika og adfærdsparametre. Parameteren bygningskarakteristika dækker over forhold som eksempelvis boligareal, opførelsesår, isolering mv, hvorimod parameteren boligkarakteristika og adfærdsparametre dækker over hvorledes boligen benyttes, det vil eksempelvis sige hvor mange soveværelser er opvarmet, men det dækker også over antal og alder af beboere mv. I tabel A.3 er de underordnede faktorerers indfyldelse på de to overordnede parametre vist, samtidig med at det er oplistet hvor stor forklaring hver enkelt af de overordnede parametre har af variationen af energiforbruget til rumopvarmning og opvarmning af varmt brugsvand.

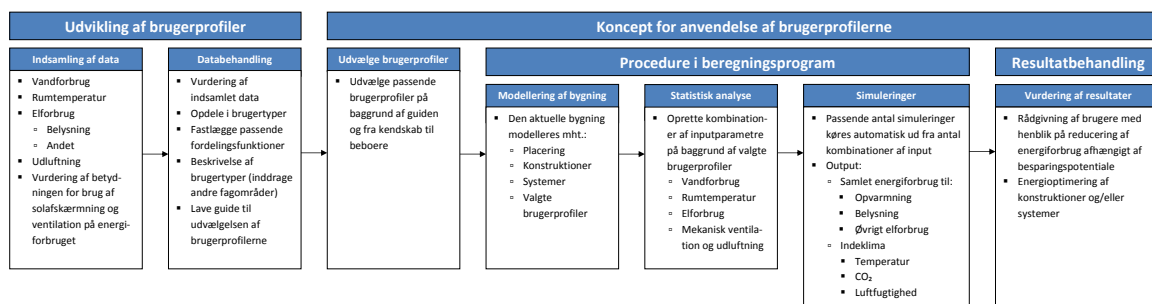
Parameter	Effekt på parameter [%]	Forklaring af variation [%]
Bygningskarakteristika		37,9
<i>Boligareal</i>	<i>32,1</i>	
<i>Isolering i facade</i>	<i>8,7</i>	
<i>Opførelsesår</i>	<i>8,2</i>	
<i>Termostat som temperaturstyring</i>	<i>7,7</i>	
<i>Termoruder</i>	<i>6,3</i>	
<i>Isolering i gulvkonstruktion</i>	<i>6,1</i>	
<i>Antal værelser</i>	<i>5,6</i>	
Boligkarakteristika og adfærdsparametre		4,2
<i>Antal opvarmede soveværelser</i>	<i>12,8</i>	
<i>Alder af beboere</i>	<i>6,4</i>	
<i>Temperatur om aftenen</i>	<i>6,1</i>	
<i>Temperatur om dagen</i>	<i>4,7</i>	
<i>Antal personer i dagtimerne</i>	<i>3,8</i>	
<i>Temperatur om natten</i>	<i>3,7</i>	
<i>Antal personer i weekenden</i>	<i>2,6</i>	

Tabel A.3: De overordnede parametres indbyrdes forklaring af variationen af energiforbruget til rumopvarmning, samt de beskrivende faktorerers effekt på de overordnede parametre [66].

Udvikling af nyt beregningsprogram

Det er i afsnit 2.3 i hovedrapporten beskrevet, at der kan knyttes nogle sandsynlighedsfordelinger til bestemte beboer- og adfærdstyper. I dette afsnit beskrives hvorledes dette kan udføres.

Der skal indsamles data for eksempelvis vandforbrug, lufttemperaturer, elforbrug samt et mønster for udluftning i boliger og styring af mekanisk ventilation. Derudover kan der blive vurderet på betydningen af beboernes styring af solafskærmning på energiforbruget og indeklimaet. Disse kan sammenholdes med viden om hvilke kendetegn beboere mere forskellige forbrug er karakteriseret ved. Den mulige fremgangsmåde er illustreret på figur B.1.



Figur B.1: Mulig fremgangsmåde for hvorledes sandsynlighedsfordelinger for forskellige parametre kan knyttes til bestemte beboer- og adfærdstyper og efterfølgende anvendes i simulering af energiforbrug og indeklima.

På baggrund af det indsamlede data skal der udføres en databehandling, således kvaliteten af måledata og eventuelle ekstreme værdier vurderes. Efterfølgende kan data sorteres og opdeles i et passende antal profiltyper. Et eksempel kan være at data for vandforbruget opdeles i et bestemt antal intervaller. Der modelleres nogle fordelinger for hvert interval, der beskriver målingerne bedst muligt. Samme fremgangsmåde kan anvendes på de øvrige parametre, der er illustreret i den første boks på figur B.1. På denne måde er der mulighed for, at beboerne eksempelvis kan have et lavt vandforbrug, et gennemsnitligt elforbrug men høje lufttemperaturer.

Der skal udarbejdes en vejledning, som kan anvendes til udvælgelse af brugerprofiler. I den forbindelse er det nødvendigt at inddrage viden fra andre fagområder, således at der eksempelvis ud fra nogle spørgsmål, der stilles til beboerne, kan vælges nogle passende profiler.

Disse profiler modelleres i et simuleringsprogram sammen med den øvrige bygningsdata, således at modellen beskriver de aktuelle forhold bedst muligt. Det vil være optimalt, hvis samme simuleringsprogram kan lave en statistisk analyse af ind- og uddata, således beregningerne ikke skal forløbe gennem flere forskellige programmer. Det vil sige, at ud fra de valgte stokastiske variable, som eksempelvis vandforbrug, lufttemperatur og elforbrug, skal programmet fremstille forskellige kombinationer af disse parametre, således der tages højde for sandsynligheden og derved også fordelingen af mulige værdier af hver enkelt parameter. De forskellige kombinationer af parametrene anvendes som inddata i simuleringerne, hvorved der fås en række sammenhængende ind- og uddata.

Når simuleringerne har fundet sted, kan det være aktuelt at undersøge hvorledes det samlede energiforbrug er delt ud på eksempelvis elforbrug til belysning, drift af ventilation eller energiforbrug til rumopvarmning, og hvorvidt der er et tilfredsstillende indeklima. Hvis nogle parametre med relation til konstruktionerne er behandlet stokastisk, kan indflydelsen på energiforbruget for hver enkelt parameter sammenholdes. Derved kan betydningen for konstruktionernes fysiske egenskaber og indflydelsen fra brugeradfærd på energiforbruget sammenholdes direkte, hvilket kan danne grundlaget for en efterfølgende rådgivnings- eller optimeringsproces.

Litteraturstudie

Nærværende kapitel indeholder et litteraturstudie der indeholder yderligere information om de emner der gennemgås i kapitel 3 i hovedrapporten.

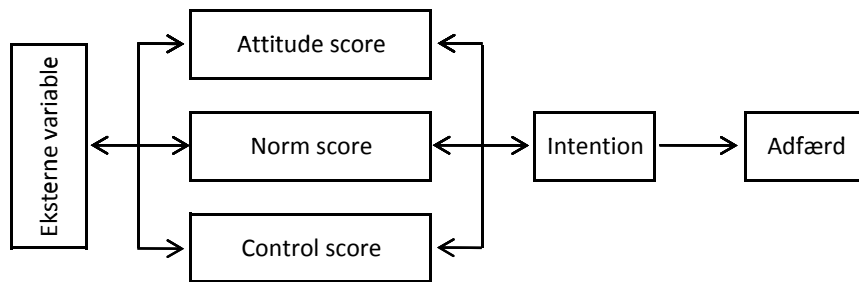
C.1 Metoder til analysering af data

Nærværende afsnit indeholder yderligere information om de metoder der er præsenteret i afsnit 3.1 i hovedrapporten.

C.1.1 Theory of Planned Behaviour

Modellen *Theory of Planned Behaviour* er udviklet af *Icek Ajzen* i 1985 [85]. Den er blandt andet anvendt inden for psykologiens verden 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.* [15], *Mari Martiskainen* [43] og *Wilson et al.* [92], vist sig at være anvendelig til at forudsige forbrugeradfærden inden for husholdningssektoren. Desuden anvendes metoden i projekter fra *Gill et al.* [22], der gennemgås senere i dette afsnit, og *Barr et al.* [5] der gennemgås i afsnit 3.1.3 i hovedrapporten.

Theory of Planned Behaviour er opbygget som illustreret i figur C.1. Modellen tager udgangspunkt i, at en persons holdninger og måde at agere påvirkes af nogle eksterne variable, der eksempelvis kunne være personens uddannelse og økonomiske situation. De eksterne variable påvirker tre overordnede dele, der betegnes *attitude score*, *norm score* og *control score*, se figur C.1. Disse tre hovedbestanddele anvendes i formel C.1, hvor der beregnes en *total behaviour score*, forklaring af de enkelte bestanddele ses i tabellerne C.1-C.3. På baggrund af *total behaviour score*, der er angivet i formel C.1, fremkommer personens intention som er et udtryk for den plan/hensigt, som personen har. Denne hensigt følges op med en adfærd, som personen udviser. [1, 22]



Figur C.1: Opbygning af modellen *Theory of Planned Behaviour* [85].

$$Total\ behaviour\ score = Attitude\ score + Norm\ score + Control\ score \quad (C.1)$$

Variabel	Beskrivelse
Attitude score	Denne variabel er et udtryk for hvilken holdning, en person har til en specifik handling eller adfærd, og om konsekvensen af denne er positiv eller negativ. Attitude score består af de to variable betegnet henholdsvis ved <i>Behavioural belief</i> og <i>Outcome evaluation</i> .
Behavioural belief	Denne variabel er et udtryk for en persons holdning til konsekvensen af en specifik handling.
Outcome evaluation	Denne variabel afhænger af, hvorvidt den omtalte handling opfattes positivt eller negativt af personen.

Tabel C.1: Forklaring af *attitude score*, der indgår i formel C.1, og bestanddelene af denne score [1, 22].

Variabel	Beskrivelse
Norm score	Denne variabel dækker over hvordan en person påvirkes af samfundet til enten at udføre/ikke udføre visse handlinger eller ændre/ikke ændre en adfærd over for en bestemt handling. Den består af de to variable betegnet henholdsvis ved <i>Normative beliefs</i> og <i>Motivation to comply</i> .
Normative beliefs	Denne variabel er et udtryk for personens holdning til en specifik adfærd. Den kan være påvirket af persons omgivelser i form af familie, venner, organisationer og lignende.
Motivation to comply	Denne variabel er et udtryk for hvorvidt personens holdninger er påvirket af omgivelserne.

Tabel C.2: Forklaring af *norm score*, der indgår i formel C.1, og bestanddelene af denne score [1, 22].

Variabel	Beskrivelse
Control score	Denne variabel beskriver i hvilken grad, individet selv tror, at dette overhovedet er i stand til at udføre en bestemt handling eller leve efter en bestemt adfærd. Den består af de to variable betegnet henholdsvis ved <i>Control belief strength</i> og <i>Control belief power</i> .
Control belief strength	Denne variabel er individets personlige holdning til, om denne har nemt eller svært ved at udføre en specifik handling eller leve efter en specifik adfærd.
Control belief power	Denne variabel er et udtryk for de fysiske og økonomisk forhold samt for faglige kompetencer, der muliggør eller besværliggør en bestemt handling eller adfærd.

Tabel C.3: Forklaring af *control score*, der indgår i formel C.1, og bestanddelene af denne score [1, 22].

C.2 Beskrivelse af brugertyper

I nærværende afsnit gennemgås de vigtigste kendetegn for de fire udviklede brugertyper fra Barr *et al.* [5], der er gennemgået i afsnit 3.1.3 i hovedrapporten.

Den engagerede type	
Kendetegn adfærd:	Høj sandsynlighed for, at denne brugertype køber sparepærer og energimærkede apparater. Personerne i denne gruppe sænker temperaturen i ubenyttede rum, og klæder sig i mere tøj før der tændes eller reguleres for varmen. De slukker lyset i ubenyttede rum og vasker kun når vaskemaskinen er fuld, se figur 3.2. i afsnit 3.1.3 i hovedrapporten
Kendetegn demografi:	Denne type har den højeste gennemsnitsalder, på 55 år. Størrelsen på husstanden, er typisk på 1 – 2 personer. Sandsynligheden for at de ejer deres eget hjem er på 83 %. Typen har størst engagement i organisatorisk/politisk arbejde med en medlemsrate på 17 %.
Kendetegn sociale parametre og miljømæssige værdier:	Den engagerede type ser vigtighed i at kunne bidrage til fællesskabet og sætter lydighed og hjælpsomhed som vigtige karaktertræk.

Tabel C.4: Kendetegn for *den engagerede type* [5].

Den generelle type	
Kendetegn adfærd:	Med hensyn til energibesparelser og indkøbsvaner har denne type stort set samme kendetegn, som den engagerede type se tabel C.4 se figur 3.3 i afsnit 3.1.3 i hovedrapporten. Den største forskel ligger i en lavere sandsynlighed for at husstandene køber energimærkede.
Kendetegn demografi:	Gennemsnitsalderen for denne type er 52 år. Typisk er størrelsen af husstandene på 1-2 personer. I denne gruppe er der en sandsynlighed på 74 %, for at de ejer deres eget hjem. næststørste engagement med hensyn til organisatorisk /politisk arbejde med en medlemsrate på 10%.
Kendetegn sociale parametre og miljømæssige værdier:	Den generelle type ser en knap så stor vigtighed i at kunne bidrage til fællesskabet, men karaktertrækkene lydighed og hjælpsomhed som lige så vigtige, som hos den engagerede type.

Tabel C.5: Kendetegn for *den generelle type* [5].

Den periodiske type	
Kendetegn adfærd:	Forskellen mellem denne type og den generelle type er at der sker en stor reduktion i den energibesparende adfærd for denne type, se figur 3.4 i afsnit 3.1.3 i i hovedrapporten. Derudover er sandsynligheden for at de køber energimærkede apparater reduceret betragteligt.
Kendetegn demografi:	Gennemsnitsalderen er for denne type er 46 år. Størrelsen på husstandene er typisk 2 – 4 personer, hvilket er større end de to typer beskrevet ovenfor. Sandsynlighed for at de ejer deres eget hjem er på 71 %, Med hensyn til organisatorisk/politisk arbejde er der for denne type en medlemsrate på 8 %.
Kendetegn sociale parametre og miljømæssige værdier:	Sammenlignet med den generelle type betyder det at kunne bidrage til fællesskabet mindre, karaktertrækket hjælpsomhed er også mindre fremtræden, hvorimod karaktertrækket lydighed ikke er lige så vigtigt.

Tabel C.6: Kendetegn for *den periodiske type* [5].

Den ikke-energibesparende type	
Kendetegn for adfærd:	Denne type ser ikke efter energimærkede apparater eller sparepærer og vil generelt kun tage del i energibesparende aktiviteter, når det slet ikke går ud over komforten. Forskellen mellem denne type og den periodiske type ses tydeligt ud fra figur 3.5, der kan ses i afsnit 3.1.3 i hovedrapporten, hvor de energibesparende adfærdsaktiviteter er kraftigt reduceret.
Kendetegn demografi:	I denne gruppe har personerne den laveste gennemsnitsalder på 43 år. I forhold til de tre øvrige typer er der markant flere mænd til stede i denne gruppe, 50 %. Størrelsen på husstandene er generelt større, med et spænd fra 2 – 5 personer. I denne gruppe er der mindst sandsynlighed for, at personerne ejer deres eget hjem 62 %. Derudover er der mindst sandsynlighed for, at de er medlem af en organisation, hvilket kommer til udtryk i en medlemsrate på 8 %.
Kendetegn sociale parametre og miljømæssige værdier:	Modsat de andre typer så rangerer hjælpsomhed og fællesskabsfølelsen mindst for denne type, hvorimod de betragter rigdom og indflydelse som vigtige parametre.

Tabel C.7: Kendetegn for *den ikke-energibesparende type* [5].

C.3 Lovgivning

C.3.1 DS 474

Normen *DS 474* [12] indeholder metoder, der skal hjælpe med at specificere, projektere og kontrollere det termiske indeklima i bygninger, hvor personer opholder sig. Ved specifikation af krav til 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* fremgår af tabel 3.4.

Krav til det termiske indeklima stilles som krav til de termiske indeklimaparametre fordelt på to kategorier; generelle påvirkninger og lokale påvirkninger, se tabel C.8 og C.9.

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 C.8: Krav til den operative temperatur, som en del af de generelle påvirkninger [12].

Lokale påvirkninger	Temperatur [°C]
Forskel i vertikal lufttemperatur mellem 0,1 m og 1,1 m over gulvet, Δt_a	$\Delta t_a < 3$
Strålingstemperaturasymmetri ved vinduer og andre kolde, vertikale overflader i forhold til en lille, vertikal flade 0,6 m over gulvet, Δt_{pr}	$\Delta t_{pr} < 10$
Strålingstemperaturasymmetri pga. et varmt loft i forhold til en lille, horisontal flade 0,6 m over gulvet, Δt_{pr}	$\Delta t_{pr} < 5$
Overfladetemperatur af gulv, t_g	$19 < t_g < 26$

Tabel C.9: Krav til de termiske indeklimaparametre for lokale påvirkninger [12].

Kravene i tabel C.8 er baseret på at mindre end 10 % er utilfredse med det generelle termiske indeklima. I tabel C.9 er grænserne angivet, så antallet af utilfredse er mindre end 5 – 10 %. For de angivne grænser ovenfor gælder, at den enkelte person kan opnå yderligere tilfredshed ved at regulere påklædningen hensigtsmæssigt.

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

C.3.2 DS/EN/ISO 7730

Den europæiske standard *DS/EN/ISO 7730* [71] indeholder nogle 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 beboerne.

Kravene, der er givet for de fire parametre i tabel C.9 ovenfor, sammenlignes med oplysningerne givet i *DS/EN/ISO 7730*. Herfra vides det, at kravene, givet i *DS 474*, svarer til kategorierne vist i tabel C.10 nedenfor. Kravene til de generelle påvirkninger fra *DS 474*, overholder begge kategori B i *DS/EN/ISO 7730*, hvorimod kravene for de lokale påvirkninger opnår enten en kategori A eller B.

Generelle og lokale påvirkninger	Kategori
Operativ temperatur i opholdszonen (vinter), t_o	B
Operativ temperatur i opholdszonen (sommer), t_o	B
Forskel i vertikal lufttemperatur mellem 0,1 m og 1,1 m over gulvet, Δt_a	B
Strålingstemperaturasymmetri ved vinduer og andre kolde vertikale overflader i forhold til en lille vertikal flade 0,6 m over gulvet, Δt_{pr}	A
Strålingstemperaturasymmetri pga. et varmt loft i forhold til en lille horisontal flade 0,6 m over gulvet, Δt_{pr}	A
Overfladetemperatur af gulv, t_g	A

Tabel C.10: Kategori for indeklimaparametrene i *DS 474* [12] når de sammenlignes med grænserne i *DS/EN/ISO 7730* [71].

De generelle påvirkninger i tabel C.10 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 samme måde [71]. 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*.¹
- Beklædningsniveau vinter: 1,0 *clo*.²

Beklædningsniveauet forudsættes, for beboerne i boliger, at kunne varieres efter behov, men de forudsatte værdier svarer godt overens med hverdagstøj for sommer- og vintersituationer. Aktivitetsniveauet i boliger kan variere fra sovende aktivitet (0,8 *met*) til rengøringsaktivitet (2,0 *met*), hvilket er afvigende fra den anvendte forudsætning. I tabel C.11 er der oplistet nogle eksempler på hvilke lufttemperaturer³, der er optimale for den givne påklædning og aktivitet. I vinterperioden er der opnået lufttemperaturer mellem $19,3 - 24,2 \text{ }^\circ\text{C}$, og for sommerperioden er der et temperaturspænd for lufttemperaturen mellem $20,3 - 26,6 \text{ }^\circ\text{C}$. Variation i de operative temperaturer er i *DS 474* mellem $20 - 24 \text{ }^\circ\text{C}$ og $23 - 26 \text{ }^\circ\text{C}$ for henholdsvis vinter- og sommerperioden. På baggrund af dette antages de ovenstående grænser for det termiske indeklima for kontorer og lignende kategori B som anvendelige grænser for boligen.

¹Svarer til: underbukser, skjorte m. korte ærmer, lette bukser, sokker og sko.

²Svarer til: underbukser, undertrøje m. korte ærmer, skjorte, bukser, v-sweater, sokker og sko.

³Jf. P.O. Fanger side 140 i [16]. Lufttemperaturen er en målbar temperatur mellem luftens reelle temperatur uden strålingspåvirkning og middelstrålingstemperaturen.

Aktivitetsniveau	Beklædningens varmeledningsmodstand	Lufttemperatur
[<i>met</i>]	[<i>clo</i>]	[°C]
1,0	1,0	24,2
1,2	1,0	22,6
1,6	1,0	19,3
1,0	0,5	26,6
1,2	0,5	25,3
1,6	0,5	22,8
2,0	0,5	20,3

Tabel C.11: Optimal lufttemperatur for forskellige aktivitetsniveauer og forskellige niveauer for beklædningens varmeledningsmodstand.

I det efterfølgende gennemgås grænserne for fire lokale påvirkninger fra *DS/EN/ISO 7730*, som er illustreret i tabel C.12 - C.14. De fire lokale påvirkninger er: Forskel i vertikal lufttemperatur mellem 0,1m og 1,1m over gulvet, tabel C.12, strålingstemperatursymmetri ved vinduer og andre kolde vertikale overflader, tabel C.13, strålingstemperatursymmetri pga. et varmt loft, tabel C.13 og overfladetemperatur af gulv, tabel C.14.

Kategori	Forskel i vertikal lufttemperatur
	[°C]
A	< 2
B	< 3
C	< 4

Tabel C.12: Krav til den vertikale lufttemperatur mellem 0,1m og 1,1m over gulvet [71].

Kategori	Strålingsassymetri [°C]	
	Varmt loft	Kold væg
A	< 5	< 10
B	< 5	< 10
C	< 7	< 13

Tabel C.13: Krav til strålingsassymetri for et varmt loft eller en kold væg. [71]

Kategori	Temperaturspænd for overfladetemperatur på gulv [°C]
A	< 19 – 29
B	< 19 – 29
C	< 17 – 31

Tabel C.14: Krav til gulvets overfladetemperatur [71].

C.3.3 Bygningsreglementet

I nærværende afsnit gennemgås kravene til atmosfærisk indeklima, træk og støj fra *Bygningsreglementet* [53].

C.3.3.1 Atmosfærisk indeklima

Det atmosfæriske indeklima påvirkes bl.a. af hvor mange mennesker der er til stede i rummet, udstyret i rummet og hvilke materialer og møbler der er anvendt. Luftkvaliteten bedømmes oftest ud fra CO_2 -koncentrationen, idet det er en indikator for hvor forurennet luften er. Dog er CO_2 -koncentrationen ikke et udtryk for luftkvaliteten, da der ved denne vurdering ikke tages højde for de øvrige parametre der har indflydelse herpå bl.a. lugt.

Der er i *Bygningsreglementet* ingen krav til CO_2 -koncentrationen i boliger, men da det i dette projekt ønskes at have mulighed for at vurdere indeklimaet, anvendes grænserne for CO_2 -koncentrationen fra *DS/CR/CEN 1752* [70], der er vist i tabel C.15.

Kategori	CO_2 -koncentration* [ppm]
A	460
B	660
C	1190

Tabel C.15: Grænser for CO_2 -koncentrationen for hhv. kategori A, B og C. * CO_2 -koncentrationen fratrasket koncentrationen udendørs [70].

C.3.3.2 Træk

Træk er en diskomfort, der kan opleves i rum hvor der luftes ud, som nedfald fra høje glasfacader eller fra ventilationsanlæg. Ved udluftningen er løsningen på problemet blot at lukke vinduet, men ved de to øvrige problemer er det en problemstilling der skal lægges fokus på i design processen af en bolig. I *Bygningsreglementet* [53] står der følgende om træk:

„Ved tilførsel af luft og ved fjernelse af luft skal det i rum, hvor personer opholder sig længere tid sikres, at der i opvarmningsperioden ikke optræder træk i opholdszonen.“

Grænsen for hvornår der opstår træk afhænger af beboernes aktivitetsniveau, lufttemperaturen og luftens turbulensniveau. Men for at undgå træk bør lufthastigheden i opholdszonen i lokaler med stillesiddende aktivitet ikke overstige $0,15\text{ m/s}$. [53]

C.3.3.3 Støj

Ud over kravene til selve anlægget og ønskede luftmængder, angiver *Bygningsreglementet* [53] også grænser for det akustiske indeklima. Heri fastslår den at funktionskravet for boliger anses for at være opfyldt, når de udføres som klasse C i den danske standard *DS 490* [72]. Grænsen der er angivet i standarden *DS 490* er på 30 dB (A-vægtet).

C.4 Rumopvarmning

Igennem litteraturstudiet er der opnået kendskab til flere forskellige forhold, der påvirkes af beboerne, der har indflydelse på energiforbruget i boliger. Disse forhold er: udluftningen, infiltrationen og brugen af solafskærmning. Flere af de oplistede forhold er enten ikke behandlet eller kun behandlet i et begrænset omfang i forhold til boliger i litteraturen. De er dog behandlet i forbindelse med analyser i kontorbyggerier, hvilket gennemgås i de efterfølgende afsnit. Ligeledes er viden fra indeklimastandarder og andre indeklima analyser sammenfattet af *Peeters et al.* [57] med det mål, at opstille en mulig metode for beregning af en lufttemperatur hvor beboerne er i komfort, se afsnit C.4.2. Derudover er brugen af rumopvarmning undersøgt af *Nicol et al.* og *Rune Vinther Andersen* [2], for henholdsvis kontorbyggeri og boliger, resultaterne fra denne undersøgelse gennemgås kort i det efterfølgende afsnit.

C.4.1 Anvendelse og styring af rumopvarmningen

I nærværende afsnit gennemgås tre forskellige analyser, der fokuserer på: brugen af rumopvarmning i forhold til udetemperaturen, klarlægning af brugeradfærd i forbindelse med justering af setpunktet for radiatortermostater og analysering af indstillingerne for programmerbare rumtermostater i fire forskellige familier.

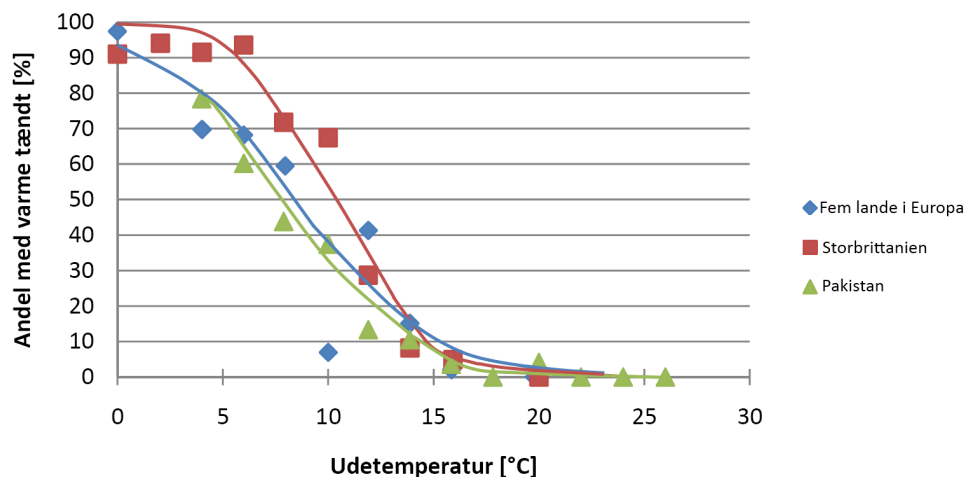
Nicol et al. [49] har på baggrund af en spørgeundersøgelse undersøgt brugen af rumopvarmningen, i forhold til udetemperaturen, i kontorbyggerier i syv forskellige lande. *Nicol et al.* har udviklet nogle algoritmer, der sammenholder sandsynligheden for, at brugerne af en bygning anvender hver af de førnævnte systemer, som funktion af inde- eller udetemperaturen. Algoritmerne er udviklet på baggrund af en database, der blandt andet indeholder spørgeundersøgelser udført i henholdsvis:

- Storbritannien - i alt 5000 besvarelser [49].
- Pakistan - i alt 7000 besvarelser [49].
- Sverige, Storbritannien, Frankrig, Portugal og Grækenland - i alt 4655 besvarelser [49].

For hver af de ovennævnte undersøgelser er der opstillet algoritmer, der er analyseret med henblik på at klarlægge en eventuel korrelation mellem enten luft- eller udetemperaturen

og hver af de førnævnte systemer. I det efterfølgende beskrives de fundne resultater for opvarmning.

Analyserne viser en høj korrelation mellem anvendelsen af varmesystemer og udetemperaturen, hvilket går igen i alle tre undersøgelser. Det er værd at bemærke, at der, trods forskel i landenes beliggenhed, er en tendens til at opvarmningen aktiveres ved ca. samme udetemperatur, se figur C.2. Det ses samtidig, at der sker en betydelig stigning i anvendelsen af opvarmning når udetemperaturen er omkring $15\text{ }^{\circ}\text{C}$, og når udetemperaturen når ned på $5\text{ }^{\circ}\text{C}$ har ca. 70 % af de fem lande i Europa og Pakistan tændt for varmen hvorimod ca. 95 % anvender varmen i Storbritannien.



Figur C.2: Sammenhæng mellem udetemperaturen og brugen af varmesystemer [49].

Rune Vinther Andersen [2] har foretaget målinger i 15 danske boliger fra januar til august 2008, med henblik på at klarlægge brugeradfærd blandt andet i forbindelse med justering af setpunktet for radiatortermostater. I denne forbindelse er der dog kun målinger for setpunktet for 13 boliger, hvor der i hver bolig er monteret måleenheder på to forskellige radiatorer. Målingerne anvendes til at opstille modeller, til anvendelse i simuleringsprogrammer, ved brug af statistiske analyser.

I første omgang endte den statistiske analyse ud med en avanceret model for estimering af setpunktet for radiatortermostaterne, eftersom resultatet var et formeludtryk bestående af 12 led med forskellige variable og yderligere 29 led med kombinationer af de variable parametre. Formeludtrykket simplificeres, således at der ender med at indgå fem forskellige parametre, hvor der for to af disse skelnes mellem hvilken tid på døgnet setpunktet skal estimeres. Der skelnes dog også mellem hverdage og weekender, således der er forskellige formeludtryk afhængig af ugedagen. De tre vigtigste parametre i forbindelse med estimeringen af setpunktet for radiatortermostater blev i projektet vurderet til at være udetemperaturen, den relative luftfugtighed udenfor samt vindhastigheden.

Sandsynligheden for, at radiatortermostater er åbne, kan ifølge projektet estimeres ud fra formel C.2 ved anvendelse af koefficienterne, der er angivet i tabel C.16.

$$\log\left(\frac{p}{1-p}\right) = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + b_4 \cdot x_4 + b_5 \cdot x_5 \quad (\text{C.2})$$

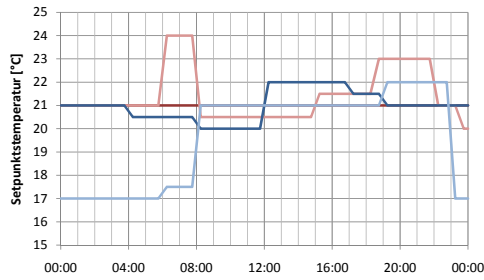
Hvor:

p	Sandsynlighed for at radiatortermostaterne er åbne [-]
a	Konstant [-]
$b_1, \dots, b_5$	Koefficienter angivet i tabel C.16
$x_1, \dots, x_5$	Variable. Der anvendes samme indeksnummerering, som ved koefficienterne

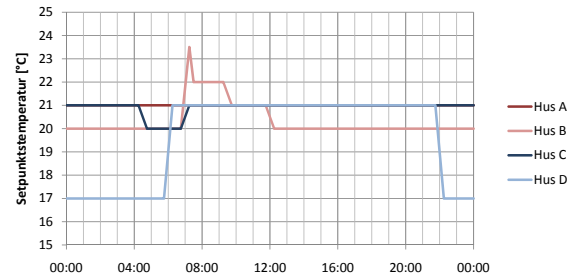
	Konstant	CO ₂ -konc.	Udetemp.	Vindhast.	Luftfugtighed	Solstråling
	a [-]	b_1 [ppm]	b_2 [°C]	b_3 [m/s]	b_4 [%RF]	b_5 [W/m ²]
Hverdag						
Morgen	23,76	0,00048	-0,30	-0,08	-0,063	-0,0006
Dag	24,82	0,00048	-0,32	-0,20	-0,063	-0,0006
Aften	23,99	0,00048	-0,33	-0,06	-0,063	-0,0006
Nat	23,29	0,00048	-0,31	0,02	-0,063	-0,0006
Weekend						
Morgen	23,80	0,00048	-0,30	-0,01	-0,063	-0,0006
Dag	24,86	0,00048	-0,32	-0,13	-0,063	-0,0006
Aften	24,02	0,00048	-0,33	0,01	-0,063	-0,0006
Nat	23,32	0,00048	-0,31	0,09	-0,063	-0,0006

Tabel C.16: Koefficienter til estimering af sandsynligheden for, at radiatortermostater er åbne [2].

Analysen af hvorledes fire husstande indstiller deres programmerbare termostater er i perioden fra september 2008 til november 2009 udført af *Gill et al.* [23] i forbindelse med et engelsk måleprojekt på boliger. Husstandene havde mulighed for at programmere termostaterne, således setpunktstemperaturen kunne varieres gennem døgnet ved forskellige indstillinger på hverdage og i weekender, hvor deres indstilling kan ses af figur C.3 og C.4. Af figurerne kan det ses, at de fire husstande har programmeret termostaterne vidt forskelligt, men fælles for dem er, at de primært svinger mellem 20 °C til 22 °C i dagtimerne.



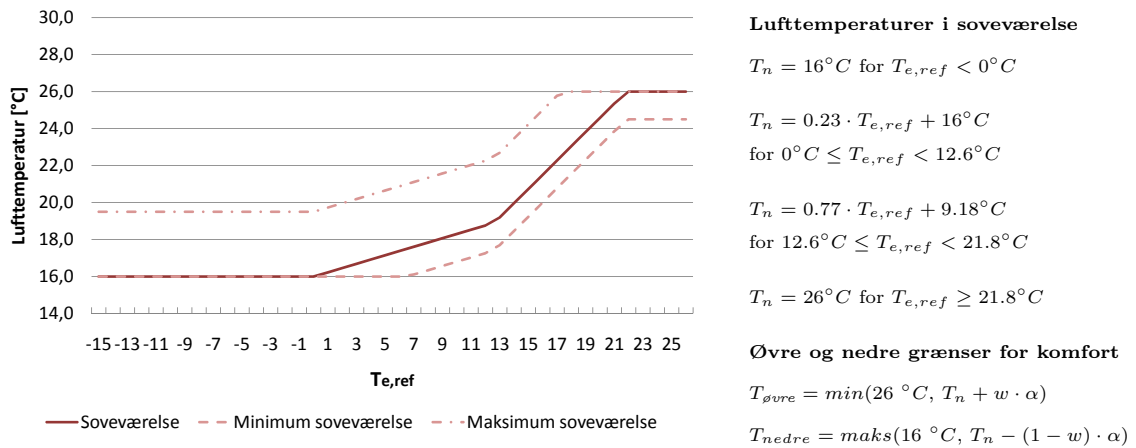
Figur C.3: Programmerede setpunkttemperaturer for fire husstande på hverdage [23].



Figur C.4: Programmerede setpunkttemperaturer for fire husstande i weekender [23].

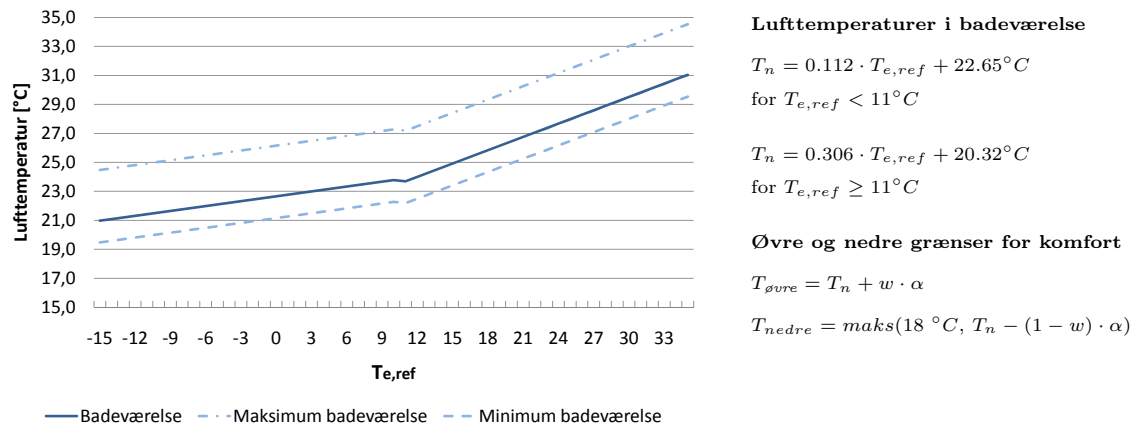
C.4.2 Praktisk anvendelse af vejledning fra normer

Peeters *et al.* [57] har, på baggrund af analyser, lovgivning for indeklima samt vejledning af temperaturer fra sundhedsorganisationer, fremlagt et forslag til en mulig metode til beregning af lufttemperaturen i tre forskellige rumtyper i boligen⁴. Der beregnes en lufttemperatur hvor beboerne føler sig i termisk komfort med en dertilhørende minimums- og maksimumsgrænse. Disse formler er præsenteret visuelt i figurerne C.5 - C.7.

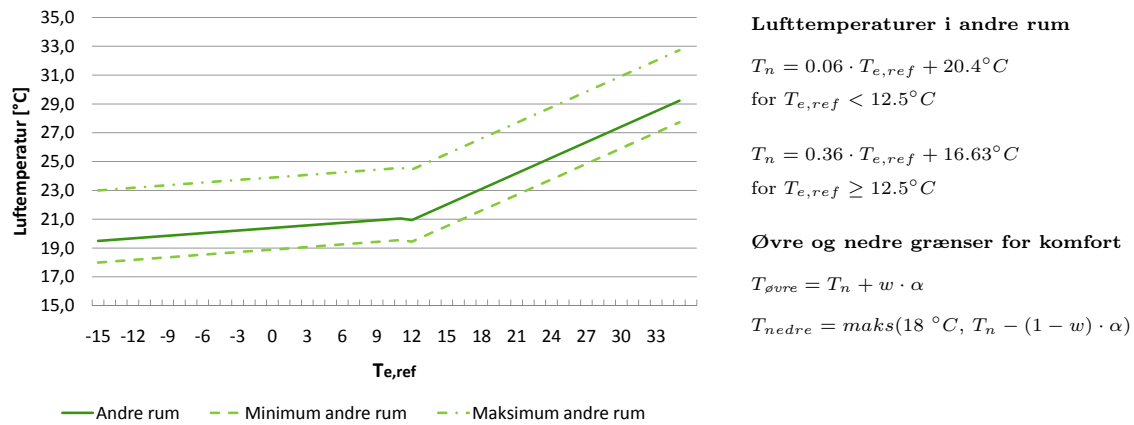


Figur C.5: Lufttemperaturspænd i soveværelse beregnet på baggrund af de angivne formler [57].

⁴Rumtyperne omfatter badeværelse, soveværelse og andre rum (stue, køkken og kontorer etc.)



Figur C.6: Lufttemperaturspænd i badeværelse beregnet på baggrund af de angivne formler [57].



Figur C.7: Lufttemperaturspænd i andre rum beregnet på baggrund af de angivne formler [57].

Hvor:

T_n	Lufttemperatur [$^{\circ}C$]
$T_{e,ref}$	Referencetemperatur bestemt af udetemperaturen [$^{\circ}C$]
$T_{\text{øvre}}$	Øvre grænse for komfort [$^{\circ}C$]
T_{nedre}	Nedre grænse for komfort [$^{\circ}C$]
w	Konstant; 5, 0 ved 10 % utilfredse [-]
α	Konstant; 0, 7 ved 10 % utilfredse [-]

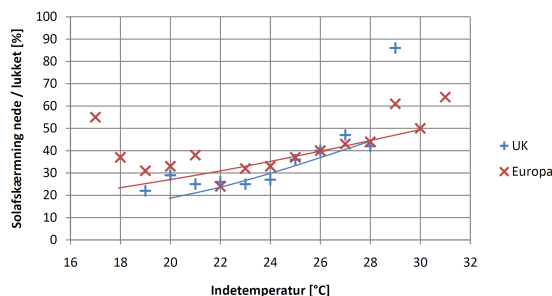
De øvre og nedre grænser for komfort, der er angivet på figur C.5 - C.7, er fremkommet gennem flere forskellige kilder. Grænserne, der er anvendt for badeværelser, er fremkommet gennem et litteraturstudie og forsøg, udført af bl.a. *Lammers et al.* [39] og *Tochihara et al.* [80]. Den nedre grænse for komfort, der er anvendt for soveværelser, er på 16 °C og stammer fra verdenssundhedsorganisations (WHO). Årsagen til valget af denne grænse for temperatur

er, at menneskets resistens imod luftvejsinfektioner falder drastisk ved temperaturer under $16\text{ }^{\circ}\text{C}$ [57]. For den anvendte maksimumsgrænse for soveværelset på $26\text{ }^{\circ}\text{C}$ henvises der til en grænse fremsat af CIBSE⁵.

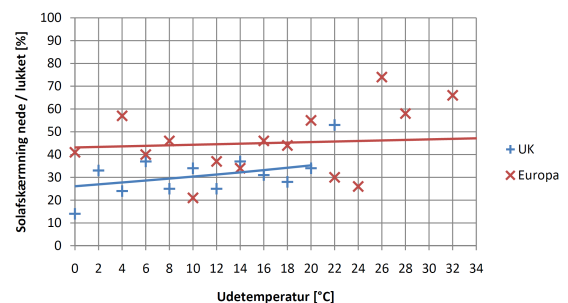
C.4.3 Solafskærmning

I dette afsnit omtales nogle af de studier, hvor brugen af solafskærmning på bygninger er undersøgt, enten i henhold til vaner for hvorledes solafskærmningen anvendes eller for undersøgelser, der kæder brugen af solafskærmning sammen med andre forhold som for eksempel udetemperaturen. Da der i dette litteraturstudie ikke er fundet målinger eller undersøgelser om brugen af solafskærmning i boliger, er det valgt at se nærmere på brugen af solafskærmning i kontorer. Det vil efterfølgende blive diskuteret i hvilket omfang den opnåede viden fra den fundne litteratur er anvendelig i forbindelse med boliger.

Nicol et al. [49] har i projektet, der er beskrevet i afsnit 3.2, undersøgt brugen af solafskærmning eller gardiner på kontorer. I analyserne er det undersøgt, hvorvidt der er en korrelation mellem enten inde- eller udetemperatur og sandsynligheden for at solafskærmning eller gardiner er i brug. Resultaterne fra dette studie viser, at der er en begrænset korrelation mellem anvendelsen af solafskærmning eller gardiner og indetemperaturen, men at korrelationen i forhold til udetemperaturen er forsvindende lille. En optegning af brugen af solafskærmning eller gardiner som funktion af temperaturerne kan ses på henholdsvis figur C.8 og C.9. Det vurderes af *Nicol et al.*, at der vil være en større korrelation mellem illuminans og anvendelsen af solafskærmning eller gardiner, eftersom hovedformålet med disse er at afskærme for den direkte solstråling. Dette kan dog ikke bekræftes i denne analyse, da der ikke er foretaget måling af illuminans i nogle af tilfældene.



Figur C.8: Brugen af solafskærmning sammenholdt med lufttemperaturen [49].

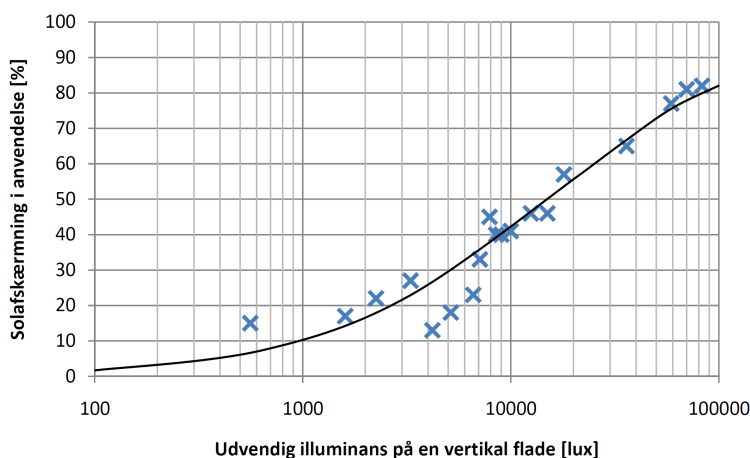


Figur C.9: Brugen af solafskærmning sammenholdt med udetemperaturen [49].

Med hensyn til brugen af solafskærmning sammenholdt med illuminansen har *Sutter et al.* [76] lavet et studie herom. Der er foretaget målinger hver 15. minut på otte kontorer over en periode på 30 uger. Sandsynligheden for at solafskærmningen er i anvendelse er afbilledet på figur C.10, hvor sandsynligheden er sammenholdt med illuminansen på en udvendig, vertikal flade. Punkterne på grafen viser kun forhold hvor solafskærmningen enten har været helt oppe eller helt nede. Sammenholdes resultaterne fra *Sutter et al.* med resultaterne fra *Nicol*

⁵The Chartered Institution of Building Services Engineers

et al., kan det ses, at når sandsynligheden for brug af solafskærmning sammenholdes med illuminansen, kan et større spænd forklares end når sandsynligheden sammenholdes med luft- eller udetemperaturen.



Figur C.10: Brugen af solafskærmning sammenholdt med illuminansen [76].

Frédéric Haldi [28] har udført målinger på solafskærmningernes position i 12 kontorer i perioden fra 1. januar 2003 til 8. september 2009. Solafskærmningen i hvert kontor er delt i en øvre og en nedre del, der kan styres separat efter brugernes ønske.

Der er efterfølgende foretaget en statistisk analyse af målingerne, for at undersøge nogle parametres signifikans i forbindelse med brug af solafskærmning. Ifølge den statistiske analyse, blev der ikke fundet nogen signifikant brugerrelateret forskel i adfærd mellem kontorerne. Analysen viser endvidere, at den øverste solafskærmning generelt anvendes og justeres oftere end den nederste solafskærmning. Den øverste del af solafskærmningen var helt oppe i 56,4 % af brugstiden og helt nede i 21,7 % af tiden. Til sammenligning var den nederste solafskærmning helt oppe i 67,1 % af brugstiden og helt nede i 5,2 % af tiden. Det kunne endvidere bemærkes, at solafskærmningens position oftest blev justeret ved ankomst til kontorerne, men der var ikke nogen signifikant stigning i anvendelsen af solafskærmning, når kontorerne forlades. Af denne årsag udviklede *Frédéric Haldi* nogle formeludtryk, der udtrykker sandsynligheden for, at der sker et skift i solafskærmningens position, hvor der skelnes mellem den øvre og nedre solafskærmning samt mellem tidsrummet ved ankomst til kontorerne og ved efterfølgende ophold i disse. Endvidere blev der opstillet udtryk, der estimerer sandsynligheden for, at solafskærmningen køres henholdsvis helt op eller helt ned, såfremt de øvrige formler angiver, at der sker et skift i solafskærmningens position. Dette betyder, at der blev opstillet seks formeludtryk for henholdsvis den øvre og den nedre del af solafskærmningen.

Disse formeludtryk tager form, som vist i formel C.3, hvor kun de variable, der har signifikant indflydelse på sandsynligheden inddrages. Dette er årsagen til, at nogle af koefficienterne i tabel C.17 er nul, hvilket betyder, at den tilhørende variabel ikke er signifikant i den pågældende formel.

$$\log \left(\frac{p}{1-p} \right) = a + b_{E_{in}} \cdot E_{in} + b_{E_{gl,hor}} \cdot E_{gl,hor} + b_{B_L} \cdot B_L + b_{B_U} \cdot B_U \quad (C.3)$$

Hvor:

p	Sandsynlighed [-]
a	Konstant [-]
$b_{E_{in}}$	Koefficient angivet i tabel C.17 [lx^{-1}]
E_{in}	Horisontal illuminans indenfor [lx]
$b_{E_{gl,hor}}$	Koefficient angivet i tabel C.17 [lx^{-1}]
$E_{gl,hor}$	Global, horisontal illuminans udenfor [lx]
b_{B_L}	Koefficient angivet i tabel C.17 [-]
B_L	Solafskærmningens position for det nederste vindue i forrige tidsskridt, hvor 0 er helt nede og 1 er helt oppe [-]
b_{B_U}	Koefficient angivet i tabel C.17 [-]
B_U	Solafskærmningens position for det øverste vindue i forrige tidsskridt, hvor 0 er helt nede og 1 er helt oppe [-]

I tabel C.17 er koefficienternes størrelse til den ovenstående formel angivet, ud fra hvilken sandsynlighed, der skal beregnes. $P_{lower,arr}$ og $P_{raise,arr}$ beskriver sandsynligheden for, at solafskærmningen køres henholdsvis ned eller op i tidsrummet umiddelbart efter ankomst til kontoret. $P_{lower,int}$ og $P_{raise,int}$ beregner sandsynligheden for, at solafskærmningen køres henholdsvis ned eller op ved efterfølgende ophold i kontoret.

Sandsynlighed	a	$b_{E_{in}}$	$b_{E_{gl,hor}}$	b_{B_L}	b_{B_U}
	[-]	[-]	[-]	[-]	[-]
Øverste vindue					
$P_{lower,arr}$	-7,290	$9,480 \cdot 10^{-4}$	$6,660 \cdot 10^{-6}$	0	2,180
$P_{raise,arr}$	-1,699	$-5,240 \cdot 10^{-4}$	$-21,800 \cdot 10^{-6}$	0	-3,916
$P_{lower,int}$	-8,211	$8,340 \cdot 10^{-4}$	$5,690 \cdot 10^{-6}$	0	1,533
$P_{raise,int}$	-3,629	$-2,900 \cdot 10^{-4}$	$-16,860 \cdot 10^{-6}$	0	-3,365
$P_{full lower}$	-0,435	0	$2,500 \cdot 10^{-6}$	0	0,150
$P_{full raise}$	1,543	0	$-2,120 \cdot 10^{-5}$	0	-0,560
Nederste vindue					
$P_{lower,arr}$	-7,410	$10,350 \cdot 10^{-4}$	0	2,170	0
$P_{raise,arr}$	-1,520	$-6,540 \cdot 10^{-4}$	0	-3,139	0
$P_{lower,int}$	-8,013	$8,410 \cdot 10^{-4}$	0	1,270	0
$P_{raise,int}$	-3,625	$-2,760 \cdot 10^{-4}$	0	-2,683	0
$P_{full lower}$	-0,270	0	$0,910 \cdot 10^{-6}$	-2,230	0
$P_{full raise}$	0,435	0	$-2,310 \cdot 10^{-5}$	1,950	0

Tabel C.17: Koefficienter til estimering af sandsynligheden justring af solafskærmning [28].

Såfremt solafskærmningen estimeres til ikke at køre helt ned eller helt op, kan ændringen af positionen beregnes ud fra formel C.4, hvis solafskærmningen køres ned. Køres solafskærmningen derimod op, kan dette estimeres ud fra en ligefordeling, hvor alle positioner har lige stor sandsynlighed.

$$f(\Delta B) = \frac{\alpha}{\lambda} \left(\frac{\Delta B}{\lambda} \right)^{\alpha-1} \exp \left(- \left(\frac{\Delta B}{\lambda} \right)^{\alpha} \right)$$

$$\lambda = \exp(-2,294 + 1,522 \cdot B_{init}) \quad (C.4)$$

Hvor:

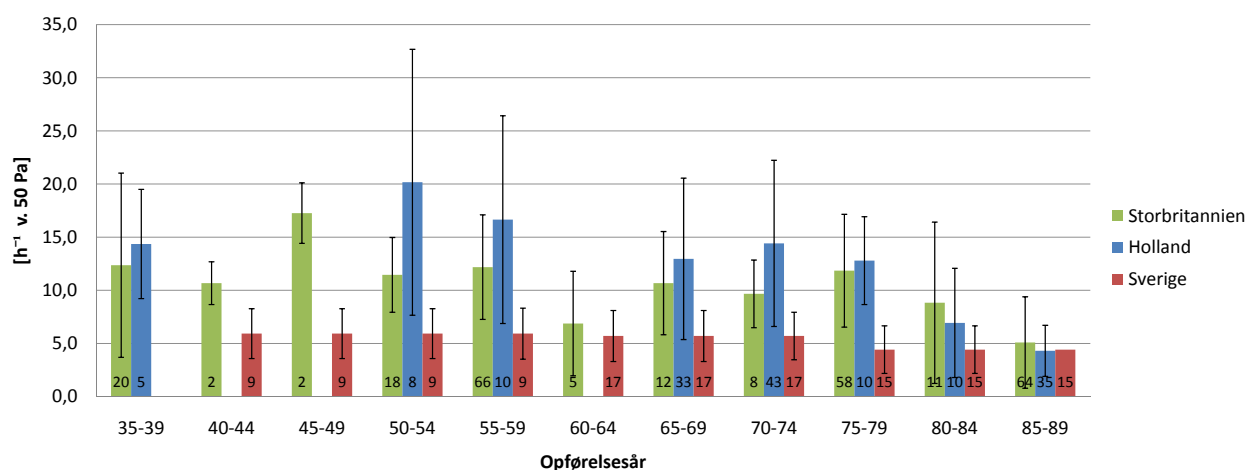
ΔB	Positionsændringen i forhold til B_{init} [-]
α	Konstant; 1,708 [-]
B_{init}	Solafskærmningens aktuelle position, hvor 0 er helt nede og 1 er helt oppe [-]

Frédéric Haldi har efterfølgende lavet en test af formeludtrykkene, der er beskrevet ovenfor. Testen viser, at 78 – 89 % af de handlinger, der kan estimeres med disse formeludtryk er korrekte, men det bør dog bemærkes, at andelen af brugstiden, hvor solafskærmningen er henholdsvis helt oppe eller helt nede underestimeres.

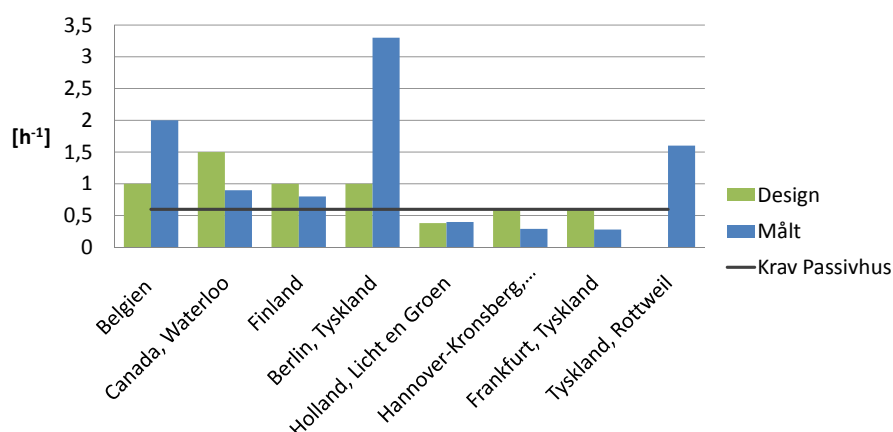
C.4.4 Infiltration

Netop boligernes tæthed har indflydelse på hvor meget der luftes ud i boligerne og igennem parameteranalysen i afsnit 2.2.1 blev det belyst at den også har betydning for energiforbruget i boliger.

Infiltrationen har bl.a. vist sig at spille en rolle i projektet fra *Kirsten Gram-Hanssen*, hvor det i undersøgelsen erfares, at boliger med større tæthed har behov for mere udluftning. Samme tendens er set i projektet fra *Soldaat et al.* [69]. På figur C.11 ses infiltrationens størrelse i nogle forskellige projekter, fra henholdsvis Holland, Storbritannien og Sverige, opført fra 1940 og frem til slutningen af 1980. Tallene, der er angivet i bunden af hver søjle på figur C.11, viser antallet af husstande, der er udført målinger på. Figur C.12 viser infiltrationens størrelse for nogle forsøgshuse, der er opført fra starten af 1990'erne og frem til 2002.



Figur C.11: Infiltrationens størrelse, ved en trykforskel på 50 Pa, i boliger opført fra 1940 – 1980 i Holland, Storbritannien og Sverige [55].



Figur C.12: Infiltrationens størrelse i forskellige byggeprojekter [78].

C.4.5 Udluftning

I afsnit 3.2.2.2 i hovedrapporten er der gennemgået hvilke typiske luftskifte, der er i boliger, samt hvilke vaner beboerne har med hensyn til udluftning af boliger. Ud over de gennemgåede emner i hovedrapporten er det observeret, at boligtypen har betydning for mængden af udluftning, samt hvordan dette gøres. Derudover introduceres der to modeller, der beskriver sandsynligheden for brugen af vinduer.

C.4.5.1 Betydning af boligtypen og vinduernes udformning

Ud over betydning af boligens tæthed for brugen af udluftning, gennemgået i afsnit 3.2.2.2 i hovedrapporten, fandt *Soldaat et al.* [69] frem til, at boligtypen også havde en betydning for

brugen af udluftning. Der var i den omtalte undersøgelse observeret en forskel i brugen af vinduer mellem parcelhuse og lejligheder. I lejligheder stod vinduerne ofte kun på klem og der blev luftet mest ud i stuen og køkkenet. I parcelhuse derimod stod vinduer under udluftning helt åbne og der blev luftet mere ud i soveværelset end i de øvrige rum. Derudover fandt *Soldaat et al.* frem til, at typen af vindue også har betydning for hvor meget der luftes ud i en bolig. Vinduer der var fikseret i bunden og åbner indad i toppen, har vist sig at være åbent dobbelt så meget som øvrige vinduestyper. Samtidig havde muligheden for justering af vinduets åbningsgrad også vist sig at være vigtig, i forhold til hvor ofte vinduet blev anvendt. Dette resultat for vinduets udformning lyder ganske fornuftigt i forhold til, at det er denne type vindue der skaber færrest trækgener, idet den køligere luft bliver opblandet med luften i rummet før den når beboeren.

C.4.5.2 Udeklimaets betydning for brugen af vinduer

Yderligere parametre der påvirker brugen af vinduer til udluftning er udeklimaet. Ifølge undersøgelsen af *Soldaat et al.* så er brugen af vinduer til udluftning lineært korreleret med udetemperaturen når den er mellem $-10-25^{\circ}\text{C}$ og omvendt korreleret med vindhastigheden. Når det regner og sneer falder mængden af udluftning. I 30 % af de undersøgte boliger lukkede beboerne vinduerne for at slippe for larm og støj fra trafik, den samme effekt har lugtgener fra trafik, industri og landbrug. [69]

Anvendelsen af udluftning afhænger som sagt af udeklimaet, og dette er beskrevet gennem nogle udviklede modeller af *Rune Vinther Andersen* og *Frédéric Haldi*, der gennemgås nedenfor.

I forbindelse med en undersøgelse af *Rune Vinther Andersen* [2], der tidligere er omtalt i afsnit 3.2, blev der udført målinger på 15 boliger af hvorvidt beboerne havde åbne eller lukkede vinduer. Gennem en statistisk analyse af målingerne, er der opstillet et formeludtryk, svarende til formel C.2 på side 22 i appendiks C.4.1, hvor der dog istedet indgår otte forskellige led. De parametre, der indgår til estimering af sandsynligheden for, at et vindue åbnes eller lukkes er: lufttemperaturen i rummet, udetemperaturen, den relative luftfugtighed både indenfor og udenfor, CO_2 -koncentrationen i rummet, vindhastigheden, solstrålingen samt en konstant. Formeludtrykket afhænger ydermere af hvornår på døgnet sandsynligheden estimeres, og om det er på en hverdag eller i en weekend. Det fremgår af analysen, at de tre vigtigste parametre i forbindelse med en estimering af sandsynligheden for at et vindue enten åbnes eller lukkes er: CO_2 -koncentrationen i rummet, udetemperaturen og lufttemperaturen i rummet.

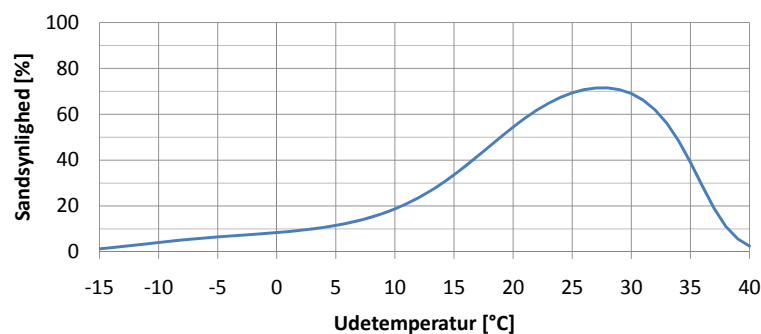
Frédéric Haldi [28] har undersøgt brugen af vinduer til udluftning, først på baggrund af et litteraturstudie og derefter på baggrund af målinger i en kontorbygning. I litteraturstudiet var der beskrevet nogle undersøgelser, hvor målet var, at undersøge udluftning ved brug af vinduer eller opstille formeludtryk, der kan beskrive dette. Resultaterne af studiet viste blandt andet, at de eksisterende modeller, der beskrev udluftning ved brug af vinduer, var baseret på målinger i kontorbygninger og at boliger sjældent blev omtalt i denne sammenhæng. *Frédéric Haldi* beskæftigede sig selv, som tidligere nævnt, med kontorbyggeri, men der fremhæves i det efterfølgende nogle af de resultater, der var opnået gennem den pågældende undersøgelse.

Litteraturstudiet af *Frédéric Haldi* viste, at der i flere tidligere undersøgelser var forsøgt at beskrive brugen af vinduer til udluftning som funktion af temperaturer eller vejrforhold. Det

blev vurderet i litteraturstudiet i det pågældende projekt, at udetemperaturen og lufttemperaturer i rum er væsentlige parametre, i forbindelse med denne beskrivelse, men det stod ikke klart, hvilken af disse parametre, der beskrev brugen af vinduer til udluftning bedst.

I undersøgelsen havde *Frédéric Haldi* anvendt målinger fra en kontorbygning, hvor der var målt på hvorvidt vinduerne i 14 mindre kontorer var åbne eller lukket i perioden fra december 2001 til september 2009. Undersøgelsen blev indledt med at andelen af åbne vinduer blev optegnet, som funktion henholdsvis lufttemperaturen i rummet, udetemperaturen, den relative luftfugtighed udenfor, vindhastigheden, vindretningen samt som funktion af regnmængden per tidsenhed.

Frédéric Haldi vurderede sammenhænge af åbne vinduer med både lufttemperaturen i rummet og udetemperaturen og fandt frem til at den bedste sammenhæng blev beskrevet ved hjælp af udetemperaturen. Denne sammenhæng fremgår af figur C.13.



Figur C.13: Sandsynlighed for et åbent vindue, som funktion af udetemperaturen [28].

Når andelen af åbne vinduer optegnes, som funktion af udetemperaturen, viste dette, at der skete en tydelige stigning op til omkring $26\text{ }^{\circ}\text{C}$, hvorefter andelen af åbne vinduer faldt. Foruden ovenstående var det undersøgt hvorvidt andelen af åbne vinduer, som funktion af henholdsvis differensen på luft- og komforttemperaturen⁶ samt differensen på ude- og komforttemperaturen, blev beskrevet bedre med disse parametre. Det viste sig dog, at disse temperaturdifferenser gav en dårligere beskrivelse af andelen af åbne vinduer, sammenholdt med lufttemperaturen i rummet eller udetemperaturen alene. Det kunne dog bemærkes, at andelen af åbne vinduer toppede, hvor komfort- og udetemperaturen havde samme størrelse. Ved at optegne andelen af åbne vinduer, som funktion af den relative luftfugtighed udenfor, kunne der ses en tendens til at andelen falder fra en relativ luftfugtighed omkring 20 % til 90 %, hvorefter der kunne ses en tendens til en stigning i andelen af åbne vinduer mellem 90 % og 95 %. Udenfor disse intervaller var der en variation i, hvorledes andelen af åbne vinduer blev fordelt.

Med hensyn til andelen af åbne vinduer, som funktion af vindhastigheden, kunne der ses en tendens, at andelen er stigende fra 0 m/s til omkring 2 m/s , hvorefter andelen var faldende op til omkring 12 m/s . Ved vindhastigheder over dette, varierede andelen af åbne vinduer betragteligt.

For andelen af åbne vinduer, optegnet som funktion af henholdsvis vindretningen og regnmængden per tidsenhed, kunne der ikke observeres nogen tydelige tendenser.

⁶Komforttemperaturen er en beregnet størrelse.

Efterfølgende blev der opstillet forskellige formeludtryk, der estimerede sandsynligheden for, at brugerne åbnede vinduerne, og disse formeludtryk blev testet mod målingerne i kontorbygningen. Det var forsøgt at opstille relativt simple formeludtryk, der kun havde én variabel, og mere komplekse udtryk, der havde flere variable og skulle have informationer om personernes opholdstid på kontorerne. Da de komplekse formeludtryk krævede en større mængde baggrundsdata, for at kunne estimere sandsynligheden for et åbent vindue, blev disse udtryk også mere relateret til den enkelte bygning samt brugerne af denne, hvilket dog også blev bemærket af *Frédéric Haldi*. Af denne årsag blev der, i dette litteraturstudie, lagt mere vægt på de relativt simple formeludtryk, der ikke var så tæt knyttet til de specifikke brugere af bygningen.

I projektet var der opstillet formeludtryk med både én, to og tre variable, men det viste sig, at det var formelen, der kun havde udetemperaturen som variabel, der gav den bedste estimering af andelen af åbne vinduer. Dette formeludtrykket ser ud som vist i formel C.5, hvor størrelsen af konstanten og koefficienterne er angivet under formelen. En visualisering af sandsynligheden for åbent vindue, angivet ved formel C.5, fremgår af figur C.13. I valideringen af dette formeludtryk beskrev *Frédéric Haldi*, at andelen af korrekte estimeringer for hvorvidt vinduer var åbne eller lukkede var 66 %, hvis beregningerne blev sammenholdt med målingerne.

$$\log \left(\frac{p}{1-p} \right) = a + b_1 \cdot t_{ude} + b_2 \cdot t_{ude}^2 + b_3 \cdot t_{ude}^3 + b_4 \cdot t_{ude}^4 \quad (C.5)$$

Hvor:

p	Sandsynlighed for åbent vindue [-]
a	Konstant; $-2,387$ [-]
b_1	Koefficient; $5,55 \cdot 10^{-2}$ [-]
b_2	Koefficient; $1,73 \cdot 10^{-3}$ [-]
b_3	Koefficient; $2,88 \cdot 10^{-4}$ [-]
b_4	Koefficient; $-9,64 \cdot 10^{-6}$ [-]
t_{ude}	Udetemperaturen [$^{\circ}C$]

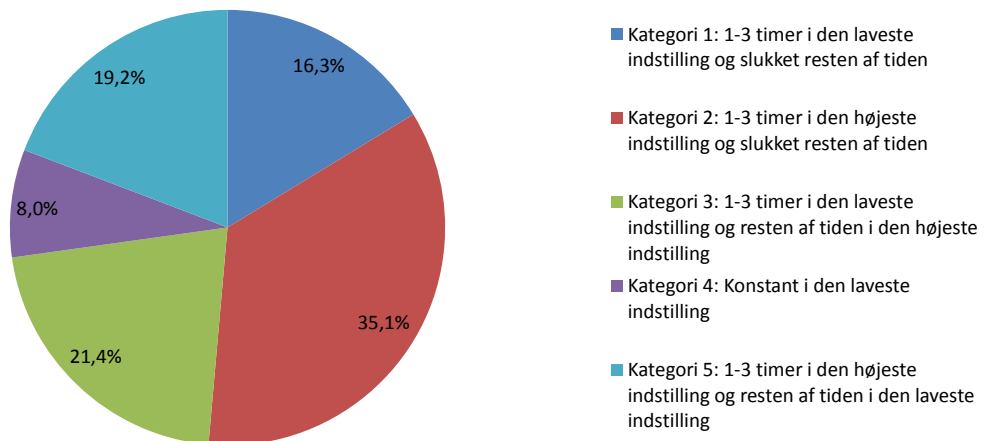
C.4.6 Mekanisk ventilation

Indenfor mekanisk ventilation kan der i nærværende projekt være tale om to forskellige typer af ventilation, mekanisk udsugning hvor udeluften kommer ind i bygningen via. udeluftsventiler og den forurenede luft fjernes ved hjælp af et udsugningsanlæg. Den anden løsning er ved brug af et anlæg med balanceret ventilation, hvor der både tilføres og fjernes luft gennem et ventilationsanlæg og varmegenvinding anvendes. Målet med dette afsnit er at beskrive brugsscenerier for hvordan beboere anvender deres ventilationsanlæg der er fundet i udførte undersøgelser samt beskrive hvilke udfordringer der er med hensyn til design af balanceret ventilation i boliger.

C.4.6.1 Brugernes betydning

Ligesom for brugen af udluftning har beboerne også betydning for hvordan ventilationssystemerne anvendes/fungerer i boligerne. Dette er gældende både med hensyn til vedligeholdelse, med henblik på styring af anlæggene og hvilke kendetegn der er for beboere, der foretrækker anderledes indstillinger for systemet.

Santin et al. [27] undersøgte i projektet, der er omtalt i afsnit 3.2.1.3 i hovedrapporten, brugen af mekanisk ventilation i boligerne. Det fremgår af undersøgelsen, at brugerne i en stor del af tiden enten slukker ventilationen eller sætter den på det laveste setpunkt, dog med undtagelse af de perioder hvor der laves mad eller tages bad. De anvendte indstillinger for styring af ventilationen ses i figur C.14. Heraf fremgår det, at 51,4 % anvender anlægget 1-3 timer om dagen i hhv. enten den laveste eller højeste indstilling og derefter slukker for anlægget. For 48,6 % betyder det, at deres anlæg kører hele tiden, men hvilken indstilling der anvendes varierer meget.



Figur C.14: Forskellige muligheder der er set anvendt for styring ventilationen [27].

Brugen af balanceret ventilation blev ligeledes undersøgt af *Soldaat et al.* [69] i en undersøgelse fra Holland udført i 2007, tidligere omtalt i afsnit 3.2.2.2 i hovedrapporten. Igennem litteraturstudiet fandt *Soldaat et al.* frem til, at fire faktorer spillede en stor rolle i adfærden omkring brugen af balanceret ventilation⁷

En af de karakteristika for beboerne der har størst betydning for anvendelsen af ventilationen er beboernes alder samt tilstedeværelsen af børn. I denne undersøgelse har det vist sig, at jo ældre beboerne er des mindre ventilation/udluftning anvendes der.

⁷Disse fire faktorer er: karakteristika for brugerne, karakteristika for boligen (se afsnit C.4.6.2 i nærværende appendiks), karakteristika for klimaet (blev gennemgået i afsnit C.4.5.2 i nærværende appendiks.) og karakteristika for ventilationssystemet (se afsnit C.4.6.2) i nærværende appendiks.

Vigtigheden af information til beboerne omkring brugen af et nyt anlæg, primært balanceret ventilation, har også vist sig at have en stor betydning. Når et ventilationsanlæg er installeret i en bolig er det vigtigt at præsentere anlægget og den optimale brug heraf til beboerne. Manglen på dette har betydning for brugen af anlægget, driften og derigennem også energiforbruget til anlægget samt tilfredsstillelsen ved at bo i en bolig med balanceret ventilation. Mængden af information til beboerne kan have stor betydning for hvordan ventilationsanlægget anvendes. I to forskellige projekter fandt *Soldaat et al.* frem til forskellig brug af ventilationen, denne forskel skyldes netop mængden af information, der blev videregivet til beboerne. Der blev i det ene projekt ikke erfaret nogen forskel i mængden af hvor meget beboerne luftede ud i boligerne, på trods af at der blev anvendt hhv. naturlig ventilation og balanceret ventilation. Derimod var der i et andet projekt forskel, boligerne med balanceret ventilation luftede mindre ud end beboerne i boligerne med naturlig ventilation. [69]

For at undersøge energiforbruget til ventilation/udluftning af boligen opstillede *Soldaat et al.* fem brugsscenarier og beregnede det primære energiforbrug for hver af disse scenarier. De fem scenarier repræsenterer fem forskellige indstillinger/anvendelser af anlægget og bygger på brugererfaringer fra interviews med beboere. Hvert scenarie gennemregnes for tre muligheder: lav infiltration, høj infiltration og lav infiltration hvor ventilatoren er slukket (ingen varmegenvinding). Resultaterne viser klart at scenarier med nedsat kapacitet i luftmængde er billigere i drift end et anlæg, der kører optimalt. Ved nedsatte luftmængder er der dog en stor risiko for at beboerne begynder at åbne vinduerne og dermed anvende naturlig ventilation også og derefter er scenarier med nedsat luftmængde ikke bedre energimæssigt end et optimalt kørende anlæg. Undersøgelsen viser, at det er vigtigt at involvere og præsentere beboerne for deres ventilationsanlæg i boligen og hvilken nødvendig vedligehold der er for at opnå det mest optimale resultat for dem og for anlægget. Hvis dette ikke gøres, er der risiko for at de energibesparelser, der var beregnet ikke opnås og at det i visse situationer kan blive dyrere med et ventilationsanlæg og indeklimaet opfattes som forringet.

C.4.6.2 Design udfordringer

Karakteristika for ventilationssystemet er vedligehold og styring. Igennem interviews erfarede *Soldaat et al.*, at beboerne i boliger med ventilation ikke var informeret nok og at dette var skyld i manglende vedligehold, forkert brug og øget energiforbrug. Vedligehold i form af rengøring/udskiftning af filtre blev kun gjort i 50 % af boligerne, hvilket betød at kapaciteten af anlægget blev reduceret med ca. 10 % hvert år og at anlægget efter 5 år kørte med en nedsat kapacitet på 65 %. Flere af beboerne ytrede, at de ikke følte, at anlægget bidrog med nok frisk luft og at indblæsningsluften var for tør, og som reaktion på dette åbnede de et eller flere vinduer.[69] Årsagen til at beboerne ikke følte at anlægget bidrog med store luftmængder nok, kunne i dette projekt skyldes designet af anlæggene. Den maksimale luftmængde anlægget kunne ventilere med, er den minimale luftmængde der stilles krav til fra bygningsreglementet der var gældende i Holland. Derved tages der hverken hensyn til beboernes mulighed for at øge luftmængden, hvis dette skulle ønskes, der tages heller ikke hånd om den faldende kapacitet ved filteret og armaturernes brug eller et behov for i visse perioder, evt. på grund af gæster, at øge ventilationsmængden.

Af karakteristika for boligen kan nævnes indretningen af boligen. Ved brug af balanceret ventilation havde det stor betydning, hvor indblæsningsarmaturerne placeres henne i rummene. Interviewundersøgelsen viste blandt andet, at hvis beboerne følte sig generet af træk eller

støj på grund af et armatur, ar løsningen for mange at lukke armaturet, reducere luftmængden eller som for nogle helt at afbryde anlægget. Resultaterne fra interviews med beboerne viste bl.a. at 17 % ventilerer mindre end de gerne ville på grund af støj. 28 % er *altid* eller *ofte* generet af støj og 37 % er *nogle gange* generet af støj.[69]

Derfor er det vigtigt, at designet af anlægget og placeringen af armaturerne gennemtænkes, før boligen opføres, for at undgå statistikker som ovenstående hvor anlægget er generende frem for rart for beboerne.

C.4.7 Diskussion

Igennem de ovenstående afsnit, er der opnået et større kendskab til brugen af rumopvarmning i boliger, gennem et studie af beboernes adfærd i forbindelse med regulering af termostater og indstilling af programmerbare termostater. Der opnås gennem to modeller, indsigt i hvorledes solafskærmning reguleres i kontorer. Derudover undersøges der forhold omkring udluftning og anvendelsen af mekanisk ventilation i boliger.

Igennem litteraturstudiet i afsnit 3.2.3 i hovedrapporten indikerede flere undersøgelser meget klart, at beboernes holdninger og vaner har stor betydning for størrelsen af deres energiforbrug til rumopvarmning. De gennemgåede 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. På trods af denne kompleksitet, er der gennem litteraturstudiet i nærværende appendiks opnået viden om andre projekter, der giver en indikation på, at der i nogle husstande særligt er tre tidsrum hvori setpunktstemperaturen for rumopvarmning justeres. Det drejer sig om morgentimerne omkring klokken 05:00 til 07:00 hvor temperaturen typisk opjusteres, for nogle personers vedkommende blot i nogle timer. Derudover kan der ses en tendens til en opjustering af temperaturen omkring klokken 15:00 til 19:00, og i flere tilfælde reduceres setpunktstemperaturen omkring klokken 22:00 til 23:00. I disse projekter er der dog også eksempler på personer, der ikke justere radiatortermostaterne, og personer, der på grund af studie eller deres tilknytning til arbejdsmarkedet, har brug for opvarmning i andre eller længere perioder i dagtimerne end de tidligere nævnte.

Der er igennem nærværende litteraturstudie opnået kendskab til flere projekter, hvori det er forsøgt, at opstille nogle matematiske modeller, der kan estimere, hvornår der anvendes rumopvarmning. Både Nicol *et al.* [49] og Rune Vinther Andersen [2] har påvist, at udetemperaturen er et vigtigt forhold for anvendelsesperioden. Det kan dog også forventes, at udetemperaturen kan indikere, hvornår der anvendes rumopvarmning, eftersom bygningers varmetab er direkte afhængigt af dette forhold. Nicol *et al.* har fremlagt nogle oplysninger, der indikerer, at der sker en stor stigning i andelen af bygninger, når udetemperaturen når under 15 °C, og når udetemperaturen når ned på omkring 5 °C er det kun et fåtal af bygningerne, hvori der ikke anvendes rumopvarmning.

Ud over udetemperaturen har Rune Vinther Andersen også inddraget andre udeforhold i sin matematiske model. Det drejer sig blandt andet om vindhastighed og udeluftens relative fugtighed, hvilket kan have betydning for bygningers varmetab, men på hvilken temperaturpåvirkning personerne føler udenfor. At Rune Vinther Andersen også har fundet CO₂-koncentrationen indenfor som en beskrivende faktor, hvor en stigning i CO₂-koncentrationen vil medføre en øget sandsynlighed for, at rumopvarmningen anvendes, kan indikere, at personernes tilstedeværelse har indflydelse på brugen af opvarmning. Dette er dog også at forvente,

eftersom tilstedeværelsen af personer i bygning gør, at lufttemperaturen, og derved brugen af rumopvarmning, bliver et grundlæggende element for hvilken komfort personerne føler.

Undersøgelser viser, at det ikke kun er de valgte lufttemperaturer og beboernes adfærd i forbindelse med regulering af denne, der har betydning for boligens energiforbrug til rumopvarmning. I de boliger, hvor der er monteret solafskærmning, har brugen af denne også en væsentlig betydning for energiforbruget, jf. figur 2.17 i hovedrapporten. Igennem litteraturstudiet er det dog bemærket, at hovedparten af undersøgelser omkring brugen af solafskærmning er foretaget i forbindelse med kontorer. Der er dog en sandsynlighed for, at en lignende adfærd kan forventes i boliger, fordi solafskærmningen, uanset om den er monteret på et kontorbyggeri eller på en bolig, typisk anvendes til to formål; at skærme af for den direkte solstråling, for at undgå generende blænding af personerne og for at reducere varmetilskuddet fra solstrålingen.

Undersøgelserne indikerer, at beboerne i nogle boliger overstyrer automatikken, når de føler sig generet af solafskærmningen. Dette forhold bliver også bemærket af *Frédéric Haldi* [28], hvor han observerede en tendens til, at solafskærmningen oftere blev justeret når personer ankommer til et rum. Det kan tyde på, at når personer oplever et betydeligt skift i omgivelserne, er der større sandsynlighed for, at personerne aktivt gør noget for at ændre de aktuelle forhold. Med dette menes eksempelvis, at når en person ankommer til sit kontor eller sin bolig, hvor solafskærmningen måske er kørt helt ned, kan de få en følelse af, at rummet er "lukket af". Dette forhold har en beboer også givet udtryk for, med en udtalelse om, at hvis huset virker åbent, giver det en fornemmelse af, at der er liv i huset, og personen føler ikke, at den er isoleret i et bestemt rum. Personens udtalelse kan findes, som det første citat i afsnit 3.2.1.3. Hvis personen får en følelse af, at være isoleret i et rum, kan den eksempelvis justere solafskærmningen, således der opnås frit udsyn til omgivelserne, hvilket kan give dele af forklaringen på *Frédéric Haldi*'s observation af, at solafskærmningen oftest justeres ved personers ankomst til rummet.

Der er opnået kendskab til undersøgelser, hvori der opstilles matematiske modeller, der kan estimere sandsynligheden for, at solafskærmningen på en bygning er i brug. I en af undersøgelserne vurderes at lufttemperaturen har en begrænset forklaringsevne i henhold til brugen af solafskærmning, men det vurderes gennem litteraturstudiet, at illuminansen har en større forklaringsevne i den forbindelse. Denne faktor benyttes da også i flere matematiske modeller, blandt andet et formeludtryk udviklet af *Nicol et al.* I *Frédéric Haldi*'s model bliver både illuminansen indenfor og udenfor medtaget, men også solafskærmningens aktuelle position, hvilket medfører, at det er den model, af dem, der er opnået kendskab til gennem litteraturstudiet, der kan tage hensyn til de fleste vigtige forhold i forbindelse med brugen af solafskærmning.

Igennem de ovenstående afsnit, er der også opnået et større kendskab til brugen af ventilation i boliger, gennem et studie boligtypens indflydelse her også infiltrationens størrelse, klimaets betydning samt design udfordringer i forhold til anvendelse af mekanisk ventilation. Diskussionen der knytter sig til brugen af udluftning og ventilation kan findes i afsnit 3.2.3 i hovedrapporten.

De vurderinger, der er beskrevet gennem denne diskussion, bygger på oplysninger fra litteraturstudiet i hovedrapporten og nærværende. Flere af vurderingerne bygger dog kun på et beskedent antal undersøgelser eller målinger, hvorfor det skal bemærkes, at der i nogle tilfæl-

de 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.

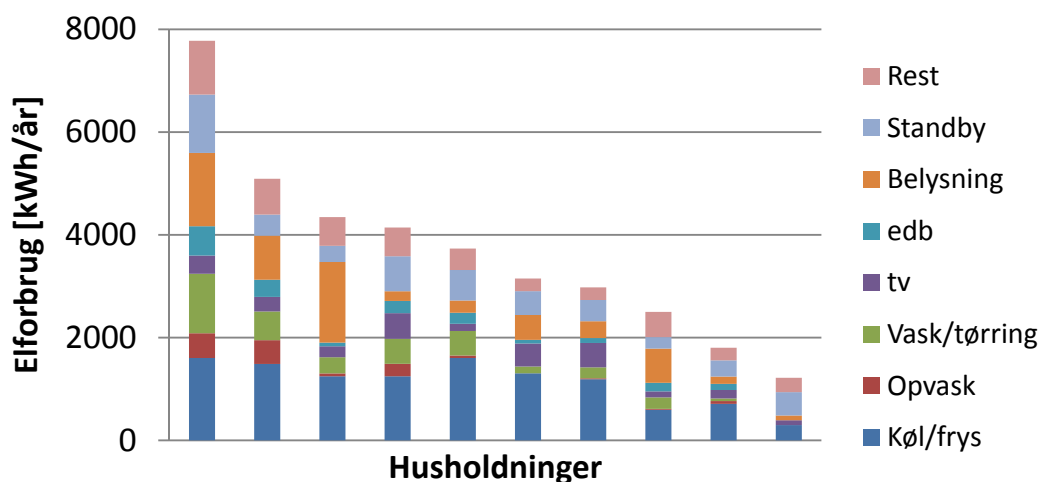
C.5 Elforbrug

I nærværende afsnit gennemgås resultaterne fra forskellige undersøgelser. Flere af resultaterne er en videreførelse af informationen der er givet i afsnit 3.3 i hovedrapporten.

C.5.1 Slutanvendelse for elforbruget

I 1999/2000 udførte *Casper Kofod* [37] en undersøgelse af elforbruget i 100 boliger i Odense. Blandt de 100 boliger 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. Det viste sig, at resultaterne for det samlede elforbrug, for både parcelhuse og lejligheder i undersøgelsen, var tæt på landsgennemsnittet, og ved en sammenligning af bestanden af elapparater med ELMODEL-bolig kunne samme tendens ses. Derfor betragtes resultaterne fra denne undersøgelse som gældende for et repræsentativt udsnit af den danske befolkning, hvilket *Casper Kofod* også bekræfter.

Elforbruget i de 100 boliger blev målt meget detaljeret, idet alle større apparater samt de fleste lamper blev målt og lagret hver 10. minut igennem den måned målingerne stod på. *Casper Kofod* har i analysen opgjort måleresultaterne for de 100 boliger på følgende slutanvendelser: køl/frys, opvask, vask/tørring, tv/video, EDB, belysning, standby og rest (komfur mm.). Af figur C.15 kan det ses, at forbruget mellem de 10 udvalgte husstandene er meget varierende.



Figur C.15: Elforbruget i 100 boliger fordelt på otte slutanvendelser [37].

I det efterfølgende beskrives nogle af de iagttagelser *Casper Kofod* har gjort for de enkelte slutanvendelser, med hensyn til forbrug og forhold, der har betydning for dette.

Køle/frys

Køleskabe og fryser er nogle af de apparater, der er fast inventar i danske boliger, men hvilken type og hvilket energimærke disse har, er dog meget forskelligt. I målingerne fra Odense varierer det årlige forbrug på dette område mellem 250 – 2000 $kWh/år$, men med et typisk forbrug mellem 500 – 1000 $kWh/år$.

Med hensyn til brugen af apparaterne er det forskellig fra familie til familie. Eksempelvis anvendes fryseren hos nogle familier med henblik på at spare penge ved storindkøb, samt som en hjælp i en travl hverdag, hvor de derved kan reducere antallet af daglige indkøb, mens den hos andre familier anvendes den til nedfrysning af fødevarer af egen avl.

Undersøgelsen viser, at mængden af el der anvendes til køleskabe og fryser er uafhængig af familiernes indkomst, da brugen af disse apparater er så forskellig. Yderligere indikerer undersøgelsen, at der generelt er et stort kendskab til energimærkning af apparaterne, og mange familier tilkendegav, at de vil vælge A- eller B-mærkede apparater, men denne holdning bliver dog ikke afspejlet i deres elforbrug. [25, 37]

Opvask

Undersøgelsen viser, at opvaskemaskiner ikke er ligeså almindelige i boliger sammenlignet med andre apparater. Blandt de 100 husstande fra Odense var der opvaskemaskine i 47 af parcelhusene og 2 af lejlighederne, hvilket er tæt på den udbredelse, der er også anvendt i ELMODEL-bolig, hvor det er omkring 46 % af husstandene, der har opvaskemaskine [37]. For de 100 husstande varierer elforbruget til opvaskemaskiner mellem 50 – 700 $kWh/år$. Gennem undersøgelsen viste det sig, at beboere mellem 30 – 60 år bruger mere el til bl.a. opvask end beboere over 60 år, hvor *Casper Kofod* vurderer at årsagen er, at husstandene med beboere mellem 30 – 60 år typisk er større på grund af hjemmeboende børn. Samtidig har undersøgelsen vist, at brugen af opvaskemaskinen stiger med stigende indkomst. [25, 37]

Vask/tørre

I undersøgelsen fra Odense er der 91 husstande, der har vaskemaskine, hvor det er standard for parcelhusene men for lejlighederne er det seks ud af femten, der ejer en. Sammenholdes dette med oplysninger fra ELMODEL-bolig viser det sig, at den samlede udbredelse af vaskemaskiner i projektet fra Odense er højere end landsgennemsnittet, hvilket kan skyldes, at der er en skæv fordeling mellem antallet af lejligheder og parcelhuse i projektet fra Odense, hvor der er højere andel af parcelhuse i forhold til landsgennemsnittet.

Med hensyn til tørretumblere var der i undersøgelsen 49 boliger, der ejede dette apparat. Da udbredelsen af tørretumblere i projektet også er højere end landsgennemsnittet, vurderer *Casper Kofod*, at det kan skyldes samme forhold som nævnt ved vaskemaskinerne, hvor fordelingen af parcelhuse og lejligheder er forskellig fra landsgennemsnittet.

Vaskemaskinernes elforbrug er i undersøgelsen fra Odense mellem 23 – 739 $kWh/år$, og for tørretumblere er dette 20 – 1200 $kWh/år$.

Tv/video

Tv/video er mere eller mindre standardapparater i danske boliger, hvor gennemsnittet per husstand er fire apparater fordelt på tv, video og HIFI. Det afgørende for størrelsen af forbruget er energieffektiviteten og specielt brugstiden af apparaterne, hvilket er meget forskelligt fra familie til familie. Forbruget i undersøgelsen fra Odense viser, at der er et forbrug mellem 4 – 844 $kWh/år$, med et gennemsnit på 211 $kWh/år$.

Rest

Af figur C.15 fremgår det, at der er en gruppe for forbruget der betegnes *rest*. Dette forbrug dækker over alt øvrigt elforbrug i boligen, der ikke tilhører en af de syv andre kategorier. Elforbruget som kategorien *rest* dækker over kan eksempelvis være: madlavning, brug af mikrobølgeovn, frituregryde, kaffemaskine, sandwich-toaster, værktøj og elektriske haveredskaber.

Belysning

Dette afsnit omhandler studier, hvor brugen af kunstig belysning er undersøgt. Det kan være i forhold til typiske vaner for hvorledes belysningen anvendes eller for undersøgelser, der kæder brugen af belysning sammen med andre faktorer som for eksempel inde- eller udetemperaturen.

Forbruget til belysning udgør, som det ses af tabel 3.11, gennemsnitligt 12,9 % af det samlede elforbrug. Ligesom for mange af de øvrige elforbrug varierer elforbruget til belysning også betragteligt fra husstand til husstand. I et projekt udført af *Casper Kofod* [37] i 1999/2000, kan der ses variationer fra 48 kWh/år til 1670 kWh/år med et gennemsnit på 452 kWh/år. Generelt fremkom det af målingerne, at hvis der i husstanden er en tendens til et højt elforbrug, er forbruget til belysning ofte også højt, hvilket bekræftes i figur C.15 på side 37. I projektet blev typerne af anvendte lyskilder ligeledes registreret for hver af de 100 husstande. I tabel C.18, kan det ses hvor stor en andel henholdsvis glødepærer, halogenpærer, lysstofrør og A-pærer, der anvendes, samt hvor stor en andel af elforbruget til belysning de udgør. Der ses ingen sammenhæng mellem størrelsen af elforbruget til belysning og typen af lyskilde, der anvendes i hjemmet.

Lyskilde	Antal	Andel	Andel af elforbrug
	[stk]	[%]	[%]
Glødepærer	1481	56	63
Halogenpærer	538	20	16
Lysstofrør	226	9	11
A-pærer	397	15	11

Tabel C.18: Udbredelsen af forskellige lyskilder samt andelen af elforbruget til belysning for de 100 husstande [37].

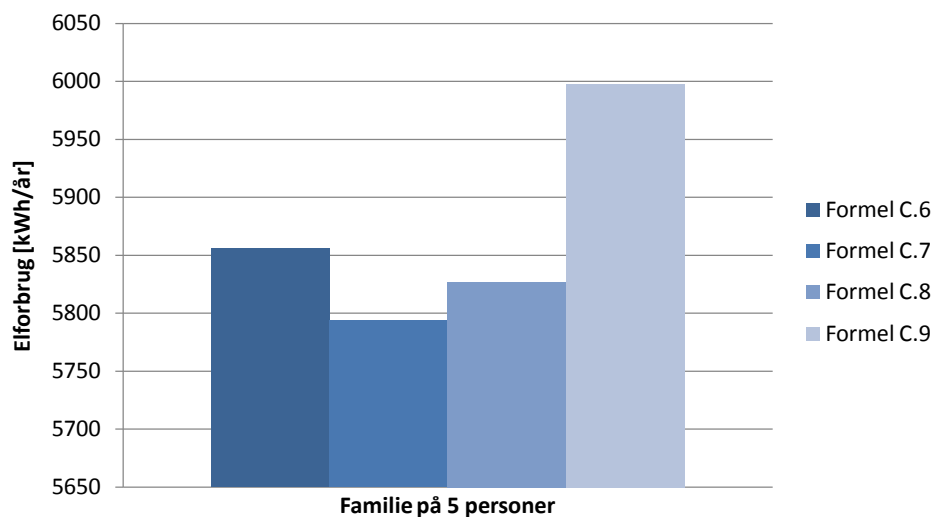
Nicol et al. [49] har i et projekt, der tidligere er omtalt i afsnit 3.2, analyseret brugen af belysning i kontorer set i sammenhæng med ude- og lufttemperaturer. Analysen viser, at der sker et fald i anvendelsen af belysning set i sammenhæng med stigende udetemperaturer, men et mere markant fald kan ses, når anvendelsen af belysning sammenholdes med lufttemperaturen. Det vurderes af *Nicol et al.*, at sammenhængen mellem temperatur og belysning skyldes årstidens variation, hvor der i de kolde vintermåneder er mørkere end i de varme sommermåneder. Analysen viser yderligere, at i de europæiske kontorer, som spørgeundersøgelsen inddrager, anvender over 60 % stadig belysning ved en udetemperatur på 30 °C. Dette er bemærkelsesværdigt, eftersom der typisk er sol på de tidspunkter, hvor udetemperaturen når denne størrelse. Dette kan enten betyde, at der er en dårlig anvendelse af dagslys i disse kontorer, hvormed kunstig belysning er nødvendig, eller at belysningsniveauet er tilfredsstillende, men den kunstige belysning vælges, bevist eller ubevist, at være tændt alligevel.

Det kan også skyldes, at solafskærmningen er anvendt for at reducere varmetilskuddet, og brugerne dermed er nødsaget til at anvende den kunstige belysning.

C.5.2 Væsentlige parametre for elforbruget

I dette afsnit gennemgås yderligere information om hvilke parametre der har indflydelse på elforbruget størrelse.

I litteraturstudiet blev det erfaret, at teenagere og børns tilstedeværelse i boliger påvirker elforbrugets størrelse. I den forbindelse har *Kirsten Gram-Hanssen* [26] på baggrund af regressionsanalyse udviklet nogle formler der beskriver elforbrugets størrelse i forhold til tilstedeværelsen af børn og teenagere. 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, se figur C.16.



Figur C.16: Illustrering af fire forskellige formlers beregning af elforbruget i boliger, udviklet af *Gram-Hanssen et al.* [26].

Analysen fra *Gram-Hanssen et al.* [26] præsenterer flere forskellige måder hvorpå elforbruget i boliger kan beregnes. Formel C.6 viser en simpel beregning hvori der blot tages højde for det samlede antal at beboere i boligen. Formel C.6-C.9 beskriver elforbruget i boliger, når der tages højde for det samlede antal beboere samt tilstedeværelsen af enten småbørn, børn og teenagere.

$$Parcelhus_{generel} = 1988 \text{ kWh/år} + \text{antal personer} \cdot 773,6 \text{ kWh/person pr. år} \quad (\text{C.6})$$

$$Parcelhus_{småbørn} = 1883 \text{ kWh/år} + \text{antal personer} \cdot 840 \text{ kWh/person pr. år} \\ + \text{antal børn} \cdot -289 \text{ kWh/barn pr. år} \quad (\text{C.7})$$

$$Parcelhus_{børn} = 1959 \text{ kWh/år} + \text{antal personer} \cdot 791 \text{ kWh/person pr. år} \\ + \text{antal børn} \cdot -87 \text{ kWh/barn pr. år} \quad (\text{C.8})$$

$$Parcelhus_{teenagere} = 2058 \text{ kWh/år} + \text{antal personer} \cdot 720 \text{ kWh/person pr. år} \\ + \text{antal børn} \cdot 340 \text{ kWh/barn pr. år} \quad (\text{C.9})$$

Formeludtrykkene er fremkommet vha. en regressionsanalyse på baggrund af indsamlede oplysninger på forbrug. Der er i undersøgelsen indsamlet data fra 8421 boliger, men det fremgår ikke af undersøgelsen, om der er anvendt data fra alle boliger, hvor beboere udenfor kategorien der beskrives, betragtes som en voksen, eller om der kun anvendes boliger med den specifikke type der beskrives. Gennemgående for formel C.7-C.9 er, at de kun tager højde for én specifik type af børn eller teenagere og ikke giver mulighed for at kunne kombinere familiesammensætningen. Sammenhængen mellem elforbruget og tilstedeværelsen af småbørn, børn og teenagere ville være mere anvendelig hvis denne var udtrykt gennem én samlet formel i stedet for tre forskellige.

C.6 Vandforbrug

I denne del gennemgås enkelte kilder omhandlende det samlede vandforbrug og det samlede varmtvandsforbrug i boliger.

C.6.1 Samlet vandforbrug

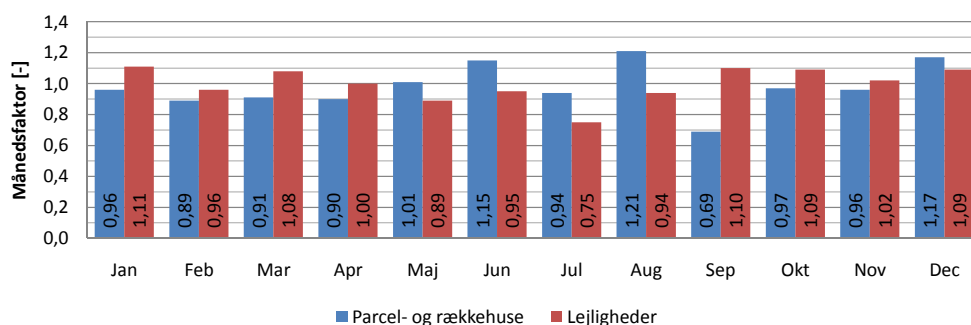
Miljøstyrelsen et al. [46] udarbejdede i 2004 et projekt omhandlende vandforbrug og forbrugsvariationer indenfor husholdninger, institutioner og erhverv, hvilket var tiltænkt at danne grundlag for en fornyelse af normen for dimensionering af forsyningsledninger til drikkevand: „*Alemene vandforsyningsanlæg*“, DS 442. Projektet fokuserede udelukkende på det samlede vandforbrug, og derfor ikke størrelser for forbruget af varmt brugsvand, der er hovedemnet for dette afsnit. Ikke desto mindre fremlægges der i projektet nogle informationer, som vurderes at være relevante for nærværende projekt.

Miljøstyrelsen et al. beskriver, på baggrund af databehandlingen i projektet, at vandforbruget i lejligheder og parcel-/rækkehuse adskiller sig så markant fra hinanden, at disse forbrug bør behandles i to separate kategorier. Denne vurdering er sket på baggrund af det i tabel C.19 angivne antal forbrugsdata, der er indsamlet over en femårig periode.

Forbrugskategori	Data			
	Time	Døgn	Måned	År
Lejligheder	11.632	10.669	567	9.831
Parcel-/rækkehuse	27.994	4.018	51	36.889
Ældreboliger	705	1.496	52	4

Tabel C.19: Antal forbrugsdata ud fra målingstype for de enkelte forbrugskategorier [46].

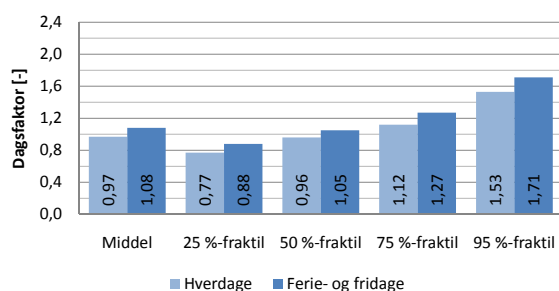
I tabel C.19 er antallet af analyserede husholdninger angivet, hvor også ældreboliger fremgår som en separat kategori. Af *Miljøstyrelsen et al.* anbefales det, at anvende samme oplysninger, som der er fundet for parcel- og rækkehuse, når det årlige vandforbrug i ældreboliger skal estimeres, hvilke vil sige, at figur 3.23 kan anvendes til dette formål. Ud over det årlige vandforbrug er variationen af dette forbrug også analyseret i projektet. Til undersøgelse af variationen af vandforbruget på månedsbasis, fremgår det af tabel C.19, at data fra omkring 23000 lejligheder og 32000 parcel- og rækkehuse kan anvendes til dette formål. Denne variation er vist for lejligheder samt parcel- og rækkehuse på figur C.17, angivet ved en månedsfaktor. Månedsfaktoren er i projektet bestemt på baggrund af forbrugsdata fra 12 på hinanden følgende måneder, hvor disse forbrug er delt med det samlede middelforbrug, under hensyntagen til antallet af dage i de enkelte måneder.



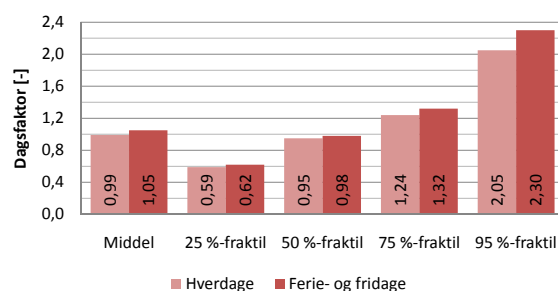
Figur C.17: Variation af vandforbrug angivet ved en månedsfaktor for hver enkelt måned, gældende for henholdsvis parcel- og rækkehuse samt lejligheder [46].

På tilsvarende vis er døgnvariationen bestemt, hvor det af *Miljøstyrelsen et al.* er valgt, at opgøre døgnvariationen ud fra to forskellige døgn typer, hvilket er hverdage og så ferie- og fridage. Af tabel C.19 fremgår det, at der er tilgængeligt data fra omtrent lige så mange husstande, som ved bestemmelse af månedsvariationen, til vurdering af døgnvariationen. På figur C.18 og C.19 fremgår døgnfaktoren for de to førnævnte døgn typer for henholdsvis parcel- og rækkehuse samt lejligheder.

C.6.1. Samlet vandforbrug



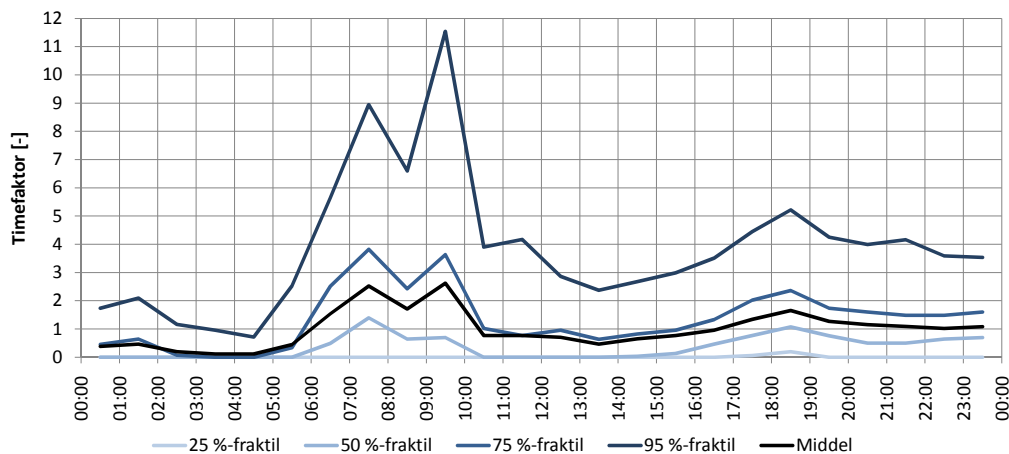
Figur C.18: Døgnvariationen af vandforbruget i parcel- og rækkehuse på hverdage samt ferie- og fridage angivet ved en dagsfaktor [46].



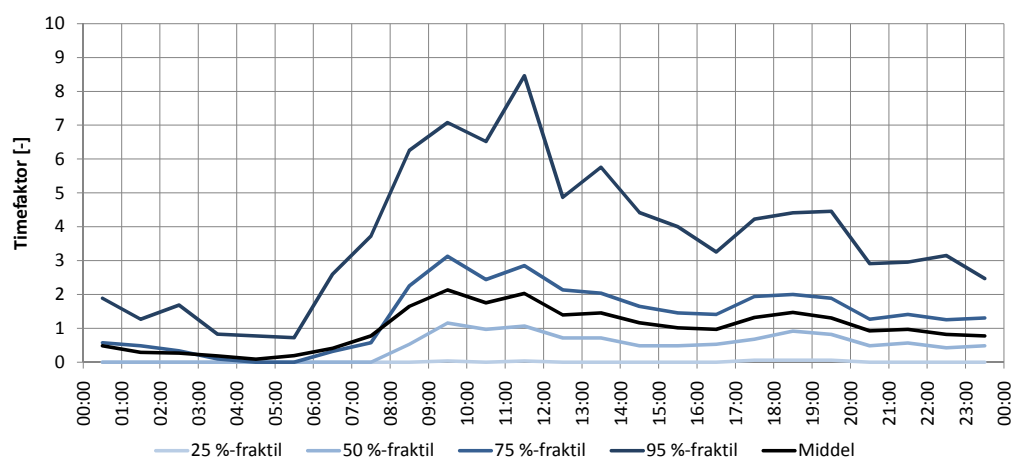
Figur C.19: Døgnvariationen af vandforbruget i lejligheder på hverdage samt ferie- og fridage angivet ved en dagsfaktor [46].

I projektet er vandforbruget på timebasis målt i omkring 11500 lejligheder og 28000, hvilket danner et udmærket grundlag for vurdering af variationen af vandforbruget gennem et døgn. På samme måde som ved døgnvariationen, er det her valgt at anvende de samme døgn typer; hverdage og ferie- og fridage.

Vandforbrugets variation i parcel- og rækkehuse på hverdage er grafisk skitseret på figur C.20 angivet ved en timefaktor. Denne timefaktor er fremkommet ved at dele timeforbruget inden for hvert døgn i forhold til gennemsnitsforbruget. På tilsvarende måde er variationen af vandforbruget i parcel- og rækkehuse på ferie- og fridage skitseret på figur C.21.

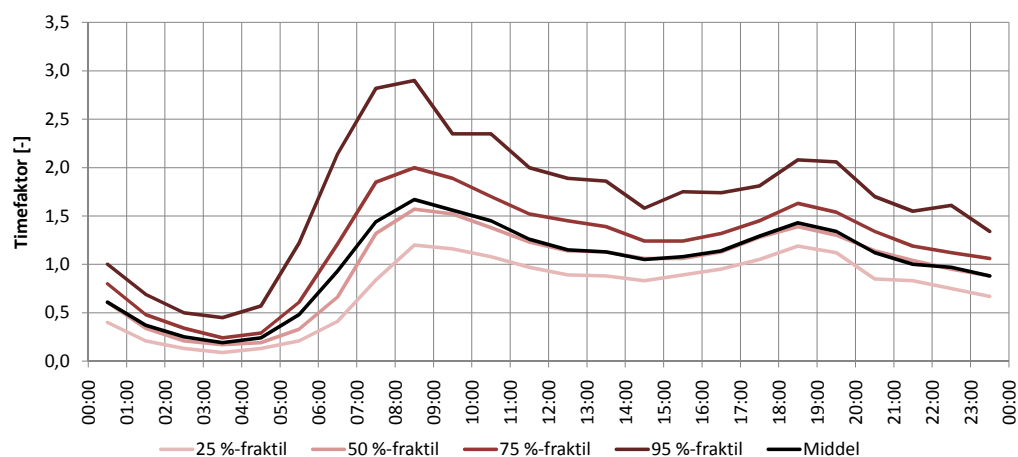


Figur C.20: Timevariationen af vandforbruget i parcel- og rækkehuse på hverdage angivet ved en timefaktor [46].

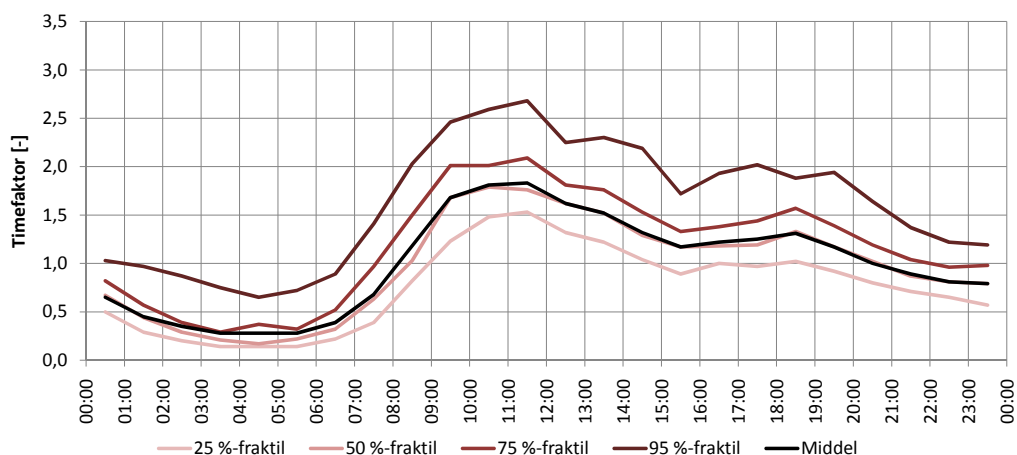


Figur C.21: Timevariationen af vandforbruget i parcel- og rækkehuse på ferie- og fridage angivet ved en timefaktor [46].

Hvad angår timevariationen af vandforbruget i lejligheder på hverdage er på tilsvarende måde angivet ved en timefaktor på figur C.22 og på figur C.23 for lejligheder på ferie- og fridage.



Figur C.22: Timevariationen af vandforbruget i lejligheder på hverdage angivet ved en timefaktor [46].



Figur C.23: Timevariationen af vandforbruget i lejligheder på ferie- og fridage angivet ved en timefaktor [46].

I en undersøgelse fra *Jesper Ole Jensen* [35] blev det erfaret, at ejerformen for husstanden i nogle tilfælde satte en begrænsning for hvilke friheder beboerne havde i forbindelse med boligen. Dette gjaldt eksempelvis udskiftning fra standard toiletter til vandbesparende toiletter eller udskiftning af varmtvandsbeholder set i relation til energiforbrug til opvarmning af varmt brugsvand. I undersøgelsen blev det fremhævet, at der hovedsagligt blev fokuseret på rentabilitet ved udskiftning af f.eks. toiletter i forbindelse med lejeboliger, hvor der i ejerboliger også kunne tages forhold som funktionalitet og miljømæssige aspekter i betragtning ved valg af toilet. På trods af at ejerformen ved ejerboliger ikke satte nogle begrænsninger for valg af toiletter, viste undersøgelsen også, at det ikke altid var det miljømæssige aspekt, der vejede mest, når beboerne skulle købe nyt.

Opsummering af indetemperaturer

Projektnummer	Beskrivelse
1	Temperaturerne er vurderet ud fra et projekt af <i>Peeters et al.</i> [57], hvor der er taget udgangspunkt i figur C.5 til C.7 i hovedrapporten.
2	Oplysningerne stammer fra en et projekt af <i>Kirsten Gram-Hanssen</i> [24] fra boligområdet Stadionparken, hvor boligerne er fra omkring 1970'erne. Byggestilen er tæt-lav byggeri, hvor temperaturerne er på baggrund af indeklimamålinger fra år 2000 over en periode på 14 dage.
3	Oplysningerne stammer fra en et projekt af <i>Kirsten Gram-Hanssen</i> [24] fra boligområdet VA-rækkehuse, hvor boligerne er fra omkring 1960'erne. Byggestilen er tæt-lav byggeri, hvor temperaturerne er på baggrund af indeklimamålinger fra år 2000 over en periode på 14 dage.
4	Oplysningerne stammer fra et projekt af <i>Henrik Tommerup</i> [83] fra et parcelhus opført i 1972. Huset er renoveret i 2005, hvor målingerne stammer fra før og efter renoveringen. Måleperioderne er henholdsvis fra 01.01.2006 til 28.02.2006 samt fra 01.01.2007 til 28.02.2007.
5	Oplysningerne stammer fra et projekt af <i>Jesper Ole Jensen</i> [35], hvor en vicevært i boligområdet Vejlbj Vænget har udtalt sig omkring beboernes typiske temperaturer. Boligerne i området er almene boliger opført i 1968-74.
6	Oplysningerne stammer fra et projekt af <i>Knudsen et al.</i> [36], hvor der er temperaturmålinger fra ni boliger, der er opført i midten af 2000'erne. Temperaturmålingerne er månedsmiddelttemperatur i perioden fra 19.03.2009 til 20.12.2009.
7	Oplysningerne stammer fra et projekt af <i>Brohus et al.</i> [9], hvor der er udført temperaturmålinger i otte rækkehuse i 2005, hvor rækkehusene er fra slutningen af 1980'erne.
8	Oplysningerne stammer fra tre forskellige projekter af <i>Henrik Tommerup</i> [81, 82, 84] for tre forskellige parcelhuse, alle sammen opført i begyndelsen af 2000'erne. Målingerne stammer fra perioden fra 24.09.2003 til 12.04.2004 for to af projekterne og fra 20.12.2003 til 30.04.2004 i det tredje.

Tabel D.1: Forklaring og kildehenvisning til de anvendte projektnumre i afsnit 3.2.3 i hovedrapporten.

Projektnummer	Beskrivelse
9	Oplysningerne stammer fra samme projekt, som nævnt under projektnummer 2, udført af <i>Kirsten Gram-Hanssen</i> [24].
10	Oplysningerne er fra dette projekt, hvor målingerne er hentet for tre passivhuse i Vejle. Husene er opført som enfamiliehuse i midten af 2000'erne. Måleperioden er fra 01.12.2009 til 28.02.2010.
11	Oplysningerne stammer fra samme projekt, som nævnt under projektnummer 3, udført af <i>Kirsten Gram-Hanssen</i> [24].
12	Oplysningerne stammer fra samme projekt, som nævnt under projektnummer 1, udført af <i>Peeters et al.</i> [57].
13	Oplysningerne stammer fra samme projekt, som nævnt under projektnummer 4, udført af <i>Henrik Tommerup</i> [83].
14	Oplysningerne stammer fra et projekt af <i>Kristiansen et al.</i> [38], hvor der er udført målinger på to rækkehuse, der er opført i 1996. Måleperioden er fra 01.12.1997 til 31.12.1997.
15	Oplysningerne stammer fra to projekter udført af henholdsvis <i>Santin et al.</i> [27] og <i>Jeeninga et al.</i> [34], hvor projektet udført af <i>Santin et al.</i> kun omhandler boliger opført efter 1995. Alle boligerne er fra Holland.
16	Oplysningerne stammer fra et projekt af <i>Gill et al.</i> [22], der har udført målinger på 11 engelske husstande bestående af både lejligheder og parcelhuse, der alle er opført i midten af 2000'erne. Der er foretaget målinger i over 1 år startende fra august 2008.
17	Oplysningerne stammer fra to forskellige projekter af <i>Feist et al.</i> [18, 19], hvor førstnævnte projekt er med målinger fra 22 passivhuse i Hannover med målinger fra 01.10.1999 til 30.04.2000. Det sidstnævnte projekt er med målinger fra 8 passivhuse, der er opført i Frankfurt i 2002, hvor målinger stammer fra 01.10.2003 til 30.04.2004.
18	Oplysningerne stammer fra et projekt af <i>Annika Nilsson</i> [50], hvor der er målinger fra tre lejligheder, hvor alle bygningerne er opført i begyndelsen af 2000'erne. Temperaturerne er målt fra slutningen af december 2002 til slutningen af april 2003.

Tabel D.2: Forklaring og kildehenvisning til de anvendte projektnumre i afsnit 3.2.3 i hovedrapporten.

Udeklima

E.1 Dansk udeklima sammenlignet med forholdene i andre projekter

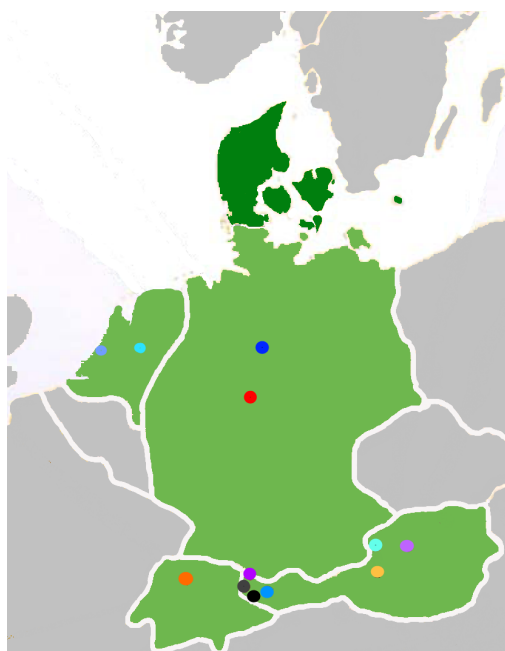
Ved udvikling af fordelingsfunktioner til temperaturforholdene i boliger blev udviklet, blev der anvendt målinger fra projekter der er udført i andre lande end Danmark. Derfor anses det som nødvendigt at sammenligne udeklimaet for Danmark med de øvrige respektive lande. Der er anvendt målinger fra følgende projekter og i figur E.1 ses projekternes placering i Europa:

- *CEPHEUS* [67], målinger fra hhv. Tyskland, Østrig og Schweiz.
- *Maeyens et al.* [42], målinger fra Holland.

Vejrdata fra de nævnte lokationer i figur E.1, er i figur E.2 og E.3 sammenlignet med udetemperaturerne i Danmark for hhv. den gennemsnitlige højste dagtemperatur og den gennemsnitlige mindste nattemperatur for de respektive måneder: november, december, januar og februar. Ud fra figurerne E.2 og E.3 ses det, at udetemperaturerne som forventet er anderledes end det danske klima. De maksimale forskelle ligger for dagtemperaturerne mellem $5 - 9\text{ }^{\circ}\text{C}$ hhv. koldere og varmere vejr. De fleste af lokationerne har stort set samme temperaturer som Danmark og ses der bort fra lokationen Luzern i Schweiz er temperaturforskellen på $5 - 6\text{ }^{\circ}\text{C}$ hhv. koldere og varme vejr. For nattemperaturerne er forskellen mellem $7 - 4\text{ }^{\circ}\text{C}$ hhv. koldere og varmere vejr. Modsat dagtemperaturernes forløb ses det her, at temperaturforholdene om natten i Luzern i Schweiz stemmer bedre overens med det danske klima, dog med undtagelse af november måned, med en afvigelse på $5\text{ }^{\circ}\text{C}$. De største afvigelser fra det danske udeklima ses i dagtimerne i Luzern i Schweiz og i nattetimerne for lokationerne Steyr og Gniß. De mindste afvigelser fra det danske udeklima ses i Hannover og Kassel med afvigelser på maksimalt $2\text{ }^{\circ}\text{C}$ for både dag- og nattemperaturer.

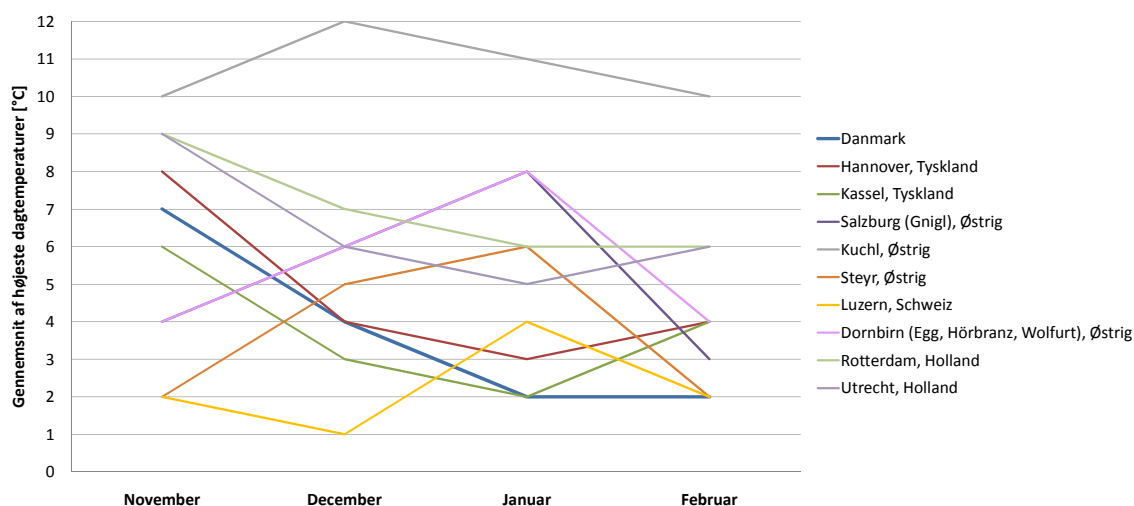
Afvigelserne betragtes som acceptable for alle lokationer og udetemperaturerne anses overordnet set som værende tæt på de danske. Derfor anses indeklimamålinger for boliger i disse lokationer som anvendelige til vurdering af indetemperaturer i boliger i Danmark.

E.1. Dansk udeklima sammenlignet med forholdene i andre projekter

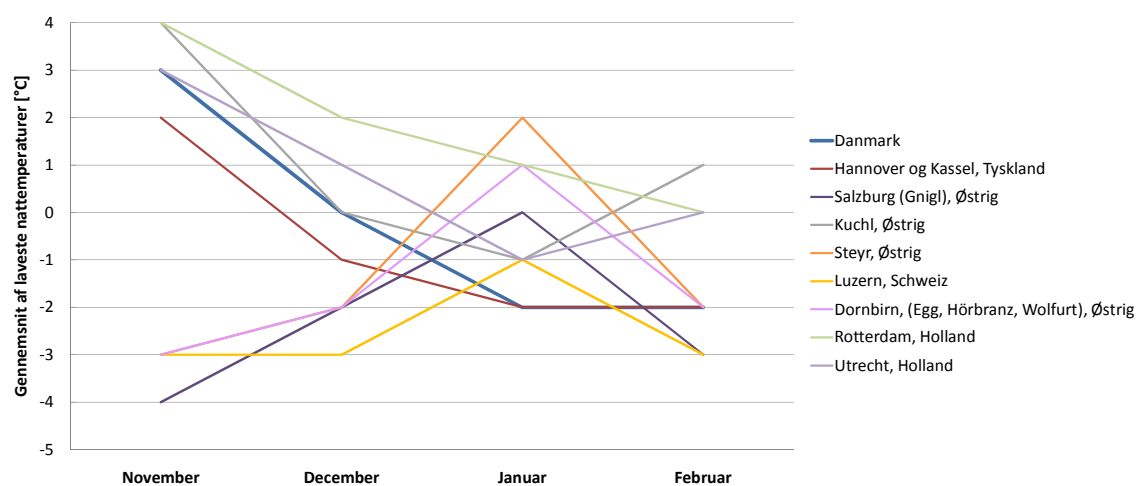


Land	By	Vejrdata
Holland	Rotterdam	ja [33]
Holland	Utrecht	ja [33]
Tyskland	Hannover	ja [33]
Tyskland	Kassel	ja [47]
Schweiz	Luzern	ja [47]
Østrig	Hörbranz	nej, Dornbirn er anvendt [47]
Østrig	Wolfurt	nej, Dornbirn er anvendt [47]
Østrig	Dornbirn*	ja [47]
Østrig	Egg	nej, Dornbirn er anvendt [47]
Østrig	Kuchl	ja [47]
Østrig	Gnigl	ja [47]
Østrig	Steyr	ja [47]

Figur E.1: De forskellige måleprojekters lokation i Europa [67], [27], [42]. * Der findes ikke statistik over gennemsnitlige højste dagtemperaturer og gennemsnitlige mindste nattemperaturer for lokationerne: Wolfurt, Hörbranz og Egg, derfor anvendes data fra Dornbirn. Der er ikke medtaget temperaturmålinger i boliger fra Dornbirn, denne lokation er udelukkende anvendt til indsamling af vejrdato.



Figur E.2: Gennemsnittet af de højeste målte dagtemperaturer i Danmark og de øvrige lokationer i Europa [33], [47].



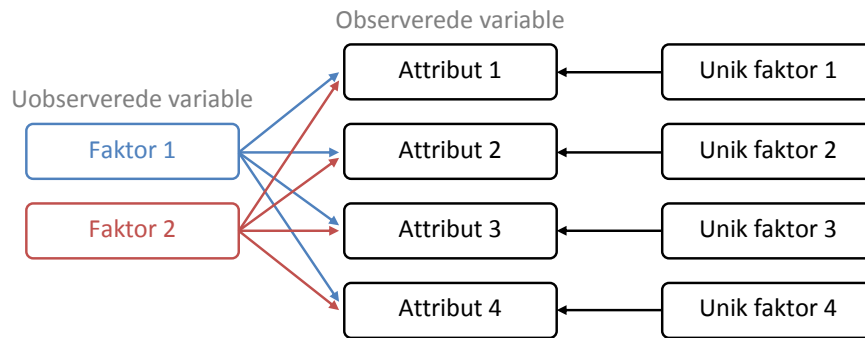
Figur E.3: Gennemsnittet af de laveste målte nattemperaturer i Danmark og de øvrige lokationer i Europa [33], [47].

Faktor- og komponentanalyse

I afsnit 3.1 i litteraturstudiet er der omtalt nogle artikler fra *Barr et al.* [5] samt *Raaij et al.* [86] omhandlende brugeradfærd i boliger. I disse artikler er der anvendt henholdsvis faktor- og komponentanalyse. Det er tidligere nævnt, at anvendelse af disse analysemetoder stadig er nyt inden for dette fagområde, hvorfor metoderne beskrives i det efterfølgende.

Faktoranalyse er en generel betegnelse for en række forskellige analysemetoder, der, sammen med komponentanalyse, befinder sig blandt de deskriptive metoder inden for data mining jf. figur 4.1. Analysemetoderne, der går under betegnelsen faktoranalyse, har til formål at klarlægge uobserverede strukturer mellem nogle observerede variable [13]. Ofte anvendes disse analysemetoder i forbindelse med spørgeskemaundersøgelser, hvor det undersøges, hvorvidt der kan findes en struktur blandt besvarelsene af spørgsmålene, der ikke direkte er registreret, men muligvis eksisterer [6]. Den uobserverede struktur udtrykkes gennem nogle *faktorer*. Idet faktoranalyse ofte anvendes i forbindelse med analyse af spørgsmål i en spørgeskemaundersøgelse [6], vil betegnelsen *attributter* blive anvendt, frem for *observerede variable*, i den efterfølgende beskrivelse. Metoden til undersøgelse af de uobserverede strukturer gør, at faktoranalyse generelt kan opdeles i to underkategorier; eksplorativ og konfirmativ faktoranalyse. Hovedforskellene mellem disse metoder beskrives i de efterfølgende afsnit.

Eksplorativ og konfirmativ faktoranalyse har det samme overordnede mål, og de to analysemetoder har ligeledes det til fælles, at de begge er baseret på en *Common Factor Model* [13]. Betydningen heraf er, at attributterne påvirkes af nogle fælles faktorer, men i forskellig grad. Dette kan eksemplificeres ved en grafisk præsentation af hovedtanken bag faktoranalyse, der er skitseret på figur F.1. Figuren viser et eksempel, hvor de fire attributter symboliserer, hvorledes besvarelsene på fire forskellige spørgsmål påvirkes af to fælles faktorer og individuelt af nogle unikke faktorer. Pilene på figuren symboliserer i hvilken retning påvirkningen sker, og de unikke faktorer anvendes til at forklare den variation af attributterne, der ikke kan forklares ved faktorerne [13]. På figur F.1 er der vist to faktorer, der påvirker samtlige attributter, men hovedtanken med faktoranalyse er, at antallet af faktorer, samt størrelsen af deres påvirkning på attributterne, skal estimeres, således at faktorerne overordnet set påvirker forskellige grupper af attributter.



Figur F.1: Grafisk præsentation af princippet bag *Common Factor Model* [13].

Den efterfølgende fremstilling af analysemetoderne har ikke til formål at udlægge en omfattende matematisk redegørelse for metoderne, men at beskrive den overordnede fremgangsmåde for metoderne således det er muligt at klarlægge fordele, ulemper, input, output og anvendelsesmuligheder for disse set i relation til det nærværende projekts overordnede mål.

F.1 Eksplorativ faktoranalyse

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 attributterne. 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. [10]

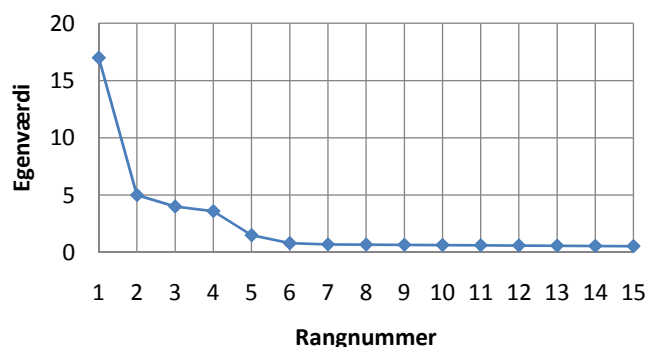
For at det er muligt at udføre en eksplorativ faktoranalyse, kræves det, at der tages udgangspunkt i nogle eksisterende data. Et eksempel på input kan være resultaterne af en spørgeskemaundersøgelse, hvor spørgsmålene, som tidligere nævnt, benævnes som attributter.

Ved en eksplorativ faktoranalyse er det korrelationen mellem attributterne i besvarelsene, der danner grundlaget for analysen [21]. Praktisk set gøres dette ved at opstille en korrelationsmatrix, hvor korrelationen beregnes parvis (bivariat korrelation) for samtlige attributter [79], jf. figur F.2. Ideen er, at attributterne, der har høj indbyrdes korrelation, påvirkes af de samme faktorer, mens de attributter der har en lav indbyrdes korrelation ikke påvirkes af de samme faktorer.

$$\text{Korrelationsmatrix} = \begin{matrix} & \begin{matrix} \text{Attribut 1} \\ \text{Attribut 2} \\ \text{Attribut 3} \\ \text{Attribut 4} \\ \text{Attribut 5} \\ \dots \end{matrix} \\ \begin{matrix} \text{Attribut 1} \\ \text{Attribut 2} \\ \text{Attribut 3} \\ \text{Attribut 4} \\ \text{Attribut 5} \\ \dots \end{matrix} & \begin{bmatrix} 1,0 & \dots & \dots & \dots & \dots & \dots \\ \dots & 1,0 & \dots & \dots & \dots & \dots \\ \dots & \dots & 1,0 & \dots & \dots & \dots \\ \dots & \dots & \dots & 1,0 & \dots & \dots \\ \dots & \dots & \dots & \dots & 1,0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \end{matrix}$$

Figur F.2: Skitsering af korrelationsmatrix.

Vurderingen af hvor mange faktorer der skal anvendes i analysen er en central del af eksplorativ faktoranalyse. Ofte foretages denne vurdering med udgangspunkt i korrelationsmatricen. Idet der ikke findes en unik måde til at vurdere, hvorvidt korrelationen mellem de forskellige attributter er høj eller lav, opstår der principielt flere mulige løsninger [21]. Der findes imidlertid nogle metoder, der kan indikere hvor mange faktorer, der kan anvendes, som for eksempel *Kaiser-kriteriet*, *scree-test*, *likelihood ratio test* m.fl. [21]. De førnævnte metoder er alle nogle matematiske regneoperationer, hvor resultatet indikerer hvor mange faktorer, der kan udtrækkes til den videre analyse. De to førstnævnte metoder er blandt de mest anvendte [11, 45], og begge metoder tager udgangspunkt i korrelationsmatricens egenverdier, der principielt set er variansen af faktorerne [21]. *Kaiser-kriteriet* giver et objektivt skøn over antallet af faktorer, da det angiver, at der skal udtrækkes et antal faktorer, der svarer til antallet af egenverdier fra korrelationsmatricen, der har en værdi større end 1 [21]. I en *scree-test* skal egenverdierne rangeres med den højste værdi først og derefter indtegnes i en graf, hvor egenverdierne er plottet mod deres rangnummer [21]. Grafen kaldes da et *scree plot*, hvor et eksempel på et sådan plot er vist på figur F.3. Det sted hvor der forekommer et stort niveauskift mellem egenverdiernes størrelse, angivet som et „knæk“ på grafen, angiver hvor mange faktorer, der skal udtrækkes [21]. Ydermere bør de efterfølgende egenverdier have et tilnærmelsesvist konstant fald [45]. I eksemplet på figur F.3 kan der udtrækkes fem faktorer til den videre analyse [45]. Valget af antal faktorer kan blive subjektivt præget, da der kan forekomme flere „knæk“ på grafen, og der må foretages en vurdering af hvor grafen begynder at flade ud.

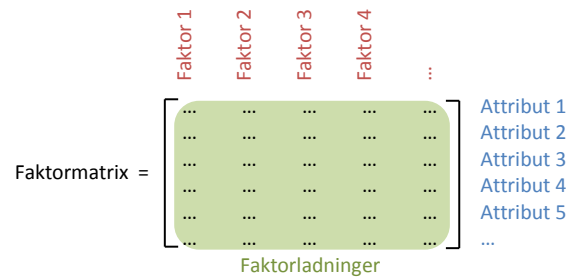


Figur F.3: Eksempel på et *scree plot*

Det er i det foregående beskrevet, at der findes både objektive og subjektive metoder, der kan anvendes til vurdering af hvor mange faktorer, der skal udtrækkes til en videre analyse. Erfaringer har vist, at disse metoder ofte giver forskellige resultater, hvorfor valget af metode kan have en betydelig indflydelse på resultatet af faktoranalysen [11]. Det betyder, at resultaterne, enten direkte eller indirekte, er subjektivt præget, da valget af metode har indflydelse på disse.

Når antallet af faktorer er valgt, skal den størrelse hvormed faktorerne påvirker attributterne estimeres. Denne størrelse beskrives ved *faktorladninger*, hvilket svarer til korrelationen mellem attributterne og faktorerne [45]. Estimeringen af faktorladningerne kan ske på baggrund af flere forskellige metoder, som for eksempel *Maximum Likelihood*, *Least-square methods*, *Alpha factoring* m.fl. [13, 45]. Resultaterne af denne estimering kan indskrives i en faktormatrix, der er en samling af faktorladningerne for samtlige faktorer, hvilket er skitseret på figur F.4. Denne faktormatrix danner i første omgang udgangspunktet for fortolkningen af fakto-

rerne, hvor resultaterne i nogle tilfælde er nemmere at fortolke end i andre. Eksempelvis kan en faktor, der har høje faktorladninger (svarende til høje korrelationer) til størstedelen af attributterne være vanskelig at fortolke, da faktoren blot vil udtrykke en „generel“ beskrivelse af attributterne. Derimod vil en faktor, der har høje faktorladninger til nogle få attributter og lave til øvrige ofte være nemmere at fortolke og tildele en „overskrift“, der er beskrivende for faktoren [45].



Figur F.4: Skitsering af faktormatrix.

I de tilfælde hvor faktormatricen ikke umiddelbart giver en meningsfuld fortolkning, kan der anvendes nogle matematiske analysemetoder, der kan lette fortolkningen. Disse matematiske metoder, der typisk betegnes som rotationsmetoder eller transformationsmetoder, har til formål at reducere antallet af høje faktorladninger blandt de enkelte faktorer [45]. Generelt set er der to forskellige metoder, hvorpå rotationen kan foregå, henholdsvis en *ortogonal rotation* samt en *oblique rotation*. En *ortogonal rotation* kan foretages på faktorer, der er indbyrdes uafhængige (ukorreleret), og som fortsat ønskes uafhængige [21]. Modsvarende er der *oblique rotation*, der er en „skæv“ transformation, som resulterer i, at faktorerne efterfølgende bliver indbyrdes afhængige [21]. I litteraturen findes ikke entydige oplysninger på, hvilken rotation der er at foretrække i hvilke situationer. *Conway et al.* [11] angiver, at oblique rotation er fordelagtig at anvende, da den ofte giver simple resultater, hvilket er at foretrække i den efterfølgende proces. Da det på forhånd ikke står klart, om faktorerne er indbyrdes korrelerede, angiver *Conway et al.*, at en ortogonal rotation medfører urealistiske resultater, der kan være vanskelige at fortolke, såfremt faktorerne er korrelerede. *Conway et al.* giver udtryk for, at oblique rotation altid bør anvendes, da det vil give simple resultater, uafhængigt af om faktorerne er korrelerede eller ej. I modsætning hertil angiver *Mario Mazzocchi* [45], at en oblique rotation i nogle tilfælde kan give nogle simple og let fortolkelige resultater, men det medfører også, at faktorladningerne ikke efterfølgende kan sammenlignes direkte. *Mario Mazzocchi* angiver, at oblique rotation *kun* bør anvendes, såfremt det ønskes at antage en korrelation mellem faktorerne.

Efter en eventuel rotation er foretaget, kan resultaterne fortolkes og tildeles en sigende „overskrift“. Ud fra nogle kriterier angående størrelser er det muligt at undlade lave faktorladninger, således fortolkningen forenkles [45]. Det betyder, at der kun er høje faktorladning tilbage ved hver faktor, hvilket er skitseret som prikker på figur F.5.

Da en faktor kun udtrykker en, i nogle tilfælde beskeden, del af den totale variation blandt attributterne, er dette endnu et punkt, hvor analysen kan blive subjektivt præget. Det skyldes, at overskriften for faktorerne tildeles ud fra subjektiv fortolkning af, hvilke attributter de enkelte faktorer dækker over.

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	...
Attribut 1	...				
Attribut 2			...		
Attribut 3		...			
Attribut 4		...			
Attribut 5				...	
...				...	...

Figur F.5: Skitsering af fortolkning af faktormatrix.

Hvis elementerne i analysen ønskes beskrevet ud fra faktorerne, der dækker over nogle uobserverede eller ikke direkte målbare strukturer i besvarelsene, skal der estimeres en *faktorscore* mellem hvert element og hver enkelt faktor. Disse faktorscores estimeres ud fra elementernes oprindelige værdier for attributterne (dvs. personers besvarelser af spørgsmål i en spørgeskemaundersøgelse), samt på baggrund af de estimerede faktorladninger. Det vil sige, at prikkerne på figur F.6 angiver de enkelte faktorscores. Idet disse faktorscores estimeres ud fra faktorladningernes størrelse har det den betydning, at de attributter der har de højeste faktorladninger, har størst indflydelse på størrelsen af de pågældende faktorscores. Når de enkelte faktorscores er beregnet, kan der foretages en analyse af elementerne ud fra disse.

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	...
Person A	...	...	...	...	...
Person B	...	...	...	...	...
Person C	...	...	...	...	...
Person D	...	...	...	...	...
...	...	...	...	...	...

Faktorscores

Figur F.6: Skitsering af hvorledes faktorscores kan anvendes til analyse af data.

F.2 Konfirmativ faktoranalyse

Konfirmativ faktoranalyse tager udgangspunkt i samme grundlæggende model som eksplorativ faktoranalyse, hvor nogle bagvedliggende faktorer antages at have indflydelse på udfaldet for attributterne [21]. En forskel i forhold til eksplorativ faktoranalyse er, at der i konfirmativ faktoranalyse på forhånd haves en hypotese om antallet af faktorer, og ofte også hvilke attributter der påvirkes af hvilke faktorer [21]. Det, der kendetegner en konfirmativ faktoranalyse, er, at analysen består af en optakt og en efterfølgende hypotesetest. Optakten består i, at faktorladningerne estimeres, ud fra nogle af de samme metoder som er beskrevet under eksplorativ faktoranalyse, eksempelvis *Maximum Likelihood*, *Least-square methods*, *Alpha factoring* m.fl. [13, 45]. Dette sker på baggrund af den oprindelige hypotese om antallet af faktorer og eventuelt strukturen blandt faktorerne og attributterne. 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 nogle aktuelle data.

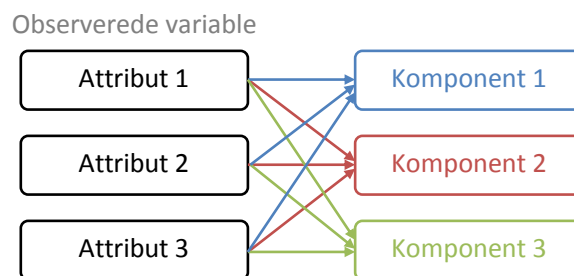
Det første skridt i en konfirmativ faktoranalyse er, at udfærdige en detaljeret beskrivelse af hvilken hypotese, og derved hvilken model, der ønskes analyseret [13]. Hypotesen kan være opstillet ud fra erfaringer, eller den kan være fremkommet gennem en tidligere udført eksplorativ faktoranalyse, hvor resultatet ønskes efterprøvet på nogle nye data. Det betyder, at der skal indsamles nogle data, der kan anvendes til at efterprøve hypotesen [13].

Når data er indsamlet, skal der foretages en estimering af faktorladninger. Estimeringen af faktorladningerne har principielt set samme fremgangsmåde, som det er beskrevet under eksplorativ faktoranalyse [13]. Det betyder, at de indsamlede data kan sammenholdes med hypotesen, som ønskes efterprøvet. Der findes flere forskellige måder hvorpå data og model kan sammenholdes, men en ofte anvendt metode er *Chi-square test for goodness of fit* [13, 21]. Princippet bag denne test er, at det forventede udfald, som er defineret i hypotesen, sammenholdes med det observerede udfald fra målingerne eller spørgeskemaundersøgelsen. Dette giver et resultat, der viser, hvorvidt den opstillede hypotese kan accepteres eller må forkastes på baggrund af et valgt signifikansniveau.

F.3 Komponentanalyse

Hovedformålet med komponentanalyser er at omskrive attributterne til et færre antal komponenter, der tager højde for hovedparten af den variation, der er mellem attributterne [13]. Ofte forveksles komponentanalyse med eksplorativ faktoranalyse grundet en lignende fremgangsmåde [13]. Det bør bemærkes, at de to analysemetoder har forskellige hovedformål; eksplorativ faktoranalyse har til formål at „udforske“ data, mens komponentanalyse har til formål at *reducere* data.

Der er nogle ganske væsentlige punkter, hvormed komponentanalyse adskiller sig fra eksplorative faktoranalyse. Helt grundlæggende adskiller de to analysemetoder sig fra hinanden på den måde, at de bygger på forskellige modeller. Eksplorativ faktoranalyse bygger på en *Common Factor Model*, hvor attributterne antages påvirket af faktorer og unikke faktorer, som fremgår af figur F.1. I komponentanalysen antages attributterne at kunne udtrykke et antal komponenter, som skitseret på figur F.7 [13]. Det vil sige, at der i eksplorativ faktoranalyse skelnes mellem delt og unik variation mellem attributterne, hvor der i komponentanalyse ikke skelnes mellem dette. Forskellene for hvilken model analysemetoderne bygger på kan grafisk set observeres ud fra pilenes retning mellem attributterne og faktorerne/komponenterne. Det bør her bemærkes, at figur F.7 viser et antal komponenter, der svarer til antallet af attributter, hvilket dog kun er tilfældet i den første del af analysen. Antallet af komponenter vil, som det beskrives senere i dette afsnit, blive reduceret.



Figur F.7: Grafisk præsentation af princippet bag komponentanalyse [13].

Foruden den ovenstående beskrivelse adskiller komponentanalyse sig fra eksplorativ faktoranalyse på den måde, at antallet af komponenter fastsættes indledningsvis. Derefter beregnes komponentladninger, og først efterfølgende reduceres antallet af komponenter. Set i relation

til eksplorativ faktoranalyse bliver antallet af faktorer vurderet som et af de første skridt i komponentanalysen, og først derefter estimeres faktorladningerne.

For at kunne udføre en komponentanalyse kræves det, at der tages udgangspunkt i nogle eksisterende data. Det kan, på samme måde som ved eksplorativ faktoranalyse, eksempelvis dreje sig om resultaterne af en spørgeskemaundersøgelse. Det første skridt i en komponentanalyse er, at antallet af komponenter vælges, således at det svarer til antallet af attributter, jf. figur F.7. Derved kan der opstilles et formeludtryk for hvert komponent, hvor formeludtrykket er en linearkombination af samtlige attributter, og hvor komponenterne er indbyrdes uafhængige [45]. Den matematiske fremgangsmåde gennemgås ikke i dette afsnit. Essensen er, at der kan findes en unik matematisk løsning til de opstillede formeludtryk, hvilket adskiller sig markant fra eksplorativ faktoranalyse, hvor de enkelte faktorladninger og faktorscores estimeres [45].

På dette punkt er dataet, der søges analyseret, beskrevet ved et antal ukorrelerede komponenter, der svarer til antallet af attributter. Eftersom et af hovedformålene med komponentanalyse er datareduktion, er det efterfølgende skridt i analysen, at antallet af komponenter skal reduceres. Dette skal imidlertid ske på en sådan måde, at de tilbageværende komponenter udtrykker hovedparten af den variation, der findes i datagrundlaget for komponentanalysen [45]. Denne reduktion kan eksempelvis ske ud fra *Kaiser-kriteriet*, *scree-test*, *likelihood ratio test* m.fl., som det ligeledes var tilfældet ved eksplorativ faktoranalyse [45]. Resultatet af reduktionen bliver, at komponenterne ikke længere udtrykker hele variationen mellem attributterne, hvilket var tilfældet, da antallet af attributter og komponenter var det samme. Ved at benytte nogle af de ovennævnte metoder til at reducere antallet af komponenter sikres det, at det er de komponenter, der udtrykker hovedparten af variationen blandt attributterne, der beholdes.

Efter reduktionen af komponentantallet er hovedformålet med komponentanalysen principielt nået. Ofte forsøges der efterfølgende foretaget en fortolkning af komponenterne, hvilket kan være en af årsagerne til, at komponentanalyse og eksplorativ faktoranalyse ofte forveksles. Hvis der stadig ønskes at foretage en fortolkning af resultaterne, kan der foretages en rotation af komponenterne, såfremt en umiddelbar fortolkning ikke giver meningsfulde resultater. Det bør dog bemærkes, at en eventuel rotation bør udføres på samtlige komponenter, og ikke kun de komponenter, der er bibeholdt. Dette skyldes, at efter rotationen er udført, er der ingen garanti for, at de udvalgte komponenter stadig er dem, der udtrykker hovedparten af variationen i datagrundlaget [45]. Når en eventuel rotation af komponenterne er foretaget, kan der, såfremt der er opnået nogle meningsfulde resultater, beregnes nogle *komponentscores*, der kan anvendes til en efterfølgende analyse af data [45]. Disse komponentscores kan forstås som en form for faktorscores, som det var tilfældet under eksplorative faktoranalyser, og kan eksempelvis anvendes hvis spørgeskemabesvarelser skal sammenholdes.

Klyngeanalyse

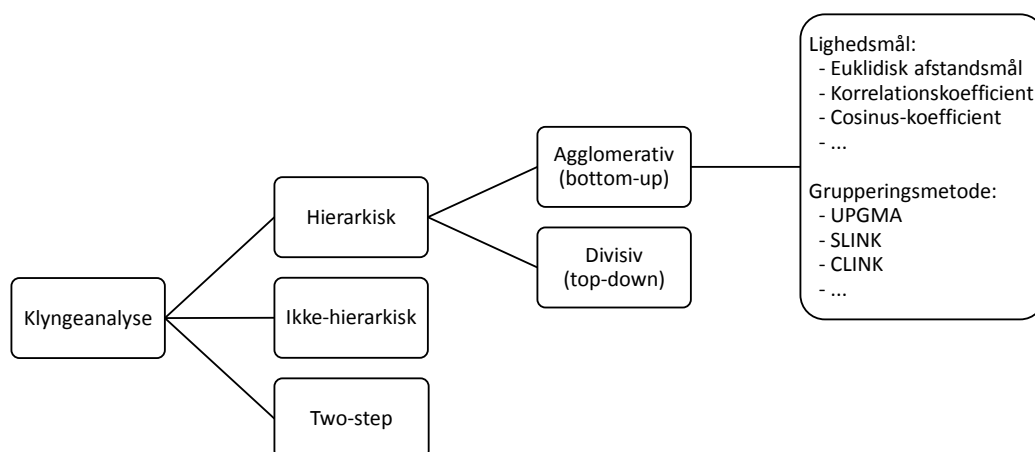
I afsnit 3.1 i litteraturstudiet er der omtalt nogle artikler fra *Barr et al.* [5] samt *Raaij et al.* [86] omhandlende brugeradfærd i boliger, hvor der i disse artikler er anvendt klyngeanalyse. Idet klyngeanalyser tidligere har været anvendt på projekter med relevans for nærværende projekt, samt at denne anvendelse stadig er ny inden for dette fagområde, vil denne analysemetode blive beskrevet nærmere. Dette sker med henblik på at beskrive den overordnede fremgangsmåde for metoden, således det er muligt at klarlægge fordele, ulemper og anvendelsesmuligheder for klyngeanalyser. Dette skal efterfølgende danne grundlag for vurdering af, hvorvidt metoden er brugbar i nærværende projekt.

Klyngeanalyse er en generel betegnelse for en række forskellige analysemetoder, der befinder sig blandt de deskriptive metoder inden for data mining, jf. figur 4.1 på side 87 [65] i hovedrapporten. De forskellige analysemetoder har et fælles mål; at identificere hvilke elementer (eksempelvis personer) i et datasæt, 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 [45]. 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 [45, 65].

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. Den efterfølgende proces er iterativ, hvor elementerne placeres i de klynger, hvortil afstanden til „tyngdepunktet“ er kortest, hvorefter klyngernes „tyngdepunkt“ beregnes på ny. Derefter omroteres elementerne således at de hele tiden tilhører den klynge, hvortil afstanden er kortest. Processen fortsætter indtil ingen elementer omroteres, eller til et given stopkriterium nås. Såfremt der er anvendt et stopkriterium, er det et punkt, hvor ikke-hierarkisk klyngeanalyse kan blive subjektivt præget. [45]

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 baggrund for at vurdere, 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. Fordelen ved denne fremgangsmåde er, at antallet af klynger vurderes, til dels subjektivt, ud fra datagrundlaget, hvorefter elementerne kan omroteres, således de grupperes i de klynger, hvortil afstanden er kortest. [45]

Af litteraturen fremgår det, at hierarkisk klyngeanalyse er den mest udbredte analysemetode [45, 65, 77], hvorfor denne vil danne grundlaget for en nærmere beskrivelse af klyngeanalyser. Mulighederne for valg af klyngeanalyse er skitseret på figur G.1, hvor det bemærkes, at flere af begreberne, der er nævnt på figur G.1, uddybes senere i det nærværende afsnit. Den efterfølgende beskrivelse fokuserer på hierarkisk klyngeanalyse, hvor beskrivelsen ikke har til formål at udlægge en omfattende matematisk redegørelse for metoderne, men at klarlægge den overordnede fremgangsmåde for disse.



Figur G.1: Skitsering af muligheder for valg af klyngeanalyse, hvor hovedfokus i den efterfølgende beskrivelse vil være blandt agglomerative hierarkiske klyngeanalyser.

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 [45]. Udgangspunktet for hierarkiske klyngeanalyser er eksisterende data fra eksempelvis en spørgeskemaundersøgelse. Spørgeskemabesvarelsene danner grundlaget for opstilling af en datamatrix, hvor respondenterne ofte omtales som elementer og spørgsmålene som attributter. Ofte er det ønsket med en klyngeanalyse at analysere elementerne med henblik på at undersøge hvilke personer, der besvarer spørgsmålene ens, og hvilke personer der besvarer spørgsmålene forskelligt. Det vil lede til en gruppering af personer, der har lignende besvarelser, og disse grupper kan beskrives ved deres „generelle“ kendetegn [65]. For at foretage denne analyse, med udgangspunkt i *H. Charles Romesburgs* [65] udlægning af metoden, skal datamatricen opskrives hvor søjlerne i matricen er elementerne, og rækkerne er besvarelserne til de enkelte attributter. Der er også muligt at analysere attributterne, hvilket blot kræver en transponering af datamatricen, hvorefter fremgangsmåden ikke afskiller sig fra analysen af elementerne. I den efterfølgende beskrivelse af analysemetoden tages der udgangspunkt i, at målet med den hierarkiske klyngeanalyse er at analysere elementerne, da dette også var udgangspunktet i de tidligere omtalte artikler fra henholdsvis *Barr et al.* og *Raaij et al.*

Efter datamatricen er opstillet, skal det vurderes, hvorvidt det er nødvendigt at udføre en standardisering af dataet i denne. Standardiseringen er nødvendig, såfremt der er anvendt forskellige måleenheder blandt attributterne. Såfremt der er anvendt forskellige måleenheder vil det medføre, at nogle attributter tildeles en større vægt i den efterfølgende fremgangsmåde sammenholdt med andre [45, 65]. En standardisering foregår ved, at der fra de enkelte

data i datamatricen subtraheres middelværdien for en pågældende attribut og efterfølgende divideres med standardafvigelsen for attributten [45].

Der skal efterfølgende beregnes nogle parvise mål for ligheden mellem elementerne, hvor der er flere forskellige metoder, der kan anvendes til dette. Nogle metoder giver en lav værdi når to elementer er tilnærmelsesvis ens, mens andre metoder giver en lav værdi for elementer, der er særdeles forskellige. Af forskellige lighedsmål kan nævnes *euklidiske afstand*, *blokafstand*, *cosinuskoefficient*, *korrelationskoefficient* m.fl. [65]. Et af de mest udbredte mål for ligheden mellem elementer i en klyngeanalyse er den såkaldte euklidiske afstand. Den giver en værdi på nul såfremt to elementer har samme attributter, og en højere værdi som funktion af større forskel mellem attributterne [45, 65]. Der beregnes ét lighedsmål for hver parvise kombination af elementer, der er beregnet på baggrund af differensen mellem samtlige attributter for de to elementer. Resultaterne kan indskrives i en såkaldt *resemblance matrix*, hvis størrelse afhænger af antallet af elementer, jf. figur G.2. Denne figur illustrerer et eksempel, hvor der er anvendt euklidisk afstand som lighedsmål. Opbygningen af matricen vil være tilsvarende, såfremt der anvendes et andet lighedsmål. Da matricen er symmetrisk, hvilket vil medføre at $e_{12} = e_{21}$, $e_{13} = e_{31}$ og så fremdeles [65]. På baggrund af dette vil der i det efterfølgende kun blive anvendt ét euklidiske afstandsmål mellem hvert elementpar, der er navngivet med det laveste indeksnummer først. Resemblance matricen danner grundlaget for grupperingen af elementer gennem klyngeanalysen. Det er nævnt tidligere, at grupperingen af elementer sker skridt for skridt, hvilket medfører, at resemblance matricen skal beregnes på ny efter hvert skridt. Det skyldes, at den parvise lighed mellem klyngerne ændres, når nogle klynger grupperes.

Resemblance matrix =

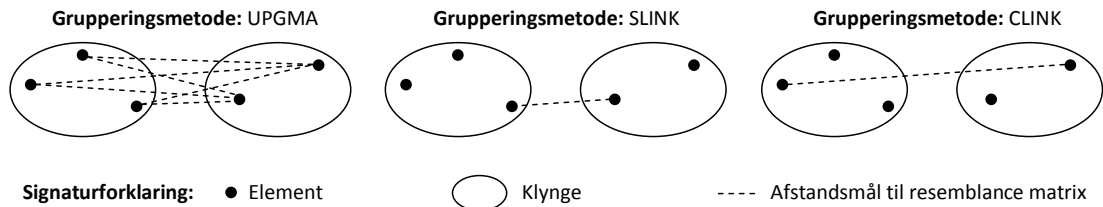
Element 1	Element 2	Element 3	Element 4	...	
e_{11}	e_{12}	e_{13}	e_{14}	...	Element 1
e_{21}	e_{22}	e_{23}	e_{24}	...	Element 2
e_{31}	e_{32}	e_{33}	e_{34}	...	Element 3
e_{41}	e_{42}	e_{43}	e_{44}	...	Element 4
...	...	...	...	...	...

Euklidiske afstandsmål

Figur G.2: Grafisk præsentation af resemblance matrix med euklidiske afstand.

Når ligheden mellem de enkelte elementer er beregnet, ud fra de førnævnte muligheder, skal der vælges en analysemetode. Inden for hierarkisk klyngeanalyse kan der skelnes mellem *agglomerative* og *divisive* analyser, hvilket kan forstås som henholdsvis en *bottom-up* eller *top-down* fremgangsmåde [45]. Den mest anvendte af disse metoder er en agglomerativ klyngeanalyse [45, 65], hvorfor der lægges mest vægt på denne i det efterfølgende. For både agglomerative og divisive klyngeanalyser skal der vælges, hvilken grupperingsmetode der skal danne grundlaget for grupperingen af klynger. Der findes flere forskellige grupperingsmetoder, som for eksempel *unweighted pair-group method using arithmetic averages (UPGMA)*, *single linkage clustering method (SLINK)*, *complete linkage clustering method (CLINK)* m.fl. [65]. Hovedforskellen blandt de tre grupperingsmetoder er måden hvorpå lighedsmålet til resemblance matricen beregnes efter en gruppering af klynger. En grafisk skitsering af hvilke elementer der indgår i lighedsmålet til resemblance matricen mellem to klynger, ved brug af de tre nævnte grupperingsmetoder, fremgår af figur G.3. Hvis UPGMA-metoden benyt-

tes skal der beregnes et lighedsmål, der er en middelværdi af samtlige parvise lighedsmål for elementerne i de to klynger [65]. Benyttes SLINK-metoden skal det laveste lighedsmål anvendes, og ved CLINK-metoden benyttes det højeste lighedsmål [65]. Blandt de mest udbredte grupperingsmetoder er UPGMA, der også er udgangspunktet for den efterfølgende beskrivelse [45, 65].

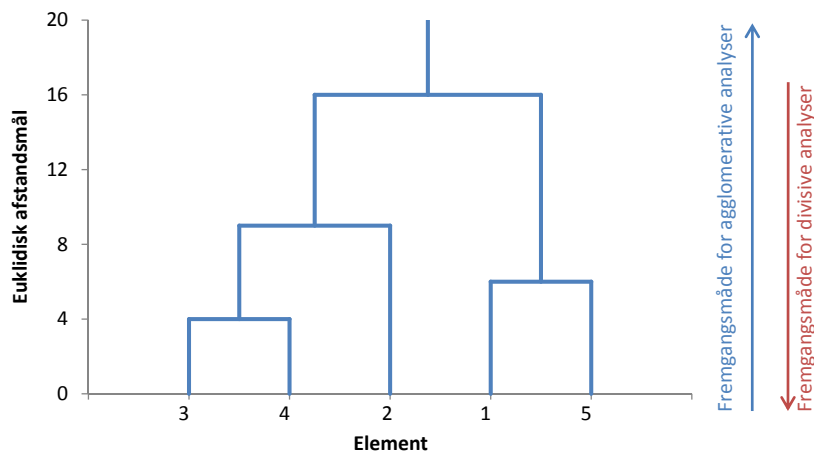


Figur G.3: Grafisk skitsering af hvorledes lighedsmålet beregnes for tre grupperingsmetoder.

I det første skridt i udførelsen af en agglomerativ hierarkisk klyngeanalyse betragtes hvert enkelt element som en klynge. Det vil sige, at der indledningsvis haves et antal klynger, der svarer til antallet af elementer. I det næste skridt grupperes de to klynger, der er mest lignedannede, hvilket drejer sig om de to klynger, der har den laveste euklidiske afstand, såfremt det er dette lighedsmål, der anvendes. Når der er foretaget en gruppering af to klynger, skal der foretages en ny beregning af de euklidiske afstande. Den nye beregning af de euklidiske afstande vil reducere den oprindelige resemblance matrix, der er skitseret på figur G.2, med én række og én kolonne. Hvis det eksempelvis er *klynge 3* og *klynge 4*, der har den laveste euklidiske afstand, vil disse grupperes til én klynge, der kan navngives *klynge 3,4*. Det vil medføre, at beregningen af den euklidiske afstand mellem *klynge 1* og *klynge 3,4* vil blive en middelværdi af e_{13} og e_{14} , såfremt UPGMA-metoden anvendes, og så fremdeles for de øvrige kombinationsmuligheder. Det næste skridt i analysen er at gruppere de klynger, der, i henhold til den seneste beregning af de euklidiske afstandsmål, er mest lignedannede. Denne proces fortsætter, frem til der haves én klynge, hvoraf betegnelsen *bottom-up* bliver meningsfuld for metoden. Formålet med at gruppere alle elementerne i én klynge er at danne grundlag for en efterfølgende subjektiv vurdering af, hvor mange klynger der skal anvendes i den følgende analyse. Vurderingen kan eksempelvis ske på baggrund af et dendrogram, der giver et overblik over resultaterne af den agglomerativ hierarkisk klyngeanalyse. Et eksempel på et dendrogram er vist på figur G.4, og en nærmere beskrivelse af dendrogrammet følger. Hovedtanken er, at det skal vurderes hvor langt oppe dendrogrammet skal stoppes, hvilket afgør, hvor mange klynger der skal arbejdes videre med. Jo flere klynger der grupperes ved lave euklidiske afstande, desto mere lignedannede er klyngerne. Valget af det antal klynger, der skal arbejdes videre med, er subjektivt præget. Det er en balancegang mellem en detaljeret model med mange klynger og en simpel model med få klynger. Der kan desuden anvendes forskellige metoder, der kan indikere, hvor mange klynger der kan anvendes. Et eksempel herpå kan være *scree-test*, der tidligere er beskrevet ved eksplorativ faktoranalyse i afsnit F.1. [45, 65]

Hvis den ovenstående fremgangsmåde sammenholdes med en divisiv hierarkisk klyngeanalyse, fremstår sidstnævnte med en fremgangsmåde, der er modsat agglomerativ hierarkisk klyngeanalyse. I en divisiv hierarkisk klyngeanalyse antages det fra begyndelsen, at samtlige elementer tilhører én klynge. Herefter opdeles klyngen skridt for skridt, således at der til sidst haves et antal klynger, der svarer til antallet af elementer.

Resultatet af både en agglomerativ og divisiv hierarkisk klyngeanalyse kan indtegnes i et såkaldt dendrogram, der er en grafisk skitsering af den foregående proces. På denne måde kan det være nemmere at få overblik over sammenhænge mellem de forskellige klynger. Et eksempel er illustreret på figur G.4. Her fremgår det, at det er *klynge 3* og *klynge 4*, der er mest ligedannede, eftersom det er de klynger, der har den laveste euklidiske afstand svarende til en værdi på 4. Af dendrogrammet fremgår det ligeledes, at i skridt to er *klynge 1* og *klynge 5* grupperet ved en euklidisk afstand på 6. Efter hvert skridt skal der, som tidligere nævnt, beregnes en ny resemblance matrix. Ved det tredje skridt, jf. figur G.4, fremgår det, at den nyeste resemblance matrix viser, at *klynge 3,4* og *klynge 2* kan grupperes sammen med en euklidisk afstand på 9 og så fremdeles.



Figur G.4: Eksempel på et dendrogram.

Foruden dendrogrammet kan resultaterne ligeledes indskrives i en såkaldt *cophenetic matrix*. Denne matrix er en samling af de euklidiske afstandsmål, der er beregnet mellem de forskellige klynger, der er grupperet gennem analysen [65]. Dendrogrammet og cophenetic matrixen viser de samme resultater, men de har hver sine fordele. Hvor dendrogrammet hovedsagligt anvendes til at skabe overblik, så anvendes cophenetic matrixen til at beregne en såkaldt *cophenetic korrelationskoefficient*, der er beregnet med udgangspunkt i Pearsons korrelationskoefficient [65]. Et eksempel, på hvorledes cophenetic matrixen kan se ud, kan ses på figur G.5, der relaterer sig til dendrogrammet på figur G.4.

	Klynge 1	Klynge 2	Klynge 3	Klynge 4	Klynge 5	
Cophenetic matrix =	-	-	-	-	-	Klynge 1
	16	-	-	-	-	Klynge 2
	16	9	-	-	-	Klynge 3
	16	9	4	-	-	Klynge 4
	6	16	16	16	-	Klynge 5
	Endelige euklidiske afstandsmål					

Figur G.5: Eksempel på en cophenetic matrix. Denne figur knytter sig til dendrogrammet på figur G.4.

Da det er ikke muligt at sammenholde outputtet fra den hierarkisk klyngeanalyse (dendrogrammet og cophenetic matricen) med oprindelige inputdata, må der i stedet anvendes de oprindelige euklidiske afstandsmål, der er skitseret i resemblance matricen på figur G.2 [65]. Derved beregnes cophenetic korrelationskoefficient ud fra cophenetic matricen og de oprindelige euklidiske afstandsmål i resemblance matricen. Resultatet viser, hvor godt outputtet fra den hierarkisk klyngeanalyse afspejler de oprindelige indbyrdes ligheder i dataet [65].

Det kan ofte være hensigtsmæssigt at lave en beskrivelse af kendetegnene for hver klynge. Denne beskrivelse udføres ofte på baggrund af en gennemsnitlig værdi for hver enkelt attribut beregnet på baggrund af elementerne i de enkelte klynger [65]. Hvis der findes oplysninger fra et andet element, og det skal vurderes hvilken klynge dette element kan tilhøre, skal der beregnes et euklidisk afstandsmål mellem det nye element og de gennemsnitlige værdier for attributterne i de enkelte klynger. Størrelsen af det euklidiske afstandsmål afgør, hvilken klynge elementet er tættest knyttet til. Denne metode gælder udelukkende, såfremt der tidligere i klyngeanalysen er anvendt euklidisk afstand og UPGMA-metoden. Såfremt der er anvendt andre lighedsmål eller analysemetoder, skal det nye element vurderes på baggrund af samme forudsætninger [65].

Boligtypens betydning i klassificeringstræet for rumopvarmning

Det er i problemanalysen i afsnit 2.2.1 erfaret, at boligtypen har indflydelse på energiforbruget til rumopvarmning. Det fremgår blandt andet at lejeboliger tidligere er opført i ringere kvalitet end ejerboliger, hvormed energiforbruget opgjort per arealenhed er større i lejeboligerne [24]. Det er i problemformuleringen i afsnit 2.5 beskrevet, at der i det nærværende projekt fokuseres på parcelhuse. Da boligtypen er essentiel i forhold til vurderingen af energiforbruget til rumopvarmning, beskrives mulighederne for at undersøge boligtypens betydning for udviklingen af et klassificeringstræ i det efterfølgende.

Der kan tages højde for boligtypens betydning ved enten at udvikle forskellige klassificeringstræer afhængigt af boligtypen, eller boligtypen kan medtages som attribut i et samlet klassificeringstræ.

Der kan differentieres mellem boligtyperne ved at udvikle to klassificeringstræer, som hver især har fokus på enten parcel-/rækkehuse eller lejligheder. Derved skabes yderligere et grundlag for, at der inden for samme boligtype, kan sammenholdes energiforbrug mellem boliger. Ved udvikling af ét samlet klassificeringstræ opstår der mulighed for at undersøge, om der er forskel på hvilke dele af brugeradfærden, der har størst indflydelse på energiforbruget, afhængig af om det er en lejlighed eller et parcel-/rækkehus. Dette bør imidlertid ikke være det primære fokusområde, idet der i tidligere undersøgelser ikke er set en entydig tendens til, at der er en direkte forskel mellem beboernes adfærd afhængig af boligtypen. Det kan ligeledes overvejes, om der er grundlag for at undersøge, hvorvidt ejerforholdet har betydning for beboernes adfærd i de forskellige boligtyper. Dette kan enten udføres ved at udvikle forskellige klassificeringstræer for de respektive ejerforhold, eller ved at indføre ejerforholdet som en attribut i hvert af klassificeringstræerne for henholdsvis parcel-/rækkehuse og lejligheder.

Hypotesetest

Grundlaget for udvikling af brugerprofiler er målinger fra forskellige husstande, hvor det er nødvendigt for den videre udvikling, at beskrive de forskellige målinger med statistiske fordelinger. For at vurdere hvorvidt målingerne kan beskrives med en given sandsynlighedsfordeling, kan der anvendes hypotesetests. I dette projekt er der anvendt to forskellige testtyper, hvor den første er en *Chi-square test for goodness of fit*, og den anden er en *Kolmogorov-Smirnov one-sample test*. Fremgangsmåden for de nævnte hypotesetests bygger i dette projekt på beskrivelser fra Ayyub et al. [4].

Dette kapitel opdeles i to afsnit, hvor de to første afsnit beskriver den grundlæggende teori bag de valgte hypotesetests, og de to sidste afsnit beskriver hovedtrækkene bag de regneark, der er lavet til hypotesetestene.

I.1 Grundlæggende teori om Chi-square test for goodness of fit

Det er, som tidligere nævnt, valgt at anvende hypotesetesten *Chi-square test for goodness of fit*, hvor der er seks grundlæggende skridt, der skal gennemgås for denne model. Disse skridt skal gennemføres, før det kan konkluderes, hvorvidt den valgte sandsynlighedsfordeling kan beskrive målingerne eller ej. Det bør dog bemærkes, at et af de krav der er til anvendelse af denne metode er, at antallet af målinger der skal testes, som minimum skal være 15, men gerne betydeligt større. Derudover er der nogle begrænsninger i forbindelse med antallet af frihedsgrader i de enkelte tests, hvilket kommenteres senere. De skridt, der skal gennemgås, beskrives kort i det efterfølgende:

Skridt 1: formulering af hypotese

Grundlaget for denne form for test er, at der opstilles to hypoteser; en nulhypotese og en alternativ hypotese. Nulhypotesen formuleres altid som en "lighedshændelse", hvilket eksempelvis kan være en hypotese om, at nogle givne målinger kan karakteriseres ved en normalfordeling med en bestemt middelværdi og standardafvigelse. Den alternative hypotese kan i dette tilfælde være, at de givne målinger ikke kan karakteriseres ved en normalfordeling med den valgte middelværdi og standardafvigelse. Hvis testen ender ud med, at nulhypotesen må forkastes, skal den alternative hypotese accepteres. I det førnævnte eksempel kan det betyde, at de givne målinger enten ikke er normalfordelte eller at de ikke kan beskrives ved de valgte fordelingsparametre, hvilket for en normalfordeling kan være enten middelværdi eller standardafvigelse.

Skridt 2: Valg af statistisk model

Eftersom det er valgt at anvende *Chi-square test for goodness of fit*, er valget af den statistiske model derved også fastlagt. Modellen beskrives ved den teststatistik, der er vist i formel I.1.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - F_i)^2}{F_i} \quad (\text{I.1})$$

Hvor:

χ^2	Beregnet teststatistik af den variable parameter karakteriseret ved en χ^2 -fordeling.
i	Intervalnummer.
k	Samlet antal intervaller.
O_i	Observeret hyppighed af den variable parameter i det givne intervalnummer.
F_i	Forventet hyppighed af den variable parameter i det givne intervalnummer.

Den beregnede størrelse χ^2 angiver den relative forskel mellem de observerede hyppigheder fra målingerne og de forventede hyppigheder, der er beregnet ud fra den i nulhypotesen valgte sandsynlighedsfordeling og de tilhørende fordelingsparametre.

Skridt 3: Valg af signifikansniveau

Der vælges et signifikansniveau, der angiver en grænse for hvor stor risikoen må være for, at resultatet af hypotesetesten er fremkommet tilfældigt. I dette projekt anvendes der et signifikansniveau på 0,05, hvilket i mange tilfælde anses som værende et passende niveau [4]. Valget af et signifikansniveau på 0,05 betyder i praksis, at nulhypotesen forkastes, hvis der er mere end 5 % risiko for, at nulhypotesen ikke er sand.

Skridt 4: Beregning af teststatistik

Målingerne inddeles i nogle valgte intervaller og hyppighederne for forekomster i hvert interval registreres. På baggrund af nulhypotesen beregnes de forventede hyppigheder for forekomster i hvert interval, som anvendes sammen med de observerede hyppigheder til beregning af teststatistikken, der er angivet ved formel I.1. Teststatistikken beregnes for hvert enkelt interval hvorefter størrelserne summeres op. Det bør dog bemærkes, at det forventede antal hyppigheder i hvert interval bør være over fem. Hvis ikke det er tilfældet, må to eller flere intervaller slås sammen, således at det forventede antal forekomster bliver større end fem.

Skridt 5: Definer forkastelsesområde

Ud fra det valgte signifikansniveau og det aktuelle antal frihedsgrader kan forkastelsesområdet defineres. Antallet af frihedsgrader beregnes på baggrund af antallet af estimerede parametre trukket fra antallet af beregnede teststatistikker. Forkastelsesområdet defineres med en kritisk χ^2 -værdi, som efterfølgende anvendes til at vurdere hvorvidt nulhypotesen må forkastes eller ej.

Skridt 6: Konklusion

Teststatistikken, der er beregnet i skridt fire ud fra formel I.1, sammenholdes med den kritiske χ^2 -værdi, der er bestemt i skridt fem. Hvis teststatistikken er lavere end den kritiske χ^2 -værdi kan nulhypotesen accepteres. Det betyder, at med det valgte signifikansniveau, kan det med 95 % sikkerhed godtgøres, at målingerne kan beskrives med den sandsynlighedsfordeling, der

er defineret i nulhypotesen. Er teststatistikken derimod højere end den kritiske χ^2 -værdi, er der mere end 5 % sandsynlighed for, at målingerne er fremkommet tilfældigt, og derved må nulhypotesen forkastes og den alternative hypotese accepteres.

I.2 Grundlæggende teori om Kolmogorov-Smirnov one-sample test

Det er beskrevet tidligere, at hypotesetesten *Chi-square test for goodness of fit* kun kan anvendes, såfremt der haves minimum 15 målinger. I nogle tilfælde haves en mindre datamængde, hvilket betyder, at der i stedet kan anvendes en *Kolmogorov-Smirnov one-sample test* som hypotesetest. Der skal for denne hypotesetest, ligesom for den foregående, gennemgås seks skridt, før det kan konkluderes, hvorvidt den valgte sandsynlighedsfordeling kan beskrive målingerne eller ej. De seks skridt er nærmest identisk med de tidligere beskrevne, hvorfor de kun gennemgås kortfattet.

Skridt 1: formulering af hypotese

Grundlaget for denne form for test er, at der opstilles to hypoteser på tilsvarende måde, som det er beskrevet i afsnit I.1.

Skridt 2: Valg af statistisk model

Da der er valgt en *Kolmogorov-Smirnov one-sample test* er valget af den statistiske model på forhånd fastlagt. Modellen kan beskrives ved den teststatistik, der er vist i formel I.2.

$$KS = \max \left\{ \begin{array}{lll} |F_x(x_1) - F_s(x_0)| & , & |F_x(x_1) - F_s(x_1)| & , & |F_x(x_2) - F_s(x_1)| \\ |F_x(x_2) - F_s(x_2)| & , & \dots & & |F_x(x_i) - F_s(x_{i-1})| \\ |F_x(x_i) - F_s(x_i)| & , & |F_x(x_n) - F_s(x_{n-1})| & , & |F_x(x_n) - F_s(x_n)| \end{array} \right\} \quad (\text{I.2})$$

Hvor:

KS	Beregnet teststatistik af den variable parameter.
F_x	Fordelingsfunktionen for den tæthedsfunktion, der er defineret i nulhypotesen.
F_s	Fordelingsfunktionen for målingerne givet ved formel I.3
i	Intervalnummer.
n	Antal målinger.

$$F_s(x) = \begin{cases} 0 & \text{for } x < x_1 \\ \frac{i}{n} & \text{for } x_i \leq x < x_{i+1} \\ 1 & \text{for } x \geq x_n \end{cases} \quad (\text{I.3})$$

Skridt 3: Valg af signifikansniveau

Der vælges et signifikansniveau, der angiver en grænse for hvor stor risikoen må være for, at resultatet af hypotesetesten er fremkommet tilfældigt. I dette projekt anvendes der et

signifikansniveau på 0,05, hvilket i mange tilfælde anses som værende et passende niveau [4]. Valget af et signifikansniveau på 0,05 betyder i praksis, at nulhypotesen forkastes, hvis der er mere end 5 % risiko for, at nulhypotesen ikke er sand.

Skridt 4: Beregning af teststatistik

Der er i det første skridt defineret en hypotese, som giver oplysninger, der kan anvendes i beregning af en kumulativ sandsynlighedsfordeling, der er i overensstemmelse med hypotesen. På baggrund af hypotesen og målingerne, der ønskes testet, kan der beregnes en kumulativ sandsynlighed for hver enkelt måling. Disse størrelser anvendes i teststatistikken, der er defineret i skridt to, hvorefter den maksimale teststatistik findes.

Skridt 5: Definer forkastelsesområde

Der beregnes en kritisk værdi, KS_α , der anvendes til vurdering af hvorvidt nulhypotesen må forkastes eller ej. Eftersom der er valgt et signifikansniveau på 0,05 bliver $KS_\alpha = KS_{0,05}$, og den kan findes ved et tabelopslag, hvor størrelsen af denne kun afhænger af antallet af målinger og det valgte signifikansniveau.

Skridt 6: Konklusion

Den maksimale teststatistik, der er beregnet i skridt fire ud fra formel I.2, sammenholdes med den kritiske $KS_{0,05}$ -værdi, der er bestemt i skridt fem. Hvis teststatistikken er lavere end den kritiske $KS_{0,05}$ -værdi kan nulhypotesen accepteres. Det betyder, at med det valgte signifikansniveau, kan det med 95 % sikkerhed godtgøres, at målingerne kan beskrives med den sandsynlighedsfordeling, der er defineret i nulhypotesen. Er teststatistikken derimod højere end den kritiske $KS_{0,05}$ -værdi, er der mere end 5 % sandsynlighed for, at målingerne er fremkommet tilfældigt, og derved må nulhypotesen forkastes og den alternative hypotese accepteres.

I.3 Regneark til udførelse af Chi-square test for goodness of fit

I dette projekt er der udarbejdet et regneark, der kan anvendes til at foretage hypotesetesten *Chi-square test for goodness of fit* med forskellige sandsynlighedsfordelinger. Regnearket er opbygget således, at i fanebladet *datagrundlag* kan indtastes grundlæggende oplysninger om de hypotesetests, der skal udføres i det aktuelle regneark. Det kan eksempelvis dreje sig om kildehenvisninger til projekter, hvor de anvendte målinger stammer fra. Efterfølgende er der mulighed for at udføre en hypotesetest med de følgende sandsynlighedsfordelinger:

- Normalfordeling (fanebladet *Chi2-test for normal*)
- Lognormalfordeling (fanebladet *Chi2-test for lognormal*)
- Eksponentialfordeling (fanebladet *Chi2-test for eksponential*)
- Weibullfordeling (fanebladet *Chi2-test for weibull*)
- Gammafordeling (fanebladet *Chi2-test for gamma*)

Fælles for de ovennævnte faneblade er, at der er anvendt tre forskellige farver i arkene; grøn, blå og rød. I cellerne, der er farvet grøn, skal der manuelt indtastes oplysninger. Det drejer

sig eksempelvis om oplysninger som formulering af nulhypotese og den alternative hypotese, indtastning af det ønskede signifikansniveau, antal estimerede parametre, måledata mv. De målte værdier indtastes som en søjlevektor startende i celle A13. I kolonne B kan de ønskede intervaller indtastes.

Den blå farve viser, at der kan anvendes Excel-funktionen *histogram*, der kan findes under *dataanalyse*, til bestemmelse af antallet af forekomster for målingerne i hvert af de angivne intervaller.

Cellerne med den røde farve indikerer, at beregningerne i disse celler foregår automatisk. Det kan eksempelvis dreje sig om beregninger af den forventede sandsynlighed for at den variable parameter falder indenfor de givne intervaller mv.

Hovedresultatet af de forskellige hypotesetests kan findes i celle H9 eller I9, hvor det angives hvorvidt nulhypotesen kan accepteres eller ej.

I.4 Regneark til udførelse af Kolmogorov-Smirnov one-sample test

I dette projekt er der udarbejdet et regneark, der kan anvendes til at foretage hypotesetesten *Kolmogorov-Smirnov one-sample test* med forskellige sandsynlighedsfordelinger. Regnearket er opbygget således, at i fanebladet *datagrundlag* kan indtastes grundlæggende oplysninger om de hypotesetests, der skal udføres i det aktuelle regneark. Det kan eksempelvis dreje sig om kildehenvisninger til projekter, hvor de anvendte målinger stammer fra.

I dette regneark er der anvendt to forskellige farvekoder; grøn og rød. I cellerne, der er farvet grøn, skal der manuelt indtastes oplysninger. Det drejer sig eksempelvis om oplysninger for det ønskede signifikansniveau, den kritiske KS_α -værdi samt måledata. De målte værdier indtastes som en søjlevektor startende i celle B14, hvor værdierne sorteres med den laveste først.

Cellerne med den røde farve indikerer, at beregningerne i disse celler foregår automatisk. Det kan eksempelvis dreje sig om beregninger af den kumulative sandsynlighed for målingerne og teststatistikken.

Hovedresultatet af hypotesetesten kan findes i celle F10 og G10, hvor det angives hvorvidt nulhypotesen kan accepteres eller ej.

Vurdering af lufttemperatur

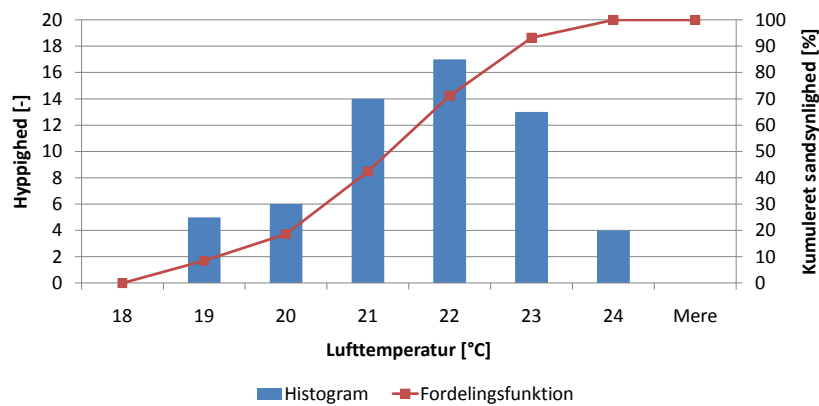
I dette afsnit beskrives udviklingen af sandsynlighedsfordelinger for lufttemperaturer i boliger. Indledningsvis skal der indsamles data, således der haves et solidt grundlag for opstilling af sandsynlighedsfordelingerne, således disse kan beskrive de variationer af lufttemperaturen, der kan forventes i boliger. For at verificere at de enkelte sandsynlighedsfordelinger kan beskrive det data, der ligger til grund for fordelingerne, bliver der foretaget en kvantitativ analyse af disse.

J.1 Overordnet temperaturspænd

I afsnit 3.2.1.3 og 3.2.3 fra litteraturstudiet i hovedrapporten er der i flere tilfælde nævnt lufttemperaturer i boliger fra forskellige artikler og rapporter. Nogle af lufttemperaturerne stammer fra målinger mens andre er fra spørgeundersøgelser og udsagn fra personer. Til at vurdere inden for hvilket temperaturspænd lufttemperaturen i boliger kan forventes at variere, er det i dette projekt valgt, at datagrundlaget skal være temperaturmålinger og ikke udsagn fra beboere eller spørgeundersøgelser. Årsagen til dette valg er, at udsagn fra beboere om hvilken temperatur der haves i boligen kan afvige fra hvilken temperatur, der rent faktisk kan måles [24, s. 104]. Datagrundlaget vurderes derved at være stærkere, hvis det overordnede temperaturspænd vurderes ud fra målinger frem for udsagn.

I dette projekt er temperaturspændet for boliger indledningsvis holdt adskilt, ud fra om målingerne er udført i danske eller udenlandske husstande. Temperaturmålingerne er så vidt muligt indsamlet fra opvarmningssæsonen, da det er i dette tidsrum der anvendes energi til rumopvarmning. For danske husstande er der indsamlet oplysninger om lufttemperaturer i Komfort Husene i Vejle samt oplysninger fra *Kirsten Gram-Hanssen* [24], *Knudsen et al.* [36], *Kristiansen et al.* [38], *Simonsen et al.* [68] og *Tommerup et al.* [81, 82, 83, 84]¹. Samlet set giver det temperaturmålinger for 59 danske husstande med både parcel- og rækkehuse for en kombination af både nye og ældre boliger. På figur J.1 er der optegnet et histogram og en kumulative fordeling over de målte lufttemperaturer. Det kan ses af figuren, at det højeste antal forekomne temperaturer ligger i intervallet mellem 21 – 22 °C, hvor den samme tendens kan ses på figur 3.15 på side 61 i hovedrapporten.

¹En sammenfatning af lufttemperaturerne kan ses i regnearket *Sandsynlighedsfordelinger\ Lufttemperatur\ Opsummering af indetemperaturer.xlsx* på den medfølgende DVD.



Figur J.1: Histogram og fordelingsfunktion for lufttemperaturer i 59 danske husstande.

Der skal foretages en kvantitativ undersøgelse af hvilken sandsynlighedsfordeling, de ovennævnte målinger af lufttemperaturer kan beskrives med. Til denne undersøgelse anvendes der i dette projekt en *Chi-square test for goodness of fit*, der er en hypotesetest, der direkte kan anvendes til vurdering af hvorvidt en given sandsynlighedsfordeling kan beskrive målingerne. Denne hypotesetest går grundlæggende ud på, at der indledningsvis defineres en hypotese for hvorvidt en given sandsynlighedsfordeling med nogle bestemte fordelingsparametre, kan beskrive den målte variable, der i dette tilfælde er lufttemperaturen. Efterfølgende inddeles den målte variable i nogle passende intervaller, hvorefter hyppigheden for forekomster af de den variable registreres for hvert interval. Resultatet af dette ser grafisk ud, som vist på figur J.1 for de målte lufttemperaturer i 59 danske boliger, men oplysningerne anvendes dog numerisk i hypotesetesten. På baggrund af den definerede hypotese estimeres det forventede antal forekomster af den variable indenfor de enkelte intervaller, hvorefter differensen mellem de målte og de forventede hyppigheder anvendes til at vurdere den i hypotesen definerede sandsynlighedsfordeling i forhold til de aktuelle målinger. Den grundlæggende teori og fremgangsmåde for denne hypotesetest er nærmere beskrevet i appendiks I.

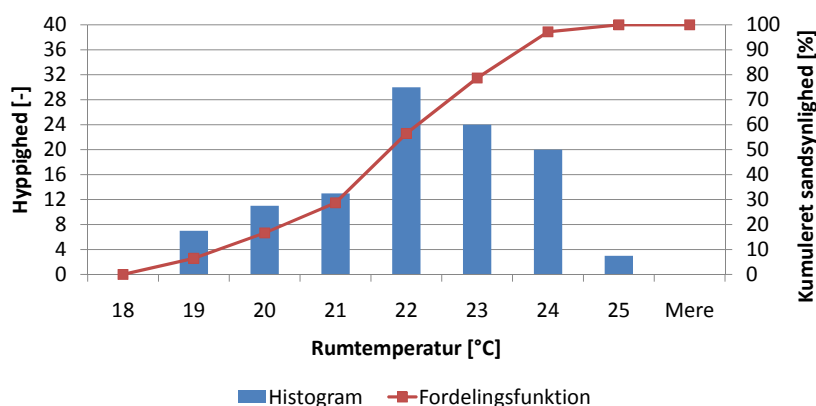
Hypotesetesten er anvendt til at vurdere fem forskellige sandsynlighedsfordelinger i forhold til de målte lufttemperaturer i de 59 danske husstande, hvilket drejer sig om en normal-, lognormal-, eksponential-, weibull- og gammafordeling. Fælles for alle de fem typer af sandsynlighedsfordelinger er, at fordelingsparametrene er fastlagt på baggrund af de 59 tilgængelige temperaturmålinger. Det betyder eksempelvis, at nulhypotesen for hypotesetesten af normalfordelingen går ud på, at temperaturmålingerne kan beskrives med en normalfordeling med en middelværdi og standardafvigelse, der er beregnet ud fra målingerne. Samtlige nulhypoteser er formuleret således, at hvis de accepteres, så kan temperaturmålingerne beskrives med den aktuelle sandsynlighedsfordeling. Til hypotesetestene er der anvendt et signifikansniveau på 0,05, hvilket betyder, at hvis nulhypoteserne accepteres, kan det med 95 % sikkerhed godtgøres, at målingerne kan beskrives med den sandsynlighedsfordeling og de fordelingsparametre, der er defineret i nulhypotesen. Resultaterne for hypotesetestene er vist i tabel J.1, hvor det kan ses, at der er flere forskellige sandsynlighedsfordelinger, der kan anvendes til beskrivelse af temperaturmålingerne for de 59 danske husstande ².

²Beregningerne kan ses i regnearket *Sandsynlighedsfordelinger\Lufttemperatur\003 DANMARK hypotesetest.xlsx* på den medfølgende DVD.

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	0,71	5,99	Middelværdi: 21,23 Standardafv.: 1,31 Nulhypotese accepteres
Lognormalfordeling	1,22	5,99	Log. middelværdi: 3,05 Log. standardafv.: 0,06 Nulhypotese accepteres
Eksponentialfordeling	518,13	3,84	Middelværdi: 21,23 Nulhypotese forkastes
Weibullfordeling	0,44	5,99	Skala-parameter: 21,83 Form-parameter: 18,21 Nulhypotese accepteres
Gammafordeling	1,01	5,99	Alfa-værdi: 261,61 Beta-værdi: 0,08 Nulhypotese accepteres

Tabel J.1: Hypotesetest af sandaynlighedsfordelinger i forhold til temperaturmålinger for 59 danske husstande, hvor nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

Der er tilsvarende ovenstående indsamlet temperaturmålinger fra et projekt af *Feist et al.* [17], hvor der er målinger af lufttemperaturen i 108 passivhuse fra Tyskland, Schweiz og Østrig³. Histogrammet over målingerne kan ses på figur J.2, hvor det fremgår, at de fleste lufttemperaturer ligger i intervallet fra 21 – 22 °C. Udeklimaet, som passivhusene er opført i, kan have indflydelse på lufttemperaturen i husstandene, hvor målingerne dog er, ligesom for de danske boliger, fra opvarmningssæsonen. I appendiks E kan der findes en sammenligning af udeklimaet for passivhusene og det danske klima.



Figur J.2: Histogram og fordelingsfunktion for lufttemperaturer i 108 passivhuse fra Tyskland, Schweiz og Østrig.

³En sammenfatning af lufttemperaturerne kan ses i regnearket *Sandsynlighedsfordelinger\ Lufttemperatur\ Opsummering af indetemperature.xls* på den medfølgende DVD.

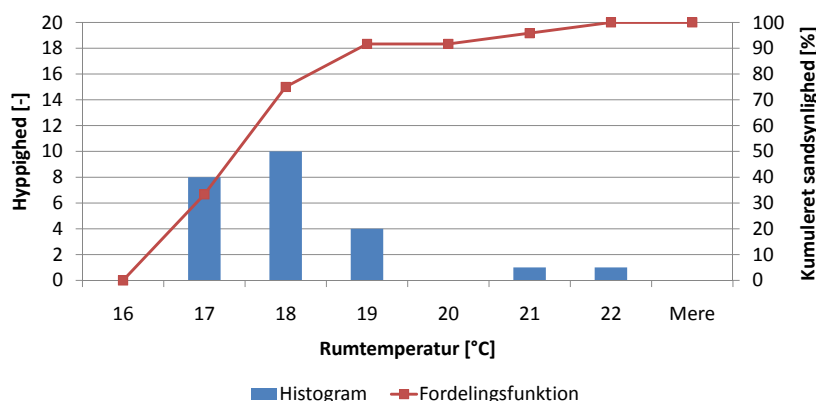
På samme måde som for målingerne i de danske husstande, er der udført hypotesetests af fem forskellige sandsynlighedsfordelinger, hvor det viser sig, at det kun er hypotesetesten med en weibullfordeling, der accepteres ⁴. Resultaterne af hypotesetestene for de 108 udenlandske husstande kan ses i tabel J.2.

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	8,76	7,81	Middelværdi: 21,70 Standardafv.: 1,52 Nulhypotese forkastes
Lognormalfordeling	11,12	7,81	Log. middelværdi: 3,07 Log. standardafv.: 0,07 Nulhypotese forkastes
Ekspontentialfordeling	908,30	5,99	Middelværdi: 21,70 Nulhypotese forkastes
Weibullfordeling	3,01	7,81	Skala-parameter: 22,39 Form-parameter: 16,78 Nulhypotese accepteres
Gammafordeling	10,29	7,81	Alfa-værdi: 202,66 Beta-værdi: 0,11 Nulhypotese forkastes

Tabel J.2: Hypotesetest af sandsynlighedsfordelinger i forhold til temperaturmålinger for 108 tyske, schweiziske og østrigske husstande, hvor nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

For et hollandsk projekt er der foretaget temperaturmålinger i 30 husstande, hvor udeklimaet for disse er sammenholdt med det danske i appendiks E. De målte lufttemperaturer er i rapporten fra *Maeyens et al.* [42] angivet som middeltemperaturer på månedsbasis. Det fremgår af rapporten, at der i nogle husstande er målt middeltemperaturer under 16°C , men det er ikke beskrevet hvorvidt de pågældende boliger er beboet eller ej. I det nærværende projekt er målinger med middeltemperaturer under 16°C udelukket fra de efterfølgende hypotesetests, da de anses som værende ubeboede. Det resulterer i, at der anvendes målinger fra 24 husstande, hvor et histogram af målingerne kan ses af figur J.3.

⁴Beregningerne kan ses i regnearket *Sandsynlighedsfordelinger\Lufttemperatur\001 CEPHEUS hypotesetest.xlsx* på den medfølgende DVD.



Figur J.3: Histogram og fordelingsfunktion for lufttemperaturer i 24 hollandske husstande.

Ud fra samme fremgangsmåde, som for de øvrige målinger, er der foretaget en hypotesetest af de 24 hollandske målinger. Det viser sig, at alle de opstillede hypotesetests bliver forkastet, hvor resultaterne kan ses i tabel J.3 ⁵. I rapporten fra *Maeyens et al.* er der beskrevet, at der er foretaget en *Chi-square test for goodness of fit* på målingerne, hvor det viser sig, at en weibullfordeling med en skala-parameter på 17,888 og en form-parameter på 7,8634 kan beskrive målingerne. I den forbindelse bør det dog nævnes, at den pågældende hypotesetest inddrager alle 30 husstande.

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	—	—	Beregning kan ikke gennemføres Antal frihedsgrader: 0 Nulhypotese forkastes
Lognormalfordeling	—	—	Beregning kan ikke gennemføres Antal frihedsgrader: 0 Nulhypotese forkastes
Ekspontentialfordeling	—	—	Beregning kan ikke gennemføres Antal frihedsgrader: 0 Nulhypotese forkastes
Weibullfordeling	7,79	3,84	Skala-parameter: 18,49 Form-parameter: 12,56 Antal frihedsgrader: 1 Nulhypotese forkastes
Gammafordeling	—	—	Beregning kan ikke gennemføres Antal frihedsgrader: 0 Nulhypotese forkastes

Tabel J.3: Hypotesetest af sandaynlighedsfordelinger i forhold til temperaturmålinger for 24 hollandske husstande, hvor nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

⁵Beregningerne kan ses i regnearket *Sandsynlighedsfordelinger\Lufttemperatur\002 HOLLAND hypotesetest.xlsx* på den medfølgende DVD.

Af ovenstående kan det diskuteres hvorvidt husstande med en månedlig middeltemperatur på $16,5^{\circ}\text{C}$ er beboet eller ej. Samtidig kan det i dette projekt ikke eftervises, at de fem afprøvede sandsynlighedsfordelinger kan beskrive målingerne fra de hollandske husstande. Det betyder, at det i dette projekt er valgt at set bort fra temperaturmålingerne fra de 24 hollandske husstande, der er nævnt i forbindelse med figur J.3, da de ligger langt fra hvad der kan forventes i danske boliger. Til sammenligning kan sandsynlighedsfordelingerne for de 59 danske husstande og de 108 tyske, schweiziske og østrigske husstande ses på figur J.5 og J.6. Figurerne viser, at sandsynlighedsfordelingerne for de danske husstande og for de tyske, schweiziske og østrigske husstande mere sammenfaldende. Der er dog en tendens til, at danske husstande har den højeste frekvens ved en lidt lavere lufttemperatur end de udenlandske.

Der er lavet en samling af de målte lufttemperaturer for 59 danske husstande og de 108 tyske, schweiziske og østrigske husstande, hvor det er valgt at set bort fra temperaturmålingerne fra de 24 hollandske husstande ⁶. Ud fra denne samling af lufttemperaturer er der foretaget hypotesetests på samme måde som de tidligere nævnte. Hypotesetestene viser, at ud af alle de opstillede hypotesetests er det kun weibullfordelingen, der kan accepteres, hvilket er vist i tabel J.4 ⁷.

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	10,03	9,49	Middelværdi: 21,53 Standardafv.: 1,46 Nulhypotese forkastes
Lognormalfordeling	14,80	9,49	Log. middelværdi: 3,07 Log. standardafv.: 0,07 Nulhypotese forkastes
Eksponentialfordeling	1597,82	7,81	Middelværdi: 21,53 Nulhypotese forkastes
Weibullfordeling	2,28	9,49	Skala-parameter: 22,20 Form-parameter: 16,78 Nulhypotese accepteres
Gammafordeling	13,26	9,49	Alfa-værdi: 214,94 Beta-værdi: 0,10 Nulhypotese forkastes

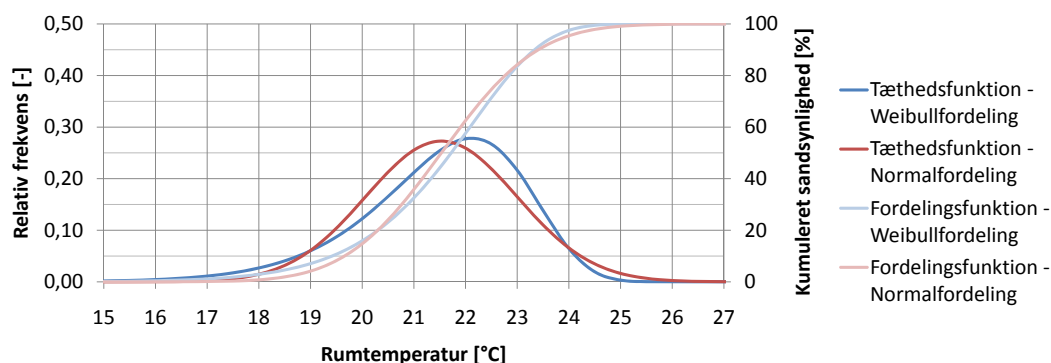
Tabel J.4: Hypotesetest af sandsynlighedsfordelinger i forhold til samlingen af temperaturmålinger for i alt 167 husstande fra Tyskland, Schweiz, Østrig og Danmark. Nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

Det bør dog bemærkes, at selvom normalfordelingen forkastes, sker det med en lille forskel mellem χ^2 og *kritisk* χ^2 . Årsagen til at normalfordelingen forkastes er primært en lavere

⁶En sammenfatning af lufttemperaturerne kan ses i regnearket *Sandsynlighedsfordelinger\Lufttemperatur\Opsumming af indetemperaturer.xlsx* på den medfølgende DVD.

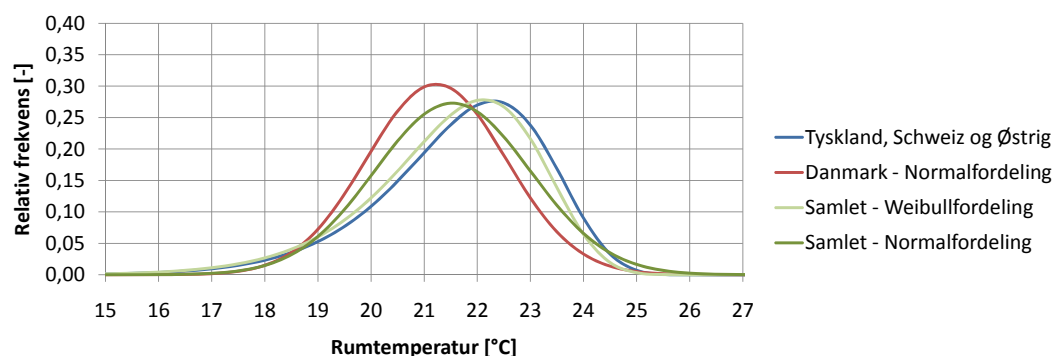
⁷Beregningerne kan ses i regnearket *Sandsynlighedsfordelinger\Lufttemperatur\004 SAMLET hypotesetest.xlsx* på den medfølgende DVD.

forventet hyppighed sammenholdt med den observerede hyppighed ved temperaturerne i intervallet 18 – 19 °C. På figur J.4 er der vist en sammenligning over de to sandsynlighedsfordelinger, der er optegnet ud fra informationerne i tabel J.4.

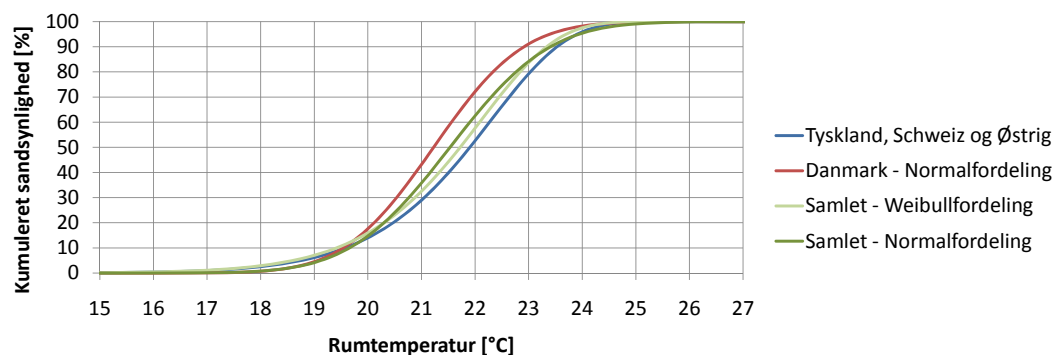


Figur J.4: Tætheds- og fordelingsfunktion for henholdsvis en normal- og weibullfordeling for 167 husstande fra Tyskland, Schweiz, Østrig og Danmark.

Der kan foretages en sammenligning af sandsynlighedsfordelingerne med de forskellige datagrundlag på figur J.5 og J.6.



Figur J.5: Tæthedsfunktion for henholdsvis de danske, husstandene fra Tyskland, Schweiz og Østrig samt samlingen af disse temperaturer.



Figur J.6: Fordelingsfunktion for henholdsvis de danske, husstandene fra Tyskland, Schweiz og Østrig samt samlingen af disse temperaturer.

Resultaterne af den samlede sandsynlighedsfordeling viser, at der er størst sandsynlighed for, at lufttemperaturen i boligerne er mellem $21 - 23\text{ }^{\circ}\text{C}$. Dette stemmer godt overens med hvad der er vist i litteraturstudiet i afsnit 3.2.1.3 og 3.2.3 i hovedrapporten. Sammenlignes resultaterne ydermere med den danske lovgivning, der beskriver, at der i vinterperioden skal kunne opretholdes lufttemperaturer på mellem $20 - 24\text{ }^{\circ}\text{C}$, hvilket er beskrevet i afsnit 3.2.1.1 i hovedrapporten, stemmer disse temperaturspænd godt overens. Det samme gælder for en sammenligning mellem resultaterne af den ovenstående hypotesetest og komforttemperaturerne, der er nævnt i tabel C.11 afsnit C.3 i appendiks C, som ligger i intervallet mellem omkring $19 - 24\text{ }^{\circ}\text{C}$.

Vurdering af elforbrug

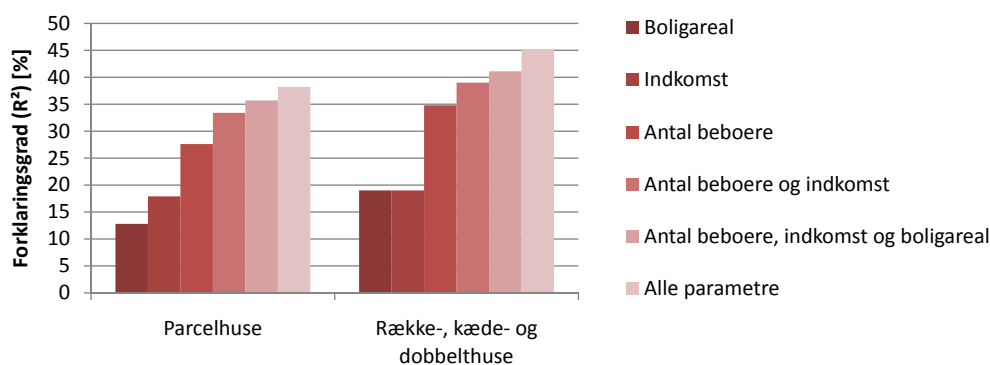
Til vurdering af boligernes samlede energiforbrug er det nødvendigt at estimere elforbrugets størrelse i de enkelte boliger. I afsnit 3.3 i hovedrapporten er det vist, at elforbrugets størrelse varierer meget fra bolig til bolig. Til at vurdere inden for hvilket interval elforbruget i boliger kan forventes at variere, opstilles der nogle sandsynlighedsfordelinger for dette. Datagrundlaget bag og opstillingen af sandsynlighedsfordelingerne beskrives nærmere i dette appendiks.

Der er anvendt oplysninger om elforbrugets størrelse, der stammer fra rapporterne af *Gram-Hanssen et al.* [26], *Jesper Ole Jensen* [35], *Casper Kofod* [37] og *Energi Nord* [51], hvor nærmere oplysninger om disse kan findes i afsnit 3.3 i hovedrapporten.

K.1 Valg af normeringsenhed

De tidligere nævnte rapporter, der er opnået kendskab til gennem litteraturstudiet, opgør hovedsagligt elforbruget ud fra husstandsstørrelse. *Jesper Ole Jensen* angiver, i den førnævnte rapport, elforbruget per person for henholdsvis en-, to-, tre- og firepersoners husstande samt et for husstande med fem eller flere beboere. I *Casper Kofods* rapport opgives husstandes samlede elforbrug, men hvor disse er holdt adskilt for henholdsvis en-, to-, tre-, fire-, fem- og sekspersoners husstande. Hvad angår *Energi Nord*, oplyser de nogle kategorier for husstandes samlede elforbrug, afhængigt af om husstandsstørrelsen er på en, to, tre, fire eller fem personer og derover. *Gram-Hanssen et al.* opgør i sin rapport elforbruget ud fra forskellige parametre, hvor det er undersøgt hvilke parametre, der har den højeste forklaringsgrad i forhold til variationen af det samlede elforbrug. Når forskellige parametre, så som boligareal, antal beboere, indkomst mv., enkeltvis sammenholdes med det samlede elforbrug for de enkelte boliger, fremgår det af *Gram-Hanssen et al.*, at den parameter, der har den højeste forklaringsgrad for variationen af det samlede elforbrug, er antallet af beboere. Det fremgår yderligere af det førnævnte projekt, at den næstvigtigste parameter er husstandens indkomst. En sammenligning af forklaringsgraden for de tre enkelte parametre samt kombinationer heraf kan ses på figur K.1.

K.2. Vurdering af elforbrugets udvikling samt betydningen heraf



Figur K.1: Forklaringsgrad når henholdsvis boligareal, indkomst, antal beboere og når disse kombineres eller når 12 forskellige parametre sammenholdes med forbrugsmålinger af det samlede elforbrug i forskellige husstande [26].

På figur K.1 fremgår betegnelsen *alle parametre*, hvor dette dækker over følgende forhold: boligens størrelse og opførelsesår, antal badeværelser, ejerforhold, antal beboere, indkomst, antal børn under 7 år, antal børn mellem 7-12 år, antal teenagere, alderen på den ældste beboer i husstanden, husstandens højeste uddannelse samt statsborgerskab.

Oplysninger om husstandes samlede elforbrug fra *Jesper Ole Jensen*, *Casper Kofod* og *Energi Nord* er alle opgjort ud fra husstandsstørrelser, og af kilderne fremgår der ikke direkte oplysninger, som kan anvendes til at opgøre elforbruget ud fra andre normeringsenheder. Da oplysningerne fra projektet af *Gram-Hanssen et al.* peger i retning af, at antallet af personer i husstandene er den enkeltstående parameter, der har den højeste forklaringsgrad, vil husstandes elforbrug hovedsagligt opgøres ud fra forskellige husstandsstørrelser i det efterfølgende.

K.2 Vurdering af elforbrugets udvikling samt betydningen heraf

Nogle af undersøgelserne, der er opnået kendskab til gennem litteraturstudiet, er foretaget med flere års mellemrum, hvilket kan have en betydning for størrelsen af de målte elforbrug i boligerne, når disse sammenholdes. For at undersøge det nærmere er der indsamlet oplysninger om udvikling i det samlede elforbrug, hvor informationer om denne udvikling er indsamlet fra *Energistyrelsen* [14]. På figur 2.5 på side 17 i hovedrapporten er der vist en graf over udviklingen af det samlede elforbrug til husholdninger. Af denne kan det ses, at elforbruget har været forholdsvis stabilt gennem de seneste 20 år. En nærmere undersøgelse af oplysningerne, der ligger til grund for grafen, viser, at elforbruget er faldet med 2,8 % fra 1999 til 2009. Hvis oplysninger fra 1989 og frem til 2009 anvendes, kan der ses et fald på 0,2 %. Denne udvikling vurderes ikke at have betydning for elforbruget, når kilder, der er skrevet med flere års mellemrum, sammenholdes. Derfor vil der i det efterfølgende ikke blive korigeret for elforbrugets udvikling.

K.3 Test af fordelingstype

Til at vurdere hvilken sandsynlighedsfordeling der kan beskrive det samlede elforbrug, udføres der hypotesetests af nogle tilgængelige målinger fra en rapport af *Jesper Ole Jensen* [35]. I forbindelse med udarbejdningen af førnævnte rapport er der foretaget en registrering af elforbruget i nogle boliger, hvor dele af disse målinger fremgik af *Jesper Ole Jensens* rapport ¹. Det drejer sig om elforbruget for i alt 244 husstande, hvor der primært er tale om rækkehuse fra 1960-70'erne. Disse oplysninger er anvendt i en *Chi-square test for goodness of fit* eller en *Kolmogorov-smirnov one-sample test*, afhængigt af antallet af målinger, hvor nærmere oplysninger om disse hypotesetests er beskrevet i appendiks I. Det bør bemærkes, at elforbruget i disse hypotesetests er opgjort per person. Resultaterne fra de første fire hypotesetests, der er vist i tabel K.1 og frem til tabel K.4, er udført som *Chi-square test for goodness of fit*. Den sidste hypotesetest, vist i tabel K.5, er udført som en *Kolmogorov-Smirnov one-sample test* ².

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	3,99	5,99	Middelværdi: 2184,41 Standardafv.: 1136,77 Nulhypotese accepteres
Lognormalfordeling	0,75	3,84	Log. middelværdi: 7,57 Log. standardafv.: 0,482 Nulhypotese accepteres
Eksponentialfordeling	21,47	7,81	Middelværdi: 2184,41 Nulhypotese forkastes
Weibullfordeling	4,24	7,81	Skala-parameter: 2476,40 Form-parameter: 2,10 Nulhypotese accepteres
Gammafordeling	1,52	5,99	Alfa-værdi: 4,51 Beta-værdi: 484,79 Nulhypotese forkastes

Tabel K.1: Hypotesetest af sandsynlighedsfordelinger i forhold til målinger af elforbruget i husstande med én beboer, hvor forbrugsmålingerne er aflæst i rapporten fra *Jesper Ole Jensen* [35]. Nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

¹Yderligere oplysninger kan ses i regnearket *Sandsynlighedsfordelinger\Elforbrug\Oplysninger til elforbrugets størrelse (JOJ).xlsx* på den medfølgende DVD.

²Yderligere oplysninger om disse hypotesetests kan ses i regnearkene fra *Sandsynlighedsfordelinger\Elforbrug\001 JOJ hypotesetest.xlsx* og frem til *005 JOJ hypotesetest.xlsx* på den medfølgende DVD.

K.3. Test af fordelingstype

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	28,79	14,07	Middelværdi: 1759,48 Standardafv.: 772,59 Nulhypotese forkastet
Lognormalfordeling	8,68	11,07	Log. middelværdi: 7,401 Log. standardafv.: 0,366 Nulhypotese accepteres
Ekspontentialfordeling	116,94	15,51	Middelværdi: 1759,48 Nulhypotese forkastes
Weibullfordeling	29,56	14,07	Skala-parameter: 1983,90 Form-parameter: 2,30 Nulhypotese forkastet
Gammafordeling	8,30	11,07	Alfa-værdi: 7,09 Beta-værdi: 248,16 Nulhypotese accepteres

Tabel K.2: Hypotesetest af sandsynlighedsfordelinger i forhold til målinger af elforbruget i husstande med to beboere, hvor forbrugsmålingerne er aflæst i rapporten fra *Jesper Ole Jensen* [35]. Nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	6,10	7,81	Middelværdi: 1316,27 Standardafv.: 394,55 Nulhypotese accepteres
Lognormalfordeling	2,03	5,99	Log. middelværdi: 7,176 Log. standardafv.: 0,286 Nulhypotese accepteres
Ekspontentialfordeling	80,81	9,49	Middelværdi: 1361,27 Nulhypotese forkastes
Weibullfordeling	2,79	5,99	Skala-parameter: 1507,90 Form-parameter: 3,70 Nulhypotese accepteret
Gammafordeling	2,24	5,99	Alfa-værdi: 12,64 Beta-værdi: 107,71 Nulhypotese accepteres

Tabel K.3: Hypotesetest af sandsynlighedsfordelinger i forhold til målinger af elforbruget i husstande med tre beboere, hvor forbrugsmålingerne er aflæst i rapporten fra *Jesper Ole Jensen* [35]. Nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	4,92	5,99	Middelværdi: 1139,43 Standardafv.: 328,34 Nulhypotese accepteres
Lognormalfordeling	3,58	5,99	Log. middelværdi: 6,995 Log. standardafv.: 0,311 Nulhypotese accepteres
Eksponentialfordeling	115,53	11,07	Middelværdi: 1139,43 Nulhypotese forkastes
Weibullfordeling	7,56	5,99	Skala-parameter: 1258,90 Form-parameter: 3,60 Nulhypotese forkastet
Gammafordeling	2,82	5,99	Alfa-værdi: 11,67 Beta-værdi: 97,63 Nulhypotese accepteres

Tabel K.4: Hypotesetest af sandsynlighedsfordelinger i forhold til målinger af elforbruget i husstande med fire beboere, hvor forbrugsmålingerne er aflæst i rapporten fra *Jesper Ole Jensen* [35]. Nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

Sandsynlighedsfordeling	KS -værdi	Kritisk KS_α -værdi	Bemærkning
Normalfordeling	0,235	0,391	Middelværdi: 1097,44 Standardafv.: 385,15 Nulhypotese accepteres
Lognormalfordeling	0,187	0,391	Log. middelværdi: 6,95 Log. standardafv.: 0,31 Nulhypotese accepteres
Weibullfordeling	0,228	0,391	Skala-parameter: 1227,10 Form-parameter: 3,10 Nulhypotese accepteres
Gammafordeling	0,203	0,391	Alfa-værdi: 10,61 Beta-værdi: 103,46 Nulhypotese accepteres

Tabel K.5: Hypotesetest af sandsynlighedsfordelinger i forhold til målinger af elforbruget i husstande med fem beboere, hvor forbrugsmålingerne er aflæst i rapporten fra *Jesper Ole Jensen* [35]. Nulhypotesen accepteres såfremt $KS < \text{kritisk } KS_\alpha$.

Foruden ovenstående er der foretaget hypotesetests af målinger for elforbruget for ni husstande, hvor der er anvendt oplysninger fra en tidligere omtalt rapport af *Knudsen et al.*

[36]. Resultaterne af disse hypotesetests viser, at hverken en normal-, lognormal-, weibull- eller gammafordeling kan afvises ved et signifikansniveau på 0,05³, som vist i tabel K.6.

Sandsynlighedsfordeling	KS -værdi	Kritisk KS_α -værdi	Bemærkning
Normalfordeling	0,284	0,521	Middelværdi: 5341,67 Standardafv.: 2402,53 Nulhypotese accepteres
Lognormalfordeling	0,211	0,521	Log. middelværdi: 8,51 Log. standardafv.: 0,40 Nulhypotese accepteres
Weibullfordeling	0,266	0,521	Skala-parameter: 6033,10 Form-parameter: 2,60 Nulhypotese accepteres
Gammafordeling	0,227	0,521	Alfa-værdi: 7,07 Beta-værdi: 755,07 Nulhypotese accepteres

Tabel K.6: Hypotesetest af sandsynlighedsfordelinger i forhold til målinger af elforbruget i ni husstande, hvor forbrugsmålingerne er aflæst i rapporten fra *Knudsen et al.* [36]. Nulhypotesen accepteres såfremt $KS < \text{kritisk } KS_\alpha$.

Det kan ses af de ovenstående hypotesetests, at normalfordelingen forkastes i ét af tilfældene, og at lognormalfordelingen er den eneste sandsynlighedsfordeling, der accepteres i alle de udførte tests. Det vurderes, at en lognormalfordeling vil vise fornuftige resultater i forbindelse med husstandes elforbrug, hvorfor denne sandsynlighedsfordeling vil blive anvendt i det efterfølgende. Parametrene til denne fordelingsstype bliver beskrevet nærmere i det efterfølgende afsnit.

K.4 Estimering af fordelingsparametre

Det er fastlagt, at der anvendes en lognormalfordeling til vurdering af det husstandes samlede elforbrug. Dertil skal der estimeres en middelværdi og standardafvigelse for denne fordeling, hvilket gøres på baggrund af oplysninger fra rapporten af *Gram-Hanssen et al.* [26]. Oplysningerne fra denne rapport anvendes, da disse stammer fra omkring 8500 husstande, hvilket er betydeligt mere end de tilgængelige oplysninger fra de tidligere omtalte rapporter. Data fra rapporten af *Gram-Hanssen et al.* er oplyst som elforbrug ved 5%-, 25%-, 50%-, 75%- og 95%-fraktiler. Disse oplysninger er anvendt til at estimere hvilken middelværdi og standardafvigelse, der kan beskrive forbruget bedst muligt. De estimerede parametre er anvendt i henholdsvis en normal- og en lognormalfordeling. Resultaterne viser, ligesom ved

³Yderligere oplysninger kan ses i regnearket *Sandsynlighedsfordelinger\Elforbrug\006 FremParc hypotesetest.xlsx* på den medfølgende DVD.

de udførte hypotesetests, at lognormalfordelingen beskriver elforbruget bedst ⁴. Parametrene fra estimeringen er angivet i tabel K.7.

Antal beboere [person]	Middelværdi [kWh/husstand per år]	Standardafvigelse [kWh/husstand per år]
1	2469	1270
2	3485	1279
3	4523	1588
4	5035	1540
5	5619	1804
6	6709	2139

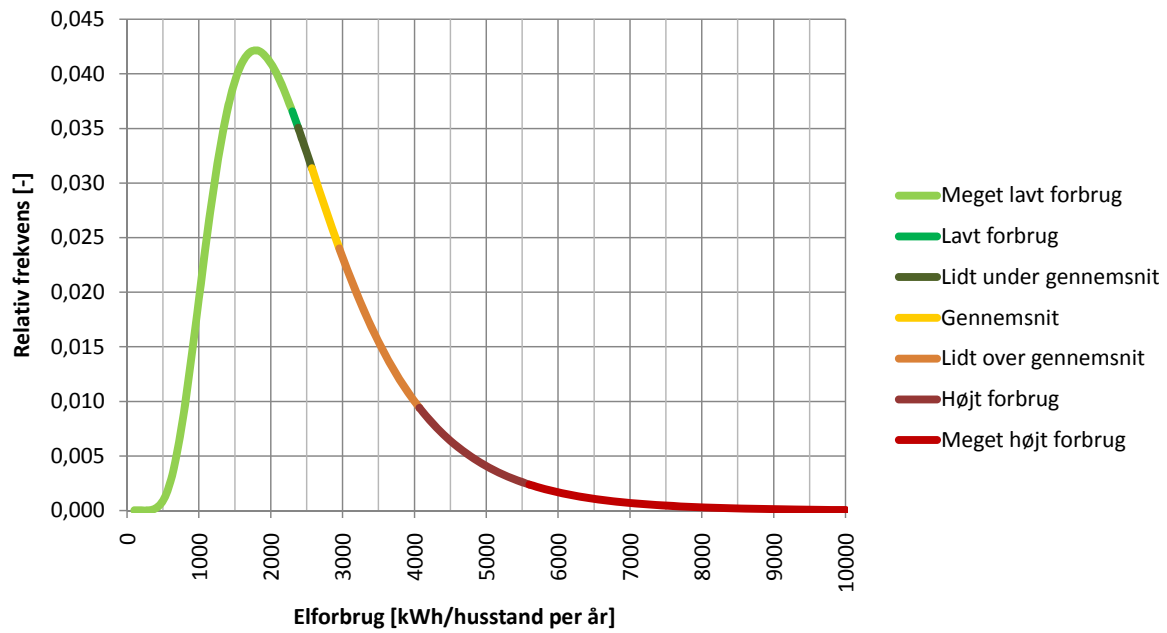
Tabel K.7: Anvendte middelværdier og standardafvigelser til lognormalfordelinger for husstandes samlede elforbrug estimeret på baggrund af oplysninger fra rapport af *Gram-Hanssen et al.* [26].

Da ovenstående parametre blot er estimeret, er der foretaget en sammenligning af disse med oplysninger fra andre kilder. Resultaterne af sammenligningen kan ses i tabel K.8, samt på figur K.2 til K.7, hvor også intervaller fra *Energi Nord* er indtegnet.

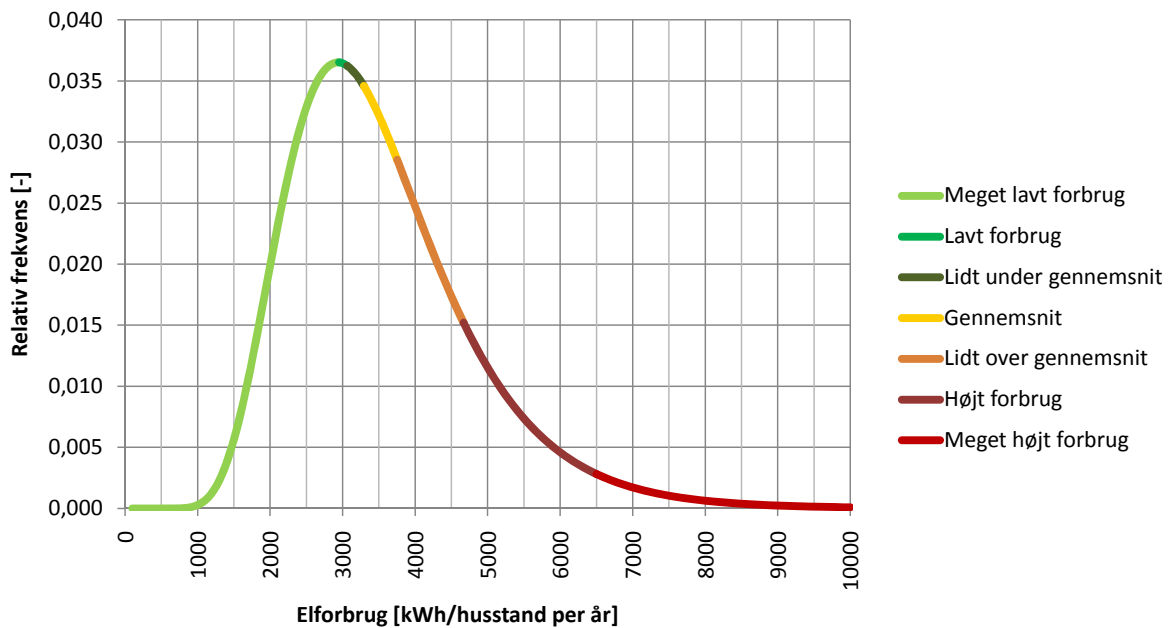
⁴Yderligere oplysninger kan ses i regnearket *Sandsynlighedsfordelinger\Elforbrug\Sandsynlighedsfordeling ud fra fraktiler (KGH).xlsx* på den medfølgende DVD.

Kilde	Minimum [kWh/år]	Middel [kWh/år]	Maksimum [kWh/år]
Boliger med 1 beboer			
Jesper Ole Jensen	789, 5	2106, 0	6432, 3
Casper Kofod	1404, 9	1818, 2	2148, 8
Energi Nord		2572, 0 – 2954, 0	
Boliger med 2 beboere			
Jesper Ole Jensen	1415, 4	3519, 0	10496, 1
Casper Kofod	1239, 7	3002, 8	5537, 2
Energi Nord		3294, 0 – 3752, 0	
Boliger med 3 beboere			
Jesper Ole Jensen	1901, 4	4083, 8	7044, 0
Casper Kofod	1983, 5	3973, 3	8760, 3
Energi Nord		4040, 0 – 4520, 0	
Boliger med 4 beboere			
Jesper Ole Jensen	1119, 7	4557, 7	8838, 5
Casper Kofod	2727, 3	4549, 2	6776, 9
Energi Nord		4677, 0 – 5181, 0	
Boliger med 5 beboere			
Jesper Ole Jensen	3609, 3	5486, 9	9915, 0
Casper Kofod	3471, 1	4451, 0	5619, 8
Energi Nord		5178, 0 – 5695, 0	

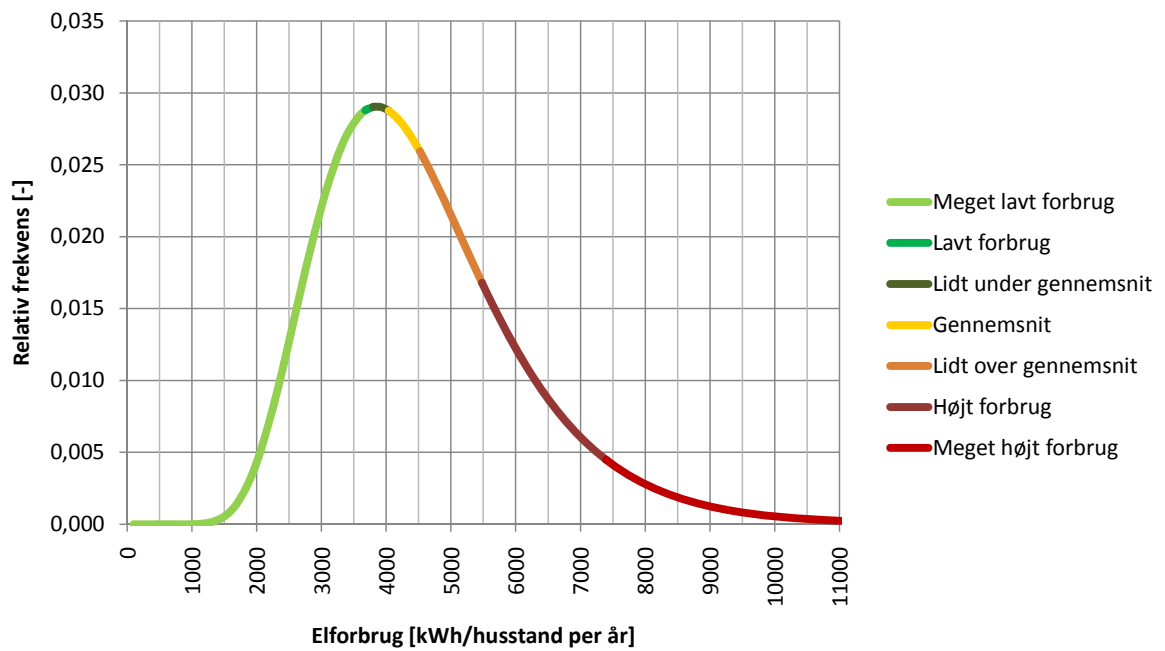
Tabel K.8: Oplysninger om elforbrug i boliger for forskellige kilder, hvor der fremgår minimums-, middel- og maksimumsforbrug fra tre forskellige projekter [26, 35, 37].



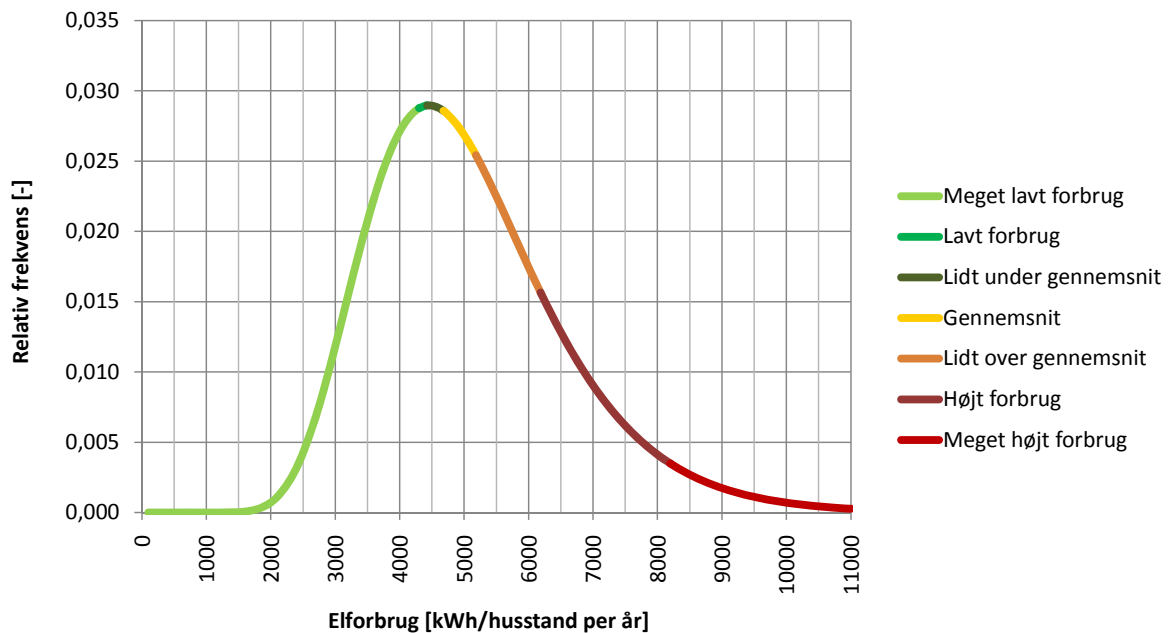
Figur K.2: Tæthedsfunktion for en lognormalfordeling af elforbrug i husstande med én beboer, hvor også intervaller fra *Energi Nord* [51] er indtegnet. Fordelingsparametrene er angivet i tabel K.7.



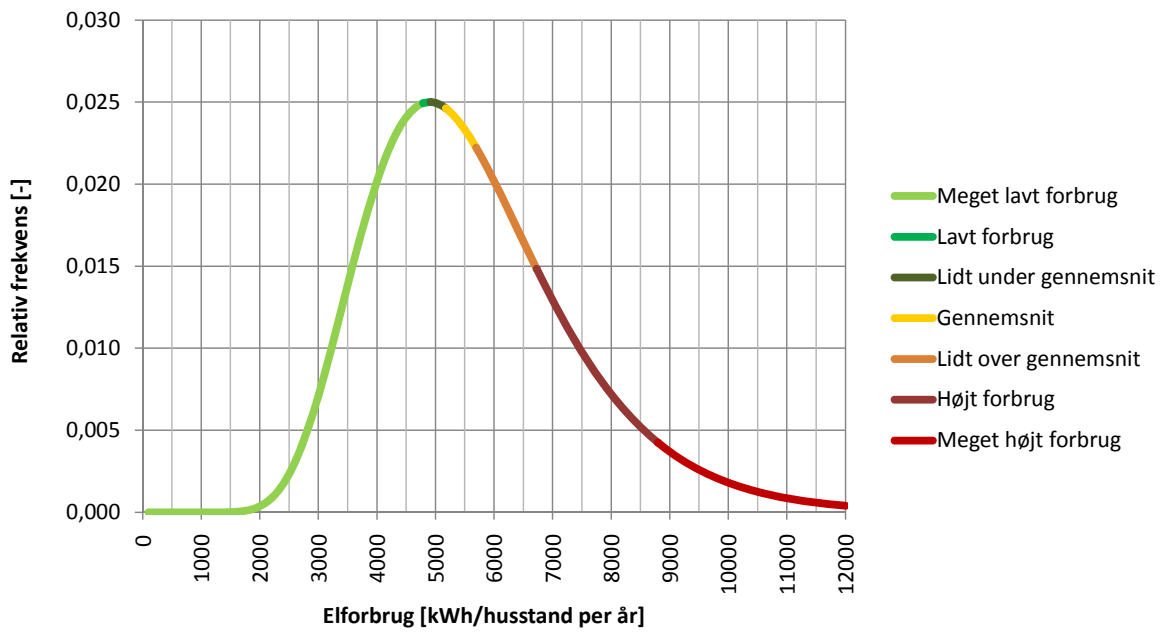
Figur K.3: Tæthedsfunktion for en lognormalfordeling af elforbrug i husstande med to beboere, hvor også intervaller fra *Energi Nord* [51] er indtegnet. Fordelingsparametrene er angivet i tabel K.7.



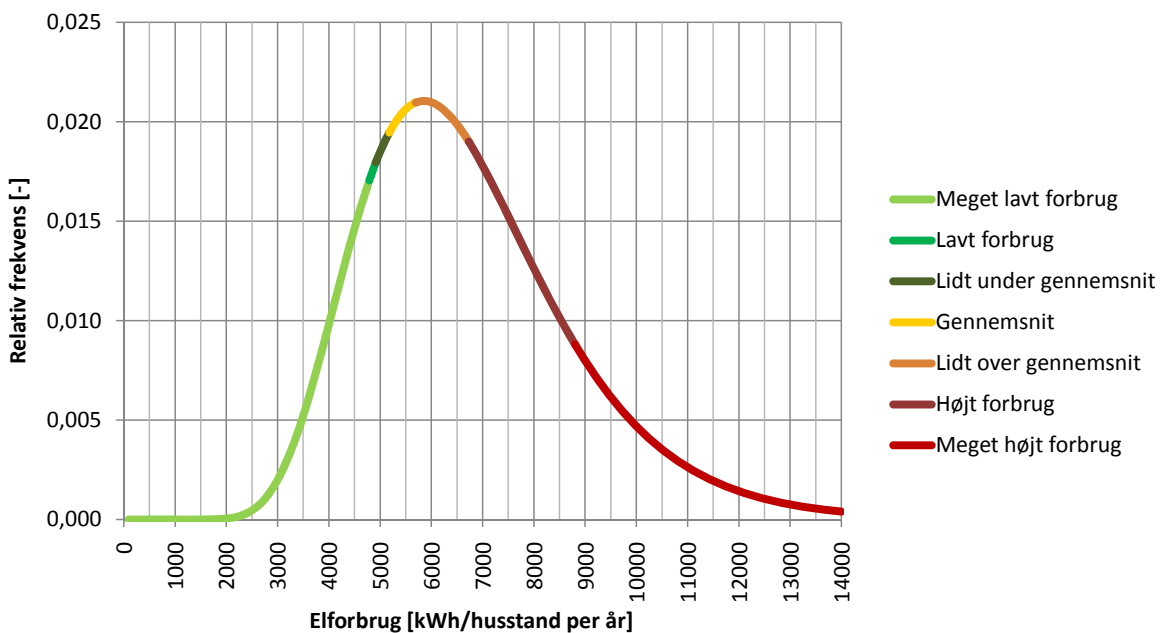
Figur K.4: Tæthedsfunktion for en lognormalfordeling af elforbrug i husstande med tre beboere, hvor også intervaller fra *Energi Nord* [51] er indtegnet. Fordelingsparametrene er angivet i tabel K.7.



Figur K.5: Tæthedsfunktion for en lognormalfordeling af elforbrug i husstande med fire beboere, hvor også intervaller fra *Energi Nord* [51] er indtegnet. Fordelingsparametrene er angivet i tabel K.7.



Figur K.6: Tæthedsfunktion for en lognormalfordeling af elforbrug i husstande med fem beboere, hvor også intervaller fra *Energi Nord* [51] er indtegnet. Fordelingsparametrene er angivet i tabel K.7.



Figur K.7: Tæthedsfunktion for en lognormalfordeling af elforbrug i husstande med seks beboere, hvor også intervaller fra *Energi Nord* [51] er indtegnet. Fordelingsparametrene er angivet i tabel K.7.

Det kan ses af figur K.2 til K.7, at de stemmer udmærket overens med oplysningerne fra tabel K.8. Derved er der opstillet sandsynlighedsfordelinger for elforbruget i husstande ud fra nogle lognormalfordelinger, med parametrene der er angivet i tabel K.7.

Vurdering af varmtvandsforbrug

Til vurdering af boligernes samlede energiforbrug er det nødvendigt at estimere forbruget af varmt brugsvand i de enkelte boliger, således energiforbruget til produktion af varmt brugsvand kan beregnes. Det har gennem litteraturstudiet vist sig vanskeligt at indsamle erfaringer om og måledata af varmtvandsforbruget og dets variation for danske boliger. Det er ligeledes begrænset hvilke undersøgelser, der findes om brugeradfærd i forbindelse med varmtvandsforbruget i boliger. Hovedparten af de undersøgelser, der er opnået kendskab til gennem litteraturstudiet, beskæftiger sig med husstandes samlede vandforbrug og hvilke af beboernes adfærdsmæssige forhold, der kan forklare dele af den variation af forbruget, der kan observeres mellem forskellige husstande. Hvis der i første omgang fokuseres på husstandes samlede vandforbrug, kan der foretages en sammenligning af hvilke vandforbrug, der er oplyst fra de forskellige kilder i litteraturstudiet. Der er dog flere forhold, der vanskeliggør en direkte sammenligning af oplysningerne. Eksempler herpå kan være at nogle kilder påpeger en betydelig forskel på størrelsen af husstandes samlede vandforbrug opgjort per beboer, når forskellige husstandsstørrelser sammenholdes. Det tager nogle af de andre kilder ikke højde for, hvor de i stedet angiver ét fast vandforbrug per beboer uafhængigt af husstandsstørrelsen. På tilsvarende vis kan det være vanskeligt at vurdere, hvorvidt undersøgelsesens alder har en betydning, når der foretages en sammenligning af husstandes samlede vandforbrug mellem forskellige kilder, når disse undersøgelser i nogle tilfælde er foretaget med flere års mellemrum.

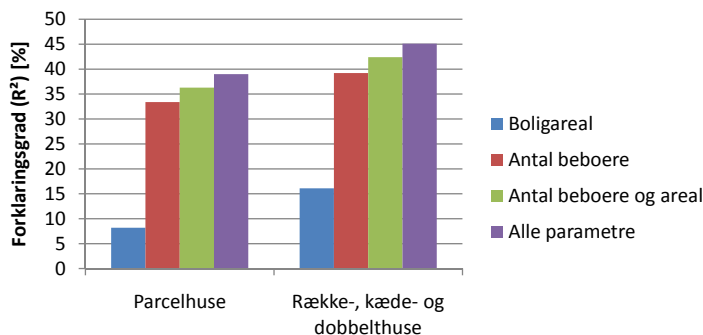
I dette appendiks præsenteres fremgangsmåden, der er anvendt i dette projekt, til at sammenholde oplysningerne fra flere af kilderne i litteraturstudiet. Formålet med dette appendiks er, at anvende eksisterende oplysninger, målinger og iagttagelser fra flere forskellige projekter og kombinere disse oplysninger således, at det munder ud i en sandsynlighedsfordeling for husstandes varmtvandsforbrug. Appendikset er opbygget således, at der tages udgangspunkt i husstandes samlede vandforbrug, når mængden af varmt brugsvand i de enkelte husstande skal estimeres. Denne opbygning skyldes, som tidligere nævnt, at hovedparten af kilderne, der er opnået kendskab til gennem litteraturstudiet, tager udgangspunkt i husstandes samlede vandforbrug og ikke direkte i varmtvandsforbruget. Til vurdering af en husstands samlede vandforbrug opstilles der flere forskellige sandsynlighedsfordelinger. Årsagen til hvorfor der opstilles flere forskellige sandsynlighedsfordelinger samt en redegørelse datagrundlaget bag disse beskrives nærmere i de efterfølgende afsnit. Eftersom det er husstandes samlede vandforbrug, der anvendes som indikator for forbruget af varmt brugsvand, vil det til sidst i dette kapitel diskuteres hvorledes størrelsen af varmtvandsforbruget kan estimeres ud fra husstandens samlede vandforbrug.

L.1 Husstandes samlede vandforbrug

I dette afsnit beskrives hvorledes oplysningerne fra litteraturstudiet, omhandlende det samlede vandforbrug i husstande, korrigeres, således de samlet set kan danne grundlaget for en sandsynlighedsfordeling, der beskriver størrelsen af dette forbrug. Oplysningerne fra de forskellige kilder kan først kombineres efter, der er foretaget en række væsentlige iagttagelser, vurdering og korrektioner. For at tydeliggøre disse er dette afsnit inddelt i en række underafsnit, der hver beskriver de enkelte vurderinger der er foretaget, før sammenligningen af det samlede vandforbrug kan finde sted.

L.1.1 Valg af normeringsenhed for det samlede vandforbrug

Gennem litteraturstudiet er der opnået kendskab til tidligere undersøgelser omhandlende vandforbrug i boliger. Hovedparten af undersøgelserne opgør det samlede vandforbrug med antallet af beboere i husstandene som normeringsenhed, hvilket vil sige, at undersøgelserne oplyser et vandforbrug per person. Det drejer sig om forbrugsoplysninger fra *Dansk Vand- og Spildevandsforening* [54], *Danmarks Statistik* [75], *Miljøstyrelsen et al.* [46], *Jesper Ole Jensen* [35] og til dels *Gram-Hanssen et al.* [26]. Sidstnævnte kilde har dog også foretaget en undersøgelse af hvilken normeringsenhed, der har den største forklaringskraft, når det samlede vandforbrug sammenholdes mellem flere forskellige husstande. Når forskellige parametre, så som boligareal, antal beboere, indkomst mv., enkeltvis sammenholdes med det samlede vandforbrug for de enkelte boliger, fremgår det af *Gram-Hanssen et al.*, at den parameter, der har det højeste forklaringsgrad for variationen af det samlede vandforbrug, er antallet af beboere. Det fremgår yderligere af det førnævnte projekt, at den næstvigtigste parameter er boligarealet. En sammenligning af forklaringsgraden for de to enkelte parametre kan ses på figur L.1 som henholdsvis *boligareal* og *antal beboere*.



Figur L.1: Forklaringsgrad når henholdsvis boligareal, antal beboere, de to førnævnte parametre eller når 13 forskellige parametre sammenholdes med forbrugsmålinger af det samlede vandforbrug i forskellige husstande [26].

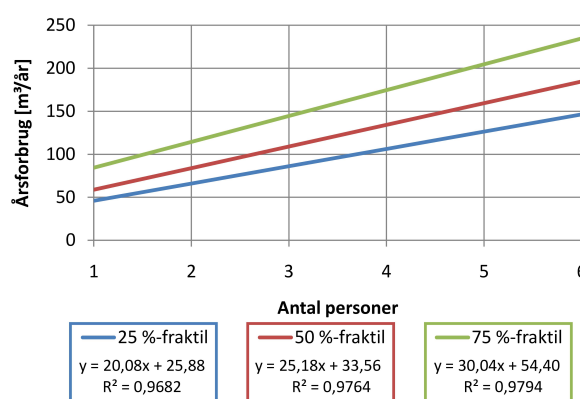
På figur L.1 fremgår betegnelsen *alle parametre*, hvor dette dækker over følgende forhold: boligens størrelse og opførelsesår, antal badeværelser og toiletter, ejerforhold, antal beboere, indkomst, antal børn under 7 år, antal børn mellem 7-12 år, antal teenagere, alderen på den ældste beboer i husstanden, husstandens højeste uddannelse samt statsborgerskab.

Oplysningerne om det samlede vandforbrug fra *Dansk Vand- og Spildevandsforening, Danmarks Statistik, Miljøstyrelsen et al.* og *Jesper Ole Jensen* er alle opgjort per person, og af kilderne fremgår der ikke direkte oplysninger, som kan anvendes til at opgøre vandforbruget ud fra andre normeringsenheder. Da oplysningerne fra projektet af *Gram-Hanssen et al.* peger i retning af, at antallet af personer i husstandene er den enkeltstående parameter, der har den højeste forklaringsgrad, vil denne normeringsenhed anvendes i det efterfølgende.

L.1.2 Korrektion af oplysninger fra kilde i litteraturstudiet

På trods af at det nu er fastlagt, at det samlede vandforbrug i husstande som udgangspunkt vurderes ud fra antallet af beboere, bliver der foretaget en korrektion af oplysninger fra en af kilderne i litteraturstudiet, før størrelsen af det forventede vandforbrug kan sammenholdes mellem de forskellige kilder.

En af de væsentligste undersøgelser omhandlende husstandes samlede vandforbrug er udarbejdet af *Miljøstyrelsen et al.*, der tidligere er beskrevet i afsnit 3.4 i hovedrapporten. På figur 3.23 på side 78 i hovedrapporten har *Miljøstyrelsen et al.* oplyst nogle formeludtryk, der kan anvendes til at skønne et forventet vandforbrug ud fra. Den førnævnte kilde giver udtryk for, at nogle af målingerne, der ligger til grund for de oplyste formler på figur 3.23, kan stamme fra flere boliger, hvilket kommer til udtryk ved, at figuren viser vandforbrug for op til 11 personer. Det betyder blandt andet, som det kan ses af figur 3.23 i hovedrapporten, at det samlede vandforbrug, beregnet med 75 %-fraktilen, i nogle tilfælde er lavere end vandforbruget beregnet ved 25 %- og 50 %-fraktilen. Tilsvarende har det skønnede vandforbruget, beregnet ved henholdsvis 25 %- og 50 %-fraktilen, tilnærmelsesvis samme størrelse ved husstande med én beboer. Ifølge indhentede oplysning fra *Danmarks Statistik* [73] har 4 ‰ af de danske husstande, svarende til omkring 10.000 husstande, en størrelse på over seks beboere. På baggrund af de førnævnte iagttagelser er det i dette projekt valgt, at omskrive formlerne på figur 3.23 i hovedrapporten, således de kun bygger på data fra husstandsstørrelser fra én til og med seks beboere. På baggrund af en lineær regression kommer de omskrevne formeludtryk da til at tage form, som vist på figur L.2 ¹.



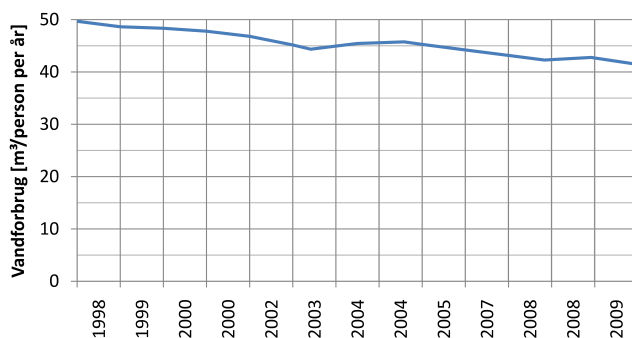
Figur L.2: Årligt vandforbrug som funktion af antal personer, gældende for parcel- og rækkehus, opstillet på baggrund af oplysninger fra *Miljøstyrelsen et al.* [46].

¹Yderligere oplysninger kan findes i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\Sammenligning af årligt vandforbrug.xlsx* på den medfølgende DVD.

Resultaterne af omskrivningen, der er vist på figur L.2, er en højere forklaringsgrad, beskrevet ved kvadrerede værdi af Pearsons korrelationskoefficient, R^2 . Ydermere er der en tydeligere adskillelse mellem vandforbruget beregnet på baggrund af henholdsvis 25 %-, 50 %- og 75 %-fraktilen. De originale formeludtryk fra *Miljøstyrelsen et al.*, angivet på figur 3.23 i hovedrapporten, anvendes ikke yderligere i dette appendiks, hvor der i det efterfølgende kun anvendes de omskrevne formeludtryk, der er angivet på figur L.2.

L.1.3 Vurdering af vandforbrugets udvikling samt betydningen heraf

Nogle af undersøgelserne, der er opnået kendskab til gennem litteraturstudiet, er foretaget med flere års mellemrum, hvilket kan have en betydning for størrelsen af de målte vandforbrug i boligerne, når disse sammenholdes. For at undersøge det nærmere er der indsamlet oplysninger om den gennemsnitlige udvikling i det samlede vandforbrug, hvor informationer omkring denne udvikling er indsamlet fra *Dansk Vand- og Spildevandsforening* [54]. Figur L.3 viser den gennemsnitlige udvikling af vandforbruget i danske husstande fra begyndelsen af 1998 og frem til slutningen af 2009. Oplysningerne på figuren stammer fra 146 vandforsyninger, der tilsammen forsyner over 2.575.000 indbygger med vand [54].



Figur L.3: Udvikling i det samlede vandforbrug i husholdninger delt ud per person fra begyndelsen af 1998 frem til slutningen af 2009 [54].

Den gennemsnitlige reduktion af vandforbruget, der kan ses på figur L.3, vurderes at have væsentlig betydning, når vandforbrugene målt med flere års mellemrum skal sammenlignes. Det betyder, at det data, som figur L.3 bygger på, er anvendt til at beregne størrelsen af den gennemsnitlige reduktion af det samlede vandforbrug. Denne reduktion er anvendt til at korrigere det vandforbrug, der er oplyst fra de enkelte litteraturkilder. Tabel L.1 viser hvilken reduktion der er anvendt, når vandforbruget fra enkelte kilder sammenlignes, hvor alle reduktionerne er beregnet som vandforbrugets udvikling fra det årstal, hvor de enkelte oplysninger stammer fra og frem til niveauet i slutningen af 2009.

Kilde	Årstal for data	Reduktion [%]
<i>Danmarks Statistik</i> [75]	2005	−8,45 %
<i>Miljøstyrelsen et al.</i> [46]	1999 – 2004 *	−12,07 %
<i>Jesper Ole Jensen</i> [35]	1998	−17,36 %
<i>Gram-Hanssen et al.</i> [26]	2000	−15,33 %

* Reduktionen er beregnet på baggrund af et gennemsnit af reduktionen for henholdsvis 1999 og 2004 og frem til slutningen af 2009.

Tabel L.1: Reduktionen af vandforbruget beregnet på baggrund af oplysninger om udviklingen af vandforbruget fra *Dansk Vand- og Spildevandsforening* [54].

I forbindelse med størrelsen af den beregnede reduktionen af de vandforbrug, der er oplyst fra de enkelte kilder, bør det dog bemærkes, at størrelsen af de reduktioner, der er angivet i tabel L.1, ikke nødvendigvis angiver den udvikling, der er sket i de husstande, hvor oplysningerne fra de enkelte kilder stammer fra.

L.1.4 Korrektion for sammenhængen mellem vandforbruget per person og husstandsstørrelsen

Undersøgelser fra *Miljøstyrelsen et al.* [46] og *Gram-Hanssen et al.* [26] viser begge, at det samlede vandforbrug opgjort per person er faldende ved stigende husstandsstørrelser. Oplysningerne fra *Miljøstyrelsen et al.* bygger på målinger i over 36.000 parcel- og rækkehuse og oplysningerne fra *Gram-Hanssen et al.* bygger på målinger i over 23.000 parcelhuse samt over 10.000 række-, kæde- og dobbelthuse. Ændringen i husstandens samlede vandforbrug opgjort per beboer, beregnet på baggrund af de to førnævnte kilder, er vist i tabel L.2.

Kilde	Ændring i husstandsstørrelse				
	1 til 2 [%]	2 til 3 [%]	3 til 4 [%]	4 til 5 [%]	5 til 6 [%]
<i>Miljøstyrelsen et al.</i> [46] *	−29,6	−14,1	−8,2	−5,4	−3,8
<i>Gram-Hanssen et al.</i> [26] **	−31,3	−15,3	−9,0	−6,0	−4,2
Middel	−30,5	−14,7	−8,6	−5,7	−4,0

* Reduktionen er beregnet som et middel af 25 %- , 50 %- og 75 %-fraktilen.

** Reduktionen er beregnet på som et middel af parcel-, række-, kæde- og dobbelthuse.

Tabel L.2: Reduktionen af vandforbruget per person fra ved en ændring af husstandsstørrelsen fra én til to beboere og så fremdeles.

På trods af at de to forskellige projekter, der er nævnt i tabellen, bygger på data fra forskellige husstande, kan der ses en ændring i vandforbruget opgjort per person, der har omtrent samme størrelse ². Idet begge kilder bygger på oplysninger fra en relativ stor mængde husstande, vil

²Yderligere oplysninger kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\ Sammenligning af årligt vandforbrug.xlsx* på den medfølgende DVD.

der i dette projekt tages højde for denne reduktion, når sammenligningen af vandforbruget fra flere forskellige kilder finder sted.

De ovenstående iagttagelser vil blive anvendt i forbindelse med data fra *Dansk Vand- og Spildevandsforening* [54] og *Danmarks Statistik* [75], eftersom disse kilder oplyser et vandforbrug opgjort per person uden hensyntagen til husstandsstørrelsen. Hvad angår oplysninger fra *Dansk Vand- og Spildevandsforening* bygger disse på data fra over 2.575.000 personer, hvilket svarer til omkring 47 % af den danske befolkning [74]. Oplysningerne om vandforbruget fra *Danmarks Statistik* bygger blandt andet på oplysninger fra *Dansk Vand- og Spildevandsforening*, men de er korrigeret af *Danmarks Statistik* efter deres oplysninger om blandt andet husholdninger uden tilknytning til almene vandforsyninger [75]. Idet ovenstående data er fra næsten halvdelen af den danske befolkning, antages det i dette projekt, at fordelingen af personerne, som dataet stammer fra, mellem forskellige husstandsstørrelser tilnærmelsesvist følger fordelingen på landsplan. Ifølge *Danmarks Statistik* [73] var 38,5 % af de danske husstande beboede af én person i 2009 og 33,0 % var beboede af to personer. På baggrund af disse oplysninger antages det i dette projekt, at de oplyste vandforbrug fra *Dansk Vand- og Spildevandsforening* [54] og *Danmarks Statistik* [75] kan anvendes for husstande med to beboere, og for de øvrige husstandsstørrelser justeres forbruget i henhold til middelværdien fra tabel L.2. Resultatet af denne justering kan ses af tabel L.3, hvor vandforbruget fra *Danmarks Statistik* er korrigeret som beskrevet i afsnit L.1.3 ³.

Husstandsstørrelse	Dansk Vand- og Spildevandsforening	Danmarks Statistik
[person]	[m ³ /person per år]	[m ³ /person per år]
1	54,08	57,25
2	41,44	43,87
3	35,36	37,44
4	32,31	34,21
5	30,38	32,27
6	29,26	30,98

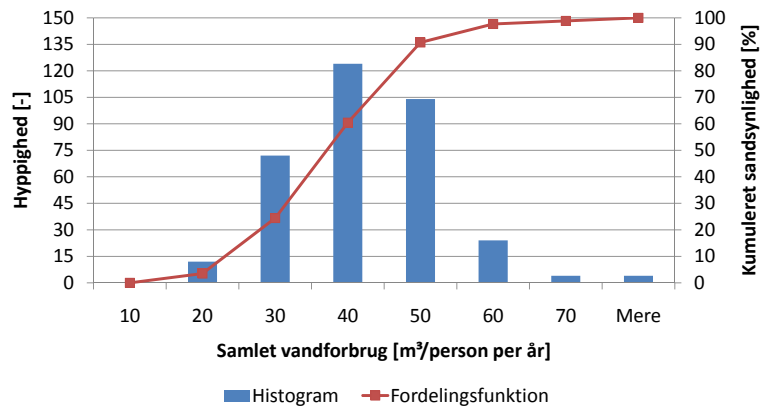
Tabel L.3: Vandforbrug per person for henholdsvis *Dansk Vand- og Spildevandsforening* og *Danmarks Statistik* korrigeret efter ændringen af vandforbruget per person som funktion af husstandsstørrelse. Oplysninger fra *Danmarks Statistik* er endvidere korrigeret efter vandforbrugets udvikling, som beskrevet i tabel L.1.

L.1.5 Test af fordelingstype

Til at vurdere hvilken sandsynlighedsfordeling der kan beskrive det samlede vandforbrug, udføres der en hypotesetest af nogle tilgængelige målinger fra en rapport af *Jesper Ole Jensen* [35]. I forbindelse med udarbejdningen af førnævnte rapport er der foretaget en registrering af vandforbruget i nogle boliger, hvor dele af disse målinger fremgik af *Jesper Ole Jensens*

³Yderligere oplysninger kan ses i regnearket *Sandsynlighedsfordelinger\Vandforbrug\Sammenligning af årligt vandforbrug.xlsx* på den medfølgende DVD.

rapport ⁴. Det drejer sig om vandforbruget for i alt 344 firepersoners husstande, hvor der primært er tale om rækkehuse fra 1960-70'erne. Disse oplysninger er anvendt i en *Chi-square test for goodness of fit*, hvor nærmere oplysninger om denne type hypotesetest er beskrevet i appendiks I. Der er foretaget en hypotesetest, hvor vandforbrugene er inddelt i intervaller på $5 \text{ m}^3/\text{person per år}$. Resultaterne viser dog at alle de opstillede hypoteser bliver forkastet ved et signifikansniveau på 0,05 ⁵. Af denne årsag udføres en hypotesetest hvor vandforbrugene er inddelt i intervaller på $10 \text{ m}^3/\text{person per år}$, hvor et histogram over boligernes samlede vandforbrug delt ud per person i de pågældende husstande er vist på figur L.4.



Figur L.4: Histogram over det samlede vandforbrug delt ud per person for i alt 344 firepersoners husstande.

På baggrund af hypotesetesten kan der foretages en kvantitativ analyse af hvilke sandsynlighedsfordelinger, der kan beskrive det måledata, der er vist på figur L.4. Resultaterne af testen med det høje interval kan ses i tabel L.4.

⁴Yderligere oplysninger kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\ Oplysninger til vandforbrugets størrelse (JOJ).xlsx* på den medfølgende DVD.

⁵Beregningerne kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\ 001 JOJ hypotesetest.xlsx* på den medfølgende DVD.

Sandsynlighedsfordeling	χ^2 -værdi	Kritisk χ^2 -værdi	Bemærkning
Normalfordeling	2,63	5,99	Middelværdi: 37,47 Standardafv.: 10,27 Nulhypotese accepteres
Lognormalfordeling	16,47	7,81	Log. middelværdi: 3,59 Log. standardafv.: 0,29 Nulhypotese forkastes
Eksponentialfordeling	611,81	12,59	Middelværdi: 37,47 Nulhypotese forkastes
Weibullfordeling	11,82	7,81	Skala-parameter: 41,31 Form-parameter: 3,81 Nulhypotese forkastes
Gammafordeling	7,27	7,81	Alfa-værdi: 12,97 Beta-værdi: 2,89 Nulhypotese accepteres

Tabel L.4: Hypotesetest af sandsynlighedsfordelinger i forhold til målinger af samlet vandforbrug for 344 danske husstande, hvor nulhypotesen accepteres såfremt $\chi^2 < \text{kritisk } \chi^2$.

Af tabellen kan det ses, at både en normalfordeling og en gammafordeling kan anvendes til beskrivelse af målingerne ⁶. Normalfordelingen har en middelværdi på omkring 37,5 $m^3/person$ *per år*, hvilket betyder, at det er denne størrelse af forbruget, der har størst sandsynlighed for at forekomme.

I afsnit L.1.2 er der beskrevet nogle formler, der kan anvendes til beregning af det samlede vandforbrug ud fra en 25 %- , 50 %- og 75 %-fraktil. Disse oplysninger er anvendt til at estimere hvilken middelværdi og standardafvigelse, der kan beskrive forbruget bedst muligt. De estimerede parametre er anvendt i henholdsvis en normal- og en lognormalfordeling. Resultaterne viser, at normalfordelingen kan beskrive forbruget med de parametre, der er angivet i tabel L.5. Ydermere viser resultaterne at en lognormalfordeling giver en bedre tilnærmelse af vandforbruget ved de forskellige fraktiler ⁷.

⁶Beregningerne kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\002 JOJ hypotesetest.xlsx* på den medfølgende DVD.

⁷Oplysninger og estimering af parametrene kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\ Sandsynlighedsfordeling ud fra fraktiler (MST).xlsx* på den medfølgende DVD.

Antal beboere [person]	Middelværdi [m ³ /person per år]	Standardafvigelse [m ³ /person per år]
1	65,20	28,52
2	45,13	17,95
3	38,44	14,43
4	35,09	12,66
5	33,09	11,61
6	31,75	10,90

Tabel L.5: Estimering af middelværdi og standardafvigelse for normalfordeling ud fra oplysninger af forbrug opgjort ved 25 %- , 50 %- og 75 %-fraktil.

Den ovenstående beskrivelse er blot en estimering af parametre for en normal- og lognormalfordeling. Den tidligere udførte hypotesetest viste, at kun en normal- og en gammafordeling kunne beskrive det samlede vandforbrug. Da det er meningen at sandsynlighedsfordeling for det samlede vandforbrug efterfølgende skal anvendes til vurdering af varmtvandsforbruget, udføres der også hypotesetests af tilgængelige data for dette forbrug. Der haves oplysninger fra *Knudsen et al.* [36] om varmtvandsforbruget i ni boliger, hvor disse oplysninger er anvendt i en *Kolmogorov-Smirnov one-sample test*. En nærmere beskrivelse af denne hypotesetest kan findes i appendiks I. Der er udført en hypotesetest på henholdsvis varmtvandsforbruget opgjort per beboer og per husstand med et signifikansniveau på 0,05. Resultaterne viser, at varmtvandsforbruget i begge tilfælde kan beskrives ved en normalfordeling, jf. tabel L.6 ⁸.

Sandsynlighedsfordeling	<i>KS</i> -værdi	<i>KS</i> _{0,05} -værdi	Bemærkning
Normalfordeling	0,16	0,43	Opjort per person Middelværdi: 9,88 Standardafv.: 4,80 Nulhypotese accepteres
Normalfordeling	0,13	0,43	Opjort per husstand Middelværdi: 35,57 Standardafv.: 19,10 Nulhypotese accepteres

Tabel L.6: Hypotesetest af normalfordelinger i forhold til målinger af varmtvandsforbruget for ni danske husstande, hvor nulhypotesen accepteres såfremt $KS < KS_{0,05}$. Hypotesetest er udført for forbruget opgjort per person og per husstand.

Ud fra iagttagelserne i dette afsnit vurderes en normalfordeling at vise fornuftige resultater for både det samlede vandforbrug men også for varmtvandsforbruget i boliger, hvorfor denne sandsynlighedsfordeling vil blive anvendt i det efterfølgende. Parametrene til denne fordelingstype bliver beskrevet nærmere i det efterfølgende afsnit.

⁸Beregningerne kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\003 FremParc hypotese-test.xlsx* på den medfølgende DVD.

L.1.6 Estimering af fordelingsparametre

Det er fastlagt, at der anvendes en normalfordeling til vurdering af det samlede vandforbrug. Dertil skal der estimeres en middelværdi for denne fordeling, hvilket gøres på baggrund af:

- Oplysninger fra *Dansk Vand- og Spildevandsforening* [54] korrigeret efter forbrugsændring som funktion af husstandsstørrelse. De anvendte forbrugsoplysninger kan ses i tabel L.3, hvor det bør bemærkes, at der er foretaget en række korrektioner af de oprindelige oplysninger fra *Dansk Vand- og Spildevandsforening*, hvilket er nærmere beskrevet i afsnit L.1.4.
- Oplysninger fra *Danmarks Statistik* [75] korrigeret efter forbrugsændring som funktion af husstandsstørrelse. De anvendte forbrugsoplysninger kan ses i tabel L.3, hvor der også er korrigeret efter vandforbrugets udvikling. Det må derfor bemærkes, at der er foretaget en række korrektioner af de oprindelige oplysninger fra *Danmarks Statistik*.
- Oplysninger fra *Miljøstyrelsen et al.* [46], hvor formelen for 50 %-fraktilen angivet i afsnit L.1.2 er anvendt til beregning af det forventede vandforbrug. Det bør bemærkes, at der er foretaget en række korrektioner af de oprindelige oplysninger fra *Miljøstyrelsen et al.*, der er nærmere beskrevet i førnævnte afsnit. Desuden er forbruget korrigeret efter vandforbrugets udvikling efter oplysningerne fra tabel L.1.
- Oplysninger fra *Jesper Ole Jensen* [35] om det samlede vandforbrug i firepersoners husstande. Der er anvendt en middelværdi fra hypotesetesten af en normalfordeling, hvor denne er angivet i tabel L.4. Det bør bemærkes, at oplysningerne er korrigeret efter vandforbrugets udvikling, hvor den anvendte korrektion er angivet i tabel L.1.
- Oplysninger fra *Gram-Hanssen et al.* [26] om det samlede vandforbrug, hvor der er foretaget en korrektion ud fra vandforbrugets udvikling, hvor den anvendte korrektion er angivet i tabel L.1.

Middelværdien for de ovenstående data kan ses i tabel L.7 ⁹. Hvad angår standardafvigelsen for det samlede vandforbrug anvendes de estimerede værdier angivet i tabel L.5, da der ikke haves andre oplysninger om variationen af dette forbrug blandt husstande ¹⁰.

⁹Beregningerne og yderligere information kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\ Sammenligning af årligt vandforbrug.xlsx* på den medfølgende DVD.

¹⁰En skitsering af sandsynlighedsfunktionerne for det samlede vandforbrug kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\ Sandsynlighedsfordelinger for samlet vandforbrug og varmt brugsvand.xlsx* på den medfølgende DVD.

Antal beboere	Middelværdi	Standardafvigelse
$[person]$	$[m^3/person\ per\ \text{år}]$	$[m^3/person\ per\ \text{år}]$
1	56,47	28,52
2	40,78	17,95
3	34,78	14,43
4	31,65	12,66
5	29,98	11,61
6	28,78	10,90

Tabel L.7: Anvendte middelværdier og standardafvigelser til normalfordelinger for det samlede vandforbrug beregnet ud fra ovenstående antagelser og korrektioner.

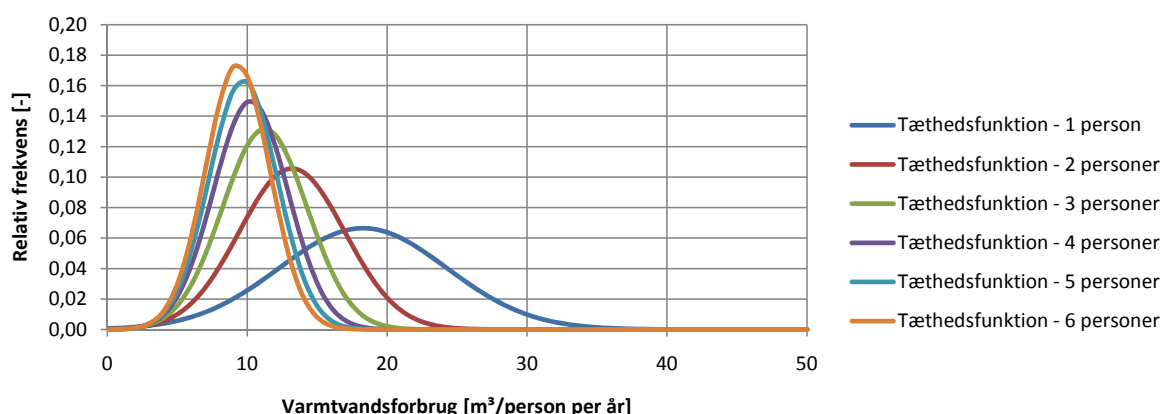
L.2 Husstandes varmtvandsforbrug

I afsnit 3.4 fra litteraturstudiet i hovedrapporten er det vist i tabel 3.20 at varmtvandsforbruget i gennemsnit udgør mellem 31,0 – 33,7 % af husstandes samlede vandforbrug i parcelhuse. Da der ikke haves andre oplysninger om varmtvandsforbrugets størrelse for parcelhuse anvendes et middel af disse to svarende til 32,35 %. På baggrund af disse oplysninger antages varmtvandsforbruget at udgøre 32,35 % af det samlede vandforbrug i parcelhuse. På trods af at middelværdierne og standardafvigelserne fra tabel L.7 er reduceret til at udgøre 32,35 % af værdierne, vurderes spredningerne stadig at være for store for varmtvandsforbruget. Det skyldes dels at intervallet, som varmtvandsforbruget varierer mellem er op mod $120\ m^3/husstand\ per\ \text{år}$ og dels at alle fordelinger angiver en sandsynlighed for et negativt forbrug. Derfor reduceres standardafvigelsen for varmtvandsforbruget med 35 %. De endelige middelværdier og standardafvigelser der anvendes til estimering af husstandes varmtvandsforbrug er angivet i tabel L.8

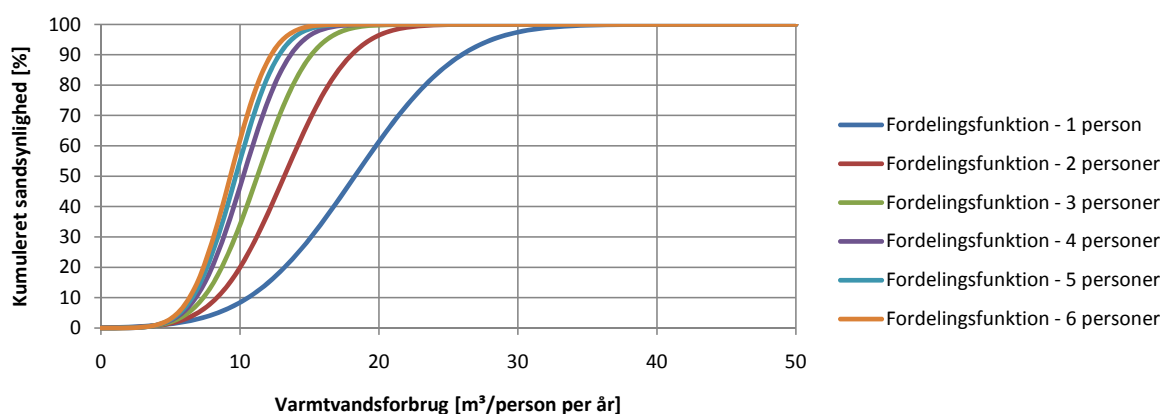
Antal beboere	Middelværdi	Standardafvigelse
$[person]$	$[m^3/person\ per\ \text{år}]$	$[m^3/person\ per\ \text{år}]$
1	18,27	9,23
2	13,19	5,81
3	11,25	4,67
4	10,24	4,10
5	9,70	3,75
6	9,31	3,53

Tabel L.8: Endelige middelværdier og standardafvigelser til normalfordeling til beregning af varmtvandsforbruget beregnet ud fra ovenstående antagelser og korrektioner.

Resultaterne er optegnet grafisk på figur L.5 til L.8 for tætheds- og fordelingsfunktion for henholdsvis varmtvandsforbruget opgjort per person og per husstand ¹¹.

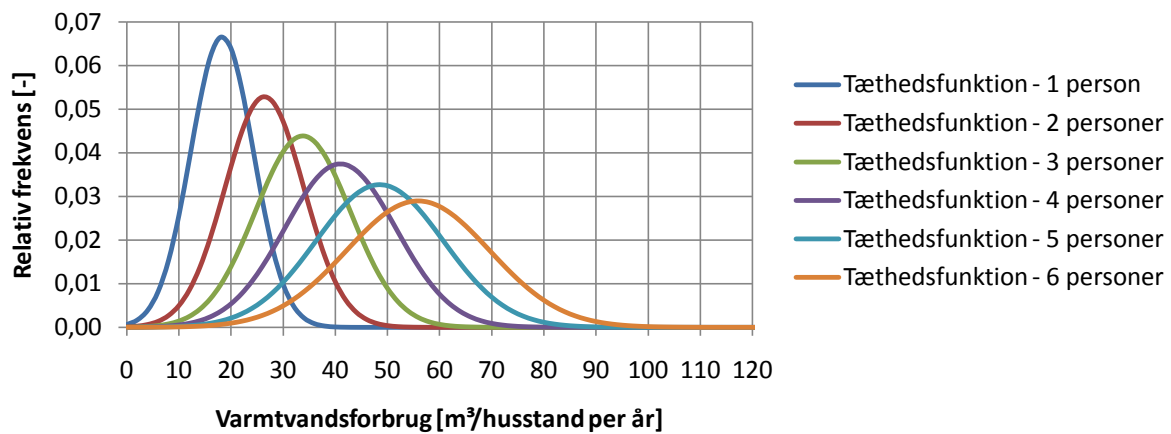


Figur L.5: Tæthedsfunktioner for varmtvandsforbrug opgjort per person for normalfordelinger med parametrene angivet i tabel L.8.

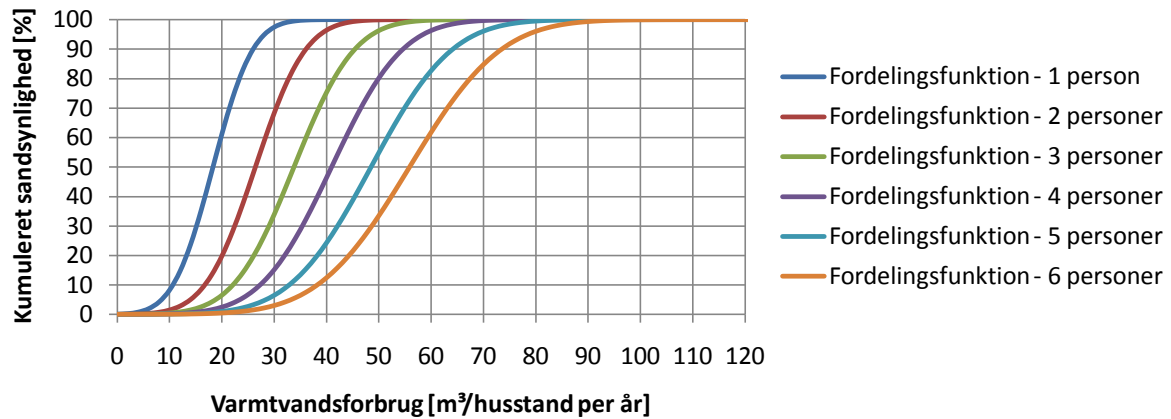


Figur L.6: Fordelingsfunktioner for varmtvandsforbrug opgjort per person for normalfordelinger med parametrene angivet i tabel L.8.

¹¹Yderligere oplysninger kan ses i regnearket *Sandsynlighedsfordelinger\ Vandforbrug\ Sandsynlighedsfordelinger for samlet vandforbrug og varmt brugsvand.xlsx* på den medfølgende DVD.



Figur L.7: Tæthedsfunktioner for varmtvandsforbrug opgjort per husstand for normalfordelinger med parametrene angivet i tabel L.8.



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

Algoritmer

I læringsdelen jf. figur 4.5 i kapitel 4.3 i hovedrapporten ses det, at der anvendes algoritmer ved udvikling af klassificeringstræ. En algoritme kan betragtes som en præcis „opskrift“ for hvordan et problem kan løses. Forklaret meget forenklet finder algoritmen automatisk en sammenhæng i en given datamængde, for herefter at visualisere sammenhængen som et beslutningstræ. Skrevet som pseudokode giver det [88] [61]:

Input: En given datamængde $\rightarrow$ Output: Et klassificeringstræ

For at denne proces foregår automatisk opstilles der i algoritmen nogle splitkriterier for hvordan data skal sorteres. Splitkriterierne der anvendes afhænger af hvilken algoritme der er valgt. I nærværende kapitel vil fokus være lagt på algoritmerne ID3, C4.5, C5.0/See5¹, CHAID, QUEST og CART, hvor fokus lægges på deres interne forskelle og deres fordele og ulemper. Efter udvælgelse af den mest anvendelige algoritme for nærværende projekt, vil det matematiske grundlag for splitkriterierne blive gennemgået i kapitel 4.3 i hovedrapporten.

M.1 QUEST

QUEST algoritmen er udviklet af *Wei-Yin Loh* og *Yu-Shan Shih* i 1997. Navnet står for *Quick, Unbiased, Efficient, Statistical Tree*. Modsat algoritmerne ovenfor udvikler denne algoritme kun binære træer, dvs. træer hvor knuderne (R- og I-knuder) maks har to grene. [32] Derudover kan denne algoritme kun anvendes ved udvikling af klassificeringstræer [41]. Splitkriteriet for denne algoritme er bestemt vha. ANOVA F-test eller Levene's test for kontinuerte attributter, hvorimod algoritmen anvender Pearson's Chi-square for nominelle attributter [64].

Da denne algoritme kun udvikler binære træer vil beskrivelsen af denne algoritme ikke blive udført yderligere, idet der i klassificeringstræet, der skal beskrive brugeradfærd i boliger kan opstå et behov for et ikke-binært træ.

¹Algoritmens navn afhænger af hvilket operativsystem der anvendes, See5 anvendes ved styresystem, Windows 2000/Xp/Vista/7 hvorimod C5.0 anvendes for styresystemet Linux/UNIX. I det efterfølgende vil betegnelsen der anvendes være See5.

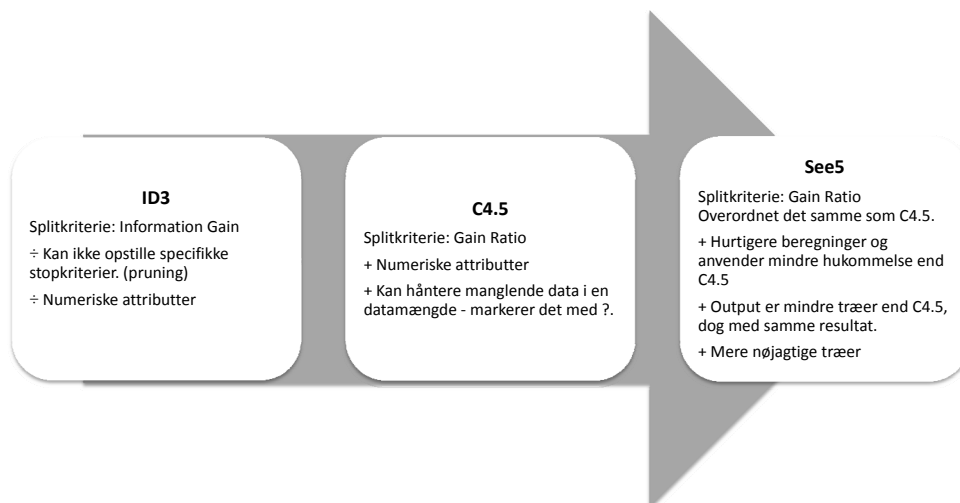
M.2 CART

CART er en sammentrækning af *Classification and Regression Trees*, der oversat til dansk betyder *Klassificerings og Regressions Træer*. Denne metode blev udviklet af *Breiman et al.* tilbage i 1984. Modsat de ovenstående algoritmer, er denne CART algoritme kendetegnet ved at den kun udvikler binære træer, dvs. træer hvor knuderne (R- og I-knuder) maks har to grene. Splitkriteriet for denne algoritme er bestemt vha. *Twoing Criteria* og stopkriteriet bestemmes vha. metoden *Cost-Complexity Pruning*. Hovedideén bag CART algoritmen er at finde den splittest der resultere i den „reneste“ inddeling af data, dvs. bestemme den attribut der inddeler data bedst i forhold til det endelige mål.

Da denne algoritme kun udvikler binære træer vil beskrivelsen af denne algoritme ikke blive udført yderligere, idet der i klassificeringstræet, der skal beskrive brugeradfærd i boliger kan opstå et behov for et ikke-binært træ.

M.3 ID3, C4.5 og See5

Algoritmerne ID3, C4.5 og See5 er alle udviklet af datalogi forskeren *Ross Quinlan* og anses som nogle af de mest populære algoritmer [44]. ID3 er den første algoritme af de tre nævnte og blev udviklet i 1986, C4.5 blev udviklet i 1993 og See5 blev introduceret i 1997. Alle tre algoritmer kan udvikle både binære og ikke-binære træer ². Algoritmen ID3 anvender splitkriteriet der hedder *Information Gain*, hvorimod C4.5 og See5 anvender splitkriteriet *Gain Ratio*. Forskellen mellem algoritmen ID3 og algoritmerne C4.5 og See5 er, at ID3 ikke kan opstille stopkriterier for det udviklede træ (pruning) og ej heller kan håndtere numeriske attributter eller manglende værdier i datamængden. De overordnede forskelle mellem de tre algoritmer kan ses i figur M.1.



Figur M.1: De overordnede forskelle mellem algoritmerne ID3, C4.5 og See5, alle udviklet af *Ross Quinlan* [61, 64].

²Binært træ: antallet af grene fra en knude er maks to. Ikke-binært træ: inddelingen af data fra en knude kan være mere end to.

Ud over de viste forskelle i figur M.1, er der også forskel på hvornår algoritmerne stopper deres inddeling af data. En algoritme fortsætter på sin opgave indtil den når et såkaldt *stopkriterie*. For algoritmen ID3 er stopkriteriet når alle målinger er inddelt og alle inddelte grupper er homogene, dvs. at al data i en inddeling er 100% ens. For C4.5 og See5 er et af stopkriterierne formet noget anderledes. Det fungerer således at algoritmen C4.5 og See5 stopper inddelingen når datamængden har nået en minimumsstørrelse, der er angivet på forhånd. [61]

Derudover er der nogle generelle stopkriterier algoritmer, der kan ses nedenfor:

- Alle datamængder i den givne knude tilhører samme klasse. En løsning for inddeling af data er fremkommet og knuden bliver en L-knude, med et navn der beskriver den klasse, der er tilstede (Klassen kan, som nævnt i kapitel 4.3 i hovedrapporten, være underopdelt i prædefinerede klasser). [94, 64]
- Der er ikke flere attributter til rådighed, der kan anvendes til at udføre en yderligere inddeling af datamængden. Konsekvensen heraf bliver, at knuden bliver en L-knude hvor flertallet af de tilstedeværende endelige klasser i datamængden er afgørende for hvilket klassenavn knuden skal tildeles. Idet knuden består af flere mulige endelige klasser kan knuden tildeles en sandsynlighedsvektor der beskriver sandsynligheden for at den ene eller anden klasse er tilstede. [94]
- Ingen af attributterne giver en værdi for information gain. Der tages et skridt baglæns i træets struktur og der skabes en L-knude, hvori flertallet af de tilstedeværende klasser bestemmer navnet til knuden. [64]
- Den maksimale brugerspecifikke træstørrelse er nået. [64]
- Efter inddeling af data indeholder en af L-knuderne et mindre antal datamængder end hvad der som minimum er tilladt, forudsat af brugeren. [64]

M.3.1 Fordele og ulemper ID3, C4.5 og See5

I dette afsnit gennemgås fordele og ulemper for algoritmerne ID3, C4.5 og See5, se tabel M.1.

Algoritme	Fordele	Ulemper
ID3 [48]	Den udvikler det hurtigste træ. Træet der udvikles er et forholdsvis lille og ukompliceret træ. Al data undersøges for at et træ kan blive udviklet. Der er kun behov for at teste det antal af attributter der er nødvendigt for at al data bliver inddelt. Når træet er udviklet er der automatisk opstillet nogle regler for hvordan data inddeles.	Ved anvendelse af en lille datamængde er der risiko for at data kan blive over-fitted eller over-classified. Der kan kun testes en attribut af gangen. Denne algoritme håndterer ikke numeriske attributter. Kan ikke håndtere kontinuerte attributter.
C4.5 [44, 62]	Kan håndtere kontinuerte attributter. Kan håndtere manglende data i en datamængde.	Risiko for over-fitting. Alle fejlplaceringer betragtes lige.
See5 [62]	Kreerer mindre beslutningstræer. Kreerer beslutningstræer med større nøjagtighed.	

Tabel M.1: Fordele og ulemper for algoritmerne ID3, C4.5 og See5 [48].

M.4 CHAID

Algoritmen CHAID ³ er udviklet af *Gordon V. Kass* i 1980, og er eftersigende en af verdens ældste algoritmer, der kan anvendes til beslutningstræer. Ligesom Id3, C4.5 og See5 algoritmerne, udviklet af *Ross Quinlan*, kan denne algoritme anvendes til at udvikle ikke-binære træer. Algoritmen kan anvendes for både klassificerings- og regressionstræer, dog er der forskel på hvilke metoder algoritmen anvender for hvert type træ mht. bestemmelse af den bedste splittest. CHAID algoritmen anvender to forskellige splitkriterier *Pearsons Chi-square test* for klassificeringstræer, og en *F-test* anvendes for regressionstræer. [41, 64] Stopkriterierne algoritmen anvender er givet nedenfor [64]:

- Den maksimale træstørrelse er nået (forudsat på forhånd).
- Det minimale antal af data (forudsat på forhånd) i en I-knude er nået og den kan derfor ikke splittes yderligere.
- Det minimale antal af data (forudsat på forhånd) i en L-knude er nået.

³Akronymet CHAID står for *Chi-squared Automatic Interaction Detector*.

M.4.1 Fordele og ulemper, CHAID

Idet det vides at CHAID algoritmen udvikler ikke-binære træer, dvs. der er flere grene for hver knude, vil denne automatisk udvikle nogle breddere træer end CART og QUEST algoritmerne. [41] Breddere træer kan skabe en risiko for et mere uoverskueligt træ, dette afhænger dog også af andre faktorer.

M.5 Opsummering

Det bedste mål for valg af algoritme vil være at vælge den der giver det bedste og dermed det mest præcise resultat når et klassificeringstræ skal udvikles. Desværre er der endnu ikke udviklet nogle generelle guidelines der kan uddybe hvilke algoritmer der er bedst anvendelige mht. nogle specifikke opgaver. [41]

Alle de præsenterede algoritmer ovenfor kan anvendes for klassificeringstræer og da dette er målet for metoden der skal udvikles i nærværende projekt, anses alle algoritmerne som anvendelige. Fra litteraturstudiet vides det, at antallet af personer i en bolig er en af de attributter der forklare store dele af spredningen for resultaterne. Denne viden lægger op til, at den algoritme der skal vælges kan udvikle ikke-binære træer. Denne viden reducerer de muligt anvendelige attributter til: ID3/C4.5/See5 og CHAID idet algoritmerne CART og QUEST, som skrevet under beskrivelserne, kun kan anvendes til udvikling af binære træer. [64, 41, 30, 32, 31]

Undersøgelse af nye mulige attributter

N.1 Rumopvarmning

Der er gennem afsnit 3.2 i litteraturstudiet beskrevet flere projekter, hvori der er undersøgt hvilke parametre, der kan have indflydelse på energiforbruget til rumopvarmning. I det efterfølgende beskrives en test på måledata fra projektet *Fremtidens Parcelhuse* [36], hvor det undersøges hvorvidt de forklarende parametre fra litteraturstudiet også gør sig gældende for den aktuelle datamængde. Foruden en test af disse parametre, er der foretaget en sammenligning vandforbruget og af andre forhold i forbindelse med boligen og beboerne, der kan have en forklarende effekt. I afsnit 3.2 i hovedrapporten er der nævnt nogle parametre, der har en forklarende effekt på energiforbruget til rumopvarmning, og disse er opsummeret nedenfor:

- Lufttemperatur i boligen
- Boligens størrelse og boligforbrug (m^2 bolig per person)
- Beboernes alder
- Holdninger til energi, miljø og økonomi
- Vaner omkring udluftning
- Indkomst

Ud fra den ovenstående liste med parametre er det ikke muligt at undersøge parametrene *vaner omkring udluftning* og *indkomst*, da disse parametre ikke har indgået i undersøgelsen.

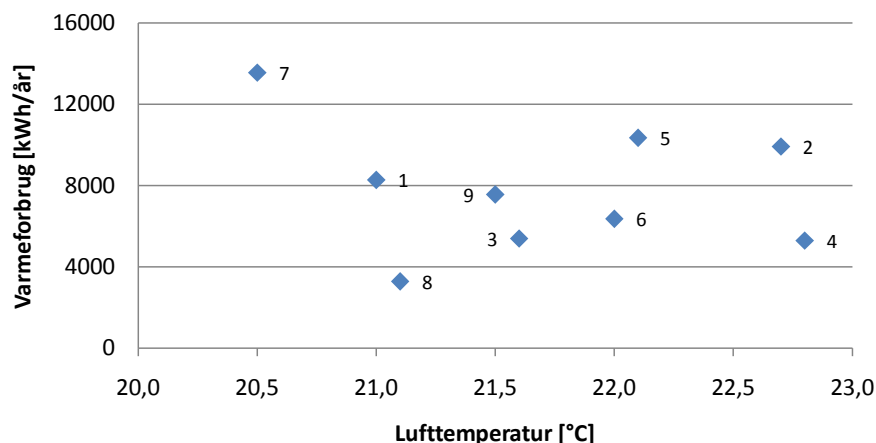
Det skal bemærkes, at testen udføres på en lille datamængde, ni-ti huse, da den tilgængelige datamængde ikke er større. Der vil derfor i nogle tilfælde ikke kunne ses et lige så klare tendenser, som det har kunne ses i de tidligere gennemgåede projekter.

N.1.1 Lufttemperatur

Der er gennem litteraturstudiet lagt stor fokus på lufttemperaturen som et væsentligt forhold i forbindelse med energiforbruget til rumopvarmning. Det skyldes, at lufttemperaturen, i

afsnit 2.2.1 i hovedrapporten og gennem litteraturstudiet i afsnit 3.2, har vist at have stor indflydelse på energiforbruget til opvarmning. På baggrund af tilgængeligt data fra *Knudsen et al.* [36] vil det vurderes, om denne sammenhæng også er til stede i dette data.

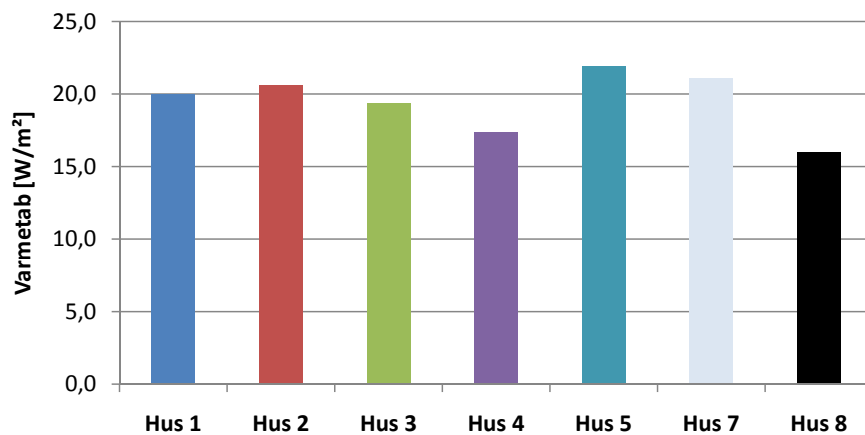
Figur N.1 viser en optegning af den gennemsnitlige lufttemperatur i boligerne, målt i perioden fra oktober 2009 til december 2009, sammenholdt med det årlige energiforbrug til rumopvarmning. Det første forhold der umiddelbart falder i øjnene, når den førnævnte figur betragtes er, at *hus 7* har den laveste gennemsnitlige lufttemperatur og samtidig det højeste forbrug, hvilket er modstridende med, hvad der er set i litteraturstudiet.



Figur N.1: Skitsering af den gennemsnitlige lufttemperatur i de ni husstande, målt i perioden fra oktober 2009 til og med december 2009, sammenholdt med energiforbruget til rumopvarmning.

Årsagen til det høje varmeforbrug i *hus 7* er ikke blevet undersøgt i rapporten af *Knudsen et al.* [36], der knytter sig til målingerne. Det kan skyldes, at beboerne i den pågældende husstand har en anderledes adfærd end de øvrige husstande i undersøgelsen. For at undersøge det nærmere er der set nærmere på boligens varmetab, udluftningsvaner og hvordan de enkelte beboere oplever indeklimaet i de pågældende husstande.

Der er foretaget en overslagsberegning af varmetabet af de enkelte boliger, hvor oplysningerne stammer fra rapporten af *Knudsen et al.* [36]. Resultaterne kan ses på figur N.2, hvor *hus 6* og *hus 9* ikke fremgår grundet manglende oplysninger til beregningen. Af figuren fremgår det, at *hus 5* og *hus 7* har de højeste varmetab, og at *hus 8* har det laveste. Disse oplysninger stemmer udmærket overens med, at *hus 5* og *hus 7* har det højeste energiforbrug til rumopvarmning, samtidig med at *hus 8* har det laveste energiforbrug. Det vurderes dog, at energiforbruget til rumopvarmning i *hus 7* er højere end hvad varmetabsrammen kompenserer for.



Figur N.2: Overslagsberegning for varmetabet fra hus 1-9, hvor beregningerne er lavet på basis af oplysninger fra *Knudsen et al.* [36].

Hvad angår infiltrationen i boligerne, er det undersøgt ved en blowerdoortest af boligerne, ikke alle målinger har været tilgængelige. Resultatet for vurderingen af infiltrationen betydning kan ses i afsnit 5.3.1 i hovedrapporten.

Da forskellene i varmetabet ikke umiddelbart forklarer den store forskel i energiforbruget, undersøges det, om der er andre forhold, der gør sig gældende. I den forbindelse undersøges beboernes vaner om udluftning og deres opfattelse af indeklimaet. I spørgeskemabesvareelserne fra husstandene fremgår det, at 100 % og 66,7 % svarer *ja* til at der luftes ud henholdsvis om dagen og om natten i sommerperioden. I vinterperioden er tallene anderledes, hvor 86,7 % og 20 % svarer *ja* til at de lufter ud henholdsvis om dagen og om natten om vinteren. Fra svarene fremgår det, at de eneste to boliger, der lufter ud om natten om vinteren er *hus 6* og *hus 7*. Beboerne i disse boliger angiver, at årsagen til udluftningen er, at de ønsker at sænke temperaturen i hjemmet, samt at de ønsker frisk luft i boligen. Det fremgår dog ikke af spørgeskemaerne hvor længe udluftningerne varer, hvorfor det er vanskeligt at vurdere hvorvidt ovenstående har stor påvirkning på varmeforbrugets størrelse. En øget udluftning kan dog være skyld i en øget ventilationsmængde, og dermed kræve yderligere energi til at opvarme luften.

Idet der er stor forskel på lufttemperaturen i boligerne undersøges beboernes vurdering af komforten, mht. lufttemperatur og luftbevægelser. På baggrund af spørgeskemabesvareelserne fremgår det, at der i *hus 6* og *hus 7* ifølge beboerne er for koldt både sommer og vinter. Det er på trods af, at beboerne holder to vidt forskellige lufttemperaturer. Med hensyn til oplevelsen af luftbevægelserne i boligen har beboerne i *hus 7*, sammen med beboerne i *hus 3*, *4* og *9*, angivet, at de også oplever træk både sommer og vinter. Årsagen til træk skyldes i *hus 9* utætheder omkring nogle stikkontakter og *hus 3* og *4* er det vinduernes konstruktion, der medfører træk ¹. I *hus 7* er der ikke angivet en specifik årsag, andet end beboerne udtrykker at: „af og til opleves træk/kuldenedfald om vinteren“. Kilden til træk er ikke angivet, men det kan enten skyldes højtplacerede vinduer eller et ventilationssystem, der ikke fungerer optimalt. Mistanken opstår idet der opleves problemer med træk både sommer og vinter, hvilket kan skyldes en lav indblæsningstemperatur. En lav indblæsningstemperatur kan forårsage et øget varmebehov, hvilket skal dækkes af gulvvarmen i den pågældende

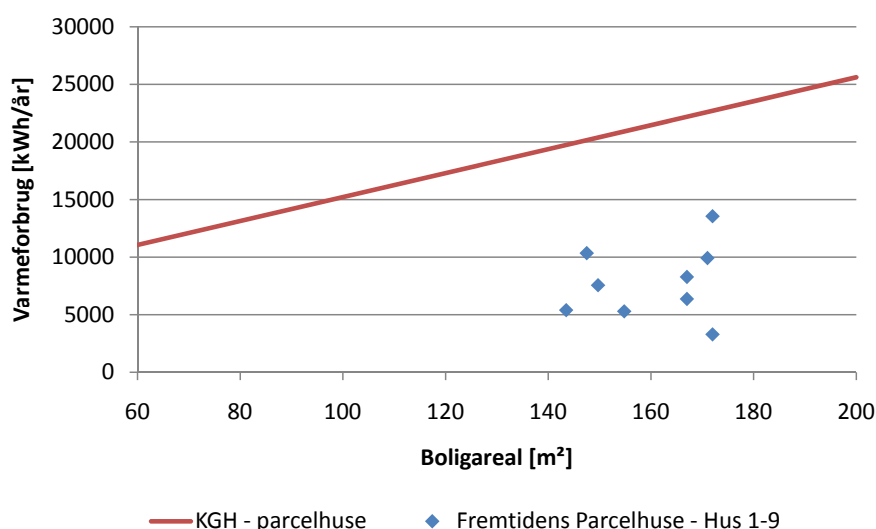
¹Der er naturlig ventilation i begge boliger

bolig, for at kunne opnå den ønskede lufttemperatur. Dette forhold kan være skyld i, at varmekorbruget for *hus 7* er højere end i de øvrige boliger. Det betyder, at årsagen til det høje varmekorbrug i *hus 7* kan skyldes en forøget ventilationsmængde grundet øget udluftning, træk fra vinduer eller et teknisk system der ikke fungerer optimalt.

Hvis der for en kort stund ses bort fra *hus 7* på N.1, kan der dog ses en beskeden tendens til, at der haves højere lufttemperaturer i husstandene med et højt energikorbrug til rumopvarmning.

N.1.2 Boligens størrelse og boligforbrug

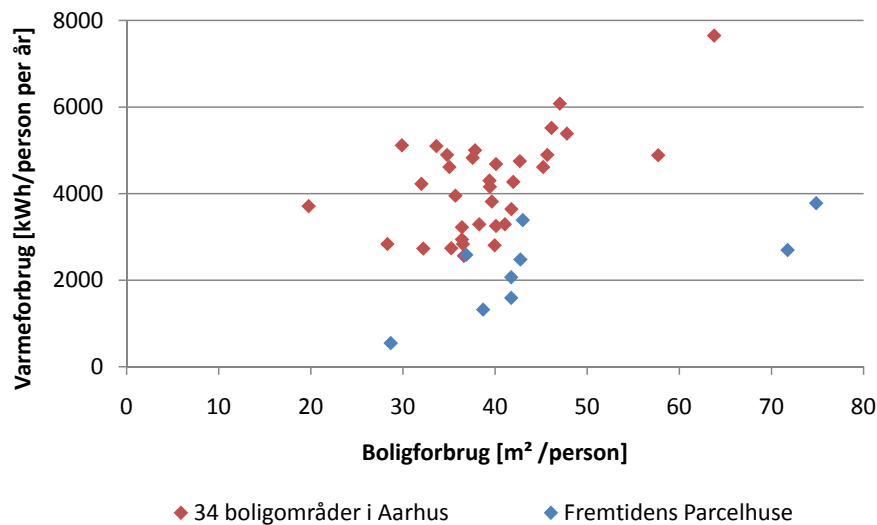
Der har fra en undersøgelse udført af *Gram-Hanssen et al.* [26] vist sig en sammenhæng mellem boligens størrelse og det samlede varmekorbrug. Af denne årsag vil sammenhængen blive testet i forhold til ni parcelhuse, hvorfra der haves tilgængeligt data. På figur N.3 ses en sammenligning af resultater fra rapporten af *Gram-Hanssen et al.* og målingerne fra de ni parcelhuse. Af figuren fremgår der ingen tydelig tendens blandt de ni husstande.



Figur N.3: Skitsering af energikorbrug til rumopvarmning blandt ni parcelhuse sammenholdt med boligarealet samt en sammenligning med resultaterne fra *Gram-Hanssen et al.* [26, 36].

Foruden boligarealet har en undersøgelse af *Jesper Ole Jensen* [35] vist en sammenhæng mellem boligforbruget og energikorbruget til rumopvarmning. Boligforbruget er opgjort som det tilgængelige boligareal delt ud per person, hvor varmekorbruget er stigende ved et stigende boligforbrug. På figur N.4 ses boligforbruget fra *Jesper Ole Jensens* rapport sammenlignet med testen udført på de ni boliger. Ud fra figuren ses det, at den førnævnte tendens også fremgår af målingerne for de ni husstande.

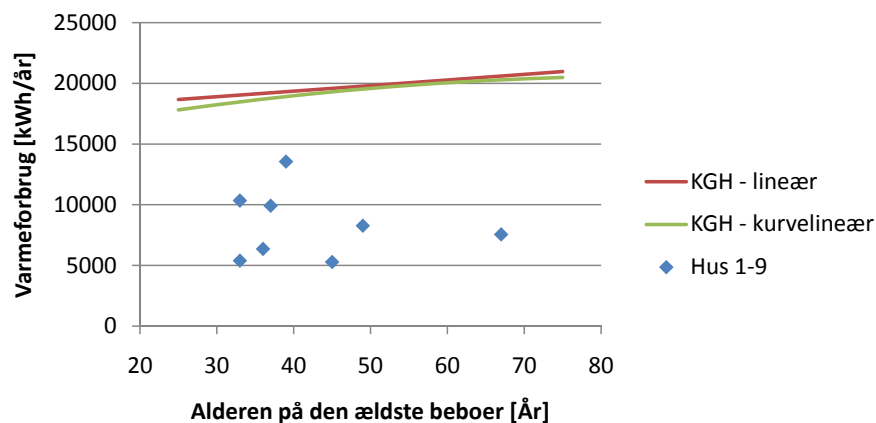
Årsagen til den lodrette forskel mellem de to typer af undersøgelser, kan skyldes boligernes isoleringsgrad. Boligområderne i Aarhus er opført i perioden fra midten af 1960'erne til midten 1990'erne og de ni boliger, hvor dataet stammer fra projektet af *Knudsen et al.*, er opført i midten af 2000'erne [35, 36].



Figur N.4: Skitsering af sammenhæng mellem boligforbrug og varmekonsum [35, 36].

N.1.3 Beboernes alder

I litteraturstudiet blev sammenhængen mellem varmekonsumet og beboernes alder gennemgået af bl.a. *Santin et al.* [27], *Gram-Hanssen et al.* [26] og *Jesper Ole Jensen* [35], hvor de fandt frem til, at beboernes alder har betydning for varmekonsumet, specielt med henblik på beboere over 65 år. *Gram-Hanssen et al.* har på basis af data fra 22379 husstande opstillet to formeludtryk, der beskriver sammenhængen mellem den ældste beboers alder og varmekonsumet. Den ene formel udtrykker en lineær sammenhæng mellem varmekonsumet og den ældste beboers alder, og den anden formel beskriver det samme forhold, men blot kurvelineært for at tage højde for beboernes livscyklus i forbrug. På figur N.5 ses en grafisk præsentation af de to formeludtryk sammen med målingerne fra ni parcelhuse.

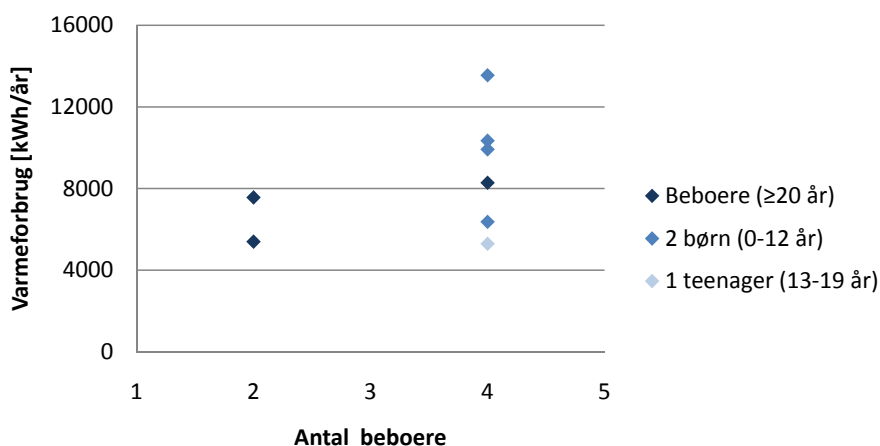


Figur N.5: Skitsering af varmekonsumet sammenholdt den ældste beboers alder for henholdsvis to forskellige formeludtryk fra *Gram-Hanssen et al.* [26] samt tilgængelige måledata fra ni parcelhuse [36]

Der ses ingen klar sammenhæng mellem varmekonsum og alderen på den ældste beboer i de ni forskellige parcelhuse, hvilket kan skyldes, at mængden af boliger der testes på er for lille. De målte varmekonsum fra de ni husstande er lavere end de to formeludtryk, der er beskrevet af *Gram-Hanssen et al.*, hvilket kan skyldes forskellen i isoleringsgraden i boligerne. Ligningerne fra *Gram-Hanssen et al.* er, som tidligere nævnt, beskrevet på baggrund af 22379 husstande fra Aarhus området, hvor bestanden af boliger generelt er ældre end de boliger de sammenlignes med.

N.1.4 Tilstedeværelse af børn og teenagere

I det efterfølgende er det undersøgt, hvorvidt der kan ses en sammenhæng mellem antallet af børn eller teenagere og energikonsumet til rumopvarmning. Ud fra spørgeskemabesvarelsene er der foretaget en opdeling af husstandene ud fra beboernes alder. Der er to husstande med to voksne beboere, samt en husstand med to voksne og to børn, der er over teenagealderen. Desuden er der fire husstande, som alle har to voksne og to børn under 12 år, samt en enkelt husstand med to voksne, ét barn og én teenager. Foruden de førnævnte husstande er der én af de ni husstande, der ikke har besvaret spørgeskemaet, hvorfor der ikke haves kendskab til beboersammensætningen i denne. På figur N.6 kan der ses en sammenligning mellem beboersammensætningen og husstandenes varmekonsum. Der fremgår ikke nogen entydig sammenhæng mellem de førnævnte parametre.

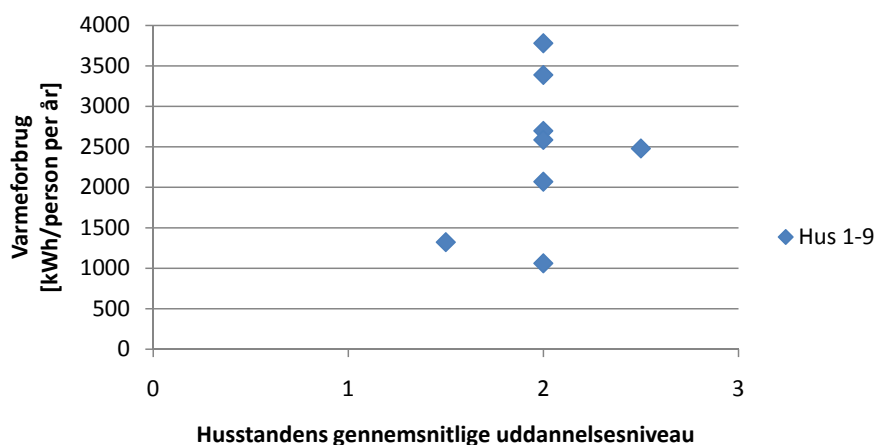


Figur N.6: Skitsering af otte husstandes energikonsum til rumopvarmning sammenholdt med antallet af børn og teenagere for de pågældende husstande.

N.1.5 Uddannelsesniveau

Det er undersøgt hvorvidt der kan ses en sammenhæng mellem uddannelsesniveauet og energikonsumet til rumopvarmning opgjort per person. Eftersom der i spørgeskemaet er stillet spørgsmål til hvorvidt beboerne har en kort, mellemlang eller lang uddannelse, gives der point efter denne skala. Det betyder at uddannelsesniveauet er opgjort således at en kort uddannelse giver ét point, en mellemlang uddannelse giver to point og en lang uddannelse giver tre point. Såfremt der er flere besvarelser fra samme husstand er der anvendt en midelværdi af de individuelle uddannelsesniveauer. Det fremgår af figur N.7, at der ikke kan

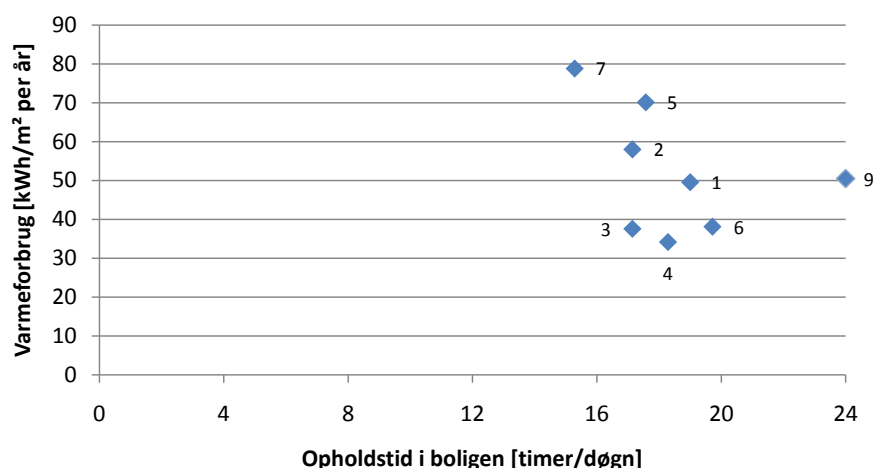
ses en direkte sammenhæng mellem husstandens gennemsnitlige uddannelsesniveau og det samlede vandforbrug.



Figur N.7: Skitsering af gennemsnitligt uddannelsesniveau i otte husstande sammenholdt med energiforbruget til rumopvarmning opgjort per person i de pågældende husstande.

N.1.6 Opholdstid i boligen

Det er undersøgt hvorvidt beboernes opholdstid i boligen har en forklarende effekt i forhold husstandens energiforbrug til rumopvarmning. I spørgeskemaet har beboerne svaret på, hvor mange timer de opholder sig i boligen på henholdsvis hverdage og i weekender. Såfremt flere beboere har besvaret spørgeskemaet, er der anvendt den længste opholdstid, der er oplyst. Opholdstiden er endvidere beregnet som et vægtet gennemsnit, således at opholdstiden i hverdagene tæller 5/7 mens opholdstiden i weekenden tæller 2/7. På figur N.8 kan der ses en sammenligning mellem husstandes samlede varmekonsum og opholdstiden. For husstandene *Hus 1-7* kan der være en tendens til, at jo kortere opholdstid i boligen, jo højere er energiforbruget til rumopvarmning opgjort per m^2 per år. Tendensen er dog ikke entydig, eftersom beboerne med den længste opholdstid, angivet som 24 timer både hverdage og weekend, ikke har det laveste forbrug.



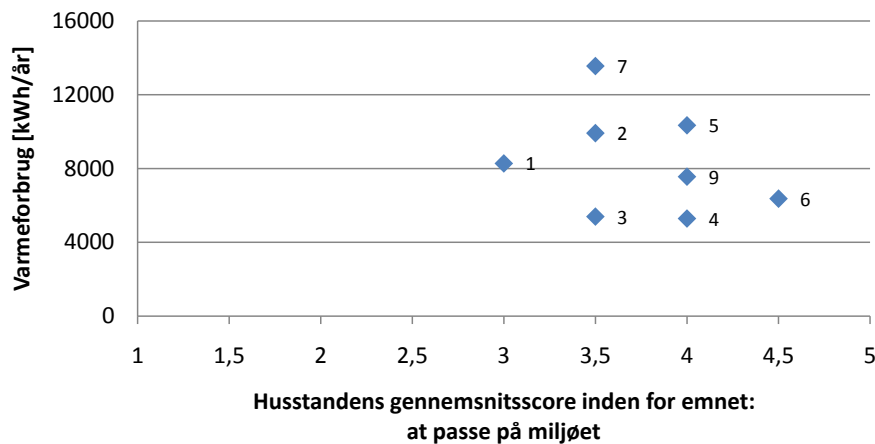
Figur N.8: Skitsering af beboernes oplyste opholdstid i otte husstande sammenholdt med energiforbruget til rumopvarmning opgjort per m^2 per år i de pågældende husstande.

N.1.7 Beboernes holdning til miljø og energibesparelser

I spørgeskemaet har beboerne svaret på i hvor høj grad de er optaget af at passe på miljøet. Besvarelserne er udført på en skala fra ét til fem, hvor ét betyder, at de er lidt optaget af at passe på miljøet, og ved fem er de meget optaget af det. Spørgsmålene lyder således:

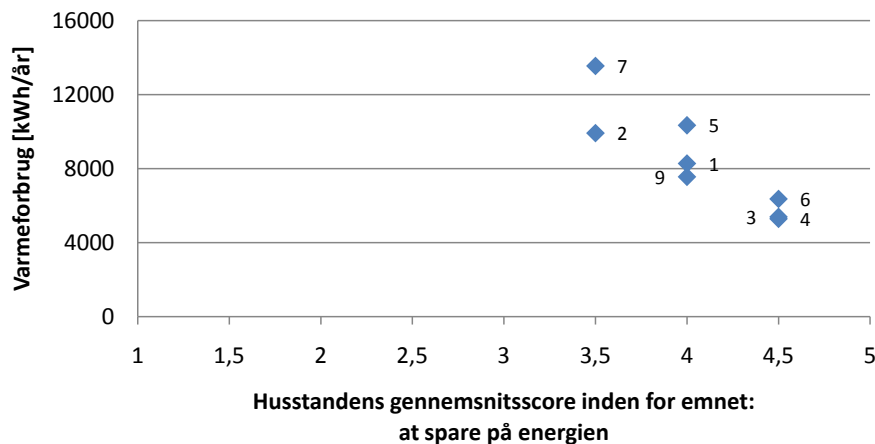
- Er du optaget af at passe på miljøet?
- Er du optaget af at spare på energien?

Besvarelserne til spørgsmålene er anvendt for at undersøge, om der kan ses en sammenhæng mellem beboernes besvarelse og energiforbruget til rumopvarmning i de pågældende husstande. I de tilfælde hvor to beboere fra samme husstand har besvaret spørgeskemaet, er der anvendt en middelværdi fra de to besvarelser. Der haves besvarelser fra otte husstande, hvor figur N.9 viser en grafisk præsentation af beboernes besvarelse for, hvor optaget der er af at passe på miljøet. Det fremgår af figuren, at der kan være en tendens til, at jo mere optaget beboerne er af at passe på miljøet, jo lavere energiforbrug til rumopvarmning er der i de pågældende husstande.



Figur N.9: Skitsering af beboernes holdning til at passe på miljøet sammenholdt med det samlede varmeforbrug i de pågældende husstande.

Hvad angår beboernes besvarelser af spørgsmålet til, hvor optaget de er af at spare på energien, kan der ses en grafisk præsentation af dette på figur N.10. Af figuren fremgår en bemærkelsesværdig tydelig sammenhæng mellem beboernes besvarelser og varmeforbrug. Sammenhængen viser, at jo mere optaget beboerne er af at spare på energien, jo lavere energiforbrug til rumopvarmning er der i de pågældende husstande. Det skal dog bemærkes, at besvarelserne kun er fra otte husstande, hvorfor undersøgelsen bør udføres på et større antal husstande.



Figur N.10: Skitsering af beboernes holdning til at spare på energien sammenholdt med det samlede varmeforbrug i de pågældende husstande.

N.2 Vandforbrug

I litteraturstudiet omhandlende boligernes vandforbrug er der i nogle projekter undersøgt hvilke forhold, der kan have en forklarende effekt i forbindelse med husstandenes samlede vandforbrug og varmtvandsforbrug. I det efterfølgende beskrives en test på måledata fra projektet *Fremtidens Parcelhuse* [36], hvor det undersøges hvorvidt de forklarende parametre fra lit-

teraturstudiet også gør sig gældende for den aktuelle datamængde. Foruden en test af disse parametre, er der foretaget en sammenligning vandforbruget og af andre forhold i forbindelse med boligen og beboerne, der kan have en forklarende effekt. I afsnit 3.4 i hovedrapporten er der nævnt nogle parametre, der har en forklarende effekt på vandforbruget, og disse er opsummeret nedenfor:

- Antal beboere
- Antal bade
- Antal tøjvaske
- Ejerform
- Tilstedeværelse af teenagere i husstanden

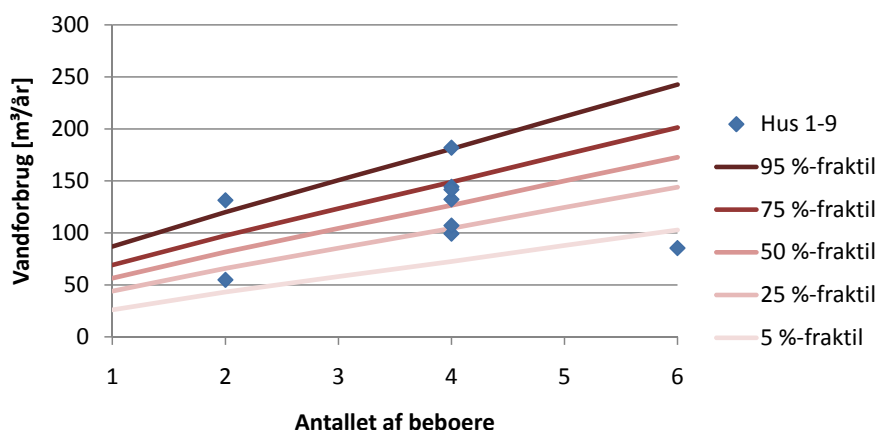
Ud fra de ovenstående parametre er det ikke muligt at undersøge ejerformens betydning i forbindelse med målingerne fra *Fremtidens Parcelhuse*, da alle de udvalgte husstande er ejerboliger.

Det skal bemærkes, at testen udføres på en lille datamængde, hvilket betyder, at der i nogle tilfælde ikke kan ses så tydelige tendenser, når testen sammenholdes med nogle af de tidligere gennemgåede kilder i litteraturstudiet, der har større datamængder til rådighed.

N.2.1 Antal beboere

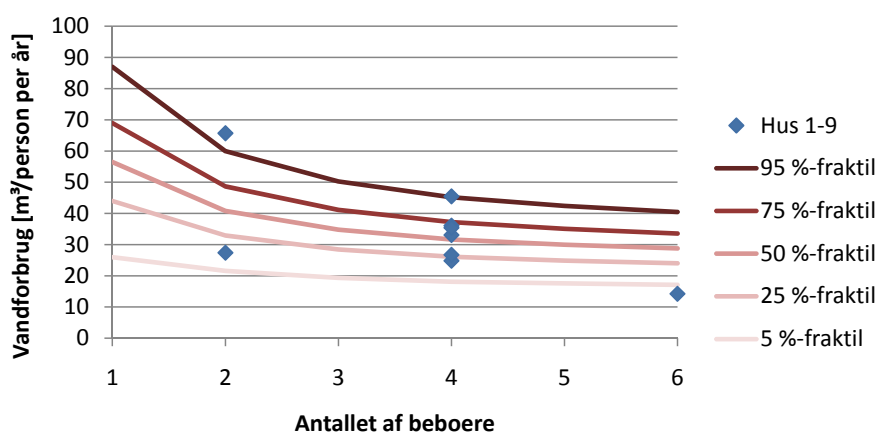
Det har gennem litteraturstudiet i afsnit 3.4 i hovedrapporten vist sig, at antallet af beboere i husstande er en af de parametre, der har den største forklarende effekt i forbindelse med variationen af vandforbruget mellem husstandene. På tilsvarende vis er det i appendiks L gjort klart, at denne parameter er væsentlig i forbindelse med estimeringen af vandforbruget. Af disse årsager undersøges det i det efterfølgende, hvorvidt der kan ses en sammenhæng mellem henholdsvis det samlede vandforbrug og varmtvandsforbruget i ni parcelhuse fra *Fremtiden Parcelhuse*.

Figur N.11 viser de ni boligers samlede vandforbrug sammenholdt med antallet af beboere. Ydermere fremgår 5 %–, 25 %–, 50 %–, 75 %– og 95 %–fraktilen for den sandsynlighedsfordeling af det samlede vandforbrug, der er opstillet og beskrevet i appendiks L.1. Af figur N.11 fremgår der ingen entydig sammenhæng mellem husstandenes samlede vandforbrug og husstandsstørrelsen. Ikke desto mindre vurderes de målte vandforbrug i de ni husstande, at svare udmærket overens med sandsynlighedsfordeling, der er opstillet i dette projekt, hvor der dog er én husstand over 95 %–fraktilen og én husstand under 5 %–fraktilen.



Figur N.11: Skitsering af de ni husstandes samlede vandforbrug fra sammenholdt med husstandsstørrelse.

I appendiks L.1 er der, i forbindelse med opstilling af en sandsynlighedsfordeling for det samlede vandforbrug, korrigeret for en sammenhæng mellem vandforbruget per person og husstandsstørrelsen. Denne korrektion vurderes at være fornuftig set i sammenhæng med målingerne fra de ni husstande, der er vist på figur N.12. På førnævnte figur kan det ses, at tendensen med et lavere vandforbrug per person i store husstande, der er set gennem litteraturstudiet, og tilsvarende ved korrektionen i forbindelse med sandsynlighedsfordelingen, ikke tydeligt gør sig gældende i forbindelse med målingerne for de ni parcelhuse.

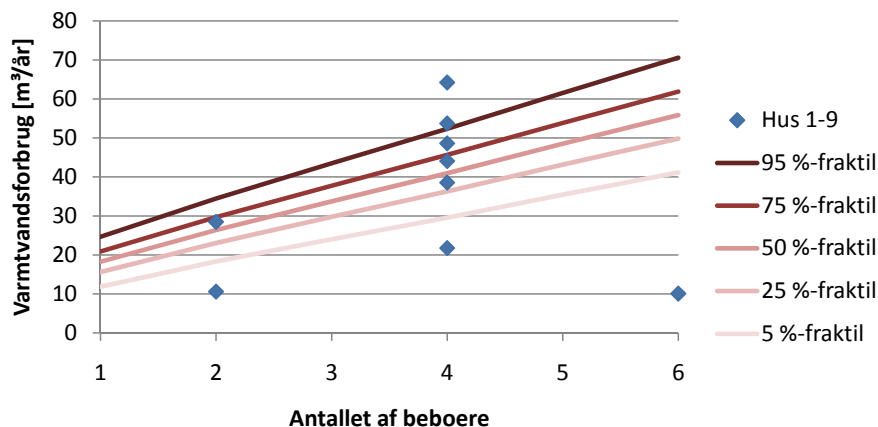


Figur N.12: Skitsering af de ni husstandes samlede vandforbrug opgjort per person sammenholdt med husstandsstørrelse.

I forbindelse med opstilling af sandsynlighedsfordelingen for varmtvandsforbruget i boliger, se appendiks L.2, er det vurderet, på baggrund af oplysninger fra litteraturstudiet, at varmtvandsforbruget udgør omkring 32 % af husstandes samlede vandforbrug. Den opstillede sandsynlighedsfordeling for varmtvandsforbruget er vist på figur N.13, hvor de ni husstandes samlede varmtvandsforbrug er sammenholdt med antallet af beboere i husstandene. På baggrund af det få antal husstande, kan der ikke, ud fra førnævnte figur, ses en entydig sammenhæng mellem husstandsstørrelsen og varmtvandsforbruget. Det fremgår også af figuren, at en simpel reduktion af det samlede vandforbrug ikke er den optimale løsning,

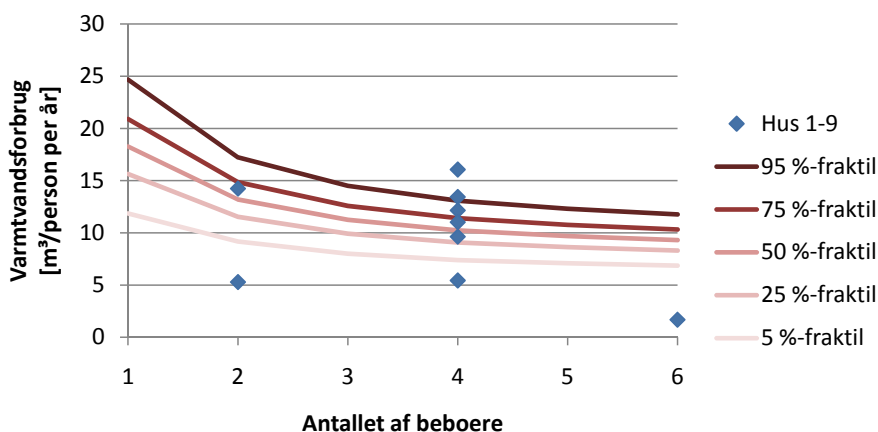
når husstandes varmtvandsforbrug skal vurderes. Det kommer til udtryk ved, at tre ud af de ni målte varmtvandsforbrug er under 5 %-fraktilen, og tilsvarende er to af de målte varmtvandsforbrug over 95 %-fraktilen.

Knudsen et al. [36] har i rapporten om *Fremtidens Parcelhuse* vurderet, at beboerne i husstanden med seks beboere har en specielt adfærd, der gør, at de har et lavt forbrug af varmt brugsvand, når de sammenholdes med de øvrige husstande. Der er blot tale om en vurdering fra *Knudsen et al.*, men det kan ses af figur N.13, at denne husstand er langt under 5 %-fraktilen.



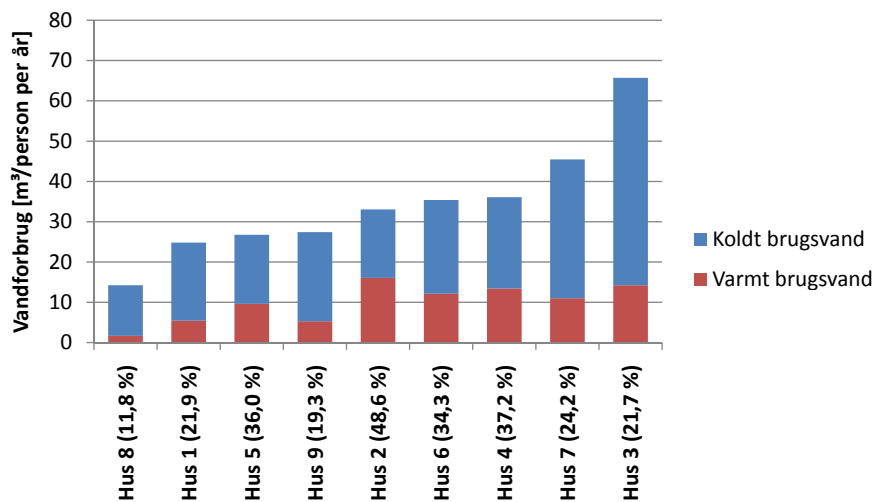
Figur N.13: Skitsering af de ni husstandes varmtvandsforbrug sammenholdt med husstandsstørrelse.

Det undersøges hvorvidt tendensen til et faldende samlet vandforbrug opgjort per person ved stigende husstandsstørrelse, også gør sig gældende for varmtvandsforbruget. Resultatet kan ses på figur N.14, hvor varmtvandsforbruget opgjort per person er sammenholdt med antallet af beboere. Denne undersøgelse giver ikke noget entydigt svar, da spredningen mellem de målte varmtvandsforbrug inden for samme husstandsstørrelser er for stor.



Figur N.14: Skitsering af de ni husstandes varmtvandsforbrug opgjort per person sammenholdt med husstandsstørrelse.

Af ovenstående kan det ses, at der ikke fremkommer nogen entydige resultater. Der er indikationer på, at det samlede vandforbrug er stigende som funktion af en stigende husstandsstørrelse, hvilket også er tilfældet i litteraturstudiet. Tilsvarende kan der ses en tendens til, at det samlede vandforbrug opgjort per person er faldende i takt med stigende husstandsstørrelser. I det tilfælde hvor antallet af beboere sammenholdes med varmtvandsforbruget, fremkommer de førnævnte tendenser mindre tydeligt. I litteraturstudiet er det endvidere beskrevet, at andelen af varmt brugsvand sammenholdt med det samlede vandforbrug, udgør omkring 31 – 34 % i parcelhuse. Hvad angår det tilgængelige data fra de ni husstande viser det, at de gennemsnitligt har et varmtvandsforbrug, for hvad der svarer til omkring 28 % af det samlede vandforbrug. Det er dog ikke ensbetydende med, at de husstande, der har et højere samlet vandforbrug også er de husstande, der forbruger mest varmt brugsvand, som det er vist på figur N.11.



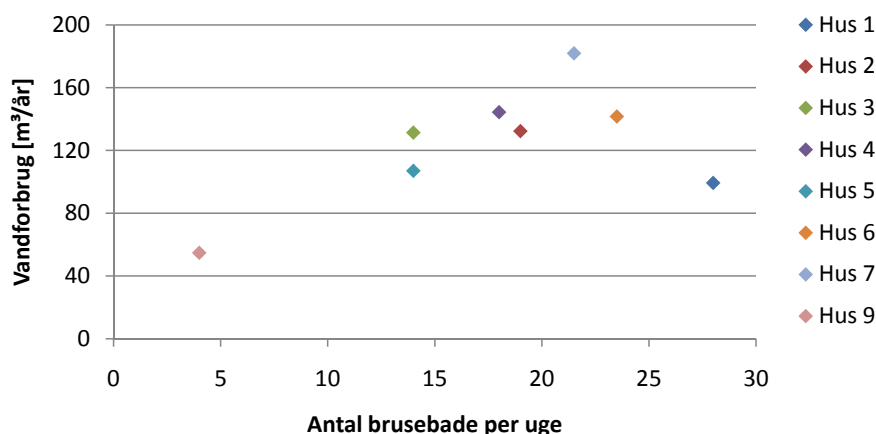
Figur N.15: Forbrug af koldt og varmt brugsvand i ni parcelhuse med angivelse af procentmæssig andel af varmt brugsvand i forhold til husstandenes samlede vandforbrug.

N.2.2 Antal bade

Der har gennem litteraturstudiet været indikationer på, at der er en sammenhæng mellem forbruget af varmt brugsvand og antallet af bade. På figur N.16 er det undersøgt, hvorvidt der kan ses en sammenhæng mellem husstandes samlede vandforbrug og det samlede antal bade. Det bør bemærkes, at *hus 8* ikke fremgår af figuren, hvilket skyldes, at der ikke haves nogen spørgeskemabesvarelse fra den pågældende husstand. I denne undersøgelse er det kun antallet af ugentlige brusebade, der indgår. Det skydes, at der kun én af husstandene tages karbade, hvorfor en undersøgelse af dette ikke vil give meningsfulde resultater. I de tilfælde hvor der foreligger besvarelser fra to personer i samme husstand anvendes en middelværdi af disse.

Af figur N.16 kan der ses en beskeden sammenhæng mellem husstandes samlede vandforbrug og det totale antal bade i husstanden. Det kan dog også ses, at *hus 1* er den husstand, der har det næstlaveste vandforbrug, samtidig med at det er den husstand, hvori der er oplyst det højeste antal ugentlige bade. Årsagen til dette udfald kendes ikke, men af mulige forklaringer kan være, at beboerne har overvurderet antallet af ugentlige bade, da besvarelsen

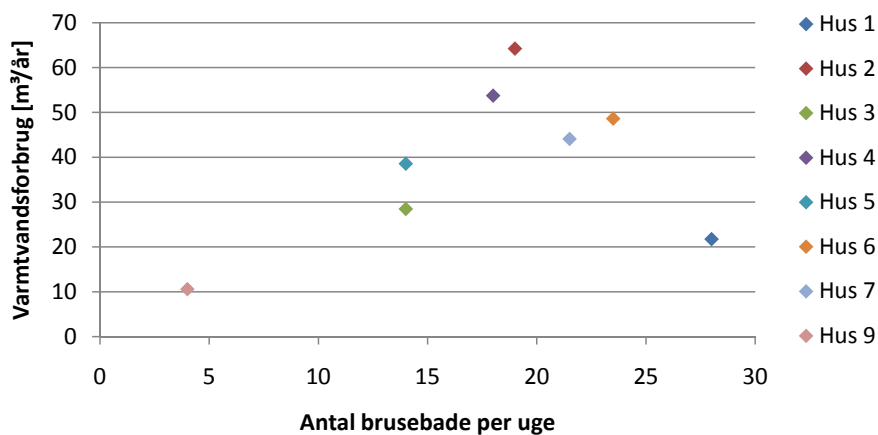
af spørgeskemaet blev udført. Det kan også skyldes, at beboerne har til vane at tage kortere bade sammenholdt med de øvrige husstande. En anden mulighed er, at beboerne i *hus 1* har nogle vaner, hvor de trods alt tager mange bade, men deres øvrige vandforbrug er lavt. Der kan som nævnt være mange mulige årsager, men det er dog bemærkelsesværdigt, at husstanden med det højeste antal ugentlige bade er blandt de husstande, der har det laveste af de registrerede forbrug.



Figur N.16: Skitsering af otte husstandes samlede vandforbrug sammenholdt med det oplyste antal ugentlige bade for de pågældende husstande.

Hvis *hus 1* for en kort stund betragtes som et „ekstrem-tilfælde“, og dermed ekskluderes fra figuren, opstår en bemærkelsesværdig høj sammenhæng mellem husstandes samlede vandforbrug og det totale antal bade. Det bør dog stadig tages i betragtning, at der stadig kun er tale om registreringer fra otte husstande. Endvidere bør det bemærkes, at der ikke haves grundlag for at betragte *hus 1* som et „ekstrem-tilfælde“. Da datamængden i forvejen er beskeden, kan de lige så vel være nogle af de øvrige husstande, der har nogle specielle adfærd og forbrug.

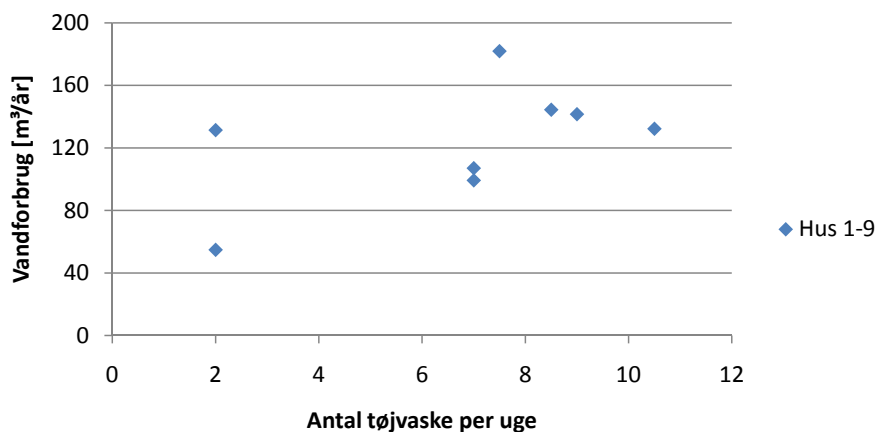
På tilsvarende måde som beskrevet ovenfor, er det undersøgt, hvorvidt der kan ses en sammenhæng mellem husstandes varmtvandsforbrug og det oplyste antal ugentlige bade. Resultatet af denne sammenligning kan ses af figur N.17, hvor der, ligesom ved det samlede vandforbrugt, ikke fremgår en entydig sammenhæng.



Figur N.17: Skitsering af otte husstandes varmtvandsforbrug sammenholdt med det oplyste antal ugentlige bade for de pågældende husstande.

N.2.3 Antal tøjvaske

Antallet af ugentlige tøjvaske har gennem litteraturstudiet vist indikationer på, at det kan have en forklarende effekt i forbindelse med det samlede vandforbrug i husstanden. Af denne årsag er denne sammenhæng undersøgt på figur N.18. I de tilfælde hvor der foreligger besvarelser fra to personer i samme husstand anvendes en middelværdi af disse. På den førnævnte figur kan der ses en tendens til, at de husstande der har et højt antal ugentlige tøjvaske, også er blandt de husstande, der har et højt vandforbrug. Figuren viser dog ikke en entydig sammenhæng eftersom én husstand har oplyst, at de foretager to ugentlige tøjvaske, og en anden husstand har et gennemsnitligt antal tøjvaske på omkring ti, men de har tilnærmelsesvis samme vandforbrug.

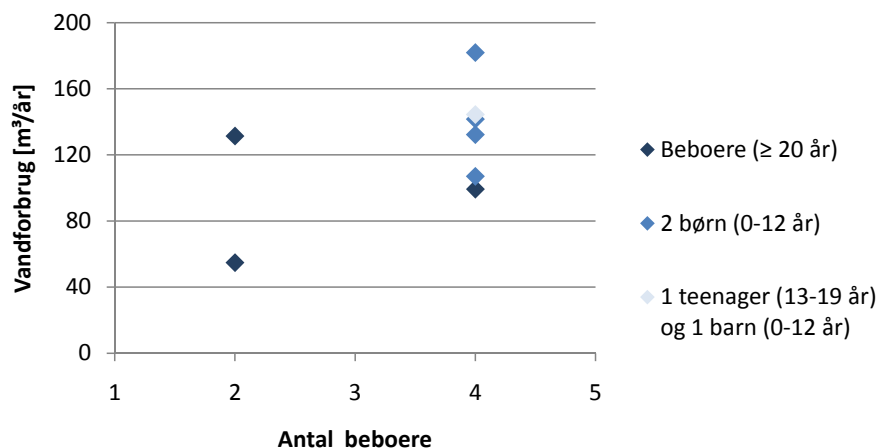


Figur N.18: Skitsering af otte husstandes samlede vandforbrug sammenholdt med det oplyste antal ugentlige tøjvaske for de pågældende husstande.

Antallet af tøjvaske er ikke sammenholdt med varmtvandsforbruget, eftersom størstedelen af vaskemaskinerne på det danske marked tilsluttes det kolde brugsvand, hvorfor der ofte ikke bør være en direkte sammenhæng mellem antallet af tøjvaske og varmtvandsforbruget.

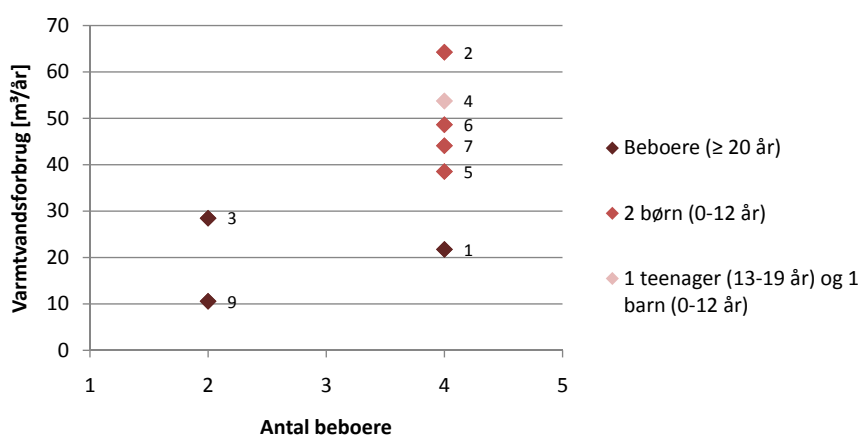
N.2.4 Tilstedeværelse af børn og teenagere

I litteraturstudiet fremgår der at være en sammenhæng mellem antallet af teenagere og vandforbruget i de pågældende husstande. Ud fra spørgeskemabesvarelsene er der foretaget en opdeling af husstandene ud fra beboernes alder. Der er to husstande med to voksne beboere, samt en husstand med to voksne og to børn, der er over teenagealderen. Desuden er der fire husstande, som alle har to voksne og to børn under 12 år, samt en enkelt husstand med to voksne, ét barn og én teenager. Foruden de førnævnte husstande er der én af de ni husstande, der ikke har besvaret spørgeskemaet, hvorfor der ikke haves kendskab til beboersammensætningen i denne. På figur N.19 kan der ses en sammenligning mellem beboersammensætningen og husstandenes samlede vandforbrug. For husstandene med fire beboere fremgår det, at husstandene børn eller teenagere ligger højere end husstanden med beboere, der er udenfor de førnævnte kategorier. Dette tilfælde kan dog ikke betragtes som et generelt tilfælde, da der er alt for få besvarelses til at underbygge det. Tilsvarende kan det ses, at en husstand med to beboere har højere vandforbrug end nogle husstande med fire beboere.



Figur N.19: Skitsering af otte husstandes samlede vandforbrug sammenholdt med antallet af børn og teenagere for de pågældende husstande.

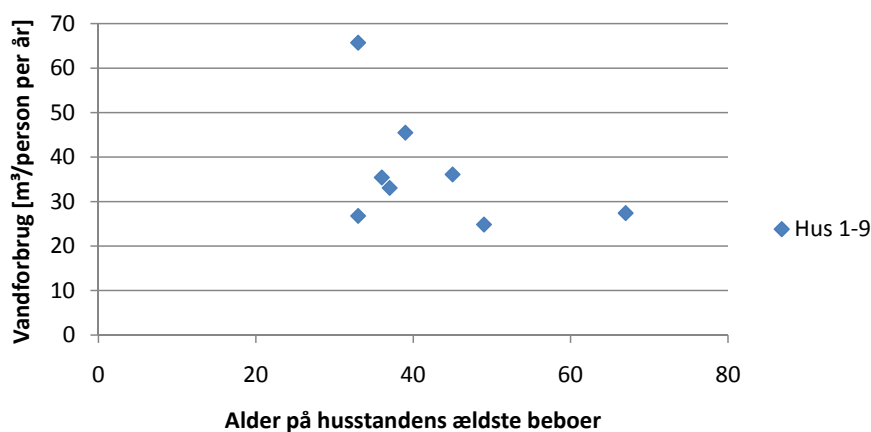
Hvad angår varmtvandsforbruget, er der en tydeligere adskillelse mellem husstande med børn eller teenagere og de øvrige husstande, hvor de førstnævnte har det højeste varmtvandsforbrug.



Figur N.20: Skitsering af otte husstandes varmtvandsforbrug sammenholdt med antallet af børn og teenagere for de pågældende husstande.

N.2.5 Beboernes alder

Det er undersøgt hvorvidt alderen på den ældste person i husstanden har en sammenhæng med henholdsvis husstandens samlede vandforbrug og varmtvandsforbruget. På figur N.21 er alderen på den ældste person i husstanden sammenholdt med det samlede vandforbrug, hvor der ikke ses en tydelig sammenhæng. I hovedparten af husstandene er den ældste beboer omkring 40 år, hvilket vanskeliggør en sammenligning, da der derved ikke er så stor en spredning husstandene imellem.



Figur N.21: Skitsering af alderen på den ældste beboer i otte husstandes sammenholdt med det samlede vandforbrug opgjort per person i de pågældende husstande.

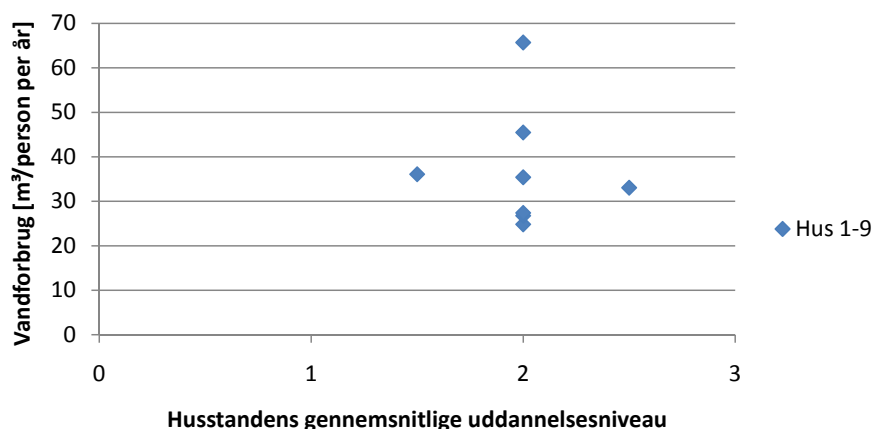
Det samme forhold gør sig gældende i sammenligningen af varmtvandsforbruget og alderen på den ældste beboer i husstandene, hvilket er skitseret på figur N.22.



Figur N.22: Skitsering af otte husstandes varmtvandsforbrug opgjort per person sammenholdt med antallet af børn og teenagere for de pågældende husstande.

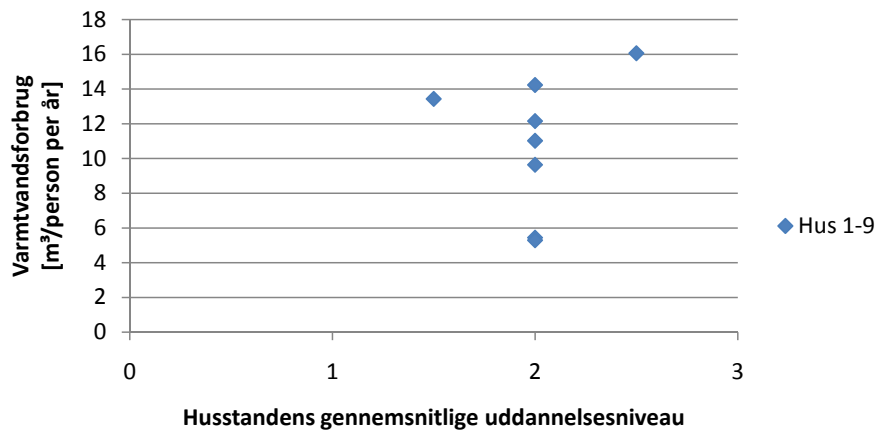
N.2.6 Uddannelsesniveau

Det er undersøgt hvorvidt der kan ses en sammenhæng mellem uddannelsesniveauet og henholdsvis det samlede vandforbrug og varmtvandsforbruget. Eftersom der i spørgeskemaet er stillet spørgsmål til hvorvidt beboerne har en kort, mellemlang eller lang uddannelse, gives der point efter denne skala. Det betyder at uddannelsesniveauet er opgjort således at en kort uddannelse giver ét point, en mellemlang uddannelse giver to point og en lang uddannelse giver tre point. Såfremt der er flere besvarelser fra samme husstand er der anvendt en midelværdi af de individuelle uddannelsesniveauer. Det fremgår af figur N.23, at der ikke kan ses en direkte sammenhæng mellem husstandens gennemsnitlige uddannelsesniveau og det samlede vandforbrug.



Figur N.23: Skitsering af gennemsnitligt uddannelsesniveau i otte husstande sammenholdt med det samlede vandforbrug opgjort per person i de pågældende husstande.

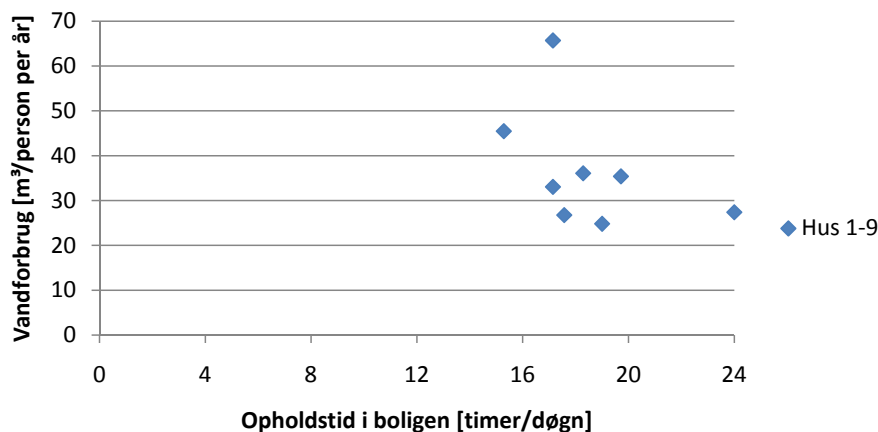
På tilsvarende vis fremgår der heller ikke en direkte sammenhæng mellem husstandens gennemsnitlige uddannelsesniveau og varmtvandsforbruget, hvilket er skitseret på figur N.24.



Figur N.24: Skitsering af gennemsnitligt uddannelsesniveau i otte husstande sammenholdt med det varmtvandsforbruget opgjort per person i de pågældende husstande.

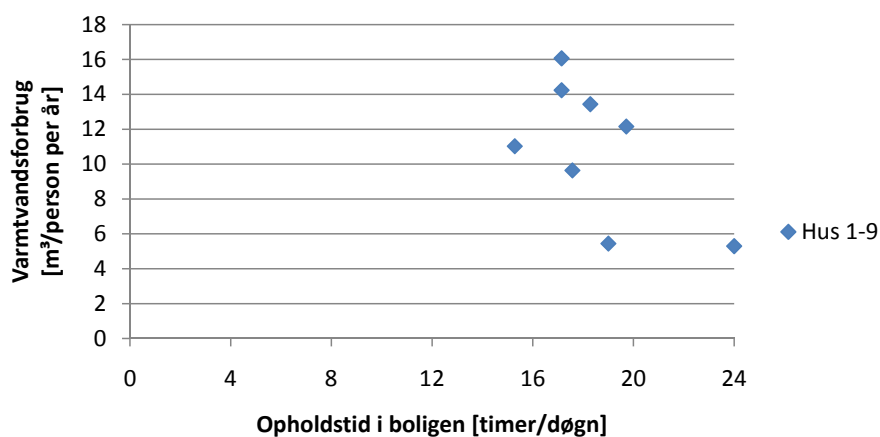
N.2.7 Opholdstid i boligen

Det er undersøgt hvorvidt beboernes opholdstid i boligen har en forklarende effekt i forhold til henholdsvis husstandens samlede vandforbrug og varmtvandsforbruget. I spørgeskemaet har beboerne svaret på hvor mange timer de opholder sig i boligen på henholdsvis hverdage og i weekender. Såfremt flere beboere har besvaret spørgeskemaet, er der anvendt den længste opholdstid, der er oplyst. Opholdstiden er endvidere beregnet som et vægtet gennemsnit, således at opholdstiden i hverdagene tæller 5/7 mens opholdstiden i weekenden tæller 2/7. På figur N.25 kan der ses en sammenligning mellem husstandes samlede vandforbrug og opholdstiden. Der ses ikke nogen tydelig sammenhæng mellem disse, hvor også en lille spredning for opholdstiden vanskeliggør en sammenligning.



Figur N.25: Skitsering af beboernes oplyste opholdstid i otte husstande sammenholdt med det samlede vandforbrug opgjort per person i de pågældende husstande.

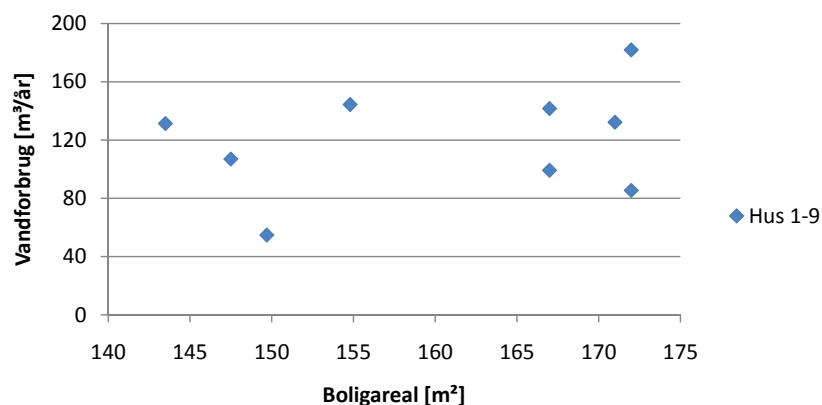
Tilsvarende er en sammenligning af opholdstiden og varmtvandsforbruget vanskelig, da spredningen mellem opholdstiderne mellem husstandene er lille, som det kan ses på figur N.26



Figur N.26: Skitsering af beboernes oplyste opholdstid i otte husstande sammenholdt med varmtvandsforbruget opgjort per person i de pågældende husstande.

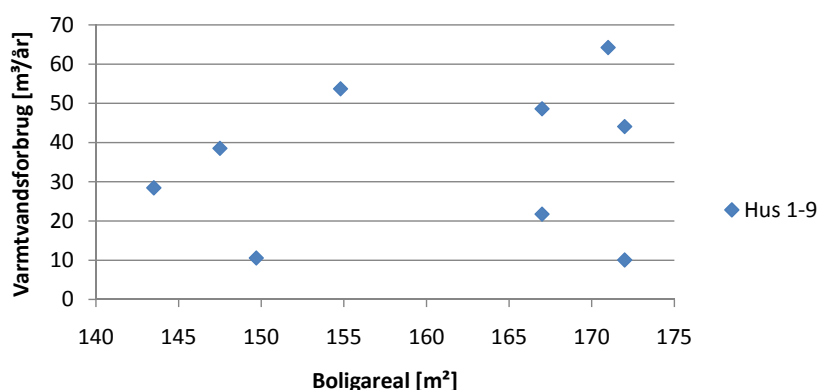
N.2.8 Boligareal

Det er undersøgt hvorvidt boligarealet kan have en forklarende effekt i forhold til henholdsvis husstandens samlede vandforbrug og varmtvandsforbruget. Af figur N.27 fremgår der ingen nævneværdig sammenhæng mellem boligarealet og det samlede vandforbrug.



Figur N.27: Skitsering af boligarealet i ni husstande sammenholdt med det samlede vandforbrug i de pågældende husstande.

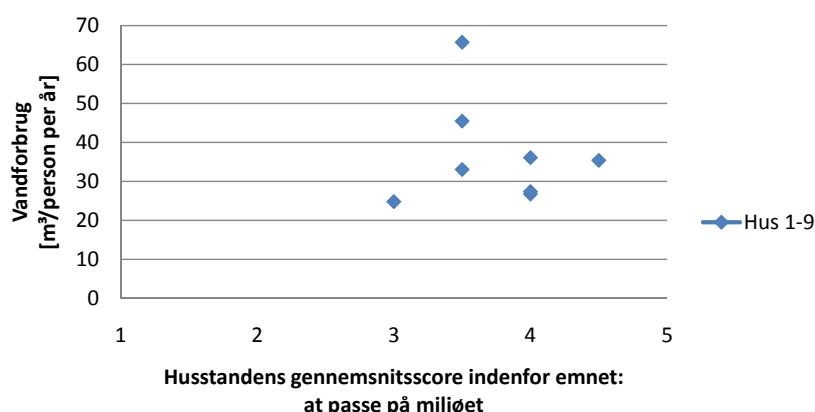
Tilsvarende kan der heller ikke ses nogen nævneværdig sammenhæng mellem boligarealet og varmtvandsforbruget.



Figur N.28: Skitsering af boligarealet i ni husstande sammenholdt med varmtvandsforbruget i de pågældende husstande.

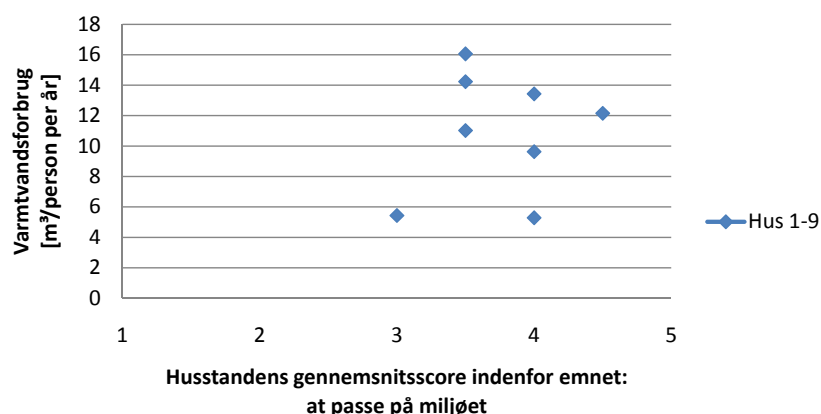
N.2.9 Beboernes holdning til miljø og energibesparelser

I spørgeskemaet har beboerne svaret på i hvor høj grad de er optaget af at passe på miljøet. Besvarelsene er udført på en skala fra ét til fem, hvor ét betyder, at de er lidt optaget af at passe på miljøet, og ved fem er de meget optaget af det. Figur N.29 fremgår der ingen sammenhæng mellem beboernes holdning til hvorvidt de er optaget af at passe på miljøet og de pågældende husstandes samlede vandforbrug.



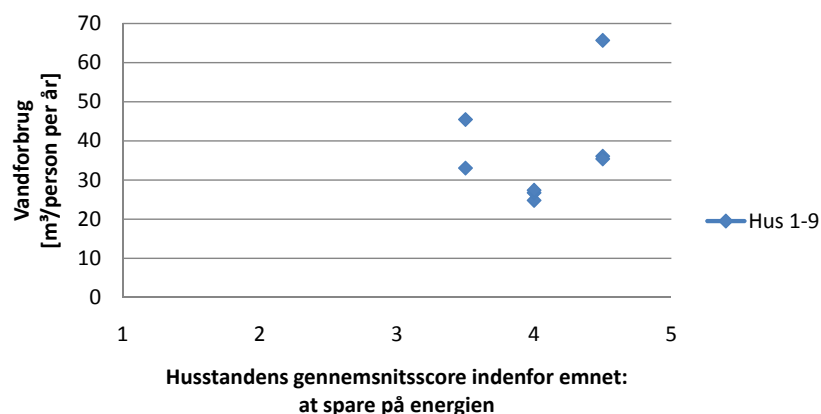
Figur N.29: Skitsering af beboernes holdning til at passe på miljøet sammenholdt med det samlede vandforbrug opgjort per person i de pågældende husstande.

Sammenholdes beboerens holdning med husstandenes varmtvandsforbrug kan der heller ikke i dette tilfælde ses en entydig sammenhæng, hvilket er skitseret på figur N.30.



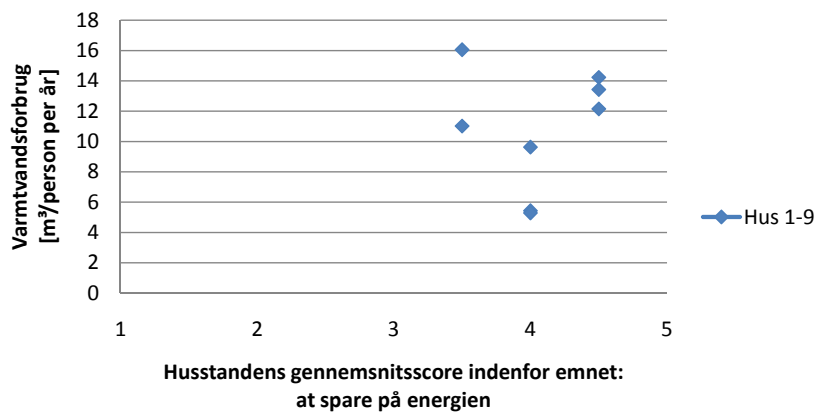
Figur N.30: Skitsering af beboernes holdning til at passe på miljøet sammenholdt med varmtvandsforbruget opgjort per person i de pågældende husstande.

På tilsvarende vis har beboerne svaret på i hvor høj grad de er optaget af at spare på energien. Besvarelsene er, ligesom ovenstående, udført på en skala fra ét til fem, hvor ét betyder, at de er lidt optaget af at spare på energien, og ved fem er de meget optaget af det. Figur N.31 fremgår der ingen sammenhæng mellem beboernes holdning og de opgældende husstandes samlede vandforbrug.



Figur N.31: Skitsering af beboernes holdning til at passe på miljøet sammenholdt med det samlede vandforbrug opgjort per person i de pågældende husstande.

Hvad angår varmtvandsforbruget kan der heller ikke i dette tilfælde ses en sammenhæng mellem dette og beboernes holdning til at spare på energien, som skitseret på figur N.32.



Figur N.32: Skitsering af beboernes holdning til at passe på miljøet sammenholdt med varmtvandsforbruget i de pågældende husstande.

N.3 Elforbrug

I litteraturstudiet omhandlende boligers elforbrug er der i nogle projekter undersøgt hvilke forhold, der kan have en forklarende effekt i forbindelse med husstandenes elforbrug. I det efterfølgende beskrives en test på måledata fra projektet *Fremtidens Parcelhuse* [36], hvor det undersøges hvorvidt de forklarende parametre fra litteraturstudiet også gør sig gældende for den aktuelle datamængde. Foruden en test af disse parametre, er der foretaget en sammenligning af elforbruget og af andre forhold i forbindelse med boligen og beboerne, der kan have en forklarende effekt. I afsnit 3.3 i hovedrapporten er der nævnt nogle parametre, der har en forklarende effekt på elforbruget, og disse er opsummeret nedenfor:

- Antal beboere
- Boligens størrelse
- Indkomst
- Beboernes alder
- Tilstedeværelse af børn og teenagere
- Apparat bestand og brugstid
- Beboeres viden om elforbrug

Parametrene *Indkomst* og *Beboernes viden om elforbruget* er ikke mulige at undersøge på basis af data fra *Fremtidens Parcelhuse*, da disse parametre ikke har indgået i undersøgelsen. Ligeledes kan *apparatbestand* og *brugstid* heller ikke vurderes direkte, derimod undersøges sammenhængen mellem elforbruget og mængden af tøjvaske/tørring af tøj samt beboernes opholdstid i boligen. Ligeledes undersøges det også om der er en sammenhæng mellem elforbruget og beboernes uddannelsesniveau.

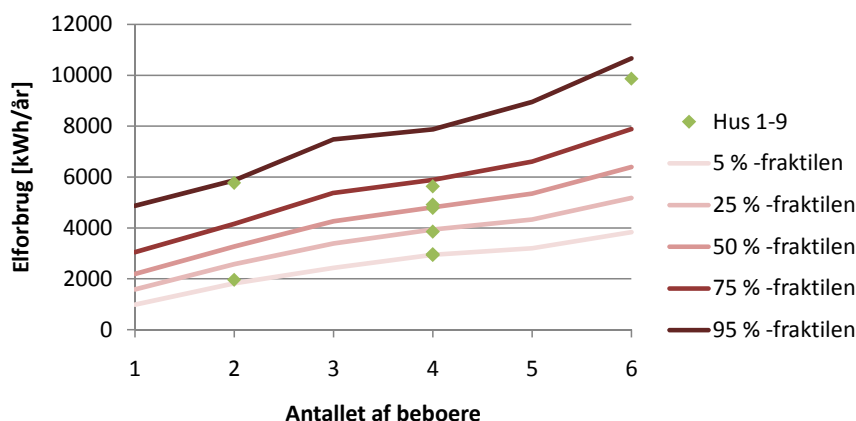
Det skal bemærkes, som i undersøgelserne af varme og vand i appendiks N.1 og N.2, at testen udføres på en lille datamængde, hvilket betyder, at der i nogle tilfælde ikke kan ses så

tydelige tendenser, når testen sammenholdes med nogle af de tidligere gennemgåede kilder i litteraturstudiet, der har haft større datamængder til rådighed. Samtidig skal den fundne sammenhæng heller ikke overfortolkes, da det ikke vides hvorledes denne sammenhæng vil figurere ved test på større datamængder.

N.3.1 Antal beboere

Det har gennem litteraturstudiet, i afsnit 3.3 i hovedrapporten, vist sig, at antallet af beboere i boligerne har en stor forklarende effekt i forbindelse med variationen af elforbruget i boligerne. Ligeledes fremgår det, af figur K.1 i appendiks K, at antallet af beboere, når variabelen er vurderet alene, har den største forklarende effekt i forhold til de øvrige parametre, hvilket gør den væsentlig i forbindelse med estimering af elforbruget. Af disse årsager undersøges det, i det efterfølgende, hvorvidt der kan ses en sammenhæng mellem henholdsvis elforbruget og antallet af beboere i de ni parcelhuse fra *Fremtidens Parcelhuse*.

Figur N.33 viser de ni boligers elforbrug sammenholdt med antallet af beboere. Ydermere fremgår 5 %-, 25 %-, 50 %-, 75 %- og 95 %-fraktilen for den sandsynlighedsfordeling af det samlede elforbrug, der er opstillet og beskrevet i appendiks K. I figur N.33 fremgår der en rimelig entydig stigning i husstandenes elforbrug, som funktion af en stigende husstandsstørrelse, dog med stor variation indenfor samme husstandsstørrelse. Dette ses specielt i boliger med to beboere, hvor forbruget, i de to husstande, enten ligger tæt ved 95 %-fraktilen eller 5 %-fraktilen. Det vurderes derudover, at de målte elforbrug i de ni husstande svarer godt overens med sandsynlighedsfordelingerne, der er opstillet i nærværende projekt, jf. appendiks K.

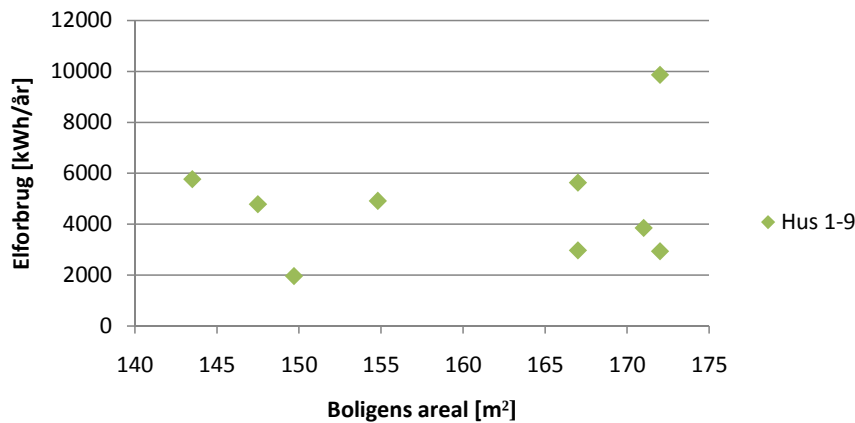


Figur N.33: Illustrering af de ni husstandes samlede elforbrug sammenholdt med antallet af beboere.

N.3.2 Boligareal

I flere projekter, fra bl.a. *Jesper Ole Jensen* [35] og *Pedersen et al.* [56], der er gennemgået i litteraturstudiet blev det påvist, at der er en sammenhæng mellem boligens størrelse og elforbrugets størrelse. På figur N.34 ses det, at der i de ni præsenterede boliger ikke ses en nævneværdig sammenhæng mellem det samlede elforbrug og boligens størrelse. Ved nærmere

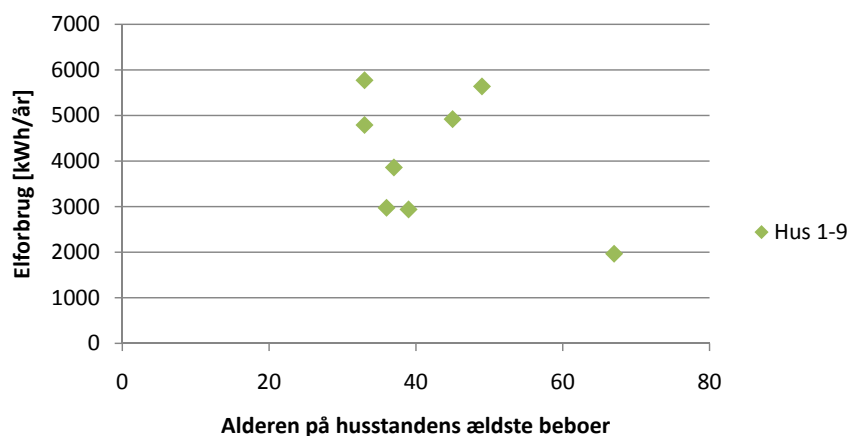
vurdering af figur 3.21 i afsnit 3.3 ses det at der er stor variation i grafen mht. elforbruget i samme størrelse bolig, dermed kan det ikke afvises, at samme tendens kan være gældende for boligerne i *Fremtidens Parcelhuse*, dette kan dog ikke konkluderes grundet størrelsen af den tilgængelige datamængde.



Figur N.34: Illustrering af de ni husstandes samlede elforbrug sammenholdt med boligens størrelse.

N.3.3 Beboernes alder

I litteraturstudiet afsnit 3.3 beskrives sammenhængen mellem elforbrug og husstandens ældste beboer vha. en kurvelineær linje, hvor forbruget i parcelhuse er stigende for perioden mellem de 20 – 50 år, herefter er den igen faldende. Det er undersøgt, for data fra *Fremtidens Parcelhuse* hvorvidt alderen på den ældste person i husstanden har en sammenhæng med elforbruget. På figur N.35 ses resultatet af testen, hvor det fremgår, at der ikke ses en tydelig sammenhæng. I hovedparten af husstandene er den ældste beboer omkring 40 år, hvilket vanskeliggør en sammenligning, da der derved ikke er så stor en spredning husstandene imellem.

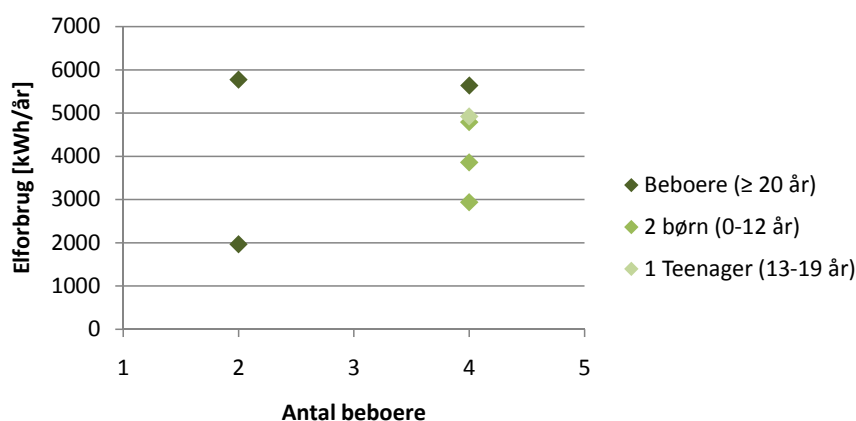


Figur N.35: Illustrering af sammenhængen mellem elforbruget og den ælste beboers alder.

N.3.4 Tilstedeværelse af børn og teenagere

Fra afsnit 3.3 i litteraturstudiet vides det, at småbørn og børn har en positiv (faldende) indflydelse på elforbruget, hvorimod teenagere har en negativ (stigende) indflydelse på elforbruget, i forhold til hvis alle i boligen var voksne. Denne tendens testes, hvor der ved nærværende test kun adskilles mellem tre kategorier: børn (0 – 12 år), teenagere (13 – 19 år) og voksne (≥ 20 år). Figur N.36 illustrerer sammenhængen mellem tilstedeværelsen af børn og teenagere og det samlede elforbrug. Testen kan ikke udføres i boliger fra *Fremtidens Parcelhuse* med to beboere, idet der ikke er børn eller teenagere tilstede.

For boliger med fire beboere ses det, at de familier med børn, er de husstande der har de laveste forbrug, hvilket stemmer godt overens med resultaterne fra *Kirsten Gram-Hanssen* [26]. Ud fra det indsamlede data fra *Fremtidens Parcelhuse* er der kun én familie med fire voksne og kun én familie med én teenager og ét barn. Familien med én teenager og ét barn ligger højere end de øvrige familier med to børn, denne type familie hvor begge kategorier af hjemmeboende børn er tilstede er ikke vurderet af *Kirsten Gram-Hanssen*, dog viste undersøgelsen at teenagere forbruger mere end børn og dermed kan det forhøjede forbrug i forhold til de øvrige husstande forklares. Familien med to hjemmeboende voksne børn har det højeste elforbrug. Ifølge undersøgelsen kan familien med fire voksne ikke sammenlignes med familien med ét barn og én teenager, idet ingen af formlerne beskriver denne sammensætning i familierne. Dog skal det bemærkes, at familien med fire beboere består af to voksne og to hjemmeboende voksne børn på hhv. 20 og 22 år, det kan diskuteres om disse ikke stadig har vaner og forbrug som en teenager og det er derfor at familien bestående af kun voksne har det største forbrug.

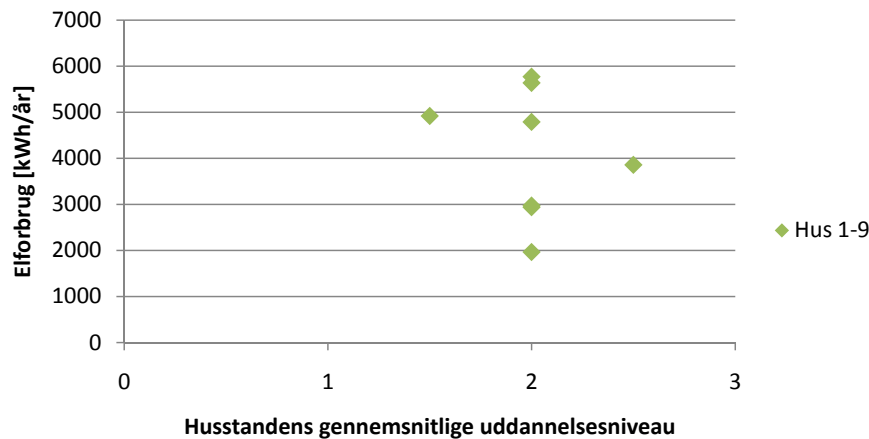


Figur N.36: Illustrering af sammenhængen mellem energiforbrug og tilstedeværelsen af børn og teenagere.

Inddelingen af børnetyper, der er anvendt i undersøgelsen fra *Kirsten Gram-Hanssen*, kunne evt. revurderes med henblik på at inddele dem i to grupper i stedet for tre: mindre børn (0–12 år) og store børn (13–når de flytter hjemmefra). Om aldersgrænsen, mellem de to grupper af børn, skal placeres lige ved 12 – 13 år er en vurderingssag, men idet ens hverdag ændrer sig omkring teenagealderen både mht. fritid og andre interesser, anses denne alder som anvendelig for inddelingen. Årsagen til at der ikke tilkendes et slutår ved gruppen for store børn skyldes, at de vaner børn tillægger sig i teenageårene sjældent laves om før børnene flytter hjemmefra.

N.3.5 Uddannelsesniveau

Det er undersøgt hvorvidt der kan ses en sammenhæng mellem uddannelsesniveauet og elforbruget. Eftersom der i spørgeskemaet er stillet spørgsmål til hvorvidt beboerne har en kort, mellemlang eller lang uddannelse, gives der point efter denne skala. Det betyder at uddannelsesniveauet er opgjort således at en kort uddannelse giver ét point, en mellemlang uddannelse giver to point og en lang uddannelse giver tre point. Såfremt der er flere besvarelser fra samme husstand er der anvendt en middelværdi af de individuelle uddannelsesniveauer. Det fremgår af figur N.37, at der ikke kan ses en direkte sammenhæng mellem husstandens gennemsnitlige uddannelsesniveau og elforbruget.

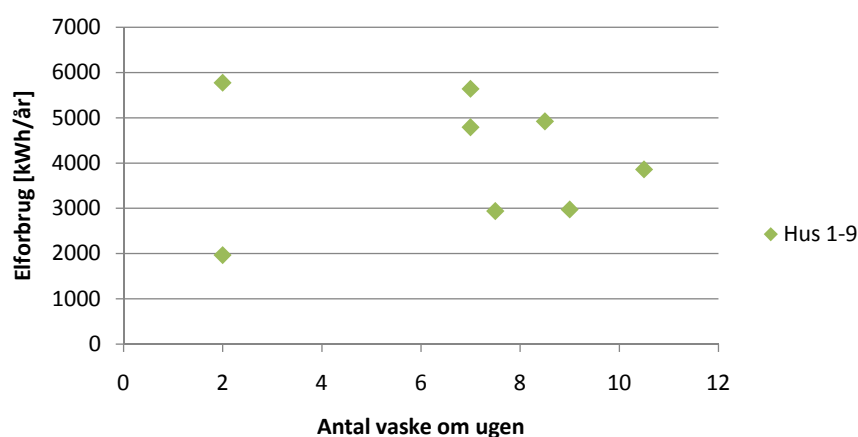


Figur N.37: Illustrering af sammenhængen mellem elforbruget og husstandens gennemsnitlige uddannelsesniveau.

N.3.6 Antal tøjvaske og tørring af tøj

Som beskrevet tidligere er der i andre undersøgelser, der er gennemgået i afsnit 3.3 i litteraturstudiet fundet en sammenhæng mellem apparatbestandens størrelse og elforbruget samt en sammenhæng mellem apparaternes brugstid og elforbruget. Hvor apparaterne der havde vist den største betydning var: vaskemaskine, tørretumbler, tv/video og opvaskemaskinen. I spørgeskemaundersøgelsen fra *Fremtidens Parcelhuse* er der ikke lagt vægt på at undersøge apparatbestandens størrelse, dog er det i stedet blevet undersøgt hvor ofte der vaskes/tørres tøj og hvilke foranstaltninger beboerne har for at reducere elforbruget i boligen, derfor er det disse variable der testes for data fra *Fremtidens Parcelhuse*.

I analysen er beboerne blevet spurgt om, hvor mange gange om ugen de vasker deres tøj. Der er i flere af husstandene angivet forskellige antal af vaske indenfor samme husstand, hvilket kan indikere, at det enten er en delt opgave hvormed overblikket forsvinder for begge parter, en anden mulighed er, at det er en opgave en af parterne står for hvormed den anden part angiver et mindre kvalificeret gæt. Idet det ikke fremgår af ovenstående, hvilken en af parterne der har givet den bedste vurdering af mængden af vaske om ugen, anvendes der i vurderingen et gennemsnit af besvarelserne. Resultatet af analysen ses på figur N.38, her fremgår det, at der ingen klar sammenhæng er mellem elforbruget og antal af tøjvaske om ugen. Idet der er set en sammenhæng i andre projekter, jf. litteraturstudiet, bør der indsamles yderligere data om dette samt overveje om mængden af vaske kan registreres på en alternativ måde.



Figur N.38: Illustrering af sammenhængen mellem elforbruget og det ugentlige antal tøjvaske.

Fra litteraturstudiet vides det, at specielt anvendelsen af tørretumbleren havde betydning for elforbrugets størrelse. I undersøgelsen blev det undersøgt på hvilke måder folk tørrer deres tøj efter tøjvask. Spørgeskemaet var udformet således at beboerne angav på hvilke måder tørringen foregik og dermed ikke hvor ofte de anvendte de forskellige måder. Alle husstande, med undtagelse af én, angav at de anvendte både tørresnor og tørretumbler. Uden en angivelse af et forhold mellem de forskellige måder hvorpå tøj tørres, er der ikke rigtig basis for, at illustrere en sammenhæng. Dog bemærkes det, at den eneste husstand der ikke anvender tørretumbleren har et mindre elforbrug end de øvrige husstande.

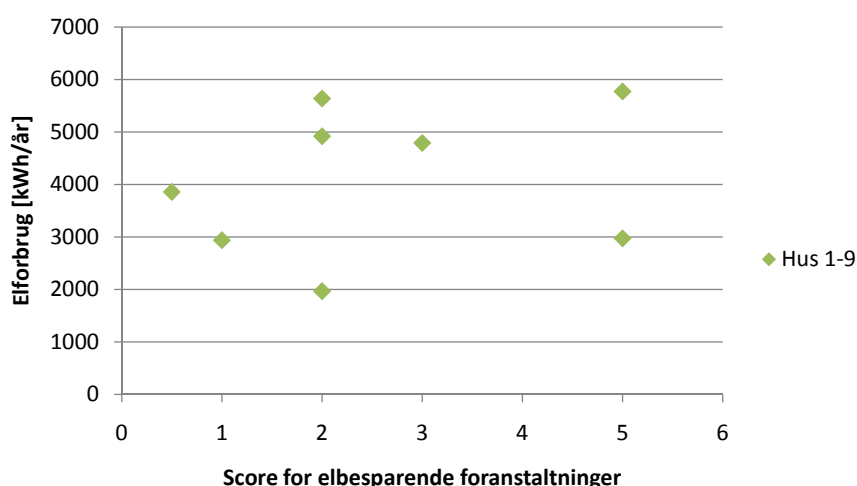
For at kunne analysere sammenhængen mellem elforbruget og apparatbestandens størrelse, i fremtidige projekter, er det nødvendigt at undersøge hvor ofte og hvor længe apparaterne anvendes, idet der kan være stor variation i hvor ofte og ikke mindst hvor længe forskellige familier anvender deres elektriske apparater.

N.3.7 Anvendelse af elbesparende foranstaltninger

I analysen fra *Fremtidens Parcelhuse* fremgår det, at de har undersøgt hvilke elbesparende foranstaltninger, der anvendes i de forskellige husstande. De foranstaltninger der undersøges omhandler følgende:

- Bruges der lavenergipærer
- Tænd/sluk ure
- Bevægelsesfølere
- Elspareskinne ved pc
- Elspareskinne ved tv
- IHC

I spørgeskemaet har beboerne svaret på hvilke foranstaltninger, der anvendes i boligen. Hvert svar, hvor de angiver, at de anvender en af ovenstående, tildeles ét point. Hvis der har været uenighed omkring besvarelsenerne indenfor samme husstand, tildeles der kun et halvt point. Alle pointene summeres, og den samlede score sammenholdes med elforbruget i boligen. Resultaterne fra undersøgelsen ses i figur N.39, hvoraf det fremgår, at der ikke ses nogen sammenhæng mellem de anvendte elspareforanstaltninger og elforbruget.

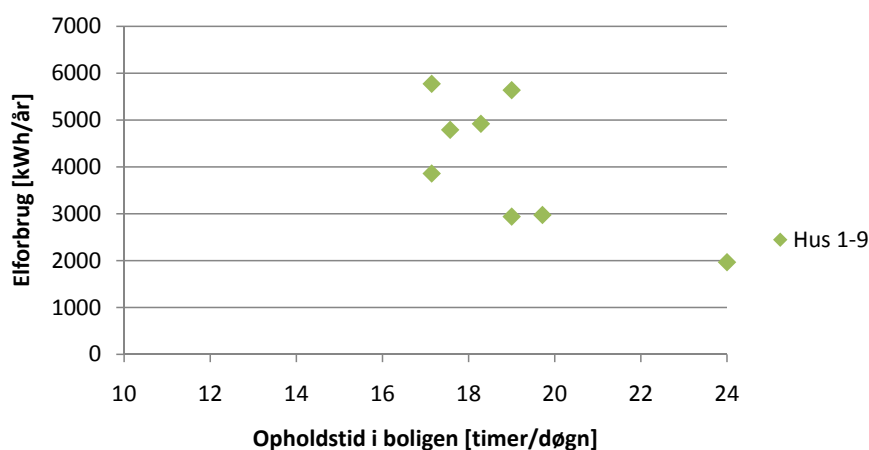


Figur N.39: Illustrering af sammenhængen mellem elforbruget og husstandens score for elbesparende foranstaltninger.

Det er ligeledes forsøgt at sammenholde de enkelte elbesparende foranstaltninger med elforbruget, hver for sig, men dette viser heller ingen sammenhæng.

N.3.8 Opholdstid i boligen

Det er undersøgt hvorvidt beboernes opholdstid i boligen har en forklarende effekt i forhold til husstandens elforbrug. I spørgeskemaet har beboerne svaret på hvor mange timer de opholder sig i boligen på henholdsvis hverdage og i weekender. Såfremt flere beboere har besvaret spørgeskemaet, er der anvendt den længste opholdstid, der er oplyst. Opholdstiden er endvidere beregnet som et vægtet gennemsnit, således at opholdstiden i hverdage tæller 5/7, mens opholdstiden i weekenden tæller 2/7. På figur N.40 kan der ses en sammenligning mellem husstandes elforbrug og opholdstiden. Der ses ikke nogen tydelig sammenhæng mellem disse, idet en lille spredning for opholdstiden vanskeliggør en sammenligning.



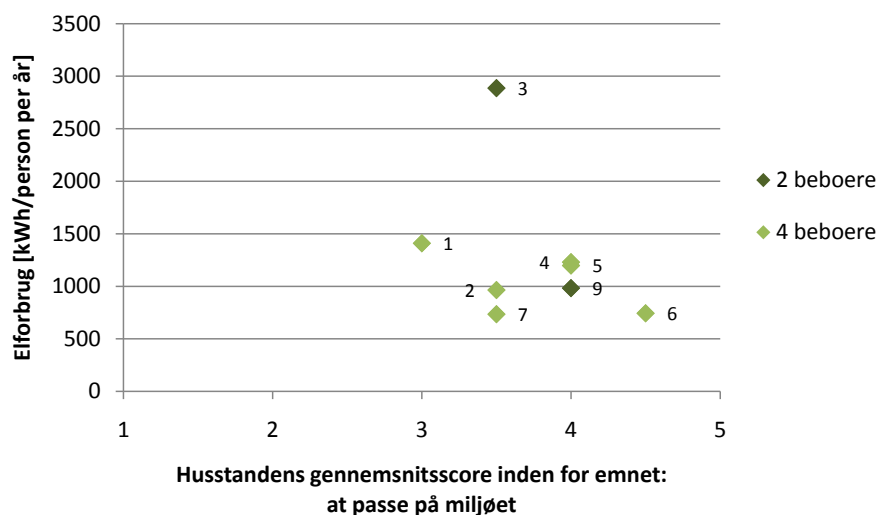
Figur N.40: Illustrering af sammenhængen mellem elforbruget og den vægtede gennemsnitlige daglige opholdstid i boligen.

N.3.9 Beboernes holdning til miljø og energibesparelser

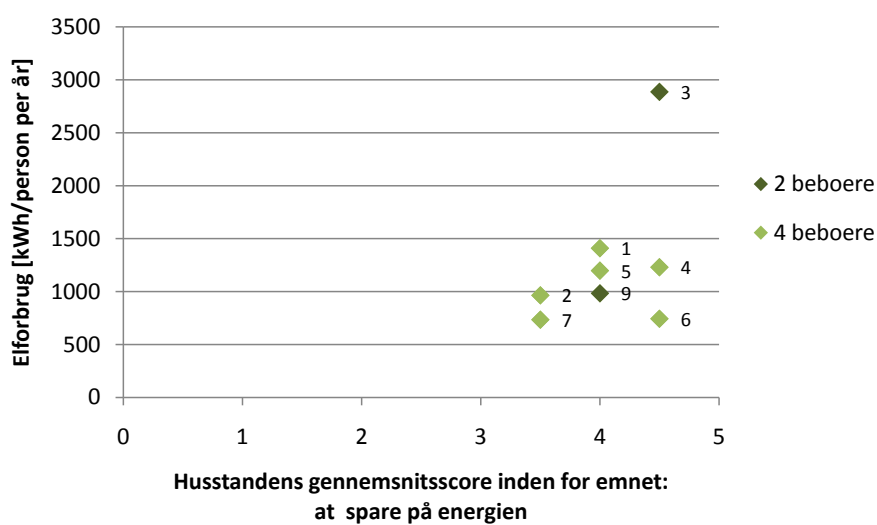
I analysen fra *Fremtiden Parcelhuse* er beboernes forhold til spørgsmålene nedenfor blevet vurderet og sammenlignet med elforbruget i boligen.

- Er du optaget af at passe på miljøet?
- Er du optaget af at spare på energien?

Beboerne blev stillet ovenstående spørgsmål og blev herefter bedt om at graduere betydningen af deres svar med en karakter mellem 1 – 5, hvor karakteren 1 er *lidt* og karakteren 5 er *meget*. Da henholdsvis 93,8 % og 97,9 % svarer **Ja** på begge ovenstående spørgsmål er det svært ud fra deres umiddelbare besvarelse at skulle vurdere størrelsen af deres elforbrug. Derfor undersøges det, om der kan findes en sammenhæng mellem deres graduerede svar indenfor hvert område og det samlede elforbrug. Da flere beboere indenfor samme husstand har svaret på spørgeskemaerne, og de flere steder angiver forskellige svar, er graferne udviklet på baggrund af husstandens gennemsnitlige svar. Det skal dog nævnes, at sammenhængen er undersøgt for både højeste og mindste score i husstanden, for begge spørgsmål, og samme tendens viser sig for begge undersøgelser. Resultatet af analyserne kan ses i figur N.41 og N.42.



Figur N.41: Illustrering af sammenhængen mellem elforbruget og beboernes gennemsnitsscore angående spørgsmålet „at passe på miljøet“.



Figur N.42: Illustrering af sammenhængen mellem elforbruget og beboernes gennemsnitsscore angående spørgsmålet „at spare på energien“.

Ud fra figur N.41 ses der ingen klar sammenhæng mellem elforbrugets størrelse per beboer og husstandens gennemsnitsscore indenfor emnet „at passe på miljøet“. I figur N.42 ses der heller ingen sammenhæng mellem elforbruget og husstandens gennemsnitsscore inden for emnet „at spare på energien“.

Beskrivelse af attributter til klassificeringstræerne

O.1 Tekniske attributter

De tekniske attributter knytter sig til boligens egenskaber. Af disse gennemgås forhold, der gennem litteratur- og casestudiet har vist indikationer på, at de kan forklare dele af forbrugsvariationerne for enten energiforbruget til rumopvarmning, varmtvandsforbruget eller elforbruget.

Boligareal

Attributten boligareal har, igennem litteraturstudiet, vist sig at være betydningsfuld for energiforbruget til rumopvarmning samt for elforbruget. Boligarealet kan oplyses enten som brutto- eller nettoareal, det vil sige arealet målt enten inklusiv eller eksklusiv ydervæggene. Arealet der anvendes til beregninger og simuleringer i henholdsvis Be10 samt BSim er bruttoarealet. Da oplysninger om bruttoarealer kan indsamles gennem BBR-registeret¹, anbefales det i nærværende projekt, at dette areal anvendes i udviklingen af klassificeringstræer.

Oplysninger om boligens varmetekniske egenskaber

Bygningers transmissionstab har stor betydning for energiforbruget til rumopvarmning i disse. Af den årsag bør der indsamles oplysninger om boligens varmetekniske egenskaber, når klassificeringstræet skal opbygges. Oplysninger om bygningers transmissionstab kan, i forbindelse med opbygning og anvendelse af klassificeringstræer, indsamles på forskellige måder. Én mulighed er en direkte anvendelse af bygningens transmissionstab som attribut, og en anden mulighed er anvendelse af oplysninger om de enkelte konstruktionsdele og deres termiske egenskaber som attributter. Begge muligheder vil medføre en række problemstillinger både ved indsamling af data til opbygning af klassificeringstræet, men også ved en efterfølgende anvendelse af klassificeringstræet.

Såfremt transmissionstabet direkte anvendes som attribut, skal der principielt foretages en varmetabsberegning, inden klassificeringstræet kan opbygges eller anvendes. Det betyder, at datagrundlaget skal indeholde oplysninger om eksempelvis U-værdier for ydervægge, vinduer, loftkonstruktion m.fl. samt arealer og orientering af disse. Disse informationer skal efterfølgende bearbejdes, inden data kan anvendes i klassificeringstræet. Resultatet bliver en relativ omstændelig proces, der sætter begrænsninger for, hvem der kan anvende klassificeringstræet efterfølgende. På trods af at det giver en omstændelig proces, har anvendelse af transmissionstabet dog den fordel, at der anvendes én parameter, som tager højde for transmissionstabet gennem hele bygningens klimaskærm.

En anden mulighed er en direkte anvendelse af oplysningerne om eksempelvis U-værdier for ydervægge, vinduer, loftkonstruktion m.fl. samt arealer af disse i klassificeringstræet. Når klassificeringstræet opbygges vil de væsentligste parametre automatisk blive inddraget og mindre væsentlige parametre

¹BBR-registeret er et landsdækkende bygnings- og boligregister med data om samtlige landets bygninger.

undlades. Resultatet heraf kan blive et større klassificeringstræ med en relativ høj detaljeringsgrad på de enkelte attributter.

I begge tilfælde skal der indsamles oplysninger om boligens varmetekniske egenskaber, hvor det er et spørgsmål om, hvor meget disse oplysninger efterfølgende skal bearbejdes. Oplysningerne kan indsamles fra en bygningsbeskrivelse, tegningsmateriale, energimærkninger eller ved tilsyn, såfremt der ønskes en relativ nøjagtig fastlæggelse af disse. Hvad angår oplysninger om U-værdier fra energimærkninger, kan disse søges i BBR-registeret, såfremt der foreligger en sådan mærkning. Ved det nærværende projekts afslutning kan det bemærkes, at der er foretaget omkring 196.000 energimærkninger af parcelhuse i Danmark siden 2006 [20]. Det betyder, at de ønskede oplysninger ikke i alle tilfælde kan søges i energimærkninger. Alternativt kan der søges oplysninger blandt beboerne, men da det højest sandsynligt er et fåtal af disse, der kan oplyse om boligens varmetekniske egenskaber, må der eventuelt suppleres med standardværdier. Med dette menes, at hvis der spørges til konstruktionernes U-værdi eller opbygning, må beboerne sammenholde deres viden med eksempelvis tabelværdier for typiske konstruktioner. Derfra kan beboerne bedømme hvilken konstruktion eller hvilke værdier, som de mener, passer bedst på deres bolig. Ved anvendelse af denne fremgangsmåde må det forventes, at oplysningernes nøjagtighed reduceres.

Boligens opførelsesår

Dette forhold har vist sig at forklare noget af forbrugsvariationen inden for energiforbruget til rumopvarmning. I denne sammenhæng knytter opførelsesåret sig til bygningens varmetekniske egenskaber, da der gennem tiderne har været forskellige byggetekniske traditioner og standarder. Oplysninger om bygningens opførelsesår kan søges i BBR-registeret, hvor der ligeledes fremgår oplysninger om årstallet for eventuelle større renoveringer. Renoveringer der ikke kræver en forudgående byggesagsbehandling eller anmeldelse fremgår dog ikke til BBR-registeret [52]. Det betyder, at forbedringer, som eksempelvis efterisolering, ikke nødvendigvis fremgår af BBR-registeret, hvorfor der må søges oplysninger om dette blandt beboerne eller ved tilsyn.

Infiltration

Bygningens tæthed har betydning for energiforbruget til rumopvarmning, og er erfaret, at have en forklarende effekt i forbindelse med de forbrugsvariationer, der kan observeres mellem husstande. Infiltrationen i bygningen afhænger af dennes tæthed, og såfremt der ønskes målinger for infiltrationen, skal der udføres en blowerdoortest. Det har dog den ulempe, at såfremt infiltrationen anvendes, skal der foretages målinger af denne, inden klassificeringstræet for estimering af energiforbruget til rumopvarmning kan anvendes. Det betyder, at såfremt informationer om infiltration medtages i et klassificeringstræ, kræver det, at der foreligger oplysning om dette, før klassificeringstræet kan anvendes. En anden mulighed er at undlade at inddrage infiltrationen i klassificeringstræet, men derved udgår et relativt væsentligt forhold i forbindelse med energiforbruget til rumopvarmning i boliger.

Styring af mekanisk ventilation

Denne attribut knytter sig dels til de tekniske forhold og dels til beboernes adfærd, og kan have indflydelse på både energiforbruget til rumopvarmning samt elforbruget. Betydningen af denne attribut afhænger af, om der er mekanisk ventilation i boligen, og i så fald hvilken type ventilation der tale om. Hvis der er et installeret et ventilationsanlæg med konstant udsugning og/eller indblæsning, har styringen af anlægget kun betydning, hvis beboerne føler sig så generet af det, at de slukker anlægget. Er der derimod tale om behovstyret ventilation, har beboernes adfærd i forbindelse med regulering af anlægget større betydning. Der bør søges nærmere oplysninger om styringen af ventilationsanlægget blandt beboerne.

O.2 Demografiske attributter

De demografiske attributter knytter sig til informationer om boligens beboere. Af disse gennemgås forhold, der gennem litteratur- og casestudiet har vist indikationer på, at de kan forklare dele af

forbrugsvariationerne for enten energiforbruget til rumopvarmning, varmtvandsforbruget eller elforbruget.

Antal beboere

Antallet af beboere har vist sig at have betydning for både varmtvandsforbrugets og elforbrugets størrelse, hvor det kan forklare noget af forbrugsvariationen husstande imellem. Informationer om antallet af beboere er et af de forhold, der let kan oplyses af beboerne selv. Disse oplysninger kan eventuelt kombineres med beboernes opholdstid i boligen, da der ellers kan opstå en problemstilling om eksempelvis delebørn.

Beboernes alder

Beboernes alder har vist sig at have betydning for energiforbruget til rumopvarmning, varmtvandsforbruget og elforbruget i boligen. Det er erfaret, at det specielt er tilstedeværelsen af børn og teenagere, der har betydning for forbruget i boligen. Hvad angår oplysninger om beboernes alder, kan det oplyses af beboerne selv.

Køn

Beboernes køn har vist sig at have betydning for elforbruget i boliger med én beboer. Der er indikationer på, at enlige kvinder har et større elforbrug end enlige mænd. Informationer om beboernes køn kan oplyses af beboerne selv.

Indtægt

Der bør søges oplysninger om husstandens indtægt, da dette forhold har vist sig at forklare dele af forbrugsvariationen for boligens energiforbrug til rumopvarmning og elforbruget i disse. Informationer om dette forhold kan søges blandt beboerne.

O.3 Adfærdsrelaterede attributter

De adfærdsrelaterede attributter knytter sig til viden om beboernes adfærd i boligen. Af disse gennemgås forhold, der gennem litteratur- og casestudiet har vist indikationer på, at de kan forklare dele af forbrugsvariationerne for enten energiforbruget til rumopvarmning, varmtvandsforbruget eller elforbruget.

Opholdstid i boligen

Beboernes opholdstid i boligen kan have betydning for samtlige forbrug. Dette forhold knytter sig til beboernes daglige vaner og rutiner, da disse i nogle tilfælde kan være styrende for, hvor stor betydning beboernes opholdstid i boligen har. Der skelnes ofte mellem hverdage og weekend/fridage, da der almindeligvis er et relativt fast mønster, som beboerne følger på hverdage. Oplysninger om hvornår der opholder sig personer i boligen, kan søges gennem et spørgeskema, eller der kan foretages målinger af eksempelvis CO₂-koncentrationen i boligen, for derigennem at registrere tilstedeværelse af personer.

Holdninger til miljø, energi og økonomi

Det er fra litteratur- og casestudiet erfaret, at beboernes engagement inden for miljø- og energimæssige forhold samt deres forhold til økonomiske aspekter kan have betydning til for de forskellige energiforbrug i boligen. Undersøgelser har vist, at beboernes holdninger til de fornævnte forhold kan afspejle sig i deres forbrug, hvorfor der bør søges informationer om denne attribut blandt beboerne. Det bør i denne sammenhæng overvejes, hvordan spørgsmålene udføres - om beboernes selv skal fortolke betydningen af miljø- og energimæssige forhold, eller der defineres nogle klare forskelle mellem disse. Dette er nærmere behandlet i afsnit 6 i hovedrapporten, hvor der fremgår en uddybning af hvorledes disse forhold kan undersøges.

Viden om størrelsen af forbrug

Det har gennem litteraturstudiet været indikationer på, at beboernes viden om størrelsen af deres elforbrug har en forklarende effekt i forhold til, hvor meget el de rent faktisk anvender. Det kan være

interessant at undersøge, hvorvidt beboernes viden om hvor stort et forbrug de har til rumopvarmning og varmt brugsvand ligeledes har en sammenhæng med deres faktiske forbrug. Undersøgelsen kan udføres således, at beboernes vurdering af hvor højt et forbrug de har, kan sammenholdes med forbrugsopgørelser fra de pågældende forsyningsvirksomheder. Ofte kan det dog ikke lade sig gøre at vurdere varmtvandsforbruget direkte, da opvarmningen af varmt brugsvand sker internt i boligen. I stedet for en vurdering af varmtvandsforbrugets størrelse kan det overvejes at spørge ind til beboernes viden om størrelsen af husstandens samlede vandforbrug.

Ønsket lufttemperatur

Beboernes ønskede lufttemperatur har i flere tilfælde gennem litteraturstudiet vist at have indflydelse på energiforbruget til rumopvarmning, hvormed den er en medvirkende årsag til forbrugsforskellene mellem boliger. Informationer om beboernes ønskede lufttemperatur kan enten opnås gennem udsagn fra beboerne selv, eller der kan foretages temperaturmålinger såfremt eksakte værdier ønskes. Det bør bemærkes, at hvis informationer om lufttemperaturen fås af beboerne, kan der være forskelle mellem de oplyste temperaturer og de faktiske temperaturer i boligerne, hvilket er observeret gennem litteraturstudiet.

Justering af termostater

Et væsentligt forhold i forbindelse med boligers energiforbrug til rumopvarmning er beboernes vaner i forbindelse med justering af varmeanlæggets termostater. Af denne årsag bør der søges informationer, der kan klarlægge beboernes daglige vaner for justering af termostater. Det drejer sig blandt andet om, hvorvidt beboerne skruer ned for varmen, når huset forlades for en længere periode, skruer ned for varmen i ubenyttede rum eller ved udluftning og lignende.

Lukke af til ubenyttede rum

Denne attribut knytter sig til beboernes vaner i forbindelse med justering af termostater, eftersom en eventuel nedjustering af setpunktstemperaturen i et rum har størst effekt, hvis der lukkes af ind til det pågældende rum. Informationer om beboernes vaner i denne forbindelse kan søges gennem en spørgeundersøgelse.

Vaner i forbindelse med udluftning

Beboernes vaner i forbindelse med udluftning har betydning for energiforbruget til rumopvarmning. Der er imidlertid flere forhold, der har betydning for hvor stor indflydelse beboernes vaner i forbindelse med udluftning har på energiforbruget. Derfor bør der søges oplysninger om, hvor ofte, i hvor lang tid af gangen og hvordan beboerne typisk lugter ud i hjemmet. Samtidig knytter disse forhold sig til beboernes vaner for justering af termostater til varmeanlægget, da det har en betydning for energiforbruget, hvorvidt beboerne slukker for varmens mens der luftes ud.

Adfærd i hjemmet hvis det opfattes koldt/varmt

Denne attribut kan være med til at beskrive forbrugsforskelle i forbindelse med energiforbruget til rumopvarmning. Det er forskelligt, om beboerne har for vane at tage mere tøj på før der skrues op for varmen, hvis de føler det for koldt. Der er set eksempler på, at det gennem en spørgeskemaundersøgelse er forsøgt at klarlægge beboernes adfærd i den forbindelse. Det drejer sig eksempelvis om, hvorvidt beboerne anvender mere tøj, et tæppe eller lignende, om de tager et bad og så videre. På tilsvarende vis kan det være forskelligt, om beboerne åbner vinduer og yderdøre, hvis de føler det for varmt, eller om de tager noget tøj af, drikker noget koldt eller lignende. Det vurderes i nærværende projekt, at der bør søges informationer blandt beboerne, om deres adfærd i de førnævnte tilfælde, da denne adfærd kan være med til at forklare forskelle i energiforbruget til rumopvarmning.

Vaner i forbindelse med hygiejne

Beboernes vaner i forbindelse med hygiejne har indflydelse på varmtvandsforbruget i boligerne. I denne forbindelse er der tale om en undersøgelse af beboernes vaner i forbindelse med antallet af brusebade og karbade. Da antallet af bade i sig selv ikke er et udtryk for størrelsen af varmtvandsforbruget, bør der også søges informationer om den typiske varighed af badene i husstanden. Det vurderes, at disse forhold tilsammen vil forklare store dele af forbrugsforskellene, der kan observeres mellem boligers varmtvandsforbrug.

Vaner i forbindelse med opvask

Foruden ovenstående attribut har beboernes vaner i forbindelse med eksempelvis opvask og afskylning af service tilsvarende betydning for varmtvandsforbrugets størrelse. Der er forskel på om beboerne eksempelvis anvender en opvaskebalje eller om det varme vand løber mens processen er i gang. Af disse årsager kan der søges informationer blandt beboere, om deres vaner i forbindelse med opvask og afskylning, når forbrugsforskelle for varmtvandsforbruget skal undersøges.

Antal og brugstid af apparater

Mængden af elektriske apparater i husstanden, så som tv, pc, køkkengrej, vaskemaskine m.fl., er erfaret, sammen med brugstiden af førnævnte, at have direkte betydning for elforbruget i boligen. Der bør søges informationer om både antal og den typiske brugstid af apparaterne, da en stor bestand i sig selv ikke er ensbetydende med et højt elforbrug.

Alder på apparater

Alderen på de elektriske apparater i boligen har en betydning for elforbruget i de pågældende husstande. Betegnelsen elektriske apparater dækker i denne sammenhæng over eksempelvis tv, pc, radio, hårde hvidevarer m.fl. Der kan dog være forskellige årsager til hvorfor et apparat udskiftes, ligesom udskiftningsfrekvensen for apparaterne kan variere. Det kan være interessant at søge informationer blandt beboerne om hvilke forhold, der har størst betydning, når de anskaffer sig et nyt apparat; om det er økonomi, design, funktionalitet eller energieffektivitet, der vægter højest, når deres valg træffes.

Standbyforbrug

Mængden af elektriske apparater med et standbyforbrug har vist sig at have en betydelig indflydelse på elforbruget. Det er forskel på, om beboerne slukker helt for apparaterne, eller lader dem stå på standby. Derved afhænger dette forhold af beboernes vaner, hvorfor informationer bør søges hos beboerne selv.

Slukke lys i ubenyttede rum

Beboernes adfærd i forbindelse med at slukke lys i rum når de ikke benyttes, har direkte indflydelse på elforbruget. Attributten har vist sig at forklare dele af forbrugsvariationen mellem husstande, hvor den dog knytter sig til flere andre adfærdsrelaterede forhold. Deriblandt er beboernes holdninger til hygge, energi, miljø og økonomi, men det har også en betydning, om beboerne føler, at den elektriske belysning har indvirkning på de førnævnte forhold.

Anvendelse af elbesparende foranstaltninger

Dette forhold er tiltænkt at undersøge om der kan findes en sammenhæng mellem beboernes anvendelse af elbesparende foranstaltninger og elforbrugets størrelse i husstanden. De elbesparende foranstaltninger kunne være anvendelse af elspareskinner, sparepærer, energimærkede apparater, IHC-anlæg og belysningsensorer til udendørs belysning. Der er mulighed for at skaffe viden om beboernes anvendelse af disse igennem et spørgeskema.

O.4 Attributter der ikke er medtaget

I dette afsnit beskrives nogle forhold, der ikke er medtaget som attributter i klassificeringstræerne, men som kan være relevante og forklare dele af forbrugsforskellene for de overordnede energiforbrug i boliger. Der er forskellige årsager til hvorfor de efterfølgende forhold ikke er medtaget som attributter. Forhold som anvendelse af vandbesparende armaturer og vandbesparende opvaske- og vaskemaskine kan forklare forbrugsforskelle for det samlede vandforbrug, men ikke nødvendigvis for forbruget af varmt brugsvand. Det betyder, hvis det i fremtidige projekter ønskes at opbygge et klassificeringstræ, der kan anvendes til estimering af husstandes samlede vandforbrug, vurderes disse forhold med fordel af kunne inddrages.

Det fremgår af de foregående afsnit om attributter til klassificeringstræerne, at flere af forholdene kan vurderes både på baggrund af udsagn fra beboere, eller der kan foretages målinger i boliger-

ne. 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 besvarelsene subjektivt præget, og der kan være afvigelser mellem besvarelsene 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 besvarelser 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.

Validering af brugen af programmet WEKA

Når et klassificeringstræ skal udvikles kan det ske vha. enten håndberegninger eller programmer. For enkelthedens skyld når datamængderne, der skal evalueres, er store, er det en fordel at anvende et beregningsprogram både for at simplificere processen men også for at undgå fejlberegninger og derigennem fejl i det udviklede klassificeringstræ.

I nærværende projekt er det ved udvikling af klassificeringstræet for elforbruget valgt at anvende programmet Weka. Weka står for *Waikato Environment for Knowledge Analysis* og er et populært program, skrevet i Java, der anvendes inden for machine learning. Programmet indeholder en samling af machine learning algoritmer, der kan håndtere flere forskellige data mining opgaver, mere specifikt drejer det sig bl.a. om: Præprocessering af data, klyngeanalyse, regression og klassificering. Programmet er udviklet af Waikato universitet i New Zealand tilbage i 1993 og er løbende blevet videre udviklet. [91, 3]

P.1 Introduktion til eksemplet fra Ross Quinlan

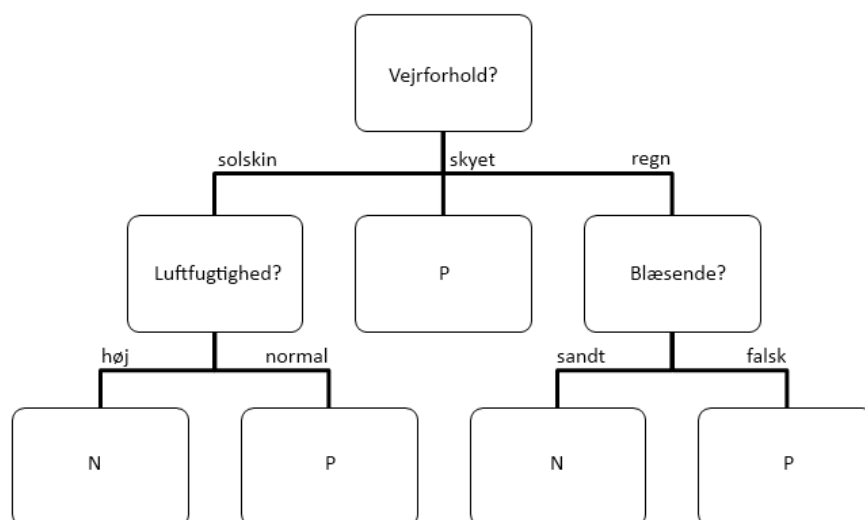
For at sikre at projektgruppen anvender det udvalgte program korrekt er det på forhånd valgt at gennemføre et simpelt eksempel. Det udvalgte eksempel er fra artiklen *Induction of Decision Trees* af Ross Quinlan [59].

Eksemplet der gennemgås i artiklen drejer sig om at vurdere vejret lørdag morgen i forhold til et kunne udføre en specifik aktivitet udendørs. Klassen i eksemplet, dvs. målet for inddelingen af data, er givet som enten **N** eller **P**, der er lig hhv. en negativ indstilling, hvor aktiviteten ikke kan udføres og en positiv indstilling, hvor aktiviteten udføres. Attributterne der er registreret er: udsigten (solskin, overskyet, regn), temperaturen udendørs (varmt, mildt, koldt), luftfugtighed udendørs (høj, normal) og blæsende (sandt, falsk). Det indsamlede data, for 14 forskellige lørdage er præsenteret i tabel P.1.

Resultatet fra artiklen giver et klassificeringstræ som er illustreret i figur P.1.

Lørdag nr.	Attributter				Klasse
	Vejrforhold	Temperatur	Luftfugtighed	Blæsende	
1	solskin	varmt	høj	falsk	N
2	solskin	varmt	høj	sandt	N
3	skyet	varmt	høj	falsk	P
4	regn	mildt	høj	falsk	P
5	regn	koldt	normal	falsk	P
6	regn	koldt	normal	sandt	N
7	skyet	koldt	normal	sandt	P
8	solskin	mildt	høj	falsk	N
9	solskin	koldt	normal	falsk	P
10	regn	mildt	normal	falsk	P
11	solskin	mildt	normal	sandt	P
12	skyet	mildt	høj	sandt	P
13	skyet	varmt	normal	falsk	P
14	regn	mildt	høj	sandt	N

Tabel P.1: Træningsdata fra eksemplet givet i artiklen „Induction of Decision Trees“ [59].



Figur P.1: Resultatet for eksemplet fra artiklen „Induction of Decision Trees“ [59].

P.2 Anvendelse af eksemplet fra Ross Quinlan

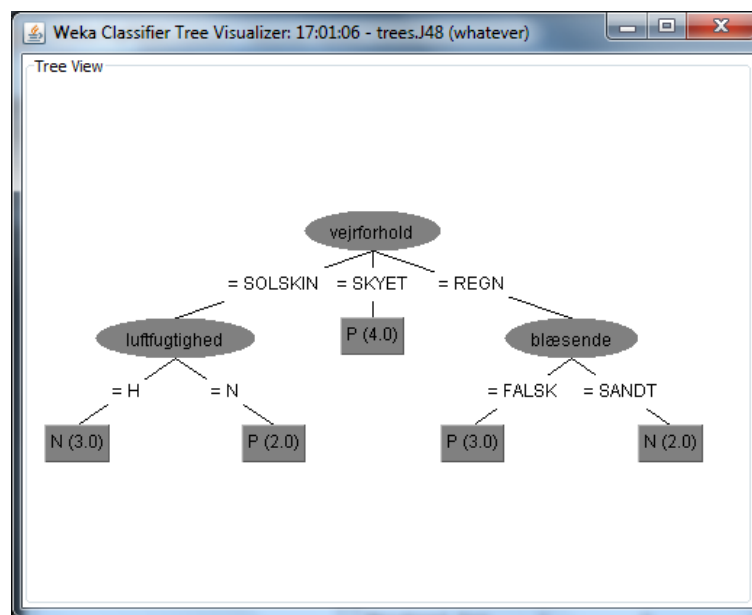
Det givne data fra tabel P.1 skrives ind, kommasepareret, i et excel ark, se filen *RQeksempel.xlsx* der kan findes på DVDen i mappen med navnet software. Dette ark konverteres til en .arff fil vha. af hjemmesiden med følgende link: <http://slavnik.fe.uni-lj.si/markot/csv2arff/csv2arff.php>. I det følgende afsnit gennemgås de nødvendige trin før konverteringen af datafilen er udført.

1. Load hjemmesiden: <http://slavnik.fe.uni-lj.si/markot/csv2arff/csv2arff.php>.
2. Start med at load den kommaseparerede excel fil og tryk „send forespørgsel“.
3. Der vises en illustration af hvad den loadede excel-fil indeholder af data. Næste skridt er, at angive hvilke attributter der skal anvendes til .arff filen, dette gøres ved at sætte et flueben i kassen ved de attributter der skal anvendes. Skal alle attributter anvendes i .arff filen kan boksen „check all“ vælges. Herefter skal det, for de udvalgte attributter, angives om de er numeriske eller kategoriske. Når ovenstående er udført for alle attributter trykkes der på knappen „send forespørgsel“.
4. Når der er trykket „send forespørgsel“ kommer der et separat vindue frem der hedder „Åbner RQeksempel.arff“. Vælg gem fil i dette vindue. Nu er excel filen konverteret til en .arff og er klar til brug i programmet Weka.

I det efterfølgende afsnit gennemgås den overordnede brug af programmet Weka. Programmet startes enten ved at dobbeltklikke på den producerede .arff fil eller vha. genvejen til programmet. Hvis genvejen anvendes skal der i det første vindue der åbnes vælges *explorer*, når programmets brugerflade åbnes trykkes der på knappen *open file* der findes øverst til venstre.

Når programmet er åbnet og .arff filen er loaded vælges fanebladet *Classify*. I området *Classifier* trykkes der på knappen *choose* hvorefter der kommer en dropdown liste frem, i denne vælges gruppen *trees* og herefter vælges *J48*, hvilket er navnet for C4.5 algoritmen i dette program. I området *Test options* vælges *Use training set*. Herefter vælges der, på drop down listen, den attribut der skal anvendes som klasse for klassificeringstræet. Dernæst trykkes der på knappen *Start* og programmet producerer klassificeringstræet, der vises som en fil, der kan ses i området *Result list*.

Den producerede fil har et navn der består af et klokkeslet for hvornår filen er genereret efterfulgt af følgende tråd: - trees.J48. Eksempelvis: 17:01:06 - trees.J48. I området *Classifier output* ses informationer om inputdata og om klassificeringstræets nøjagtighed præsenteret enten vha. *Evaluation on training set* hvor det fremgår hvor mange elementer der er klassificeret rigtigt og forkert. En anden mulighed for præsentation af klassificeringstræets nøjagtighed er gennem *Confusion Matrix*, der ved hjælp af en matrix viser hvor mange elementer der er klassificeret korrekt/forkert inden for hver mulig inddeling af klassen, i dette tilfælde N eller P. Ved at højreklikke på den fremkomne fil i *resultat listen* fremkommer der en liste af valgmuligheder, ved at vælge *Visualize tree* visualiseres det genererede klassificeringstræ. Resultatet fra Weka når data fra tabel P.1 anvendes kan ses i figur P.2.



Figur P.2: Screenprint af det genererede klassificeringstræ på basis af data fra tabel P.1.

P.2.1 Sammenfatning

Ved sammenligning af figur P.1 og P.2 ses det, at det genererede klassificeringstræ er lig det klassificeringstræ *Ross Quinlan* fandt frem til. Hermed kan det konkluderes, at projektgruppens anvendelse af programmet Weka er godkendt.

Resultatfil fra Weka

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    whatever
Instances:   100
Attributes:  15
              Boligareal      Optaget_el
              Indtaegt         Underholdningsapparater
              Alder_v1         Brugstid_UA
              Alder_v2         Toejvaske
              Alder_b1         Toerringer
              Alder_b2         Opvask
              BES              Standbyforbrug
                              Klasse

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      85           85      %
Incorrectly Classified Instances    15           15      %
Kappa statistic                    0.773
Mean absolute error                 0.1139
Root mean squared error             0.2386
Relative absolute error             33.8011 %
Root relative squared error         58.2658 %
Total Number of Instances          100

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
              0.886    0.077    0.861      0.886    0.873      0.958    Lavt
              0.881    0.121    0.841      0.881    0.86       0.936    Gennemsnitligt
              0.571     0       1          0.571    0.727      0.986    Meget_hoejt
              0.813    0.036    0.813      0.813    0.813      0.973    Hoejt
Weighted Avg.   0.85     0.083    0.855      0.85     0.848      0.953

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
31  4  0  0 | a = Lavt
 5 37  0  0 | b = Gennemsnitligt
 0  0  4  3 | c = Meget_hoejt
 0  3  0 13 | d = Hoejt

```

Figur Q.1: Resultatfil fra Weka.

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

Optaget_el = Ja_for_det_meste
|   Toerringer = 0-3
|   |   Boligareal = <135: Lavt (14.0/2.0)
|   |   Boligareal = 135-160
|   |   |   Opvask = 5-7: Gennemsnitligt (6.0/2.0)
|   |   |   Opvask = 0-4: Lavt (4.0)
|   |   |   Opvask = >7: Lavt (0.0)
|   |   Boligareal = >160: Hoejt (2.0)
|   Toerringer = 4-5
|   |   BES = Hoejt: Gennemsnitligt (3.0/1.0)
|   |   BES = Meget_hoejt: Hoejt (3.0)
|   |   BES = Lavt: Gennemsnitligt (2.0)
|   |   BES = Gennemsnitligt: Gennemsnitligt (1.0)
|   Toerringer = >5: Gennemsnitligt (2.0)
Optaget_el = Sommetider
|   Toejvaske = 0-3: Gennemsnitligt (5.0)
|   Toejvaske = 4-5: Gennemsnitligt (15.0/2.0)
|   Toejvaske = >5
|   |   Brugstid_UA = 0-5: Meget_hoejt (0.0)
|   |   Brugstid_UA = >10: Meget_hoejt (2.0)
|   |   Brugstid_UA = 5-10: Hoejt (2.0)
Optaget_el = Nej_slet_ikke
|   Opvask = 5-7: Hoejt (9.0/3.0)
|   Opvask = 0-4: Hoejt (0.0)
|   Opvask = >7: Meget_hoejt (2.0)
Optaget_el = Ja_altid
|   Toerringer = 0-3: Lavt (16.0/3.0)
|   Toerringer = 4-5
|   |   Alder_b2 = Barn: Gennemsnitligt (8.0/2.0)
|   |   Alder_b2 = Teenager: Lavt (2.0)
|   Toerringer = >5: Gennemsnitligt (2.0)

Number of Leaves   :    22

Size of the tree   :    32

```

Figur Q.2: Resultatfil fra Weka.

Begrænsning af klassificeringstræets størrelse

I afsnit 7.3 i hovedrapporten fremgår der en beskrivelse af fremgangsmåden for begrænsning af klassificeringstræets størrelse. Metoden tager udgangspunkt i formel 7.15 i hovedrapporten, hvor denne er gengivet i formel R.1 for overskuelighedens skyld. Formel R.1 skal, ifølge C4.5 algoritmen, anvendes til beregning af fejlraten, der er grundlaget for vurderingen af hvor meget klassificeringstræets størrelse skal begrænses. Da formelen kun omhandles overfladisk i bogen om C4.5 algoritmen, skrevet af udvikleren *John Ross Quinlan* [60], er denne undersøgt nærmere. Denne undersøgelse er foretaget med henblik på at vurdere, hvorvidt antagelserne bag formelen er overholdt ved den aktuelle anvendelse.

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

Hvor:

CF	Faktor, der som standard er 0,25 i C4.5 algoritmen [60].
E	Antal fejlplacerede elementer i knudepunktet.
N	Samlet antal elementer i knudepunktet.
P	Fejlrate.

Formel R.1 angiver hvorledes CF -faktoren kan beregnes på baggrund af N , E og P , eller omvendt hvorledes fejlraten kan beregnes ud fra et kendskab til de øvrige parametre. Formlen til beregning af CF -faktoren er identisk med *Ayyub et al.*'s [4] udlægningen af metoden til beregning af den kumulerede sandsynlighed i en binomialfordeling. En binomialfordeling kan anvendes til beregning af sandsynligheder for diskrete variable, og omtaltes af denne årsag ofte som en diskret sandsynlighedsfordeling [4]. Det vil sige, at de N elementer i formel R.1 antages at kunne beskrives ved en binomialfordeling. For at en variabel kan repræsenteres ved en binomialfordeling skal denne overholde følgende antagelser:

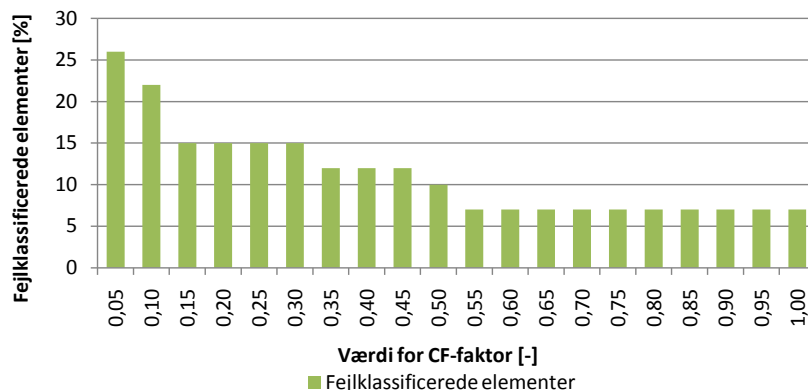
1. Binomialfordelingen er baseret på N Bernoulli-forsøg, hvor disse kun har to mulige udfald [4].
2. De N Bernoulli-forsøg der udføres skal være uafhængige [4].
3. Sandsynligheden for de to mulige udfald forbliver konstant henholdsvis P og $(1-P)$ i samtlige udførte Bernoulli-forsøg [4].

Et Bernoulli-forsøg er kendetegnet ved, at det kun kan have to mulige udfald. Hvis sandsynligheden for det ene udfald betegnes P , vil sandsynligheden for det andet udfald være $(1 - P)$ [4]. Det vil sige, at hvis en variabel skal overholde det første kriterium, skal det kun have to mulige udfald. Hvis dette relateres til hvorledes klassificeringstræet anvendes, vil det første kriterium kun være overholdt, såfremt klassen i klassificeringstræet har to inddelinger. I nærværende projekt er klassen for elforbruget inddelt i fire; *lavt*, *gennemsnitligt*, *højt* og *meget højt*. Det betyder, at det ikke er matematisk korrekt at beskrive elementerne ved en binomialfordeling.

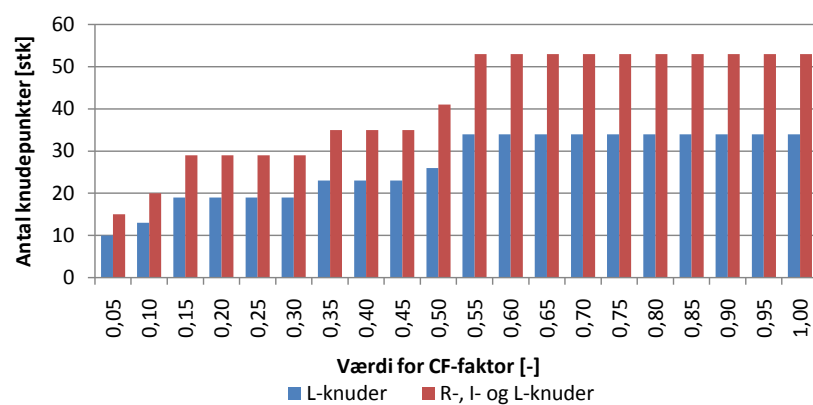
Et andet kriterium der skal være overholdt, for at en variabel kan beskrives ved en binomialfordeling, er, som tidligere nævnt, at Bernoulli-forsøgene skal være uafhængige. Relateret til projektet betyder det, at de N elementer, i det aktuelle knudepunkt, skal være uafhængige. Dette kriterium vil ikke være overholdt i klassificeringstræet, da elementerne på forhånd er inddelt efter en række forskellige attributter og er derved ikke uafhængige.

Det sidste kriterium, omhandlende sandsynligheden for udfaldene, overholdes heller ikke. Det skyldes, at der, som tidligere nævnt, er mere end to klasseinddelinger i det aktuelle tilfælde.

Af ovenstående fremgår det, at antagelserne for en binomialfordeling ikke overholdes i det aktuelle tilfælde. Det betyder, at det er matematisk forkert at anvende antagelsen om en binomialfordeling for elementerne i klassificeringstræet, men ikke desto mindre anvendes metoden stadig i C4.5 algoritmen. Det er undersøgt hvor stor betydning CF -faktoren har på størrelsen af klassificeringstræet for elforbruget. Af figur R.1 fremgår det, at andelen af fejlklassificerede elementer i klassificeringstræet er faldende som funktion af en stigende CF -værdi. Modsat fremgår de af figur R.2, at antallet af knudepunkter i klassificeringstræet er stigende, som funktion af en stigende CF -værdi. Sammenhængen er dog forventelig, da en høj CF -værdi er et udtryk for, at der tillades en lav fejlrate. Den lave fejlrate betyder, at klassificeringstræet skal have en vis størrelse, for at inddele elementerne korrekt.



Figur R.1: Illustrering af CF -faktorens betydning for andelen af fejlklassificerede elementer i klassificeringstræet for elforbruget.



Figur R.2: Illustrering af CF -faktorens betydning for antallet af L-knuder og det samlede antal knudepunkter i klassificeringstræet for elforbruget.

Håndtering af numeriske attributter i klassificeringstræet

I klassificeringstræet, der er udviklet og beskrevet i kapitel 7 i hovedrapporten, anvendes der kun kategoriske attributter. 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 frem for kategoriske. C4.5 algoritmen kan håndtere både numeriske og kategoriske attributter, men algoritmen har en bestemt metode til håndtering af numeriske attributter. Derfor vil den efterfølgende beskrivelse redegøre for, hvorledes en inddeling for en numerisk attribut udføres i henhold til C4.5 algoritmen.

Flere af værdierne for attributterne i datamængden, der er genereret på baggrund af beskrivelsen i kapitel 6 i hovedrapporten, er som udgangspunkt genereret med numeriske værdier. Disse er, som tidligere omtalt, efterfølgende konverteret til kategoriske værdier. I nærværende kapitel tages der udgangspunkt i det data, der er beskrevet i kapitel 7 i hovedrapporten, men hvor attributten omhandlende den ugentlige anvendelse af opvaskemaskine er numerisk¹.

C4.5 algoritmen anvender en binær splittest på data ved numeriske attributter. Splittesten udføres således, at der fastlægges en grænseværdi, som dataet inddeles efter. Inddelingen sker på baggrund af følgende test: $v_i \leq G$ eller $v_i > G$. Hvor v_i er attributtens numeriske værdi og G er en fastsat grænseværdi. Formålet er at vurdere hvilken grænseværdi, der giver den bedste klasseinddeling af elementerne i den aktuelle attribut. Grænseværdien bestemmes på baggrund af hvilke værdier der optræder i den numeriske attribut. For at anskueliggøre dette anvendes eksemplet, der fremgår af figur S.1 på side 157. Det bør dog bemærkes, at den førnævnte figur kun viser et udsnit af klassificeringstræet. Ved I-knuden omhandlende den ugentlige anvendelse af opvaskemaskine er der ti elementer. Besvarelsen fra de ti elementer fremgår af tabel S.1. På baggrund af besvarelsene til attributten skal følgende grænseværdier undersøges: 3, 4 og 5.

¹Datasættet med den numeriske attribut fremgår af filen *Klassificeringstræ \ Udvikling af klassificeringstræ, numerisk attribut* på den medfølgende DVD.

	Element									
	1	2	3	4	5	6	7	8	9	10
Besvarelse til attribut:	3	4	4	4	5	5	5	5	6	6
Klasseinddeling*:	L	L	L	L	G	L	G	L	G	G

L: Tilhører klasseinddelingen *Lavt*.

G: Tilhører klasseinddelingen *Gennemsnitligt*.

Tabel S.1: Ugentlig anvendelse af opvaskemaskine for ti fiktive elementer.

På baggrund af data fra tabel S.1 kan der beregnes en entropi før et split ved brug af formel 7.1 på side 138 i hovedrapporten. Beregning af denne entropi fremgår af formel S.1².

$$\begin{aligned}
 E(S) &= - \sum_{j=1}^k \left(\frac{\text{freq}(C_j, S)}{|S|} \cdot \log_2 \left(\frac{\text{freq}(C_j, S)}{|S|} \right) \right) \Rightarrow \\
 E(S) &= - \left(\frac{6}{10} \cdot \log_2 \left(\frac{6}{10} \right) + \frac{4}{10} \cdot \log_2 \left(\frac{4}{10} \right) \right) \Rightarrow \\
 E(S) &= 0,971 \text{ bits}
 \end{aligned} \tag{S.1}$$

Der skal beregnes en entropi efter en inddeling ved hver af de tre enkelte grænseværdier, hvilket udføres på baggrund af formel 7.3 i hovedrapporten. For en grænseværdi på tre vil følgende resultater fremkomme:

$$\begin{aligned}
 E_X(S) &= \sum_{i=1}^n \left(\frac{|S_i|}{|S|} \cdot E(S_i) \right) \Rightarrow \\
 E(S_{\leq 3}) &= -\frac{1}{1} \cdot \log_2 \left(\frac{1}{1} \right) = 0 \text{ bits} \Rightarrow \\
 E_{\leq 3}(S) &= \frac{1}{10} \cdot 0 = 0 \text{ bits} \\
 E(S_{> 3}) &= - \left(\frac{5}{9} \cdot \log_2 \left(\frac{5}{9} \right) + \frac{4}{9} \cdot \log_2 \left(\frac{4}{9} \right) \right) = 0,991 \text{ bits} \Rightarrow \\
 E_{> 3}(S) &= \frac{9}{10} \cdot 0,991 = 0,892 \text{ bits} \\
 E_X(S) &= E_{\leq 3}(S) + E_{> 3}(S) = 0,892 \text{ bits}
 \end{aligned} \tag{S.2}$$

Tilsvarende de kategoriske attributter kan der beregnes *information gain* med udgangspunkt i formel 7.5 i hovedrapporten:

²Håndberegninger for den numeriske attribut fremgår af regnearket *Klassificeringstræ \ Udvikling af klassificeringstræ, numerisk attribut \ Udvikling af klassificeringstræ, numerisk attribut.xlsx* på den medfølgende DVD.

$$\begin{aligned} IG(X) &= E(S) - E_X(S) = 0,971 \text{ bits} - 0,892 \text{ bits} \Rightarrow \\ IG(X) &= 0,079 \text{ bits} \end{aligned} \quad (\text{S.3})$$

Split information beregnes på baggrund formel 7.6 i hovedrapporten:

$$\begin{aligned} SI(X) &= - \left(\frac{1}{10} \cdot \log_2 \left(\frac{1}{10} \right) + \frac{9}{10} \cdot \log_2 \left(\frac{9}{10} \right) \right) \Rightarrow \\ SI(X) &= 0,469 \text{ bits} \end{aligned} \quad (\text{S.4})$$

Gain ratio beregnes på baggrund formel 7.8 i hovedrapporten:

$$\begin{aligned} GR(X) &= \frac{IG(X)}{SI(X)} \Rightarrow \\ GR(X) &= \frac{0,079 \text{ bits}}{0,469 \text{ bits}} \Rightarrow \\ GR(X) &= 0,168 \end{aligned} \quad (\text{S.5})$$

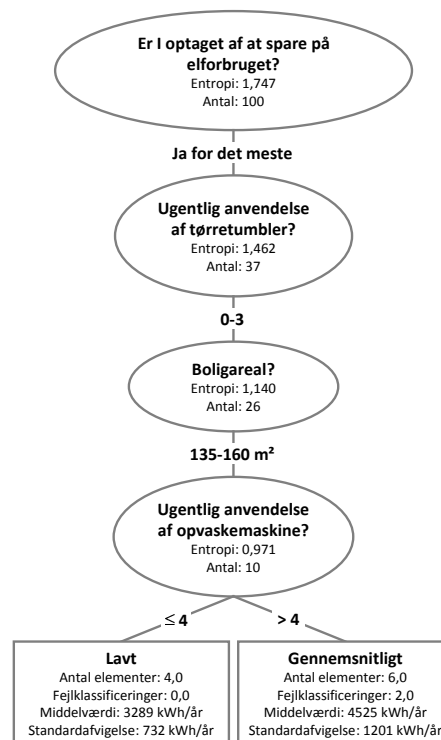
Ved tilsvarende beregninger for de øvrige grænseværdier fremkommer resultaterne, der er oplistet i tabel S.2.

Grænseværdi	Information gain	Split information	Gain ratio
[<i>gange/uge</i>]	[<i>bits</i>]	[<i>bits</i>]	[–]
3	0,079	0,469	0,168
4	0,420	0,971	0,433
5	0,322	0,722	0,446

Tabel S.2: Resultater fra beregning af *information gain*, *split information* og *gain ratio* for forskellige grænseværdier af den numeriske attribut.

Af tabel S.2 fremgår det, at det er inddelingen ved en grænseværdi på fem, der giver den højeste *gain ratio*. På baggrund af oplysningerne i bogen fra *John Ross Quinlan* [60] skal inddelingen foretages efter denne grænseværdi. Af figur S.1 fremgår det, at data minings programmet Weka anvender en grænseværdi på fire til inddeling af data. Dette skyldes, at Weka anvender *information gain* som splitkriterium [40, 93]. Af tabel S.2 fremgår det af håndberegningerne, at det er ved en grænseværdi på fire, at *information gain* er højest, hvilket stemmer overens med inddelingen fra Weka.

På baggrund af ovenstående beskrivelse er der redegjort for, hvorledes numeriske attributter håndteres i henholdsvis C4.5 algoritmen og i data minings programmet Weka.



Figur S.1: Udsnit af klassificeringstræ udviklet på baggrund af fiktivt data med én numerisk attribut.

S.1 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 [7]. Målene for nøjagtigheden, der fokuseres på i nærværende afsnit, er at finde i under *Detailed accuracy by class* i resultatfilen. Blandt disse fremgår af tabel S.3. 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. [63]
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. [63]
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. [7, 90]
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. [7, 90]
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. [89]
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. [89]

Tabel S.3: Definition af nøglebegreber som anvendes til vurdering af et klassificeringstræs nøjagtighed i Weka.

S.1.1 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 S.4. 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 S.4, 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 S.4. 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 matricen 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. [63]

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 S.4: Confusion matrix for det udviklede klassificeringstræ der fremgår af *tegning A*.

Til beregning af de tidligere omtalte mål for nøjagtigheden anvendes der fem betegnelser, der knytter sig til elementerne i confusion matrixen. Til beskrivelsen af disse betegnelser tages der i det efterfølgende udgangspunkt i elementerne med klasseinddelingen *lavt*, hvor de fem betegnelser fremgår af tabel S.5. I tabel S.6 er betegnelserne illustreret i confusion matrixen.

Parameter	Beskrivelse
True positive (TP)	Denne betegnelse omfatter de elementer, hvis virkelige klasseinddeling er <i>lavt</i> og som tilsvarende estimeres til klasseinddelingen <i>lavt</i> ved brug af klassificeringstræet.
False positive (FP)	Denne betegnelse omfatter de elementer, der estimeres til at have klasseinddelingen <i>lavt</i> , men hvor deres virkelige klasseinddeling er <i>gennemsnitligt</i> , <i>højt</i> eller <i>meget højt</i> .
False negative (FN)	Denne betegnelse omfatter de elementer, hvis virkelige klasseinddeling er <i>lavt</i> , men som klassificeres som værende en del af klasseinddelingerne <i>gennemsnitligt</i> , <i>højt</i> eller <i>meget højt</i> .
True negative (TN)	Denne betegnelse omfatter de elementer, der er klassificeret korrekt, men for henholdsvis klasseinddelingerne <i>gennemsnitligt</i> , <i>højt</i> eller <i>meget højt</i> .
Negative (N)	Denne betegnelse omfatter de elementer, der er fejlklassificeret ved klasseinddelingerne <i>gennemsnitligt</i> , <i>højt</i> eller <i>meget højt</i> .

Tabel S.5: Definition af elementer fra confusion matrixen, som anvendes til beregning af forskellige mål for et klassificeringstræs nøjagtighed i Weka [58, 90].

Confusion matrix				
↓ virkelig klasse / → klassificeret som:	a	b	c	d
a	TP	FN	FN	FN
b	FP	TN	N	N
c	FP	N	TN	N
d	FP	N	N	TN

Tabel S.6: Illustration af anvendelse af benævnelserne *TP*, *TN*, *FP*, *FN* ved vurdering af klasseinddelingen *lavt*.

S.1.2 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 besluttet 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.

Nøjagtigheden (*accuracy*) af et klassificeringstræ afhænger af hvor mange hvor elementer, ud af den samlede datamængde, der er klassificeret korrekt. Dette kan beregnes ud fra formel S.6 [63], hvori parametrene, der er defineret i tabel S.5, er anvendt. *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.

$$Accuracy = \frac{TP + TN + N}{TP + TN + FP + FN + N} \quad (S.6)$$

Modsat *accuracy* kan andelen af fejlklassificerede elementer ligeledes bestemmes. Dette foregår ved, at udskifte tælleren i brøken med $(FP + FN + N)$. Resultat modsvarer naturligvis andelen af korrekt klassificerede elementer, hvorfor den er lig 15 % for det udviklede klassificeringstræ. Da klassificeringstræet er udviklet på baggrund af fiktivt træningsdata, kan der ikke konkluderes direkte på noget, men resultatet anvendes til at eksemplificere beregningerne.

S.1.3 True positive rate og False positive rate

True positive rate (*TP-rate*) beskriver klassificeringstræets „egenskab“ til at klassificere elementer inden for en given klasseinddeling korrekt i forhold til det samlede antal elementer med den pågældende klasseinddeling. *TP-rate* kan variere mellem 0 – 1, hvor en høj TP-værdi betyder, at klassificeringstræet egenskab til at klassificere elementerne med den pågældende klasseinddeling korrekt er god. *TP-rate* beregnes, som illustreret i formel S.7, og resultaterne fra Weka fremgår af tabel S.7. [7, 90]

$$TP - rate = \frac{TP}{TP + FN} \quad (S.7)$$

False positive rate (*FP-rate*) 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. *FP-rate* kan variere mellem 0 – 1 hvor en lav værdi betyder, at klassificeringstræet er god til ikke at fejlklassificere elementer som tilhørende eksempelvis klasseinddeling *lavt*, i forhold til det samlede antal elementer der tilhører klasseinddelingerne *gennemsnitligt*, *højt* og *meget højt*. *FP-rate* bestemmes som illustrere i formel S.8, og resultaterne fra Weka fremgår af tabel S.7. [7, 90]

$$FP - rate = \frac{FP}{FP + TN + N} \quad (S.8)$$

Begreb	Klasseinddeling			
	Lavt	Gennemsnitligt	Højt	Meget højt
TP-rate	0,886	0,881	0,813	0,571
FP-rate	0,077	0,121	0,036	0,0

Tabel S.7: Værdier for *TP-rate* og *FP-rate* fra det udviklede klassificeringstræ.

Det fremgår af tabel S.7, 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 0,571, 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 værdierne er lave for alle fire klasseinddelinger. Dette betyder, at der for hver klasseinddeling ikke fejlklassificeres specielt mange elementer med en anden klasseinddeling.

S.1.4 Precision

Precision beskriver sammenhængen mellem antallet af elementer, der er klassificeret korrekt, i forhold til mængden af elementer der totalt er klassificeret med klasseinddelingen, som det er udtrykt i formel S.9. Resultaterne fra Weka fremgår af tabel S.8 [89]. *Precision* kan variere mellem værdien 0 – 1, hvor en høj værdi for *precision* betyder, at klassificeringstræet er god til at klassificere elementerne i den pågældende klasseinddeling korrekt.

$$Precision = \frac{TP}{TP + FP} \quad (S.9)$$

Begreb	Klasseinddeling			
	Lavt	Gennemsnitligt	Højt	Meget højt
Precision	0,861	0,881	0,813	1,0

Tabel S.8: Værdier for precision for det udviklede klassificeringstræ, jf. appendiks Q.

Ved vurdering af resultaterne fra Weka fremgår det, at der opnås en maksimal værdi for *precision* for klasseinddelingen *meget højt*, idet 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.

S.1.5 F-measure

F-measure 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 [89]. Ved beregning af *F-measure* anvendes mængden af elementer i hhv. TP, FP og FN, der er beskrevet i det foregående. Værdien for *F-measure* varierer imellem 0 – 1, hvor nøjagtigheden for klassificeringstræet er bedst ved en værdi på 1. Denne værdi opnås, når der

ingen fejlklassificerede elementer er for den pågældende klasseinddeling. Formlen for *F-measure* er illustreret i formel S.10 og resultaterne fra Weka fremgår af tabel S.9.

$$F\text{-measure} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (\text{S.10})$$

Begreb	Klasseinddeling			
	Lavt	Gennemsnitligt	Højt	Meget højt
F-measure	0,873	0,86	0,813	0,727

Tabel S.9: Værdier for *F-measure* fra Weka, jf. appendiks Q.

Af tabel S.9 fremgår det, at værdierne for *F-measure* varierer imellem 0,727 – 0,873 for klassificeringstræet, der er udviklet i kapitel 7. Værdierne må for klasseinddelingerne *lavt*, *gennemsnitligt* og *højt* anses som værende relativt høje, hvorimod værdien for klasseinddelingen *meget højt* er en anelse lavere.

Litteratur

- [1] Icek Ajzen. *TPB Diagram*.
<http://people.umass.edu/aizen/tpb.diag.html#null-link>, 2011.
- [2] Rune Vinther Andersen. *Occupant behaviour with regard to control of the indoor environment*. Ph.d.-afhandling, 2009.
- [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] Bilal M. Ayyub og Richard H. McCuen. *Probability, Statistics, and reliability for engineers and scientists*. Chapman & Hall/CRC Press LLC, 2003.
- [5] 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.
- [6] Niels J. Blunch. *Faktoranalyse - en metode til datakomprimering*. Ledelse og Erhvervsøkonomi/Handelsvidenskabeligt Tidsskrift/Erhvervsøkonomisk Tidsskrift, 36, 1972.
- [7] 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 of Waikato, 2010.
- [8] Henrik Brohus, Per Heiselberg, Allan T. Hesselholt og Henrik E. S. Rasmussen. *Application of Partial Safety Factors in Building Energy Performance Assessment*. Proceedings of Building Simulation 2009, University of Strathclyde, Glasgow, UK, 27.-30. juli 2009.
- [9] 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.
- [10] 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.
- [11] James M. Conway og Allen I. Huffcutt. *A Review and Evaluation of Exploratory Factor Analysis Practices in Organizational Research*. Organizational Research Methods, 6:147–168, 2003.
- [12] Dansk Standard. *DS 474 - Norm for specifikation af termisk indeklima*. isbn ICS: 91.040.01. Byggedata, 1993.
- [13] Jamie DeCoster. *Overview of Factor Analysis*.
<http://www.stat-help.com/factor.pdf>, 1998.
- [14] Energistyrelsen. *Energistatistik 2009*. isbn 978-87-7844-872-9 www. Energistyrelsen, 2010.
- [15] 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.
- [16] P.O. Fanger. *Thermal Comfort - Analysis and Application in Environmental Engineering*. isbn 0-07-019915-9. McGraw-Hill Book Company, 1970.
- [17] Dr. Wolfgang Feist, Søren Peper og Manfred Görg. *CEPHEUS - Projectinformation No. 36*. Teknisk rapport, Passiv Haus Institut, 2001.
- [18] Dr. Wolfgang Feist, Søren Peper, Oliver Kah og Matthias von Oesen. *CEPHEUS - PEP Project Information No. 1 - Climate Neutral Passive House Estate in Hannover-Kronsberg: Construction and Measurement Results*. Teknisk rapport, Passiv Haus Institut, 2005.

-
- [19] Dr. Wolfgang Feist, Søren Peper og Dr.-Ing. Rainer Pfluger. *Ein nordorientiertes Passivhaus. Messtechnische Untersuchung und Auswertung von 19 Wohnungen im Passivhaus-Standard in Frankfurt-Bockenheim, GremptstraSe*. Teknisk rapport, Passiv Haus Institut, 2004.
- [20] Sekretariatet for energieffektive bygninger. *Statistik - Indberettede energimærker*. <http://www.maerkdinbygning.dk/Menu/Konsulent/Statistik>, year = 2011.
- [21] Åge Andersen. *Speciale om: Latente variable, faktor analyse og LISREL modeller anvendt på kommuners tilkendelse af førtidspensioner*. Specialeafhandling, Københavns Universitet - Økonomisk institut, 2004.
- [22] 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.
- [23] Zachary M. Gill, Michael J. Tierney, Ian M. Pegg og Neil Allan. *Measured energy and water performance of an aspiring low energy/carbon affordable housing site in the UK*. Energy and Buildings, 43:117–125, 2011.
- [24] Kirsten Gram-Hanssen. *Boligers energiforbrug - sociale og tekniske forklaringer på forskelle*. Teknisk rapport 87-563-1170-2, 2003.
- [25] Kirsten Gram-Hanssen. *Husholdningers elforbrug - hvem bruger hvor meget, til hvad og hvorfor?* Teknisk rapport 87-563-1235-0, 2005.
- [26] Kirsten Gram-Hanssen og Kirstine Nærvig Petersen. *Husholdningers energi- og vandforbrug*. Teknisk rapport 87-563-1231-8, 2005.
- [27] Olivia Guerra-Santin og Laure Itard. *Occupants' behaviour: determinants and effects on residential heating consumption*. (38:3, 318-338), 2010.
- [28] Frédéric Haldi. *Towards a Unified Model of Occupants' Behaviour and Comfort for Building Energy Simulation*. Ph.d.-afhandling, 2010.
- [29] P. Hoes, J.L.M. Hensen, M.G.L.C. Loomans, B. de Vries og D. Bourgeois. *User behavior in whole building simulation*. Energy and Buildings, 41:295–302, 2009.
- [30] SPSS IBM. *CART Algorithm*. <http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/TREE-CART.pdf>, 2011.
- [31] SPSS IBM. *CHAID and Exhaustive CHAID Algorithms*. <http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/TREE-CHAID.pdf>, 2011.
- [32] SPSS IBM. *QUEST Algorithm*. <http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/TREE-QUEST.pdf>, 2011.
- [33] Danmarks Meteorologiske Institut. *Verdensvejr*. <http://www.dmi.dk/dmi/index/verden.htm>, 2011.
- [34] H. Jeeninga, M. Uytendinck og J. Uitzinger. *Energy use of energy efficient residences*. Teknisk rapport, 2001.
- [35] Jesper Ole Jensen. *Livsstil, boform og ressourceforbrug*. Ph.d.-afhandling, By og Byg, Statens Byggeforskningsinstitut, 2002.
- [36] 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.
- [37] Casper Kofod. *Elforbrugets sammensætning, slutforbrugsanalyse på basis af målinger i 100 boliger i Odense*. Teknisk rapport.

-
- [38] Finn Kristiansen, Bjarne Saxhof, Kirsten Engelund Thomsen, Jørgen Schultz, Mogens Jørgensen, Poul Dedenroth og Lars Kæstel Jørgensen. *Lavenergirækkehuse, IEA - task 13, målinger og beregninger*. Teknisk rapport, 2000.
 - [39] J. Lammers. *Human factors, energy conservation and design practice*. Ph.d.-afhandling, 1978.
 - [40] K. Ming Leung. *Decision Trees and Decision Rules*. <http://cis.poly.edu/~mleung/FRE7851/f07/decisionTrees.pdf>, 2007. Polytechnic University - Department of Computer Science / Finance and Risk Engineering.
 - [41] Pawel Lewicki og Thomas Hill. *STATISTICS: Methods and Applications*. isbn 1-884233-59-7. StatSoft, 2006.
 - [42] J. Maeyens og A. Janssens. *A stochastic ventilation model regarding leakage and user behaviour*. isbn 90-5809-565-7. Research in building physics, 2003.
 - [43] Mari Martiskainen. *Affecting Consumer Behaviour on Energy Demand*. Teknisk rapport, Sussex Energy Group, Brighton, 2007.
 - [44] Mohammed M Mazid, A B M Shawkat Ali og Kevin S Tickle. *Improved C4.5 Algorithm for Rule Based Classification*. 9th WSEAS international conference on Artificial intelligence, knowledge engineering and data bases.
 - [45] Mario Mazzocchi. *Statistics for Marketing and Consumer Research*. isbn 978-1-4129-1122-1. SAGE Publications, 2008.
 - [46] Miljøstyrelsen og Watertech a/s. *Vandforbrug og forbrugsvariationer, Miljøprojekt Nr. 998*. Teknisk rapport, 2005.
 - [47] Myweather2. *Historical Monthly Averages*. <http://www.myweather2.com/>, 2011.
 - [48] Allan Neymark. *ID3 Algorithm*. www.cs.sjsu.edu/~lee/cs157b/ID3-AllanNeymark.ppt, 2007.
 - [49] J. Fergus Nicol og Michael A. Humphreys. *A Stochastic Approach to Thermal Comfort - Occupant Behavior and Energy Use in Buildings*. (NA-04-2-3), 2004.
 - [50] Annika Nilsson. *Energianvändning i nybyggda flerbostadshus på Bo01-området i Malmö*. Teknisk rapport 91-88722-30-9, Lunds Tekniska Högskola, Lunds universitet, 2003.
 - [51] Energi Nord. *Sammenlign dit elforbrug med gennemsnittet*. http://www.energinord.dk/dk/privat/spar_el/sammenlign_dit_elforbrug.htm, 2011.
 - [52] Erhvervs og Boligstyrelsen. *BBR-instruks - Retningslinier for føring af Bygnings- og Boligregistret (BBR)*. <http://www.ebst.dk/download/html/BBR/ren.htm>, year = 2011.
 - [53] Erhvervs og Byggestyrelsen. *Bygningsreglement 2010 (30.06.10 ->)*. <http://www.ebst.dk/bygningsreglementet.dk/br10/0/42>, 2011.
 - [54] Dansk Vand og Spildevandsforening. *Vand i tal 2010 - DANVAs benchmarking og vandstatistik 2010*. Teknisk rapport, DANVA, 2010.
 - [55] Malcolm Orme, Martin W. Lissament og Andrew Wilson. *Numerical Data for Air Infiltration & Natural Ventilation Calculations*. isbn 1-946075-97-2. International Energy Agency, AIVC, 1998.
 - [56] 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.
 - [57] 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.

-
- [58] University of Waikato NZ Peter Reutemann Dept. of Computer Science. *Confusion matrix*. <http://old.nabble.com/I%27m-confused-with-Confusion-Matrix,-Please-help-td21939603.html>, 2011.
- [59] John Ross Quinlan. *Induction of Decision Trees*. Machine Learning, 1:81–106, 1986.
- [60] John Ross Quinlan. *C4.5: programs for machine learning*. isbn 1-55860-238-0. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [61] John Ross Quinlan. *Data Mining Tools See5 and C5.0*. <http://www.rulequest.com/see5-info.html>, 2011.
- [62] John Ross Quinlan. *Is See5/C5.0 Better than C4.5?* <http://www.rulequest.com/see5-comparison.html>, 2011.
- [63] Andrew Roberts. *AI32 - Guide to Weka*. <http://www.andy-roberts.net/teaching/ai32/weka.pdf>, 2005.
- [64] Lior Rokach og Oded Maimon. *Data Mining with Decision Trees: Theory and Applications*. isbn 981-277-171-9. World Scientific, 2007.
- [65] H. Charles Romesburg. *Cluster Analysis for Researchers*. isbn 0-534-03248-6. Lifetime Learning Publications, 1984.
- [66] Olivia Guerra Santin, Laure Itard og Henk Visscher. *The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock*. Energy and Buildings, 41:1223–1232, 2009.
- [67] J. Schneiders og A. Hermelink. *CEPHEUS results, measurements and occupants satisfaction provide evidence for Passive Houses being an option for sustainable building*. Energy Policy, 34:151–171, 2006.
- [68] Allan Simonsen og Kim Christian Sørensen. *Usikkerhed ved energiforbruget i Danmark*. Specialeafhandling, Aalborg Universitet, 2007.
- [69] Karin Soldaat og Laure Itard. *Influence of occupants on the energy use of balanced ventilation*. Proceedings of Clima 2007.
- [70] Dansk Standard. *DS/CEN/CR 1752 - Ventilation i bygninger: Projekteringskriterier for indeklimaet*. Byggedata, 2001.
- [71] 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.
- [72] Dansk Standard. *DS490 - Lydklassifikation af boliger*. Byggedata, 2007.
- [73] Danmarks Statistik. *BOL522: Beboede boliger efter område, boligart, antal værelser og husstandsstørrelse*. <http://www.statistikbanken.dk/BOL522>, 2010.
- [74] Danmarks Statistik. *HISB3: Nøgletal om befolkningen*. <http://www.statistikbanken.dk/HISB3>, 2010.
- [75] Danmarks Statistik. *VAND1: Forbruget af drikke- og råvand efter område og anvendelse*. <http://www.statistikbanken.dk/vand1>, 2010.
- [76] Yannick Sutter, Dominique Dumortier og Marc Fontoynt. *The use of shading systems in VDU task offices: A pilot study*. Energy and Buildings, 38:780–789, 2006.
- [77] Pang-Ning Tan, Michael Steinbach og Vipin Kumar. *Introduction to data mining*. isbn 0-321-32136-7. Addison-Wesley, 2006.

-
- [78] 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.
- [79] Søren Risbjerg Thomsen. *Stat 4 - Faktoranalyse og klyngeanalyse*. <http://www.mit.ps.au.dk/srt/Seminar/Stat4.pdf>, 2001.
- [80] Y. Tochihara, Y. Kimura, I. Yadoguchi og M. Nomura. *Thermal responses to air temperature before, during and after bathing*. Proceedings of international conference of environmental ergonomics VIII, side 309–313, 1998.
- [81] Henrik Tommerup. *Måling af bruttoenergiforbrug i nybyggeri svarende til BR2005 energikrav, Byggesystem: Skalmurede porebetonelementer 1*. Teknisk rapport 87-7877-162-5, BYG-DTU, 2004.
- [82] Henrik Tommerup. *Måling af bruttoenergiforbrug i nybyggeri svarende til BR2005 energikrav, Byggesystem: Skalmurede porebetonelementer 2*. Teknisk rapport 87-7877-163-3, BYG-DTU, 2004.
- [83] Henrik Tommerup. *Energirenovering af typisk parcelhus opført i perioden 1960-1980, Del 2: Målinger*. Teknisk rapport 9788778772404, BYG-DTU, 2007.
- [84] Henrik Tommerup og Jørgen Rose. *Forsøgshus med nye typer klimaskærmskonstruktioner, Byggesystem: Lette ydervægselementer i stålskelet*. Teknisk rapport 87-7877-099-8, BYG-DTU, 2002.
- [85] Toolbox. *TPB - Theory of Planned Behavior (A)*. <http://toolbox.systime.dk/index.php?id=852>, 2011.
- [86] W. Fred van Raaij og Theo M.M. Verhallen. *Patterns of residential energy behavior*.
- [87] Theo M.M. Verhallen og W. Fred van Raaij. *Household behavior and the use of natural gas for home heating*. The Journal of Consumer Research, 8:253–257, 1981.
- [88] Wikipedia. *Algoritmer*. <http://da.wikipedia.org/wiki/Algoritme>, 2011.
- [89] Wikipedia. *Precision and recall*. http://en.wikipedia.org/wiki/Precision_and_recall, 2011.
- [90] Wikipedia. *Receiver operating characteristic*. http://en.wikipedia.org/wiki/Receiver_operating_characteristic, 2011.
- [91] Wikipedia. *Weka (machine learning)*. http://en.wikipedia.org/wiki/Weka_%28machine_learning%29, 2011.
- [92] Charlie Wilson og Hadi Dowlatabadi. *Models of decision making and residential energy use*. Annual Review of Environment and Resources, 32:169–203, 2007.
- [93] I. H. Witten, E. Frank og M. A. Hall. *Decision Trees and Decision Rules*. <http://www.cs.waikato.ac.nz/ml/weka/book.html>, 2011. Machine Learning Group at University of Waikato - slides til bog.
- [94] 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.