

Masteroppgave i kjemometri

**Nær- infrarød deteksjon av avgasser fra fremtidig
brenselcelle- konsept.**

Andreas Linge Tomren



Kjemisk Institutt
Universitetet i Bergen
Våren 2008

Forord

Mitt masterstudie ved Kjemisk Institutt, Universitetet i Bergen, begynte i samråd med Bjørn Grung høsten 2006. Masteroppgaven er et samarbeid mellom Christian Michelsen Research (CMR) og Kjemisk Institutt, og problemstillinger ble diskutert i fellesskap med Christian Michelsen Research. I oppgaven benyttes nær- infrarød spektroskopi og kjemometriske metoder for kvantitativ analyse av gasskomponenter i blanding.

Jeg vil takke min veileder Bjørn Grung for god veiledning i tiden på mastergrad. Takk rettes også til Egil Nodland for sin tålmodighet i sammenheng med laboratoriearbeidet på CMR, for sine spektroskopiske bidrag, og for at han tok seg tid til å hjelpe med oppgaveskrivingen helt på tampen.

Takk rettes til de ansatte ved Christian Michelsen Research for god støtte og innspill gjennom masteroppgaven.

Jeg vil takke mine medstudenter ved Kjemisk Institutt for stimulerende diskusjoner og støtte.

Til slutt vil jeg takke min samboer Susanne Nordvik for støtte gjennom hele studietiden min, og for at hun tok seg tid til å korrekturlese oppgaven min..

Bergen

Andreas Linge Tomren

Innholdsfortegnelse

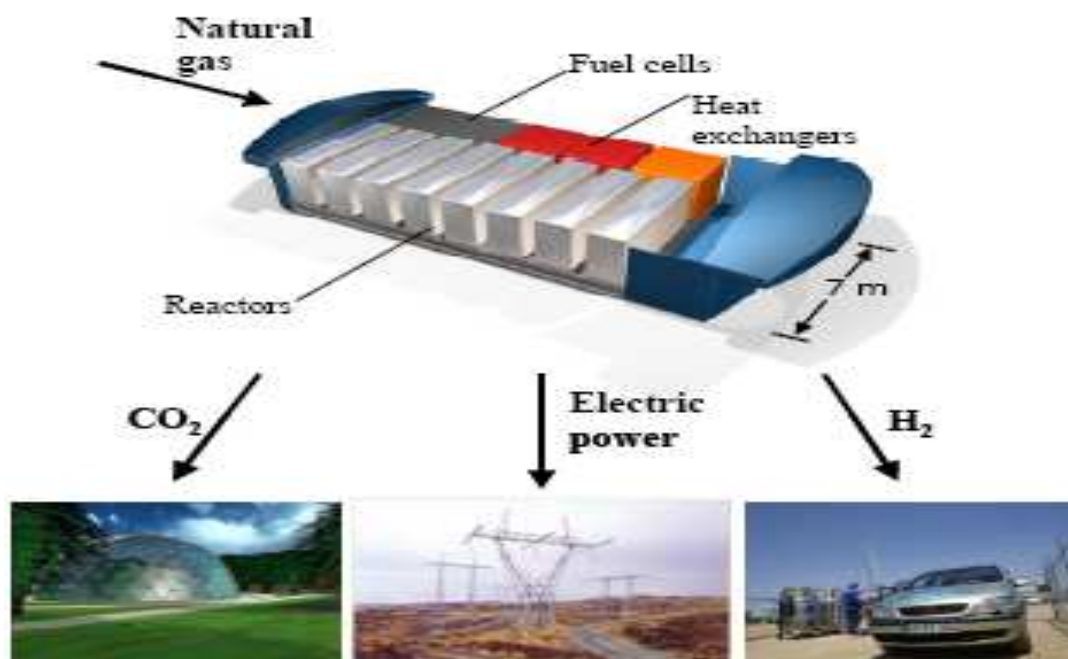
FORORD	1
INNHALDSFORTEGNELSE	2
1. INNLEDNING	5
2. NÆR- INFRARØD SPEKTROSKOPI (NIR)	8
2.1 Det elektromagnetiske spektrum	8
2.2 Vibrasjonsspektroskopi	10
2.2.1 Molekylære vibrasjoner	10
2.2.2 Absorpsjon av elektromagnetisk stråling	11
2.3.2 Harmoniske svingninger	12
2.3.3 Anharmoniske svingninger	14
2.3 IR som analytisk metode	16
2.4:Infrarøde spektrofotometre	17
2.4.1 Lyskilder	17
2.4.2 Bølgelengdeseparator	17
2.4.3 Prøvebeholder	20
2.4.4 Detektor	21
2.5 Måleteknikker	23
2.5.1 Transmittans	23
2.5.2 Diffus refleksjon	25
2.5.3 Dempet totalrefleksjon, ATR	26
2.5.4 Analysemetoder	27
2.6 Temperatureffekter	28
2.7 Den ideelle gasslov	29
3. MULTIVARIAT DATAANALYSE	32
3.1 Multivariat dataanalyse/kjemometri	32

3.2	Objektrom og variabelrom	32
3.3	Forbehandling	34
3.3.1	Sentrering	34
3.3.2	Variabelreduksjon	34
3.3.3	Derivasjon	35
3.4	Prinsipalkomponent analyse (PCA)	37
3.4.1	Latente variable	37
3.4.2	PCA	38
3.4.3	Skårer og ladninger	39
3.4.4	RSD (Residual Standard Deviation)	40
3.4.5	Uteliggere	41
3.5	Multivariat kalibrering	43
3.5.1	Partial Least Squares (PLS)	43
3.5.2	Mål for avvik	45
3.5.3	Intern validering/kryssvalidering	46
3.5.4	Ekstern validering	46
3.6	Instrumentstandardisering	47
3.6.1	Piece- wise Direct Standardization (PDS)	47
3.6.2	Generalized Least Squares Weighting (GLSW)	49
4.	EKSPERIMENTELT	51
4.1	Gassblandinger	51
4.2	Spekteropptak	52
4.2.1	Modellspektre	52
4.2.2	Valideringsspektre	53
4.2.3	Strømningsspektre	53
4.2.4	Høytemperaturspektre	53
4.2.5	Høytemperaturvalidering	54
4.2.6	Celle 2- spektre	55
5.	RESULTATER OG DISKUSJON	56
5.1	Modellbygging	57
5.2	Modellvalidering	63

5.3 Bakgrunnseffekter	64
5.4 Strømningseffekter	67
5.5 Modellbygging del 2	69
5.6 Temperatureffekter	76
5.6.1 Temperaturmodeller	78
5.6.2 Modeller uten oppdeling i høy og lav konsentrasjon	82
5.6.3 Modeller med oppdeling i høy og lav konsentrasjon	83
5.6.4 Temperaturvalidering	85
5.7 Kalibreringsoverføring	90
5.7.1 GLSW- standardisering	95
5.7.2 PDS- standardisering	97
5.7.3 Derivasjon	103
6. KONKLUSJONER OG DISKUSJON	105
6.1 Undersøkte problemer	105
6.1.1 Bakgrunnseffekter	105
6.1.2 Strømningseffekter	106
6.1.3 Vann sin innvirkning på modellene	106
6.1.4 Temperatureffekter	107
6.1.5 Kalibreringsoverføring	107
6.2 Konklusjon	108
6.3 Videre arbeid	109
REFERANSER	110
APPENDIX	113

1. INNLEDNING

Bakgrunnen for denne oppgaven er et samarbeidsprosjekt mellom Institutt for Energiteknikk (IFE), Christian Michelsen Research (CMR) og Prototech. Prosjektet heter Zero Emission Gas Power Project, ZEG¹.



Figur 1.1: Oversikt over samarbeidsprosjektet mellom IFE, CMR og Prototech.

Prosjektet innebærer et nytt konsept for produksjon av både elektrisitet og hydrogen gjennom integrering av høytemperatur- brenselcelleteknologi og reformering/kalsinering- teknologi med CO₂- rensing/lagring fra naturgass.

Naturgass er primært bestående av metan (CH₄), men også små mengder etan, propan, butan, pentan, karbondioksid (CO₂), nitrogen (N₂), helium og dihydrogensulfid².

Når naturgass brenner, frigjøres det CO₂ og vanndamp. Dersom det er et problem i forbrenningsprosessen vil det i tillegg være rester av metan og karbonmonoksid i avgassene grunnet ufullstendig forbrenning. Ved fullstendig forbrenning er situasjonen som i formel 1.1.



Ved ufullstendig forbrenning av metan kan situasjonen være som i formel 1.2.



Før prosjektet settes ut i full skala vil et laboratorieskalert "kraftanlegg" utprøves for å demonstrere og evaluere hele konseptet. Laboratorie- eksperimenter og simuleringer vil bli utført for å oppnå optimale forhold for reformeringsreaksjonen. For å oppnå optimale forhold for reformerings-reaksjonen vil online nærinfrarød (NIR) overvåking av avgasser fra brenselcellen (før reformeringsreaksjonen) utføres. Men først må modeller for prediksjon av avgasser bygges ved hjelp av kjemometrisk modellering, på bakgrunn av NIR-målinger av forskjellige gassblandinger av CH₄, CO₂ og CO. Dette er første steg mot en online- situasjon i reell drift, og forsøkes løst med laboratorieforsøk.

Noen av de potensielle problemene som kan oppstå, og dermed er undersøkt, er:

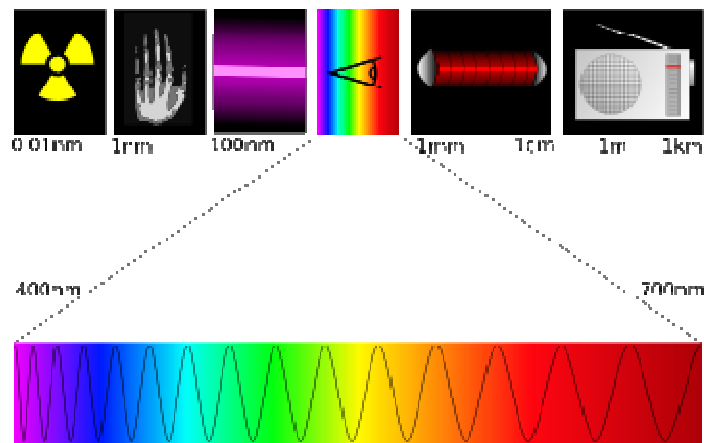
- *Bakgrunnseffekter:* Endres bakgrunnsspektrene vesentlig over tid? Hvilken effekt har forskjell i gasstrykk mellom bakgrunn og spekteret som korrigeres?
- *Strømningseffekter:* Vil prediksjoner av spektre tatt opp med strømmende gass være tilfredstillende og kan slike spektre korrigeres med bakgrunns-spektre tatt opp med stillestående gass?
- *Modellbygging:* Vil absorpsjonsbåndene fra vanndamp overskygge signalene til CO₂, CH₄ og CO i det nær- infrarøde området?
- *Temperatureffekter:* Vil toppintensiteter, båndbredder og topp- posisjoner endres vesentlig ved økt temperatur? Hvilke effekter vil det ha på modellering og prediksjon?
- *Kalibreringsoverføring:* Vil forskjellene mellom målecellene være store? Er det behov for instrumentstandardisering?

CMR har bygget to NIR- måleceller for prosjektet som benyttes i arbeidet med å bygge modeller for prediksjon av avgasser. Denne oppgaven er dermed et samarbeid mellom CMR og Universitetet i Bergen, ved student Andreas Linge Tomren.



2. NÆR- INFRARØD SPEKTROSKOPI (NIR)

2.1 Det elektromagnetiske spektrum



Figur 2.1: Det elektromagnetiske spektrum³.

Det elektromagnetiske spektrum (figur 2.1) er et kontinuum av alle elektro-magnetiske bølger sortert etter frekvens og bølgelengde. Synlig lys er en spesiell type av elektromagnetisk stråling som kan bli sett av menneskets øye, dette er området uthevet i figuren⁴. Infrarødt lys ligger mellom synlig lys og mikrobølger i det elektromagnetiske spektrum. Det nære- infrarøde området ligger nærmest synlig lys i bølgelengder⁵. Synlig lys innebærer bølgelengder fra ca 400 nm til ca 700 nm (nanometer). Tabell 2.1 viser hvor grensene mellom de forskjellige områdene går.

Tabell 2.1: Infrarøde områder

Område	Bølgetall, cm^{-1}	Bølgelengde, nm
Nær- infrarød	4000-12800	780- 2500
Midt- infrarød	200-4000	2500- 50000
Fjern- infrarød	10-200	50000- 100000

Disse grensene er ikke absolutte, det finnes for eksempel mennesker som kan se lys med både kortere bølgelengder enn 400 nm og lengre bølgelengder enn 700 nm. Det eksisterer også vitenskapsmenn som mener at det nærinfrarøde området går fra fra 690 nm til 3000 nm^6 .

Elektromagnetisk stråling kan betegnes som en harmonisk vekslende bølge av elektriske og magnetisk felter, som spres ut fra en kilde i rette linjer til den brytes eller reflekteres. Bølgemodellen beskriver den elektromagnetiske strålingen som bølger med egenskapene bølgelengde, λ , frekvens, ν , og amplitude, A . Bølgelengden er avstanden mellom to nabobølgetopper og oppgis i nanometer, frekvensen er antallet svingninger per sekund og oppgis i hertz (Hz). Bølgelengde og frekvens er omvendt proporsjonale; stor bølgelengde gir lav frekvens og motsatt. Amplitude er avstanden fra en bølges største utslag (toppunkt eller bunnpunkt) til likevektstilstanden.

Bølgeformelen gir sammenhengen mellom lysets hastighet c , bølgelengde og frekvens på denne måten:

$$c = \lambda \cdot \nu \quad (2.1)$$

Energien i det elektromagnetiske spektrum er gitt ved sammenhengen mellom strålingens bølgelengde og frekvens:

$$E = h\nu = \frac{hc}{\lambda} \quad (2.2)$$

der h =Plancks konstant= 6.63×10^{-34} Js.

Elektromagnetisk stråling kan i tillegg ansees som små masseløse fotoner med høy energi.

2.2 Vibrasjonsspektroskopi

Bindingene som holder molekyler sammen er aldri i ro, men vibrerer kontinuerlig. Når et molekyl blir bestrålt med lys vil bindingene absorbere energi (lys) med spesielle bølgelengder. I det nær- infrarøde området er det spesielt bindingene C-H, N-H, O-H og C=O som absorberer energi. I det infrarøde området skjer absorpsjonen som strekk- og bøyvibrasjoner i molekylet, mens i det nærinfrarøde området skjer absorpsjonen i overtoner og kombinasjonstoner⁷.

Spektroskopi kan defineres som studiet av absorpsjon eller emisjon av elektromagnetisk stråling⁸.

2.2.1 Molekylære vibrasjoner

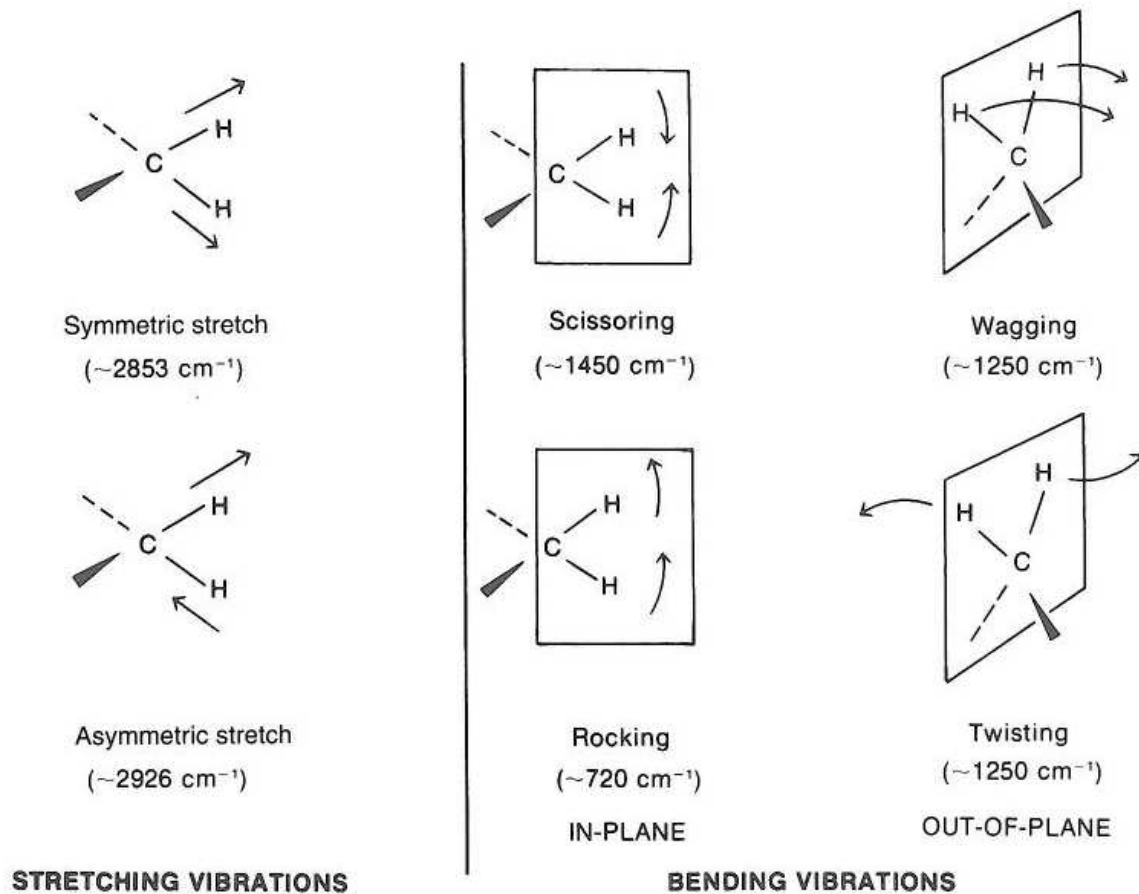
Molekylers energi kan deles inn som følger⁹:

$$E_{\text{molekyl}} = E_{\text{translasjon}} + E_{\text{rotasjon}} + E_{\text{vibrasjon}} + E_{\text{elektronisk}} \quad (2.3)$$

I vibrasjonsspektroskopi er det E_{rotasjon} (0,1- 30 cm^{-1}) og $E_{\text{vibrasjon}}$ (100- 4000 cm^{-1}) som studeres. Hver av disse kan deles inn i flere vibrasjonsnivåer hvor molekylet har fått økt energi, men uten at elektronene eksiteres. Disse vibrasjonsnivåene, E_0 =grunntilstand, og E_x =økt energi, kan igjen deles inn i flere ulike rotasjonsnivåer.

Overganger mellom E_0 og E_x finner sted når energi tilsvarende $\Delta E = E_x - E_0 = h\nu$, enten absorberes eller frigis fra molekylet. Blir molekylet bestrålt med lys av en bestemt bølgelengde $\lambda = c/\nu$ vil molekylet absorbere energien og dets indre energi øker fra grunntilstanden til det økte energinivået E_x . Samme energimengde vil sendes ut fra molekylet når en slutter å bestråle molekylet.

Figur 2.2 viser de forskjellige typene av vibrasjoner.



Figur 2.2.: Forskjellige typer vibrasjoner¹⁰.

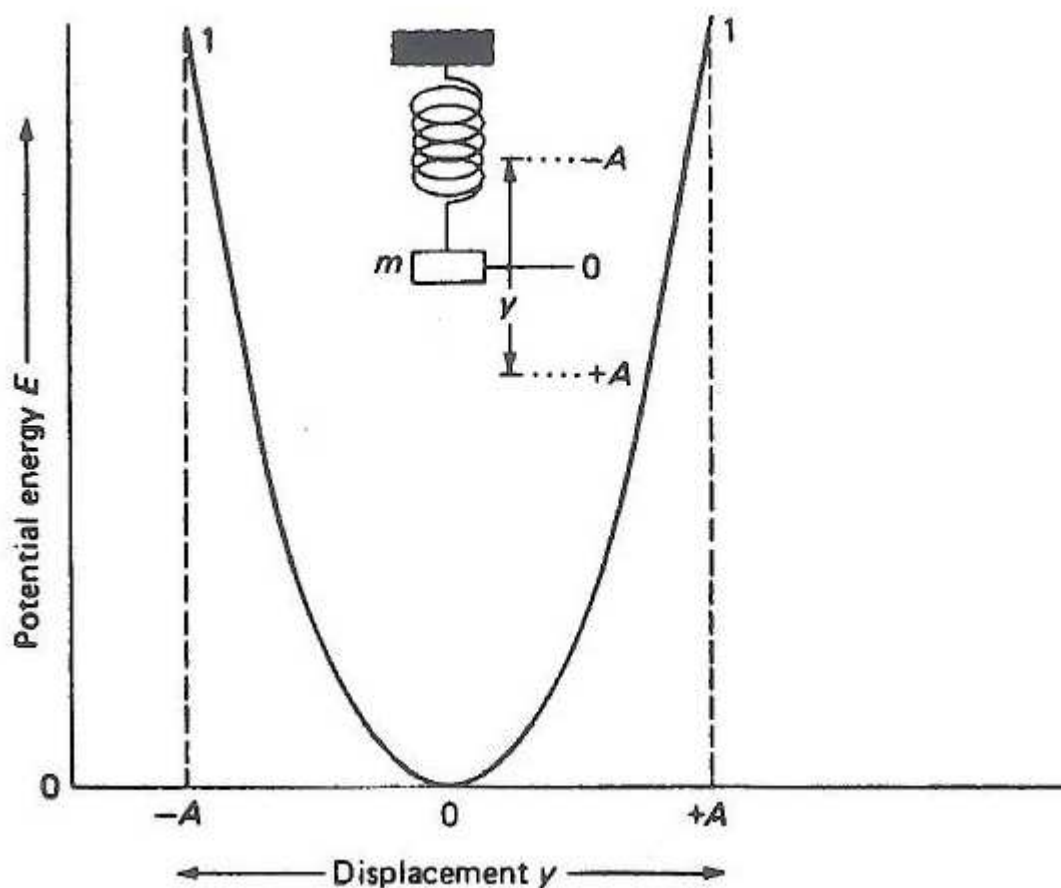
2.2.2 Absorpsjon av elektromagnetisk stråling

Dersom frekvensen mellom to vibrerende atomkjerner sammenfaller med det svingende elektriske feltet fra innsendt elektromagnetiske stråling og vibrasjonen medfører en netto endring i det molekulære dipolmomentet, vil infrarød stråling absorberes.

Dette medfører at homonukleære molekyler som N₂ og H₂, som har symmetrisk ladningsdistribusjon, ikke kan absorbere IR- stråling¹¹. Videre vil et molekyl med kraftig dipol absorbere mer infrarød stråling enn et molekyl med svak dipol.

2.2.3 Harmoniske svingninger

Ved hjelp av en harmonisk svingningsgenerator¹², en mekanisk modell som består av en masse m som holdes fast av en fjær (se figur 2.3), kan en tenke seg hvordan to atomer i et molekyl vibrerer.



Figur 2.3: Harmonisk svingningsgenerator. A er amplitude, y er avstand fra likevektsposisjon.

Energi kreves for å strekke eller presse samme fjæren. Når fjæren slippes vil den strekkes ut og presses sammen i sykluser, og bevegelsene til massen m vil i teorien være harmoniske. I praksis er alle svingesystemer mer eller mindre dempet¹³.

Bevegelsen langs aksene av fjæren kalles en harmonisk bevegelse. Den potensielle energien til en harmonisk oscillator er gitt ved:

$$V = \frac{1}{2}ky^2 \quad (2.4)$$

k er kraftkonstanten til fjæren, y er distansen fra likevektsposisjonen. Overført til molekylnivå er kraftkonstanten forskjellig fra binding til binding mellom ulike atomer. Den potensielle energien er 0 i likevektsposisjonen og øker når fjæren blir strukket eller presset sammen.

Hookes lov sier at vibrasjonsfrekvensen for en kjemisk binding øker når bindingsstyrken øker og når den reduserte massen av vibrasjonssystemet avtar:

$$\nu_m = \frac{1}{2\pi} \sqrt{\frac{k}{m_1 m_2 / (m_1 + m_2)}} \quad (2.5)$$

Der ν_m =vibrasjonsfrekvens, k =konstant relatert til bindingsstyrke, m_1 og m_2 =massene til de to atomene som utgjør vibrasjonssystemet, og $m_1 m_2 / (m_1 + m_2)$ =redusert masse.

Separasjonen mellom to energinivå er gitt ved:

$$E_{v+1} - E_v = h\nu_m \quad (2.6)$$

Separasjonen er lik for alle vibrasjonsfrekvensene, og siden den laveste tillatte verdien for den potensielle energien til en harmonisk svingningsgenerator er 0 vil en harmonisk svingningsgenerator ha nullpunktsenergi

$$E_0 = \frac{1}{2}h\nu_m \quad (2.7)$$

For å eksitere molekylet til en høyere energitilstand må en tilføre elektromagnetisk stråling med en frekvens som tilsvarer energiforskjellen mellom de to nivåene:

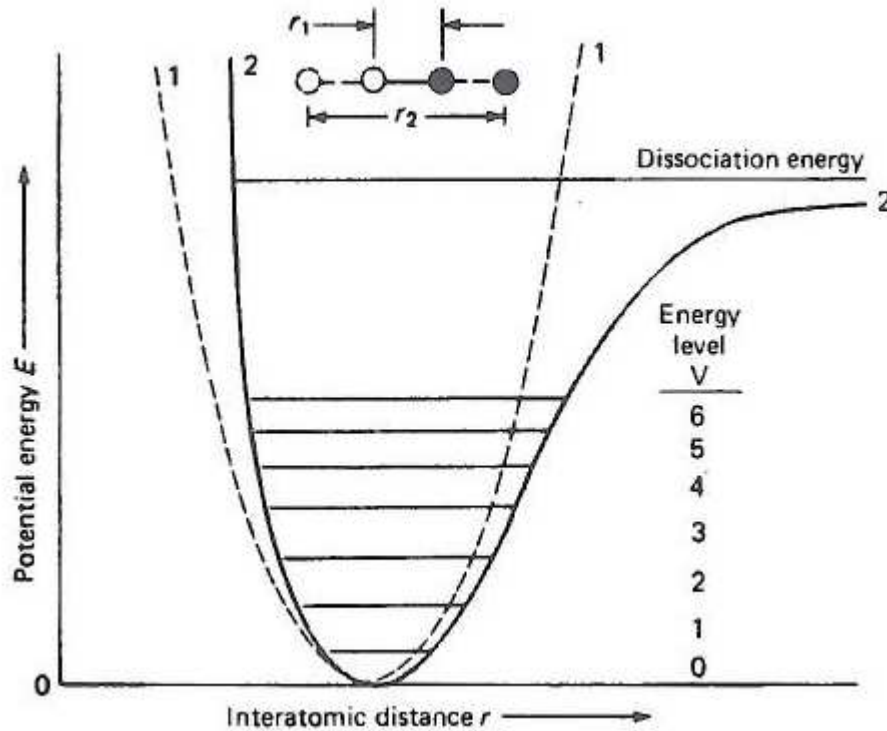
$$\Delta E = h\nu \quad (2.8)$$

I modellen for en harmonisk svingningsgenerator er det bare tillatt med transisjoner som fører til at energinivået blir endret med en enhet. Den mest vanlige transisjonen er fra nivå 0 til nivå 1, ettersom de fleste molekyler befinner seg i grunntilstanden før de blir bestrålt. Denne transisjonen kalles fundamentaltransisjonen og den tilhørende vibrasjonsfrekvensen kalles fundamentalvibrasjonen. Det er disse fundamentalvibrasjonene som er årsak til absorpsjonsbåndene i det midt- infrarøde området.

2.2.4 Anharmoniske svingninger

Den harmoniske svingningsgeneratoren er en tilnærming som gjelder nær likevektstilstanden for en kjemisk binding. Harmoniske svingninger mellom to atomer forekommer kun for svake vibrasjoner.

For å forklare absorpsjonsbåndene i det nærinfrarøde området kan en se på en anharmonisk svingningsgenerator:



Figur 2.4: Anharmonisk svingningsgenerator. Potensialkurven for harmonisk svingningsgenerator er vist med stiplet linje, den heltrukne linjen gjelder for anharmonisk svingningsgenerator.

For et molekyl som består av to atomer vil en bestemt mengde energi kunne strekke en binding lengre enn den kan komprimeres, i tillegg vil den potensielle energien minke ettersom avstanden mellom to atomer øker fordi bindingsstyrken minker kraftig. Dette gjenspeiles i figur 2.4. Bevegelsen blir anharmonisk fordi kraften F ikke lenger er proporsjonal med forflytningen y .

Med redusert potensiell energi vil ΔE bli mindre slik at regelen om at tillatte transisjoner skal ha endring i energinivå $= \pm 1$ ikke blir oppfylt. Transisjoner med endring i energinivå $= \pm 2$ og ± 3 blir observert, omtalt som henholdsvis 1. og 2. overtone.

Kombinasjons- og differanse- bånd vil oppstå i tillegg til overtone- bånd. To eller flere vibrasjoner kan påvirke hverandre og vibrasjonsbånd der frekvensene er tilnærmet lik summen eller differansen av fundamentalfrekvensene

$$V_{\text{kombinasjon}} = g_1 V_1 \pm g_2 V_2 \pm \dots \tag{2.12}$$

der g_1 og g_2 er positive heltall. Et kombinasjonsbånd er bare tillatt dersom det involverer vibrasjoner fra samme funksjonelle gruppe i molekylet.

Kombinasjoner og overtoner er såkalte forbudte transisjoner, intensiteten på disse båndene er derfor mellom 10- 100 ganger svakere enn fundamentalbåndene.

I nærinfrarød- spektroskopi er det disse kombinasjons- og overtone- båndene som blir observert.

For eksempel har C-H utslag i området $4070-4650 \text{ cm}^{-1}$ som stammer fra første kombinasjonstone¹⁴.

2.3 IR som analytisk metode

William Herschel oppdaget den infrarøde delen av det elektromagnetiske spektrum i 1800. De infrarøde bølgene inneholder ikke nok energi til å eksitere elektroner i kjemiske forbindelser, men kan få atomene til å vibrere. Et molekyl som absorberer elektromagnetisk stråling vil få økt indre energi, som må frigis for å komme tilbake til det mer stabile energinivået. Denne energien frigis ved å bøye og strekke molekylets bindinger.

IR- spektroskopi er mye brukt til kvalitativ og kvantitativ analyse, mens det nær- infrarøde området lenge lå ubrukt. Først på slutten av 1960- årene og frem til i dag har NIR- spektroskopi blitt mer interessant. Årsaken til at NIR var lite brukt er at spektrene er lite selektive, med brede topper og har lav intensitet. Men etter at en fant ut at en kan bruke multivariat kalibrering for å trekke ut nyttig informasjon av spektrene ble det mer interessant å bruke NIR.

2.4 Infrarøde spektrofotometre

Det eksisterer en stor mengde forskjellige infrarøde spektrofotometre, bruksområder er blant annet prosessanalyse og laboratorieanalyse. Produsenter av infrarøde spektrofotometre er blant andre ABB Bomem og Angstrom Advanced Inc.

2.4.1 Lyskilder

I nærinfrarød- spektroskopi er det vanlig å benytte kontinuerlige lyskilder, det vil si lyskilder som sender ut stråling med ett kontinuerlig bånd av bølgelengder. Det eksisterer i tillegg en annen type lyskilde som består av ett eller flere bånd av diskrete bølgelengder, denne type lyskilde kalles linjekilde.

Hvilken lyskilde en bør bruke avhenger av bølgelengdeområdet en bruker, ettersom forskjellige lyskilder sender ut lys med forskjellige bølgelengder. Lyskilden bør sende ut stråling med tilstrekkelig styrke for god deteksjon og måling. For å oppnå reproducerbare målinger bør i tillegg lysstyrken være konstant. De vanligste lyskildene i de midt- infrarøde området er inerte faste stoff som blir varmet opp til 1500- 2000 K, for eksempel nikkel- krom vaier eller silisiumkarbid. For det nær- infrarøde området er wolfram- filament lamper og wolfram- halogen lamper vanlige. Rundt 2900 K er vanlig temperatur for nær- infrarød spektroskopi, ettersom mesteparten av strålingen vil være i det nær- infrarøde området.

Det benyttes en halogen- lyskilde i denne oppgaven.

2.4.2 Bølgelengdeseparator

For å separere bølgelengdene i den infrarøde strålen kan en benytte en av tre måter:

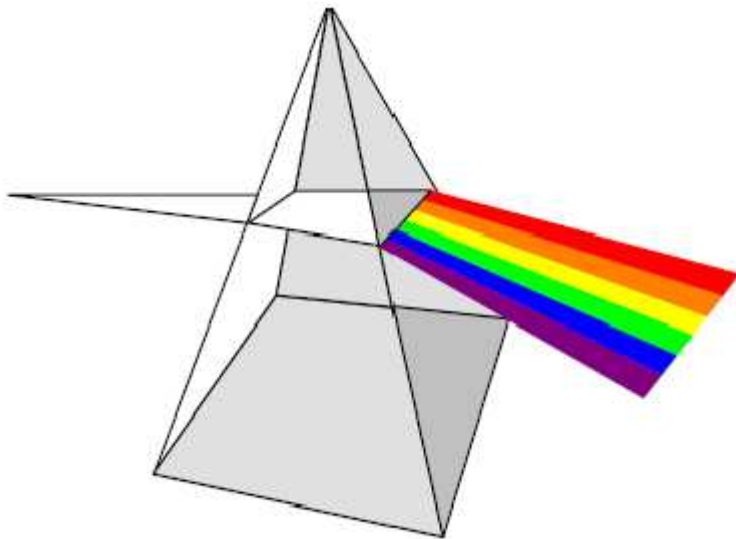
- Et filter
- En monokromator (bølgelengdevelger)
- Et interferometer

I filterbaserte instrumenter sendes strålen mot et gitter som separerer og reflekterer lyset. Diode Array Detector (DAD) benyttes av filterbaserte instrumenter, og består av topols elektronrør oppstilt på rekke som bestråles av alle bølgelengder samtidig. Hele bølgelengderegionen detekteres dermed med en gang.

En monokromator (bølgelengdevelger) består av 4 komponenter:

- Inngangsspalte
- Speil eller linse
- Prisme eller gitter
- Utgangsspalte

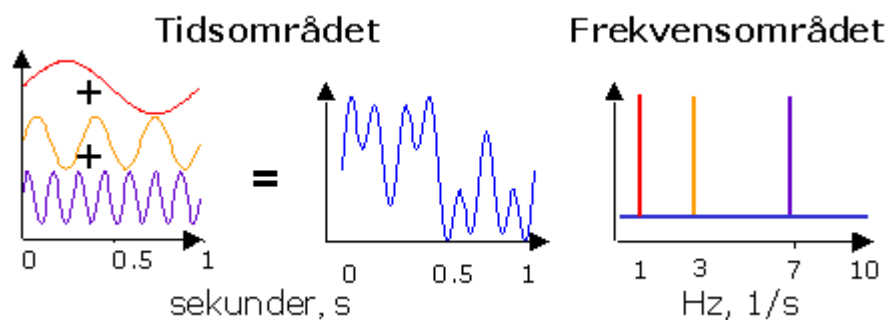
Inngangsspaltan gir rektangulær lyskilde. Speilet eller linsen samler og fokuserer strålingen i parallelle stråler. Prismet eller gitteret sprer strålingen til de aktuelle bølgelengdene. Utgangsspaltan isolerer det spektrale båndet som skal brukes. Figur 2.5 viser separasjonen.



Figur 2.5: Separasjon av bølgelengder ved hjelp av et prisme¹⁵.

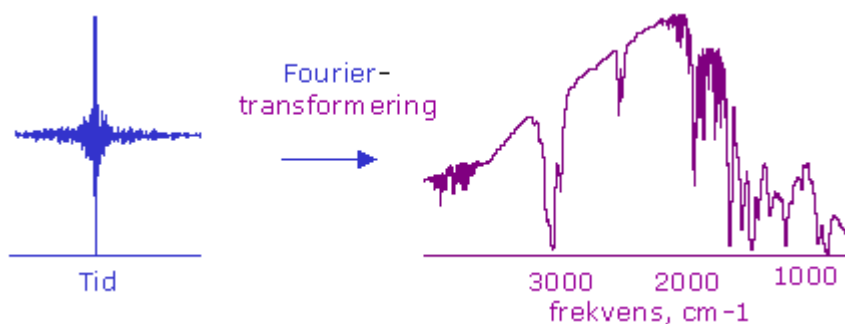
I et interferometer deles lysstrålen opp av en strålesplitter i to like porsjoner, der den ene delen av lyset går gjennom strålesplitteren og treffer et bevegelig speil, mens den andre delen reflekteres 90 grader og treffer et fastmontert speil. Ved å flytte det bevegelige speilet kan lysveien endres slik at veilengden blir lik for de to lysstrålene. De to delene av lyset har ulik

fase og ulik intensitet, og interferogrammet dannes ved at bølger med lik lengde slår hverandre ut. Et interferogram er et signal satt sammen av mange bølger med ulik frekvens. Figur 2.6 viser hvordan fouriertransformering brukes til å separere bølger med ulike frekvenser.



Figur 2.6: Bølger med ulik frekvens (til venstre) danner et signal (i midten) som ved bruk av fourier- transformasjoner skiller frekvenser fra hverandre (til høyre)¹⁶.

I infrarøde måleinstrumenter er det tusenvis av bølgelengder som gir tusenvis av frekvenser, som vist i figur 2.7.



Figur 2.7: Tusenvis av bølgelengder gir tusenvis av frekvenser¹⁶.

2.4.3 Prøvebeholder

Det finnes flere typer prøvebeholdere, blant annet den vanlige kyvetten. Et annet instrument er ikke utstyrt med kyvette eller annen type prøvebeholder, men en probe som er koplet til instrumentet gjennom en fiberoptisk kabel.

Proben stikkes ned i glasset/beholderen der prøven oppholdes. Tuppen på proben kan byttes ut avhengig av om en skal måle transmittans eller reflektans, og en kan justere veglengden ved å endre avstanden mellom refleksjonsspeilet og safirvinduet, slik at veglengden blir to ganger avstanden mellom refleksjonsspeilet og safirvinduet. Figur 2.8 viser hvordan tuppen ser ut.



Figur 2.8: Probetupp¹⁶.

I denne oppgaven har CMR bygget en målecelle som skal tåle 5-6 atmosfærers trykk og 70-80 grader celsius, ettersom det er gass med forskjellig trykk som skal måles. Gassene som kommer fra brenselcellen skal avkjøles slik at det er maksimalt 60-70 grader når gassene føres inn i målecellen. Prøvebeholderen, eller målecellen, som er benyttet i denne oppgaven er vist i figur 2.9.



Figur 2.9: Prøvebeholder/ målecelle benyttet i denne oppgaven.

2.4.4 Detektor

Detektoren oversetter mottatt stråling til signaler, og bør ha høy selektivitet samt stort signal/støy- forhold i den regionen som måles, samt at den konverterer strålingen til signaler raskt og gir minimalt signal ved fravær av stråling.

Det er to hovedtyper av detektorer: termoelektriske og fotoelektriske.

Termoelektriske detektorer overfører den mottatte strålingen til varme, som konverteres til elektrisk signal og relateres til den infrarøde strålingen.

Fotoelektriske detektorer benytter en aktiv overflate som kan absorbere stråling. Fotoner fra den elektromagnetiske strålingen har nok energi til å frigi elektroner fra detektoroverflaten, og

endringen i ledningsevnen til overflaten kan registreres og relateres til den infrarøde strålingen.

I denne oppgaven er en fotoelektrisk detektor benyttet: InAs (Indium Arsenid) 3500G.

Instrumentet som benyttes i denne oppgaven er delt i to deler. Den ene delen er en utmodul som sender ut lys; ABB Bomem FTPA (Fourier Transform Process Analysis) 2000, MOD SYM 3300G. Den andre delen er en innmodul der lys går inn etter å vært i kontakt med prøven; ABB Bomem TEC Multi- Detector module, MOD SNH 2700GS. Figur 2.10 viser instrumentet.



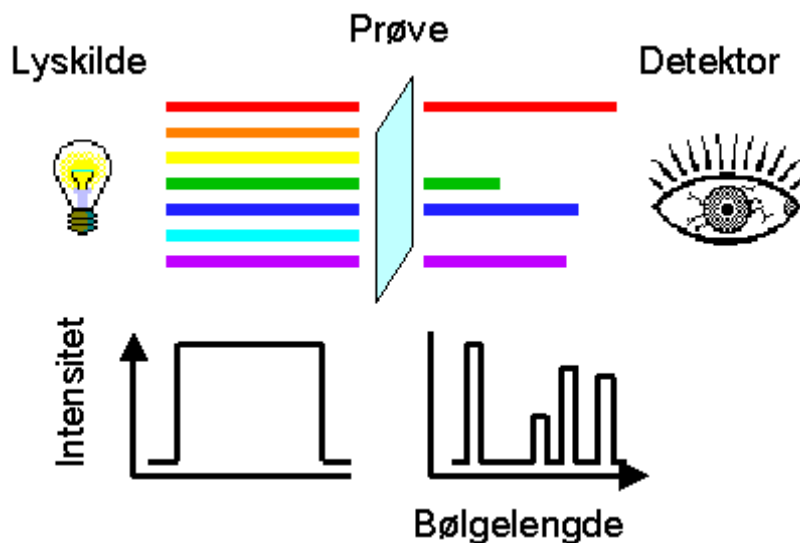
Figur 2.10: Måleinstrumentet som er benyttet i denne oppgaven. Utmodulen er underst, innmodulen er øverst.

2.5 Måleteknikker

Det er flere typer måleteknikker, hvilken som passer best avhenger av problemstillingen og hvilken fase som måles. Enkelte metoder brukes kun til analyse av klare løsninger og gasser, mens andre benyttes kun til analyse av faste stoffer.

2.5.1 Transmittans

Dette er den eldste og mest kjente teknikken. Den blir brukt på blanke væskeprøver og væsker som gir mindre lysspredning samt prøver i gassfase. Transmisjon kan måles ved å sende lys gjennom en prøve, og ut fra analyttens molekylstruktur vil ulike deler av den infrarøde strålingen absorberes, som vist i figur 2.11.



Figur 2.11: Illustrasjon av hvordan en prøve absorberer lys¹⁶.

For å eliminere naturlige forekomster av vanddamp, CO₂ og eventuelle andre kilder til signal som kan forurenske målspekteret, tar en opp et bakgrunnsspekter uten analytten tilstede. Deretter tar en opp spekter av analytten, dividerer dette spekteret på bakgrunnen og multipliserer med 100%. Ingen absorpsjon kan dermed settes til 100% transmittans.

$$\text{Transmittans} = \frac{I}{I_0}, \%T = \frac{\text{Spekter}}{\text{Bakgrunn}} \times 100\% \quad (2.13)$$

der I =intensitet av stråling som kommer ut av prøven og I_0 =intensiteten av strålingen som går inn i prøven. Enheten oppgis som transmittans eller % transmittans fordi det beregnes et forholdstall mellom to målinger.

Spektroskopi er en god måte for kvantitativ analyse ettersom responsen i et spekter avhenger av konsentrasjon av analytt i prøven, som kan vises ved å benytte Beer- Lamberts lov:

$$A = \varepsilon \times C \times l \quad (2.14)$$

der A =absorpsjon, C =konsentrasjon (mol/l), l =veilengde (cm) og ε =molar absorptivitet ($M^{-1}cm^{-1}$). A finner vi ved hjelp av transmittansen:

$$A = -\log(T) = \varepsilon \times C \times l \quad (2.15)$$

der $T = \frac{I}{I_0}$. Absorpsjon er dermed en størrelse som varierer lineært med konsentrasjonen av analytten.

Beer- Lamberts lov kan brukes til kvantitativ analyse av blandinger dersom det ikke er interferens mellom de ulike konstituentene i blandingen. Den totale absorbansen er gitt ved summen av absorbansen for hver konstituent.

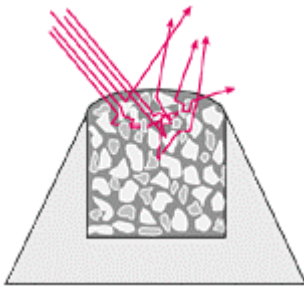
I praksis avviker målinger av transmittans fra Beer- Lamberts lov på grunn av interaksjoner mellom molekyler og forskjeller i trykk og temperatur. Fullstendig samsvar med Beer- Lamberts lov er bare blitt observert ved bruk av monokromatisk lys.

Ved anvendelse av FTIR kan man måle både CO , CO_2 og CH_4 samtidig, og bestemme konsentrasjon helt ned til noen få ppm (parts per million)¹⁷.

NIR- analyse av CH_4 og CO_2 i gass fase har blitt utført og er blitt funnet til å være en ganske direkte og enkel analyse¹⁸.

2.5.2 Diffus refleksjon

Diffus refleksjon blir brukt på faste stoff og væsker som gir kraftig lysspredning. Noe av strålingen som blir sendt mot overflaten av prøven vil reflekteres direkte fra overflaten av partiklene uten å trenge inn i partiklene. Resten av strålingen vil trenge inn i prøven og reflektere på overflaten til andre partikler i prøven, og deretter reflekteres tilbake til detektoren. Dette kalles diffus refleksjon, og vises i figur 2.12.



Figur 2.12: Diffus refleksjon¹⁶.

Målet på hvor mye av strålingen som reflekteres av en prøve er gitt ved:

$$R = \frac{I}{I_0} \quad (2.16)$$

der R = reflektans, I =intensitet av lyset som treffer detektoren etter refleksjon fra prøven og I_0 =intensitet av lyset som treffer detektoren uten å være hindret av prøven. Målet på hvor mye av strålingen som absorberes av en prøve er gitt ved:

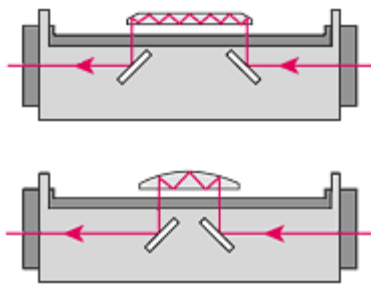
$$A = \log\left(\frac{1}{R}\right) = -\log(R) = \log\frac{I_0}{I} \quad (2.17)$$

Det vil si at dersom en prøve absorberer 25% av strålingen ved en gitt bølgelengde, er reflektansen ved denne bølgelengden 75%.

2.5.3 Dempet totalrefleksjon (ATR)

Som diffus refleksjon er ATR en refleksjonsteknikk. En infrarød lysstråle blir sendt gjennom en krystall som er gjennomsiktig for infrarød stråling. Denne krystallen må ha en høyere brytningsindeks enn prøven det skal måles på.

Den interne refleksjonen av den infrarøde strålingen i krystallen danner en kortlivet bølge som strekker seg ut fra krystallens overflate, og inn i en prøve som er i god kontakt med krystallen. Figur 2.13 viser hvordan lyset går gjennom krystallen.



Figur 2.13: Lysets gang gjennom krystallen¹⁶.

Bølgen blir dempet i områder av det infrarøde spekteret hvor prøven absorberer energi. Jo flere ganger strålen sendes inn i prøven, jo mer dempes bølgen. Endringer oppstår i den infrarøde strålingen som kommer ut i motsatt ende av krystallen. Endringene måles av en detektor og danner grunnlaget for å beregne prøvens infrarøde spektrum. Vanlige krystaller er germanium (Ge), zinkselenid (ZnSe) og diamant (C).

ATR benyttes for prøver som absorberer mye eller er tykke, for eksempel prøver som inneholder svart karbon og væsker som inneholder vann. Slike prøver gir ofte svært intense og ubrukelige toppler når de måles ved hjelp av infrarød transmisjon. Intensiteten i den kortlivede bølgen ved ATR- teknikken avtar raskt med avstanden fra krystallens overflate og fungerer derfor bra for slike prøver.. Dette medfører at teknikken er lite følsom for ulike prøvetykkelser. Andre typer faste materiale som er velegnet for å måle ved hjelp av ATR er homogene faste prøver, overflatelaget på flerlags prøver eller belegget på en fast prøve. En forutsetning for suksess er at prøven er flat nok eller så fleksibel at det er god kontakt mellom den og krystallen. I tillegg er ATR en enkel teknikk for væskeanalyse fordi målingen trenger kun en dråpe av væsken på krystallen.

Bruk av diamantkrystall krever omtrent ingen prøvepreparering og er veldig hardfør. Den tåler dermed at fast prøve presses hardt mot diamanten for å sikre god kontakt.

2.5.4 Analysemetoder

Det eksisterer fire metoder å utføre en analyse på:

- Offline
- At- line
- Online
- In- line

Tradisjonelt utføres analyserer offline. Det vil si at en prøve hentes ut fra produksjonsområdet, prøven transporteres til et analyselaboratorium og analyseres uten tilknytning til produksjonsutstyret. Dette er tidkrevende, og ikke minst lite økonomisk dersom en må stanse en hel bedrift mens en venter på analyseresultatet.

På den annen side påvirkes ikke følsomme analyseinstrumenter av støy fra produksjonutstyr og annen elektronikk. Offline- analyse utføres vanligvis i leting etter variable som er signifikante for prosessen, og i arbeidet med å utvikle verktøy for bruk av at- eller online- analyse.

At- line betyr at analyseutstyret befinner seg like ved produksjonsutstyret, men er uten direkte tilknytning til produksjonslinjen. En ulempe med at- line er at idet en tar ut en prøve kan forholdene endres (trykk, temperatur og lignende), det kan medføre systematiske feil. Det gjelder også offline- analyse.

Online betyr at analyseinstrumentet er koblet til produksjonslinjen, ved at produktene blir ført gjennom instrumentet og deretter videre i produksjonslinjen. Ulemper ved online analyse er at turbulens, luftbobler og andre partikler kan bringe med seg støy og gi dårlig kvalitet på signalene. Innretninger som reduserer virkningen av disse problemene kan monteres.

In- line betyr at en gjør målinger inne i reaktoren ved å benytte fiberoptiske kabler med måleinstrumenter, ved å senke det ned i reaktoren. Prosessen kan dermed bli fulgt veldig nært, men en kan derimot ikke gjøre stort med turbulens, luftbobler og partikler i reaktoren.

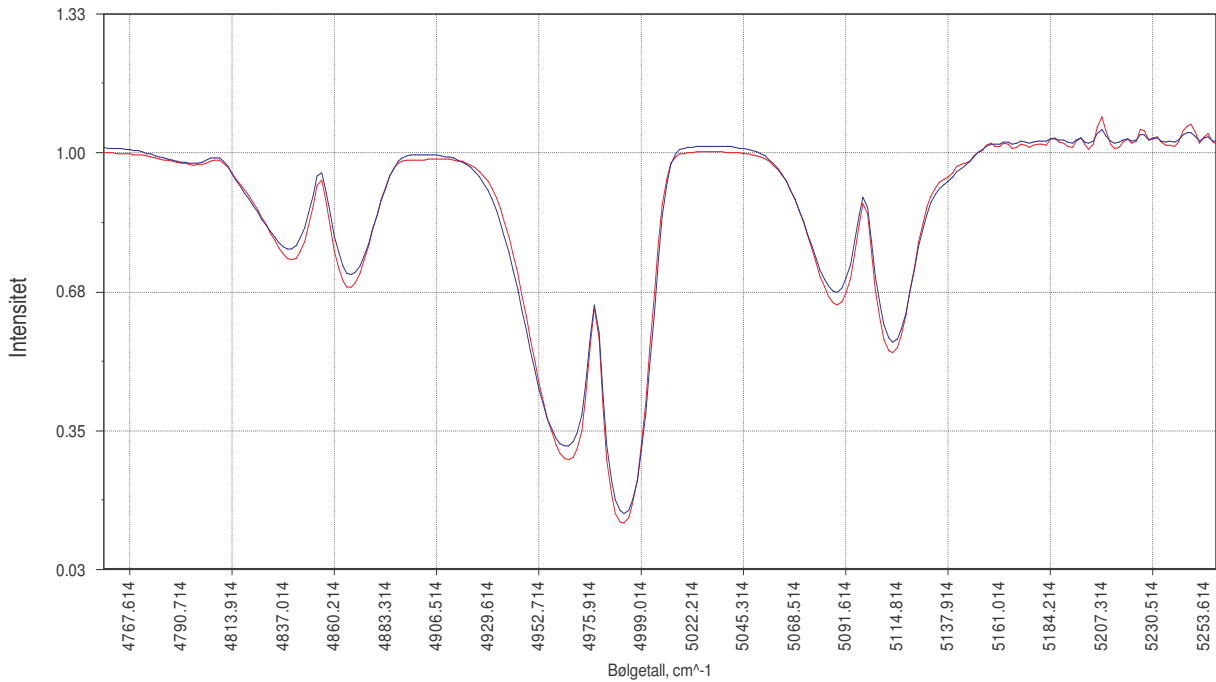
Denne oppgaven er første steg mot en online- situasjon i reell drift, i dette steget utføres analysene offline.

2.6 Temperatureffekter

Ved oppvarming blir generelt absorpsjonsbånd i IR bredere og mindre intense¹⁹. Hvor mye intensiteten synker varierer med det individuelle båndet^{20,21}. Samtidig blir bredden på båndet ved halv topphøyde økt. Dette medfører at arealet under toppene er tilnærmet konstant. Orlando et al.²² konstaterte at en reduksjon på ca. 5% ikke er signifikant innen eksperimentell usikkerhet.

Oftest skifter båndene posisjon i tillegg til at båndene blir bredere ved økt temperatur²³.

Figur 2.14 viser et eksempel av temperatureffekten, der to spektre av ren CO₂ er tatt opp ved to forskjellige temperaturer og med samme konsentrasjon. Det røde spekteret er tatt opp ved romtemperatur, det blå er tatt opp ved 60°Celsius.



Figur 2.14: Temperatureffekten. Rødt spekter=romtemperatur, blått spekter=60°Celsius.

Det blå spekteret har lavere toppe og bredere bånd som følge av økt temperatur.

2.7 Den ideelle gasslov

I det syttende århundre studerte Robert Boyle oppførselen til gasser²⁴. Han fant ut at trykk og volum er omvendt proporsjonale størrelser, gitt som

$$V \propto \frac{1}{P}, \text{ ved konstant } n \text{ (stoffmengde) og } T \text{ (temperatur, i } ^\circ \text{ Kelvin)} \quad (2.18)$$

der P =trykk og V =volum. Denne sammenhengen er kalt Boyles lov. Det eksisterer i tillegg en sammenheng mellom temperatur og volum/trykk, den sammenhengen er kalt Charles og Gay-Lussacs lov, og sier at volumet/trykket er proporsjonalt med temperaturen:

$$V \propto T, \text{ ved konstant } n \text{ og } P \quad (2.19)$$

Amedeo Avogadro fant i begynnelsen av 1800- tallet ut at volumet av en gass er proporsjonalt med stoffmengden, denne sammenhengen kalles Avogadros lov:

$$V \propto n, \text{ ved konstant } P \text{ og } T \quad (2.20)$$

Ved å kombinere de tre lovene, og dermed likning 2.18, 2.19 og 2.20, får en

$$V \propto \frac{nT}{P} \quad (2.21)$$

som kan omformes til

$$V = R \frac{nT}{P} \quad (2.22)$$

der R=gasskonstanten. Omformer en 2.22 litt til får en

$$PV = nRT \quad (2.23)$$

Likning 2.23 kalles *den ideelle gasslov*, og beskriver forholdet mellom de fire variablene P, V, n og T. En kan anta at reelle gasser oppfører seg som ideelle gasser, men ikke ved alle forhold. Ved økende trykk, 5 atm (1 atm=1.01325 bar) og høyere, avviker gassene mer og mer fra den ideelle gasslov²⁴. Van der Waal foreslo i 1873 å legge til et korreksjonsledd slik at den ideelle gasslov kan skrives som:

$$\left(P + \frac{an^2}{V^2} \right) (V - nb) = nRT \quad (2.24)$$

der a og b er van der Waal konstanter. Likning 2.24 kalles van der Waals lov.

Ettersom en arbeider med trykk mindre enn 5 atm vil ikke van der Waals lov, og dermed korreksjonsleddet, benyttes.

Formel 2.23 kan skrives som:

$$P = CRT \quad (2.25)$$

der C=konsentrasjon. Trykket avhenger dermed av konsentrasjonen til gassen som er i prøvebeholderen og temperaturen. Økes en av variablene mens den andre holdes konstant vil trykket øke.

I denne oppgaven regner en ut trykkene som hver av gasskomponentene består av ved hjelp av følgende formel:

$$P = \left(\frac{\text{Lufttrykk}(hPa)}{1000} \right) + \text{overtrykk}(bar) \quad (2.26)$$

Trykket som regnes ut er dermed gitt i bar. Ved å omforme formelen 2.25 kan en få konsentrasjonen gitt som:

$$C = \frac{P}{RT} \quad (2.27)$$

der C er gitt i mol/L, P er gitt i atm, R er gitt i $L \cdot atm \cdot K^{-1} \cdot mol^{-1}$ og T er gitt i Kelvin.(K).

I denne oppgaven ble først gasstrykk benyttet som respons for de forskjellige gasskomponentene, og senere ble gasstrykkene omregnet til konsentrasjon.

3. MULTIVARIAT DATAANALYSE/ KJEMOMETRI

Følgende notasjon gjelder for de neste kapitlene:

Fet skrift og stor bokstav: \mathbf{X} = matrise

Fet skrift og liten bokstav: \mathbf{t} = vektor

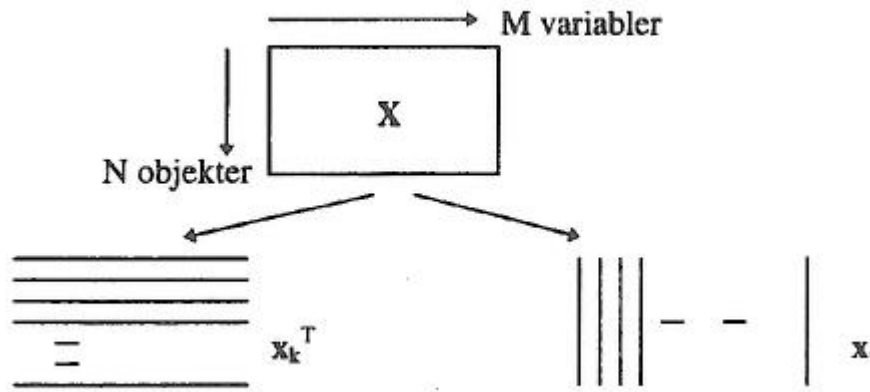
Bokstav opphøyd i T: \mathbf{X}^T =kolonner er transponert til rekker eller motsatt

3.1 Multivariat dataanalyse/ kjemometri

Multivariat dataanalyse innebærer analyse av data som inneholder mange variable. Når en har store datamengder, for eksempel ved spektroskopi der en benytter tusenvis av bølgelengder, blir det vanskelig å hente ut viktig informasjon. Da kan en benytte kjemometri. Kjemometri er et fagområde der en anvender avanserte matematiske og statistiske metoder for å planlegge og optimalisere forsøk, gjenkjenne mønstre og trekke ut relevant informasjon fra kjemiske data²⁵.

3.2 Objektrom og variabelrom

Kjemiske data, for eksempel i form av en spekterserie, kan bli representert som en matrise \mathbf{X} , kalt datamatriksen. Denne matrisen er bygget opp av rekker, tilsvarende N spektre eller objekter, og kolonner, tilsvarende M bølgelengder eller variable, og utgjør tilsammen en N x M- matrise, som vist i figur 3.1.



Figur 3.1: Datamatriksen X består av N rekkevektorer \mathbf{x}_k^T og M kolonnevektorer \mathbf{x}_i .

Ved å plote alle objektvektorer i et M - dimensjonalt koordinatsystem der aksene er de M variablene får en et koordinatsystem som kalles *variabelrommet*.

På samme måte får en *objektrommet* ved å plote variabelvektorene i et N - dimensjonalt koordinatsystem der aksene er de N objektene.

Til sammen utgjør variabelrommet og objektrommet all informasjon i datasettet. Variabelrommet inneholder informasjon om objektene i X , for eksempel korrelasjoner mellom de forskjellige objektene. Korrelasjonen mellom to objekter a og b er gitt ved:

$$\cos \varphi = \frac{\mathbf{x}_a^T \times \mathbf{x}_b}{\|\mathbf{x}_a\| \times \|\mathbf{x}_b\|} \quad (3.1)$$

Dersom vinkelen er 0° er objektene perfekt positivt korrelert, er vinkelen 180° er objektene perfekt negativt korrelert, og dersom vinkelen er 90° er det ikke noe felles informasjon objektene imellom. Likheter mellom objektene kan en finne ved å se på vinklene og avstanden mellom objektene: Liten avstand mellom to objekter forteller om likhet mellom objektene, og liten vinkel mellom to objekter forteller om likhet mellom objektene.

De samme prinsippene om likhet kan benyttes for objektrommet.

3.3 Forbehandling

Data som er hentet direkte fra måleinstrument kalles rådata. Rådata kan inneholde støy, drivende baselinje, spredningseffekter, dominerende variable og andre faktorer som overskygger den informasjonen som er signifikant. Derfor kan det være nødvendig å forbehandle data, for å fjerne effekter som ikke representerer kjemiske, fysiske eller biologiske egenskaper i måleprøven.

3.3.1 Sentrering

For at variasjonen mellom objektene skal fanges opp må rådata sentreres, og siden objektene posisjon i forhold til hverandre er av større interesse enn objektene posisjon i forhold til senteret i koordinatsystemet, er det viktig å sentrere data. Dersom en ikke sentrerer data trenger en vanligvis minst en ekstra latent variabel (forklares i kapittel 3.4) Sentrering kan gjøres på flere måter, blant annet ved kolonnesentrering, der en sentrerer variabler. En regner først ut gjennomsnittsverdien for hver variabel og deretter trekker denne fra de opprinnelige variablene:

$$\mathbf{X}_c(i, j) = \mathbf{X}(i, j) - \frac{1}{N} \sum_{i=1}^N x(i, j) \quad (3.2)$$

En flytter i praksis origo i koordinatsystemet til tyngdepunktet i datasettet.

3.3.2 Variabelreduksjon

Ofte inneholder datamatriksen veldig mange variabler, der ikke alle er like interessante. Modellene forbedres ofte dersom en fjerner variabler med mye støy, liten eller ingen variasjon og variabler med små regressjonskoeffisienter i modellene.

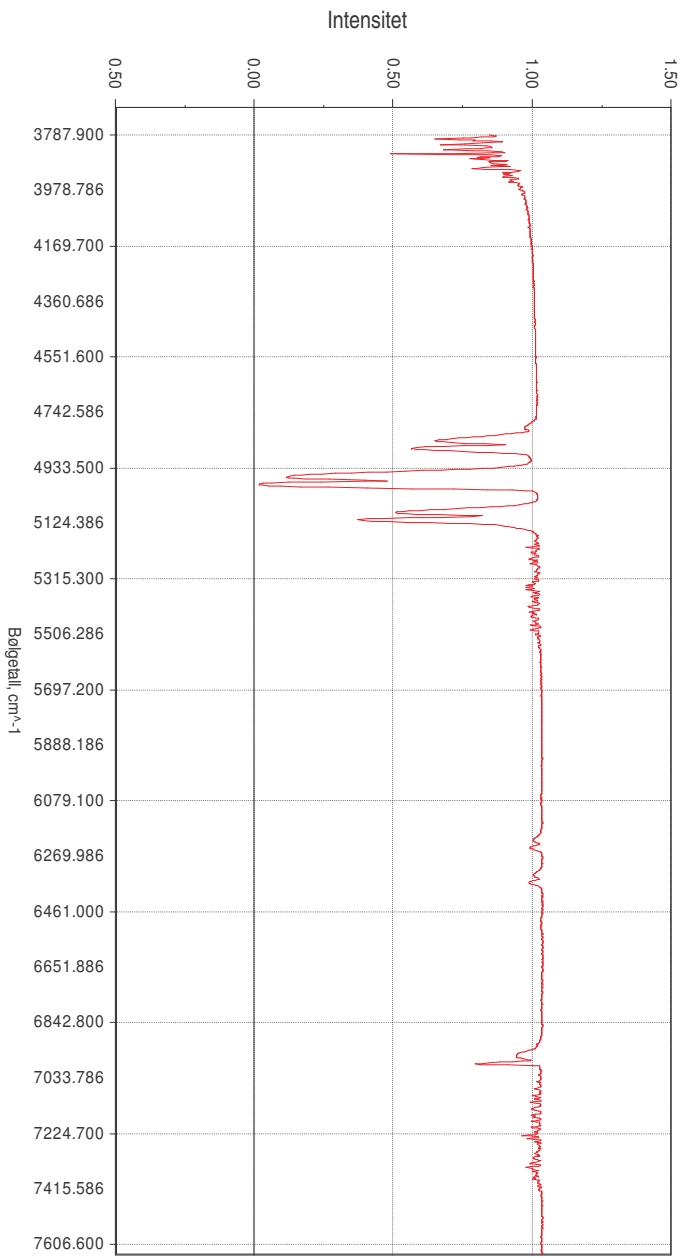
3.3.3 Derivasjon

Ved å derivere spektrene kan en fjerne additive baselinjer og drivende baselinjer. Ved første ordens derivasjon fjernes additive baselinjer og ved andre ordens derivasjon fjernes drivende baselinjer. Med additive baselinjer menes at spektrene er flyttet oppover eller nedover i forhold til hverandre.

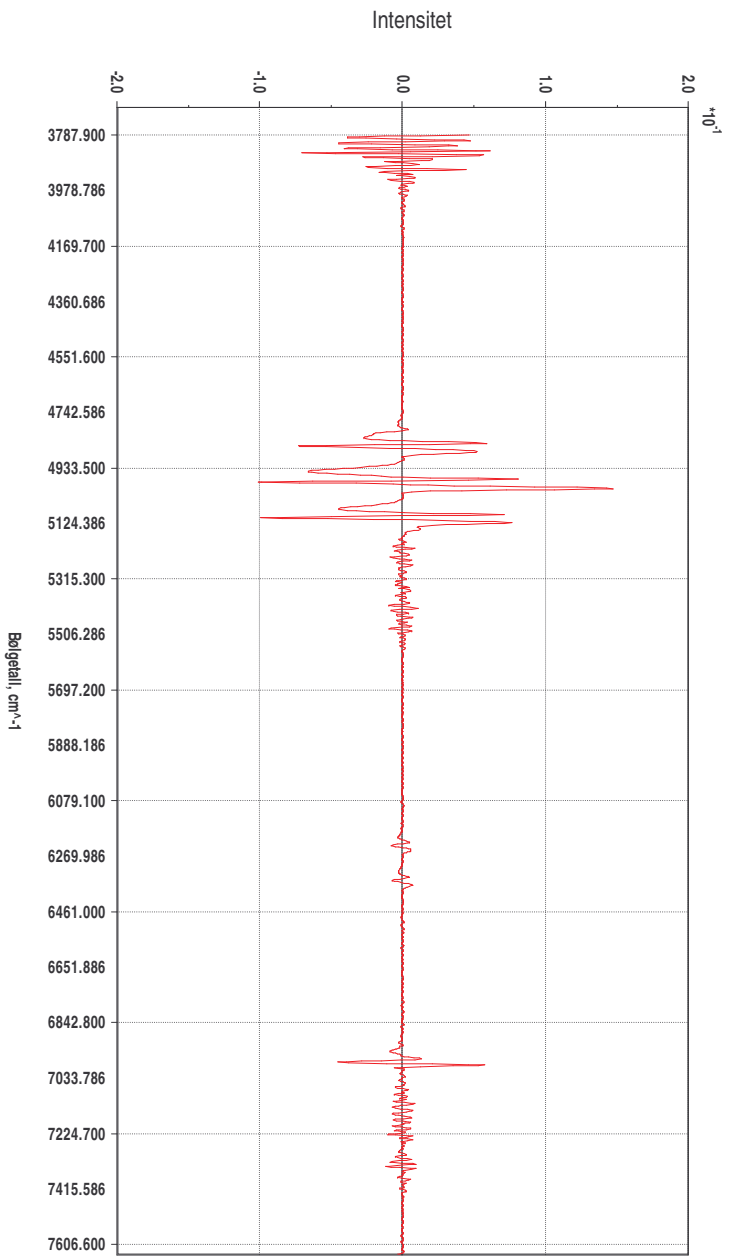
Det eksisterer flere typer derivasjon, blant annet numerisk derivasjon som Savitzky- Golay derivasjon²⁶.

Ved numerisk derivasjon deriveres punkt for punkt, og det er viktig at en har kontinuerlige data. De nye profilene kan imidlertid være vanskelige å tolke, i tillegg blir forholdet mellom signal og støy forverret ved hver derivasjon.

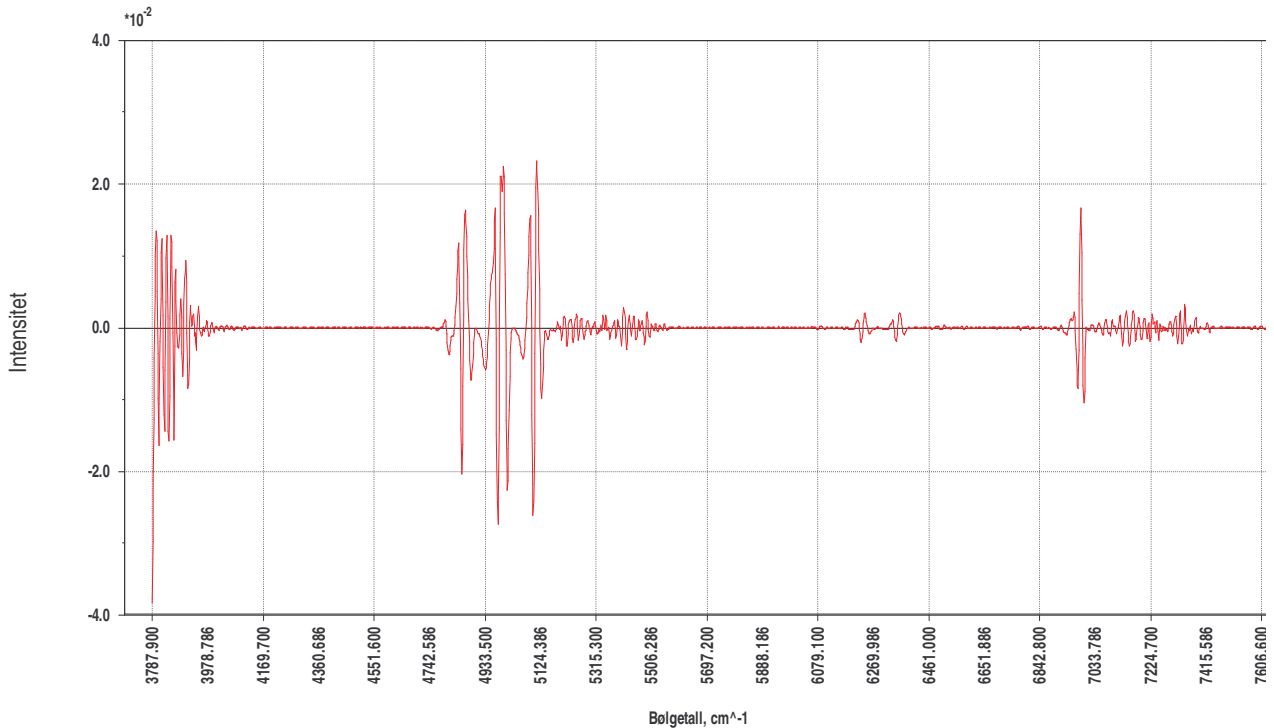
Savitzky- Golay derivasjon utfører i tillegg til derivasjon en glatting av data for å minske støyen som oppstår ved derivasjon. Ved Savitzky- Golay velges først en vindusstørrelse der et laveregrads polynom tilpasses, for eksempel ved en størrelse på 5 punkter vil polynomet tilpasses de 5 punktene. Den deriverte til polynomet regnes ut for midtpunktet i vinduet. Deretter flyttes vinduet fra første rekke eller kolonne til siste rekke eller kolonne og den deriverte til hvert punkt blir det nye signalet. Figur 3.2 viser hvordan et vanlig spekter av ren CO₂ ser ut, figur 3.3 viser første ordens derivasjon på CO₂- spekteret og figur 3.4 viser andre ordens derivasjon av CO₂- spekteret.



Figur 3.2: Spekter av ren CO_2 .



Figur 3.3: Første ordens derivasjon av figur 3.2.



Figur: 3.4: Andre ordens derivasjon av figur 3.2

3.4 Prinsippkomponentanalyse (PCA)

Når en analyserer store datasett, for eksempel spektre fra nærinfrarød spektroskopi, vil variabelrommet bli veldig stort. Dersom en har 1000 bølgelengder vil variabelrommet bli 1000 dimensjoner, noe som er umulig for mennesket å tolke. Ved å samle den relevante informasjonen i prinsipale komponenter (PC), eller latente variabler, kan en redusere antallet variabler til et nivå som det er mulig å tolke.

3.4.1 Latente variable

De ulike variablene i datasett fra spektroskopi inneholder ofte samme informasjon - de er avhengige av hverandre. Dette kalles kolinearitet, og gjør det mulig å samle ett mindre sett med variabler som forklarer all systematisk variasjon i datasettet. Disse variablene er lineære

kombinasjoner av de opprinnelige variablene, og kalles *latente variabler*. Det eksisterer mange typer dekomponeringsmetoder som gir forskjellige latente variabler, og prinsipalkomponentanalyse er en av disse. De latente variablene kalles her prinsipalkomponenter, eller PC. I utgangspunktet skal antallet latente variable for et kjemisk system være lik antallet uavhengige kjemiske komponenter, men på grunn av instrumentelle og eksperimentelle usikkerheter og støy kan antallet avvike.

3.4.2 PCA

Prinsipalkomponentanalyse²⁵ på en datamatrix \mathbf{X} kan skrives som:

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a + \mathbf{E} = \mathbf{TP}^T + \mathbf{E} \quad (3.3)$$

der A er antallet prinsipalkomponenter, \mathbf{X} er startmatrisen med dimensjonene $N \times M$, \mathbf{T} er en $N \times A$ – matrise som inneholder ortogonale skårevektorer \mathbf{t} mens \mathbf{P} er en $A \times M$ – matrise som inneholder ortogonale ladningsvektorer \mathbf{p} .

Matrisen \mathbf{E} er residualmatrisen, har samme dimensjon som \mathbf{X} og regnes som støy. De latente variablene kan hverken skrives som multiplum eller lineær kombinasjon av hverandre; de er lineært uavhengige. Datasettet inneholder ikke lenger kolinearitet. I det flervariabel rom peker de latente variablene i ulike retninger og står vinkelrett på hverandre.

Den første prinsipalkomponenten blir plassert i den retningen der det er størst variasjon. Den neste prinsipalkomponenten blir plassert ortogonalt på den forrige og i den retningen variansen er nest størst, helt frem til all variansen er forklart. De siste prinsipalkomponentene vil mest sannsynlig ligge i områder der det er liten variasjon og beskriver dermed støy i datasettet.

NIPALS- algoritmen for PCA er gitt i Appendiks A.

3.4.3 Skårer og ladninger

Hver prinsipalkomponent består av en skårevektor og en ladningsvektor.

Ladningsvektorene beskriver retningene til prinsipalkomponentene i variabelrommet, og skårevektorene inneholder koordinatene til de ulike objektene på prinsipalkomponentene.

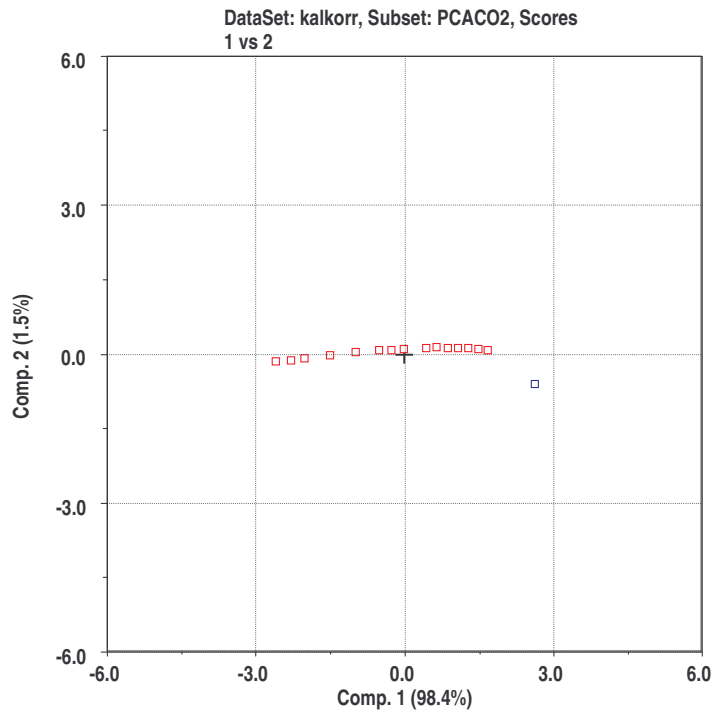
Når en utfører en prinsipalkomponentanalyse dannes det et nytt koordinatsystem der prinsipalkomponentene utgjør aksene, hvor skårevektoren har dimensjon lik antall rekker i datamatriksen og ladningsvektoren har lengde lik antall kolonner i datamatriksen.

Ved å projisere et objekt ned på en prinsipalkomponent kan en finne skåren til objektet. Skåren til objektet er avstanden fra origo til punktet. På samme måte kan en finne ladningen til forskjellige variable.

Når en projiserer variabelrommet ned i rommet utspent av prinsipalkomponentene får en et skåreplott, der en kan se på likheter mellom objekter ved å bruke prinsippene om avstand og vinkler (avsnitt 3.2).

I figur 3.5 vises et skåreplott fra et datasett som består av en serie med spektrale avren CO₂ (i rødt), og ett spekter med ren CO (i blått).

CO₂-spektrene samler seg på en linje mens CO-spekteret distanserer seg fra linjen. Det blir med andre ord et skille mellom CO₂-objekter og CO-objektet. Vinkelen mellom for eksempel CO-objektet og ett av CO₂-objektene finnes ved å trekke en rett linje fra CO₂-objektet og inn til origo, og deretter ut til CO-objektet. Vinkelen som da dannes er vinkelen mellom objektene.



Figur 3.5: Skåreplott av CO_2 - objekter og ett CO - objekt.

På samme måte som skåreplott får en ladningsplott ved å projisere objektrommet ned i rommet utstøpt av prinsipalkomponentene, hvor likheter mellom variablene kan studeres.

Ved å plote skåreverdiene og ladningsverdiene i samme figur får en et biplot. Dette plottet viser hvilke objekter som forklares av hvilke variable, og gir et bedre bilde av helheten i situasjonen.

3.4.4 RSD (Residual Standard Deviation)

Når en dekomponerer datamatriksen får en skårer, ladninger og et restledd. Dette restleddet regnes som støy, og kalles residualer. Residualvektoren for et objekt i er gitt ved:

$$\mathbf{e}_i^T = \mathbf{x}_i^T - \sum_{a=1}^A \mathbf{t}_{ia} \mathbf{p}_a^T \quad (3.4)$$

For å måle hvor god modellen er kan en benytte RSD, som er gitt ved:

$$RSD = \sqrt{\frac{\mathbf{e}_i^T \cdot \mathbf{e}_i}{M - A}} \quad (3.5)$$

der \mathbf{e}_i er residualvektoren til objekt i , M er antall variable i den opprinnelige datamatriksen og A er antall prinsipalkomponenter. RSD gir informasjon om hvert objekts avstand til modellen.

Ved å samle alle objektene sine RSD- verdier i en vektor \mathbf{s} med lengde N , for N objekter, kan en regne ut karakteristisk RSD for hele modellen:

$$RSD_{modell} = \frac{\mathbf{s}^T \cdot \mathbf{s}}{N - A - 1} \quad (3.6)$$

3.4.5 Uteliggere

En uteligger er et objekt eller en variabel som skiller seg fra resten av datasettet. I figur 3.5 ligger objektet med CO et stykke utenfor de andre objektene fra serien med CO₂, og ligner mistenkelig på en uteligger. Et annet eksempel er om en studerer et datasett med katter som objekter og forskjellige egenskaper hos katten som variable. Dersom en hund, for eksempel en Grand Danois, som er en stor hund, skulle bli registrert i datasettet ville den skilt seg kraftig ut fra kattene. Blant annet vil størrelse og kroppsspråk være ting som skiller hunden fra kattene.

Når en dekomponerer et datasett og finner latente variable vil uteliggere påvirke modellen, og det er ikke ønskelig.

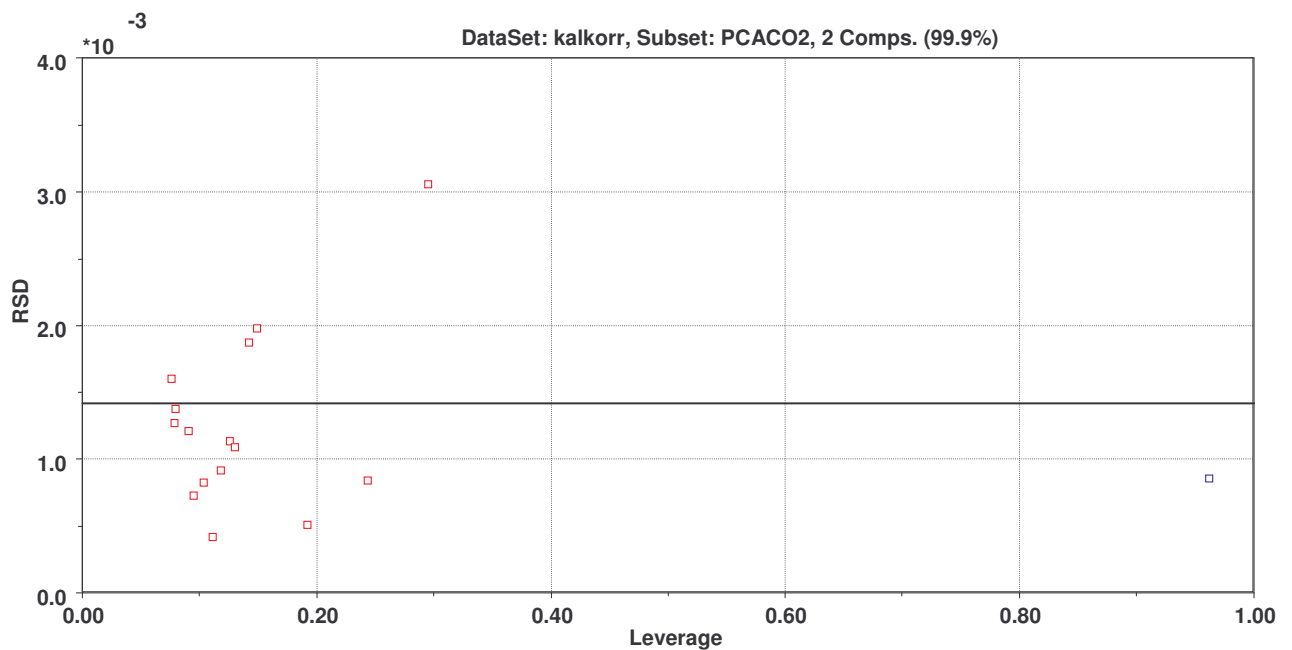
Det eksisterer flere måter å oppdage uteliggere på, blant annet ved å studere skåreplott. På den måten er det lett å oppdage tydelige uteliggere, mindre tydelige uteliggere kan detekteres ved å se på RSD- verdiene til de forskjellige objektene. En øvre grense for akseptable RSD- avvik defineres, ofte ved hjelp av formel 3.6. De objektene som har vesentlig større avvik enn grensen bør vurderes som uteliggere.

En endelig avslutning om objektene med store RSD- verdier kan taes ved å studere hvert objekts innflytelse, eller leverage på engelsk, på modellen. Et objekt har forskjellig innflytelse på modellen utifra hvor de er plassert i det flervariable rom. Et objekts innflytelse er gitt ved²⁵:

$$h_i = \frac{1}{N} + \sum_{a=1}^A \frac{t_{ia}^2}{\mathbf{t}_a^T \mathbf{t}_a} \quad (3.7)$$

der t_{ia} er skåren til objekt i på prinsipalkomponent a og \mathbf{t}_a skårevektoren til prinsipalkomponent a . Verdien til h vil alltid være mellom 0 og 1, dersom et objekt er langt fra modellsenteret vil det få leverage nær 1 mens et objekt nærmt modellsenteret vil få leverage nær 0.

Ved å plotte RSD mot leverage får en oversikt over den totale situasjonen. Små uteliggere har liten innvirkning på modellen og detekteres ved lav leverage og høy RSD. Store uteliggere har gjerne lav RSD men høy leverage, ettersom modellen må strekke seg for å kunne forklare objektet med prinsipale komponenter. Figur 3.6 viser et plott av RSD mot leverage for serien med CO₂- objekter og ett CO- objekt.



Figur 3.6: Plott av RSD mot leverage for serien med CO₂- objekter og ett CO- objekt.

Objektet med CO har høy leverage og lav RSD og er dermed en klar uteligger. Den bør utelukkes fra modellen.

Fjerning av uteliggere er uproblematisk i store datasett, men når datasettene blir små, 5 objekter og 3 variable for eksempel, bør en være varsom med fjerning av uteliggere. Disse datasettene er ustabile, og dersom en fjerner objekter vil det få store konsekvenser for de prinsipale komponentene og RSD_{modell} .

3.5 Multivariat kalibrering

Ved å uttrykke en eller flere responsvariabler y som en funksjon av et gitt antall x - variabler er det mulig å predikere responser. En prøver da å tilpasse regresjons-/kalibrerings- modellen:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{f} \quad (3.8)$$

der \mathbf{y} er responsvariabelen som vi ønsker å predikere, \mathbf{X} inneholder målte variabler, \mathbf{b} inneholder estimerte regresjonskoeffisienter og \mathbf{f} inneholder residualer fra \mathbf{y} .

3.5.1 Partial Least Squares (PLS)

Partial Least Squares^{27,28}, også kalt Projection to Latent Structures, er en av mange metoder å uttrykke en responsvariabel y som en funksjon av et antall x - variabler.

Vi trenger en vektor \mathbf{y} med responsvariablene som en ønsker å predikere, og en matrise \mathbf{X} med prediktorvariabler. De latente variablene fra PLS forklarer mest mulig av kovariansen mellom \mathbf{X} og \mathbf{y} , mens de latente variablene fra PCA forklarer mest mulig av variansen i \mathbf{X} . Kovarians er et mål på den lineære avhengigheten mellom to varierende størrelser, i dette tilfellet \mathbf{X} og \mathbf{y} .

Den enkleste måten å forstå PLS er ved å forestille seg at en foretar en PCA- analyse på \mathbf{X} og en på \mathbf{y} , og deretter roterer to og to latente variabler (en fra \mathbf{X} og en fra \mathbf{y}) mot hverandre til korrelasjonen blir maksimal, det vil si når vinkelen mellom skårevektoren i \mathbf{X} - rommet og skårevektoren i \mathbf{y} - rommet er 0. Da blir sammenhengen mellom skårene:

$$\mathbf{u}_a = \mathbf{b}_a \mathbf{t}_a \quad (3.9)$$

der \mathbf{u} er skårevektor i \mathbf{y} - rommet, \mathbf{t} er skårevektor i \mathbf{X} - rommet og \mathbf{b} er indre relasjons regresjonskoeffisienten.

Dekomponeringen som PLS utfører skjer i ett steg, samtidig på \mathbf{X} og \mathbf{y} ²⁹. For hvert par av latente variable søkes en vektor \mathbf{w} , kalt ladningsvektor, som skal være et kompromiss mellom den latente variabelen i \mathbf{y} - rommet og den latente variabelen i \mathbf{X} - rommet. PLS finner en vektor om gangen, den mest signifikante blir først funnet, deretter trekkes den informasjonen som blir forklart av den faktoren fra \mathbf{X} - og \mathbf{y} - matrisene. Deretter finnes neste faktor.

En ender opp med en modell der regresjonsvektoren er gitt ved:

$$\mathbf{B} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{Q}^T \quad (3.10)$$

der \mathbf{W} er ladningsvektene, \mathbf{P} er ladninger i \mathbf{X} - rommet og \mathbf{Q} er ladninger i \mathbf{y} - rommet.

For spektroskopiske data, der konsentrasjonen er kjent, vil spektrene med høyere konsentrasjon bli vektet sterkere enn spektrene med lav konsentrasjon, ettersom PLS bruker konsentrasjonsinformasjonen under dekomponeringsprosessen.

Det eksisterer flere typer PLS, blant annet PLS-1 og PLS-2. I PLS- 1 blir en og en respons modellert, i PLS- 2 blir alle responsene modellert samtidig. PLS- 2 gir vanligvis litt dårligere prediksjoner enn PLS-1.

Algoritmen for PLS-1, som er benyttet i denne oppgaven, er gitt i Appendiks A.

3.5.2 Mål for avvik

Det eksisterer flere mål på feil for multivariate kalibreringsmodeller. Noen av dem er $RMSEC^{30}$, $RMSEP^{30}$ og $PRESS^{31}$:

$RMSEC$ (Root Mean Squared Error of Calibration) er et mål på gjennomsnittlig differanse mellom predikert og målt verdi når en bruker modellen til predikere responsen til prøvene i treningssettet, med andre ord prediksjonsfeilen til de objektene som er benyttet i modellen. Det er derfor et mål for modellfeil, og regnes ut slik:

$$RMSEC = \sqrt{\left(\frac{\sum_{i=1}^n (\hat{y}_{i,cal} - y_{i,cal})^2}{n} \right)} \quad (3.11)$$

$RMSEP$ (Root Mean Squared Error of Prediction) er et mål på gjennomsnittlig differanse mellom predikert og målt verdi når en bruker modellen på prøver i et valideringssett, det vil si nye prøver som ikke er med i modellen. Feilen en får her er det en kan forvente seg å få når en tilpasser nye prøver til modellen, gitt at prøvene er innenfor området modellen gjelder. $RMSEP$ regnes ut som følger:

$$RMSEP = \sqrt{\left(\frac{\sum_{i=1}^n (\hat{y}_{i,cal} - y_{i,val})^2}{n} \right)} \quad (3.12)$$

$PRESS$ (Predicted Residual Sum of Squares) blir vanligvis brukt om den prediksjonsfeilen en regner ut når en utfører kryssvalidering. $PRESS$ er kvadratsummen av alle responsresidualene når en tester modellen på et sett med prøver som ikke er med i modellen. $PRESS$ regnes ut som følger:

$$PRESS = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3.13)$$

3.5.3 Intern validering/kryssvalidering

En intern metode for validering heter kryssvalidering. Den bruker objektene i treningssettet, og bygger på at data anvendes mange ganger i like modeller.

I kryssvalidering deles data inn i grupper. Deretter holdes den første gruppen utenfor, og modellen bygges på de resterende data. Etterpå brukes modellen til å predikere objektene i gruppen som ble holdt utenfor, og avvikene mellom den predikerte og observerte verdien regnes ut. Kvadratene av avvikene summeres til PRESS. Deretter holdes gruppe nummer 2 utenfor og prosedyren gjentas, deretter gruppe 3 og så videre inntil alle objektene har blitt holdt utenfor en gang og PRESS inneholder bidrag fra alle objektene.

Ved å benytte kryssvalidering kan en finne ut hvor stor del av variasjonen i y som er predikterbar.

3.5.4 Ekstern validering

Ved å lage modeller av hele treningssettet og benytte et helt nytt test/valideringssett for å finne prediksjonsfeil, validerer en modellene eksternt. På denne måten finner en den forventede prediksjonsfeilen for nye prøver, med andre ord hvor gode resultater en kan forvente seg.

Det er viktig at treningssett og valideringssett er representative prøver; at de ikke inneholder uteliggere, har samme konsentrasjonsområde og har generelt like forhold ved prøvetaking som for fremtidig prøver. Dess større datasett en bruker for å validere modellen, dess bedre totalinntrykk av den fremtidige prediksjonen får en.

Det eksisterer en konflikt mellom hva som er best av intern og ekstern validering: Ved å benytte intern validering på flest mulig objekter, ved for eksempel å benytte objekter som var tiltenkt ekstern validering, blir ikke modellene robust nok til å ta høyde for fremtidige objekter. Ved ekstern validering kan en gjøre det motsatte; å benytte objekter ment for intern

validering til ekstern validering. Det totale resultatet blir det samme; modellene blir ikke robuste nok til å predikere fremtidige objekter.

3.6 Instrumentstandardisering

Når en har flere instrumenter, som i hovedsak er like, og som skal benyttes til å måle det samme, kan det være tidsparende å overføre en kalibreringsmodell bygget på ett av instrumentene over til det andre instrumentet. Instrumenter er aldri helt identiske, derfor vil en direkte overføring sjelden være tilfredsstillende. En trenger derfor metoder for å standardisere instrumentene.

Generelt faller metodene inn i to klasser:

- 1) Teknikker som prøver å predikere responsen av hovedinstrumentet basert på responsen til det nye instrumentet.
- 2) Teknikker som eliminerer forskjellen mellom instrumentene, slik at bare delene som instrumentene har felles er igjen.

I det første tilfellet kjøres data fra begge instrumentene gjennom en kalibreringsmodell som er bygget på hovedinstrumentet. I det andre tilfellet bygges en modell på transformerte data fra begge instrumenter.

En må i begge tilfeller ha en serie med like prøver som er kjørt på begge instrumenter. Disse prøvene bør helst være prøver som har størst effekt på modellene, med andre ord objekter med høy leverage.

3.6.1 Piece- wise Direct Standardization (PDS)

PDS (Piece- wise Direct Standardization)^{32,33,34} er blant den første typen standardiseringsmetoder. Den lager "lokale" lineære modeller som relaterer responsen av

instrumentet som skal standardiseres over en rekke frekvenser til responsen av hovedinstrumentet ved en enkelt frekvens.

Kalibrasjonsmodellen er formulert som følger:

$$\mathbf{y} = \mathbf{X}_1 \mathbf{b} + \mathbf{1} \beta_1 \quad (3.14)$$

der \mathbf{y} er konsentrasjonsvektoren for den aktuelle analytten, \mathbf{X}_1 er x- responsmatrisen, \mathbf{b} er regresjonsvektoren, β_1 er offset fra modellen og $\mathbf{1}$ er en $m \times 1$ - vektor med ett- tall, hvor m er antallet prøver. Etter sentrering:

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}_1 \mathbf{b} \quad (3.15)$$

Forholdet mellom instrumentene kan modelleres som:

$$S_1 = S_2 \mathbf{F}_b + \mathbf{1} \mathbf{b}_s^T \quad (3.16)$$

der S_1 er x- responsen fra instrument 1, S_2 er x- responsen fra instrument 2, \mathbf{F}_b er transformasjonsmatrisen, \mathbf{b}_s er bakgrunnskorreksjonen, $\mathbf{1}$ er en $p \times 1$ - vektor med ett- tall, der p er antallet prøver i S_1 og S_2 . Transformasjonsmatrisen \mathbf{F}_b kan dermed kalkuleres for å tilfredsstill

$$\bar{S}_1 = \bar{S}_2 \mathbf{F}_2 \quad (3.17)$$

Transformasjonsmatrisen benyttes deretter til å standardisere spektrene fra instrument 2, slik at modellene som er bygget på spektre tatt opp på hovedinstrumentet fungerer bedre. For alle nye spektre som stammer fra instrument 2 må denne standardiseringen utføres.

Denne standardiseringen blir utført med funksjoner fra PLS Toolbox 4.1³⁵ i Matlab R2006b³⁶, fremgangsmåten i Matlab er gitt i PLS Manual fra PLS Toolbox 4.1.

3.6.2 Generalized Least Squares Weighting (GLSW)

GLSW (Generalized Least Squares Weighting) er en standardiseringsmetode som fjerner variasjon som instrumentene ikke har felles. Det gjør den ved å kalkulere en filtermatrise basert på forskjeller mellom par eller grupper av prøver som i utgangspunktet skulle vært like. Fremgangsmåten er som følger³⁷: Estimer offset- termen som differansen mellom gjennomsnittet av instrumentene:

$$\bar{\mathbf{x}}_d = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 \quad (3.18)$$

Den sentrerte differansen mellom prøvene er dermed:

$$\mathbf{X}_d = \mathbf{X}_1 - \mathbf{X}_2 + \mathbf{1}\mathbf{x}_d \quad (3.19)$$

der $\mathbf{1}$ er en vektor med ett- tall. Kovariansen til differansen blir estimert på vanlig måte men en liten verdi α blir lagt til diagonalen:

$$\mathbf{C}_d = \frac{\mathbf{X}_d^T \mathbf{X}_d}{(m-1)} + \alpha \mathbf{I} \quad (3.20)$$

der m er antallet prøver. Konstanten α legges til for å sikre at kovariansmatrisen har full rang. \mathbf{C}_d blir dekomponert ved å bruke en singularverdi- dekomponering:

$$\mathbf{USV}^T = \mathbf{C}_d \quad (3.21)$$

Diagonalene av \mathbf{S} , "Eigenvalues", justeres til å ha gjennomsnitt lik 1, og kvadratroten tas for å lage matrise av justerte singularverdier:

$$\mathbf{S}_{adj} = \sqrt{\frac{\mathbf{S} * n}{\text{trace}(\mathbf{S})}} \quad (3.22)$$

der n er antall variable. Vektings- matrisen lages ved:

$$\mathbf{W} = \mathbf{V}\mathbf{S}_{adj}^+ \mathbf{V}^T \quad (3.23)$$

der \mathbf{S}_{adj}^+ er pseudo- inversen av \mathbf{S}_{adj} , og singularverdier nær 0 har ikke blitt invertert.

Instrumentene kan nå standardiseres: Det sentrerte hovedinstrumentet kan bli preprossesert ved å multiplisere med vekten \mathbf{W} :

$$\mathbf{X}_{1adj} = (\mathbf{X}_1 - 1\mathbf{x}_1)\mathbf{W} \quad (3.24)$$

Dermed kan regresjonsmodeller bygges, og kan benyttes på det justerte instrument 2:

$$\mathbf{X}_{2adj} = (\mathbf{X}_2 - 1\mathbf{x}_1 + 1\mathbf{x}_d)\mathbf{W} \quad (3.25)$$

Fremgangsmåten i Matlab er gitt i PLS Manual fra PLS Toolbox 4.1.

I denne oppgaven er både PDS og GLSW testet, men det har ikke blitt tid til å utføre GLSW-standardiseringen på riktig måte.

4. EKSPERIMENTELT

Den praktiske delen av denne oppgaven er utført hos Christian Michelsen Research på Fantoft i Bergen. To måleceller for nærinfrarød spektroskopi har blitt bygget der, og har blitt benyttet i denne oppgaven. Spektre taes opp på gassblandinger som er levert av Yara, som skal simulere utslipp fra brenselcellen. Ved deretter å bruke multivariat dataanalyse/kjemometri skal det undersøkes hvor bra en kan predikere en eller flere responser på bakgrunn av et spekter.

Området som benyttes er i bølgetall fra 3787 cm^{-1} til 7641 cm^{-1} .

4.1 Gassblandinger

Gassblandingene som er benyttet i denne oppgaven er som følger:

- 100% CO_2
- 1.0% CO_2 i N_2
- 1.01% CH_4 i N_2
- 1.01% CO i N_2
- 9.92% CH_4 i CO_2
- 9.99% CO i CO_2
- 0.94% CO + 1.03% CH_4 i N_2
- 1.0% CO + 10.03% CH_4 i N_2

4.2 Spekteropptak

Inne i målecellene er det montert en trykkmåler, slik at en hele tiden kan overvåke hva trykket er i målecellen. Gass blir overført til målecellen frem til ønsket trykk er oppnådd, deretter stenges målecellen igjen, både inngang og utgang, og spekter taes opp.

Bakgrunnsspektre blir tatt opp på samme dag som en eller flere serier med spektre av gassblandinger blir tatt opp. Det blir tatt opp 8 bakgrunnsspektre for hver dag spektre taes opp, med ca. samme overtrykk av N_2 som for de andre spektrene som taes opp. For spektrene tatt opp ved høy temperatur blir det tatt opp bakgrunnsspektre som har samme temperatur og trykk som de andre spektrene som blir tatt opp.

Mellom hvert spekteropptak blir målecellen skylt med ren N_2 først, og deretter med gassblandingen som skal måles.

4.2.1 Modellspektre

Først taes det opp spektre som skal brukes til å bygge modeller. De ble tatt opp i september 2006, og bestod av 15 spektre fra hver av gassblandingene. Trykkintervallet som ble observert av trykkmåleren gikk fra ca. 4.6 bar til ca 1 bar, det vil si fra ca 3.6 bar overtrykk til 0 overtrykk. Først fylte en cellen til måleren viste ca. 4.6 bar og stengte cellen, deretter ble spekter tatt opp. Cellen ble deretter tømt for enten 0.2 eller 0.4 bar, stengt igjen og nytt spekter ble tatt opp. Dette ble gjentatt til en hadde 15 spektre. Samme fremgangsmåte ble gjentatt for hver gassblanding, slik at trykket i første spekter i en serie var tilnærmet likt trykket i første spekter for en annen serie, det samme for spekter to og så videre.

Dette ga totalt 120 spektre.

4.2.2 Valideringsspektre

I oktober 2006 ble valideringsspektrene tatt opp. Samme prosedyre som september ble benyttet, men istedenfor 15 spektre tok en opp to serier med 8 spektre for hver gassblanding. I tillegg ble to spektre tatt opp med flyt, det vil si at målecellen var åpen i begge ender mens gassen fløt fra gassbeholderen og gjennom målecellen og ut, mens spektre ble tatt opp.

Dette ga totalt 129 spektre.

4.2.3 Strømningspektre

I oktober 2006 ble som sagt to spektre tatt opp med strømming. Dette ble utprøvd for å teste strømnings- effekter i analysedelen, for å se om det lar seg gjøre med online- måling av spektre. I januar 2007 ble en serie med spektre av CO₂ tatt opp der tre av spektrene ble tatt opp med gasstrømming, og i april 2007 ble fire serier med spektre tatt opp, i hver serie var tre av spektrene tatt opp med strømming. Disse seriene bestod av gassblandingene 9.99% CO i CO₂, 1.0% CO₂ i N₂, 100% CO₂ og 0.94% CO + 1.03% CH₄ i N₂.

Dette ga totalt 55 spektre.

4.2.4 Høytemperaturspektre

Avgassene som kommer ut av brensecellene vil holde høy temperatur, og det er derfor nødvendig å undersøke temperaturens effekt på spektrene. Avgassene skal kjøles ned slik at de holder max 70° Celsius når spektre taes opp. I juni og juli 2007 ble det derfor tatt opp tre serier med spektre, av gassblandingene 0.94% CO + 1.03% CH₄ i N₂, 9.99% CO i CO₂ og 100% CO₂. Dette ble gjort ved å legge målecellen i en stor varmeovn. Varmeovnen er utstyrt med et termometer som er plassert høyt i ovnene, og et termometer som er plassert under en hylle i ovnene, hvor målecellen er plassert oppå. Varmeovnen har i tillegg en kjølefunksjon.

En fyller målecellen inntil totaltrykket er ca. 4.6 bar med den ene gassblandingen, lukker cellen og tar opp spekter ved romtemperatur. Deretter økes temperaturen til 45° Celsius. Når temperaturen til termometeret som er plassert under hyllen i ovnen er nådd denne temperaturen ventes det i 10 minutter før spekter taes opp. Deretter økes temperaturen til 60° Celsius, og det samme gjentas. Når spekteret er tatt opp, tømmes litt av gassen ut inntil totaltrykket i cellen er ca. 3 bar. Deretter taes det opp spekter, temperaturen settes til 45° Celsius, ovnen kjøles ned, og spekter taes opp 10 minutter etter at temperaturen til termometeret som er plassert under hyllen har nådd 45° Celsius. Dermed settes temperaturen til ca 20° Celsius og samme prosedyre gjentas. All gassen tømmes deretter ut av cellen, slik at det er 0 overtrykk av gassblandingen, og spektre ved romtemperatur, 45° Celsius og 60° Celsius taes opp.

Samme prosedyre utføres for de to andre gassblandingene.

Årsaken til at en valgte å vente med å ta opp spekter ti minutter etter temperaturen til termometeret under hyllen i ovnen hadde nådd ønsket temperatur, var at en etter ti minutter ikke observerte noe særlig økning i trykkmåleren som er plassert inne i målecellen, og dermed antok en at temperaturen inne i målecellen var tilnærmet lik temperaturen til termometeret som var plassert under hyllen i ovnen.

Dette ga totalt 32 spektre.

4.2.5 Høytemperaturvalidering

Et tema som kom opp under og etter opptaket av høytemperaturspektre var om en hadde ventet lenge nok med å ta opp spektre, med andre ord; var temperaturen i målecellen lavere enn antatt for spektrene tatt opp ved høy temperatur på grunn av at spektrene ble tatt opp for tidlig?

Et annet tema var om gassen som kommer ut av gassbeholderene er kaldere enn romtemperatur.

Dette undersøkes ved å fylle CO₂ i målecellen til ca. 3 bar, ta opp spekter, og deretter ta opp nytt spekter hvert femte minutt inntil 40 minutter er gått. På den måten vil en se både på spektrene og på trykkmåleren hvordan det utvikler seg.

Deretter økes temperaturen til 45° Celsius, og en tar opp spekter hvert femte minutt inntil 50 minutter er gått. En har nå ventet i 35 minutter etter at temperaturen til termometeret under hyllen i ovnen har nådd ønsket temperatur, det burde være tilstrekkelig. Temperaturen økes til 60° Celsius, og spekter taes opp hvert femte minutt inntil 45 minutter er gått. Igjen har en ventet i 35 minutter etter at temperaturen har nådd ønsket verdi.

Dette ga totalt 28 spektre.

4.2.6 Celle 2- spektre

Spektrene som til nå er tatt opp er tatt opp på en og samme celle (hovedcellen), kalt celle 1. Det skal benyttes minst to måleceller i ”kraftanlegget” i laboratorieskala, derfor må celle nummer 2 også kalibreres. Noenlunde samme prosedyre som for valideringsspektrene benyttes; en serie med 8 spektre taes opp for alle gassblandingene.

Dette ga totalt 64 spektre.

5. RESULTATER

Formålet med oppgaven er å undersøke om det lar seg gjøre å predikere responsene CH₄, CO₂ og CO ved hjelp av NIR- spektre. Noen av de potensielle problemene som har blitt undersøkt i denne oppgaven er:

- *Bakgrunnseffekter:* Endres bakgrunnsspektrene vesentlig over tid? Hvilken effekt har forskjell i gasstrykk mellom bakgrunn og spekteret som korrigeres?
- *Strømningseffekter:* Vil prediksjoner av spektre tatt opp med strømmende gass være tilfredstillende og kan slike spektre korrigeres med bakgrunns-spektre tatt opp med stillestående gass?
- *Modellbygging:* Vil absorpsjonsbåndene fra vanndamp overskygge signalene til CO₂, CH₄ og CO i det nær- infrarøde området?
- *Temperatureffekter:* Vil toppintensiteter, båndbredder og topp- posisjoner endres vesentlig ved økt temperatur? Hvilke effekter vil det ha på modellering og prediksjon?
- *Kalibreringsoverføring:* Vil forskjellene mellom målecellene være store? Er det behov for instrumentstandardisering?

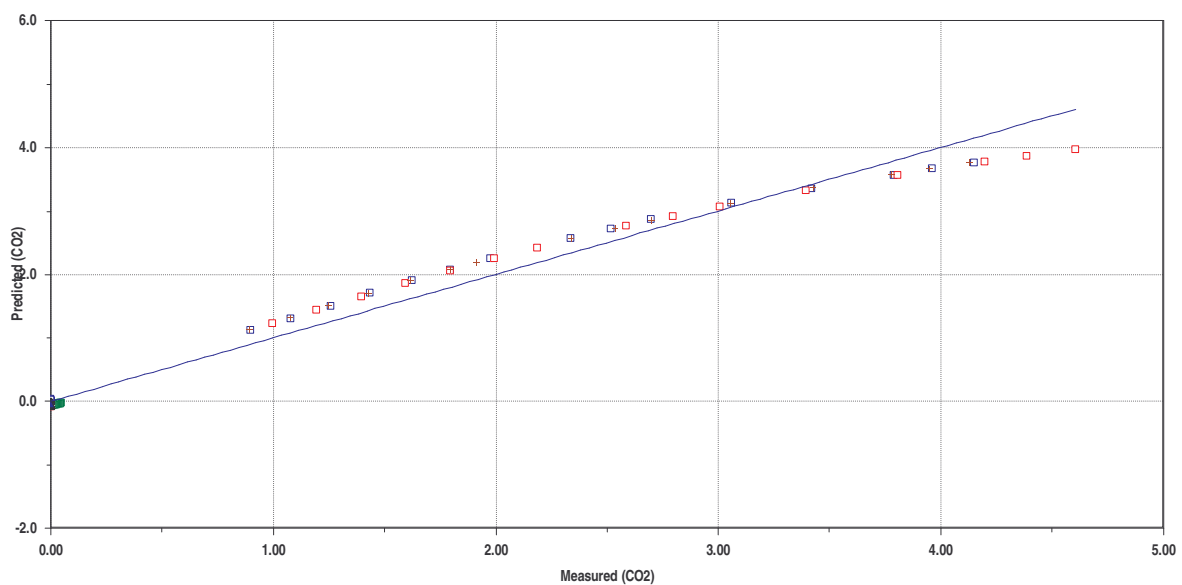
I denne delen er MatLab R2006b³⁵ benyttet for bakgrunnskorreksjon og instrumentstandardisering, resten av dataanalysen er gjort med Sirius 8.0³⁸.

I det følgende undersøkes effekten av bakgrunn, strømning, vann sin effekt på modellbygging, temperatureffekter og kalibreringsoverføring i den rekkefølgen.

Betydningen av de forskjellige effektene er testet ved å undersøke hvordan prediksjonsevnen til PLS- modeller varierer.

5.1 Modellbygging

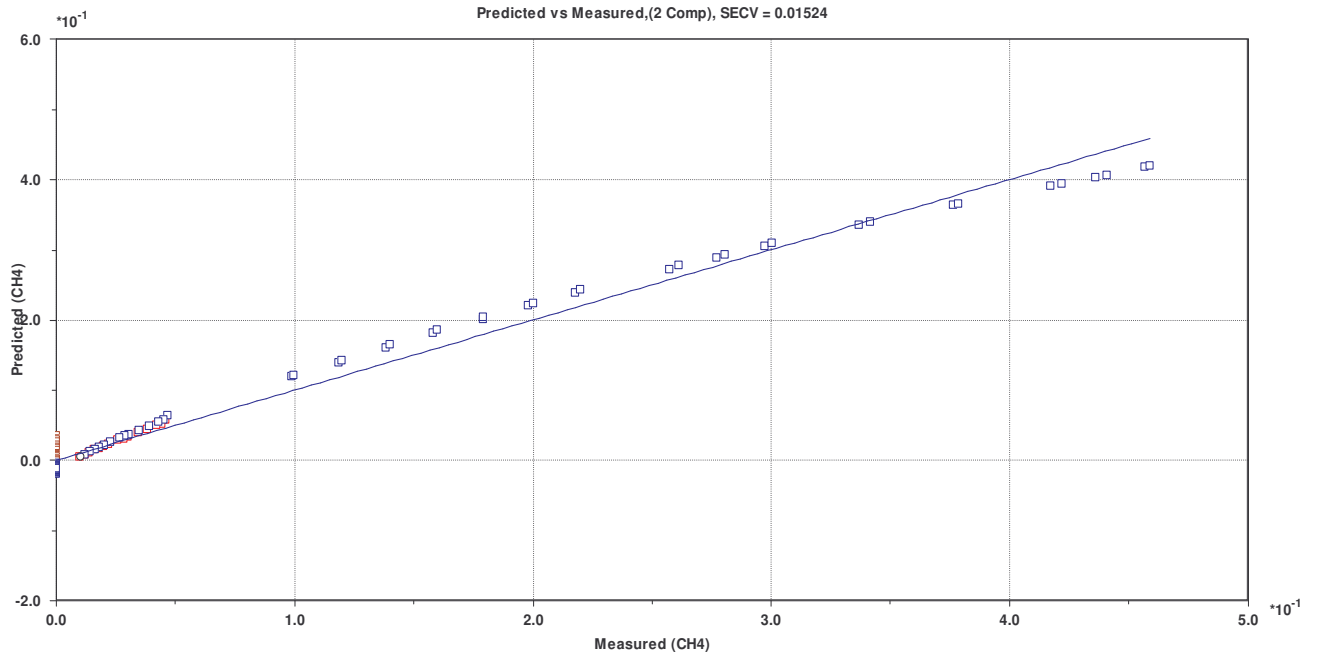
Spektre tatt opp i september 2006 benyttes til å bygge kalibreringsmodeller. Først undersøkes det om det er nødvendig å dele opp modellene i høy- og lav- konsentrasjonsområder, ettersom gassene varierer over et stort konsentrasjons- område. CO_2 har et sprang fra 1% til opptil 100%, og for CH_4 og CO er det et sprang fra 1% til 10%. Dersom en prøver å lage en PLS-modell av alle variabler og objekter fra modellspektrene for CO_2 , får en et plott av predikert mot målt verdi som i figur 5.1:



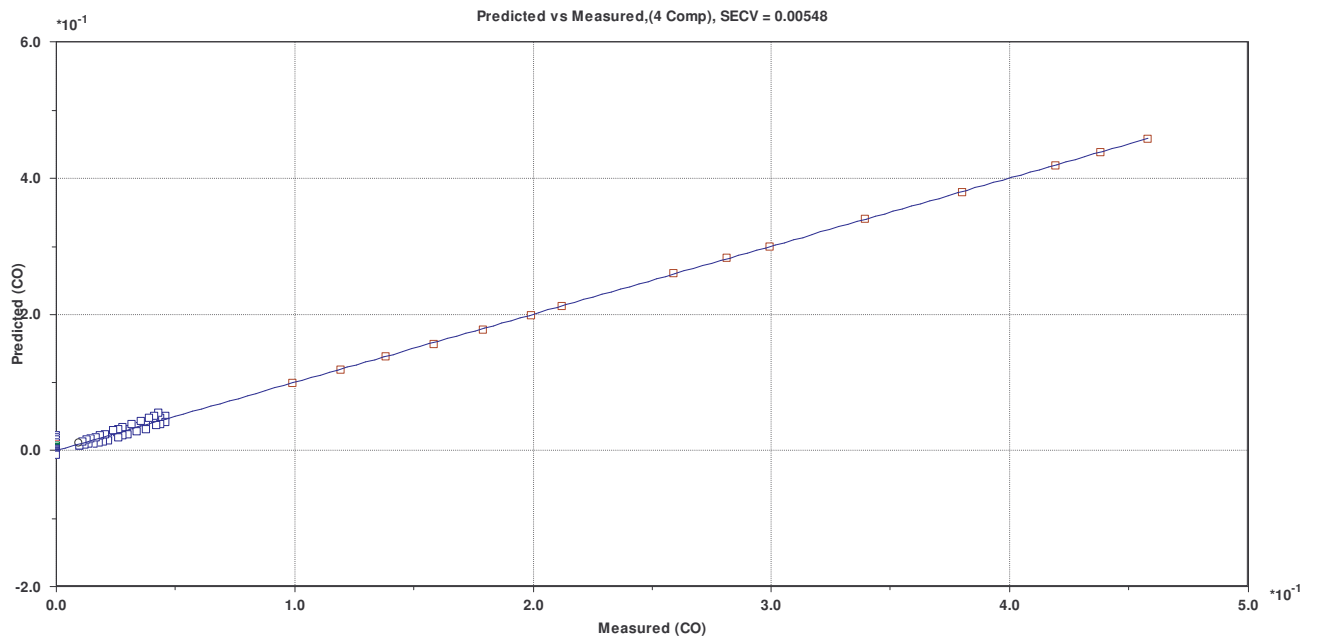
Figur 5.1: Predikert verdi mot målt verdi for CO_2 .

Alle punktene skulle optimalt dannet en rett linje, men det er en tydelig krummet linje som blir dannet. For CO_2 bør det derfor deles opp i en høy- og en lav- konsentrasjonsmodell. For CH_4 ser det likt ut, figur 5.2, mens for CO er det ikke noen klar krumning, figur 5.3.

Resultater



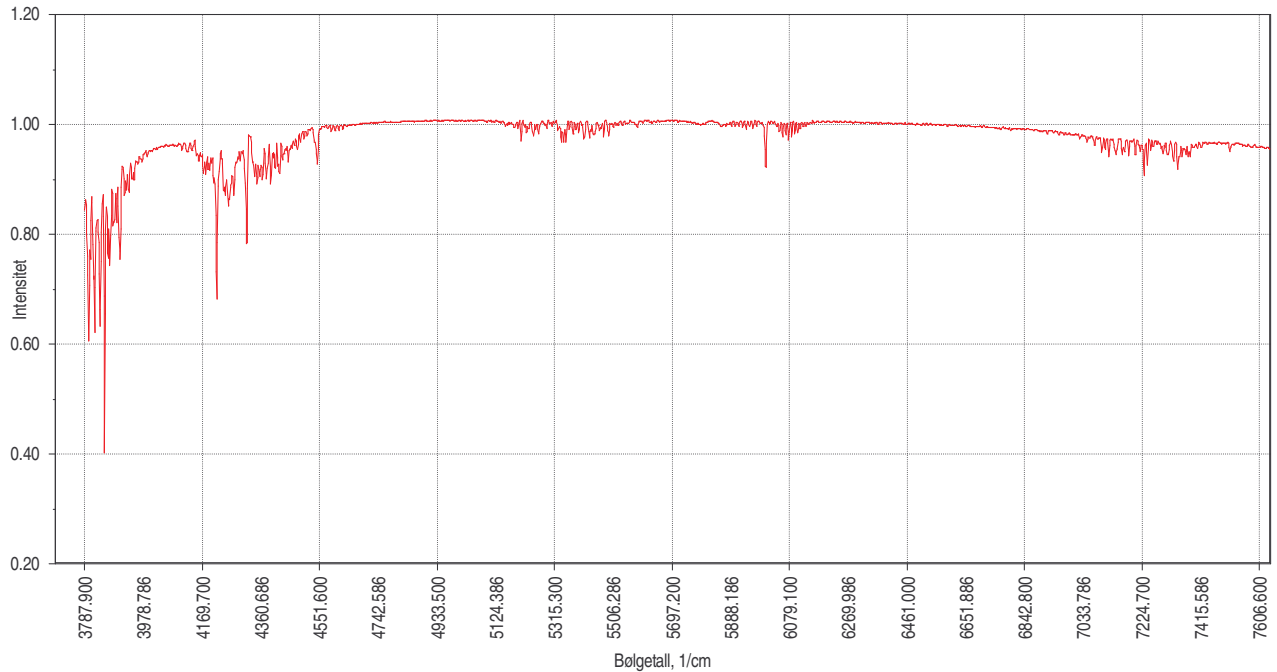
Figur 5.2: Predikert verdi mot målt verdi for CH₄.



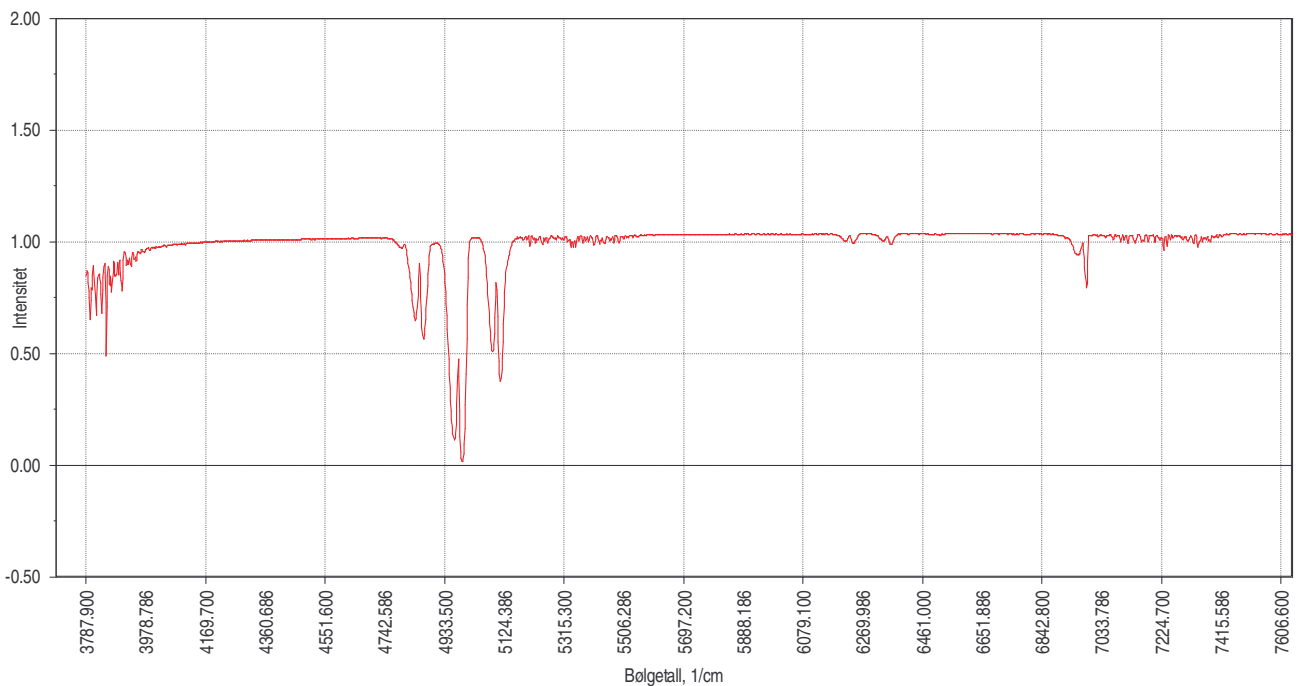
Figur 5.3: Predikert verdi mot målt verdi for CO.

For å systematisk undersøke en eventuell effekt av store sprang i konsentrasjon ble det besluttet å bygge to modeller (lav og høy konsentrasjon) for hver av stoffene.

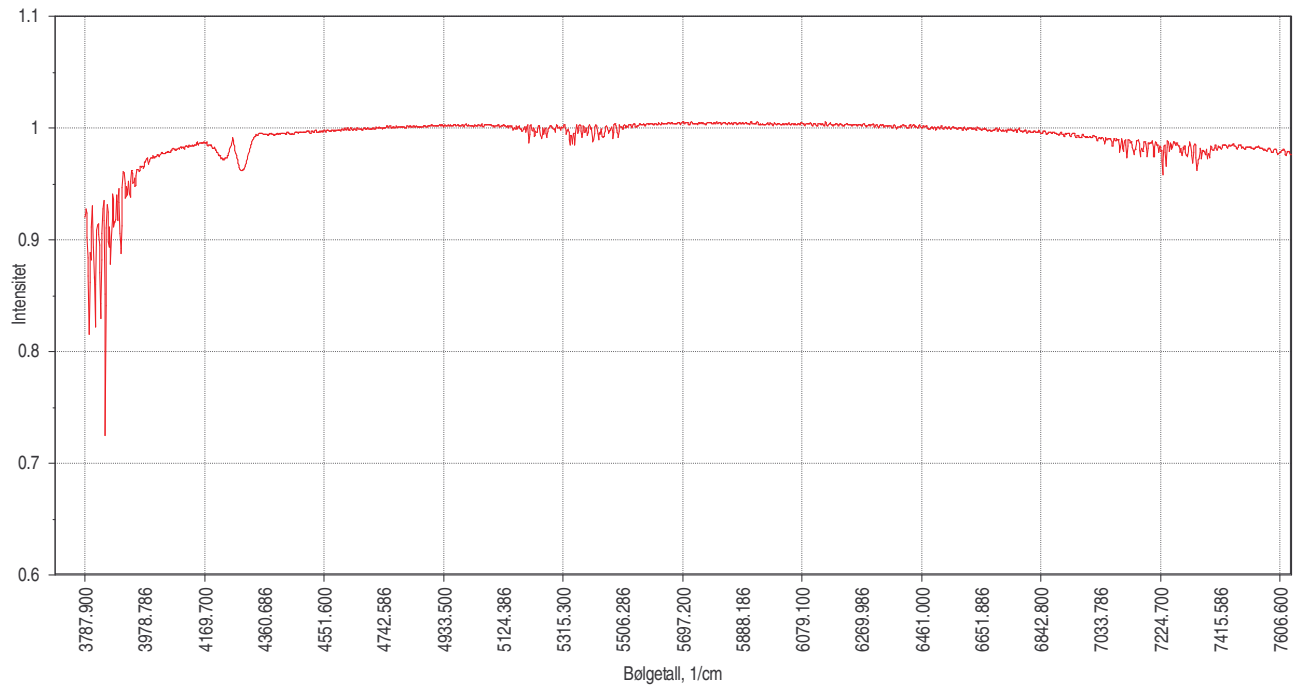
Deretter undersøkes det for hver respons, CO₂, CH₄ og CO, hvilke bølgelengder som er aktuelle å benytte i modellene. Rene spektrere av hver av responsene undersøkes; figur 5.4 viser et rent spekter av CH₄, figur 5.5 viser et rent spekter av CO₂ og figur 5.6 viser et rent spekter av CO.



Figur 5.4: Rent spekter av CH₄; CH₄ i N₂.



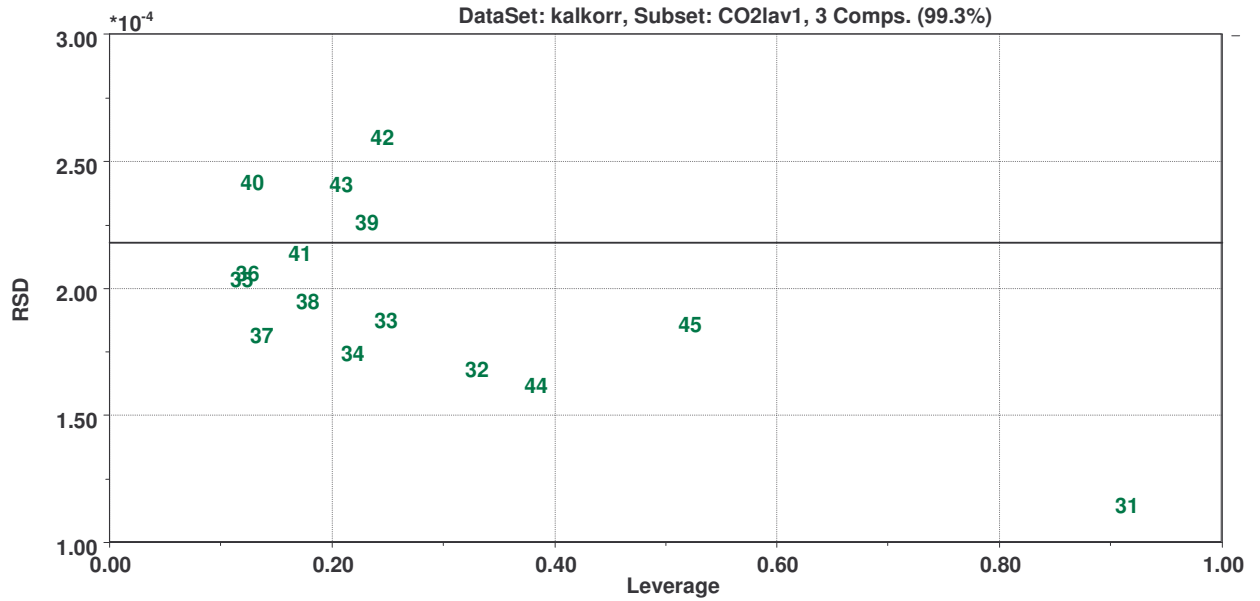
Figur 5.5: Rent spekter av CO₂.



Figur 5.6: Rent spekter av CO; CO i N₂.

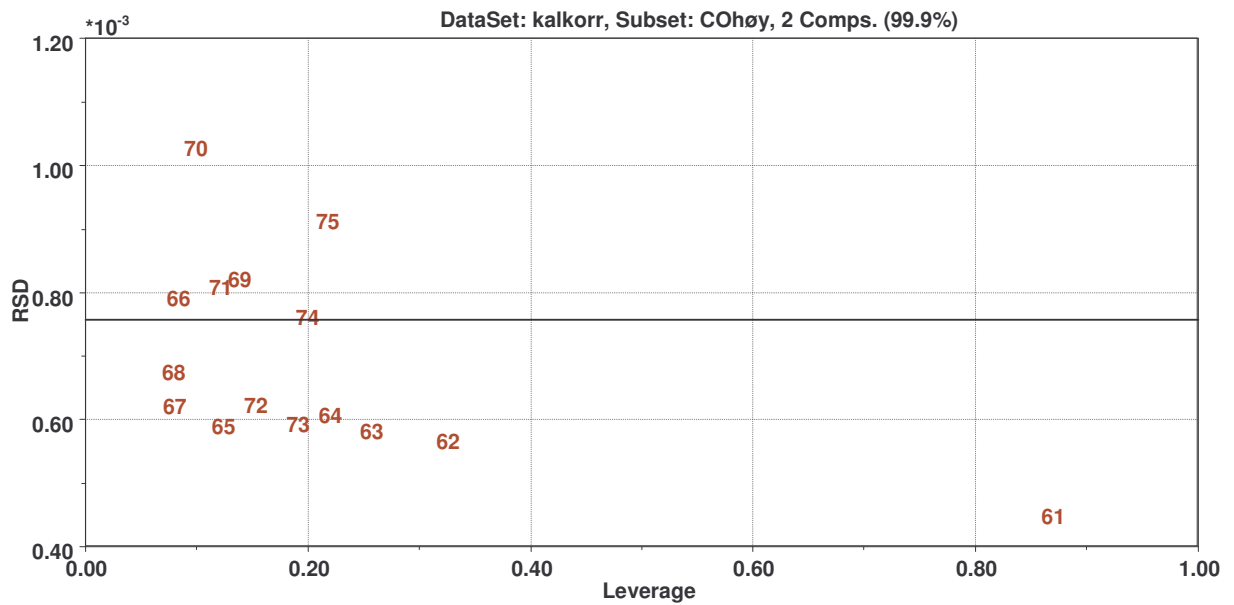
For hver av modellene benyttes alle objektene som inneholder responser som er i det området modellen gjelder. Det vil si, i modell CH₄- høy er det bare objekter som har høy konsentrasjon av CH₄, og tilsvarende for de andre modellene.

Ved å benytte PCA på modellene søkes det etter uteliggere i modellene. For CO₂- lav detekteres en klar uteligger; nederst til høyre i figur 5.7 Det objektet fjernes fra modellen.



Figur 5.7: Lav RSD og høy leverage; klar uteligger

Samme situasjon er det for CO- høy, figur 5.8.



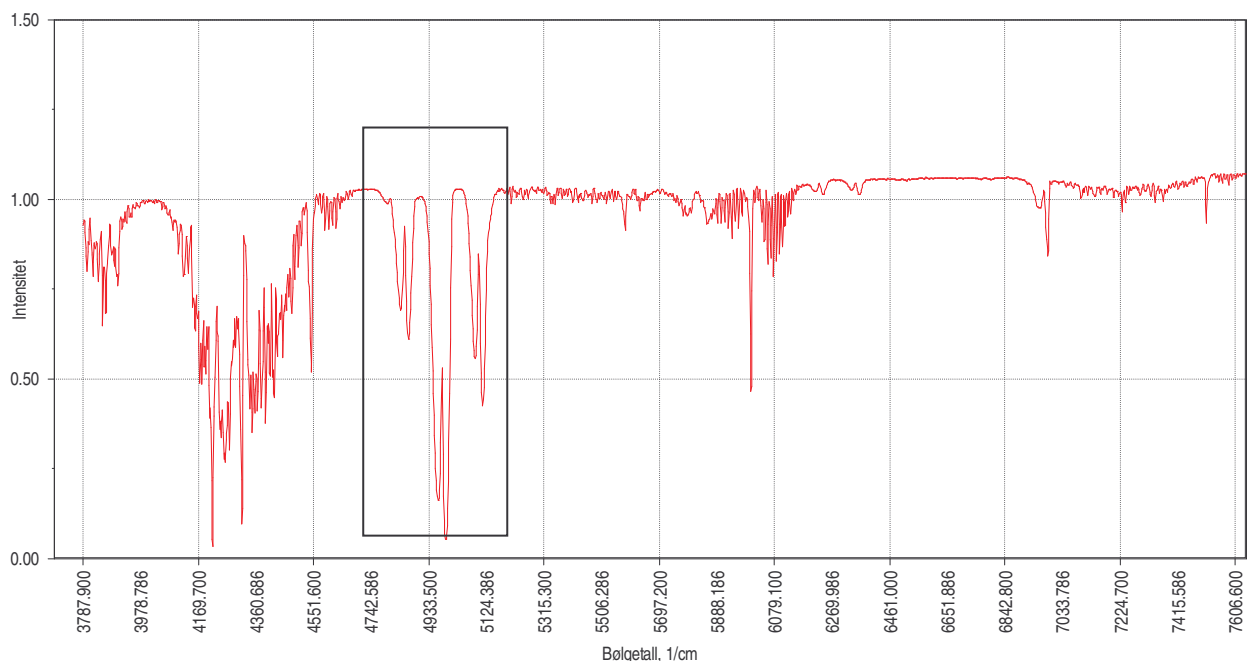
Figur 5.8: Uteligger for CO- høy; lav RSD og høy leverage.

Intervaller og antall objekter som er benyttet i modellene er da gitt som i tabell 5.1.

Tabell 5.1: Bølgelengdeområder og objekter benyttet i modellene

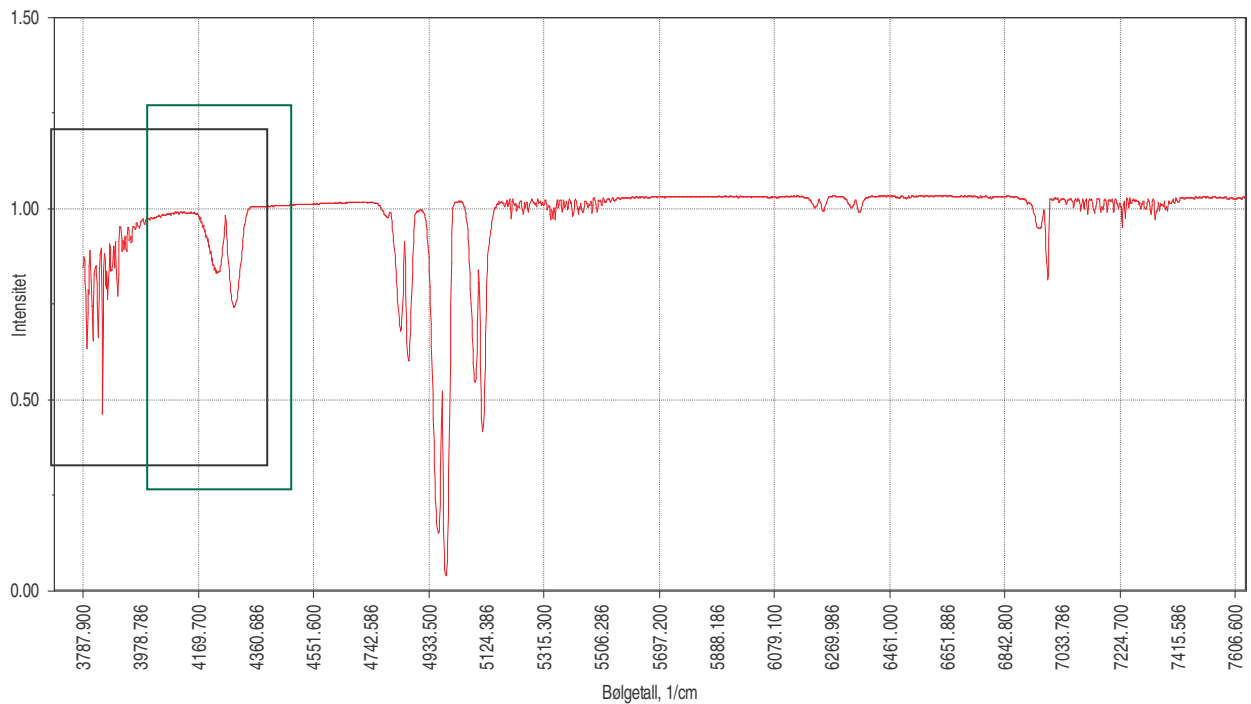
Modell	PLS- Komponenter	Objekter	Intervall, cm ⁻¹
CH4- høy	3	30 CH4- høy- objekter	3787- 7641
CH4- lav	4	30 CH4- lav- objekter	3787- 7641
CO2- høy	2	45 CO2- høy- objekter	4800- 5130
CO2- lav	2	14 CO2- lav- objekter – en uteligger	4800- 5130
CO- høy	1	14 CO- høy- objekter – en uteligger	3787- 4335
CO- lav	5	45 CO- lav- objekter	4121- 4638

Figur 5.9 viser et spekter med en blanding av høy konsentrasjon av CH₄ og høyt konsentrasjon av CO₂. Området som er markert med en svart firkant er området som benyttes i modellene for CO₂, CH₄ bruker hele området i spekteret.



Figur 5.9: Områdene som brukes av CO₂ (svart firkant) og CH₄ (hele området i spekteret).

Figur 5.10 viser områdene som benyttes i modellene for CO- høy og CO- lav i et spekter med blanding av høy konsentrasjon av CO og høy konsentrasjon av CO₂. Svart firkant angir området for CO- høy, grønn firkant angir området for CO- lav.



Figur 5.10: Områdene som brukes i modellene for CO- høy (svart firkant) og områdene som brukes for CO- lav (grønn firkant).

5.2 Modellvalidering

For å validere modellene tilpasses valideringsspektrene tatt opp i oktober 2006 til modellene bygget av spektre fra september 2006. Prediksjons- avvikene en får er et mål på hvor gode modellene er. Tabell 5.2 viser en oversikt over avvikene. Gjennomsnittlig og maksimalt prosentavvik er regnet ut på bakgrunn av de objektene som modellen er gyldig for. Det vil si, for CH₄- høy er gjennomsnittlig og maksimalt prosentavvik regnet ut på bakgrunn av objekter som har høyt trykk av CH₄. Objekter med 0 eller lav konsentrasjon av CH₄ er utelatt fra utregningene, ettersom de vil regnes som uteliggere.

Likedan vil gjennomsnittlig og maksimalt avvik for CH₄- lav være regnet ut på bakgrunn av objekter med lav konsentrasjon av CH₄, objekter med høy konsentrasjon og 0 konsentrasjon vil være utelatt i utregningen. Objektene med 0 konsentrasjon skal imidlertid fanges opp av lavkonsentrasjons- modellene.

Tabell 5.2: Modellvalidering

Modell	Gjennomsnittlig %- avvik	Max %- avvik	Antall objekter
CH4- høy	5.4	12.2	31
CH4- lav	11.1	32.1	32
CO2- høy	3.3	13.9	47
CO2- lav	4.8	10.4	16
CO- høy	4.9	19.0	16
CO- lav	2.7	10.8	48

For CH4- lav er det maksimale avviket litt høyt, men det skyldes objektene som har lavest konsentrasjon. Objektene har tilnærmet like respons- residualer, og når en regner ut prosentavvik vil avviket være større for objekter med høyt trykk enn med lavt trykk dersom residual- avviket er det samme.

Ifølge CMR er prosentavvik på 10% akseptabelt, og resultatene er dermed akseptable.

5.3 Bakgrunnseffekter

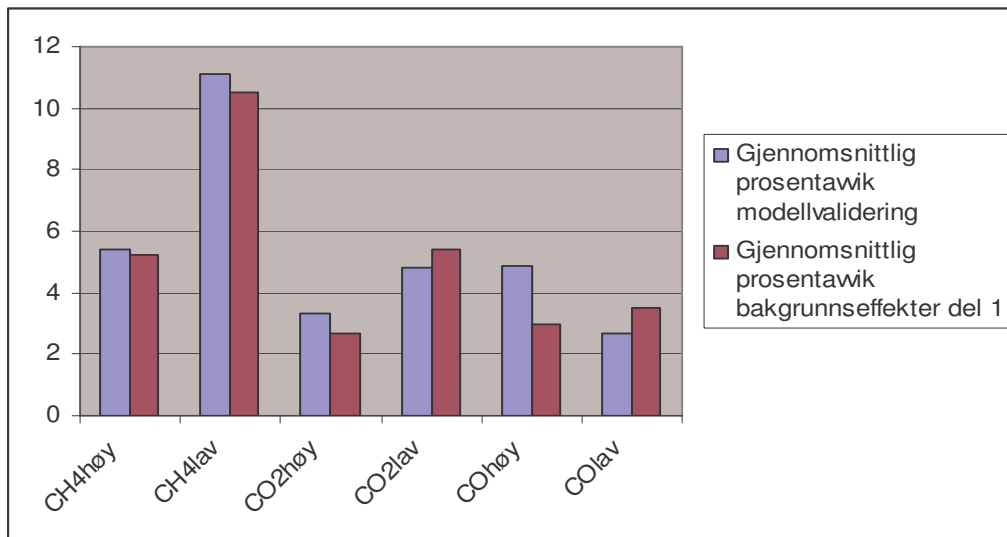
Bakgrunnsspektrene som taes opp har forskjellig overtrykk med N₂, som skal tilsvare overtrykks- situasjonene en har når en tar opp spektre av gassblandingene. Dersom et spekter har 3.6 bar overtrykk med for eksempel CO₂ ville det bli bakgrunnskorrigert med et bakgrunnsspekter som har ca. samme overtrykk av N₂. Dette var situasjonen for spektrene tatt opp i september og oktober 2006. En av effektene som har blitt undersøkt, er om en trenger å gjøre bakgrunnskorreksjonen med samme trykk i bakgrunnsspekteret som i spekteret en skal korrigere. De 64 siste spektrene fra oktober 2006, valideringsspektrene, blir derfor korrigeret med tilfeldige bakgrunnsspektre. Resultatene av den tilfeldige korreksjonen blir at 2 spektre blir korrigeret med bakgrunnsspektre med samme overtrykk, 16 spektre blir korrigeret med bakgrunnsspektre slik at forskjellen er mindre enn 0.4 bar mellom dem, 9 objekter blir korrigeret med bakgrunnsspektre slik at forskjellen blir mer enn 3 bar mellom dem, og resten

blir korrigert slik at forskjellen mellom spekter og bakgrunnsspekter blir mellom 0.4 og 3 bar. Med andre ord er både små og store forskjeller i trykk for bakgrunnsskorreksjonen oppnådd. Spektrene blir dermed tilpasset modellene, og resultatene er vist i tabell 5.3.

Tabell 5.3: Bakgrunnseffekter del 1

Modell	Gjennomsnittlig %-avvik	Max %- avvik	Antall objekter
CH4- høy	5.2	12.5	16
CH4- lav	10.5	27.5	16
CO2- høy	2.7	6.9	24
CO2- lav	5.4	11.2	8
CO- høy	3.0	8.5	8
CO- lav	3.5	10.3	24

Avvikene er tilnærmet like som i tabell 5.2 for modellvalideringen, figur 5.11 viser en sammenligning av resultatene.



Figur 5.11: Sammenligning av gjennomsnittlig prosentavvik for modellvalidering og bakgrunnseffekter del 1

Konklusjonen blir dermed at det ikke spiller noen rolle om det er forskjell i trykk mellom spekteret som skal korrigeres, og bakgrunnsspekteret. En kan dermed nøye seg med ta opp *ett* bakgrunnsspekter, for eksempel med middels trykk av N₂.

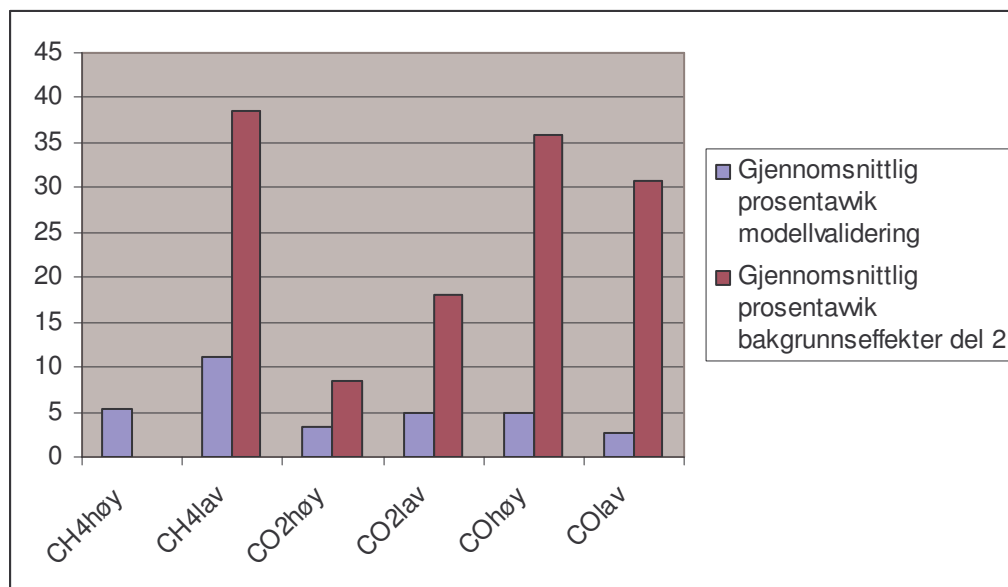
Neste bakgrunnseffekt som undersøkes er å korrigere et spekter med et bakgrunnsspekter som er tatt opp flere måneder før. Dette er ikke det vanlige å gjøre ved bakgrunnskorreksjon av spektre. Bakgrunnsspektre blir vanligvis tatt opp samme dag, helst rett før en har tenkt å ta opp spektre. Dette er derfor en test på om bakgrunnsspektrene er vesentlig forskjellige fra tid til annen, med andre ord om bakgrunnsspektrene ”driver”.

Spektre tatt opp i april 2007 korrigeres med bakgrunnsspektre tatt opp i januar 2007. Objektene tilpasses deretter modellene som er bygget på data fra september 2006, resultatene er vist i tabell 5.4. De to kolonnene til høyre i tabellen er resultatene for de samme spektrene tatt opp i april, men som er korrigert med bakgrunnsspektre tatt opp samme dag. De fungerer dermed som kontroll for testen av bakgrunnseffekter. Ingen spektre med høy konsentrasjon av CH₄ ble tatt opp ettersom gassbeholderen var tom på det tidspunktet.

Tabell 5.4: Bakgrunnseffekter del 2

Modell	Gjennomsnittlig %-avvik feilkorrigert	Max %-avvik feilkorrigert	Gjennomsnittlig %-avvik	Max %-avvik
CH4- høy	-	-	-	-
CH4- lav	38.6	80.6	3.8	6.1
CO2- høy	8.4	18.0	7.0	11.1
CO2- lav	18.1	23.5	5.6	7.3
CO- høy	35.8	65.1	18.3	33.3
CO- lav	30.7	60.7	8.9	16.0

For de aller fleste modellene blir det store avvik, for CO₂- høy blir det litt bedre. En sammenligning av tabell 5.2 og 5.4 er gitt i figur 5.12.



Figur 5.12: Sammenligning av gjennomsnittlig prosentavvik for modellvalideringen og bakgrunnseffekter del 2.

Konklusjonen er at en ikke bør bakgrunnskorrigere spektre med bakgrunnsspektre tatt opp flere måneder i forveien. Bakgrunnsspektre bør taes opp samme dag som spektre taes opp.

5.4 Strømningseffekter

I oktober 2006 ble det tatt opp to spektre av CO₂ med strømning av gass gjennom målecellen. Inngangen til cellen, der gassen blir tilført, var åpen og en "strupte", eller delvis stengt, utgangen for å skape høyt trykk i målecellen. Disse spektrene ga prosentvise avvik på 25 og 26% i modelleringen. Det var derfor nødvendig å undersøke strømningseffektene nærmere.

I januar 2007 ble derfor tre nye spektre av CO₂ tatt opp med gjennomstrømning, og i april 2007 ble enda flere spektre tatt opp gjennomstrømning. For spektrene tatt opp i januar ble to bakgrunns- korreksjoner utført: en der spektrene ble bakgrunnskorrigert med bakgrunnsspektre tatt opp med strømning, og en der spektrene ble bakgrunnskorrigert med bakgrunnsspektre tatt opp uten strømning. På bakgrunn av tidligere resultater ble bare sistnevnte korreksjon utført for spektrene fra april. Spektrene som er tatt opp på vanlig måte i

april, og strømningspektrene, tilpasses modellene fra september 2006. Resultatene er vist i tabell 5.5.

Tabell 5.5: Strømningseffekter

Modell	Gjennomsnittlig %- avvik	Gjennomsnittlig avvik objekter	%- strømming- objekter	Antall strømnings- objekter
CH4- høy	-	-	-	0
CH4- lav	4.0	3.1	-	3
CO2- høy	6.0	8.5	-	6
CO2- lav	5.7	5.4	-	3
CO- høy	17.0	22.0	-	3
CO- lav	7.0	13.9	-	3

For januar- spektrene var gjennomsnittlig prosentvis avvik for strømningsobjektene som ble korrigert med bakgrunnsspektre som ble tatt opp med gjennomstrømning 8.6%, mot 8.7% for objektene som ble korrigert med bakgrunner tatt opp uten strømning, og resten av objektene, som ikke ble tatt opp med strømning, hadde prosentvis avvik på 5.9%.

For enkelte av modellene ser det ut som at gjennomsnittlig prosentvis avvik er litt høyere for strømningsobjektene enn for de vanlige objektene. Dette skyldes at gjennomsnittet er regnet ut på tre strømnings- objekter med lav respons, mens for de vanlige objektene er gjennomsnittlig prosentavvik regnet ut på bakgrunn av 8 objekter med både høy og lav respons. Et eksempel på prediksjon er vist i tabell 5.6. I denne tabellen blir CO predikert for to objekter med blanding av høy konsentrasjon av CO og høy konsentrasjon av CO₂. Objekt nummer 332 er et vanlig spekter, mens objekt nummer 338 er et spekter som er tatt opp med gasstrømning. Residualene er like for de to objektene.

Tabell 5.6: Prediksjon av CO i en blanding av høy konsentrasjon av CO og høy konsentrasjon av CO₂.

Objektnr	Objektnavn	Predikert trykk (bar)	Målt trykk (bar)	Residual (bar)	%- avvik
332	10CO i CO2VER05april07	0,254	0,219	-0,035	16,013
338	10CO i CO2VERF03april07	0,276	0,241	-0,035	14,558

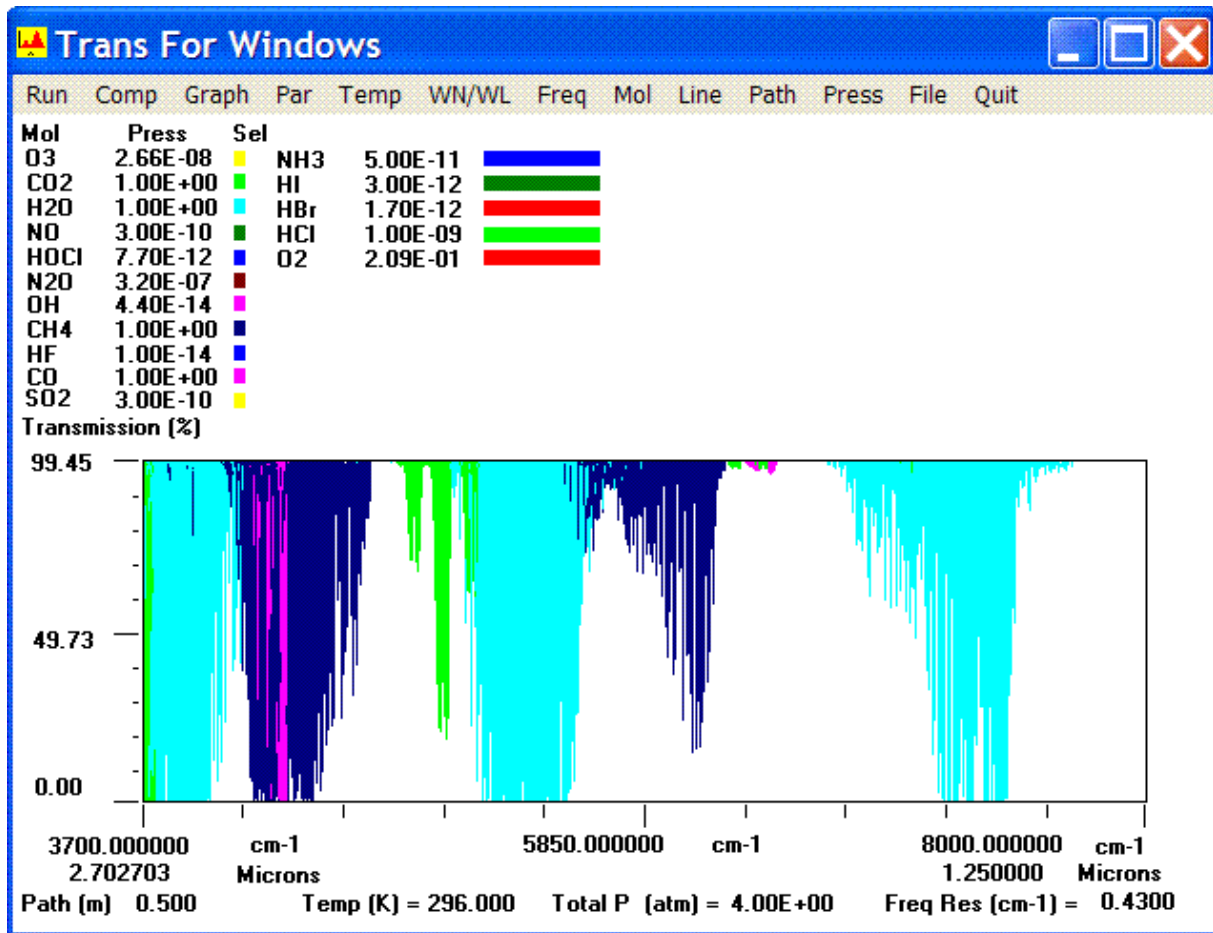
Totalt er det dermed ingen signifikant forskjell mellom å korrigere strømmingsspektre med bakgrunnsspektre tatt opp med gasstrømning, og mellom å korrigere strømmingsspektre med bakgrunnsspektre tatt opp uten strømning. Videre er det ingen signifikant forskjell mellom de vanlige objektene og strømmingsspektrene som ble korrigert med bakgrunnsspektre tatt opp uten gasstrømning.

Konklusjonen blir dermed at spektre tatt opp med gass- gjennomstrømning kan korrigeres med bakgrunnsspektre tatt opp uten strømning.

5.5 Modellbygging del 2

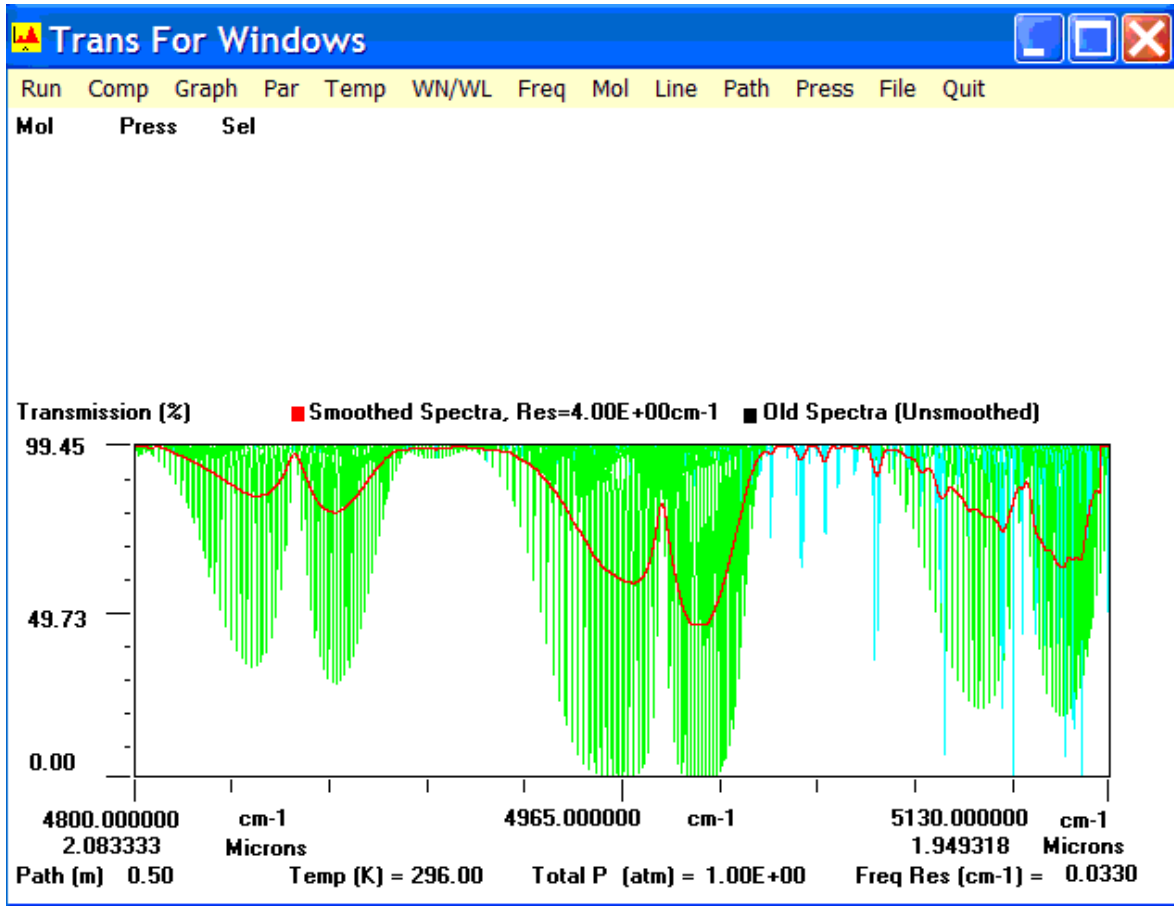
Ved forbrenning av naturgass dannes det vanndamp i tillegg til CO₂. Det er derfor nødvendig å undersøke hvor vanndamp gir signal i området som blir undersøkt. Den eneste gassblandingen som var tilgjengelig inneholdt bare 20 ppm vanndamp, og det var ikke mulig å observere noen signaler i spektre av blandingen, uansett hvor mye overtrykk en benyttet. Jon Oddvar Hellevang ved CMR benyttet derfor et simuleringsprogram, HitranPC Simulations Version 3.0³⁹, for å simulere spektre av vanndamp.

Figur 5.13 viser et simulert spekter der alle de aktuelle gassene er med: CH₄, CO₂, CO og H₂O. Blandingen inneholder 1 bar av hver av gassene.



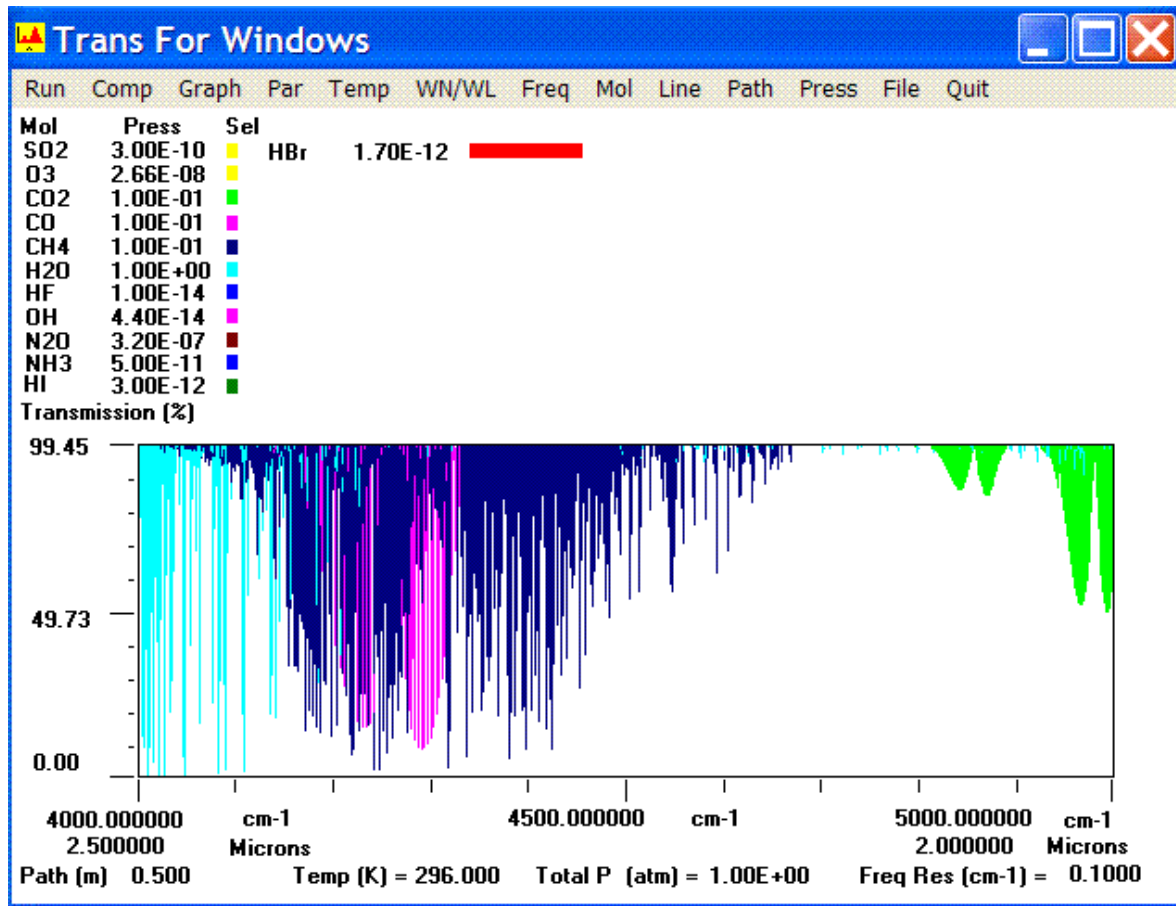
Figur 5.13: Simulert spekter av CH_4 , CO_2 , CO og H_2O i blanding. $\text{CH}_4=\text{CO}_2=\text{CO}=\text{H}_2\text{O}=1$ bar. $\text{CH}_4=\text{blå}$, $\text{CO}_2=\text{grønn}$, $\text{CO}=\text{rosa}$ og $\text{H}_2\text{O}=\text{turkis}$.

Ved å se nærmere på simulerte spektre av vann og hver av de andre gassene kan en se hvor vann vil dominere og dermed hvilke områder en ikke bør benytte i modellene. Situasjonen for CO_2 er vist i figur 5.14 En hale av H_2O dukker opp i høyre ende av CO_2 sine signaler; det området bør utelates fra modellen.

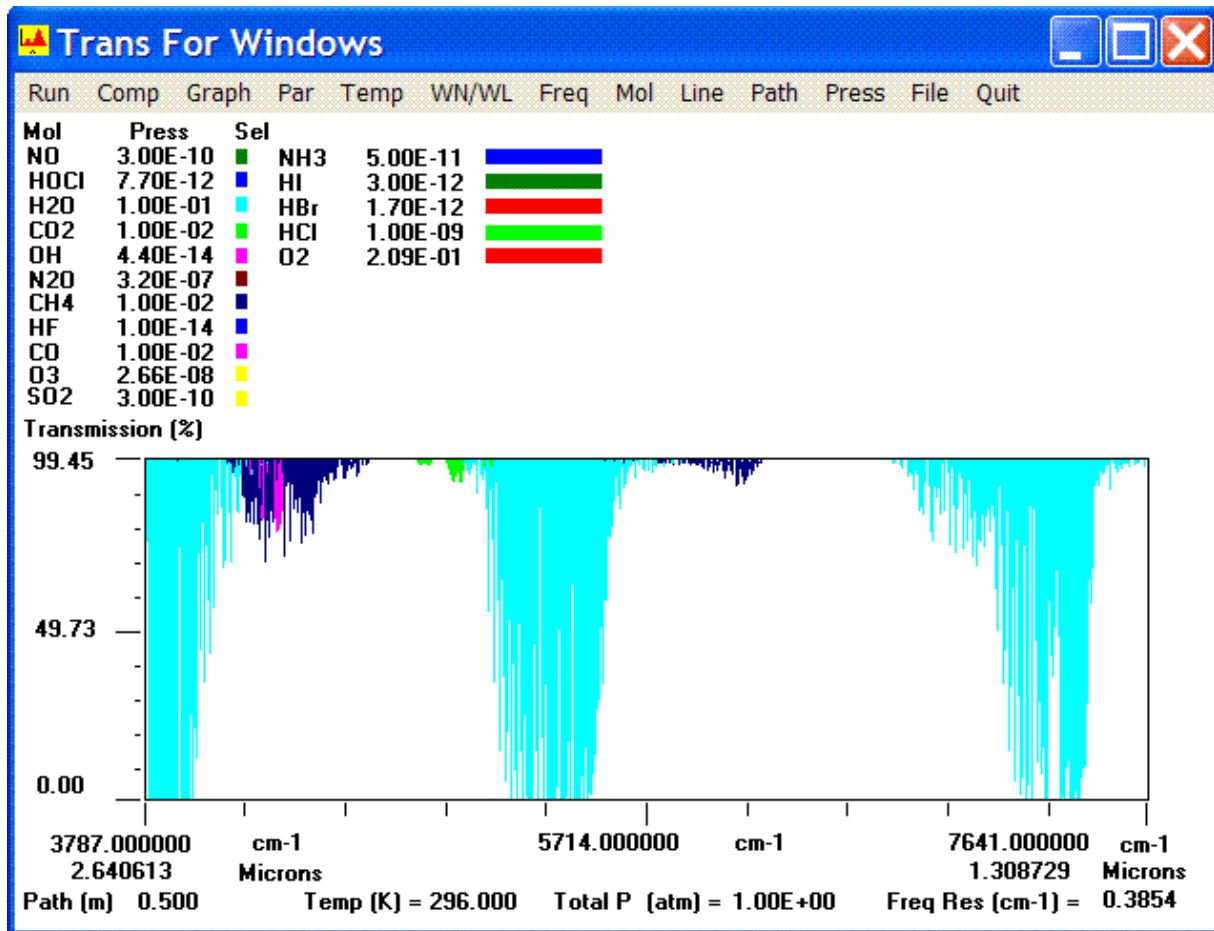


Figur 5.14: CO_2 og H_2O . $CO_2=H_2O=1$ bar. $CO_2=gr\ddot{o}nn$ og $H_2O=turkis$.

For CO sammen med CH_4 blandes signalene som vist i figur 5.15 og 5.16.



Figur 5.15: CO og CH₄ sine signaler blandes. H₂O=1 bar,
 CH₄=CO₂=CO=0.1 bar. CH₄=blå, CO₂=grønn, CO=rosa og H₂O=turkis.



Figur 5.16: Alle 4 gassene blandet. $H_2O=CH_4=CO_2=CO=0.01$ bar. $CH_4=bl\ddot{a}$, $CO_2=gr\ddot{o}nn$, $CO=rosa$ og $H_2O=turkis$.

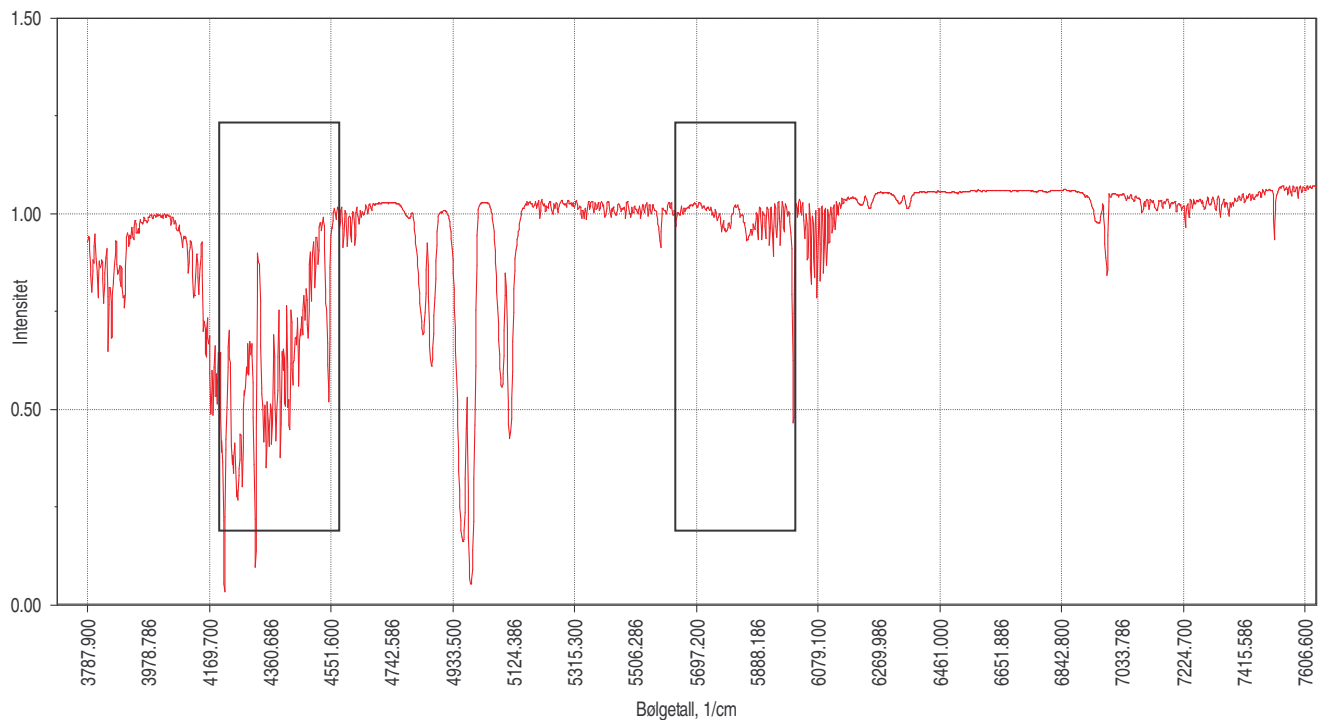
Basert på disse figurene velges nye bølgelengdeområder for de forskjellige modellene, og nye modeller bygges. Tabell 5.7 viser hvilke intervaller som benyttes.

Tabell 5.7: Nye bølgelengdeintervaller i modellene

Modell	Bølgelengde- intervall	PLS- komponenter	Objekter
CH4- høy	4351- 4601 + 5751- 6200	2	30
CH4- lav	4351- 4601 + 5751- 6200	2	30
CO2- høy	4800- 5020	2	45
CO2- lav	4800- 5020	2	14
CO- høy	4100- 4501	2	14
CO- lav	4100- 4501	4	45

Med disse modellene blir det færre PLS- komponenter enn for modellene som er benyttet til nå, dette er fordi variabelmengden er redusert. Dermed fjernes variasjon som fører med seg støy.

Modellene for CO₂ og CO er blitt litt justert, mens modellene for CH₄ har fått litt større endringer. Figur 5.17 viser hvilke områder som nå gjelder for den modellen.



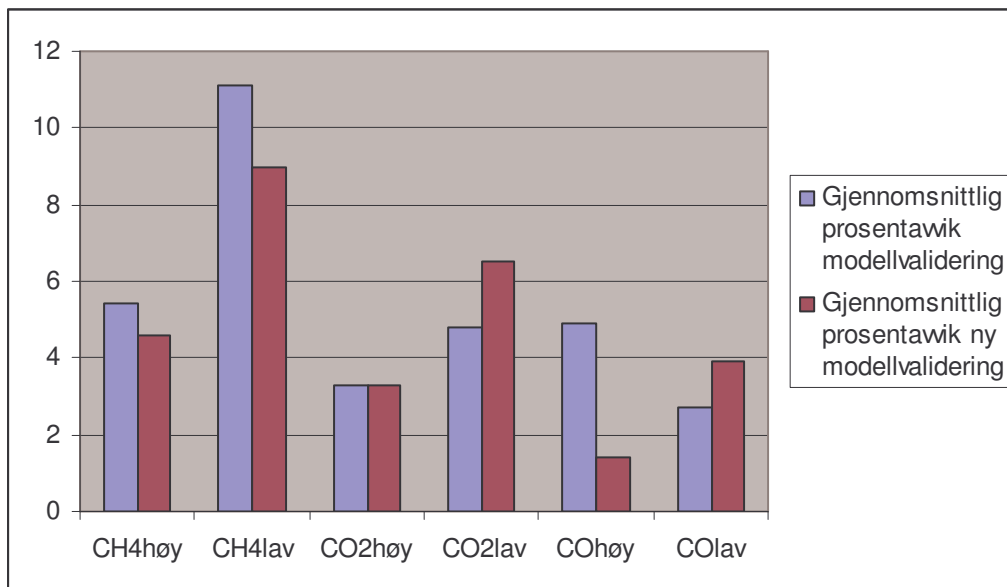
Figur 5.17: Bølgetallområdene som benyttes i modellene for CH₄.

De nye modellene testes deretter ut på valideringsobjektene fra oktober 2006, resultatene vises i tabell 5.8

Tabell 5.8 Modellvalidering

Modell	Gjennomsnittlig %- avvik	Max %- avvik
CH4- høy	4.6	12.4
CH4- lav	9.0	24.8
CO2- høy	3.3	14.3
CO2- lav	6.5	16.2
CO- høy	1.4	5.6
CO- lav	3.9	15.0

Ved å sammenligne disse resultatene med valideringsresultatene fra tabell 5.2 får en figur 5.18.



Figur 5.18: Sammenligning av modellvalidering for nye og gamle modeller.

De nye modellene er i enkelte tilfeller litt bedre, i enkelte tilfeller litt dårligere. Modellene må sies å være godkjente.

5.6 Temperatureffekter

Avgassene som kommer ut av en brenselcelle vil holde temperaturer opptil 1000°Celsius, de skal avkjøles til romtemperatur, eller maksimalt 75° Celsius. Økt temperatur fører til at spektre i nær- infrarød spektroskopi får lavere og bredere topper^{21,40}. Denne effekten må undersøkes, og spektre av noen av gassblandingene taes opp med temperaturer på opptil 65° Celsius. Målecellen som er benyttet for opptak av disse spektrene er konstruert for å tåle 75°Celsius, men en vil ikke presse cellen til yttergrensen med en gang.

Spektrene blir bakgrunnsskorrigert med bakgrunnsspektre tatt opp med samme temperatur. Modellene er bygget på objekter som har holdt romtemperatur, og for å undersøke om det lar seg gjøre å bruke de på spektre som holder opptil 65° Celsius. De nye modellene fra tabell 5.7 blir benyttet til å predikere responsene til de nye spektrene, tabell 5.9 viser resultatene.

Tabell 5.9: Høytemperatur- spektre

Modell	Gjennomsnittlig %- avvik	Max %- avvik
CH4- høy	-	-
CH4- lav	12.9	23.7
CO2- høy	3.3	14.3
CO2- lav	-	-
CO- høy	10.2	20.1
CO- lav	6.9	10.5

Ingen spektre med høy konsentrasjon av CH₄ og ingen med lav konsentrasjon av CO₂ ble tatt opp på grunn av henholdsvis eksplosjonsfare og mangel av gass.

Totalt er resultatene akseptable, men det var en stund usikkert om avgassene ville holde enda høyere temperatur, og dermed om temperatur- effektene ville være enda sterkere. Det ble derfor nødvendig å omforme responsen fra trykk til konsentrasjon.

Alle objektene sine responser overføres dermed fra trykk til konsentrasjon ved hjelp av likning 2.26:

$$C = \frac{P}{RT} \quad (2.27)$$

For et CO₂- objekt blir regnestykket slik:

$$C_{CO_2} = \left(\frac{P_{CO_2} (bar)}{1.01325 bar / atm} \right) / (T_{CO_2} (Kelvin) \cdot 0.08205784 \cdot L \cdot atm \cdot K^{-1} \cdot mol^{-1}) \quad (5.1)$$

Temperaturen til hvert objekt må inkluderes i datasettet. Til nå har spektrene blitt benyttet for å predikere et totaltrykk. Dette trykket er avhengig av konsentrasjonen og temperaturen til gassen. For at en skal predikere konsentrasjon, må en derfor inkludere temperaturen i modellene.

Først undersøkes det hvordan objektene tilpasses til den ideelle gasslov. Ved opptak av spekter ble målecellen fylt med gass ved romtemperatur og spekter ble tatt opp. Deretter ble temperaturen økt først til 45°Celsius, og en ventet i ti minutter etter at temperaturen hadde nådd 45°Celsius før spekter ble tatt opp. Til slutt ble temperaturen økt til 60°Celsius, og spekter ble tatt opp ti minutter etter at temperaturen hadde nådd 60°Celsius. De tre spektrene skal ha lik konsentrasjon. Men dersom spektrene er tatt opp for tidlig er ikke temperaturen i gassen inne i målecellen lik temperaturen til termometeret. Termometeret som brukes er plassert under hyllen i varmeovnen. Dersom spektrene er tatt opp for tidlig vil trykket som observeres på trykkmåleren være lavere enn det ville vært ved riktig temperatur. Dermed vil konsentrasjonen som regnes ut for spektrene som er tatt opp med høy temperatur være lavere enn for romtemperatur- spekteret.

Konsentrasjonene regnes ut på bakgrunn av trykk og temperatur som er observert, og for serien med størst avvik finner en et avvik på 6% fra den ideelle gasslov. I den serien hadde romtemperatur- objektet en konsentrasjon på 0.184, mens objektet med størst avvik hadde konsentrasjon på 0.174. Om dette avviket skyldes at den ideelle gasslov ikke er helt ideell eller om det skyldes at spektrene ble tatt opp for tidlig er uvisst. Dette undersøkes senere i oppgaven.

For å se effekten av den ideelle gasslov regnes konsentrasjonene for de samme objektene ut, dersom det hadde vært romtemperatur for begge objektene. For de to objektene finner en konsentrasjoner på henholdsvis 0.184 og 0.194.

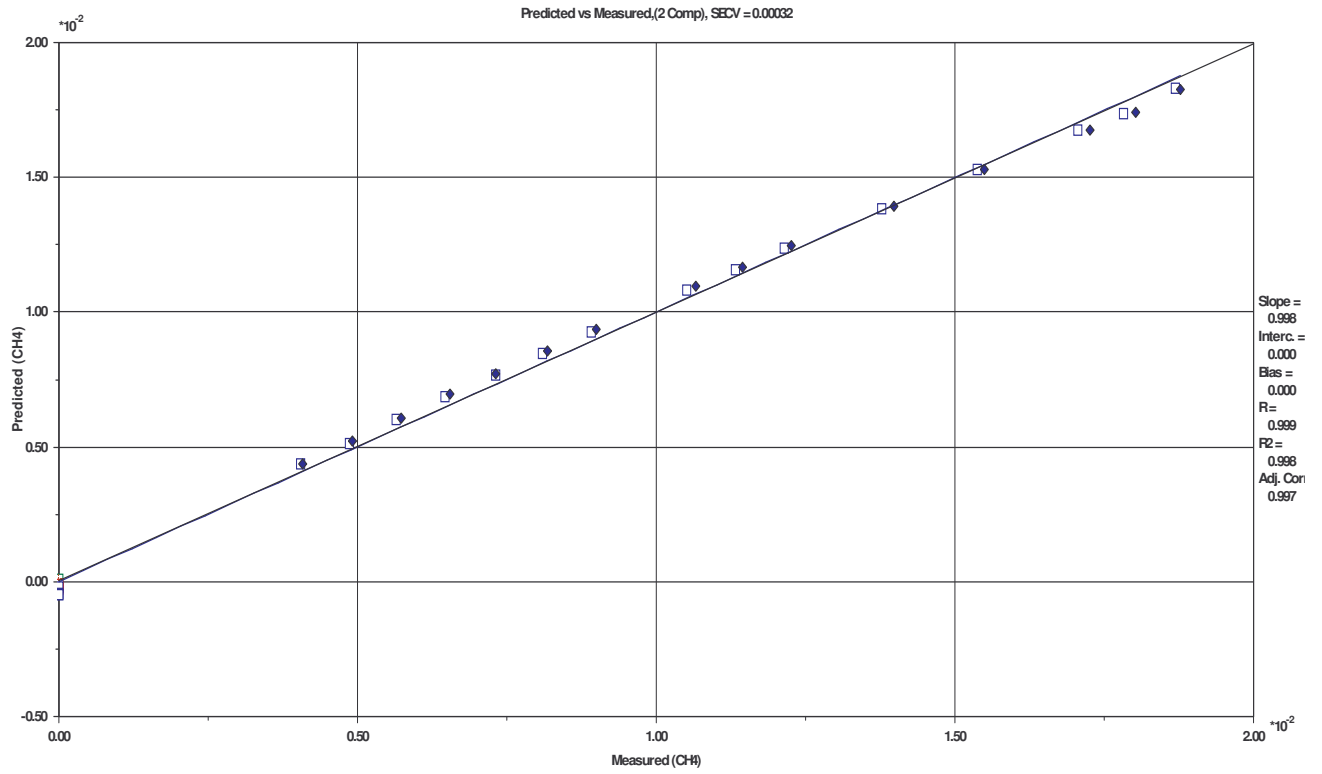
For spektrene med høy temperatur settes konsentrasjonen til å være lik konsentrasjonen til det første spekteret i hver serie. Det antas dermed at den ideelle gasslov gjelder.

5.6.1 Temperaturmodeller

Nye modeller ble bygget for å håndtere spektrere med varierende temperatur. I disse modellene benyttes spektrere med høytemperatur- objekter der det er mulig. Ingen spektrere med høy konsentrasjon av CH_4 og ingen spektrere med lav konsentrasjon av CO_2 er tatt opp med høy temperatur, på grunn av henholdsvis eksplosjonsfare og mangel av gass. Nytt for modellene er at en benytter logaritmisk transformasjon (se 2.5.1) på datasettet. Ved å benytte logaritmisk transformasjon blir forskjellen mellom store tall redusert. For eksempel om en har to tall, 10 og 100, og benytter logaritmisk transformasjon på de tallene. Da får en tilbake tallene 1 og 2. Istedenfor at det ene tallet er 10 ganger så stort som det andre, er det nå bare dobbelt så stort.

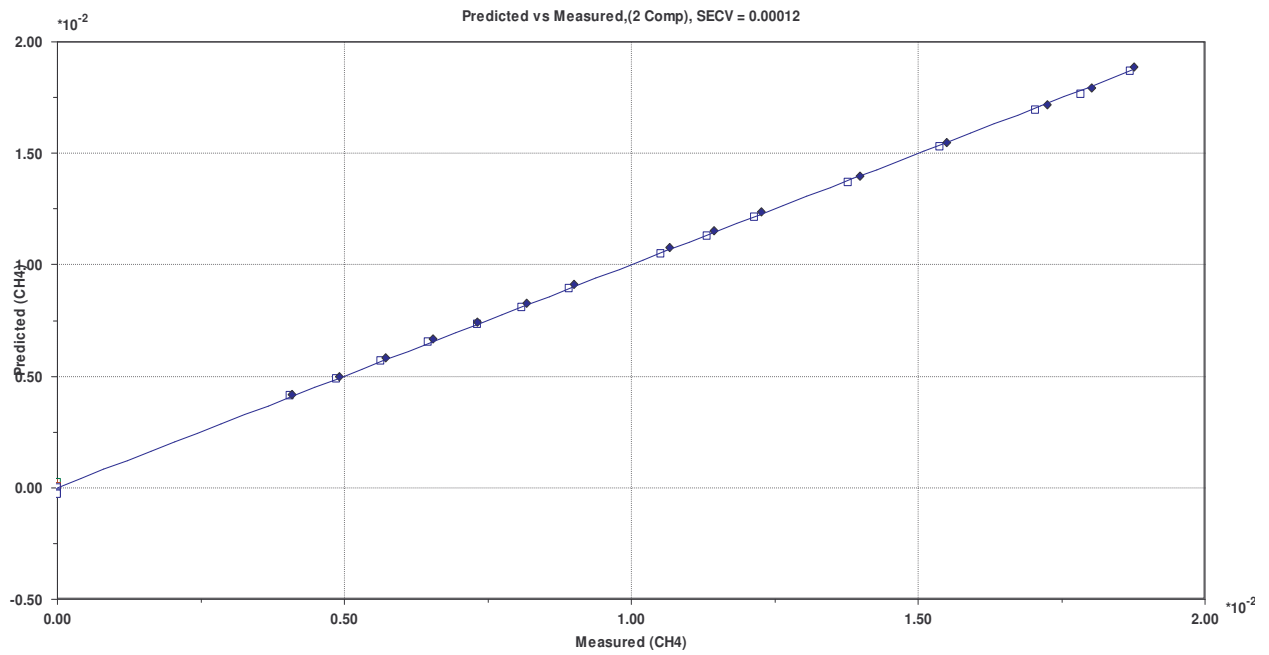
For å vise effekten av å benytte logaritmisk transformasjon kan en sammenligne figur 5.19, der logaritmisk transformasjon ikke er benyttet, og figur 5.20, der logaritmisk transformasjon er benyttet. Disse figurene stammer fra bygging av modell for CH_4 - høy, der en ser predikert respons mot målt respons for objektene i modellen.

Resultater



Created: 11/22/07 18:00:54

Figur 5.19: Predikert respons mot målt respons for CH₄- høy, uten logaritmisk transformasjon.

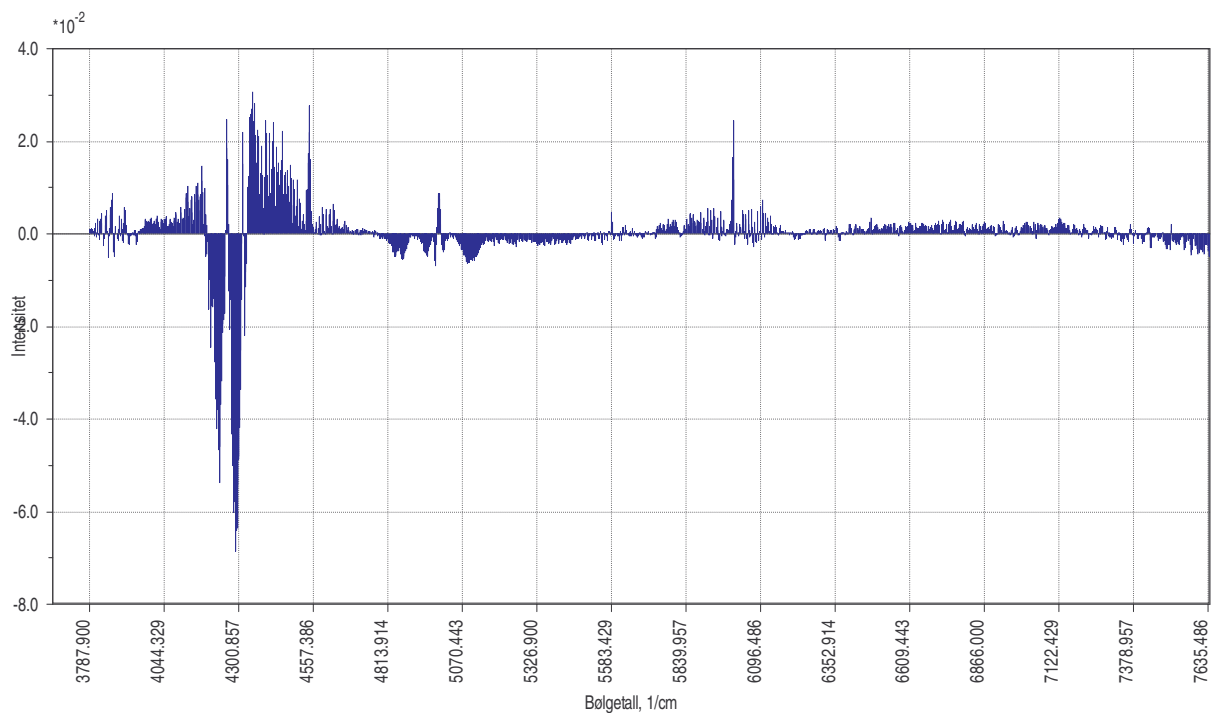


Figur 5.20: Predikert respons mot målt respons for CH₄- høy, med logaritmisk transformasjon.

I figur 5.19 danner objektene en krummet linje, mens i figur 5.20 danner objektene en rett linje. Dette tyder på at det kan være mulig å modellere høy- og lav- konsentrasjonsobjekten med samme modell.

Ettersom signalene til CO og CH₄ overlapper, derfor studeres regresjonskoeffisientene til modellene for CO og CH₄.

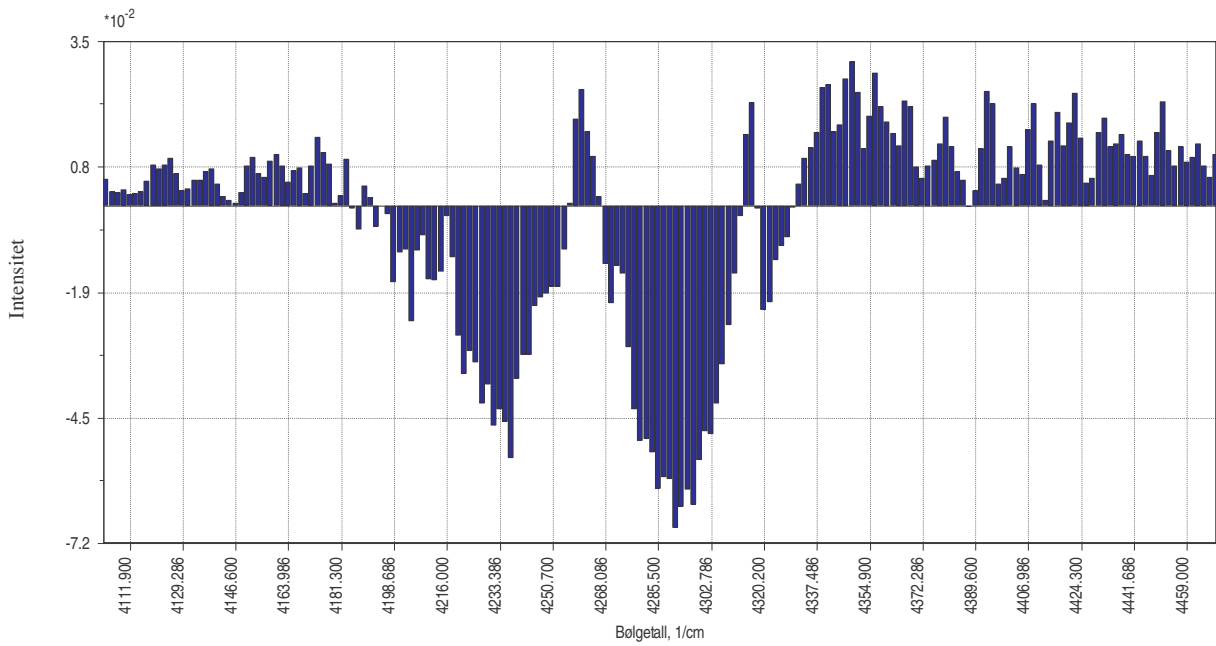
Ved å bygge en modell for CO der alle spektre med både CO og CH₄ er med og deretter se på regresjonskoeffisientene kan en se hvor stor effekt hver variabel har på prediksjonen. En kan da for eksempel velge ut variabler som har bare har positiv effekt på prediksjonen, eller bare negativ effekt, eller en blanding. Figur 5.21 viser regresjonskoeffisientene til alle variablene fra 3787 cm⁻¹ til 7641 cm⁻¹.



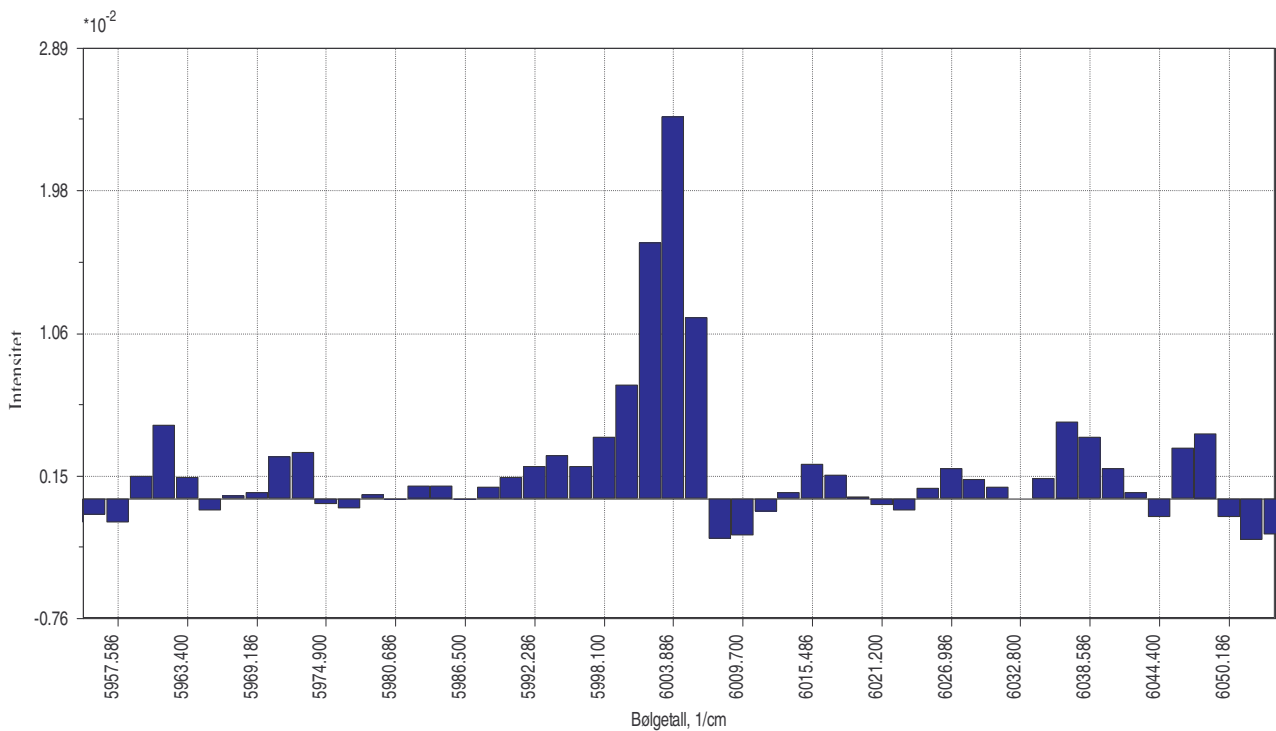
Figur 5.21: Regresjonskoeffisienter for CO.

Ved å zoome inn i de områdene som er aktuelle, nemlig i områdene der CO og CH₄ har signal, får en figur 5.22 og 5.23.

Resultater



Figur 5.22: Regresjonskoeffisienter for bølgetall fra 4110 cm^{-1} til 4440 cm^{-1} .

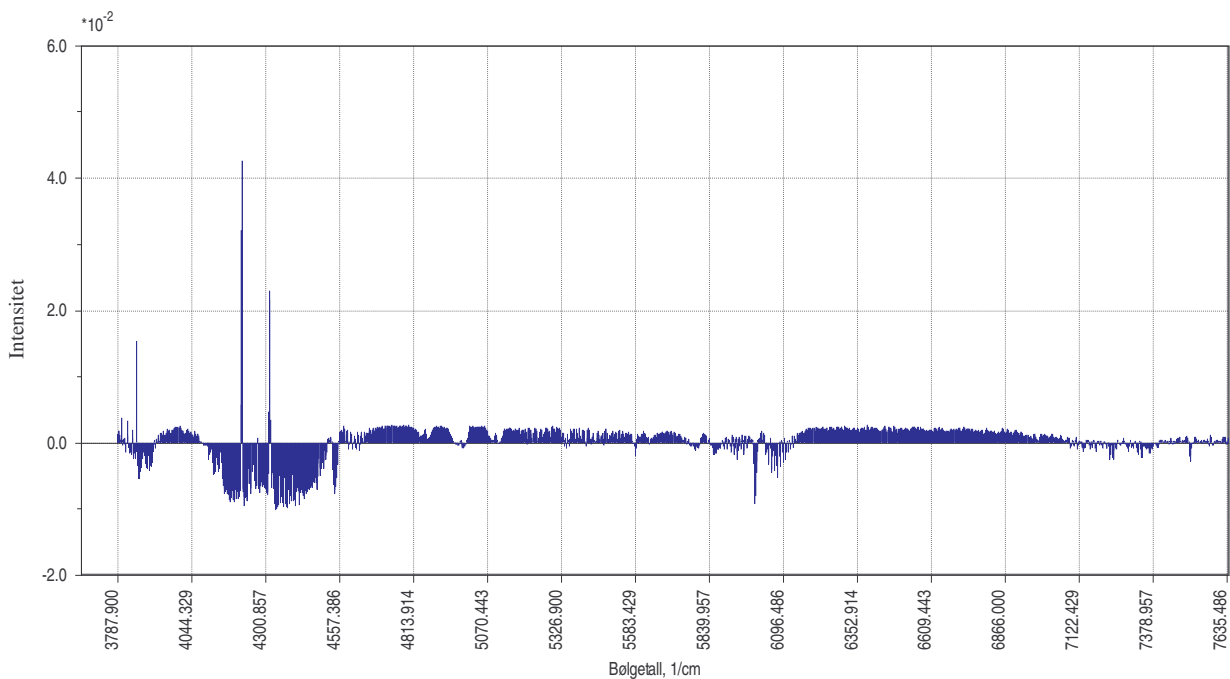


Figur 5.23: Regresjonskoeffisienter for bølgetall fra 5960 cm^{-1} til 6090 cm^{-1} .

Ved å teste ut forskjellige varianter av CO- modeller finner en de beste intervallene. Testingen foregår ved å bruke de samme objektene fra modell til modell, men forskjellige bølgetall-områder. Deretter tilpasses alle CO- objekter til modellene, den som har lavest SEP- verdi regnes som den beste varianten. En variant er å bruke området fra 4198 cm^{-1} til 4329 cm^{-1} pluss området fra 6000 cm^{-1} til 6005 cm^{-1} , en annen er å kutte ut området fra 6000 cm^{-1} til

6005 cm^{-1} . Den beste varianten fås imidlertid ved å kutte ut områdene som er positive i figur 5.22 i området mellom 4198 cm^{-1} til 4329 cm^{-1} , i tillegg til å ta med området fra 6000 cm^{-1} til 6005 cm^{-1} .

Samme prosedyre utføres for CH_4 : en modell bygges for CH_4 der alle objektene med CO og CH_4 er med og regresjonskoeffisientene studeres. Figur 5.24 viser situasjonen for CH_4 .



Figur 5.24: Regresjonskoeffisienter for CH_4 .

For CH_4 finner en at området fra 4320 cm^{-1} til 4557 cm^{-1} pluss området fra 5712 cm^{-1} til 6152 cm^{-1} gir lavest SEP- verdi.

5.6.2 Modeller uten oppdeling i høy og lav konsentrasjon

Med bølgetall- intervaller bestemt kan det undersøkes om det lar seg gjøre å ikke dele modellene opp i høy- og lav- konsentrasjonsområder. Altså, kun en modell per gasskomponent. I modellene benyttes objekter både med høy og lav konsentrasjon, høy og

lav temperatur og objekter som ikke inneholder den gasskomponenten som predikeres. Resultatene er vist i tabell 5.10

Tabell 5.10: Resultater for modeller som ikke er delt opp i høy- og lav- konsentrasjon

Modell	Gjennomsnittlig %- avvik høykonsentrasjons- objekter	Gjennomsnittlig %- avvik lavkonsentrasjons- objekter
CH₄	2.1	22.9
CO₂	4.4	87.5
CO	4.1	30.3

Med andre ord, høykonsentrasjonsobjektene blir bra predikert, mens lavkonsentrasjonsobjektene blir dårlig predikert.. Dersom det ikke er viktig å predikere nøyaktig for objektene med lav konsentrasjon går det an å benytte modellene som ikke er delt opp i høy og lav konsentrasjon.

5.6.3 Modeller med oppdeling i høy og lav konsentrasjon

Ettersom resultatene for å bare benytte en modell per gasskomponent ikke ga gode resultater for både objekter med høy og lav konsentrasjon bygges det nå separate modeller. Altså, to modeller per gasskomponent, en for høy og en for lav konsentrasjon. De nye modellene bygges og intervaller, antall PLS- komponenter og objekter for hver av de er vist i tabell 5.11.

Tabell 5.11: Nye modeller

Modell	Bølgelengde- intervall	PLS- komponenter	Objekter
CH4- høy	4320- 4557 + 5712- 6152	2	46
CH4- lav	4320- 4557 + 5712- 6152	4	21
CO2- høy	4800- 5105	3	20
CO2- lav	4800- 5105	2	39
CO- høy	4198- 4214 + 4217- 4254 + 4268- 4312 + 4320- 4329 + 6000- 6005	3	22
CO- lav	4198- 4214 + 4217- 4254 + 4268- 4312 + 4320- 4329 + 6000- 6005	6	39

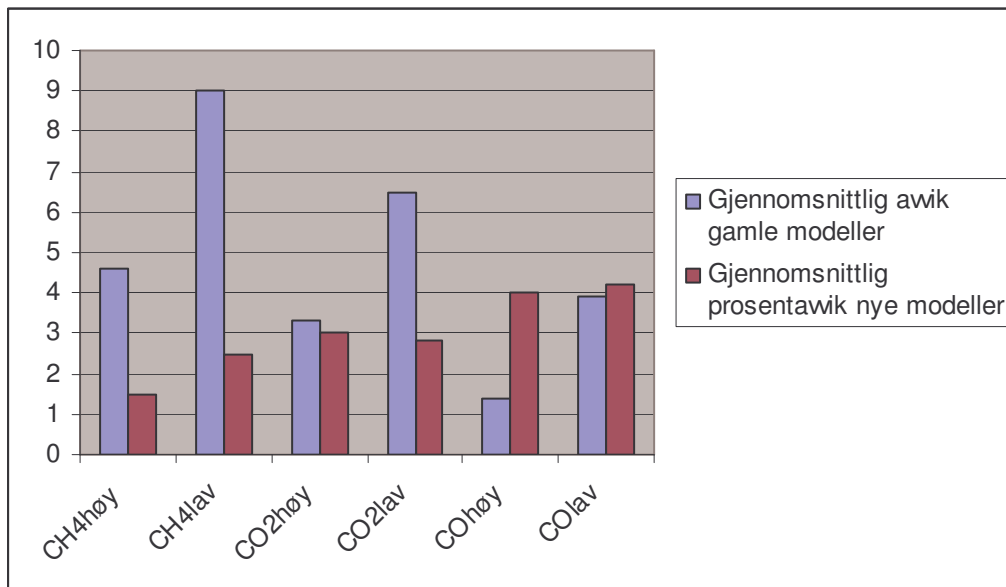
Modellene for CH₄ og CO₂ blir finjustert, mens modellene for CO- lav blir kuttet litt opp. Disse modellene testes ut på alle objektene, fra september 2006 til høytemperaturobjekten, med resultater vist i tabell 5.12:

Tabell 5.12: Resultater for nye modeller

Modell	Gjennomsnittlig %- avvik	Max %- avvik
CH4- høy	1.5	9.5
CH4- lav	2.5	8.4
CO2- høy	3.0	10.9
CO2- lav	2.8	10.0
CO- høy	4.0	17.1
CO- lav	4.2	24.6

Noe som er verdt å legge merke til er at spektrene fra april 2007 som ble korrigert med bakgrunnsspektre fra januar 2007 nå blir predikert bra. Det vil ikke si at en bør bakgrunnskorrigere spektrene sine med bakgrunner tatt opp flere måneder i forveien, men

enkelte modeller klarer allikevel å fange opp slike objekter. Figur 5.25 viser hvordan resultatene er i forhold til resultatene fra modellene fra tabell 5.7.



Figur 5.25: Resultater for gamle modeller i forhold til de nye modellene.

Konklusjonen blir at de nye modellene med finjusterte variabelområder, temperatur, logaritmisk transformasjon og ikke minst med responser omformet til konsentrasjon, fungerer veldig bra.

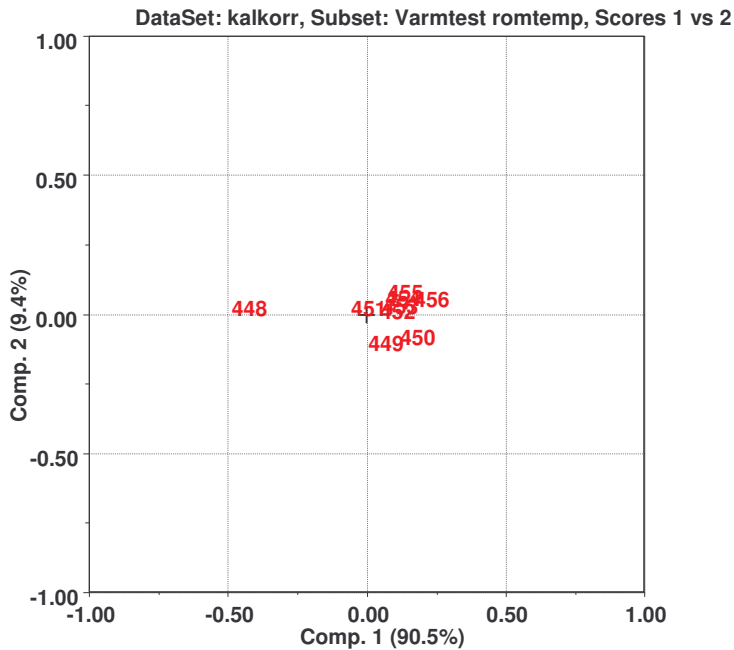
5.6.4 Temperaturvalidering

I desember 2007 ble en serie med spektrere av CO₂ tatt opp, der en studerte hvor lang tid en bør vente med å ta opp spektrere etter å ha fylt målecellen og økt temperaturen.

Det første som ble undersøkt, var om gassen holdt en annen temperatur enn romtemperatur idet den kom ut av gassbeholderen og inn i målecellen. En gass under lavt trykk vil ha lav temperatur, som i gassbeholderen. Når gassen kommer ut av gassbeholderen vil trykket økes drastisk, gassen vil ekspandere og dermed økes temperaturen.

Målecellen ble fylt med ren CO₂, spektrere ble tatt opp med en gang, og deretter hvert femte minutt inntil 40 minutter var gått. En umiddelbar effekt som ble observert var at overtrykket

sank fra 1.977 bar til 1.941 bar fra første til siste spekter. Temperaturen sank i tillegg litt, fra 19° Celsius til 18°Celsius. Det kan være at det er en ubetydelig endring, eller en endring som skyldes at temperaturen generelt sank i lokalet. Totalt fører dette til at konsentrasjonen synker fra 0.12379 til 0.12256, en endring på 1%. Ved å utføre en PCA på spektrene får en figur 5.26.



Figur 5.26: PCA på CO₂- spektre fra desember 2007.

Ett objekt skiller seg klart ut: det er nemlig det første spekteret. Dette spekteret hadde høyest trykk, konsentrasjon og temperatur av alle spektrene. Årsaken til dette er uviss. Det kan skyldes turbulens i målecellen, at gassen ikke er stabil før den har ”hvilt” litt i målecellen. Tabell 5.13 viser hvordan trykket varierer over de 40 minuttene.

Tabell 5.13: Oversikt over trykkvariasjon

Minutter i målecelle	0	5	10	15	20	25	30	35	40
Overtrykk (bar)	1.977	1.965	1.957	1.952	1.948	1.945	1.944	1.943	1.941

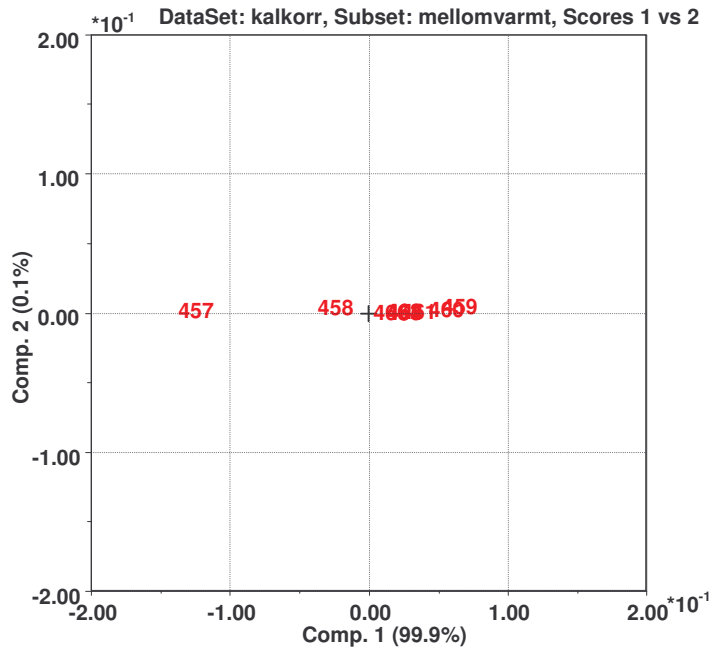
Situasjonen for alle spektrene som er tatt opp ved romtemperatur er at de er tatt opp med en gang målecellen er fylt, med andre ord etter 0 minutter ”hviletid” i målecellen. Det beste hadde vært å vente i 10- 20 minutter med å ta opp spekter hver gang, slik at gassen i målecellen får stabilisert seg. Det er imidlertid et tidsaspekt med i bildet: Dersom en skal ta opp 300 spektre ved romtemperatur ved å vente i 10 minutter mellom hvert spekter sier det seg selv at det blir mange dager på laboratoriet. Ved å ta opp spekter på samme måte hver gang, nemlig med en gang gassen er kommet inn i målecellen, regner en med at det ikke har noe å si for analysen at gassen ikke er helt stabilisert for romtemperaturspektrene.

Neste punkt er å øke temperaturen til 45°Celsius og deretter ta opp spekter hvert 5 minutt inntil 50 minutter er gått. Tabell 5.14 viser hvordan overtrykk og temperatur varierer.

Tabell 5.14: Variasjoner i overtrykk og temperatur for 45°Celsius- objektene

Minutter i målecelle	Overtrykk (bar)	Temperatur (°C)
5	1.957	27-28
10	2.015	37-39
15	2.087	47
20	2.128	46
25	2.140	44
30	2.150	43
35	2.157	43
40	2.164	43
45	2.168	43
50	2.168	42

Spektrene som er benyttet i analysen er tatt opp etter at gassen vært i målecellen i 10 minutter etter at temperaturen har nådd ønsket temperatur. Det tilsvarer objektet som har vært i målecellen i 25 minutter i tabell 5.14. Som tabell 5.14 viser har trolig temperaturen til gassen i spektrene som er benyttet ikke nådd den ønskete temperaturen. PCA på disse objektene er vist i figur 5.27, og viser samme situasjon som for romtemperaturobjektene: det første objektet skiller seg klart ut. Årsaken til det er at en nettopp har økt temperaturen, og temperaturen i gassen er i ferd med å endre seg, og dermed overtrykket.



Figur 5.27: PCA på 45°Celsius- objekter

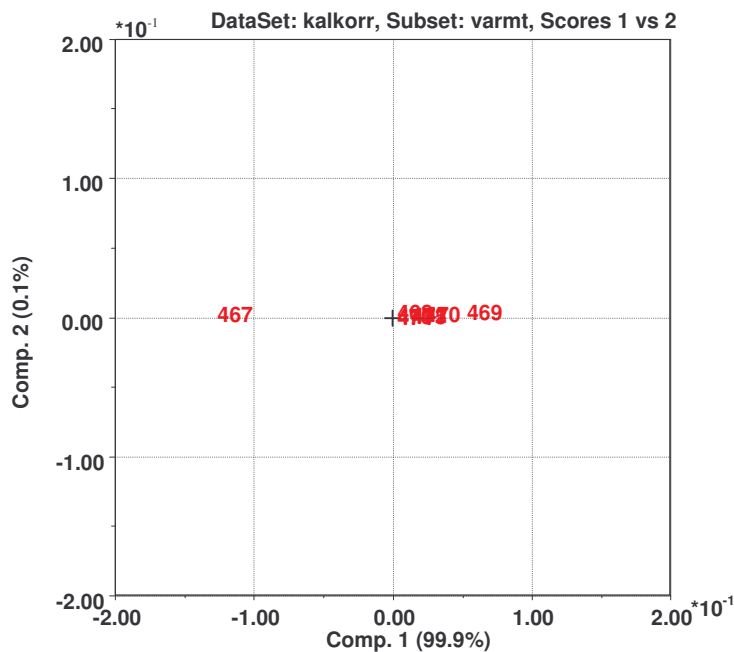
Det beste ville vært å vente med å ta opp spekter i 20- 25 minutter etter at temperaturen har nådd ønsket temperatur. Ettersom en benytter samme måte for å ta opp spekter hver gang, regner en med at det ikke får store effekter.

Siste del er å øke temperaturen til 60°Celsius og ta opp spekter hvert 5 minutt inntil 45 minutter er gått. Tabell 5.15 viser variasjonene i overtrykk og temperatur.

Tabell 5.15: Variasjoner i overtrykk og temperatur for 60°Celsius- objektene

Minutter i målecelle	Overtrykk (bar)	Temperatur (°C)
5	2.177	47-48
10	2.229	60
15	2.299	65
20	2.310	62
25	2.326	61
30	2.330	61
35	2.332	60
40	2.331	60
45	2.331	61

Igjen observeres det at spektrene som er benyttet i analysen, som er tatt opp 10 minutter etter at temperaturen har nådd ønsket temperatur, ikke er helt stabile. Det tilsvarer spekteret som har vært 20 minutter i målecellen. En bør vente i 15-20 minutter fra temperaturen har nådd ønsket temperatur til en tar opp spekter. PCA på spektrene viser samme situasjon som for romtemperatur- og 45°Celsius- objektene, og er vist i figur 5.28.



Figur 5.28: PCA på 60°Celsius- objekter.

Til slutt sammenlignes de objektene som ville blitt brukt i analysen med de som burde blitt brukt ved å tilpasse objektene til modellen for CO₂- høy, resultatene er vist i tabell 5.16. For romtemperaturobjektene er objektet som ville blitt benyttet i analysen objektet som har vært 0 minutter i målecellen (fra tabell 5.13), og det som burde blitt brukt er objektet som har vært i målecellen i 20 minutter. For 45°Celsius- objektene er objektet som ville blitt benyttet i analysen objektet som har vært i målecellen i 25 minutter (fra tabell 5.14), mens objektet som burde blitt benyttet er objektet som har vært i målecellen i 35 minutter. For 60°Celsius- objektene er objektet som ville blitt benyttet i analysen objektet som har vært i målecellen i 20 minutter (fra tabell 5.15), mens objektet som burde blitt benyttet er objektet som har vært i målecellen i 25 minutter.

Tabell 5.16: Tilpassing av analyseobjekter og ”riktige” objekter.

Objekt	%- avvik
CO2 analyse romtemp	10,62052
CO2 riktig romtemp	9,30135
CO2 analyse 45°C	7,13849
CO2 riktig 45°C	8,81061
CO2 analyse 60°C	6,32741
CO2 riktig 60°C	7,33284

Som tabellen viser er det små, om noen, forskjeller mellom objektene. For to tilfeller er tilpassingen bedre, for det siste tilfellet er tilpassingen dårligere. Prosentavviket er under det maksimale avviket for andre objekter (tabell 5.12).

Konklusjonen blir at ”feilen” ved å ikke vente lenge nok med å ta opp spektre er neglisjerbar.

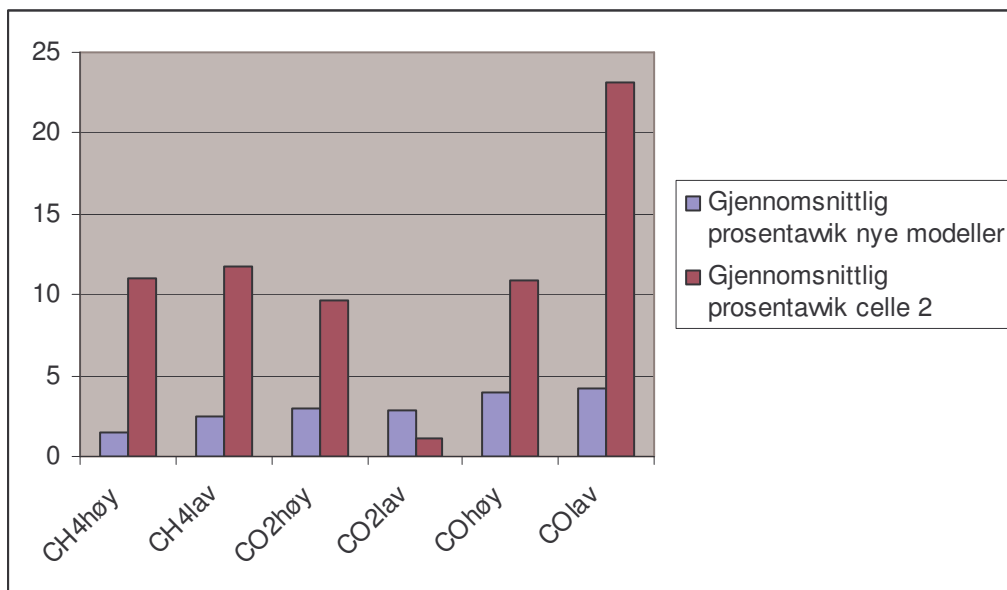
5.7 Kalibreringsoverføring

Så langt har det blitt jobbet med én målecelle. I laboratoriekraftverket som skal testes skal to måleceller benyttes på forskjellige steder i kraftverket. To måleceller er aldri like, og det er viktig å undersøke om modeller bygget på ett instrument kan overføres til et annet. I desember 2007 ble 8 spektre av hver av gassblandingene tatt opp på målecelle 2, totalt 64 spektre. Disse spektrene tilpasses modellene som er bygget på celle 1, for å se om modellene kan brukes direkte. Resultatene er gitt i tabell 5.17.

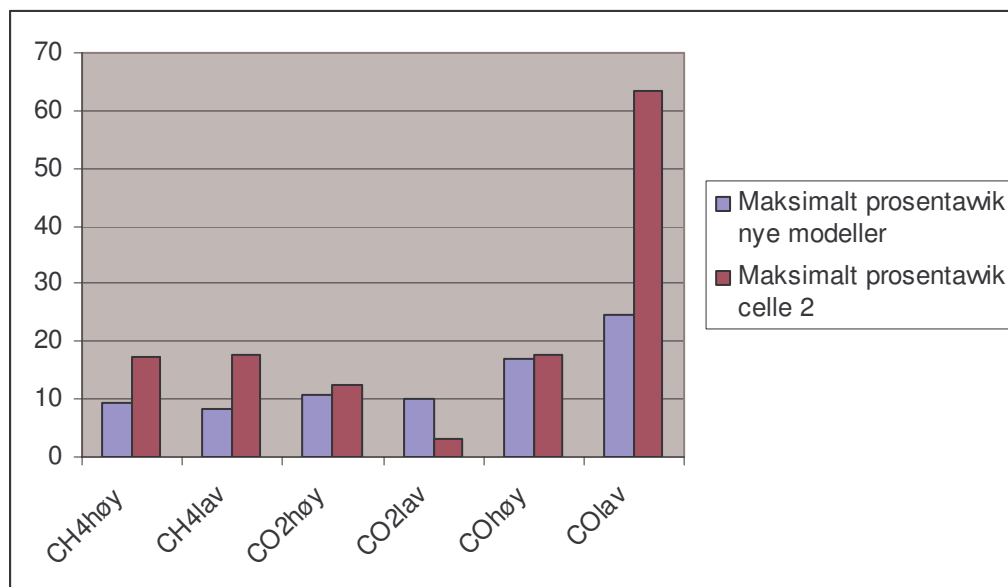
Tabell 5.17: Resultater fra tilpassing av spektre fra celle 2 til modeller bygget på celle 1

Modell	Gjennomsnittlig %- avvik	Max %- avvik
CH4- høy	11.0	17.5
CH4- lav	11.7	17.8
CO2- høy	9.6	12.6
CO2- lav	1.1	3.1
CO- høy	10.9	17.7
CO- lav	23.1	63.3

Sammenlignes resultatene for celle 2 mot resultatene for celle 1 fra tabell 5.12 får en stolpediagram som i figur 5.29 og 5.30.



Figur 5.29: Gjennomsnittlig prosentavvik for modellvalidering av nye modeller og tilpassing av celle 2 til modellene bygget på celle 1.

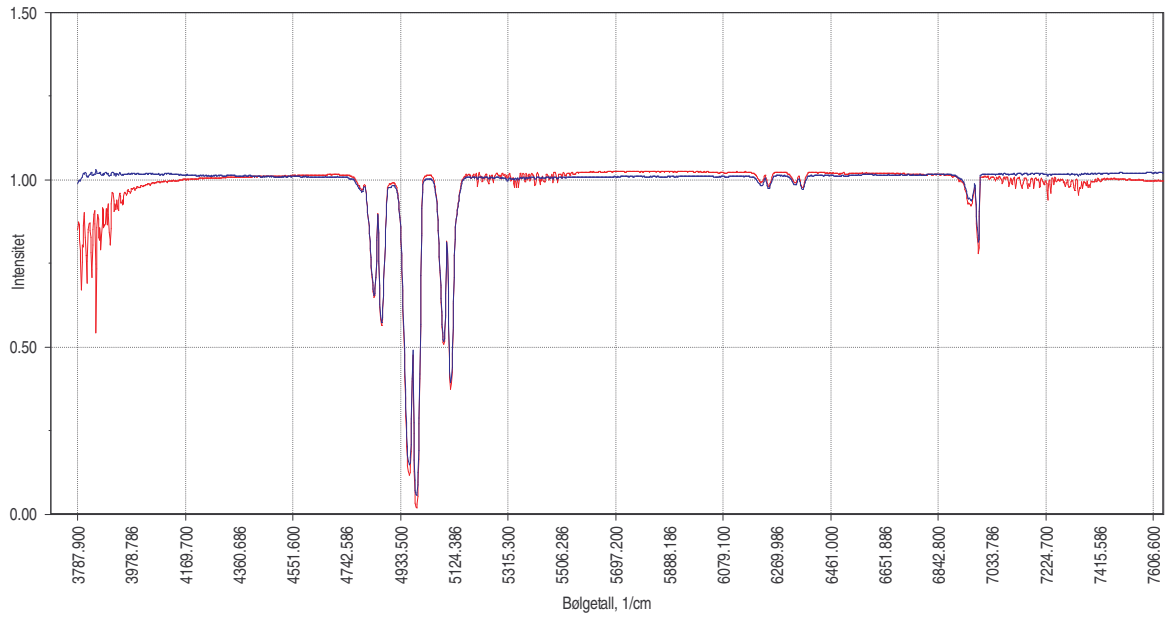


Figur 5.30: Maksimalt prosentavvik for modellvalidering av nye modeller og tilpassing av celle 2 til modellene bygget på celle 1.

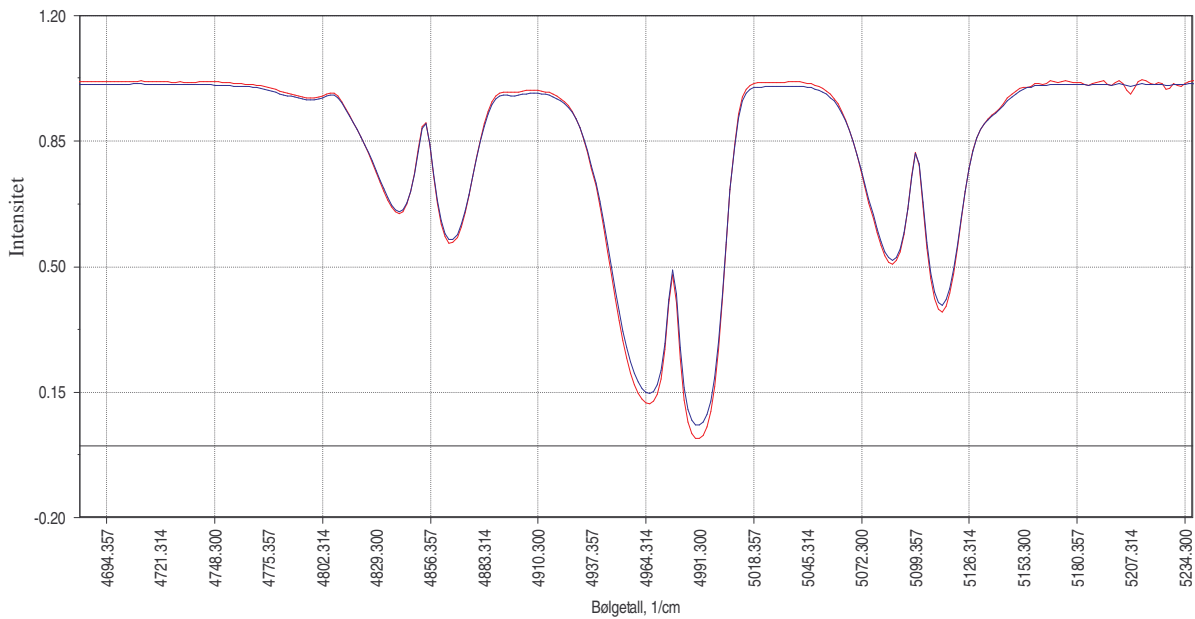
De gjennomsnittlige avvikene er blitt større for alle modellene unntatt CO₂-lav. Samme situasjon observeres for maksimalt avvik har økt litt i tillegg. Totalt er ikke situasjonen altfor sterkt forverret, men det må undersøkes om en standardisering av spektrene fra celle 2 fører til bedre resultater.

En gassblanding med samme trykk tatt opp på to forskjellige instrumenter burde gi to spektre som er helt like. På grunn av små forskjeller i måle-instrumentene blir de sjelden helt like. Situasjonen for de to målecellene er vist i figur 5.31. Spektrene har tilnærmet samme overtrykk av ren CO₂ (høy konsentrasjon av CO₂) og skulle vært tilnærmet like. Det røde spekteret stammer fra celle 1, og det blå stammer fra celle 2. Enkelte områder er tydelig annerledes for de to cellene. Dette er i områder som ikke er med i modellene for CO₂. Ved se litt nærmere på områdene som gjelder for modellene får en figur 5.32. Der ser en at det er noe forskjeller mellom de to spektrene.

Resultater

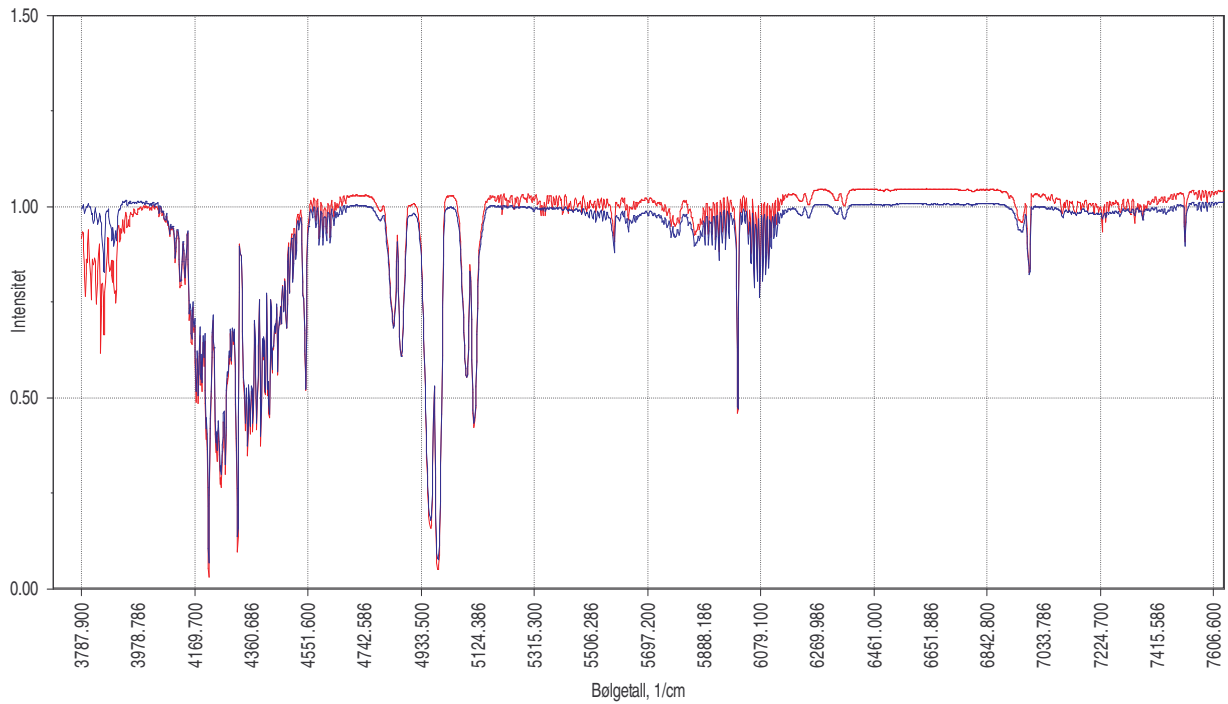


Figur 5.31: Spekter av CO₂ fra de to målecellene. Celle 1 har rød farge, celle 2 har blå farge.



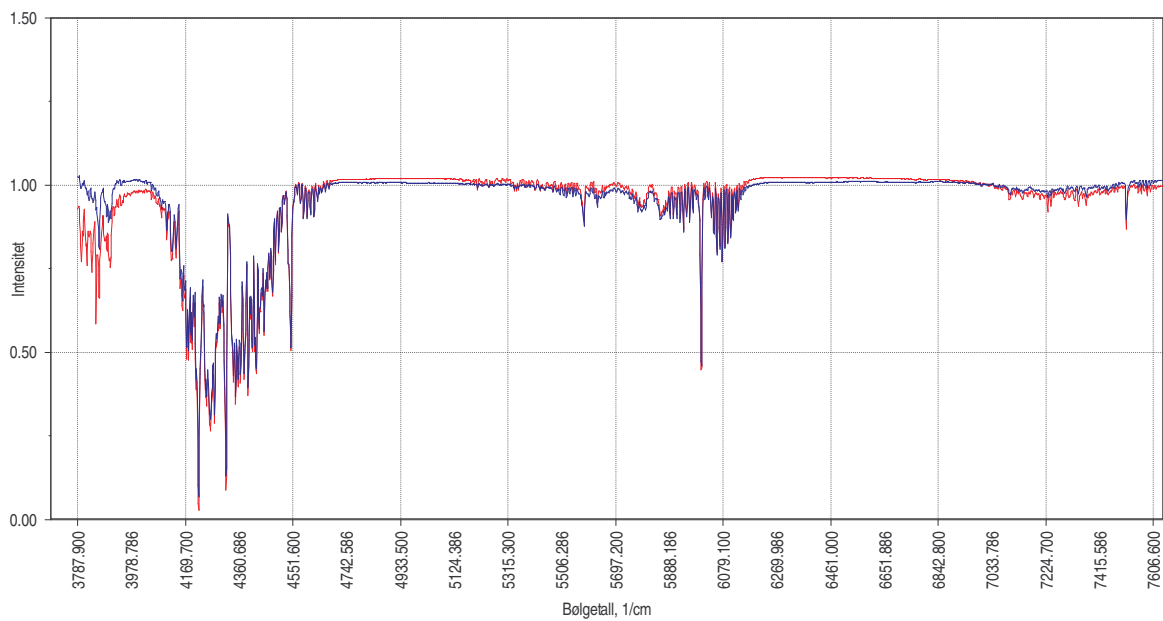
Figur 5.32: Spekter av CO₂ fra de to målecellene med fokus på området som brukes i modellene.

Sammenligning av celle 1 og 2 for blanding av CH₄ med høy konsentrasjon og CO₂ med høy konsentrasjon er vist i figur 5.33.



Figur 5.33: Spektre av blanding av høy konsentrasjon av CH_4 og høy konsentrasjon av CO_2 for celle 1 (rød) og celle 2 (blå).

En sammenligning for blanding av CH_4 med høy konsentrasjon og CO med lav konsentrasjon er vist i figur 5.34.



Figur 5.34: Spektre av blanding av høy konsentrasjon av CH_4 og lav konsentrasjon av CO for celle 1 (rød) og celle 2 (blå).

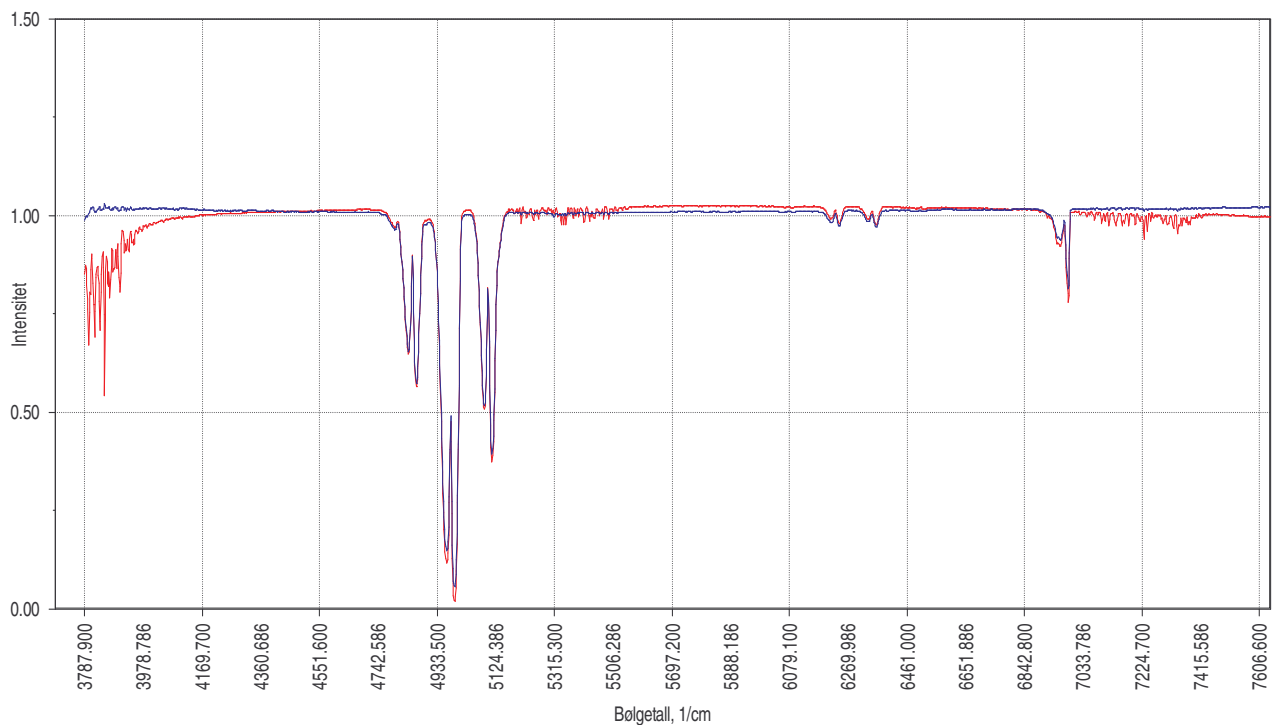
Disse figurene viser at det er en del forskjeller mellom spektrene.

En standardisering av spektrene fra celle 2 kan fjerne disse forskjellene, og to standardiseringer prøves ut: GLSW og PDS. Derivasjon av PDS- standardiserte data undersøkes i tillegg.

5.7.1 GLSW- standardisering

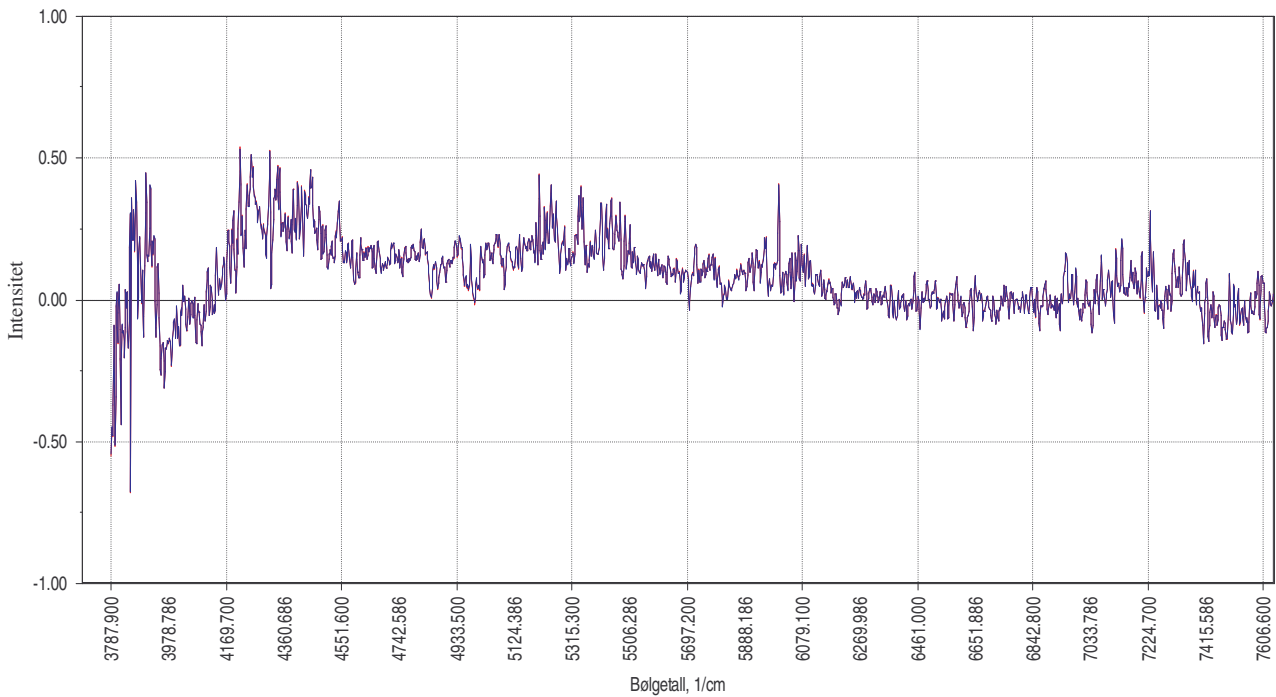
Ved å bruke Generalized Least Squares Weigthing standardiseres begge instrumentene ved å fjerne variasjon som instrumentene ikke har felles. Det blir dermed nødvendig å lage nye modeller. Dette har det ikke vært nok tid til å gjøre, modellene og forbehandlingen fra kapittel 5.6.3 vil derfor benyttes.

Før standardiseringen ser ett spekter av ren CO₂ fra celle 1 (rødt) og ett spekter med samme konsentrasjon av CO₂ fra celle 2 (blått) ut som i figur 5.35.



Figur 5.35: Spektre av ren CO₂ med samme konsentrasjon fra celle 1(rødt) og celle 2(blått).

Standardiseringen fører til at signalene for samme spektre får form som vist i figur 5.36.



Figur 5.36: "Spekter" av standardiserte spektre fra celle 1 og 2 for ren CO₂, samme overtrykk.

Figuren viser et spekter av CO₂ tatt opp med celle 1 og et spekter tatt opp med celle 2 med samme overtrykk, etter standardisering. Det ser ut som én linje, men det er faktisk to linjer. Det vil si at standardiseringen har klart å fjerne den variasjonen som instrumentene ikke har felles. De nye standardiserte signalene tilpasses deretter til modellene bygget på ikke-standardiserte signaler. Resultatene er vist i tabell 5.18.

Tabell 5.18: Resultater fra tilpassing av standardiserte spektre til modeller bygget på ikke-standardiserte spektre

Modell	Celle 1		Celle 2	
	Gjennomsnittlig %- avvik	Max %- avvik	Gjennomsnittlig %- avvik	Max %- avvik
CH4- høy	4.5	16.3	4.7	12.4
CH4- lav	6.3	29.2	7.1	11.6
CO2- høy	5.4	24.2	3.5	12.6
CO2- lav	6.1	21.6	39.5	89.4
CO- høy	3.5	15.1	3.7	5.3
CO- lav	16.1	52.6	10.2	22.4

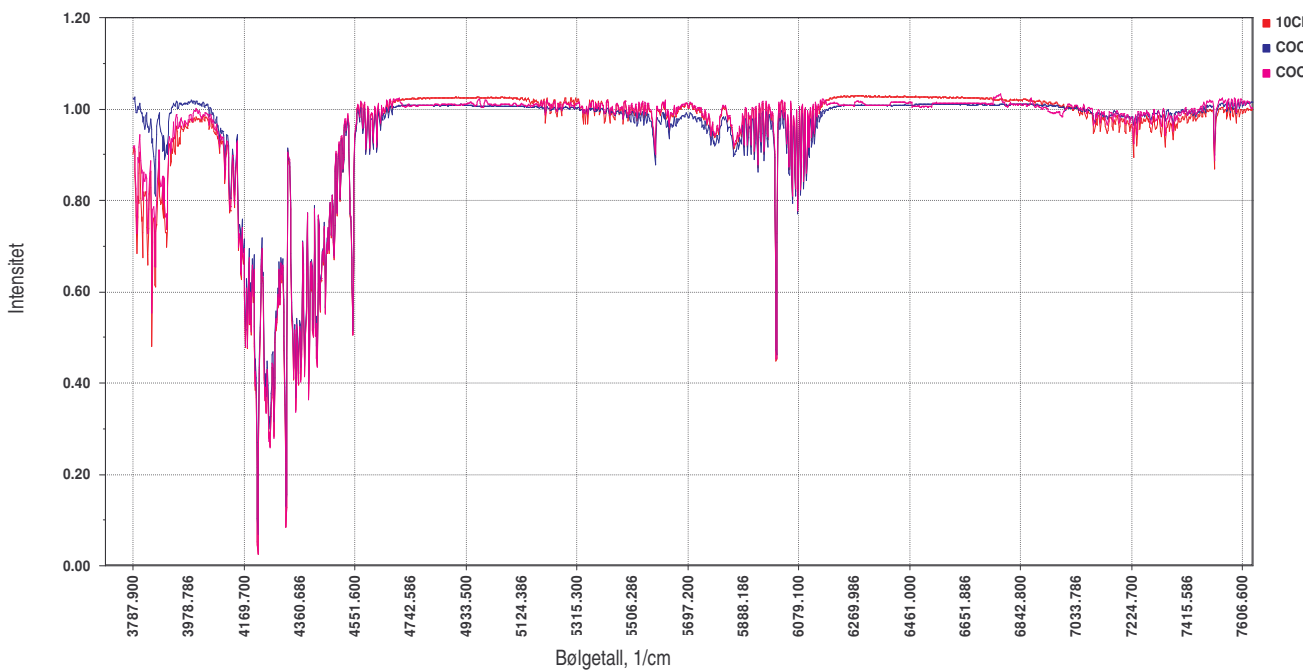
Det er dårlige resultater for objektene til CO2- lav fra celle 2, og objektene til CO- lav fra celle 1. Ellers er det ganske gode resultater. Alle nye spektre, både fra celle 1 og celle 2, må standardiseres dersom denne standardiserings- metoden skal brukes for fremtidige spektre.

Meningen med GLSW- standardisering er å bygge modeller på standardiserte data. Resultatene ble allikevel rimelig gode ved å benytte modeller som var bygget før standardiseringen, og det er derfor rimelig å tro at resultatene kunne blitt ytterligere forbedret dersom en hadde fått tid til å bygge modeller.

5.7.2 PDS- standardisering

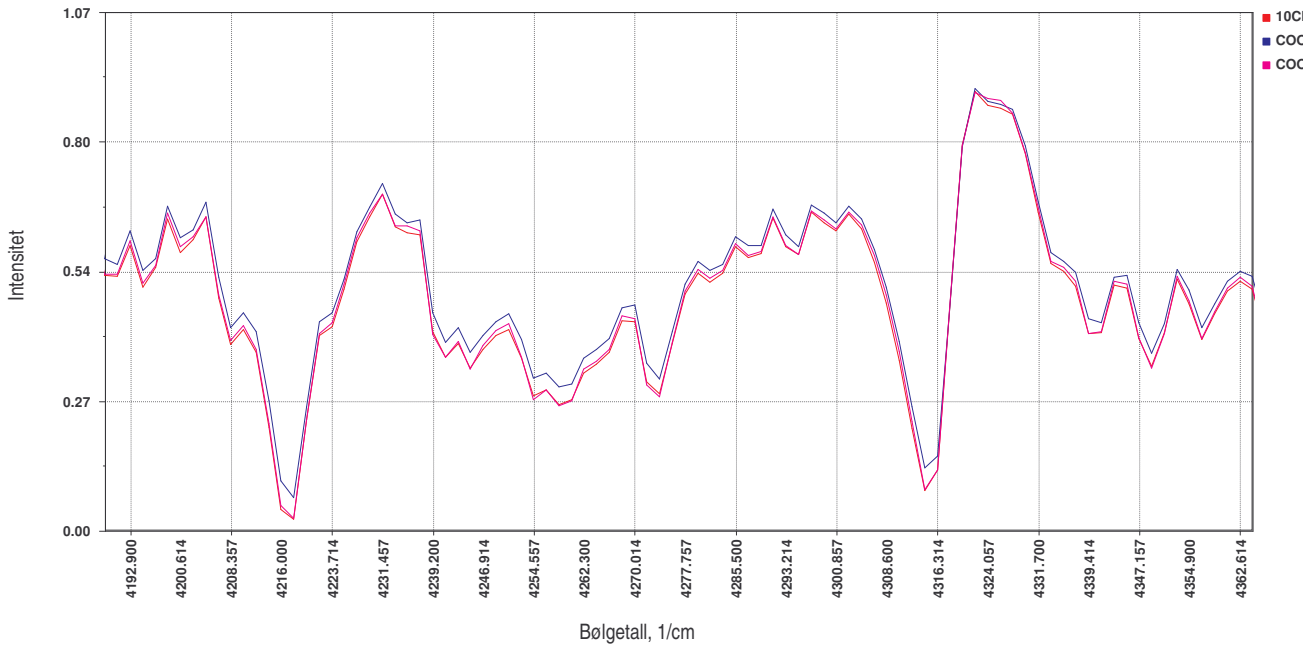
Ved PDS- standardisering prøver en å relatere responsen fra instrumentet som skal standardiseres til responsen av hovedinstrumentet. Flere elementer er blitt undersøkt ved PDS- standardiseringen. Det første som ble prøvd var å benytte hele bølgetall- området som er benyttet, det vil si fra 3787 cm^{-1} til 7641 cm^{-1} . Deretter ble bare de områdene som gjelder for modellene benyttet. Dermed er det utført tre standardiseringer. En siste del som er prøvd, var å la celle 2 være hovedinstrumentet og celle 1 instrumentet som skulle standardiseres. Det vil forklares senere.

Først standardiseres hele bølgetall- området for celle 2. Effekten av det vises i figur 5.37. Figuren viser spektre av samme gassblanding (høy $[\text{CH}_4]$ med lav $[\text{CO}]$) med samme overtrykk fra celle 1, celle 2 og fra den standardiserte celle 2. Det røde spekteret er fra celle 1, det blå er fra celle 2 og det rosa stammer fra den standardiserte celle 2. Det standardiserte spekteret fra celle 2 er som ønsket og forventet mer sammenfallende med spekteret fra celle 1, enn det originale spekteret fra celle 2 er. I figur 5.38 vises kun det bølgetallsområdet som inngår i modellen for CH_4 .



Figur 5.37: Spektre av samme gassblanding med samme overtrykk, for celle 1(rød), celle 2(blå) og standardisert celle 2(rosa).

-----Resultater-----



Figur 5.38: Bølgetallsområdet som er med i modellen for CH₄.

De standardiserte spektrene fra celle 2 tilpasses deretter til modellene, og i tabell 5.19 er resultatene vist.

Tabell 5.19: Tilpassing av standardiserte spektre til modellene fra celle 1.

Modell	% - avvik celle 1		% - avvik celle 2		% - avvik celle 2 standardisert	
	Gjennomsnittlig	Max	Gjennomsnittlig	Max	Gjennomsnittlig	Max
CH ₄ - høy	1.5	9.5	11.0	17.5	3.0	6.0
CH ₄ - lav	2.5	8.4	11.7	17.8	16.9	36.3
CO ₂ - høy	3.0	10.9	9.6	12.6	6.6	15.2
CO ₂ - lav	2.8	10.0	1.1	3.1	43.5	109.0
CO- høy	4.0	17.1	10.9	17.7	8.8	10.4
CO- lav	4.2	24.6	23.1	63.3	21.3	93.0

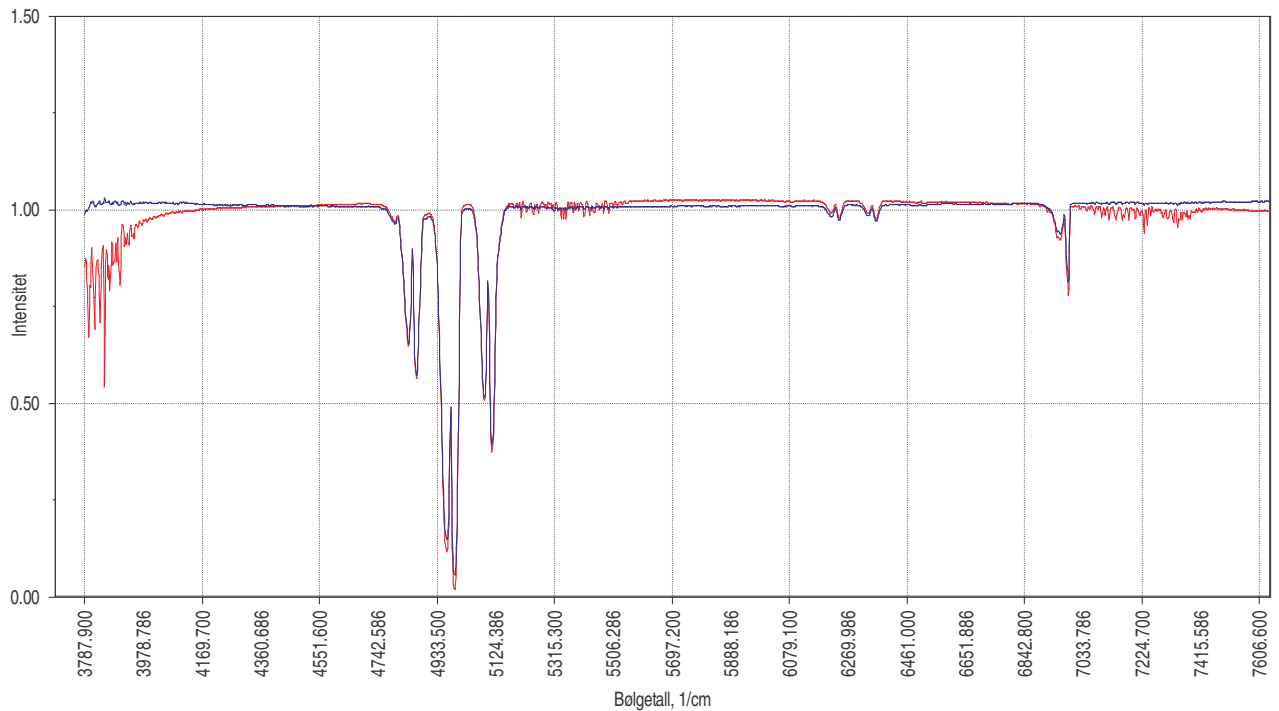
For den standardiserte celle 2 er det spesielt objekter med bare lav konsentrasjon av CH₄ (CH₄iN₂) som får store avvik i modellen for CH₄- lav. For CO₂- lav er det lavresponsobjektene som gir dårlige resultater. For CO- lav er det spesielt objekter med en blanding av høy konsentrasjon av CH₄ og lav konsentrasjon av CO som gir store avvik.

Sammenlignet med rådata fra celle 2 går de gjennomsnittlige avvikene ned for 4 modeller og opp for 2 modeller. For 3 av 6 modeller går det maksimale avviket opp. Dette er ikke ensartete resultater, og derfor ble det undersøkt om standardisering av kun modell- områdene ville gi bedre resultater. Resultatene er vist i tabell 5.20.

Tabell 5.20: Standardisering av bare modell- områder.

Modell	% - avvik celle 1		% - avvik celle 2		% - avvik celle 2 standardisert	
	Gjennom- snittlig	Max	Gjennom- snittlig	Max	Gjennom- snittlig	Max
CH4- høy	1.5	9.5	11.0	17.5	3.5	7.5
CH4- lav	2.5	8.4	11.7	17.8	12.7	27.5
CO2- høy	3.0	10.9	9.6	12.6	7.5	14.0
CO2- lav	2.8	10.0	1.1	3.1	13.4	42.4
CO- høy	4.0	17.1	10.9	17.7	5.9	9.9
CO- lav	4.2	24.6	23.1	63.3	19.1	61.9

Sammenlignet med rådata fra celle 2 går avviket ned for 4 modeller og opp for 2 modeller. Maksimale avvik går opp for 3 modeller. Avvikene er derimot bedre for denne standardiseringen enn ved å standardisere hele bølgetallsområdet. Dette kan forklares med at avvikene mellom spektrene fra de to målecellene er størst i områder som ikke er med i modellene, og standardiseringsmetoden har derfor større avvik å ta hånd om ved bruk av hele bølgetallsområdet. Dette vises i figur 5.39.



Figur 5.39: Spekter av ren CO₂ med samme overtrykk. Celle 1=rød, celle 2=blå.

I hver ende av spekteret fra celle 1 er det signal som ikke vises i spekteret fra celle 2. Spekteret fra celle 2 har mindre intense eller ingen signaler i endene som forventet. Årsaken til dette kan være ukontrollerte mengder vanddamp i celle 1. Når en bygger modeller vil en ta utgangspunkt i spektre med minst mulig støy. Siden spektrene fra celle 2 er mindre støyfulle enn spektrene fra celle 1, ble det forsøkt å bygge modeller på spektre fra celle 2, og standardisere celle 1. De samme bølgetallsområder som for modellene fra celle 1 ble benyttet, og alle objektene som var relevante for hver modelltype. Dette gjelder, CO₂- høy har objekter med høy konsentrasjon av CO₂ i tillegg til en del objekter med 0 CO₂, CO₂- lav har objekter med lav konsentrasjon av CO₂ i tillegg til en del objekter med 0 CO₂. Resultatene er vist i tabell 5.21.

Tabell 5.21: Modeller på celle 2, celle 1 standardisert.

Modell	% - avvik celle 1		% - avvik celle 2		% - avvik celle 1 standardisert	
	Gjennom- snittlig	Max	Gjennom- snittlig	Max	Gjennom- snittlig	Max
CH4- høy	8.1	24.3	1.5	2.9	3.7	17.1
CH4- lav	11.8	24.9	1.2	24.9	19.7	56.0
CO2- høy	10.9	29.1	2.0	5.7	5.6	37.2
CO2- lav	6.6	37.0	1.7	4.7	76.0	225
CO- høy	8.7	16.7	2.3	4.0	3.7	12.6
CO- lav	18.8	65.5	4.7	20.4	9.1	36.8

Modellering på celle 2, og standardisering av celle 1 gir dårligere resultater. De gjennomsnittlige avvikene er bedre for 4 modeller etter standardisering av celle 1, men det vil være umulig å predikere lave konsentrasjoner av CO₂.

Ved å legge til et par flere objekter med blanding av høy konsentrasjon av CH₄ og lav konsentrasjon av CO til modellen for CO- lav som er bygget på celle 1, og deretter predikere på både standardisert og ikke- standardisert celle 2 blir resultatene litt bedre.

Resultatene av å tilpasse alle objektene til de nye modellene er vist i tabell 5.22.

Tabell 5.22: Bedre resultater for modeller på celle 1 og standardiser celle 2

Modell	% - avvik celle 1		% - avvik celle 2		% - avvik celle 2 standardisert	
	Gjennom- snittlig	Max	Gjennom- snittlig	Max	Gjennom- snittlig	Max
CH4- høy	1.5	9.5	11.0	17.5	3.5	7.5
CH4- lav	2.5	8.4	11.7	17.8	12.7	27.5
CO2- høy	3.0	10.9	9.6	12.6	7.5	14.0
CO2- lav	2.8	10.0	1.1	3.1	13.4	42.4
CO- høy	4.0	17.1	10.9	17.7	5.9	9.9
CO- lav	10.7	42.6	34.0	71.0	16.2	65.1

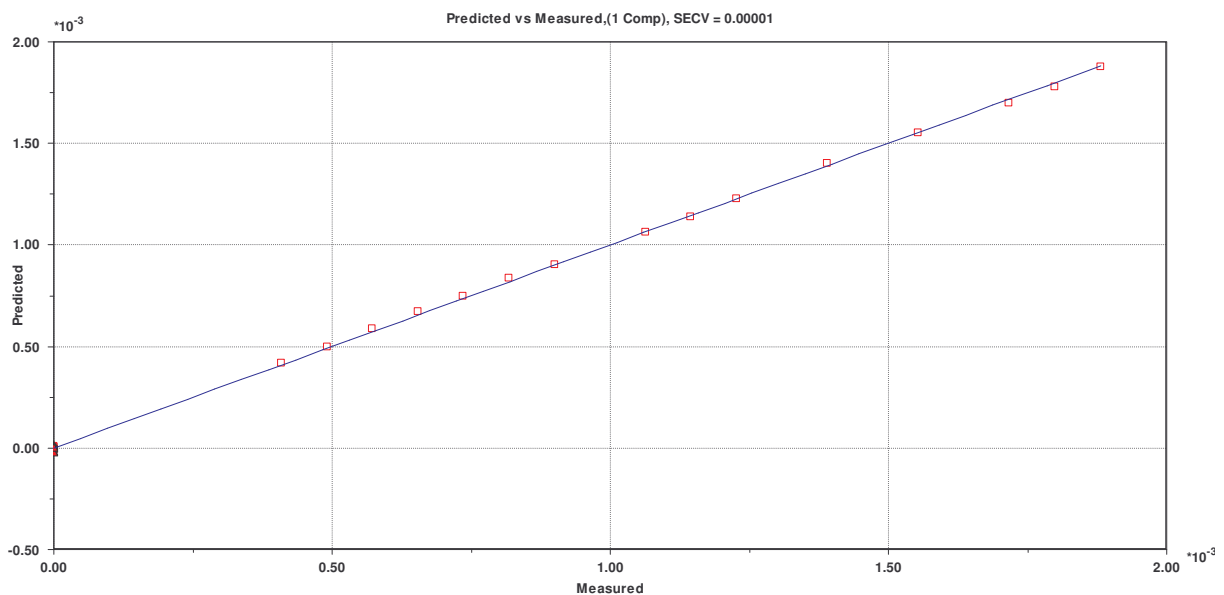
Resultatene blir bedre for CO₂-lav sammenlignet med tidligere, se tabell 5.20: Gjennomsnittlig prosentavvik går ned fra 19.1 til 16.1 for standardisert celle, og det er 4 objekter av 24 som får prosentvis avvik over 30%. Gjennomsnittlig avvik for celle 1 går litt opp, men 10.7% gjennomsnittlig avvik akseptabelt.

Plott av predikert respons mot målt respons for disse tilpassingene for tabell 5.22 er gitt i Appendix B.1.

De beste resultatene oppnådd ved å velge å standardisere med PDS, er å benytte celle 1 som hovedinstrument og standardisere celle 2.

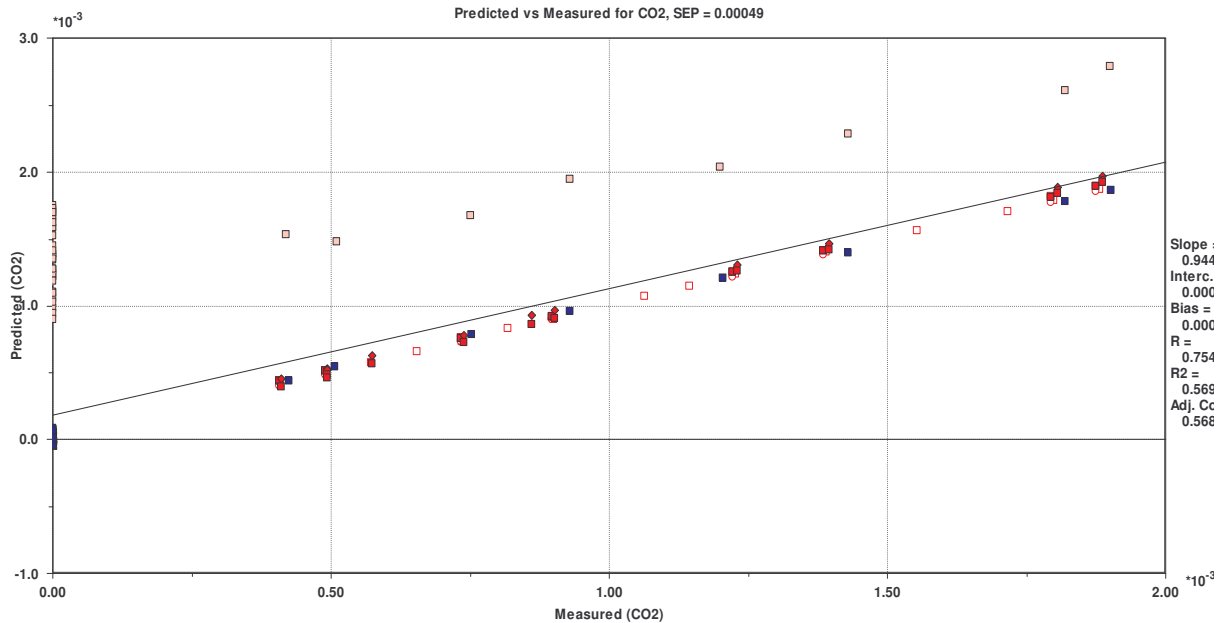
5.7.3 Derivasjon

For å teste effekten av derivasjon på spektrene, prøves det å benytte første- og andreordensderivasjon på CO₂-lav sine spektrere, som er standardisert med PDS. 1. grads derivasjon prøves først. En effekt som observeres er at antall nødvendige PLS-komponenter synker til 1, samtidig som 99.97% av variansen forklares. Objektene i modellen danner en rett linje og det viser at modellen er god, se figur 5.40:



Figur 5.40: Predikert respons mot målt respons for modellobjektene til CO₂-lav etter 1. ordens derivasjon.

For objektene som tilpasses ser situasjonen ut som i figur 5.41:



Created: 05/29/08 18:05:48

Figur 5.41: Predikert respons mot målt respons for alle CO₂- lav- objekter og 0-responsobjekter. Rosa firkanter=standardiserte celle 2- objekter, blå firkanter= ikke-standardiserte celle 2- objekter, røde firkanter=celle 1- objekter.

Prediksjoner for standardiserte objekter har alt for store avvik.

For 2. ordens derivasjon er situasjonen tilsvarende.

En kan konkludere med at derivasjon ikke gjør modellen for CO₂- lav bedre.

6. Konklusjoner og diskusjon

6.1 Undersøkte problemer

Noen av de potensielle problemene som har blitt undersøkt er:

- *Bakgrunnseffekter:* Endres bakgrunnsspektrene vesentlig over tid? Hvilken effekt har forskjell i gasstrykk mellom bakgrunn og spekteret som korrigeres?
- *Strømningseffekter:* Vil prediksjoner av spektre tatt opp med strømmende gass være tilfredstillende og kan slike spektre korrigeres med bakgrunns-spektre tatt opp med stillestående gass?
- *Modellbygging:* Vil absorpsjonsbåndene fra vanndamp overskygge signalene til CO₂, CH₄ og CO i det nær- infrarøde området?
- *Temperatureffekter:* Vil toppintensiteter, båndbredder og topp- posisjoner endres vesentlig ved økt temperatur? Hvilke effekter vil det ha på modellering og prediksjon?
- *Kalibreringsoverføring:* Vil forskjellene mellom målecellene være store? Er det behov for instrumentstandardisering?

6.1.1 Bakgrunnseffekter

Første bakgrunnseffekt som ble undersøkt var om det var nødvendig å ta opp bakgrunnsspektre ved alle trykk som benyttes i oppgaven. I tabell 5.3 (se s. 65) vises at det ikke var noen tendens til at spektre som ble korrigerert med bakgrunnsspektre tatt opp med vidt forskjellig trykk, hadde større avvik enn andre prøver. Det konkluderes dermed med at det

bare trengs å ta opp ett bakgrunnsspekter, for eksempel med et trykk midt mellom maksimalt overtrykk og 0 overtrykk.

Andre bakgrunnseffekter som ble undersøkt var å bakgrunnskorrigere et spekter med et bakgrunnsspekter tatt opp flere måneder i forveien. Tabell 5.4 (se s. 66) viser at det for enkelte modeller blir store avvik, for andre ikke så store. Etter å ha endret modellene, viste det seg at disse spektrene ikke hadde noe vesentlig høyere avvik enn andre spektre.

Konklusjonen blir at enkelte modeller klarer å fange opp spektre som er korrigert med bakgrunnsspektre tatt opp flere måneder før, mens noen klarer det ikke. Derfor bør spektre korrigeres med bakgrunnsspektre tatt opp på om lag samme tidspunkt.

6.1.2 Strømningseffekter

Strømningseffektene ble undersøkt ved å korrigere spektre tatt opp med gass strømmende gjennom cellen, med bakgrunnsspektre av stasjonær N₂- gass. Tabell 5.5 (se s. 68) viser at avvikene til disse objektene i forhold til strømningsobjekter som er korrigert med bakgrunnsspektre som er tatt opp med gasstrømning, ikke er signifikante. Det er heller ikke signifikant forskjell i forhold til vanlige objekter, som er tatt opp uten strømning og korrigert med bakgrunnsspektre tatt opp uten strømning.

Konklusjonen blir dermed at effektene av strømmende gass kan neglisjeres; spektre tatt opp med gasstrømning kan bakgrunnskorrigeres med bakgrunnsspektre tatt opp uten gasstrømning.

6.1.3 Vann sin innvirkning på modellene

For å undersøke vanndamp sin innvirkning på modellene simulerte Jon Oddvar Hellevang ved CMR en del situasjoner som skal tilsvare noen av gassblandingene som er benyttet, i tillegg til å legge inn vanndamp sine signaler. Disse er vist i figur 5.13, 5.14, 5.15 og 5.16, henholdsvis

s. 70, 71, 72 og 73. Figurene viser at vanndamp mest sannsynlig vil overskygge enkelte deler av modellområdene som på det tidspunkt var valgt. Områdene ble dermed justert for alle modellene. Signaler fra vanndamp overskygger delvis signalene til henholdsvis CO₂, CO og CH₄. Modeller bygget på de reduserte områdene gav ikke signifikant dårligere resultater enn de originale modellene. Med mindre de simulerte spektrene er feile, eller forholdene i målecellen blir annerledes enn de som er simulert, vil ikke de intense signalene fra vanndamp redusere presisjonen og nøyaktigheten i prediksjonene.

6.1.4 Temperatureffekter

Tabell 5.9 (se s. 76) viser at det ikke er store forskjeller mellom avvikene til objekter med romtemperatur og avvikene til objekter med høy temperatur. Det er allikevel en tendens til større avvik for objekter med høy temperatur enn for romtemperatur. På dette tidspunkt var det uvisst om en skulle benytte målecellen ved enda høyere temperaturer. Responsene ble derfor omformet til konsentrasjon, og modeller basert på høytemperatur- objekter ble bygget. Disse modellene ga gode resultater.

6.1.5 Kalibreringsoverføring

Figur 5.31 og 5.32 (se s. 93) viser at det er litt forskjell mellom de to målecellene. Tabell 5.17 (se s. 91) viser at det er litt avvik for enkelte modeller når modellene fra celle 1 benyttes på celle 2. Dette gjelder spesielt for modellen for CO- lav. GLSW- og PDS- standardisering er blitt benyttet for å forsøke å forbedre resultatene.

GLSW- standardiseringen er ikke blitt utført på den måten den er tenkt. Modeller skal bygges etter standardiseringen, her er modellene som ble bygget før standardiseringen benyttet. Llikevel gir standardiseringen bare litt dårligere resultater enn for PDS, se tabell 5.18 (s. 96). Sannsynligvis hadde resultatene blitt bedre dersom en hadde fått tid til å bygge nye modeller. En må med denne standardiserings- metoden standardisere alle nye spektrere som blir tatt opp, både fra celle 1 og celle 2.

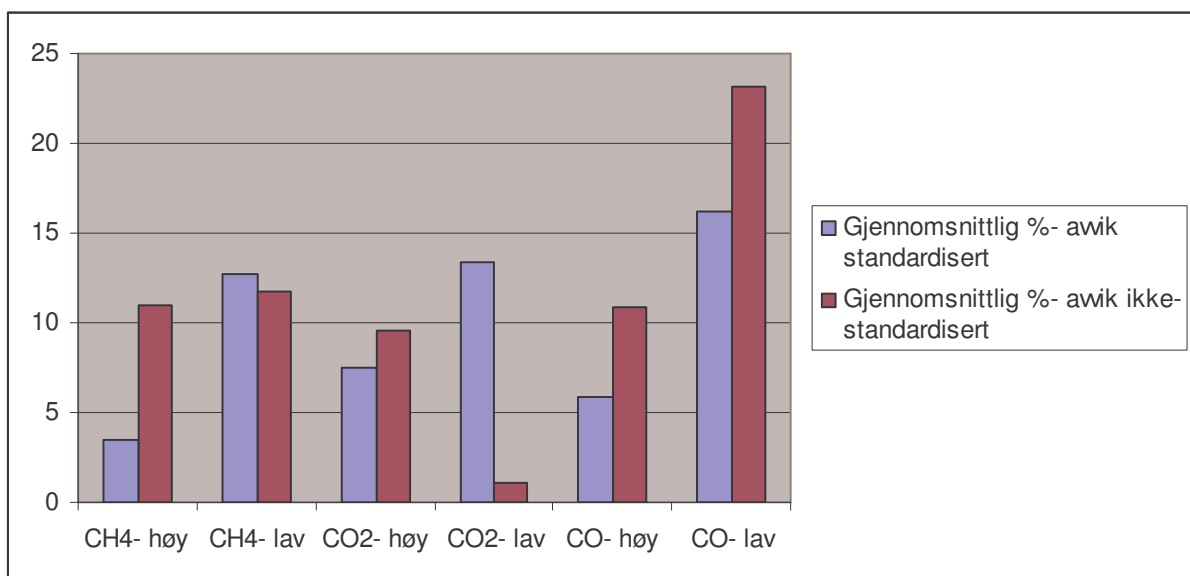
Resultater fra PDS- standardisering er vist i tabell 5.22 (se s. 102). Det blir dårlige resultater for CO- lav og CO₂- lav, med 16% gjennomsnittlig avvik og 65% maksimalt avvik for CO- lav som de dårligste. 10% avvik er den ønskete grensen, men det bør gå an å bruke modellene.

6.2 Konklusjon

Hoveddelen av denne oppgaven gikk ut på å undersøke om det lar seg gjøre å lage PLS-modeller for å predikere konsentrasjoner av henholdsvis CH₄, CO₂ og CO i forskjellige blandinger.

For celle 1 får en samme resultater ved å standardisere celle 2 som ved å ikke standardisere, med unntak av objekter med lave konsentrasjoner av CO. Modellene for CO- lav er ulik i de to tilfellene. Gjennomsnittlig prosentavvik går fra 4.2% for den opprinnelige modellen til 10.2% for modellen som er endret for standardiseringen.

For celle 2 ser resultatene ut som i figur 6.1:



Figur 6.1: Sammenligning av gjennomsnittlig prosentavvik for celle 2 for standardiserte og ikke- standardiserte spektre.

Figur 6.1 viser at ved å standardisere spektrene får 4 modeller bedre resultater.

En bør derfor standardisere målecelle 2. Ved å standardisere hvert framtidige spekter tatt opp på celle 2 blir responsen relatert til responsen av hovedinstrumentet. Dermed unngår en dårligere prediksjoner på grunn av små endringer i spektre fra celle 2.

PLS- modeller for å predikere høye konsentrasjoner av henholdsvis CH₄, CO₂ og CO gir gode resultater. PLS- modeller for å predikere lave konsentrasjone gir dårligere resultater, men bør være anvendbare.

Det lot seg gjøre å lage PLS- modeller for prediksjon av henholdsvis CH₄, CO₂ og CO.

6.3 Videre arbeid

Figur 5.31 og 5.32 viser at det er forskjeller mellom spektre av like gassblandinger ved samme overtrykk, tatt opp på de to målecellene. Spektrene fra celle 2 ser ut til å inneholde mindre støy enn spektrene fra celle 1. Det optimale hadde vært om begge cellene hadde hatt tilnærmet like støyforhold, mindre støy ville dermed blitt tatt med i modellene.

Årsaken til at celle 1 har disse støyområdene som celle 2 ikke har bør finnes. Dersom celle 1 endres kan det være at en må sette igang en ny prosess med spekteropptakt og modellbygging. Det er kanskje ikke ønskelig. Dersom dette ikke kan gjøres, kan det om nødvendig bygges nye modeller av eksisterende spektra.

Referanser

1. <http://www.cmr.no/avd10/project-ZEG.shtml>, april 2008.
2. <http://www.naturalgas.org/overview/background.asp>, mai 2008.
3. http://no.wikipedia.org/wiki/Visuelt_spektrum, april 2008.
4. <http://www.geo.mtu.edu/rs/back/spectrum/>, april 2008.
5. <http://science.hq.nasa.gov/kids/imagers/ems/infrared.html>, april 2008.
6. <http://www.spectroscopynow.com/coi/cda/detail.cda?id=1881&type=EducationFeature&chId=2&page=1>, april 2008.
7. <http://www.skretting.no/web/skrettingnorway/Interweb.nsf/wPrIdPrint/3D5689CB6A7EF983C1256C7E00502CF7?OpenDocument>, april 2008.
8. http://www.kjemi.uio.no/09_spekt/spektr.htm, april 2008.
9. A. A. Christy, FT- IR/NIR- spectroscopy in process analysis. Theory practice and data processing, Dept. Of chemistry, University of Bergen, 1994.
10. Donald P. Lavia, Gary M. Lampmann, George S. Kris: Introduction to spectroscopy, third edition, 2001, Washington.
11. Dudley H. Williams, Ian Fleming: Spectroscopic methods in organic chemistry, 5th edition, McGraw- Hill publishing Company (1995).
12. P.W. Atkins: Physical Chemistry, 6th edition, Oxford University, 1999.
13. H.D. Young, R.A. Freedman: University Physics with Modern Physics, 11th Ed., Pearson, Addison Wesley 2004.
14. N.Makhoukhi, E. Péré, R. Creff, C. Pouchan: Determination of the composition of gases by infrared analysis and chemometric methods. Journal of Molecular Structure vol. 744-747 (2005), s. 855-859.
15. [http://library.abb.com/global/scot/scot205.nsf/veritydisplay/caf7c25fc0f837c585256cbd0049f888/\\$File/generalspectroscopy.pdf](http://library.abb.com/global/scot/scot205.nsf/veritydisplay/caf7c25fc0f837c585256cbd0049f888/$File/generalspectroscopy.pdf), mai 2008
16. <http://www.kj.uib.no/ft-ir-norsk/>, april 2008.
17. J. Bak, A. Larsen: Quantitative analysis of the gases CO, CO₂ and CH₄ by Fourier Transform Infrared (FTIR) spectroscopy, *Dansk Kemi*, volume 75, number 8, 16-21 1994
18. Maria A. van Agthoven, Go Fujisawa, Philip Rabbito, and Oliver C. Mullins, Near-Infrared Spectral Analysis of Gas Mixtures, *Applied Spectroscopy*, Volume 56, Number 5, 2002.

19. Egil Nodland: Multivariate Analysis of Infrared Profiles: On- line Techniques, Kinetics and Effects of Temperature, Department of Chemistry, University of Bergen, 1995.
20. G. Dijkstra: The influence of temperature variations on the intensities of absorption bands in the infra- red spectra of aliphatic compounds, *Spectrochimica Acta*, 11 (1957) 618-622.
21. J. E. Keaton, S. R. Lobo and J. C. Simpson: The temperature dependence of band maxima molar absorptivities of the carbon- halogen stretching mode in tertiary butyl chloride, bromide and iodide in the liquid phase, *Applied Spectroscopy*, 41 (1987) 468-471.
22. J. J. Orlando, G. S. Tyndall, A. Huang and J. G. Calvert, *Geophys. Res. Lett.*, 19 (1992) 1005-1008.
23. R. A. Nyquist: The significance of temperature and solvent effects upon infrared spectra and group frequencies, *Applied Spectroscopy*, 40 (1986) 79-85.
24. Raymond Chang: General Chemistry, 3rd edition, Williams College, 2003
25. R. Nordtvedt, F. Brakstad, O. M. Kvalheim, T. Lunstedt. Anvendelse av kjemometri innen forskning og industri, Tidsforlaget Kjemi AS, 1996.
26. A. Savitzky, M. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry* (1964), 36: 1627-1639.
27. H. Martens, T. Næs. Multivariate Calibration, John Wiley & Sons (1991).
28. R. Manne: Analysis of two partial least- squares algorithms for multivariate calibration, *Chemometrics and Intelligent Laboratory Systems*, 2, 187-197, 1987
29. http://www.thermo.com/com/cda/resources/resources_detail/1,2166,13429,00.html, mai 2008.
30. Kim H. Esbensen: Multivariate Data Analysis in Practice, 4th edition, CAMO ASA (1990)
31. R. Kramer: Chemometric Techniques for Quantitative Analysis, Marcel Dekker Inc, (1998)
32. Wang, Y., D.J. Velkamp and B.R. Kowalski: Multivariate Instrument Standardization, *Anal. Chem.*, **63** (23), 2750-2756, 1991.
33. Wang, Y., M.J. Lysaght and B.R. Kowalski: Improvement of Multivariate Calibration through Instrument Standardization, *Anal. Chem.*, **64** (5), 562-565, 1992.
34. Wang, Z., T. Isaksson and B.R. Kowalski: New Approach for Distance Measurement in Locally Weighted Regression, *Anal. Chem.*, **66**, 249-260, 1994.
35. <http://www.eigenvector.com/>, mai 2008.
36. <http://www.mathworks.com/>, mai 2008
37. Barry M. Wise, Harald Martens, Martin Høy, Rasmus Bro, Per B. Brockhoff: Calibration Transfer by Generalized Least Squares, fra <http://citeseer.ist.psu.edu/596709.html> mai 2008.

38. <http://www.prs.no/>, mai 2008.

39. <http://www.ontar.com/>, mai 2008.

40. Go Fujisawa, Maria A. Van Agthoven, Fredrick Jenet, Philip A. Rabbito, and Oliver C. Mullins: Near- Infrared Compositional Analysis of Gas and Condensate Reservoir Fluids at Elevated Pressures and Temperatures, *Applied Spectroscopy*, Volume 56, Number 12, 2002.

Appendix

Appendiks A – Algoritmer

A.1: NIPALS algoritme for PCA

I NIPALS algoritmen benyttes kriteriet om maksimalvariasjon. Skal tilpasse modellen:

$$\mathbf{X} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a + \mathbf{E} = \mathbf{TP}^T + \mathbf{E}$$

Sentrerer \mathbf{X} : $\mathbf{X}_s = \mathbf{X} - \mathbf{1}\mathbf{x}^{-T}$

1) Velger startverdi \mathbf{t}_a =kolonnevektoren i datamatriksen med størst varians.

2) Regner ut ladninger som tilhører \mathbf{t}_a :

$$p_a = \frac{\mathbf{t}_a \cdot \mathbf{X}_{a-1}}{\mathbf{t}_a^T \cdot \mathbf{t}_a}$$

3) Normaliserer ladningene til lengde 1:

$$\mathbf{p}_a = \frac{\mathbf{p}_a}{\sqrt{\mathbf{p}_a^T \mathbf{p}_a}}$$

4) Finner bedre tilnærming av skåren:

$$\mathbf{t}_a = \mathbf{X}_{a-1} \mathbf{p}_a$$

5) Estimerer egenverdi: $\tau_a = \mathbf{t}_a^T \mathbf{t}_a$

Dersom τ_a for denne iterasjonen minus τ_a for forrige iterasjon er større enn en konstant bestemt på forhånd går en tilbake til steg 2.

6) Trekk fra effekten av denne faktoren:

$$\mathbf{X}_a = \mathbf{X}_{a-1} - \mathbf{t}_a \mathbf{p}_a^T$$

Gå tilbake til steg 1. Antallet komponenter bestemmes ved hjelp av kryssvalidering eller en annen valideringsmetode.

A.2: NIPALS algoritme for PLS- 1.

Sentrerer \mathbf{X} og \mathbf{y} :

$$\mathbf{X}_0 = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^{-T}$$

$$\mathbf{y}_0 = \mathbf{y} - \mathbf{1}\bar{\mathbf{y}}$$

For faktor $a=1,2,\dots,A$:

1) Setter startverdier for u- skårer: $\mathbf{u}_a = \mathbf{y}_{a-1}$

2) Finner ladningsvektor: $\mathbf{w}_a^T = \frac{\mathbf{u}_a^T \mathbf{X}_{a-1}}{\mathbf{u}_a^T \mathbf{u}_a}$

3) Normaliserer til lengde 1: $\mathbf{w}_a = \frac{\mathbf{w}_a}{\sqrt{(\mathbf{w}_a^T \mathbf{w}_a)}}$

4) Finner skårer for \mathbf{X} : $\mathbf{t}_a = \mathbf{X}_{a-1} \mathbf{w}_a$

5) Finner ladninger for \mathbf{X} : $\mathbf{p}_a^T = \frac{\mathbf{t}_a^T \mathbf{X}_{a-1}}{\mathbf{t}_a^T \mathbf{t}_a}$

6) Finner ladninger for \mathbf{y} : $\mathbf{q}_a^T = \frac{\mathbf{t}_a^T \mathbf{y}_{a-1}}{\mathbf{t}_a^T \mathbf{t}_a}$

7) Trekker fra effekten av denne faktoren i \mathbf{X} : $\mathbf{X}_a = \mathbf{X}_{a-1} - \mathbf{t}_a \mathbf{p}_a^T$

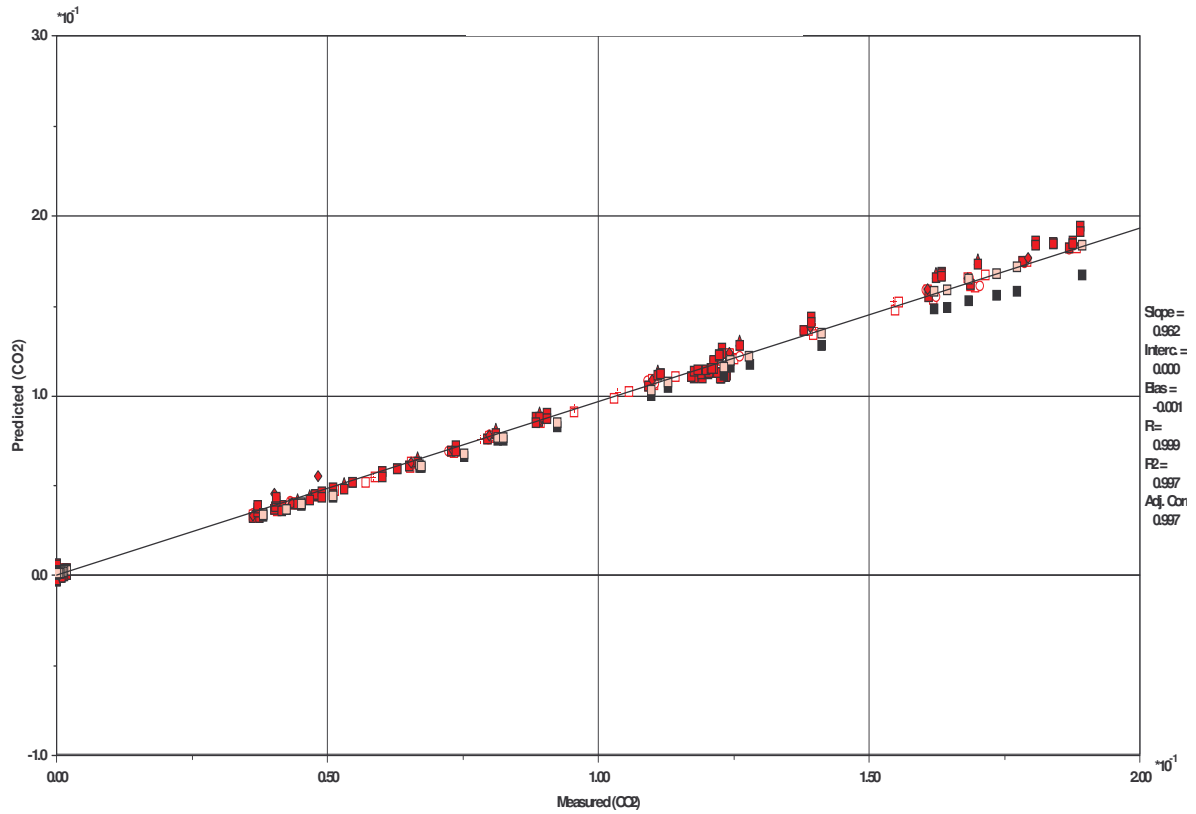
8) Trekker fra effekten av denne faktoren i \mathbf{y} : $\mathbf{y}_a = \mathbf{y}_{a-1} - \mathbf{t}_a \mathbf{q}_a^T$

Går til 1) for å estimere neste faktor.

Bruk en valideringsmetode for å bestemme optimalt antall faktorer.

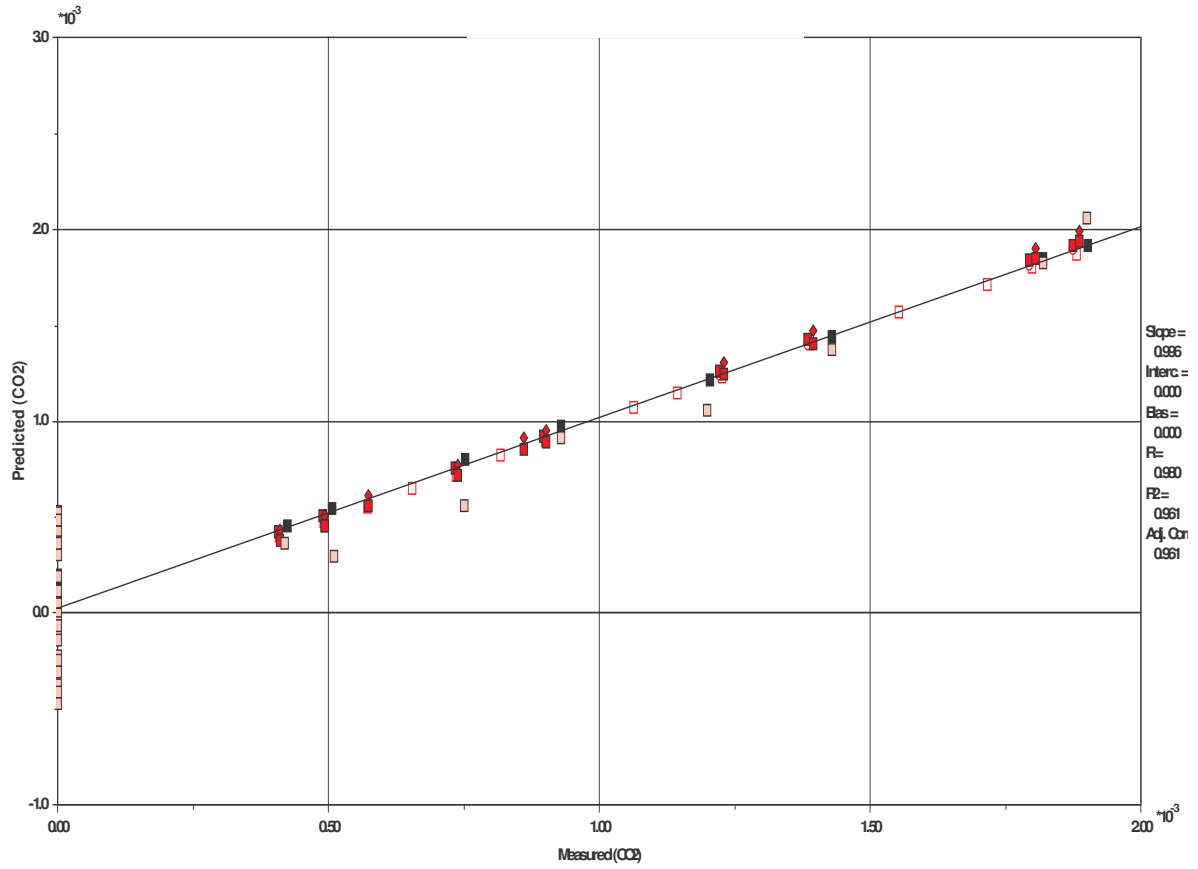
Appendix B – Plott

B.1: Plott av predikert respons mot målt respons for modeller som fører til tabell 5.22



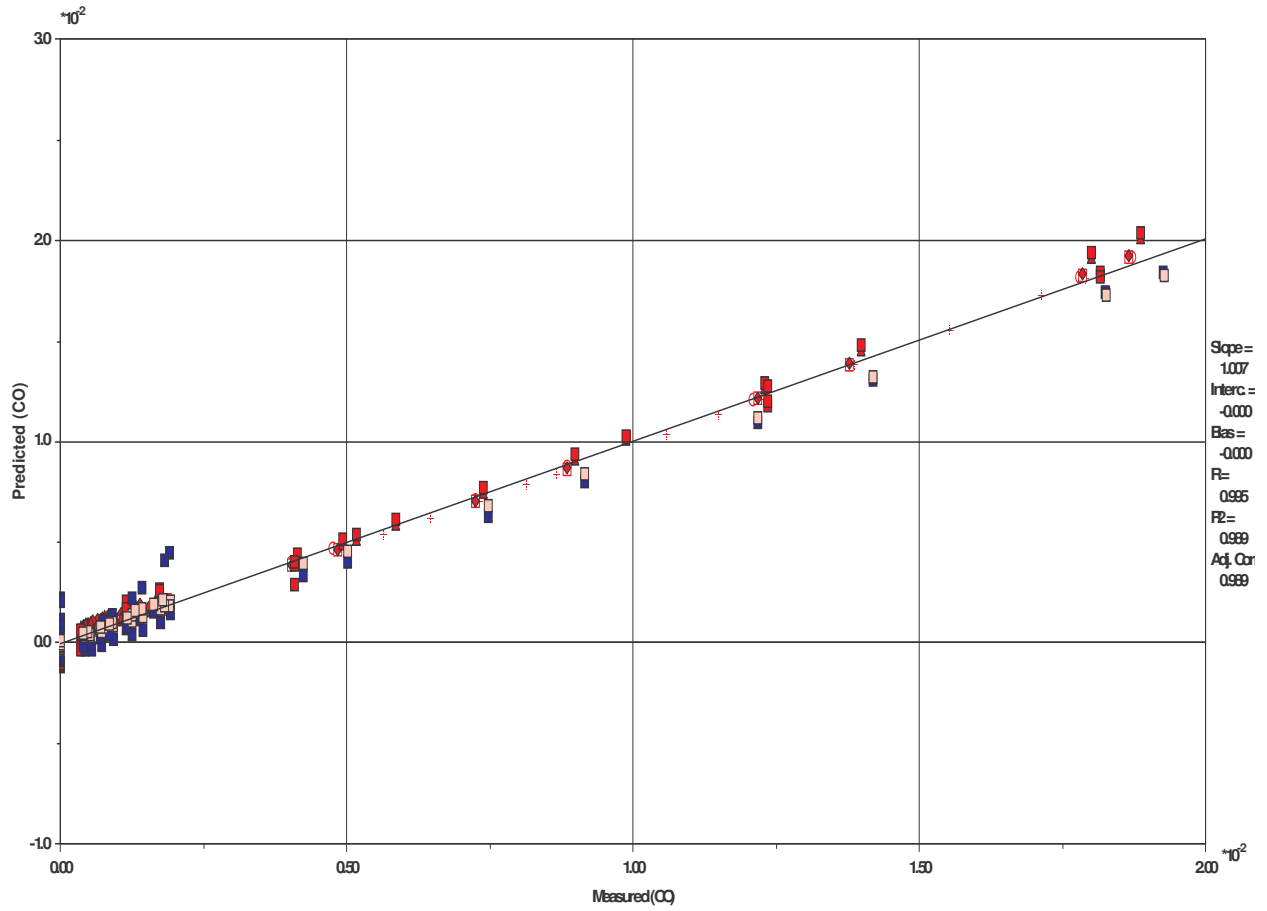
Created: 04/09/08 11:34:33

Figur B.1.1: Predikert respons mot målt respons for CO₂- høy. Røde firkanter=Celle1-objekter, svarte firkanter=celle2-objekter, rosa firkanter=standardiserte celle2-objekter.



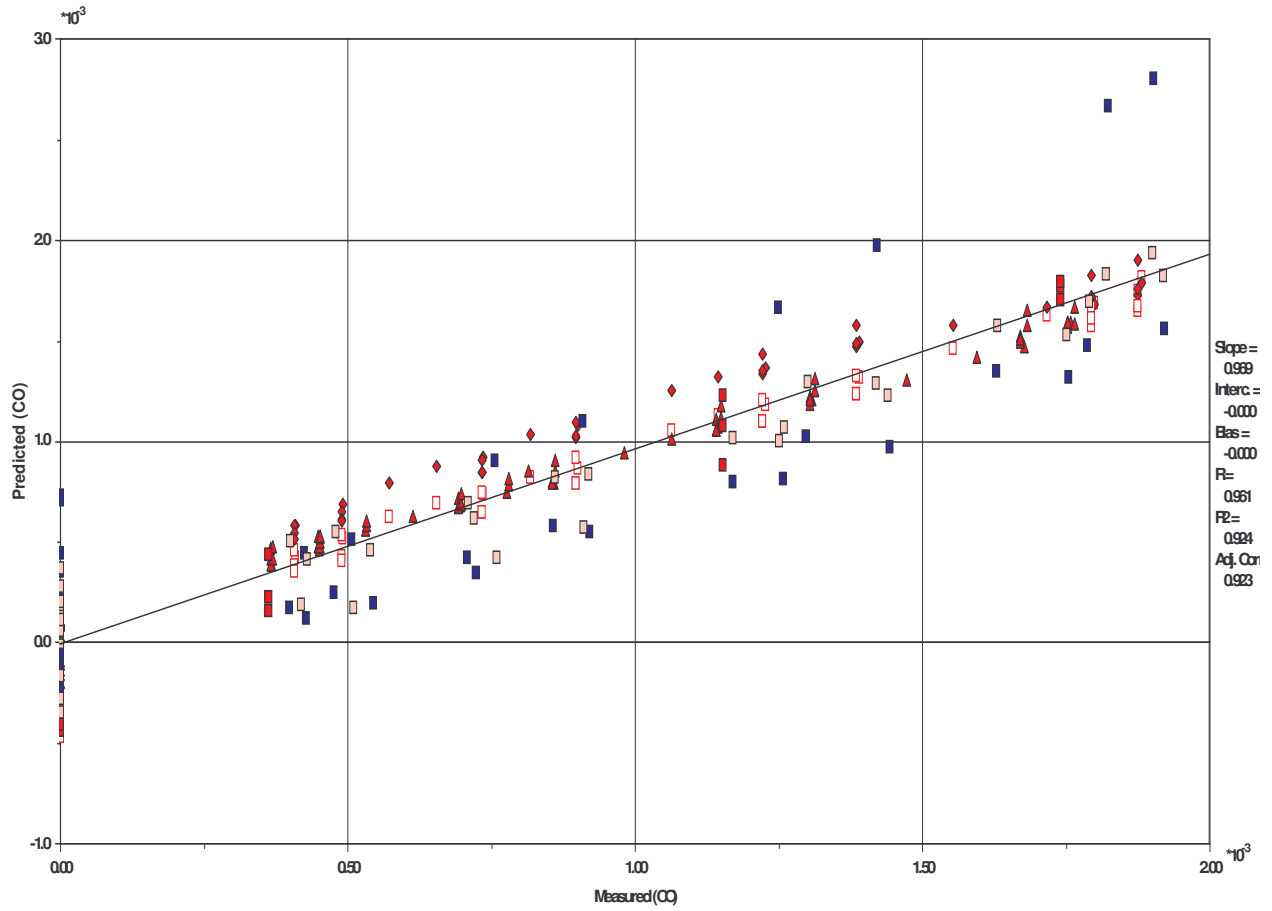
Created: 04/09/08 11:57:47

Figur B.1.2: Predikert respons mot målt respons for CO₂-lav. Røde firkanter=Celle1-objekter, svarte firkanter=celle2-objekter, rosa firkanter=standardiserte celle2-objekter.



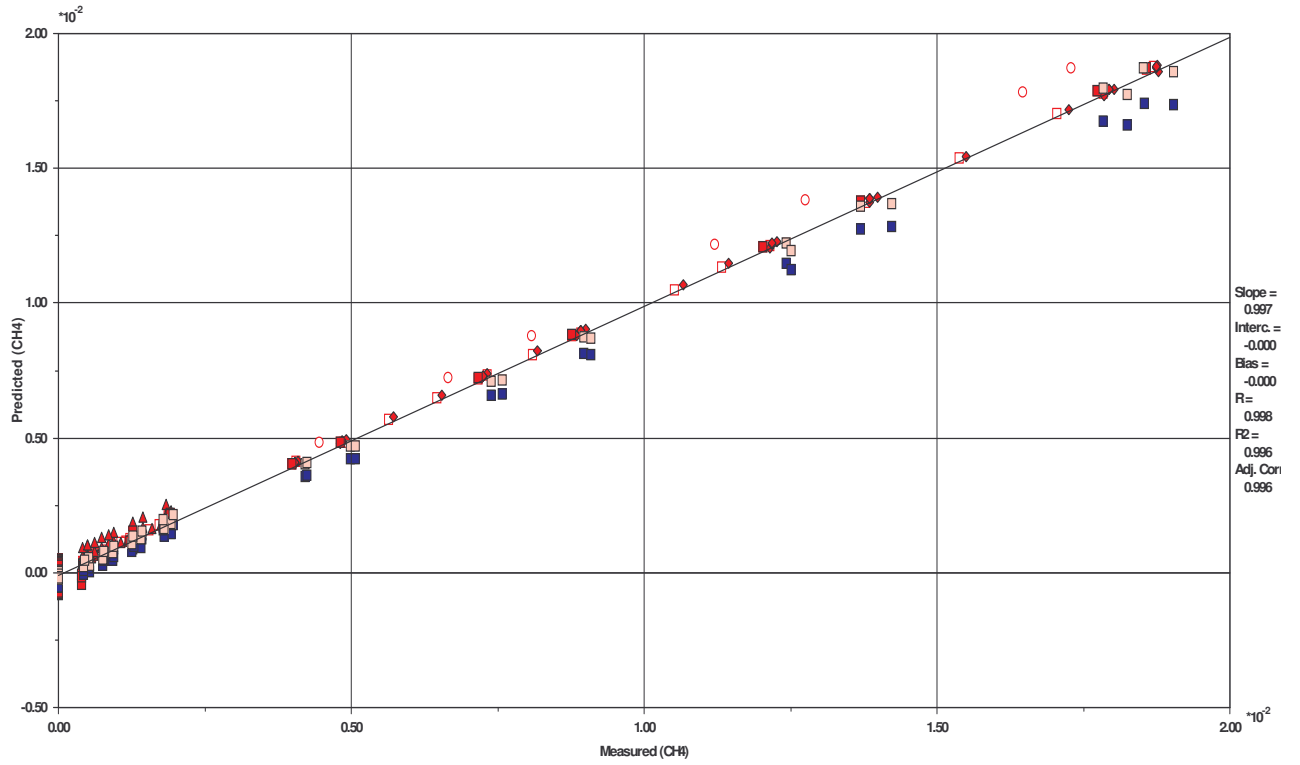
Created: 04/16/08 13:15:38

Figur B.1.3: Predikert respons målt respons for CO- høy. Røde firkanter=Celle1-objekter, blå firkanter=celle2-objekter, rosa firkanter=standardiserte celle2-objekter.



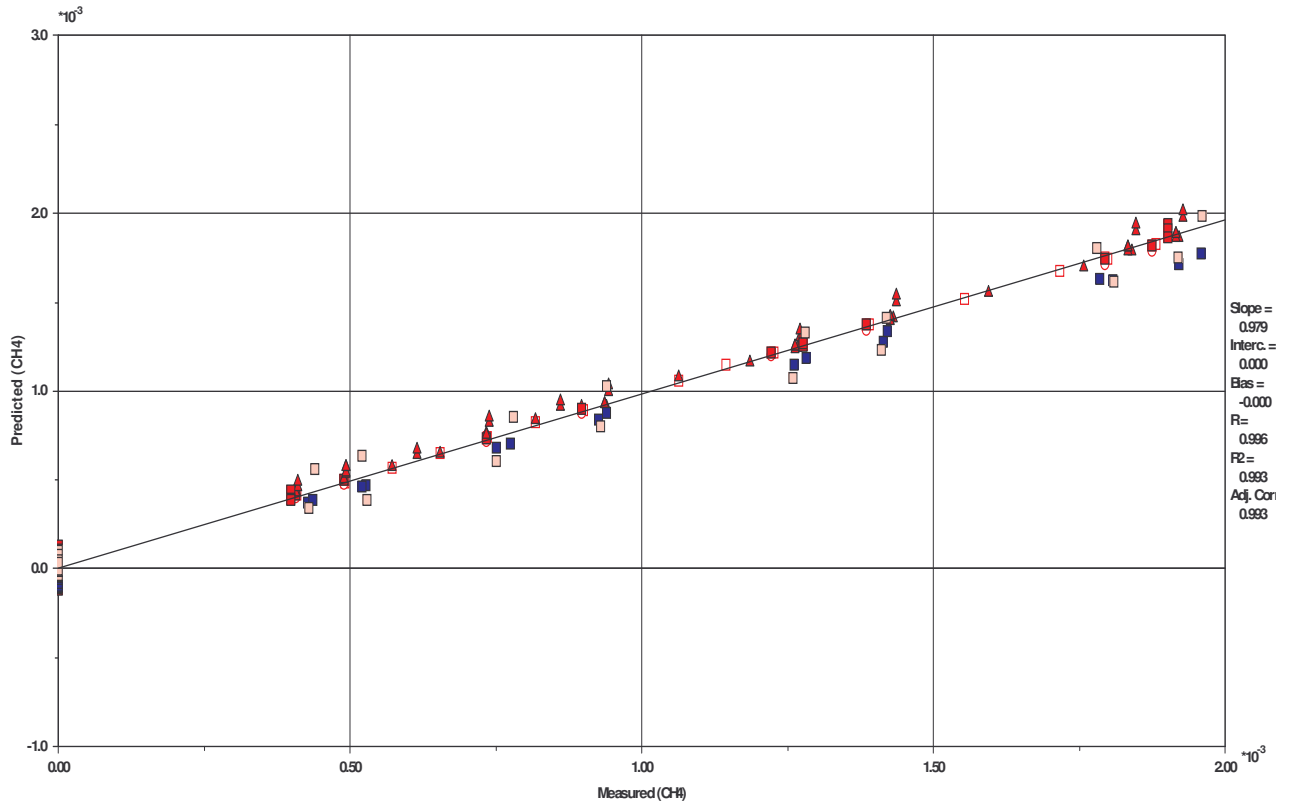
Created: 04/16/08 13:56:44

Figur B.1.4: Predikert respons mot målt respons for CO- lav. Røde firkanter=Celle1-objekter, blå firkanter=celle2-objekter, rosa firkanter=standardiserte celle2-objekter.



Created: 04/14/08 12:06:21

Figur B.1.5: Predikert respons mot målt respons for CH₄- høy. Røde firkanter=Celle1-objekter, blå firkanter=celle2-objekter, rosa firkanter=standardiserte celle2-objekter.



Created: 04/14/08 12:55:09

Figur B.1.6: Predikert respons mot målt respons for CH₄- lav. Røde firkanter=Celle1-objekter, blå firkanter=celle2-objekter, rosa firkanter=standardiserte celle2-objekter.