

PROSJEKT NR. 13
TILGJENGELIGHET Åpent

OsloMet – storbyuniversitetet

Fakultet for teknologi, kunst og design

Institutt for maskin, elektronikk og kjemi

Postboks 4 St. Olavs plass, 0130 Oslo

Telefon: 67 23 50 00

BACHELOROPPGAVE

BACHELOROPPGAVENS TITTEL Kjemiskscreening av husstøvprøver.	DATO: 23.05.19
	ANTALL SIDER / BILAG 53/2
FORFATTERE Tharsana Sellaturai og Finan Fitsum Gebremariam	INTERN VEILEDER Per Ola Rønning

UTFØRT I SAMMARBEID MED NILU	EKSTERN VEILEDER Pawel Rostkowski
---------------------------------	--------------------------------------

<p>SAMMENDRAG</p> <p>Det ble utført en kjemiskscreening av 26 husstøvprøver, kjørt på en GC-HR-QTOF MS ved en non-target analyse. Dataene fra analysen er ekstrahert ved bruk av programmet Unknowns Analysis påkoblet til MS biblioteker som referanse. Vi fikk ekstrahert komponenter som er både identifiserte og uidentifiserte. For de identifiserte ble det utført en manuell kvalitetssikring, for fjerning av en del duplikater. En pivottabell ble brukt til å vise hvor mange ulike identifisert komponenter vi har og hvor mange ganger de gjentok seg. Et PCA-plott ble laget for å vise samvariasjonen av ekstraherte komponenter, basert på variablene; komponent RT, « Base Peak M/Z » og komponent areal. I PCA figuren, var det 12 ukjente komponenter som skilte seg mest. Vi klarte å identifisere tre av disse ved bruk av <i>in silico</i> metode. De ble klassifisert som « (4S)-2,2-Dimethyloxane-4-⁻ (C₈H₁₅NOS), N-[2-(2-hydroxyethoxy) propyl]octadecanamide (C₂₃H₄₇NO₃) og 2-Methyl-4-propyl-1,3-oxathiane (C₈H₁₆OS) ».</p>
--

3 STIKKORD:
Kjemiskscreening
Non-target analyse
GC-HR-QTOF MS

5/23/2019

KJEMISKSCREENING AV HUSSTØVPRØVER



Tharsana Sellathurai og Finan Gebremariam
OSLOMET

FORORD

I løpet av tiden vi har tilbrakt å utføre bacheloroppgaven vår i NILU, har vi lært og fått erfart veldig mye. Dermed vil vi først og fremst takke den eksterne veilederen vår Pawel Rostkowski fra NILU og den interne veilederen vår Per Ola Rønning fra OsloMet. Pawel har gitt oss en god oppfølging og veiledning av dataprogrammene. Per Ola har fulgt oss opp ved rapportskrivingsprosessen vår og gitt gode forslag. Men viktigst av alt, har veilederne vært veldig hjelpsom i de faglige aspektene.

Videre vil vi takke Anders R. Borgen og Hans Gundersen fra NILU. Anders har analysert støvprøvene for oss og forklart teorien bak ulike analyseinstrumenter. Hans har blant annet veiledet oss gjennom hvilke betingelser som bør optimaliseres i programmet brukt, samt tolking av massespektre. Tilslutt må vi få takke alle på NILU for et fint og hyggelig opphold.

Oslo, 23 Mai 2019

Innholdsfortegnelse:

1. INNLEDNING	1
2. TEORI	3
2.1 MILJØGIFTER OG PESTICIDER	3
2.1.1 Miljødirektoratets prioritetsliste	3
2.2 INSTRUMENT, AGILENT TECHNOLOGIES	5
2.2.1 Gasskromatografi og massespektrometri	6
2.2.2 Elektronionisasjon- EI	7
2.2.3 Kvadrupol- TOF.....	8
2.3 UNKNOWNNS ANALYSIS	8
2.3.1 SureMass.....	10
2.3.2 Biblioteker	11
2.4 QUALITATIVE ANALYSIS	12
2.5 KVALITETSSIKRING	12
2.5.1 Kovats RI	13
2.5.2 Statistiske Analyser.....	13
2.5.3 <i>IN-SILICO</i> , «Target» identifikasjon;	15
3. MATERIALER OG METODE	17
3.1 PROGRAMMER:	17
3.2 NETTBASERTE MODELLERINGSVERKTØY	17
3.3 DATABASERTE-LØSNINGER	17
3.4 PROSEDYRE	17
3.2.1 Unknowns Analysis	17
3.2.2 Qualitative Analysis- Identifikasjon Alkanmix.....	19
3.2.3 Manuell kvalitetssikring av identifiserte komponenter	19
3.2.4 Pivottabell	20
3.2.5 Principal Component Analysis.....	20
4. RESULTATER	22
4.1 UNKNOWNNS ANALYSIS	22
4.1.1 Bibliotek.....	23
4.2 MANUELL KVALITETSSIKRING AV IDENTIFISERTE KOMPONENTER	24
4.2.1 Oversikt identifiserte komponenter.....	25
4.2.3 Konfidens identifiserte komponenter	28

4.3 PCA- ANALYSE AV ALLE KOMPONENTER (IDENTIFISERTE OG UIDENTIFISERTE)	29
4.3.1 Identifikasjon av ukjente komponenter fra PCA-plott	31
5. DISKUSJON	34
5.1 Bearbeiding identifiserte komponenter fra Unknowns Analysis	35
5.2 Klassifisering av identifiserte stoffer	36
5.3 Bearbeiding, alle komponenter fra Unknowns Analysis	36
5.4 Arbeidsflyt;	37
6. KONKLUSJON	39
<i>Referanser :</i>	<i>40</i>
<i>Supplement 1;</i>	<i>44</i>
<i>Vedlegg 1; Oversikt ID støvprøver</i>	<i>44</i>
<i>Vedlegg 2, Alkanmiks</i>	<i>45</i>

SAMMENDRAG

Denne rapporten handler om resultater som har blitt bearbeidet fra en kjemiskscreeningsanalyse gjort på innendørs støvprøver. Her drøftes ulike statistiske analyseringsprogrammer, kvantifiseringsteknikker, samt noen data visualiseringsmetoder som er brukt i dataopparbeidingsprosessen. Hensikten med oppgaven er, å identifisere kjemiske komponenter som finnes i støvprøvene og optimalisere noen betingelser i en non-target analyse. I tillegg skal det utarbeides gode kvalitetssikringsmetoder og en arbeidsflyt.

En kjemiskscreening ble utført på 26 forskjellige støvprøver. Selve prøveopparbeidelsen og prøveanalysen av støvprøvene er gjort av ansatte på NILU. Prøvene ble kjørt på en GC-HR-QTOF-MS instrument med en elektron-positiv ionisasjon. Støvet er samlet fra ulike steder innendørs. Dermed var tilstedeværelsen og mengden av komponentene varierende i de ulike prøvene. Resultatet fra GC-analysen ble bearbeidet videre for å ekstrahere komponenter i programmet, Unknowns Analysis. Ekstraksjonen av komponenter ble gjort ved bestemte betingelser. En del av disse ble identifisert ved hjelp av MS- bibliotekene; NIST 17, MassBank og PCDL. Resterende ekstraherte komponenter gjenstod som uidentifiserte. For å forsikre at ekstraherte komponenter ble identifisert riktig, har vi i et nytt forsøk snevret betingelsene ved optimalisering. I det nye forsøket, ble det ekstrahert totalt 423 457 komponenter, hvorav 3047 ble identifisert. For fjerning av duplikatene i lik prøve og øking av konfidens, ble en manuell kvalitetssikring gjennomført ved å ta utgangspunkt i Kovats RI-verdier og dobbeltsjekke MS-spektre. Dermed gjensto det 2842 identifiserte komponenter, hvor 2% var ftalat, 2% fosfatforbindelser og 4% andre halogenerte forbindelser. En pivottabell ble lagd av de 2842 komponentene. Tabellen viste at det totalt har blitt identifisert 909 ulike komponenter. Blant disse gjentok tert-butylamin seg mest.

Fra en PCA analyse av alle de ekstraherte komponenter viste det seg, at 12 komponenter skilte seg veldig ut. Tre av disse ble identifisert ved hjelp av «Agilent Mass Hunter Qualitative Analysis», nettsidene ChemSpider og PubChem, samt modellen CFM-ID. Disse ble brukt for å finne bestemte spektre ved en *in-silco* identifikasjon. Komponentene ble identifisert som mindre fragmenter med opphav fra følgende større fragmenter; «(4S)-2,2-Dimethyloxane-4-‘ (C₈H₁₅NOS), N-[2-(2-hydroxyethoxy) propyl]octadecanamide (C₂₃H₄₇NO₃) og 2-Methyl-4-propyl-1,3-oxathiane (C₈H₁₆OS)» .

LISTE OVER FORKORTELSER OG SYMBOLER

NILU: Norsk institutt for luftforskning

GC: Gas Chromatography

LC: Liquid Chromatography

MS: Mass Spectroscopy

QTOF: Quadrupole time-of-flight

EI: Electron ionization

PCDL: Personal Compound Database and Library

NIST: National institute of standards and technology

SNR: Signal-noise-ratio threshold

MW: Molecule Weight

RI: Retention Indices

RT: Retention Time

M/Z: Mass ion/ Ion charge ratio

PCA: Principal Component Analysis

- **PCn:** Principal Component n, where $n = \{1, 2, 3 \text{ etc.}\}$

IUPAC: International Union of Pure and Applied Chemistry

INCHI: IUPAC -NIST Chemical Identifier

CFM-ID: Competitive Fragmentation Modeling for metabolite identification

PubChem: Database for chemical molecules

ChemSpider: Database for chemical molecules

Kjemikalier:

POP: Persistent Organic Pollutants

PBT: Persistent Bioaccumulative Toxic

PCB: Polychlorinated Biphenyl

BDE: Polybrominated Diphenyl Ethers

TBBPA: Tetrabrom-bobisphenol A

TCEP: (Tris(2-chloroethyl) phosphate)

DEHP: Bis(2-ethylhexyl) Phthalate

BBP: Benzyl butyl Phthalate

DBP: Di-n- Butyl Phthalate

DIBP: Diisobutyl Phthalate

PCP: Phenol, Pentachloro-

PAH: Polycyclic Aromatic Hydrocarbons

PVC: Poly (Vinyl Chloride)

1. INNLEDNING

Undersøkelser på husholdningsstøv har lenge vært et sentralt tema i luftforskning rundt i verden. Blant annet har USA, Kina og Canada gjennomført ulike analyser for sikring av innendørs miljø og helseskadelige miljøgifter gjemt i støvet (Zhang Mengtao, 2019) (Sun Yu, 2019) (Rantakokko, et al., 2019) (Liu & Mabury, 2019) (Shoeib, et al., 2019). I den siste tiden har bekymringer rundt miljøgifter i omgivelser og dets påvirkning på mennesker, dyr og klima økt. Flyktige og semi-flyktige miljøgifter ender vanligvis opp i matkjeden, drikkevannet og lufta. Store mengder av slike stoffer i næringskilder gir en betydelig helserisiko (Velázquez-Gómez, et al., 2018). Før i tiden, var forskning på miljø hypotesebaserte som peker mot target-analyser. I dag finnes det moderne og mye mer følsomme massespektrometriske-metoder for kvalitative og kvantitative non-target analyser (Samanipour, et al., 2019)

I Norge har vi NILU, Norsk institutt for luft undersøkelser, som utfører ulike analyser. Instituttet ble etablert i 1969 og tar for seg analyser av luftkvalitet og undersøkelser av miljøgifter i atmosfæren. De utvikler nye innovative løsninger og forslag til tiltak for å redusere forurensinger i miljøet. I tillegg bidrar de til økt kunnskap og bevissthet om utslipp av miljøgifter og deres konsekvenser for samfunnet (NILU, 2019) .

Støv er en stor matriks som lett fanger opp de fleste miljøgiftene og andre stoffer man finner innendørs. Støv består av flere tusener av partikler, som befinner seg i ulike steder i lufta avhengig av deres tyngde. Store, tunge partikler vil raskt falle og ligge på overflater av møbler og gulv. De mellomstore partikler vil også ende opp på overflater, men vil lett virvle rundt i et rom ved luftbevegelser. Støv som består av små partikler vil for det meste sveve i luften og vil kunne pustes inn av mennesker og dyr (Ledermann, 2018a).

Støv har generelt ikke en stor påvirkning på vårt respirasjonssystem, med mindre man puster inn store mengder eller har overfølsomhet for luften. Barn er mest utsatt for påvirkning fra støv, fordi de ikke har utviklet immunforsvaret sitt fullt ut. I tillegg oppholder de seg nærmere bakken hvor mest parten av støvet befinner seg (Rantakokko, et al., 2019). Stor eksponering og innpusting av støv kan føre til astma eller allergiske utfall hos enkelte (Ledermann, 2018b). I støv finnes det miljøgifter og allergener som kan ha helseskadelige effekter. Effektene kan komme i form av kreftfremkallelse, nerveskader, hormonforstyrrelser, fosterskader og skader på reproduksjonsevnen (Velázquez-Gómez, et al., 2018) (Miljødirektoratet, 2019).

Reduksjon av utslippet av miljøgifter er et av tiltakene som blitt avtalt, for å begrense mengde av miljøgifter. Stockholmkonvensjonen er et av eksemplene. I tillegg blir stadig nye og bedre analysemetoder utviklet innen dette forskningsfeltet. Dette er viktig for å oppnå bedre, nøyaktige og kvalitetssikre resultater. Kjemiskscreening er en av de viktige analysemetodene for miljøgifter. De fleste kjemiskscreeningsanalysene blir i dag gjort på vannprøver. I motsetning til vannprøver har støvprøver en mer kompleks matriks som gjør det vanskeligere å detektere komponentene man er interessert i (Moschet, et al., 2018) (Schlabach, 2013).

I dette forsøket ble det gjennomført en non- target kjemiskscreening av støvprøver. Formålet med dette forsøket var å identifisere og kvalitetssikre så mange som mulig komponenter fra støvprøver. Støvprøvene hadde gått gjennom et høyt oppløselig, GC- HR-QTOF instrument med EI ionekilde. Screeningen gjøres ved bruk av ulike analytiske og statistiske dataprogrammer. I tillegg skal det forsøkes å lage ulike forslag for å oppnå en bedre arbeidsflyt videre.

Dataprogrammene som ble benyttet innebærer kvantifiseringsteknikker, kvalitative analyser og MS- spektroskopi. Hvor målinger som retensjonstid, komponent areal, «base peak m/z» og retensjonsindekser har store roller. Programmet Unknowns Analysis sammen med «Suremass deconvolution» brukes for å ekstrahere ulike komponentsignaler fra GC-kromatogrammer. «Suremass deconvolution » er en algoritme, som kan utnytte GC-QTOFs høye evne til å separere ulike molekyler.

En lignende oppgave har tidligere, i vår 2018, blitt gjennomført av to studenter fra OsloMet (Stensbøl & Kirkemo, 2018). I oppgaven deres, hadde de et annet antall med støvprøver og brukte en annen prøveopparbeidelsesmetode. De bearbeidet data for 3 paralleller av 9 ulike støvprøver, hvorav 8 prøver var spikket. Vi fikk utdelt data av 26 analyserte støvprøver. Disse prøvene har blitt samlet fra stue, kjøkken, soverom og trapp. Prøvetaking, prøveopparbeiding og analysedelen i GC-QTOF- en var gjort av ansatte på NILU. Carl Stensbøl og Per Ivar Kirkemo hadde noen prøver med kjent stoffer. Vi har derimot ikke hatt noen prøver med kjente stoffer. Dermed har vi bare delvis tatt utgangspunkt i deres arbeidsflyt og har tatt fokus på å videreutvikle den. I forsøket deres, ble det oppdaget at det bør brukes mykere ionisasjonsteknikk i tillegg til EI. De har fått signalmetning i MS-spektrene for noen av de ekstraherte punkter. I tillegg bør Kovats RI beregnes og RI-verdier fra NIST hentes direkte fra Unknowns Analysis.

2. TEORI

2.1 MILJØGIFTER OG PESTICIDER

Noen miljøgifter er persistente, bioakkumulerende og toksiske. Dermed er slike miljøgifter ofte forkortet som PBT. Persistent gjør stoffene tungt nedbrytbare (irreversible), bioakkumulerende gjør at de lett hoper seg opp i levende organismer. Mens toksisitet gjør dem ekstra skadelige for miljø, mennesker og dyr (Miljødirektoratet, 2017).

En del miljøgifter grupperes inn i POP-er (Persistente Organiske Miljøgifter). Disse skiller seg ut fra de fleste miljøgiftene, fordi de kan transporteres langt. Eksempler på viktige persistente organiske miljøgifter er PCB, dioksiner, en del PAH-er, bromerte flammehemmere, og perfluorerte forbindelser (Miljødirektoratet, 2017). I Arktis har det vært funn av en del POP-er som utgjør en trussel for det arktiske økosystemet. POP-ene funnet har ikke opphav fra Arktis. Men har klart å bli transportert med hav og luftstrømmer fra mulige produksjonskilder rundt om i verden (Norsk Polarinstitutt, 2017).

Det meste av de helsefarlige kjemikaliene vi får i oss er ført gjennom drikke og mat. Som blant annet fisk, kjøtt osv. Stoffene blir lagret i næringskjeden hovedsakelig på grunn av den bioakkumulerende egenskapen de har. I tillegg til miljøgiftene, kan vi også få i oss pesticider. Pesticider er stoffer som brukes for å drepe insekter. De blir sett på som helseskadelige. Pesticider innebærer stoffene; herbicid, insektmidler (insektvekstregulatorer), nematicid, baktericid osv. Blant pesticider er herbicider mest brukt. Pesticider er benyttet som plantevernmidlene (Friedrich, 2016).

2.1.1 Miljødirektoratets prioritetsliste

Hvilke kjemikalier som utgjør de største truslene i Norge er ført opp i prioritetsliste gitt av miljødirektoratet. I denne listen er det ført opp 30 ulike kjemikalier. Målet er å stanse utslippene innen 2020 (Miljødirektoratet, 2019).

De fleste av disse er redusert kraftig ved regulering av produkter, industriutslipp, avfallshåndtering og tiltak for forurensings opprydding. Miljøgifter, funnet oftest i dag kommer fra importerte produkter eller transport av hav og luftstrømmer mot Norge (Miljødirektoratet, 2019).

2.1.1.1 Flammehemmere;

I prioritetslisten har vi blant annet bromerte flammehemmere, fosfororganiske flammehemmere og dekloraner. Stoffene benyttes i produkter for å gjøre dem mindre brennbare. De finnes vanligvis i elektroniske, elektriske produkter, biler, isolasjonsprodukter og tekstiler (Miljødirektoratet, 2017a).

De bromerte flammehemmerne som utgjør de største truslene er Penta-BDE, Okta- BDE, Dekka-BDE og TBBPA. Disse kan medføre nedsatt fruktbarhet, fosterskade, forplantningsevne, skade nervesystemet og hormonsystemet. En del av disse klassifiseres også som POP-er (Miljødirektoratet, 2017a).

De fosfororganiske flammehemmere brukes for å myk-gjøre plast. De finnes vanligvis i maling og plast. Den fosfororganiske flammehemmeren som er satt opp i prioritetslisten er TCEP (tris(2-kloretyl) fosfat). Den har lik effekt på mennesker som bromerte flammehemmere (Miljødirektoratet, 2019).

2.1.1.2 Ftalater;

Av de prioriterte miljøgiftene er det ftalater som utgjør den største mengden av miljøgifter funnet i Norge. De finner man vanligvis i plastprodukter som er bygd opp av PVC monomere, som gulvbelegg, kabler, bygningsmaterialer, vesker, klær, skosåler, regntøy, leketøy, innpakkingsmateriale og medisinsk utstyr (Miljødirektoratet, 2018).

Ftalatene som er oppført på prioritetslisten er, de-(2-etylheksyl) ftalat (DEHP), benzyl butyl ftalat (BBP), dibutyl ftalat (DBP) og diisobutyl ftalat (DIBP). Siden 1995 har utslippet av DEHP blitt redusert med 50 %. I dag er de fleste ftalatene forbudt i leker, småbarnsprodukt og kosmetikk. Mest av ftalater kommer fra importerte produkter. Reproduksjonsskader, miljøskader og hormonforstyrrelser er noen av innvirkningene ftalatene kan medfører. Ftalatene lekker vanligvis ut fra produkter til miljøet, når de er i bruk eller kastet som søppel (Miljødirektoratet, 2018).

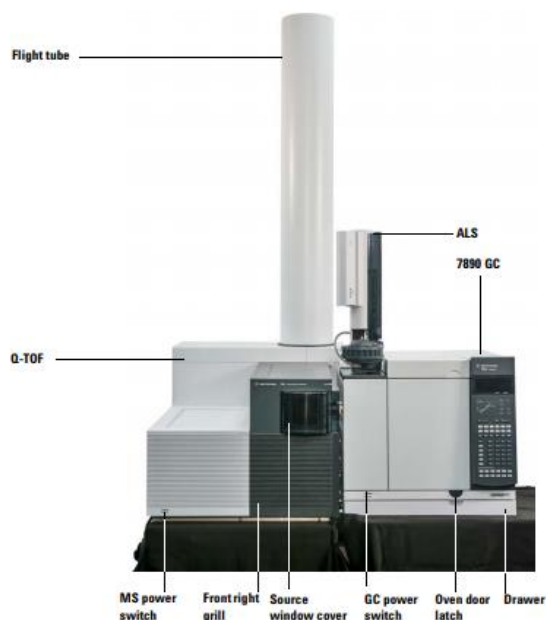
2.1.1.3 Dioksiner og furaner;

Dioksiner og furaner er to andre prioriterte miljøgifter. Disse oppstår i omgivelser under forbrenning. Siden 1995 er utslippet redusert med ca. 30 %. Klorerte, bromerte dioksiner og furaner kan fortsatt bli funnet. Klorerte dioksinene og furanene blir sett på som de farligste miljøgiftene. De har like effekter på miljø og helse som flammehemmere. Men de kan være dødelige ved svelging og gi alvorlige øyeirritasjoner (Miljødirektoratet, 2017b).

I tillegg til flammehemmere, ftalater, dioksiner, furaner har vi miljøgiftene PCB (polyklorerte bisfenoler), PAH, Bisfenol A og pentaklorfenol (PCP), som man finner i tekstil og treverk. Siden 1995, har PCP redusert med 99% (Miljødirektoratet, 2017) (Flem, 2019).

2.2 INSTRUMENT, AGILENT TECHNOLOGIES

Støvprøvene i denne oppgaven er analysert på en GC-QTOF fra Agilent Technologies, se figur 1. Hvor det er brukt en kapillær GC-kolonne «Agilent J&W GC Columns» og helium som bæregass.



Figur 1: GC-QTOF, Agilent Technologie, (Agilent Technologies, 2017)

2.2.1 Gasskromatografi og massespektrometri

Gasskromatografi sammen med massekromatografi er en god kvantifiseringsteknikk for organiske flyktige forbindelser som lett lar seg forgasse og er termisk stabile. Noe som kjennetegner enn del miljøgifter. Av de fleste tidligere studier innenfor kjemisk screening av miljøgifter har kromatografi, LC og GC, koblet opp mot MS, oftest MS i tandem (Velázquez-Gómez, et al., 2018) eller to dimensjonale kromatografer (GCxGC og LCxLC) blitt brukt (Ouyang, et al., 2017)

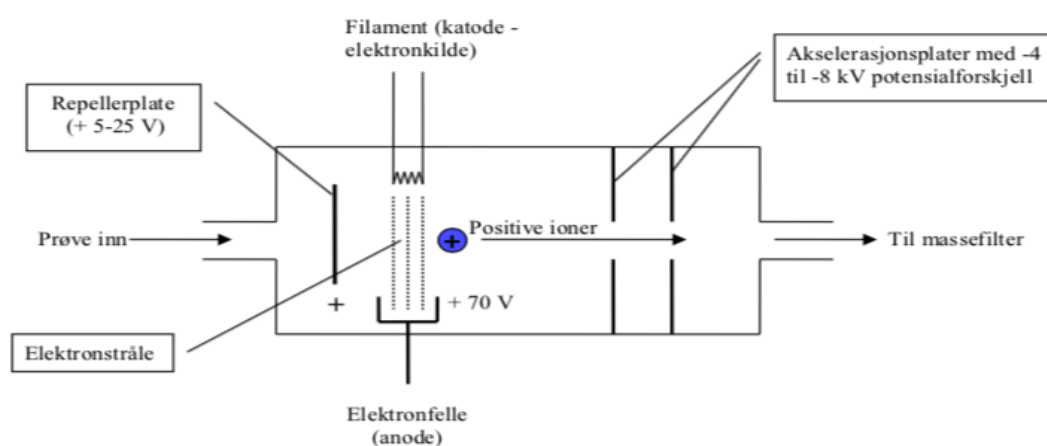
I en GC-MS fokuserer GC-en på separasjonen av de ulike stoffene i prøven. Separasjonen av de ulike stoffene i blir uttrykt i et kromatogram som viser retensjonstiden og mengden signal detektert. Retensjonstiden blir bestemt av ulikheter i fysiske og kjemiske egenskaper, i hovedsak kokepunkt, polaritet og molekylvekt (Harris, 2016).

Massespekteret fokuserer derimot på deteksjon av ulike ioner, som molekylioner og datterioner, i form av masse og ladningsforhold (m/z). Deteksjonen fra massespekteret blir uttrykt i et massespekter og fra dette spekteret kan det være mulig å bestemme molekylets masse og strukturformel. Et massespekter viser signaltopper til ulike ion med masse og ladningsforhold (m/z) i første aksene og den relative intensiteten i andre aksene. Linjesignalet med høyest blir referert til som «base peak» m/z og er ofte molekylionet (Harris, 2016) (Reusch, 2013). I forhold til isotoper som Cl og Br kan man få flere linjesignaler på grunn av isotopenes ulike masser, og intensitetsforholdene til de ulike linjesignalene baserer seg på deres naturlige forekomst. Denne informasjonen kan være nyttig ved identifisering av massespektre. ^{35}Cl har en naturlig forekomst på 75.77 % og ^{37}Cl på 24.23. ^{79}Br har en naturlig forekomst på 50.69% og ^{81}Br på 49.31% (Rønning, 2017).

Ionene massespekteret detekterer er ioniserte forbindelser av stoffene fra prøven separert i GC-en. Før MS-deteksjonen har stoffene gått gjennom et ionisasjonsområde samt en masseanalysator. I dette instrumentet er ionisasjonsteknikken brukt, elektronionisasjon, og masseanalysatoren er, kvadrupol Q-TOF.

2.2.2 Elektronionisasjon- EI

Ved elektronionisasjon (EI) blir stoff ionisert av en elektronstråle med høy spenning. Fra elektronstrålen vil en haug med høyenergiske elektroner bombardere molekyler og slå ut elektroner, slik at det dannes et positivt radikalt molekylion. Molekylionet vil fragmentere videre og gi opphav til datterioner. Ved dannelse av positive radikale molekylioner, har vi positiv ionisasjon, se ligning 1 og figur 2 (Harris, 2016).



Figur 2: Illustrasjon av innsiden til EI, hentet fra (Harris, 2016). Det er en elektronkilde (filamentet) som ioniserer kommende stoffene fra en prøve fra venstre side. Stoffene eller molekylene blir positivt ionisert ved elektronstrålen og kan gå videre til en massefilter og bli videre detektert ved en masseanalyzer (massedetektor).

EI er en veldig hard ionisasjonsteknikk og kan resultere til at molekylionene til noen av stoffene forsvinner helt siden er så radikale. Dette kan ha konsekvenser når kjemikalier skal identifisere ut i fra massespekteret. En ionisasjonsteknikk som er mykere og som ikke danner svært radikale molekyler er en kjemisk ionisasjon, CI (Harris, 2016).

2.2.3 Kvadrupol- TOF

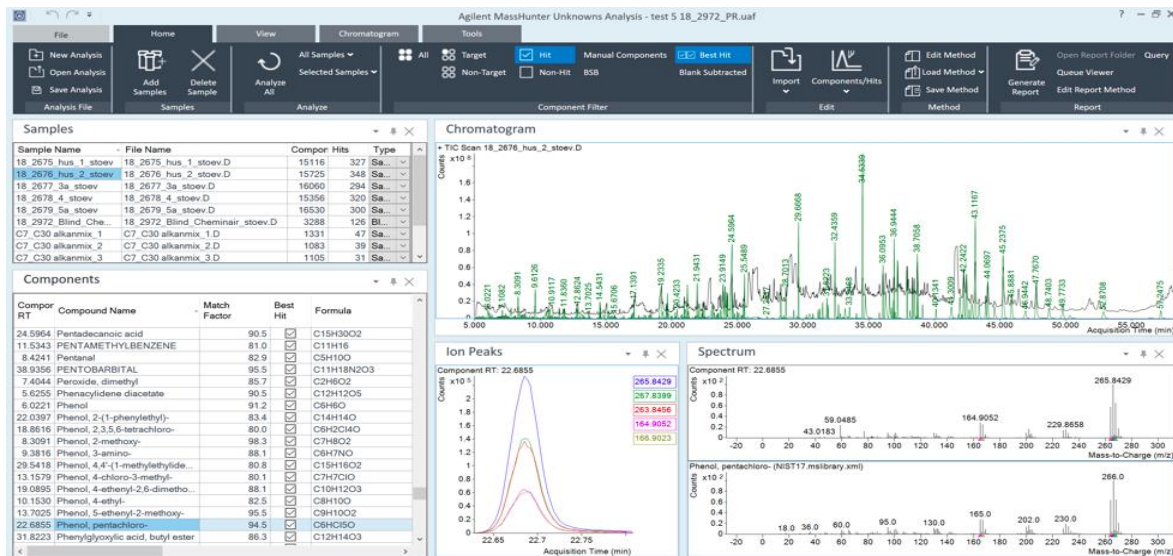
Q-TOF, «kvadrupol- Time of Flight», er et massefilter som består av en kvadrupol, en kollisjonscelle og en TOF. TOF er et feltfritt rør, der ioner blir separert ut ifra den kinetiske energien deres, som er avhengig av masse og ladningsforholdet (m/z). Ioner med lik masse skal ha lik kinetisk energi, men det kan være avvik. For å redusere avviket er TOF-en utstyrt med en refleksjon-TOF (Harris, 2016). Kvadrupolen sender ioner inn i TOF-en, den består av fire stavformede poler og gir en bedre retensjon. Polene kan bli påsatt spenning og ved å ha på en bestemt spenning er det mulig å kjøre en target analyse. En Q-TOF er svært følsom masseanalysator, og kan detekterer ioner med høy oppløsning og nøyaktighet (McKenna, 2015)

2.3 UNKNOWNNS ANALYSIS

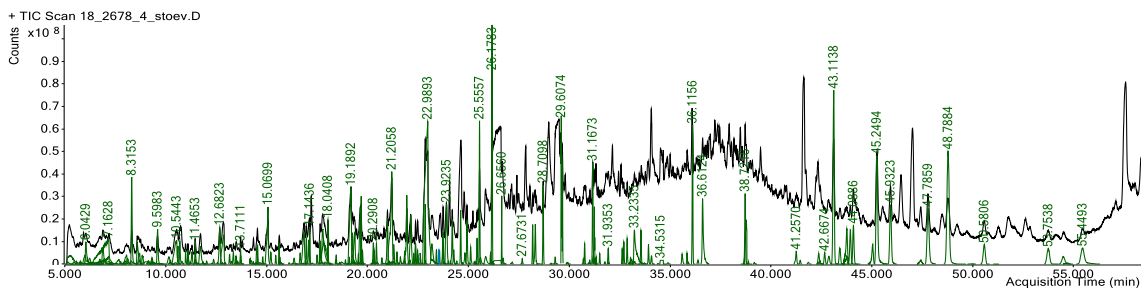
Unknowns Analysis er et analyseprogram fra produsenten Agilent. Her kan det velges ulike analyseringsmetoder. En av de nye brukte metodene er; Mass Hunter Quantitative Analysis med en Peak-deteksjons metode – SureMass Deconvolution. Agilent Masshunter workstation quantitative analysis, brukes for å lage en batchtabell med en valgt kvantifiseringsmetode for en target komponent ved hjelp av de samlede datafiler (Agilent Technologies, 2010) (Agilent Technologies, 2011).

Ved bruk av disse kombinasjon kan man gjennomføre en non-target-analyse. Med deconvolution, menes det å dekonvolverer komponentene i kromatogrammet. Muliggjøre å ekstrahere de rene spektrene fra bakgrunnsstøy basert på både retensjonstid og toppform. Dersom man velger TIC analysemetode, identifiserer man kromatografisk topp ved hjelp av integrasjon i stedet for dekonvolusjon. Signalene som gis fra de separerte og ioniserte komponentene i massespektrometrene, detekteres og kan ekstraheres fra kromatogrammer i programmet. Programmet kan påkobles til en eller flere biblioteker (library search). Det tar utgangspunkt i MS-spektre for de ekstraherte komponenter, som videre sammenliknes med en liknende funnet MS spektre fra et referansebibliotek. Slik fås en liste over de identifiserte og uidentifiserte komponenter, se figur 3. Hvor stor overstemmens det er mellom MS spektrene, kan bestemmes ved en parameter som heter «Minimum matchfaktor», se figur 3 (Agilent Technologies, 2015).

Figur 3 og 4 er bare tatt med fra forsøket vårt for å vise oppsettet.



Figur 3: Viser oppsettet av dataprogrammet, Unknowns Analysis. Den er hentet fra en av de kjørte analysene våre, for å vise en visuell forklaring. Bildet viser sammenligning av MS-spektre av den eksperimentelle komponenten mot et treff fra et MS- bibliotek. Betingelsen matchfaktor gir en overbestemmelse av likheten mellom MS-spektrene (se høyere side). Nederste til venstre, vises det en liste over ekstraherte komponenter i en prøve. Nederst i midten vises «ion peaks» av en komponent med bestemt retensjonstid.

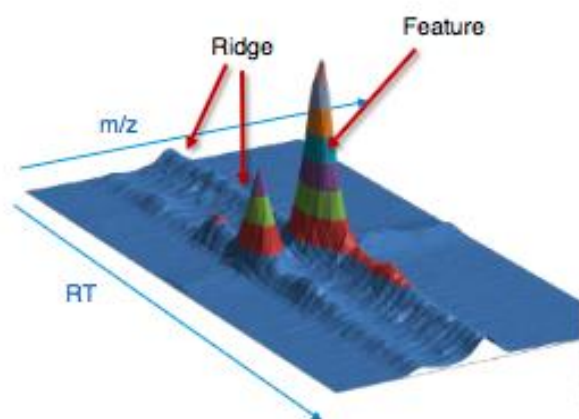


Figur 4: Kromatogrammet av TIC skannet til en av de kjørte analysene, med ekstraherte signaler. Y-aksen viser mengde i «counts» og x-aksen viser retensjonstiden i min. De grønne linjene er de ekstraherte signalene.

I programmet finnes også parameterne; «SNR threshold» (signal/støy forhold), «absolutte area», og ekstraksjonsvindu, som er viktige faktorer for å begrense det man ser etter i analysen. SNR er forhold mellom intensiteten av signal og signalstøy. Det er brukt for å karakterisere den relative styrken av signalene man er interessert i (Johnson, 2006). Parameterne «absolute area» og ekstraksjonsvindu hjelper å filtrere det vindu man er interessert i (Agilent Technologies, 2015).

2.3.1 SureMass

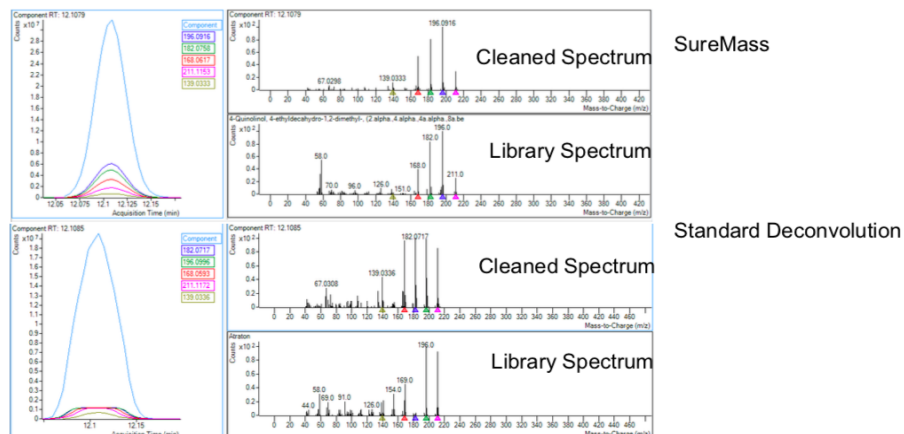
SureMass er en algoritme software basert på dekonvolusjonsteknikk. Den brukes for å ekstrahere presise signaler på et fullt spekter av en kjemiskskanning, generert fra et analyseinstrument. Denne softwaren er ofte brukt på et høyoppløselig GC-HR-MS- QTOF instrument ved kjemiskscreening av prøver. Dette for å kunne samle inn alle de høyttoppløselige dataene fra instrumentet. SureMass viser et to-dimensjonalt kromatogram. Dette kromatogrammet representerer en kontinuerlig detektert tre-dimensjonal MS-profildata med ulike variabler. For eksempel variablene; retensjonstid, «base peak m/z» og komponent areal. Dersom det dukker opp eksperimentelle endringer under prosessen vil metoden tilpasse seg forholdene raskt. SureMass er nøyaktig, sensitivt, reproduserbar, lineær og kompakt. For tiden er SureMass vanligvis brukt for GC-QTOF men forventes å bli tilgjengelig for andre typer analysemetoder i fremtiden. Samt er algoritmen bare brukt i Masshunter Qualitative Analysis og Unknown Analysis (Riener, 2016) .



Figur 5: Illustrasjon av hvordan SureMass algoritmen detektert signaler ser ut i en tre-dimensjonal graf. Bildet er hentet fra (Riener, 2016).

Figur 5 illustrerer hvordan SureMass detekterer signaler. Vanligvis oppstår det « Ridges » som er signalstøy fra lavt-nivå komponenter. Disse ses bort i fra når SureMass ekstrahere de viktigste signalene, ved å spore endringen i målte m/z verdier. Da oppnås ekstraherte SureMass kromatogram (ESM) (Agilent Technologies, 2017).

ESM er målt i m/z , altså en gjennomsnittlig m/z verdi fra spektre som tilsvarer signalet. Formålene av SureMass er at det gir redusert analyse tid, samt effektiv i innsamlinger og prosessering av relevante informasjoner fra data. Det er omtrent 100 ganger effektivt enn tidligere teknikker, se figur 6. Den detekterer og ekstraherer komponentsignaler med større sensitivitet, nøyaktighet og er mer reproduserbar (Agilent Technologies, 2017) (Agilent Technologies, 2015).



Figur 6: Sammenligning av MS-spektre for lukt komponent ved Suremass og tidligere brukt standard dekonvolusjon. Det viser kvaliteten på det nye utviklet metode; SureMass gir ion-topper og MS-spektre med bedre fasong og struktur, på grunn av bedre deteksjon og ekstraksjon. Bildet er hentet fra (Riener, 2016).

2.3.2 Biblioteker

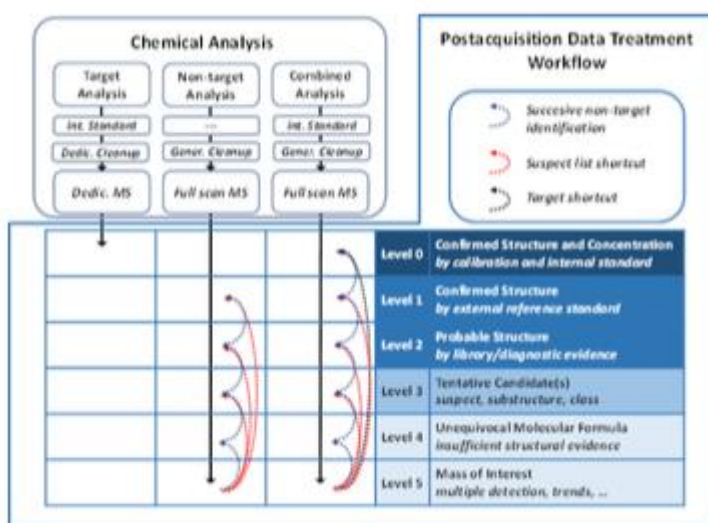
Unknown Analysis programmet kan påkobles til en eller flere MS-biblioteker, de virker som referansebiblioteker. Her kan detekterte MS-spektre av komponenter sammenlignes med et MS-spekter fra bibliotekene. Ved en non-target-støv analyse brukes vanligvis bibliotekene PCDL, NIST 17, Mass Bank og pesticidbiblioteker. PCDL er et format fra Agilent for tidligere funnet pesticider. Pesticidbiblioteket er høyt oppløselig og komponenter identifisert herfra har stor troverdighet. Mens de to andre bibliotekene er lavoppløselige og komponenter herfra må bli godkjent manuelt (Agilent Technologies, 2011) (SIS, 2017a).

2.4 QUALITATIVE ANALYSIS

Qualitative Analysis er et program som brukes for kvantifisering av en kjørt blanding eller løsnings. Det skjer en kvalitativ kvantifisering hvor eluering av en komponent fra en kolonne detekteres ved å ekstrahere signal ved en bestemt retensjonstid. Molekyler med høyere molekylvekt går tregere gjennom kolonnen enn molekyler med lavere molekylvekt. Dermed har de større molekyler lengere elueringstid (Agilent Technologies, 2019).

2.5 KVALITETSSIKRING

For å oppnå gode, troverdige resultater med god kvalitet er det viktig med kvalitetssikring. Kvalitetssikring kan utføres ved å gjøre ulike systematiske tester og analyser. Det vil øke konfidensen på resultater. Angivelse på konfidens kan skje ved bruk av konfidenstabeller. Konfidenstabell er en standard godkjenningstabell for funnet komponenter som viser hvor sikre identifisert de er. I figur 8, har vi en konfidenstabell foreslått av (Schymanski et al.) Den går fra nivå null til fem. Der nivåene med lavest tall viser mest sikre komponent.



Figur 7: Konfidenstabell foreslått av (Schymanski et al)

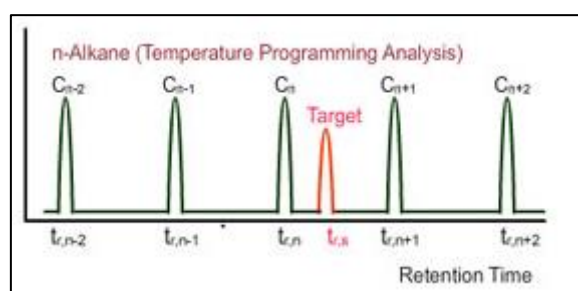
2.5.1 Kovats RI

Kovats retensjonsindeksen for et rent alkan uttrykker antall karbonatomer multiplisert med 100. Retensjonsindeksen beregnes ved bruk av retensjonstiden til ukjent komponent samt retensjonstiden til en eluert n-alkanmiks- standard, bestående av en homolog serie av alkaner, hvor hvert alkan har kjent retensjonstid. Retensjonstiden til alkanene avhenger hovedsakelig av kokepunkt og polaritet, i forhold til stasjonærfasen. I en homolog alkanmikstur vil alkaner med færre c-atomer ha raskere eluering enn de med flere C-atomer (IUPAC, 2014)

I likning 2 har vi ligningen for Kovats RI, den basere seg på en lineær funksjon hvor antall karbonatomer er proporsjonal med logaritmen til retensjonstiden (Vonk, et al., 2016). Figur 8 viser en visuell forklaring av parameterne i likningen (SHIMADZU, n.d.).

$$RI(\text{mål komponenten}) = 100 \times \left(\frac{tr,s-tr,n}{tr,n+1-tr,n} + n \right) \quad (2)$$

- tr,s : Retensjonstiden til mål komponenten RI skal beregnes for
- tr,n : Retensjonstiden til et alkan fra alkanmiksen med mindre antall C-atomer en mål komponenten.
- $tr,n+1$: Retensjonstiden for et høyere alkan, hentet fra alkanmiksen.



Figur 8: Viser hvor tr,n og $tr,n+1$ ligger i forhold til tr,s . Er hentet fra (SHIMADZU, n.d.)

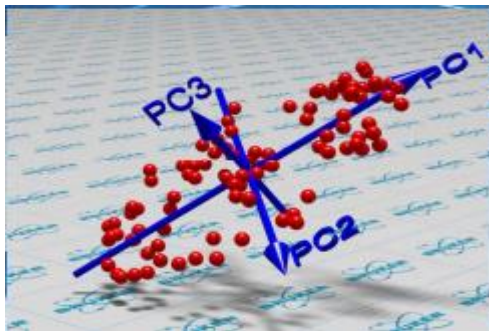
2.5.2 Statistiske Analyser

De statistiske analysene, som det settes mest fokus på i denne oppgaven er pivottabeller, PCA, og konfidenstabell. Pivottabell og PCA gir en mer strukturert oversikt over en stor mengde med data. Pivottabell, i form av tall og hvor mange ganger lik data gjentar seg over flere prøver, og PCA, i form av et plott.

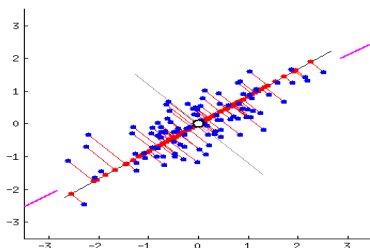
PCA står for «Principal Component Analysis» og viser samvariasjonen av flere variabler for ulike prøver, og kan brukes for å vise om det finnes noen komponenter som skiller seg ut.

Dette er en multivariabel analyse metode, som baserer seg på uavhengige korrelerte punkter som ved regresjon har blitt omgjort til lineære ukorrelerte hovedkomponenter kalt PC1, PC2, PC3 osv. Med korrelasjon menes det snakk om samvariasjon (Frøslie, 2018).

Hovedkomponentene er lineære linjer i et plan med antall akser som stemmer overens med antall variabler man har valgt å basere PCA-plottet på. I planet er ulike punkter plassert i forhold til variablene deres. Hvordan de ulike hovedkomponentlinjene ligger i planet kan ses på egenvektoren. Hver hovedkomponent uttrykker variansen til de ulike punktene i ulike grader, se figur 9 og 10 Den første hovedkomponenten PC1 viser normalt punkter med høyest varians og deretter kommer PC2 og PC3 (Zhang & Adela , 2017).



Figur 9: Viser visuelt i 3D hvordan PC1, PC2, PC3 er orientert (Bruker Daltonics, u.d.).



Figur 10: Viser hvordan en PC linje ligger blant punktene; hentet fra (Jaardi, 2019).

Et PCA-plott vil vise en to-dimensjonalfigur, som er et utsnitt av det tre-dimensjonale planet med PC1, PC2 og PC3. Utsnittet tas fra vinkelen hvor det er en god nok rotasjon av PC1, PC2 og PC3 aksene til å vise variansen og fordelingen av punktene på en best mulig måte. Hvor stor grad av hovedkomponentene, PC1, PC2 og PC3 som blir vist i PCA-plottet, samt

fordelingen av variablene i hver hovedkomponent kan vises i egenanalyser av korrelasjon og egenvektorer (Zhang & Adela , 2017).

2.5.3 IN-SILICO, «Target» identifikasjon;

In-silico er et uttrykk som vil si «utført på data eller via datasimulering» (Stein, 2012) . In-silico identifisering er en type analysemetode som baserer seg på sammenligning av oppnådde MS-spektre med *in-silico* spektre, hentet fra bestemte modeller; CFM-ID og MetFrag. In-silico MS-spekteret for et bestemt stoff søkes opp ved bruk av InCHI koder (Blaženović, et al., 2017).

Vanlige MS-biblioteker eksperimentelle MS-spektre ved non-target analyser, dekker mindre enn 1 % av komponentene dra Chempider og PubChem, med 50-90 millioner forbindelser. I motsetning til MS-biblioteker gir *in-silico* fragmenteringsverktøyene tilgang til flere teoretiske MS-spektre, som kan brukes som referanse spektre for identifikasjon. (Blaženović, et al., 2017) , (Yap, et al., 2018)

INCHI

IUPAC international chemical identifier (InCHI) er et begrep brukt av Inchi Trust, for identifisering av kjemiske komponenter ved bruk av bokstaver og tall. InCHI TM er et program software, som gir koder som representerer ulike komponenter (IUPAC, 2019).

Programmet er standardisert av IUPAC, slik at den koder basert på den molekylære informasjon av komponentene. INCHI tar hensyn til atomer, bindinger, tautomer, isotoper, stereokjemi og elektroniske ladninger som er tilstede i en komponent. Med andre ord, er det samme tankegang som når man brukes CAS-nummer for identifisering av komponenter. Eneste forskjellen er at, InCHI er lett brukbar, lett forståelig og ikke er patentert (IUPAC, 2019)

Ved å søke molekylformelen av en komponent i en av følgende nettsider; PubChem, Chemical Identifier Resolver (NCI) og ChemSpider kan man generere en InCHI-beskrivelse av en komponent (IUPAC, 2019).

CFM-ID;

Programmet, «Competitive Fragmentation Modeling for metabolite identification » (CFM-ID) brukes for å identifisere komponenter ved hjelp av en *in-silico* prediksjon. MS-spektre av fragmentene fra komponenten søkes ved å valg en type ionisasjonsmetode. Programmet viser en peak-tabell og fragmentstruktur. Den genererer en sannsynlig MS/MS fragmenteringsprosess, altså spekterprediksjon for en gitt komponent. Ved å tolke signaltoppene arrangerer programmet mulige kandidater for funnet spektra (CFM-ID, n.d.) .

3. MATERIALER OG METODE

3.1 PROGRAMMER:

- Agilent Mass Hunter Unknowns Analysis B 10 (Quant-My-Way)
- Agilent Mass Hunter Qualitative Analysis B.07.00
- Minitab Ltd, Minitab 18
- Microsoft, Excel 2013
- NIST 2017
- Mass Bank
- Agilent PCDL Pesticide Library

3.2 NETTBASERTE MODELLERINGSVERKTØY

- CFM-ID, <http://cfmid.wishartlab.com/>

3.3 DATABASERTE-LØSNINGER

- ChemSpider, <http://www.chemspider.com/>
- PubChem, <https://pubchem.ncbi.nlm.nih.gov/>

3.4 PROSEDYRE

Det ble analysert 26 støvprøver med GC-HR-QTOF MS, samlet fra stue, kjøkken, soverom, trapp osv. (se vedlegg 1). Under analysen ble prøvene delt inn i fire undergrupper, som ble kjørt med fire ulike blindprøver. Dataene fra analysen ble overført til Unknowns Analysis og bearbeidet videre.

3.2.1 Unknowns Analysis

I dette programmet ble det ekstrahert ulike signaler fra TIC kromatogrammene til alle de 26 ukjente støvprøvene og de 4 blindprøvene hentet inn fra GC-TOF-MS analysen.

Analysekjøringen i programmet av hver prøve ble kjørt med «blank subtraction» i henhold til deres rette blindprøve. Ved bruk av algoritme, SureMass Deconvolution, ble ulike signaler detektert. I programmet har hvert ekstraherte signal fått sitt eget MS-spekter. MS- spektrene

ble sammenlignet med MS-spektre fra bibliotekene; Mass Bank, NIST17 og pesticidbiblioteket for å kunne identifisere de mulige signalene.

Visse filtreringsbetingelser ble satt før kjøring;

- SNR threshold: 20
- Matchfaktor:80%,
- Ekskludert m/z: 28 m/z
- Absolutt areal: 5000 counts,
- Ekstraksjonsvindu på venstre side: 0,3 delta m/z
- Ekstraksjonsvindu på høyre side: 0,7 delta m/z.
- Minimum antall ion peaks: 3
- Maksimum antall ion peaks: 10

For å begrense antall funn av identifiserte og uidentifiserte komponenter med sikrere kvalitet, ble kjøringen gjort på nytt. I denne omgangen ble noen av betingelsene optimalisert.

Matchfaktor ble økt til 85% og minimum antall ion peaks ble økt til 5.

Før kjøring ble disse parameterne lagt til slik at de ble tilgjengelig for hvert ekstraherte signal;

- Component RT
- Compound Name
- Match Factor
- CAS#
- Base Peak Area
- Library File
- Library Type
- Component Area
- Base Peak MZ
- Component RI

I begge kjøringene ble en alkanmiks fil lastet opp i programmet. Filen gir en oversikt over identifiserte alkaner i en alkanmiks, se vedlegg 2. Alkanmiksen er kjørt i GC-QTOF-en med like betingelser som støvprøvene. Sammen med de identifiserte alkanene er retensjonstiden ble skrevet opp. Denne filen ble brukt i dataprogrammet Unknowns Analysis for å oppgi en komponent RI- verdi for alle identifiserte komponentsignalene. Den oppgitte RI-verdi ble videre brukt for kvalitetssikring, se kap. 3.2.3 for kvalitetssikring. Se kap. 3.2.2 for identifikasjon av alkaner med riktig retensjonstid.

Etter kjøringene ble parameterne for de identifiserte og uidentifiserte komponentene eksportert til CSV-filer. Det ble opprettet en Excel-fil over alle de ekstraherte komponentene, se supplement 1.

3.2.2 Qualitative Analysis- Identifikasjon Alkanmix

Programmet Qualitative Analysis ble brukt for å identifisere alkantoppene i en alkanmiks, kalt C7-C13. Massespekteret for den siste eluerte alkantoppen i kromatogrammet ble ekstrahert. Det ekstraherte spektret ble sammenlignet med et bibliotek for å finne ut hvilke topp tilsvarende hvilke typer alkan i kromatogrammet. For å spare tid, ble de resterende toppene identifisert i følge teorien om at større molekyler elueres tregere enn de små. Det ble anslått hvilke alkan-komponenter som tilsvarende kromatogramsignalene ved bruk av en gitt liste over hvilke alkaner som er tilstede i alkanmiksen. Retensjonstiden til «base peaken» for hver topp ble lest av og ble notert med tilsvarende alkantype for beregning av Kovats RI, se kap. 3.2.3. Vedlegg 2 er det tatt med en tabell for retensjonstiden til de ulike alkanmiksen.

3.2.3 Manuell kvalitetssikring av identifiserte komponenter

En manuell kvalitetssikring av de identifiserte komponentene ble utført ved å kontrollere MS-spektre, sammenfallende «ion peaks», «matchfactor», og sammenligne en beregnet og oppgitt kovats RI-verdi. Kvalitetssikringen ble utført for å fjerne mulige urealistiske/ ufornuftige ekstraherte komponenter og duplikater av ett komponent i lik prøve. Se supplement 1 for kvalitetssikrede identifiserte komponenter.

Kovats retensjonsindeks (RI)

Eksperimentelle RI-verdier ble beregnet ved bruk av likning 2, se kap 2.5.1. Likningen ble brukt sammen med retensjonstiden til komponentene, samt retensjonsverdiene til alkantopper identifisert i alkanmiksen. Se supplement 1 for de beregnede kovats RI verdiene.

RI verdiene for de identifiserte komponenter oppgitt fra biblioteket NIST 17 i Unknowns Analysis, ble sammenlignet med retensjon indekser beregnet manuelt. Komponenter ble godkjent ved kvalitetssikringen dersom $-0.7 < \text{delta RI verdi} < 0,7$.

3.2.4 Pivottabell

Fra samlet excel-fil av identifiserte komponenter ble det lagd en pivottabell. Denne viser totalt antall ulike stoffer identifisert, samt gjentakelsene av de ulike komponentene i hver støvprøve. Pivottabell ble lagd for de identifiserte komponentene fra begge forsøkene i Unknowns Analysis. I tillegg ble det lagd en pivottabell etter kvalitetssikringen av de identifiserte komponentene fra forsøk nr. 2 i Unknowns Analysis. Se supplement 1 for pivottabellen lagd for forsøk 2.

3.2.5 Principal Component Analysis

Fra samlet data over uidentifiserte komponenter, samt bearbejdede (kvalitetssikrede) identifiserte komponenter, ble det lagd et PCA- plott i Minitab 18, se figur 20. Variablene brukt for å lage plottet er; Komponent RT, «Base Peak m/z» og komponentareal. Hvert punkt i plottet ble gruppert etter prøvenummer. Punkter som skilte seg ut i plottet ble markert og uttrykt i form av bestemte ID variabler. ID-variablene ble satt til prøvenummer, komponentnavn, komponent RT, «Base Peak m/z» og komponentareal.

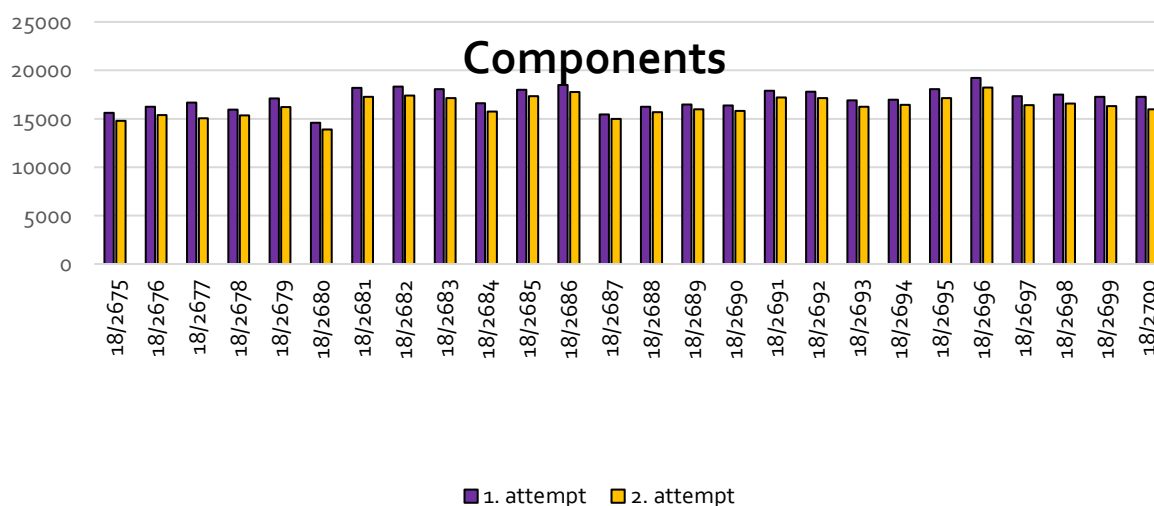
Av punktene som skilte seg ut ble det tatt beslutninger om det var mulig å identifisere noen uidentifiserte punkter ved bruk av identifiserte punkter i nærheten. Andre uidentifiserte punkter som skilte seg i plottet ble identifisert ved en *in-silico* identifikasjon. Hvert av punktene ble bearbejdet i Qualitative Analysis programmet. Her ble først EIC kromatogrammet som tilhører et av de bestemte punktene ekstrahert med punktets «base peak m/z». Betingelsen «Single m/z expansion» for kromatogrammet ble satt til $\pm 0,2$. Toppen som tilhører det bestemte punktet ble funnet på kromatogrammet med punktets komponent RT verdi. Videre ble MS-spekteret til toppen ekstrahert og det ble generert formel for ulike fragmenter.

Det ble antatt at linjesignalet med høyest m/z er nærest m/z -verdien for molekylionet til det bestemte punktet. Molekylformelen for dette linjesignalet ble søkt opp i ChemSpider. Dette for å finne treff blant stoffer som har lik molekylformel. De med flest referanser ble valgt til å ha høyst sannsynlighet for å være molekylet. Navnet på molekylet ble søkt opp på PubChem for å finne INCHI-koden. INCHI-koden ble videre søkt opp på modellen, CFM-ID for å få *in-silico* massespekteret for det bestemte stoffet. m/z -verdien til linjesignalene med høyst intensitet i disse *in-silico* MS-spektrene ble søkt opp på kromatogrammene i Qualitative Analysis, for å foreta en kvalitetssikring. Se kap. 4.3.1 for de identifiserte punktene.

4. RESULTATER

4.1 UNKNOWNNS ANALYSIS

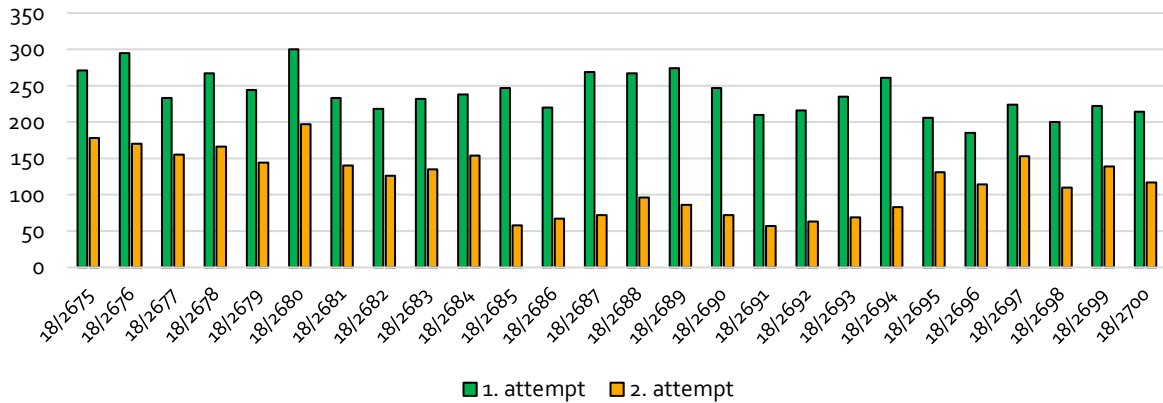
Unknown Analysis ga en liste over identifiserte og uidentifiserte komponenter fra detekterte signaltopper fra TIC-en til hver støvprøve, kjørt på GC- QTOF MS instrumentet. Antallet identifiserte og uidentifiserte komponenter påvirkes av valgte betingelser. Figur 11 og 12 viser resultater fra begge to forsøkene. Diagrammet i figur 11 viser variasjonen av ekstraherte komponenter for hver prøve i de to forsøkene. I x-aksen er type støvprøve oppgitt og antall komponenter funnet står i y-aksen. Stolpene farget lilla representerer resultatene fra første forsøk og stolpene farget gule representerer resultatene fra andre forsøket. Totalt er det ekstrahert 423 457 komponenter i forsøk 2.



Figur 11: Viser antall komponenter ekstrahert fra hver prøve med Unknowns Analysis i to ulike forsøk, med ulike betingelser. Lilla for antall komponentene ved første forsøk og oransje ved den andre forsøk. I x-aksen vises prøvenummeret og i y-aksen vises antallet.

Figur 12, viser antall komponenter som har gitt treff i bibliotekene; NIST 17, Pesticidbiblioteket, MassBank, og blitt identifisert. De grønne stolpene representere resultater fra første forsøk og de gule stolpene resultater fra andre forsøk. Her er det observert stor reduksjon. Totalt er det identifisert 3047 komponenter i forsøk 2.

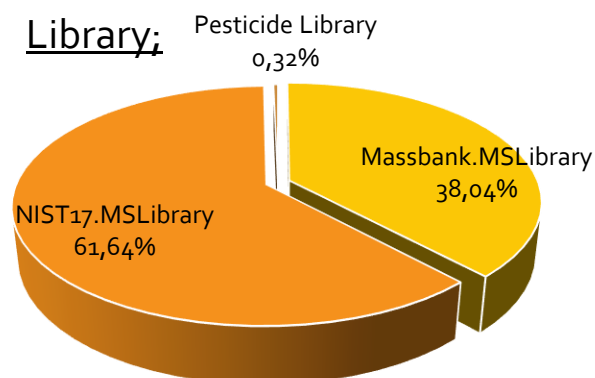
HITS



Figur 12: Viser antall hits ved første forsøk og andre forsøk. Grønne stolper for første forsøk og gule stolper for andre forsøk. I x-aksen vises prøvenummeret og i y-aksen vises antallet.

4.1.1 Bibliotek

Diagrammet i figur 13 fremstiller hvor stor andel av de identifiserte komponentene som er funnet med bibliotekene; NIST 17, Pesticid, Massbank. Kjemiske komponenter identifisert ved hjelp av NIST 17 utgjør 61,64%, Pesticide bibliotek lik 0,32% og fra Massbank bibliotek var det 38,04%.



Figur 13: Viser prosentvis fordeling av hvilke biblioteker komponentene har ble identifisert med.

4.2 MANUELL KVALITETSSIKRING AV IDENTIFISERTE KOMPONENTER

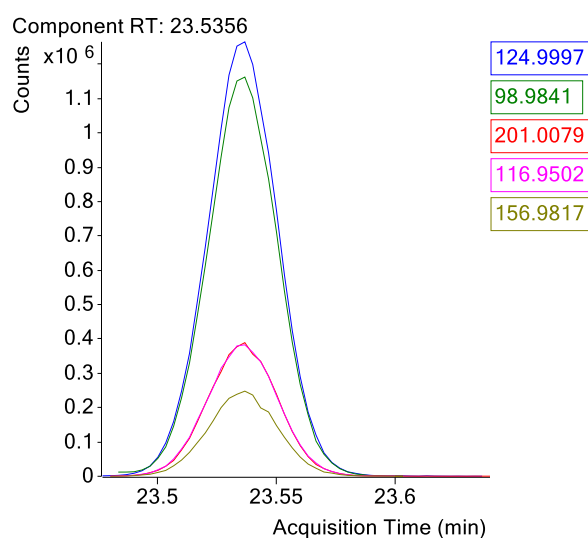
Tabell 1 viser hvor mange totalt av alle de ekstraherte komponentene som ble identifisert, og hvor mange som har blitt fjernet etter den manuelle kvalitetssikringen.

Tabell 1: Viser antall identifiserte komponenter og antall fjernet etter manuell kvalitetssikring, for forsøk 2.

Kjemikalieliste	Totalt identifiserte komponenter (i alle prøver)	Antall fjernet
Samlet data, forsøk 2	3047	205

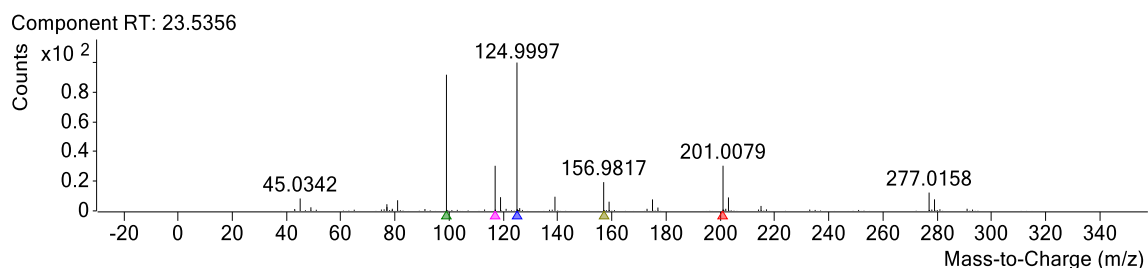
De identifiserte komponenter utgjør totalt antall komponenter lik 3047. Ved sjekking av MS-spektre og feil ble 205 av disse fjernet. Dermed er det igjen 2842 identifiserte komponenter,

I figur 14, har vi tatt med et utsnitt av «ion peaks» for en identifisert komponent. Denne er fra prøve 18_2678 og ble identifisert som 2-propanol,1-klor-, fosfat (3:1) med en retensjonstid på ~23.54 min. Utsnittet viser en fin fragmentfordeling av komponenten, og det er omtrent slikt «ion peaks» for resterende komponenter ser ut etter en manuell kvalitetssikring.

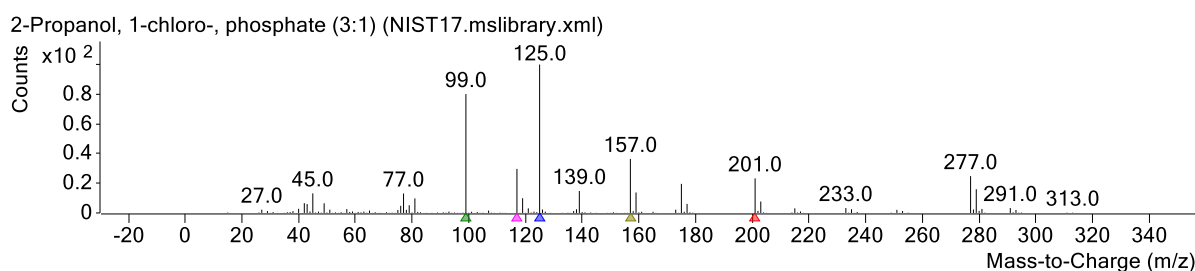


Figur 14: Ion peaks for 2-propanol, 1-chloro-, phosphate (3:1). Hver topp tilhører hvert sitt fragment fra likt molekyl. De har hovedtopp med lik retensjonstid, siden molekylet de stammer fra har eluert ved komponent RT 23.5356 min. Øverst til høyre vises m/s for hvert fragment.

Figur 15 viser massespekteret som tilhører det ekstrahert signalet 2-propanol, 1-klor-, ftalat (3:1). Spekteret stemmer godt overens med massespekteret fra NIST 17 biblioteket det har blitt sammenlignet med (se figur 16).



Figur 15: Massespekteret for ekstrahert signalet i prøve 18_2678 ved 23.5356 min.

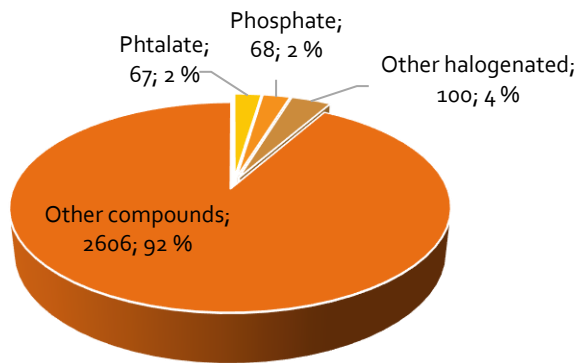


Figur 16: Massespekteret for 2-Propanol, 1-chloro-, phosphate (3:1), en komponent som har omtrent lik retensjonstid og lignende MS spektre som det ekstrahert spektre for en komponent i kjørt analyse. Overbestemmelse av likhet mellom disse spektrene er bestemt av betingelsen; matchfaktor. Jo høy matchfaktoren er mellom dem, jo tydelig er det at komponenten er 2-propanol, 1-chloro-, phosphate (3:1).

4.2.1 Oversikt identifiserte komponenter

Av de 2842 identifiserte komponentene, ble det funnet noen miljøgifter. Blant annet ftalater, flammehemmere og andre halogenerte komponenter. Hvor stor prosentandel de utgjør av det totale komponenter er det vist i figur 17.

Overview identified compounds



Figur 17: Viser en kategoriskoversikt over identifiserte komponenter i forhold til ftalater, fosfater samt resterende halogenerte forbindelser.

I tabell 2 er det mulig å se noen av de mest prioriterte miljøgiftene i samfunnet. Disse miljøgiftene er noen av komponentene som befinner seg i diagrammet vist i figur 16, i blant ftalat, fosfat og andre halogenerte komponenter.

Tabell 2; Pivottabell over miljøgifter fra prioritetslisten, funnet i de 26 prøvene.

Number of Compounds	Sample																		Grand Total				
	18_2675	18_2676	18_2677	18_2678	18_2679	18_2681	18_2682	18_2683	18_2684	18_2685	18_2686	18_2687	18_2688	18_2689	18_2690	18_2691	18_2694	18_2695		18_2696	18_2697	18_2698	18_2699
Compound Name																							
Benzyl butyl phthalate	1				1		1															1	4
DIISOBUTYL PHTHALATE										1	1	1	1	1	1	1	1						8
Phenol, 4,4'-(1-methylethylidene)bis- (Bisfenol A)				1																		1	2
Phenol, pentachloro- (PCP)		1		1				1															3
Tri(2-chloroethyl) phosphate	1		1		1	1			1									1	1	1	1		9
Triclosan	1					1																	2
Grand Total	3	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	28

4.2.2 Pivottabell av identifiserte komponenter

Fra excel-arket med kvalitetssikret identifiserte komponenter, uten duplikater, ble det lagd en pivottabell, se supplement 1. Pivottabellen består av 909 ulike identifiserte komponenter og viser gjentakelse av samme komponent i hver prøve, samt den totale gjentakelsen. Navnet på komponentene står i første kolonne og type støvprøve står i øverste rad. Tabell 3 viser et utsnitt av pivottabellens øverste del, med de mest gjentatte kjemikaliene i lista. Blant disse, står kjemikaliene tert-butylamine øverst og er den eneste som har dukket opp i alle prøvene.

Tabell 3; Viser en pivottabell over de første kjemikaliene som gjentar seg mest totalt i alle de 26 prøvene.

Number of Compounds	Sample																										Grand total
	18_2675	18_2676	18_2677	18_2678	18_2679	18_2680	18_2681	18_2682	18_2683	18_2684	18_2685	18_2686	18_2687	18_2688	18_2689	18_2690	18_2691	18_2692	18_2693	18_2694	18_2695	18_2696	18_2697	18_2698	18_2699	18_2700	
TERT-BUTYLAMINE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	28
Phthalimide	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	25
Diethyl Phthalate		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	25
Benzophenone	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	25
INDOLE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	23
Benzoic acid	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	23
Triphenyl phosphate	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1			1	1	1			19	
METHYL ANTHRANILATE						1	1	1		1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	19
Nonanal	1	1	1	1	1	1				1	1			1	1	1	1	1	1		1		1	1	1	1	19
Campesterol	1	1	1	1	1	1	1	1	1	1					1	1		1		1	1	1	1	1	1	1	19
Phenol	1	1	1	1			1	1		1				1	1	1	1	1	1	1	1	1	1	1	1	1	18
ORTHO-METHOXYPHENOL						1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1					16
Tri(2-chloroethyl) phosphate	1	1	1	1	1	1	1	1	1	1											1	1	1	1	1	1	16
Dibutyl phthalate	1		1	1	1	1	1	1	1	1											1	1	1	1	2	1	16
2-Propanol, 1-chloro-, phosphate (3:1)	1	1	1	1	1	1	1	1	1	1											1	1	1	1	1	1	16

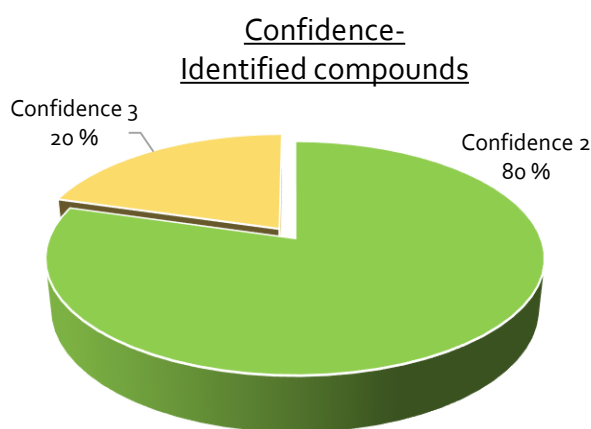
Det ble også lagd en pivottabell for samlet data fra det første forsøket i Unknowns Analysis. Pivottabellen viste at antall total ulike identifiserte komponenter er lik 1591. Se tabell 4.

Tabell 4; Viser antall ulike identifiserte komponenter i forsøk 1 og 2 i Unknowns Analysis

Kjemikalieliste	Antall total ulike identifiserte komponenter
Samlet data, forsøk 1	1591
Samlet data, forsøk 2	909

4.2.3 Konfidens identifiserte komponenter

Den prosentvise fordelingen av de identifiserte komponentene med konfidens nivå 2 og 3 er vist i figur 17. Dette ble gjort etter kvalitetssikringen og godkjenningen av de identifiserte komponentene fra Unknowns Analysis. Konfidensen er anslått ved hjelp av konfidens Tabellen vist i kapittel 2.5, figur 7. 80% av de 2842 identifiserte komponentene har konfidensnivå 2. Resten. Mens resten, 20% har konfidensnivå 3.



Figur 18: Viser prosentfordeling av konfidens til identifiserte komponentene, etter manuell kvalitetssikring.

4.3 PCA- ANALYSE AV ALLE KOMPONENTER (IDENTIFISERTE OG UIDENTIFISERTE)

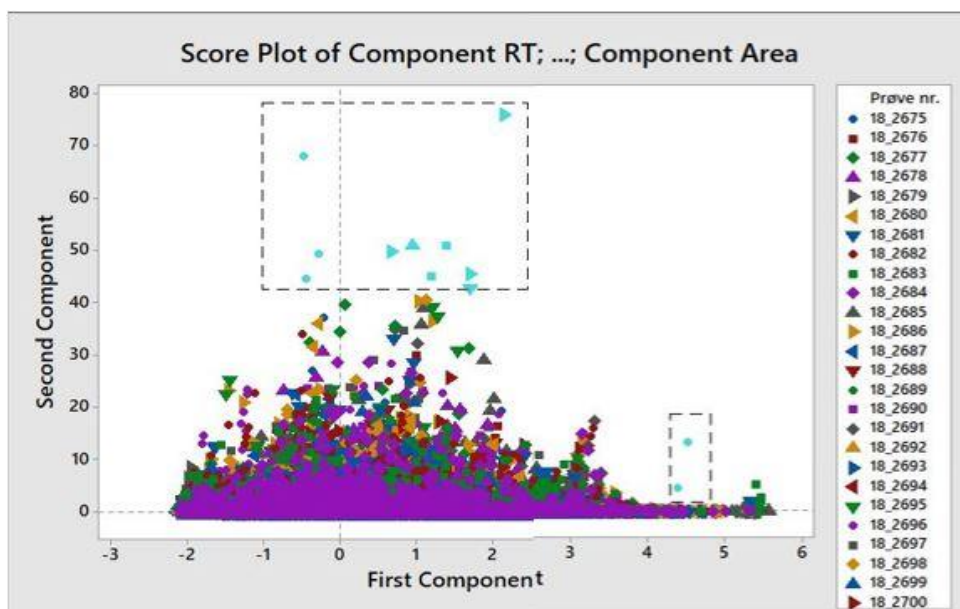
Et PCA- plott ble lagd ved bruk av variablene på komponent RT, «Base Peak m/z» og Komponentareal for de identifiserte samt en del uidentifiserte komponenter. Plottet er vist i figur 20. PCA-plottet består av 347 509 punkter og de er gruppert i form av farge og form i forhold til prøvenummer.

Standardavviket, andel varians og den dynamiske delen av varians for PC1, PC2, og PC3, samt fordelingen av de ulike variablene er vist i figur 19.

Eigenanalysis of the Correlation Matrix			
Eigenvalue	1,4463	1,0011	0,5527
Proportion	0,482	0,334	0,184
Cumulative	0,482	0,816	1,000

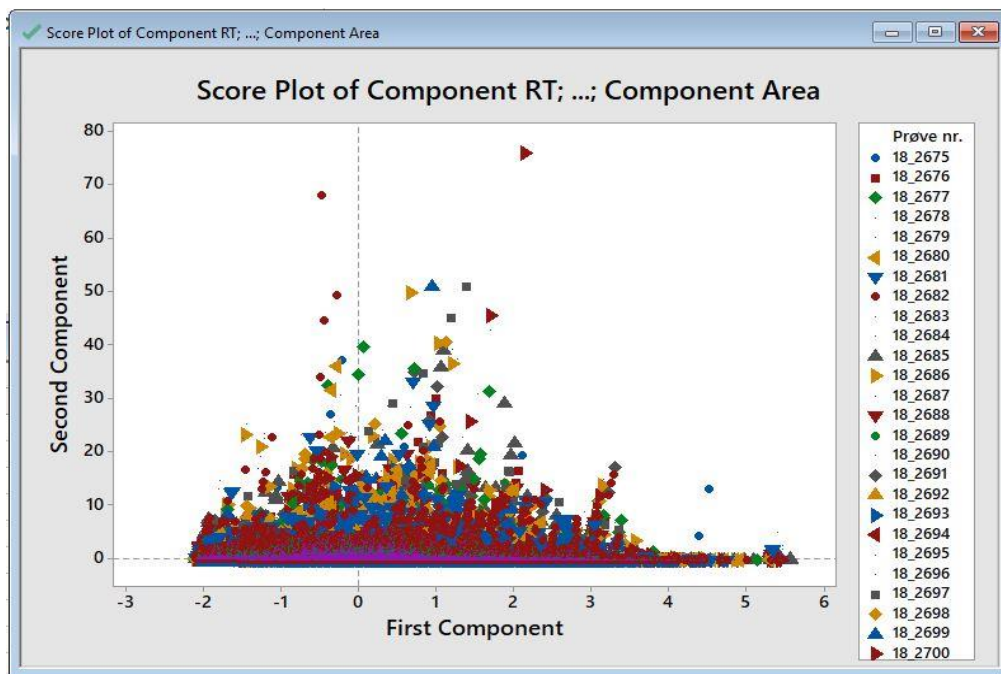
Eigenvectors			
Variable	PC1	PC2	PC3
Component RT	0,707	0,026	0,706
Base Peak MZ	0,707	-0,046	-0,706
Component Area	0,014	0,999	-0,051

Figur 19: Viser egenanalysen og egenvektorene til de ulike variablene for PCA-plottet.



Figur 20: Viser et PCA-plott med 347 509 punkter. Plottet er basert på punktenes komponent RT, «Base Peak m/z» og komponentareal. De to aksene, «First Component» og «Second Component» er en kombinasjon av PC1, PC2 og PC3.

I figur 21 har vi det samme PCA-plottet, men med færre punkter for å kunne se om noen punkter har skjult eller skillet seg ut under det lilla farget haug med punkter.



Figur 21: Viser samme PCA-plott som vist i figur 17, men med færre synlige punkter.

Figur 22 og 23 viser data for punktene markert i figur 20, i form av ulike ID variabler. ID variablene er prøve nummer, komponent navn, komponent RT, «Base Peak m/z» og komponentareal

Figur 22 viser data for de to punktene nederst til venstre. Fra likheten i Prøve nr., «Component RI» og «Base Peak MZ» er det mulig å identifisere «Unknown 13554» som et duplikat av; «Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, octadecyl ester»

Prøve nr.	Compound Name	Component RT	Base Peak MZ	Component Area
18_2675	Unknown 13554	54,1499	530,468	145450501
18_2675	Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, octadecyl ester	54,1490	530,468	434067954

Figur 22: Viser ID variablene, prøve nr., «Compound Name, Component RT, Base Peak MZ, Component Area», over de markerte punktene nederst til høyre i figur 20.

Figur 23 viser data for punktene markert i øverst midten i PCA-plottet, se figur 20. Alle er ukjente utenom en, prøve 18_2699, som er «dioctylphtalate». Det er ikke mulig å trekke noen likheter mellom dette punktet og de andre ukjente punktene.

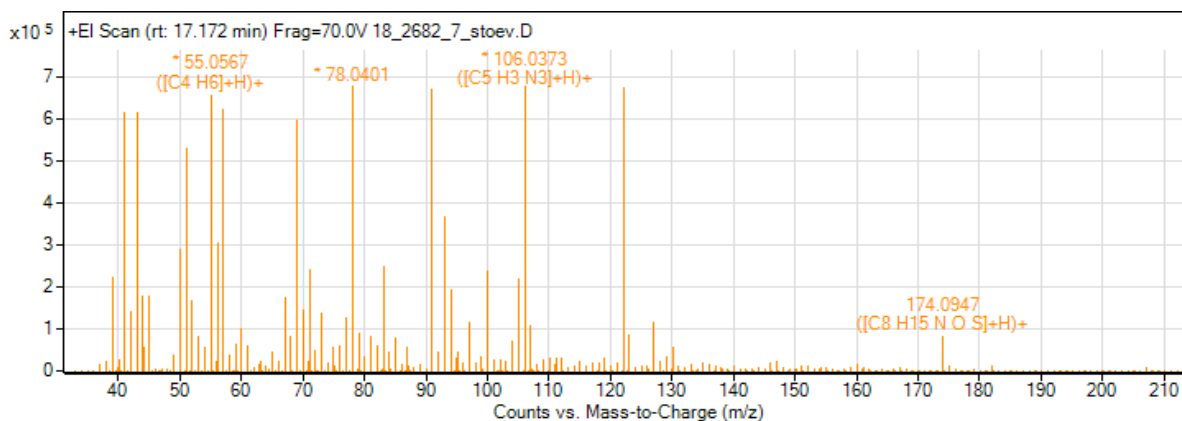
Prøve nr.	Compound Name	Component		Base	Component Area
		RT	Peak MZ	Peak MZ	
18_2686	Unknown 6162	23,2014	185,156	185,156	1613907973
18_2682	Unknown 3154	17,1040	94,054	94,054	1445865050
18_2682	Unknown 3155	17,1040	50,016	50,016	2201469177
18_2682	Unknown 3156	17,1040	106,030	106,030	1599052002
18_2695	Unknown 11876	41,6093	213,169	213,169	1391678887
18_2697	Unknown 11376	41,8965	145,104	145,104	1463000423
18_2697	Unknown 11380	41,9071	159,120	159,120	1652654075
18_2700	Unknown 11323	41,9101	207,038	207,038	2462217464
18_2700	Unknown 11324	41,9101	207,038	207,038	1475066821
18_2699	dioctylphtalate	34,6117	149,027	149,027	1648152174

Figur 23: Viser ID variablene, prøve nr., «Compound Name, Component RT, Base Peak MZ, Component Area», over de markerte punktene øverst i midten i figur 20.

4.3.1 Identifikasjon av ukjente komponenter fra PCA-plott

I PCA-plottet er det en del punkter som skiller seg ut. De fleste er uidentifiserte punkter. Ved *in-silico* identifisering har 3 større fragmenter av punktene blitt identifisert. Det er antatt at punktene ikke representerer molekylionet og dermed heller ikke stoffet i støvprøven punktet stammer fra. *in-silico* identifisering tar utgangspunkt i punktene «Unknown 3154, Unknown 3155 og Unknown 11323», se figur 23. Fra Qualitative Analysis ble først MS-spektre de tre punktene stammer fra ekstrahert.

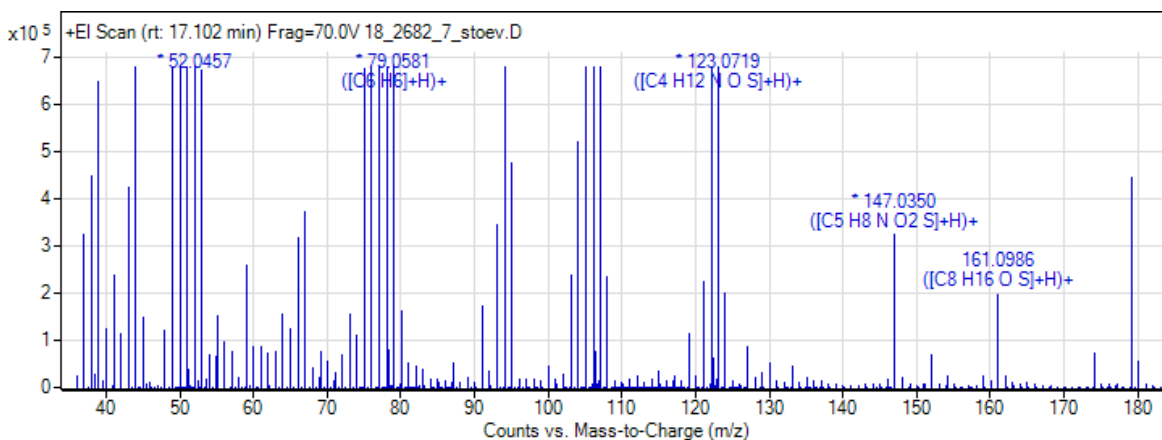
Figur 24-25 viser ekstraherte MS- spektre som hvert av de tre punktene. De ulike ekstraherte massespektrene inneholder fragmenter med lik masse som de ukjente punktene.



Figur 24: Viser massespekteret for uidentifisert komponent 1 med retensjonstid lik 17.172 min.

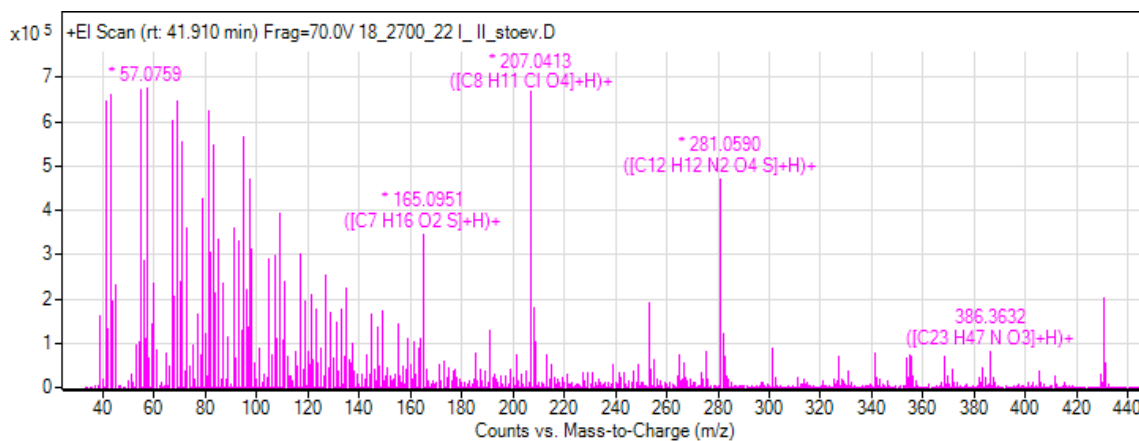
Figur 24 viser massespekteret til «Unknown 3154». Massespektret er ekstrahert fra en topp med retensjonstid omtrent lik 17,172 min. Toppen er funnet fra et kromatogram, tilhørende prøvenummeret for punktet, som har fått ekstrahert «Base Peak» som tilhører punktet. «Base Peak» til «Unknown 3154» er 94,054 m/z, se figur 23. I MS-spekteret i figur 24 er ikke «Base peak» oppgitt for punktet det linjesignalet med høyest m/z. Linjesignalet med 94,054 m/z ligger blant de mindre fragmentene. Lenger ned i teksten vises det identifiserte navnet på molekylformelen med m/z lik 174.0947.

Figur 25 viser massespekteret for «Unknown 3155» og figur 26 viser «Unknown 11323». Begge MS-spektrene følger likt prinsipp som MS-spekteret i figur 24. Observasjonene er også like, de ukjente punktene er mindre fragmenter av større fragmenter.



Figur 25: Viser massespekteret for en uidentifisert komponent 2 med retensjonstid lik 17.102 min.

I figur 25 er MS-spekteret ekstrahert for en topp omtrent ved 17,102 min. MS-spekteret er ekstrahert med hensyn til punktet sitt «base peak» på 50,016 m/z. Navnet for molekylformelen med 161,0986 m/z er vist lenger nedi teksten.



Figur 26: Viser massespekteret for en uidentifisert komponent 3 med retensjonstid lik 41.910 min.

Figur 26 er MS-spekteret ekstrahert for en topp omtrent ved 41,910 min. MS-spekteret er ekstrahert med hensyn til punktet sitt «base peak» på 207,038 m/z. Navnet for molekylformelen med 386,363 m/z er vist lenger ned i teksten.

Etter å ha søkt molekylformelene til de tre ukjente større fragmentene, fra figur 24-26, i ChemSpider, kom det opp liste av kjente komponenter. Disse er arrangert etter hvilke av de som har høyest sannsynlighet til å være en riktig komponent. Ved å plukke komponentene med størst sannsynlighet fra PubChem og sammenligne *in-silico* MS-spektre fra CFM-ID og MS-spektrene i figur 24-26, ble fragmentene identifisert, se tabell 5.

Tabell 5: Viser hvilket større fragment hvert av de 3 punktene som skiller seg ut, stammer fra. Samt vises antall referanser tilgjengelig av hvert molekyl i ChemSpider. Høyere referanse gir høyere troverdighet til det bestemte navnet til molekylformlene.

Punkt	Navn på større fragment	Antall referanser
Unknown 3154,	« (4S)-2,2-Dimethyloxane-4- ^c (C ₈ H ₁₅ NOS) »	31
Unknown 3155	« 2-Methyl-4-propyl-1,3-oxathiane (C ₈ H ₁₆ OS) »	98
Unknown 11323	« N-[2-(2-hydroxyethoxy) propyl]octadecanamide (C ₂₃ H ₄₇ NO ₃) »	12

5. DISKUSJON

Ved bruk av Unknowns Analysis ble det ekstrahert omtrent 14 000- 20 000 komponenter i hver støvprøve, se figur 11. Ved å stramme inn betingelsene har antall ekstraherte komponenter minket med 600 – 1000 komponenter i hver prøve.

Antall komponenter som har gitt hits i bibliotekene er omtrent 200-300 komponenter for hver prøve, se figur 12. Etter endret betingelser har antallet blitt halvert. Vi kan se at endringene i betingelsene har hatt større betydning for antall komponenter som får treff i bibliotekene, enn totalt ekstraherte komponenter.

Av de optimaliserte betingelsene i forsøk 2 har «ion peaks hatt» en betydning for antall ekstraherte komponenter. Mens «matchfactor» har filtrert likheten mellom massespektre ekstrahert for hver komponent og massespektrene fra bibliotekene. Her har vi «matchfactor» som ble optimalisert fra 80% til 85%. «Minimum ion Peaks» som ble optimalisert fra 3 til 5. Videre i arbeidet har vi oppdaget at «matchfactoren» ikke gir et helt godt grunnlag for hvor høy konfidens det er til den identifiserte komponenten. Det har blitt oppdaget at identifiserte komponenter med lav «matchfactor» kan høyere troverdighet, enn komponenter med høyere «matchfactor».

«Matchfactor» gir et tall på hvor godt de ulike massespektrene ligner hverandre, men for å gi et godt nok grunnlag bør den basere seg på hvor stor likhet det er i intensiteten til like linjesignaler som er til stedet i begge massespektrene. I den manuelle kvalitetssikringen, har vi derfor ikke tatt så stort hensyn til matchfactoren.

Komponentene ekstrahert med optimaliserte betingelser i Unknowns Analysis ble bearbeidet videre. Av de identifiserte komponentene ble 0.32% identifisert med det høyt oppløselige pesticidbiblioteket og har en konfidens på 2. Resterende ble identifisert med de lavoppløselige bibliotekene: NIST.17 og MassBank, se figur 13. Disse har en konfidens på 3 og har videre blitt kvalitetssikret gjennom en kombinasjon av manuell gjennomgang av komponentenes «ionpeaks», massespekter i forhold til det sammenlignede massespekteret fra bibliotekene, samt beregnede Kovart RI verdier. Sett på den beregnede Kovats RI verdien for en komponent som gjentar seg i de andre prøvene, skulle verdien være omtrent likt. Kovats RI verdier ble ikke beregnet for komponenter med komponent RT under 8.54 min og over

39.84 min. Disse ble ikke kvalitetssikret, fordi alkanmiksen ikke dekket retensjonstider utenfor disse intervallene, som gjorde det umulig å bruke likning 2.

5.1 Bearbeiding identifiserte komponenter fra Unknowns Analysis

Den manuelle kvalitetssikringen fjernet for det meste duplikater av ulike komponenter. Totalt ble det fjernet 205 komponenter. De resterende komponentene som har blitt manuell kvalitetssikret har fått økt sin konfidens til 2. Komponentene som ikke fikk beregnet Kovats RI ligger fortsatt på konfidens 3, siden vi ikke hadde et godt grunnlag for å øke konfidensen. Totalt har det blitt identifisert 909 ulike komponenter.

For identifisering av alkantoppene i alkanmiksen ble dataprogrammet Qualitative Analysis brukt. Alkanmiksen skulle inneholde 24 alkantyper, men det ble bare detektert 19 topper. Dataprogrammet var koblet opp mot et NIST bibliotek som viser en liste over hvilke stoffer i hver topp det er detektert etter mengde. Stoffer med høyst mengde viste seg høyest i lista. I flere av toppene var det mulig å se at like stoffer gjentok seg øverst i listene. Noen av stoffene kan ha hatt metningsproblemer i GC-kolonnen, og det kan ha skjedd koeluering.

Fra pivottabellen er det mulig å se hvilke av de identifiserte komponentene som gjentar seg totalt mest i prøvene, se tabell 3. Det høyeste tallet ligger på 28, mens det laveste ligger på 1. Det er større troverdighet til komponenten som gjentar seg mest. Tert-butylamin er komponenten som gjentok seg mest totalt, med 28 ganger. Dette stoffet er blant annet et mellomprodukt som brukes for å danne gummi, insektsmidler fargestoffer, legemidler (NIH, U.S. National Library of Medicine, 2017). To andre stoffer høyt på pivottabellen, med 25 gjentakelser, er Dietyl ftalat (DEP) og Benzofenon. Dietyl ftalat (DEP) finnes i en del plantevernmidler og kosmetikk (NIH, U.S. National Library of Medicine, 2008). Benzofenon brukes som duftforstærker og finnes i såpe, samt brukes det i blekk, bilder, maling for UV-beskyttelse (NIH, U.S. National Library of Medicine, 2015)

Siden vi har en hard ionisasjonsteknikk, er det en stor mulighet for at molekylionene i de forskjellige massespektrene har forsvunnet, og at vi heller har fått identifisert en del mulige derivater av det egentlige stoffet. Fra GC-instrumentet er det også mulig at det har skjedd

koeluering i kolonnen. En del ion-peaks kan dermed ha fragmenter som tilhører et annet molekyl.

5.2 Klassifisering av identifiserte stoffer

Av de identifiserte stoffene er 2% ftalater, 2% fosfater og 4 % andre halogenerte komponenter. I disse identifiserte stoffene ble 5 av miljøgiftene fra prioritetslisten funnet (stoffer som er satt til å bli fjernet innen 2020 av miljødirektoratet). Miljøgiften med flest gjentakelser er den fosfororganiske flammehemmeren og plastmykningsmiddelet, tri(2-kloretyl) fosfat. Deretter kommer Benzyl butyl ftalat (DBP) og diisobutyl ftalat (DIBP). Disse er brukt å lage vesker, klær, sko, leketøy, kabler som inneholder PVC plast. Videre har vi Fenol, pentaklor- (PCP), som finnes i treverk, Bisfenol A, som kommer primært fra plast, og triklosan, som man finner i kosmetikk og såpe.

5.3 Bearbeiding, alle komponenter fra Unknowns Analysis

PCA- plottet ga en god visualisering av samvariasjonene til alle de uidentifiserte og identifiserte komponentene. Komponent RT, «Base Peak m/z» er omtrent lik for de like komponenter. Mens den tredje variabelen; komponentareal for noen like komponenter kan variere. Fra figuren 19, kan vi se at andel varians i den to-dimensjonale PCA-plottet baserer seg størst på PC1, med 48,2 %. PC1 er en lineær vektor som endrer seg mest i Component RT og Base Peak MZ i økende retningen, med 0.707 for begge.

Fra plottet kan vi se at de fleste punktene ligger i en vannrett lineær haug rundt første aksene. Punkter som ligger og avviker mest fra denne haugen er unike punkter som skiller seg mest ut. I motsetning til disse punktene er det vanskelig å si i god grad hvilke av punktene som skiller seg ut, siden vi har så mange. Det er 12 komponenter som skiller seg mest ut i PCA-plottet.

Det er totalt 347 509 punkter i plottet. Ved slutten av arbeidet har det blitt oppdaget at PCA- figuren mangler en del av komponentene fra prøve 18_2689 til og med 18_2694. Dette påvirker ikke resultater fått fra analysen av PCA- plottet. Punktene som skiller seg ut er

tilhører komponenter som kommer fra andre prøver. Dersom vi hadde tatt med punktene vi mangler, kan det hende at vi ville hatt flere punkter som skiller seg ut.

Av de uidentifiserte komponentene er det en del duplikater. På grunn av stor andel uidentifiserte komponenter (ca. 344 666) ville det vært tidskrevende å gå gjennom dem og fjerne duplikatene. En mulig forbedring kan være eventuelt å fjerne disse duplikatene, fordele punktene fra de 26 prøvene i flere PCA-plott. Samt gruppere dem i farge og form etter komponentnavn («Base Peak», komponent RT). Isteden for prøvenummer som er gjort i dette tilfelle. Da er det mulig å se klynger med punkter som kjemisk bør ligne på hverandre. Et annet tiltak for å få bedre visualisering er å få til et 3 dimensjonalt PCA plott. Plottet ville da vist variansen til punktene sett fra PC1, PC2 og PC3.

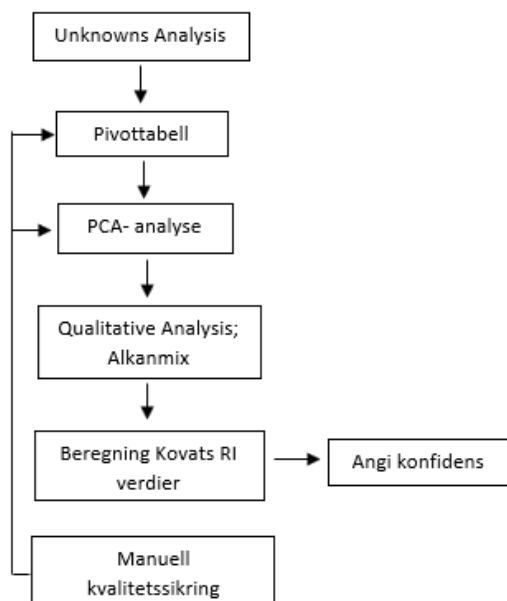
Et av de unike punktene som skilte seg ut, ble identifisert til å være et duplikat av «Benzenepropanoic acid, 3,5-bis(1,1-dimethyl)-4hydroxy-, octadecyl ester». Tre av punktene som skilte seg mest ut ble bearbeidet videre og identifisert ved *in-silico* spektre. Disse ble bestemt til å være fragmenter av større molekyler « (4S)-2,2-Dimethyloxane-4-‘ ($C_8H_{15}NOS$), N-[2-(2-hydroxyethoxy)propyl]octadecanamide ($C_{23}H_{47}NO_3$) og 2-Methyl-4-propyl-1,3-oxathiane ($C_8H_{16}OS$) ».

I spektrene kan vi se at de genererte formlene til noen av linjesignalene som har blitt ekstrahert med de ulike «Base Peak» verdiene og riktig komponent RT har et stjerne-kryss. Krysset indikerer at det har skjedd en metning av signalene. Metning gir avvik, men siden vi har valgt en ppm på 0,20 for ekstraksjonen, har det ikke mye å si i dette tilfelle. Spektrene for hvert stoff hentet fra CFM-ID som har ekstrahert spekter med *in-silico* teknikk, ble sammenlignet med spektre hentet fra Qualitative Analysis. Hvor alle ga positivt resultat. Denne metoden kan tilføyes i arbeidsflyt for å identifisere resterende komponenter som skiller seg ut i PCA-figuren.

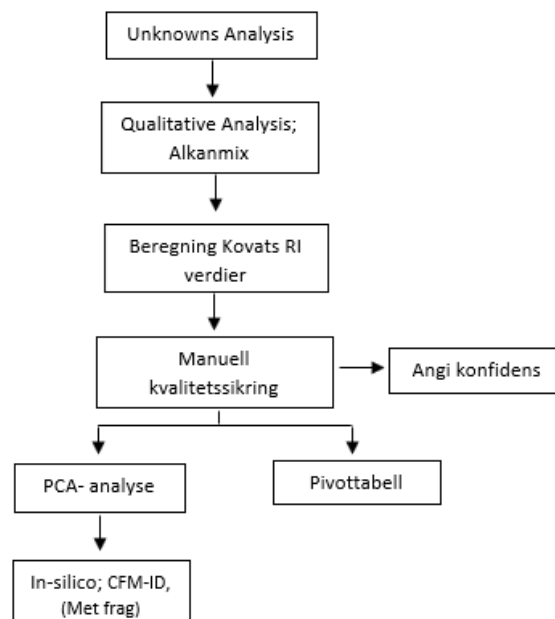
5.4 Arbeidsflyt;

Arbeidsflyt har blitt satt sammen underveis i oppgaven. Over de første komponentene som ble ekstrahert fra Unknowns Analysis ved de første betingelsene, ble det lagd en pivottabell og PCA- plott. Fra disse resultatene så vi at vi burde lage pivottabellen og PCA-plottet på nytt etter en manuell kvalitetssikring. Fordi, det ble for mange duplikater av lik komponent.

Duplikatene ga for eksempel for mange unødvendige punkter i PCA-plottet. Ved den manuelle kvalitetssikringen ble det bestemt å utføre ekstraksjonen på Unknowns Analysis på nytt, med strengere krav. Siden de fleste identifiserte komponenter med mindre enn «5 ion peaks» var utroverdige komponenter. Ved andre forsøk ble pivottabell og PCA- analyse flyttet lenger ned i arbeidsflyt. Den manuelle kvalitetssikringen ble også mindre tidskrevende fordi vi hadde færre komponenter å forholde oss til. I arbeidsflyt har også *in-silico* metoden blitt tilføyd for å bearbeide resultater fra PCA- analysen.



Figur 28: Viser den første arbeidsflyt for forsøket



Figur 27: Viser den siste fornyet arbeidsflyt, brukt ved 2. forsøk.

Arbeidsflyten kan forbedres ved å tilføye en metode for å visualisere de halogenerte forbindelsene i prøvene, og sjekke om det er andre uidentifiserte halogenerte forbindelser i prøven. Et eksempel er Kendrick massdefekt analyse. Denne ville passet inn som en egen del etter den manuelle kvalitetssikringen. Analysen går ut på å lage et plott hvor en beregnet Kendrick massdefekt for ulike komponenter er plottet som en funksjon av beregnede nominelle Kendrick masse-verdier. Ioner fra en bestemt kjemiskfamilie og en homologserie vil ha lik Kendrick massdefekt. Tatt hensyn til dette er det mulig å plukke fram hvilke punkter som er halogenerte forbindelser (Ubukata, et al., u.d.).

6. KONKLUSJON

I denne kjemiskscreeningen av støvprøvene har det blitt påvist en del komponenter. Hvor noen har blitt merket som helseskadelige miljøgifter. Ved miljø og luft forskning er det viktig å kjøre en non-target analyse, siden vi ikke vet hva vi er ute etter. Med en slik analyse er det muligheten for å ekstrahere alle komponenter som opprinnelig finnes i støvprøvene våre og det er mindre tidskrevende og kostbart enn å kjøre hypotesebaserte target analyser.

Programmet Unknowns Analysis ekstraherte mange gode komponenter, som ga et godt grunnlag for non-target analyse og dataprosessering. Av de identifiserte ekstraherte komponenter var det 6,7% som ble fjernet ved den manuelle kvalitetssikringen. Det ble oppdaget at strengere optimaliserte betingelser i Unknowns Analysis ga bedre og sikrere resultater. Totalt ble det identifisert 909 ulike komponenter. Totalt ble det identifisert 2842 komponenter, hvorav 8% er; ftalater, fosfater og halogenerte komponenter. Av de kvalitetssikret og godkjente komponenter, har vi anslått at 80% av de ligger på konfidensnivå lik 2 og 20% av de på konfidensnivå lik 3. Av de ekstraherte komponentene er bare ca. 0.9% identifisert med bibliotekene.

En PCA- analyse ga en god visualisering av alle de ekstraherte komponentene og viste at 12 komponenter skilte seg mest ut. Bruk av *in-silico* MS-spektre gjort det mulig å identifisere de ukjente komponentene som skilte seg i PCA figuren. Tre av disse ukjente komponentene ble identifisert som mindre fragmenter av « (4S)-2,2-Dimethyloxane-4-‘ ($C_8H_{15}NOS$), N-[2-(2-hydroxyethoxy)propyl]octadecanamide ($C_{23}H_{47}NO_3$) og 2-Methyl-4-propyl-1,3-oxathiane ($C_8H_{16}OS$) ».

Referanser :

Agilent Technologies, 2010. *Agilent MassHunter Workstation Software- Quantitative Analysis*, Santa Clara: Agilent Technologies Inc.

Agilent Technologies, 2011. *MassHunter Personal Compound Database and Library Manager*, s.l.: Agilent Technologies Inc.

Agilent Technologies, 2015. *Agilent MassHunter Workstation - Unknowns Analysis*, Santa Clara: Agilent Technologies Inc.

Agilent Technologies, 2017. *Agilent 7250 Accurate- Mass/ Q-TOF GC/MS System*, USA: Agilent Technologies, Inc..

Agilent Technologies, 2017. *SureMass*, USA: Agilent Technologies Inc.

Agilent Technologies, 2019. *MassHunter Software*. [Online]
Available at: <https://www.agilent.com/en/products/software-informatics/masshunter-suite/masshunter/masshunter-software>
[Accessed 18 February 2019].

Blaženović, I. et al., 2017. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: database boosting is needed to achieve 93% accuracy. *Journal of Cheminformatics*, 15 May.

Bruker Daltonics, s.d. *Principal Component Analysis Basics*, s.l.: Bruker Daltonics.

CFM-ID, n.d. *Competitive Fragmentation Modeling for Metabolite Identification*. [Online]
Available at: <http://cfmid3.wishartlab.com>
[Accessed 2019 April 8].

Friedrich, D., 2016. The problems won't go away: Persistent Organic Pollutants (POPs) in the Arctic. *THE ARCTIC INSTITUTE*, 1 July.

Frøslie, K. F., 2018. *Korrelasjon*. [Online]
Available at: <https://snl.no/korrelasjon>
[Accessed May 6 2019].

Harris, D. C., 2016. *Quantitative Chemical Analysis*. 9 éd. New York: W. H. Freeman and Company.

IUPAC, 2014. *Retention Index*. [Online]
Available at: <https://goldbook.iupac.org/html/R/R05360.html>
[Accessed 4 February 2019].

IUPAC, 2019. *THE IUPAC INTERNATIONAL CHEMICAL IDENTIFIER (INCHI)*. [Online]
Available at: <https://iupac.org/who-we-are/divisions/division-details/inchi/>
[Accessed 8 April 2019].

Jaardi, Z., 2019. *A step by step explanation of Principal Component Analysis*. [Online]
Available at: <https://towardsdatascience.com/a-step-by-step-explanation-of-principal-component-analysis-b836fb9c97e2?gi=d286a3fc1c8a>
[Accessed 6 May 2019].

Johnson, D. H., 2006. Signal-to-noise ratio. *Scholarpedia*.

- Ledermann, N., 2018a. *Fakta om husstøv*. [Online]
Available at: <https://www.lhl.no/sunnere-bolig/rengjoring/fakta-om-husstov/>
[Accessed 23 January 2019].
- Ledermann, N., 2018b. *Husstøv og innemiljø*. [Online]
Available at: <https://www.lhl.no/sunnere-bolig/rengjoring/husstov-og-innemiljo/>
[Accessed 23 January 2019].
- Liu, R. & Mabury, S. A., 2019. Identification of Photoinitiators, Including Novel Phosphine Oxides, and Their Transformation Products in Food Packaging Materials and Indoor Dust in Canada. *Environmental science & technology*, 16 April, 53(8), pp. 4109-4118.
- McKenna, A., 2015. Tandem Mass Spectrometry (MS/MS). *NATIONAL MAGNETIC FIELD LABORATORY*, 15 September.
- Miljødirektoratet, 2017a. *Bromerte flammehemmere*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/prioritetslisten/bromerte-flammehemmere/>
[Accessed 28 January 2019].
- Miljødirektoratet, 2017b. *Dioksiner*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/prioritetslisten/dioksiner/>
[Accessed 28 January 2019].
- Miljødirektoratet, 2017. *Miljøgifter*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/>
[Accessed 21 January 2019].
- Miljødirektoratet, 2017. *Pentaklorfenol (PCP)*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/prioritetslisten/pentaklorfenol-pcp/>
[Accessed 30 January 2019].
- Miljødirektoratet, 2018. *Ftalater*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/prioritetslisten/ftalater/>
[Accessed 28 January 2019].
- Miljødirektoratet, 2019. *Fosfororganiske flammehemmere*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/prioritetslisten/fosfororganiske-flammehemmere/>
[Accessed 28 January 2019].
- Miljødirektoratet, 2019. *Prioritetslisten*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/prioritetslisten/>
[Accessed 28 January 2019].
- Moschet, C. et al., 2018. Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution Mass Spectrometric Study. 6 March.
- NIH, U.S. National Library of Medicine, 2008. *HSDB: DIETHYL PHTHALATE*. [Online]
Available at: <https://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+84-66-2>
[Accessed 9 May 2019].

NIH, U.S. National Library of Medicine, 2015. *HSDB; BENZOPHENONE*. [Online]
Available at: <https://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+119-61-9>
[Accessed 9 May 2019].

NIH, U.S. National Library of Medicine, 2017. *HSDB: TERT-BUTYLAMINE*. [Online]
Available at: <https://toxnet.nlm.nih.gov/cgi-bin/sis/search2/r?dbs+hsdb:@term+@rn+@rel+75-64-9>
[Accessed 9 May 2019].

NILU, 2019. *NILU*. [Online]
Available at: <https://www.nilu.no>
[Accessed 6 May 2019].

Norges Geologiske Undersøkelse, 2019. Organiske miljøgifter. *Norges Geologiske Undersøkelse*, 8 May.

Norsk Polarinstitutt, 2017. *Miljøgifter i Arktis*. [Online]
Available at: <https://www.miljostatus.no/tema/kjemikalier/miljogifter-i-arktis/>
[Accessed 27 February 2019].

Ouyang, X. et al., 2017. Non-target analysis of household dust and laundry dryer lint using comprehensive two-dimensional liquid chromatography coupled with time-of-flight mass spectrometry. *Chemosphere*, January, pp. 431-437.

Rantakokko, P. et al., 2019. Concentrations of brominated and phosphorous flame retardants in Finnish house dust and insights into children's exposure. *Chemosphere*, May, Volume 223, pp. 99-107.

Reusch, W., 2013. *Chemistry msu education*. [Online]
Available at:
<https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/Spectrpy/MassSpec/masspec1.htm>
[Accessed 6 March 2019].

Riener, D. J., 2016. *The Latest Development in MassHunter for GC/Q-TOF*. [Online]
Available at: <https://docplayer.net/54978713-The-latest-developments-in-masshunter-for-gc-q-tof-insights-from-an-agilent-expert-speaker-dr-jorg-riener-senior-gc-ms-scientist.html>
[Accessed May 2019].

Rønning, P. O., 2017. *Innføring i massespektrometri*, Oslo: OsloMet.

Samanipour, S. et al., 2019. Optimism for nonterget analysis in enviromental Chemistry. *Enviromental Science Technology*, 9 May .

Schlabach, M., 2013. *Non-target screening - A powerful tool for selecting environmental pollutants*, Oslo: Miljødirektoratet.

SHIMADZU, n.d. *Linear Retention Index (LRI)*. [Online]
Available at: https://www.shimadzu.com/an/linear_retention_index.html
[Accessed 21 May 2019].

Shoeib, T., Webster, G. M., Hassan, Y. & Tepe, S., 2019. Organophosphate esters in house dust: A comparative study between Canada, Turkey and Egypt. *The Science of the total environment*, 10 February, Volume 650, pp. 193-201.

SIS, 2017a. *NIST 17- Mass Spectral Library*. [Online]
Available at: <https://www.sisweb.com/software/ms/nist.htm>
[Accessed 18 February 2019].

Stein, 2012. *In silico*. [Online]
Available at: <https://fremmedord.org/hva-betyr/in-silico/>
[Accessed 20 May 2019].

Stensbøl, C. C. F. & Kirkemo, P. I. B., 2018. *Global non-target screening av støvprøver*, Oslo: OsloMet-Storbyuniversitetet.

Sun Yu, L. L.-y., 2019. Organophosphate flame retardants in college dormitory dust of northern Chinese cities: Occurrence, human exposure and risk assessment. *The science of the total environment*, 15 May, Volume 665, pp. 731-738.

Ubukata, M. et al., s.d. *Analysis of electronic waste*, Eindhoven: JBS.

Velázquez-Gómez, M., Lacorte, S. & Hurtado-Fernández, E., 2018. Comprehensive method for the analysis of multi-class organic micropollutants in indoor dust. *Science of The Total Environment*, 1 September, pp. 1484-1494.

Vonk, N., Avans & Breda, 2016. Qualitative analysis: identification. *CHEMISTRY LibreTexts*, 24 October.

Yap, B. K. et al., 2018. In Silico Identification of Novel Inhibitors. *Encyclopedia of Bioinformatics and Computational Biology*, 6 September, pp. 761-779.

Zhang Mengtao, S. J. M. Y. G. W. L. h., 2019. Occupational exposure characteristics and health risk of PBDEs at different domestic e-waste recycling workshops in china. *Ecotoxicology and Environmental safety*, 15 June, Volume 174, pp. 532-539.

Zhang, Z. & Adela, C., 2017. Principal components analysis in clinical studies. *Annals of translational medicine*, 28 June.

Supplement 1;

Rådata ekstraherte komponenter + kvalitetssikret identifiserte komponenter + pivot (forsøk 2)

Vedlegg 1; Oversikt ID støvprøver

Tabell 6; Viser hvor støvprøvene er tatt, samt hvor stort området er

Prøver		
Prøvenr.	Område	Areal [m ²]
18/2675	Stue	2,2
18/2676	Stue	4,5
18/2677	Stue	7
18/2678	Stue	1,2
18/2679	Stue	2,1
18/2680	Soverom	1
18/2681	Stue	2,2
18/2682	Stue	20
18/2683	Stue	27+3,4
18/2684	Stue	2,5
18/2685	Stue + soverom	5+5
18/2686	Stue + soverom	7+4
18/2687	Stue	5
18/2688	Stue + soverom	Ukjent
18/2689	Stue + trapp	2,5+1,3
18/2690	1+2	2,5+3
18/2691	Stue	4+3
18/2692	Stue	45
18/2693	Kjellerstue	30
18/2694	Stue	21
18/2695	Barnerom	6
18/2696	Stue	3
18/2697	Soverom	1,5
18/2698	Stue	7
18/2699	Stue	Ukjent
18/2700	Stue	8

Vedlegg 2, Alkanmiks

Tabell 7: Viser retensjonstiden i min for de ulike alkanene i alkanmiksen, C7-C13

Alkanmiks, C7-C13	
Antall C-atomer	RT [min]
12	8,54
13	10,96
14	13,36
15	15,62
16	17,74
17	19,74
18	21,62
19	23,41
20	25,11
21	26,72
22	28,27
23	29,75
24	31,17
25	32,54
26	33,86
27	35,12
28	37,53
29	38,66
30	39,84