

# **Bitumin vanhenemismekanismit ja elvytysprosessien arviointi FTIR-spektroskopiolla**

Miila Hautamäki

**Perustieteiden korkeakoulu**

Kandidaatintyö

Espoo 22. joulukuuta 2025

**Työn ohjaaja**

Jarno Schmidt, Peab Industri Oy

**Vastuopettaja**

Vanhempi yliopistonlehtori Jani-Petri  
Martikainen

Copyright © 2025 Miila Hautamäki

The document can be stored and made available to the public on the open internet pages of Aalto University.  
All other rights are reserved.

---

**Tekijä** Miila Hautamäki

---

**Työn nimi** Bitumin vanhenemismekanismit ja elvytysprosessien arviointi  
FTIR-spektroskopiolla

---

**Koulutusohjelma** Teknistieteellinen kandidaattiohjelma

---

**Pääaine** Teknillinen fysiikka

**Pääaineen koodi** SCI3028

---

**Työn ohjaaja** Jarno Schmidt, Peab Industri Oy

---

**Vastuopettaja** Vanhempi yliopistonlehtori Jani-Petri Martikainen

---

**Päivämäärä** 22. joulukuuta 2025

**Sivumäärä** 30+3

**Kieli** Suomi

---

### **Tiivistelmä**

Bitumi on asfalttipäällysteiden sideaine, jonka fysikaaliset ja kemialliset ominaisuudet heikkenevät ajan ja ympäristöolosuhteiden vaikutuksesta. Bitumin vanheneminen johtuu pääasiassa hapettumisesta, mikä lisää bitumin polaarisuutta ja vahvistaa asfalteenien välisiä vuorovaikutuksia. Elvyttämisen tavoitteena on palauttaa bitumin alkuperäisiä ominaisuuksia tasapainottamalla asfalteenien ja malteenien suhdetta ja vähentämällä hapettumistuotteita. Elvytetyn bitumin kemiallisia muutoksia voidaan havaita FTIR-spektroskopiolla, mutta sen tarkkuus on rippuvainen näytteen valmistelusta ja tulosten tulkinnan ongelma on absorptiospektripiikkien epäselvyys ja päällekkäisyys. ATR-FTIR-tekniikka parantaa mittaustarkkuutta, mutta rajoittuu näytteen pintakerroksen analysointiin. Näiden rajoitteiden vuoksi FTIR-spektroskopioiden rinnalle tarvitaan laskennallisia menetelmiä, jotta monimutkaisten ja suurten spektridatojen tulkinta on mahdollista ja luotettavampaa.

Monimuuttuja-analyysiä hyödynnetään FTIR-datan käsittelyyn ja bitumin kemiallisten muutosten luokitteluun. Koneoppimista ja data-analytiikkaa hyödyntävät esikäsitteily- ja erottelumenetelmät parantavat dataa ja löytävät näytteiden väliltä korrelaatioita. Tutkimukset ovat hyötyneet monimuuttuja-analyysin käytöstä, sillä spektridatan matemaattinen tarkastelu on parantanut tulosten laajuutta ja tarkkuutta. Yleisesti menetelmän käyttö elvytetyn bitumin tutkimisessa on vielä varhaisessa vaiheessa ja saatavilla olevat tutkimusaineistot ovat pieniä, analyysimenetelmien valintaperusteet epäyhtenäisiä ja menetelmien toistettavuus puutteellisesti raportoitua. Kirjallisuuskatsauksen perusteella monimuuttuja-analyysillä on potentiaalia selittää elvytetyn bitumin monimutkaisia absorptiospektrejä, mutta tämänhetkinen tutkimusaineisto ei kata riittävästi erilaisia elvytin- ja bitumilaatuja. Tutkimuksissa käytettyjen analyysimenetelmien välille tarvitaan yhtenäisempiä käytäntöjä ja mahdollisia epäoptimaalisia menetelmiä pitäisi pysyvä välttämään tehokkaammin.

---

**Avainsanat** bitumi, vanheneminen, elvyttäminen, FTIR, ATR-FTIR,  
monimuuttuja-analyysi

---

# Sisällys

|  |           |
|--|-----------|
| <b>Tiivistelmä</b>   | <b>3</b>  |
| <b>Sisällys</b>  | <b>4</b>  |
| <b>Symbolit ja lyhenteet</b>   | <b>5</b>  |
| <b>1 Johdanto</b>  | <b>6</b>  |
| <b>2 Teoreettinen tausta</b>   | <b>7</b>  |
| 2.1 Bitumin kemiallinen koostumus ja fysikaaliset ominaisuudet . . . . . | 7         |
| 2.2 Bitumin vanhentuminen . . . . .                                      | 7         |
| 2.3 Elvyttimet ja ominaisuuksien palauttaminen . . . . .                 | 9         |
| 2.4 IR-spektroskopia . . . . .   | 9         |
| 2.5 Fourier muunnettu absorptiospektri . . . . .                         | 11        |
| 2.5.1 FTIR-analyysin heikkoudet . . . . .                                | 13        |
| 2.5.2 ATR-FTIR . . . . .   | 13        |
| <b>3 Monimuuttuja-analyysi</b>   | <b>14</b> |
| 3.1 Esikäsittely . . . . .   | 14        |
| 3.1.1 Pääkomponenttianalyysi . . . . .                                   | 16        |
| 3.1.2 Faktorianalyysi . . . . .  | 16        |
| 3.1.3 Osittaisen pienimmän neliösumman regressio . . . . .               | 18        |
| 3.1.4 Luokittelumenetelmä . . . . .                                      | 18        |
| <b>4 Tutkimusongelma</b>   | <b>19</b> |
| <b>5 Tulokset</b>  | <b>20</b> |
| 5.1 Elvyttimien erottelu . . . . .                                       | 21        |
| 5.2 Elvyttämisen tehokkuus . . . . .                                     | 23        |
| 5.3 Kriittinen tarkastelu . . . . .                                      | 25        |
| <b>6 Johtopäätökset</b>  | <b>26</b> |
| <b>A Monimuuttuja-analyysin matemaattinen tarkastelu</b>                 | <b>31</b> |
| A.1 Pääkomponenttianalyysi . . . . .                                     | 31        |
| A.2 Faktorianalyysi . . . . .  | 31        |
| A.3 Osittaisen pienimmän neliösumman regressio . . . . .                 | 31        |
| A.4 LDA . . . . .  | 32        |

## Symbolit ja lyhenteet

### Symbolit

|               |   |
|---------------|---|
| $A$           | absorbanssi   |
| $I$           | näytteen läpäisseen säteilyn intensiteetti                                    |
| $I_0$         | alkuperäisen säteilyn intensiteetti   |
| $T$           | läpäisevyys   |
| $\varepsilon$ | molaarinen absorptiokerroin   |
| $b$           | optinen polunpituus / näytteen paksuus  |
| $c$           | konsentraatio   |
| $\lambda$     | aallonpituus  |
| $\bar{\nu}$   | aaltoluku [ $\text{cm}^{-1}$ ]  |
| PC            | pääkomponentti (Principal Component)  |
| LV            | latenttimuuttuja (Latent Variable)  |
| VIP           | muuttujan tärkeysindeksi PLSR-analyysissä (Variable Importance in Projection) |

### Lyhenteet

|      |   |
|------|---|
| FTIR | Fourier Transform Infrared Spectroscopy                                       |
| ATR  | Attenuated Total Reflection   |
| IR   | infrapunasäteily (Infrared Radiation)   |
| MVA  | monimuuttuja-analyysi (Multivariate Analysis)                                 |
| PCA  | pääkomponenttianalyysi (Principal Component Analysis)                         |
| FA   | faktorianalyysi (Factor Analysis)   |
| PLSR | osittaisen pienimmän neliösumman regressio (Partial Least Squares Regression) |
| LDA  | lineaarinen diskriminanttianalyysi (Linear Discriminant Analysis)             |
| SNV  | Standard Normal Variate   |
| SG   | Savitzky–Golay -suodatus  |
| UV   | ultraviolettisäteily (Ultraviolet Radiation)                                  |
| IRE  | Internal Reflection Element   |

# 1 Johdanto

Asfalttipäällysteiden kestävyys ja elinkaareen vaikuttaa muunmuassa niiden sisältämä sideaine, bitumi. Bitumi on monimutkainen seos orgaanisia yhdisteitä, joiden rakenne ja ominaisuudet muuttuvat ajan ja ympäristötekijöiden vaikutuksesta [1, 2]. Ulkoiset rasitukset, kuten lämpötilavaihtelut, UV-säteily, kosteus ja hapettuminen, johtavat bitumin vanhenemiseen [1]. Tämä ilmenee sen jäykistymisenä ja haurastumisena mikä lopulta heikentää päällysteen kestävyyttä. Viime vuosina kasvava tarve kierrättää asfalttimateriaaleja ja pidentää päällysteiden käyttöikää on lisännyt kiinnostusta bitumin elvyttämiseen [3]. Työn aihe kytkeytyy siten globaaliin ongelmaan ja tutkii, miten bitumin kemiallisia ominaisuuksia voidaan palauttaa tehokkaasti ja kestävästi, jotta asfaltin uusiokäyttö olisi mahdollisimman kannattavaa.

Bitumin vanhenemisen tutkimiseen liittyy kuitenkin vielä useita haasteita. Vanheneminen ei ole yksittäinen ilmiö, vaan seurausta useista samanaikaisista fysikaalisista ja kemiallisista prosesseista, joiden mekanismeja ei vielä täysin tunneta [2, 4]. Perinteiset laboratoriomenetelmät antavat tietoa yksittäisten funktionaalisten ryhmien muutoksista, mutta ne eivät pysty kuvaamaan monimutkaisia rakenteellisia vuorovaikutuksia bitumin eri komponenttien välillä [2]. Viime vuosina Fourier'n muunnokseen perustuva FTIR-spektroskopia on vakiintunut toimivaksi menetelmäksi bitumin kemiallisten muutosten havaitsemiseen. Menetelmällä voidaan tunnistaa hapettumisen ja elvytyksen vaikutuksia esimerkiksi karbonyyli- ja sulfoksidiryhmien perusteella [5]. Silti FTIR-analyysin tulkinta on haasteellista, sillä spektrien päällekkäiset piikit, taustakohina ja näytteen valmistelun vaihtelu voivat hämärtää tuloksia [6]. Näitä rajoitteita on pyritty ratkaisemaan hyödyntämällä kemometrisia eli monimuuttuja-analyysiin perustuvia menetelmiä, mutta niiden soveltaminen bitumitutkimuksessa on vasta alkuvaiheessa [7]. Kirjallisuudessa on esitetty lupaavia tuloksia, mutta tutkimuksia on tehty pienillä aineistoilla ja menetelmien valintaperusteet vaihtelevat, mikä rajoittaa tulosten vertailtavuutta ja toistettavuutta [8].

Tässä kirjallisuuskatsauksessa tarkastellaan bitumin vanhenemismekanismeja ja arvioidaan elvyttimien toimintaa FTIR-spektroskopian avulla sekä selvitetään, miten kemometria voi parantaa analyysin tulkintaa. Työssä arvioidaan FTIR:n vahvuuksia ja heikkouksia bitumin tutkimuksessa sekä vertaillaan monimuuttuja-analyysin menetelmiä, kuten pääkomponenttianalyysiä (PCA), faktorianalyysiä (FA), osittaisen pienimmän neliösumman regressiota (PLSR) ja lineaarista diskriminanttianalyysiä (LDA). Erityistä huomiota kiinnitetään siihen, miten näitä menetelmiä on tähän mennessä käytetty elvytetyn bitumin tutkimuksessa ja millaisia ongelmia niissä on. Työssä keskitytään tunnistamaan kemometrian hyödyntämisen nykyiset puutteet ja selvittämään, miten FTIR- ja MVA-menetelmien yhdistäminen voisi tulevaisuudessa tuottaa luotettavampaa tietoa bitumin elvytyksen tehokkuudesta.

## 2 Teoreettinen tausta

### 2.1 Bitumin kemiallinen koostumus ja fysikaaliset ominaisuudet

Bitumia voidaan tuottaa teollisuuden raakaöljyn tislauksjäännöksestä. Tislauksessa raakaöljy jaotellaan fraktioiksi kiehumispisteiden perusteella ja tiheydeltään suurin fraktio on bitumin valmistukseen käytettävä hiilivetyseos [9]. Jalostusprosessilla tästä jäännöksestä saadaan mustaa bitumia, jota käytetään asfalttimassan sideaineena. Bitumin sovelluskohteita ovat myös vesitiiviit eristeet ja -materiaalit, mutta niiden osuus tuotannossa on pientä asfalttipäällysteiden rinnalla [1].

Kemialliselta koostumukseltaan bitumi on monimutkainen yhdiste, joka sisältää molekyyliä melko epäpolaarisista vahamaisista hiilivedyistä, aina erittäin polaarisiin ja helposti polarisoituihin aromaatteihin. Painoprosentuaalisesti se koostuu pääosin hiilestä (80 % – 88 %), vedystä (8 % – 12 %), hapestä (0 – 2 %) sekä typestä (0 – 2 %). Kompleksinen koostumus sisältää polaarisia tiiviitä aromaattisia rengasrakenteita sekä yksittäisiä heteroatomeja, kuten happea, typpeä ja reaktiivista rikkiä. Bitumi voidaan siis jakaa molekyylipainon ja kemiallisen koostumuksen mukaan neljään komponenttiin: asfalteenihin, hartseihin, aromaatteihin ja saturaatteihin. Hartseista, aromaateista ja saturaateista voidaan käyttää yhteisnimitystä malteenit. Nämä kemiallisesti toisistaan eroavat rakenteet käyttäytyvät ryhminä eri tavoin ja jokainen komponentti vaikuttaa osaltaan bitumin fysikaalisiin ominaisuuksiin. [2]

Polaariset asfalteenit sisältävät usein happea, typpeä ja rikkiä. Niiden klusterimainen ryhmittymä pysyy koossa vetysidosten, dipoli–dipoli-vuorovaikutusten sekä  $\pi - \pi$ -pinoutumisen avulla [2]. Näiden heikkojen vuorovaikutusten kautta syntyy nanoskaalan kokoisia rakenteita, jotka muistuttavat polymeerimolekyylien rakenteita ja antavat bitumille sen tyypillisen viskoelastisen käyttäytymisen. Korkeissa lämpötiloissa (100 – 200°C) asfalteenien ryhmittymien heikot sidokset voivat katketa, mikä pienetää ryhmittymien kokoa ja siten lisää niiden liikkuvuutta [9]. Tämä näkyy bitumin viskositeetin alenemisena ja nestemäisempänä olemuksena. Matalissa lämpötiloissa sidokset puolestaan vahvistuvat ja bitumi jäykistyy. Lämpötilasta riippuvat fyysikaaliset ominaisuudet ovat suurilta osin palautuvia, mikä antaa bitumille sen muokattavuuden.

Asfalteenien lisäksi bitumin fysikaaliseen olemukseen vaikuttavat hartsit, aromaattit ja saturaatit, jotka muodostavat asfalteenien ympärille malteenifaasin. Hartsit ovat pääasiassa polaarisia amorfisia yhdisteitä. Ne ovat reaktiivisia ja toimivat bitumissa staabiloivana komponenttina asfalteenien välillä. Saturaatit ovat heikosti polaarisia suoraketjuisia tai haaroittuneita alifaattisia hiilivetyjä, mutta niiden suuri pitoisuus voi aiheuttaa asfalteenien saostumisen. Aromaattit ovat puolestaan rengasrakenteisia hiilivetyjä. Nämä toimivat bitumin pehmentävinä komponentteina. [2]

### 2.2 Bitumin vanhentuminen

Bitumin vanhentuminen muuttaa sen kemiallista koostumusta ja vaikuttaa suoraan asfaltin fysikaalisiin ominaisuuksiin. Keskeisin vanhenemismekanismi on hapettumi-

nen, joka muokkaa bitumin kemiallista rakennetta ja tekee siitä vähitellen jäykempää, epäelastisempaa ja hauraampaa [1]. Kun bitumi on hapettumisen vuoksi jo kovettunut, muut ympäristötekijät, kuten UV-säteily, kosteus, lämpötilanvaihtelut ja mekaaninen rasitus voimistavat entisestään sen fysikaalista vanhenemista ja nopeuttavat päällysteen heikkenemistä. Jo bitumin valmistusvaiheessa voi tapahtua vanhenemista, kun lämpötila pidetään korkeana. Kevyitä komponentteja haihtuu, mikä vaikuttaa asfalteenien ja malteeneiden väliseen tasapainoon [4].

Bitumi on orgaaninen yhdiste, joka hapettuu altistuessaan ilman happiyhdisteille. Hapettuminen johtaa uusien funktionaalisten ryhmien muodostumiseen ja sitä kautta muutoksiin bitumin kemiallisessa rakenteessa. Hapettumista pidetään nykyisin tärkeimpänä vanhentumismekanismina ja siitä aiheutuvat muutokset kemiallisessa rakenteessa ovat pysyviä. Yleisimmät hapettumistuotteet ovat polaarisia ketoneita ja rikkioksideja. Hapettumistuotteiden konsentraatio voi nousta jopa yhteen mooliin litrassa, mikä vastaa tilannetta, jossa keskimäärin yhtä noin 1000 Daltonin kokoista bitumimolekyyliä kohden muodostuu yksi uusi funktionaalinen ryhmä. Hapettuminen ei jakaudu tasaisesti kaikkiin bitumin fraktioihin, joten vähäpolaariset komponentit, kuten saturaatit ja osa aromaateista eivät juurikaan hapetu. Polaarisiin osiin eli asfalteeneihin voi sen sijaan muodostua jopa yli yksi uusi funktionaalinen ryhmä yhtä molekyyliä kohden. [2]

Hapettumistuotteiden painoprosentuaalinen osuus molekyylistä ei kuitenkaan ole suoraan bitumin ikääntymistä ja kovettumista määräävä tekijä. Ratkaisevaa on se, miten hapettumisesta johtuvat uudet polaariset ryhmät vaikuttavat komponenttien välisiin sidoksiin. Hapettumistuotteet pystyvät muodostamaan vetysidoksia, jolloin heikot vuorovaikutukset voimistuvat [2]. Näiden heikkojen, mutta lukumäärältään suurten sidosten lisääntyminen johtaa molekyylien ryhmittymiseen ja kovettavien asfalteeniklustereiden kasvuun. Bitumilaaduilla, joiden rakenne perustuu valmiiksi suuriin määriin heikkoja vuorovaikutuksia, polarisaation kasvu vauhdittaa kovettumista [2]. Sen sijaan rakenteissa, joissa molekyylit ovat jo valmiiksi tiiviisti sitoutuneita ja joissa vahvat vuorovaikutukset dominoivat, hapettumisen vaikutus jää vähäisemmäksi. Näin ollen hapettumisnopeus ja sen vaikutus riippuvat paitsi uusien funktionaalisten ryhmien määrästä, myös siitä, kuinka herkkä bitumin alkuperäinen rakenne on lisääntyneelle polaarisuudelle.

Ilman hapen aiheuttama hapettuminen korostuu auringon UV-säteilyn vaikutuksesta. UV-säteilyn ja hapen yhteisvaikutuksesta bitumin molekyylit käyvät läpi foto-oksidaation eli säteilyn aiheuttaman hapettumisen [10]. UV-säteily virittää bitumin sidoksia ja helpottaa hapen liittymistä molekyyliin. Tämä voi johtaa vetyatomien irtoamiseen hiilestä sekä C=C-, C-S- ja osin C-O-sidosten katkeamiseen. Näin syntyvät reaktiiviset kohdat sitovat happea ja muodostavat polaarisia ja jäykkiä karbonyyli- sekä sulfoksidiryhmille ominaisia C=O- ja S=O-sidoksia [10]. Ne lisäävät hapellisten funktionaalisten ryhmien suhteellista määrää. Kevyemmät fraktiot vähenevät ja molekyylit reagoivat muodostaen raskaampia ja polaarisempia yhdisteitä [4]. Tämä kasvattaa bitumin viskositeettiä, kovettaen sitä entisestään.

## 2.3 Elvyttimet ja ominaisuuksien palauttaminen

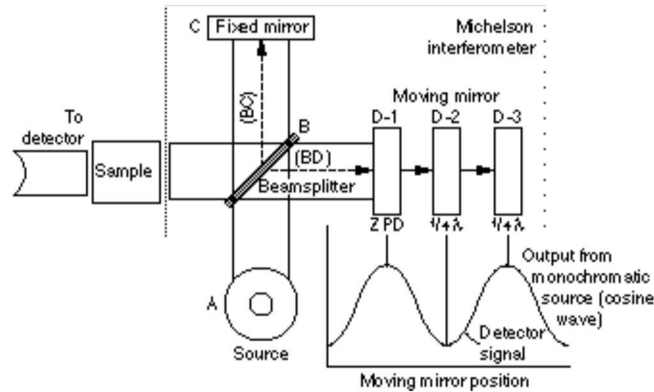
Bitumin elvyttimet ovat aineita, joiden avulla pyritään palauttamaan vanhentuneen bitumin alkuperäisiä ominaisuuksia. Tarkoituksena on vähentää bitumin jäykkyyttä ja viskositeettiä sekä lisätä sen muokkautuvuutta. Vaikka bitumin ikääntymiseen liittyy monia samanaikaisesti tapahtuvia ja osittain toisiinsa kytkeytyviä mekanismeja, elvyttäminen aloitetaan pienentämällä asfaltenien suhdetta malteeneihin verrattuna [11].

Elvyttäminen voidaan toteuttaa pehmentävillä, öljypohjaisilla aineilla. Ne eivät muovaa bitumin alkuperäistä kemiallista rakennetta, vaan lisäävät pehmentävien malteenifraktioiden osuutta [3, 12]. Tämän kaltaisia elvyttimiä kutsutaan pehmentimiksi (fluxing agent). Vaihtoehtoisesti todellisilla elvyttimillä (real rejuvenator) voidaan palauttaa bitumin vanhentunut kemiallinen rakenne. Ne vaikuttavat suoraan bitumin kemiallisiin ominaisuuksiin hajottamalla suuria asfalteeniklustereita ja palauttamalla haihtuneita komponentteja [3, 12]. Pienennetyt asfalteeniryhmittymät edistävät asfalteenien sekoittumista tasaisesti malteenien kanssa, jolloin ideaali tasapaino fraktioiden välillä löytyy. Koostumukseltaan todellisten elvyttimien tulisi siis sisältää pehmittäviä ja asfalteenien pinoutumista estäviä aromaatteja [11]. Saturaattien pitoisuus pitäisi pysyä alhaisena, jotta asfalteenit eivät saostu [11]. Elvyttimien tehokkuuteen vaikuttaa myös niiden tasainen sekoittuminen bitumiin. Diffuusion tehokkuutta säätelee pääasiassa malteenien viskositeetti, joten sekoittumista voidaan tehostaa nostamalla lämpötilaa [3]. Tutkimuskirjallisuudessa käytetään usein pehmittimistä ja todellisista elvyttimistä yhteisnimitystä elvytin, ottamatta kantaa sen toimintaperiaatteeseen.

## 2.4 IR-spektroskopia

Bitumi koostuu suurista ja monimutkaisista orgaanisista molekyyleistä, jotka eroavat toisistaan sekä rakenteeltaan, että kemialliselta koostumukseltaan. Useat molekyylit sisältävät monia funktionaalisia ryhmiä, mikä tekee niiden yksilöllisestä tunnistamisesta ja erottelusta hyvin haastavaa [2]. Funktionaalisia ryhmiä voidaan yrittää tunnistaa infrapunaspektroskopiolla. Infrapunasäteily (IR-säteily) kattaa sähkömagneettisen spektrin alueen, jonka aaltoluvut  $\bar{\nu}$  ovat noin  $13000 - 10 \text{ cm}^{-1}$  ja aallonpituus  $\lambda$  välillä  $0,78 - 1000 \mu\text{m}$  [13]. Käytännössä tämä on punaisen valon ja mikroaaltojen väliin jäävä alue.

IR-spektroskopioteknologioiden toimintaperiaate mukailee Michelssonin interferometriä, joka esitetään kuvassa 1. Siinä IR-säteily jaetaan säteenjakajan avulla kahteen kulkureittiin. Toisen reitin pituus pidetään vakiona, kun taas toista muutetaan liikkuvalla peilillä. Kun säteet yhdistetään takaisin säteenjakajassa, niiden välille muodostuu vaihe-ero, joka riippuu kulkureittien erotuksesta. Tämän seurauksena osa säteilystä ohjautuu takaisin lähteelle ja osa detektorille, kuljettuaan näytteen läpi. Vaihe-eron ollessa nolla, kiinteä ja liikkuva peili ovat yhtä kaukana säteenjakajasta ja säteet vahvistavat toisensa konstruktiivisesti. Tällöin kaikki säteily kulkeutuu detektorille. Kun kulkureittien ero on puoli aallonpituutta, tapahtuu destruktiivinen interferenssi ja suurin osa säteilystä ohjautuu takaisin lähteeseen. [13]



Kuva 1: Yksinkertaistettu kaaviokuva Michelsonin interferometrissä. Kuva: [13]

IR-spektroskopia perustuu molekyylien kykyyn absorboida infrapunasäteilyä ja siirtyä korkeammille värähtelytasolle. Kun IR-säteily osuu näytteeseen, sen molekyylien atomit alkavat värähdellä toisiinsa nähden. Nämä värähtelyt voivat ilmetä sidosten venymisenä, taipumisena tai kiertymisena [14]. Näytteen jokaisella funktionaalisella ryhmällä on joukko sille ominaisia värähtelytiloja, joiden lukumäärä ja luonne riippuvat atomien massoista, lukumäärästä ja sidostyypeistä [14].

Kaikki molekyylit eivät kuitenkaan absorboi IR-säteilyä. Esimerkiksi symmetriset kahden atomin molekyylit eivät ole IR-aktiivisia, koska niiden värähtely ei muuta molekyylin dipolimomenttia [15]. Sen sijaan epäsymmetrisissä molekyyleissä värähtely aiheuttaa dipolimomentin muutoksen, jolloin IR-säteilyn absorptio on mahdollista [15]. Monimutkaisemmissa molekyyleissä useat sidokset ja funktionaaliset ryhmät mahdollistavat useita erilaisia värähtelytiloja, joista kukin vastaa tiettyä ominaistajuutta [14]. Kun näytteeseen kohdistuvan IR-säteilyn taajuus vastaa jonkin näistä värähtelytaajuuksista, molekyyli absorboi säteilyenergian.

Näytemolekyylien absorboiman energian perusteella saadaan näytteen koostumuksesta tietoa. Detektori mittaa läpäisevyyttä  $T$ , joka kertoo kuinka suuri osa IR-säteistä kulkee näytteen läpi [14]. Läpäisevyys lasketaan näytteen läpäisseen säteilyn tehon  $I$  ja alkuperäisen säteilyn tehon  $I_0$  suhteena kaavan 1 mukaisesti

$$T = \frac{I}{I_0} \times 100\%. \quad (1)$$

Absorbanssi  $A$  puolestaan kuvaa näytteeseen absorboituneen säteilyn määrää. Lambert-Beerin lain (2) mukaisesti se on lineaarisesti verrannollinen näytteen konsentraatioon  $c$ , valon kulkemaan matkaan  $b$  ja näytteen molaariseen absorptio-koefisiettiin  $\epsilon$ .

$$A = \epsilon bc \quad (2)$$

Absorbanssi saadaan laskettua läpäisevyyden logaritmisena käänteisarvona kaavan (3) mukaisesti [14]

$$A = \log_{10}\left(\frac{I_0}{I}\right). \quad (3)$$

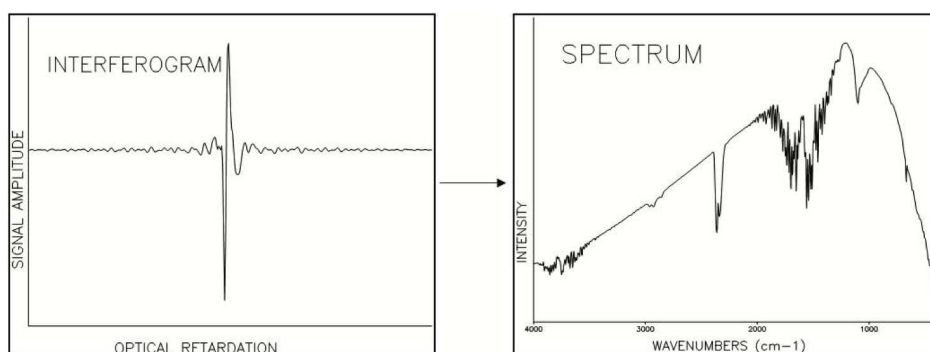
Kun IR-säteily kulkee näytteen läpi, niiden taajuuksien amplitudit vaimenevat, joita vastaavilla värähtelytaajuuksilla näyte absorboi säteilyä. Absorptiospektriin muodostuu siis ylöspäin nousevia piikkejä jotka erottuvat toisistaan logaritmisesta riippuvuudesta ansiosta [13]. Läpäisevyysspektrissä piikit näkyvät alaspäin suuntautuneina [13]. Absorptiospektrin etuna on sen lineaarinen riippuvuus konsentraation suhteen, jonka vuoksi se on yleisesti käytetty spektrien esitysmuoto.

## 2.5 Fourier muunnettu absorptiospektri

IR-mittauksessa säteily kohdistetaan näytteeseen ja liikkuvan peilin kulkiessa vakionopeudella, detektorille saapuvan säteilyn intensiteetti vaihtelee ajan funktiona [13, 14]. Kukin säteilyn taajuus tuottaa detektorille sinimuotoisen intensiteettivaihtelun, jonka amplitudi riippuu näytteen absorptiosta kyseisellä taajuudella [13, 14]. Kuvassa 1 havainnollistetaan, kuinka peilin ajasta riippuva sijainti vastaa mitattavaa interferenssisignaalia. Mikäli näyte absorboi jollakin mitatuista taajuuksista, kyseisen aallon amplitudi pienenee suhteessa näytteessä olevaan aineen määrään.

Koska mittauksessa käytetty IR-säteily on polykromaattista, detektorille mitattava signaali muodostuu useiden eri taajuuksia vastaavien sinimuotoisten komponenttien summasta [13]. Näiden komponenttien yhteisvaikutuksena syntyy monimutkainen interferogrammi, joka sisältää informaation koko alueesta, jolle detektori on herkkä [14].

Michelsonin interferometriin perustuvaa mittaustekniikkaa hyödynnetään Fourier Transform Infrared-spektroskopiassa (FTIR). FTIR-menetelmän etuna on kyky mitata koko IR-spektri samanaikaisesti, ilman yksittäisten taajuuksien erillistä summaamista [16]. Fourier-muunnoksen avulla interferogrammi, joka on esitetty ajan tai optisen polunpituuseron funktiona, muunnetaan taajuusavaruuteen [14]. Tällöin saadaan varsinainen infrapunaspektri eli näytteen absorptio tai läpäisevyys aaltoluvun funktiona. Fourier-muunnos muokkaa interferogrammin visuaalisesti helpommin tulkittavaan spektrimuotoon, kuten kuvassa 2 on esitetty.



Kuva 2: Fourier-muunnoksen tuottama visuaalinen muutos, jossa vasemmalla on yhden säteen interferogrammi ja oikealla Fourier muunnettu spektri. Kuva: [17]

Absorptiospektrissä esiintyvän piikin paikan avulla voidaan tunnistaa funktionaalinen ryhmä. Piikin pinta-ala kuvastaa kyseisen ryhmän konsentraatiota [5].

Taustaspektri mitataan referenssiksi, koska näytteen liuotin tai ympäristön kaasut voivat sisältää yhdisteitä, jotka aiheuttavat ei-toivottuja absorptiosignaaleja [5]. Taustaspektri vähennetään näytteen mitatusta spektristä, jotta saadaan esiin näytteen oma absorptio. Bitumin yleisimpien funktionaalisten ryhmien suuntaa-antavat spektriipiikit on taulukoituna taulukossa 1.

Taulukko 1: FTIR-spektrien yleisimmät aaltoluvut ja niitä vastaavat kemialliset sidokset sekä vaikutus spektrin muotoon [5].

| Aaltoluku<br>( $\text{cm}^{-1}$ ) | Tyypillinen sidos / funktio-<br>ryhmä | Esiintyminen spektrissä          |
|-----------------------------------|---------------------------------------|----------------------------------|
| 2955                              | C-H                                   | Alifaattinen $\text{CH}_3$       |
| 3005                              | =C-H                                  | Aromaattinen                     |
| 1735-1750                         | C=O (esteri)                          | -                                |
| 1720-1740                         | C=O (Aldehydi)                        | -                                |
| 1715-1675                         | C=O (ketoni, karbonyyli)              | Hapettumistuote                  |
| 1600-1580                         | Aromaattinen C=C                      | Aromaattiset rakenteet           |
| 1550-1640                         | N-H (Amidi)                           | -                                |
| 1460-1450                         | $\text{CH}_2/\text{CH}_3$             | Alifaattinen                     |
| 1375                              | $\text{CH}_3$                         | Alifaattinen                     |
| 1300-1000                         | C-O (esteri, eetteri)                 | Vähenee vanhentuneessa bitumissa |
| 1031                              | S=O (sulfoksidi)                      | Hapettumistuote                  |
| 727                               | - $\text{CH}_2$                       | -                                |

### 2.5.1 FTIR-analyysin heikkoudet

FTIR-spektroskopian transmissiotavassa (transmission mode) valonsäde kulkee koko näytteen läpi ennen detektoriin osumista. Kriteeri valonläpäisevyydestä kuitenkin rajoittaa näytteen olomuotoa. Tällaisessa tutkimustavassa bituminäytteen pitää olla riittävän ohut kalvo, mikä voidaan saavuttaa levittämällä kuuma bitumipisara infarpunasäteen läpäisevälle kirkkaalle alustalle ja mahdollisesti hyödyntää liuotinta levittämisessä. Kaasumaisen näytteen käyttö on myös mahdollista, mutta se ei ole yleistä bitumin tutkimisessa korkean kiehumispisteen vuoksi. [18]

Transmissiotapaan sopivan näytteen saamiseksi voidaan myös muodostaa liuos, joka suljetaan kirkkaiden ikkunoiden väliin säiliöön ja tiivistetään tiivisterenkaalla. Säteilystä saadaan kulkemaan tehokkaasti ikkunoista näytteen läpi, mutta liuos voi vaikuttaa menetelmän tarkkuuteen, mikäli se muuttaa bitumin kemiallista rakennetta. [18]

Tiivisterenkaan paksuus vastaa säteiden kulkeman matkan pituutta  $b$  ja Lambert-Beerin lain (2) mukaisesti sillä on suora riippuvuus absorbanssiin. Erityisesti liuottimien kanssa, joiden molaarinen absorptiivisuus  $\epsilon$  on suuri, tämä muodostuu merkittäväksi virhelähteeksi paksuilla näytekerroksilla [5]. Tiivisterenkaan paksuus pitäisi siis minimoida, jotta kyllästymistä ei tapahtuisi. Tämä aiheuttaa kuitenkin ongelmia tiivisteiden paksuuden vakioimisessa ja siten työn toistettavuudessa. Hyvin ohut näyte lisää myös ikkunoita vasten olevien molekyylien suhteellista osuutta koko näytteen kokoon verrattuna [18]. Nämä molekyylit voivat vuorovaikuttaa ikkunamateriaalin kanssa vääristäen spektriä. Mikäli näyte sisältää vettä, sen absorptiointensiteetti ei muutu lineaarisesti pitoisuuden kasvaessa, mikä täytyy huomioida tuloksia tarkasteltaessa [18]. Myös ilmakuplien mahdollisuus nestemäisessä näytteessä täytyy ottaa huomioon [18].

### 2.5.2 ATR-FTIR

Transmissiotavan ongelmia on ratkottu kokonaisheijastumista hyödyntävällä FTIR menetelmällä, ATR-FTIR (The Attenuated Total Reflection). ATR-FTIR perustuu IR-säteiden kokonaisheijastumiseen, kun näyte on kosketuksissa heijastavan elementin IRE (Internal reflection element) kanssa. Bitumin testauksessa yleisimmin käytetyt prismamateriaalit ovat ZnSe ja timantti [5]. IRE kokonaisheijastaa valon elementin rajapinnasta sisäänpäin, mikäli heijastuneen säteen taitekulma on pienempi, kuin tulokulma [5]. Rajapinnassa tapahtuva kokonaisheijastunut säde voidaan ohjata IR detektoriin.

Bituminäytteen ja IRE-prisman välille on tärkeä saada riittävän tiivis kosketuspinta, jotta saadaan luotettava FTIR-mittaus. Käytännössä bitumi voidaan levittää suoraan prisman pinnalle esimerkiksi lastalla, mutta näytettä voidaan lämmittää kevyesti levityksen helpottamiseksi [5]. Lämmitys voi olla hyödyllistä varmistettaessa tasainen ja ilmakuplaton kontakti prisman pintaan. On kuitenkin huomioitava, että lämmin näyte saattaa aiheuttaa mittaustuloksiin epävarmuutta ja lisätä kohinaa spektriin [5].

ATR-FTIR näytteen olomuoto ei ole niin rajoitettu, kuin transmissiotavassa. On huomioitava kuitenkin, että ATR-FTIR:ssä IR-säde tunkeutuu ainoastaan näytteen

pintakerrokseen. Bituminäytteen pintakerros ei välttämättä edusta koko näytettä. Säteiden tyypillinen tunkeutumissyvyys on vain  $0,2 - 0,5 \mu\text{m}$  [5]. Useita eri funktionaalisia ryhmiä sisältävästä bitumista pelkkä pintakerros ei välttämättä anna kattavaa ja luotettavaa spektriä. Toiseksi, kuten transmissotavan analyysissä, myös ATR-FTIR:n käytössä haasteeksi muodostuu veden osuus näytteessä ja sen epälineaarinen absorptiokäyttäytyminen [5].

### 3 Monimuuttuja-analyysi

Monimuuttuja-analyysi (MVA) hyödyntää data-analytiikan ja koneoppimisen menetelmiä bitumipohjaisten näytteiden luokitteluun ja erotteluun niiden kemiallisten ominaisuuksien perusteella. Menetelmä tunnistaa FTIR-spektrin aaltolukujen välisiä riippuvuuksia, huomioimalla muuttujia, kuten vanhentumisasteen ja elvytintyyppin [19]. FTIR-spektroskopia tuottaa jokaisesta aaltoluvusta yhden mittauspisteen, ja siten erittäin suuren määrän muuttujia [20]. Kun näytekohtaisia muuttujia on tuhansia, datan käsittely sellaisenaan on haastavaa. Tämän vuoksi data tyypillisesti esikäsitellään ennen analyysiä. Esikäsitelyn tavoitteena on poistaa poikkeamia, käsitellä voimakkaita huippuja ja vähentää taustakohinaa [19, 20]. Koska esikäsitely säilyttää datan alkuperäiset dimensiot, muuttujamäärää on tarpeen vähentää projisoimalla aineisto uuteen koordinaatistoon [19].

Dimensioiden pienentämiseen käytetään erilaisia algoritmeja, kuten faktorianaalyysiä (FA), pääkomponenttianaalyysiä (PCA) ja osittaisen pienimmän neliösumman regressiota (PLSR) [21]. Kaikki ulottuvuuksien vähennysmenetelmät pyrkivät luomaan uusia piirteitä, kuten latentteja muuttujia (LV) tai pääkomponentteja (PC) yksinkertaistaakseen monimutkaista spektridataa säilyttäen samalla suurimman osan varianssista [21]. Vaikka dimensioiden vähentäminen helpottaa ryhmien erottamista, erot luokkien välillä pyritään maksimoimaan. Bitumialalla lineaarinen diskriminanttianaalyysi (LDA) on yksi käytetyimmistä erottelumenetelmistä ja tässä työssä keskitytään vain siihen [19].

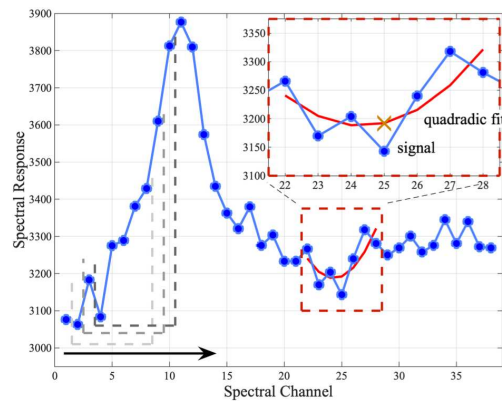
#### 3.1 Esikäsitely

Esikäsitelyn tavoite on parantaa spektriaineiston laatua ja helpottaa erojen havaitsemista analysoitavien näytteiden välillä. Raakaspektrit voivat sisältää esimerkiksi kohinaa, baseline-vaihtelua, intensiteettieroja ja häiriöitä, jotka eivät liity varsinaiseen kemialliseen koostumukseen, mutta voivat heikentää myöhempien analyysien tarkkuutta [22]. Epäolennaisen vaihtelun poistaminen vahvistaa spektrien informatiivisia piirteitä, jotta mallinnusmenetelmät voivat erottaa näytteet luotettavammin. Erilaisia esikäsitelymenetelmiä on lukuisia, mutta elvytetyn bitumin tutkimisessa keskeisimmät ovat SNV (Standard normal variate), SG (Savitzky-Golay) sekä ensimmäisen ja toisen asteen derivaattoihin perustuvat käsittelyt [22].

SNV normalisoi spektrin siten, että jokaisesta spektristä vähennetään sen oma keskiarvo ja jaetaan sen keskihajonnalla. Käsittelyn jälkeen jokaisella spektrillä on siis keskiarvo 0 ja keskihajonta 1. Tämän tarkoituksena on tehdä kaikki spektrit

vertailukelpoisiksi intensiteettien tai absorbanssien suhteen. Mentelmän avulla pystytään korjaamaan spektridatasta optisen polunpituuden tai valon sironnan vaihteluista aiheutuvia epätarkkuuksia. SVN voi olla hyödyllinen kompensoimaan esimerkiksi näytteen pinnan karheudesta aiheutuvia muutoksia tai näytteiden paksuusvaihteluista johtuvia absorbanssieroja. [22]

SG-suodatinta käytetään diskreetin spektrin silottamiseen. Se on erityisen suosittu suodatin spektroskopiassa, sillä se säilyttää spektrihiippujen korkeuden ja leveyden hyvin, mutta poistaa tehokkaasti taustakohinaa ja epäoleennaisia vaihteluita [22]. Suodatin määritellään käyttämällä kiinteälevyistä ikkunaa  $w$ , jonka sisällä lasketaan ja sovitetaan polynomiasteen  $o$  polynomi. Ikkuna siirtyy yksi datapiste kerrallaan koko signaalin yli. Suodatettu arvo ikkunan keskellä saadaan sovitetun polynomin arvosta kyseisessä keskikohdassa [23]. Tekniikka on havainnollistettu kuvassa 3, jossa spektridatasta on valittu seitsemän datapisteen levyinen ikkuna. Ikkunan sisälle jääviin pisteisiin on sovitettu polynomifunktio, joka antaa uuden tasoitetun signaalin. Käytännössä  $w$  valitaan parittomaksi, jotta ikkunalla on yksiselitteinen keskikohta ja ikkunan leveys määrittää, kuinka paljon dataa silotetaan. Pieni ikkuna säilyttää huiput ja yksityiskohdat, mutta sallii enemmän kohinaa [23]. Leveämpi ikkuna voi puolestaan vaimentaa huippuja ja pienempiä piirteitä [23]. Leveyden valinta riippuu siis analyysin tavoitteesta, halutaanko keskittyä huippujen erotteluun vai taustan tasoittamiseen. Suodattimen minimivaatimus on kuitenkin, että ikkunan leveys on vähintään yhtä suuri, kuin polynomin aste lisättynä yhdellä [23].



Kuva 3: SG-suodatin, johon on valittu seitsemän datapisteen levyinen ikkuna. Alikuvassa on esitetty polynomin sovitus ikkunan pisteisiin. Kuva: [23]

Suodatinikkunassa voidaan arvioida polynomitasoituksen lisäksi signaalin derivaattaa. Sovitetusta polynomista laskettu derivaatta korostaa kaikkia kohtia, joissa signaalin käyttäytyminen muuttuu nopeasti. Ensimmäisen asteen derivaatan arvon muuttuminen positiivisesta negatiiviseen kertoo piikistä, jonka huippu löytyy derivaatan nollakohdasta [23]. Toinen derivaatta erottelee jyrkästi muuttuvat huiput negatiivisina minimeinä. Loivasti tai lineaarisesti muuttuva taustakohina suodattuu pois, koska derivaatan muutos on pientä. Derivaatan avulla on mahdollista löytää osittain päällekkäisten piikkien erillisten huippujen nollakohdat [23]. Derivaattaa

kannattaa hyödyntää esimerkiksi juuri SG-suodattimen kanssa, sillä ilman polynomisovituksen tuomaa tasoitusta, derivaatta korostaa kaikkea kohinaa ja tulos on epäselvä [23].

### 3.1.1 Pääkomponenttianalyysi

Pääkomponenttianalyysi (PCA) on lineaarinen dimensioiden pienennysmenetelmä, joka muuntaa datajoukon uuteen koordinaatistoon. Uudet koordinaatit, eli pääkomponentit (PC), ovat lineaarisia yhdistelmiä alkuperäisistä muuttujista. Pääkomponentit järjestetään siten, että ensimmäinen pääkomponentti kertoo suurimman osan datan varianssista. Toinen pääkomponentti selittää jäljelle jääneen datan vaihtelusta suurimman osan ja niin edelleen. FTIR-spektroskopian yhteydessä PCA auttaa tiivistämään tuhansia mitattuja aaltolukuja muutamaksi pääkomponentiksi, jolloin datan käsittely ja visualisointi helpottuu, erityisesti käytettäessä kahta ensimmäistä pääkomponenttia. [24]

Spektridatasta muodostetaan matriisi  $X(n \times p) \in \mathbb{R}^{n \times p}$ , missä  $n$  on FTIR-spektrimittausten lukumäärä ja  $p$  mitattujen aallonlukujen määrä. Jotta muuttujat olisivat vertailukelpoisia keskenään, data standardoidaan vähentämällä keskiarvo  $\mu$  ja skaalaamalla se keskihajonnan  $\sigma$  suhteen. Tällöin saadaan kaavan (4) mukainen standardoitu matriisi  $X_{std(n \times p)}$  [24]

$$X_{std(n \times p)} = \frac{X_{(n \times p)} - \mu}{\sigma}. \quad (4)$$

PCA perustuu standardoidun datan kovarianssimatriisiin  $\Sigma$ , joka määritellään yhtälössä (5). Kovarianssimatriisi kuvaa eri aallonlukujen välistä lineaarista riippuvuutta ja jokainen kovarianssimatriisin alkio ilmaisee, kuinka datan pisteparit muuttuvat yhdessä [24].

$$\Sigma_{(p \times p)} = \frac{1}{n-1} X_{std}^T X_{std} \quad (5)$$

Kovarianssimatriisin ominaisarvohajotelma tuottaa pääkomponenttien suunnat sekä niitä vastaavat ominaisarvot, jotka kuvaavat komponenttien selittämää varianssia. Varsinaisen pääkomponenttimatriisin laskemiseksi standardoitu  $X_{std}$  projisoidaan yhtälön (6) mukaisesti ominaisvektoreille  $V_{(p \times p)}$  [22]. Ominaisvektori sisältää kovarianssimatriisin ominaisvektorit suuruusjärjestyksessä ja edustaa suurimman vaihtelun suuntia. Jokainen  $Z_{(n \times p)}$ -matriisin sarake muodostaa yhden pääkomponentin, jotka järjestetään niiden selittämän varianssin mukaan (Liite A.1) [22].

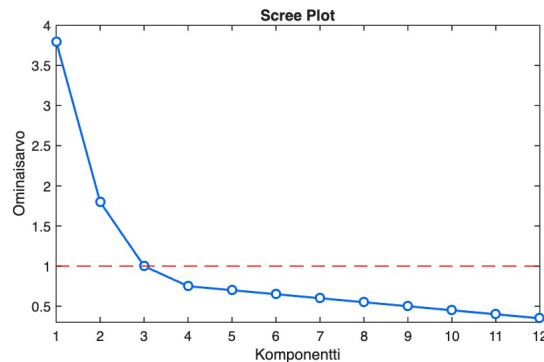
$$Z_{(n \times p)} = X_{(n \times p)}^{std} V_{(p \times p)} \quad (6)$$

### 3.1.2 Faktoriantalyysi

Faktoriantalyysiä (FA) käytetään aineiston rakenteen ymmärtämiseen tunnistamalla muuttujien välisiä yhteyksiä ja kuvaamalla havaittua vaihtelua pienemmällä määrällä uusia muuttujia, joita kutsutaan latenttimuuttujiksi (latent variables, LV) [21]. Toisin kuin PCA, joka keskittyy kuvaamaan mahdollisimman suuren osan datan

kokonaisvaihtelusta, faktorianalyysin tavoitteena on selittää, miksi tietyt muuttujat käyttäytyvät samankaltaisesti [21]. FTIR-spektroskopian yhteydessä alkuperäisen datajoukon aaltolukujen oletetaan riippuvan taustalla olevista kemiallisista tekijöistä eli latenttimuuttujista. Nämä latenttimuuttujat vastaavat esimerkiksi erilaisia kemiallisia sidoksia tai funktionaalisia ryhmiä, jotka vaikuttavat useisiin aaltolukuihin samanaikaisesti [24]. Faktorianalyysin avulla voidaan siten tunnistaa, mitkä aaltoluvut liittyvät toisiinsa ja selittää, miksi samaan kemialliseen ryhmään liittyvät spektriipiirteet korreloivat keskenään [24].

Dataa käsitellään samankaltaisesti, kuin PCA:n tapauksessa. Aineisto standardoidaan yhtälön (4) mukaisesti ja siitä muodostetaan kovarianssimatriisi, kuten yhtälössä (5). Kovarianssimatriisille suoritetaan ominaisarvohajotelma, jonka avulla voidaan arvioida latenttien muuttujien määrä. Tätä varten tarkastellaan ominaisarvoja sekä niistä muodostettua Sree-kuvaajaa, joka on havainnollistettu kuvassa 4. Kuvaajasta nähdään, kuinka paljon tietty LV selittää havaittujen muuttujien yhteisestä vaihtelusta. Käytännössä tämä nähdään kuvaajan jyrkästä muuttumisesta. Kuvaajan notkokohdan jälkeen ominaisarvot pienenevät jyrkästi, minkä jälkeen LV:t selittävät vain satunnaista tai vähäistä vaihtelua, kuten taustakohinaa. Usein valitaan ominaisarvot, jotka ovat ennen notkoa ja joiden ominaisarvo on yli yksi. [22]



Kuva 4: Matlabilla piirretty havainnollistava Scree-kuvaaja, jossa yli yhden olevat ominaisarvot vastaavat ensimmäistä kolmea latenttipistettä.

Latenttien muuttujien tulkittavuutta parannetaan kiertämällä alkuperäisiä faktoriakseleita, jolloin saadaan pyöritetty matriisi  $V_{rot}$ . Pyöritysmenetelmä ja rotaatiomatriisin laskentatapa on esitetty liitteessä (Liite A.2). Lopulliset faktoripisteet lasketaan projektioina standardoidusta datasta yhtälön (7) mukaisesti [22]

$$F_{(n \times m)} = X_{std(n \times p)} V_{(p \times m)}^{rot}. \quad (7)$$

Faktoripistematriisi  $F_{(n \times m)}$  kuvaa FTIR-spektristä johdettuja latentteja kemiallisia rakenteita eli varsinaisia faktoreita, joita voidaan käyttää myöhemmässä analyysissä tai luokittelussa [22].

### 3.1.3 Osittaisen pienimmän neliösumman regressio

Osittaisen pienimmän neliösumman regressio PLSR (Partial Least Square Regression) on menetelmä sekä regressiomallinnukseen, että dimensioiden pienentämiseen. Toisin kuin PCA, joka keskittyy pääasiassa selittävien muuttujien varianssin arviointiin, PLSR:n tavoitteena on mallintaa selittävien muuttujien ja vasteiden välinen yhteys [21]. Menetelmä muodostaa lineaarisen mallin, jossa sekä selittävät muuttujat, kuten aaltoluvut, että vasteet, esimerkiksi bitumin vanhentumisaste, projisoidaan uuteen latenttiin koordinaatistoon. PLSR pyrkii maksimoimaan selittävien muuttujien ja vasteiden välisen kovarianssin [21]. Tämän ansiosta menetelmä soveltuu erityisesti tilanteisiin, joissa selittävien muuttujien määrä on suuri ja ne ovat keskenään voimakkaasti korreloituneita [24].

Ennen analyysiä selittävien muuttujien matriisi  $X_{(n \times p)}$  ja vasteiden matriisi  $Y_{(n \times q)}$  standardoidaan yhtälön (4) mukaisesti. PLSR hajottaa nämä matriisit iteratiivisesti scores- $T_{(n \times k)}$  ja loading- $P_{(p \times k)}$  -matriiseiksi. Selittävien muuttujien matriisi esitetään yhtälön (8) mukaisesti [22]

$$X_{std(n \times p)} = T_{(n \times k)} P_{(p \times k)}^T + e \quad (8)$$

ja vasteiden matriisi yhtälön (9) mukaisesti

$$Y_{std(n \times q)} = U_{(n \times k)} Q_{(q \times k)}^T + e. \quad (9)$$

Scores matriisit  $T$  ja  $U$  sisältävät latenttimuuttujat, jotka ovat muodostettu siten, että  $X$ - ja  $Y$ - matriisien välinen kovarianssi maksimoituu. Latenttimuuttujien määrä valitaan sen perusteella, kuinka hyvin malli selittää aineiston vaihtelua [22]. Käytetyt menetelmät latenttimuuttujien määrän valintaan on esitetty liitteessä (Liite A.3).

Mallin muodostamisen jälkeen PLSR:n tulkinta keskittyy siihen, mitkä selittävistä muuttujista vaikuttavat eniten vasteiden ennustamiseen. Tätä varten hyödynnetään VIP-pisteitä (Variable Importance in Projection), jotka kuvaavat, mitkä aaltoluvut ovat mallin kannalta tärkeimpiä [22]. VIP-pisteiden laskenta on esitetty liitteessä (Liite A.3). Yleisesti muuttujat, joiden VIP-arvo ylittää arvon 1, katsotaan merkittäväksi vasteen kannalta ja luokkien erottelussa [25]. Menetelmän heikkoutena on sen herkkyys epälineaarisuuksille ja kohinalle. Mikäli data ei ole hyvin esikäsitelty tai jos vasteiden määrä on pieni, mallista voi tulla ylisovitettu [26].

### 3.1.4 Luokittelumenetelmä

Linear Discriminant Analysis (LDA) on valvottu koneoppimismenetelmä, jota käytetään sekä luokitteluun, että dimensioiden vähentämiseen. Menetelmän tavoitteena on löytää datan ominaisuuksien lineaarinen yhdistelmä, joka erottaa eri luokat mahdollisimman hyvin. Tämä saavutetaan maksimoimalla luokkien välinen hajonta ja minimoimalla luokkien sisäinen hajonta. FTIR-analyysissä LDA auttaa tunnistamaan, mitkä aaltoluvut tai kemialliset piirteet vaikuttavat voimakkaimmin näytteiden erotteluun. LDA:n heikkoutena on sen oletus datan lineaarisuudesta ja normaalijakamasta. Jos bitumin FTIR-data on ei-lineaarista tai sisältää voimakkaita poikkeamia, menetelmän erottelukyky heikkenee. [22]

LDA perustuu kahteen hajontamatriisiin, luokkien sisäiseen hajontaan  $S_W$  ja luokkien väliseen hajontaan  $S_B$ , jotka määritellään yhtälöissä (10) ja (11). Luokkien sisäinen hajonta kuvaa datapisteiden vaihtelua omien luokkiensa ympärillä, kun taas luokkien välinen hajonta kuvaa luokkakeskisarvojen etäisyyttä koko aineiston keskiarvosta. Hajontamatriisit ovat

$$S_W = \sum_{i=1}^c S_i = \sum_{i=1}^c \sum_{x \in X_i} (x - m_i)(x - m_i)^T \quad (10)$$

$$S_B = \sum_{i=1}^c N_i(\mu_i - \mu)(\mu_i - \mu)^T, \quad (11)$$

missä  $c$  on luokkien lukumäärä,  $x$  on yksittäinen näyte,  $X_i$  on luokan  $i$  näytejoukko,  $N_i$  on luokan  $i$  näytteiden määrä,  $\mu_i$  on luokan  $i$  keskiarvovektori ja  $\mu$  on koko aineiston keskiarvovektori. [24]

LDA:n optimointitehtävä perustuu tavoitefunktioon, joka maksimoi luokkien välisen ja sisäisen hajonnan suhteen. Optimaalinen projektiomatriisi saadaan ratkaisemalla hajontamatriiseihin perustuva yleistetty ominaisarvotehtävä. Suurimmat ominaisarvot kuvaavat merkittävimmät latenttimuuttujat ja niitä vastaavat ominaisvektorit muodostavat LDA-projektion suunnat. Tavoitefunktion ja optimaalisten ominaisarvojen laskenta on esitetty liitteessä (Liite A.4). Löydettyjä ominaisarvoja vastaavat ominaisvektorit projisoivat datan pienempään ulottuvuuteen, jossa luokkien välinen ero maksimoituu yhtälön (12) mukaisesti [24]

$$Y_{(n \times k)} = X_{(n \times p)}W_{(p \times k)}. \quad (12)$$

Lopullinen luokittelu tehdään vertaamalla kunkin näytteen etäisyyttä projektiossa määriteltyihin luokkakeskisarvoihin. Näyte sijoittuu siihen luokkaan, jonka keskiarvoa se on lähimpänä [22]. Tällöin otetaan huomioon sekä luokkien keskiarvot, että luokkien sisäinen hajonta.

## 4 Tutkimusongelma

Vanhentuneen bitumin FTIR-spektrin tulkinnan perusideana on tunnistaa hapettumisen tuottamia funktionaalisia ryhmiä, kuten karbonyyliyhdisteitä (C=O) ja sulfoksiedeita (S=O). Elvytetyissä bitumeissa tilanne on kuitenkin monimutkaisempi, sillä elvyttimet voivat itse sisältää samoja funktionaalisia ryhmiä, jolloin FTIR-spektrissä näkyvä signaali ei enää erota, ovatko havaitut ryhmät peräisin bitumin hapettumisesta vai elvyttimen kemiallisesta rakenteesta. Spektrin tulkintaa hankaloittaa lisäksi muun muassa spektriipiikkien päällekkäisyys, epäinformatiiviset alueet, taustakohina ja näytteenvalmistelun vaihtelevuus. [11]

Elvytetyn bitumin spektrin monimutkaisuudesta johtuvan ongelman osoittivat Nayak ja Sahoo tutkimuksessaan [6], jossa he tarkastelivat kahden erilaisen pehmitinöljyn toimintaa bitumin elvyttäjinä. He havaitsivat, että ensimmäinen tutkittu elvytin pienensi odotetusti C=O-ryhmien suhteellista määrää, mikä viittaa hapettumisen vähenemiseen elvytyksen seurauksena. Ristiriitainen tulos saatiin kuitenkin

toisella elvyttämiseen käytetyllä öljyllä, sillä bitumin C=O-pitoisuus kasvoi selkeästi, vaikka kemiallisesti elvytyksen olisi pitänyt tuottaa päinvastainen ilmiö. Tulokset osoittivat voimakkaan absorptio-alueen keuhonien aaltolukualueella ja siksi FTIR antaa näennäisesti virheellisen signaalin elvytyksen tehokkuudesta. Tämä lisääntyminen ei kuitenkaan johtunut bitumin vanhenemisestä, vaan pehmittimen omasta kemiallisesta koostumuksesta [6].

Tämä esimerkki tiivistää keskeisen ongelman. FTIR-spektrin yksittäisten huippujen tulkinta ei riitä arvioimaan elvyttimien vaikutusta, koska elvyttimien tuomat funktionaaliset ryhmät voivat peittää alleen bitumin todelliset muutokset ja luoda harhaanjohtavia tulkintoja. Elvytettäessä, bitumissa tapahtuu useita kemiallisia muutoksia, kuten asfalteenien liukeneminen, malteeni-asfalteeni-suhteen muutos ja elvyttimen sekoittuminen bitumiin [6]. Nayak ja Sahoo painottavat, että spektrialueiden suhteelliset pinta-alat voivat olla käyttökelpoisia, mutta nekin on tulkittava varoen, jos elvyttimessä on samoja absorboivia ryhmiä [6]. Tästä syystä FTIR-analyysiä täytyy täydentää matemaattisilla ja tilastollisilla menetelmillä, jotka pystyvät erottamaan päällekkäiset signaalit ja tunnistamaan heikkoja tai osittain peittyviä piirteitä. Erityisesti monimuuttuja-analyysin (MVA) potentiaali on herättänyt kiinnostusta bitumialalla, sillä sen avulla voitaisiin tunnistaa vähemmän näkyviä piirteitä spektristä [8].

Monimuuttuja-analyysiä on hyödynnetty jo useita vuosikymmeniä muilla tieteenaloilla, mutta menetelmä on otettu vasta hiljattain käyttöön bitumien ja bitumisten materiaalien tutkimuksessa [7]. Kirjallisuudessa on toistaiseksi esitelty vain rajallinen määrä tutkimuksia, joissa MVA-menetelmiä on sovellettu elvytetyn bitumin analysointiin [8]. Tämän kirjallisuuskatsauksen tavoitteena on tarkastella, kuinka tehokkaasti monimuuttuja-analyysiä on hyödynnetty elvytetyn bitumin ja eri elvyttimien ominaisuuksien tunnistamisessa sekä arvioida, missä kohdin nykyiset tekniikat ovat riittämättömiä. Erityistä huomiota kiinnitetään siihen, miten FTIR-mittauksia on tähän saakka tulkittu visuaalisesti ja miksi tämä lähestymistapa saattaa johtaa virheellisiin johtopäätöksiin.

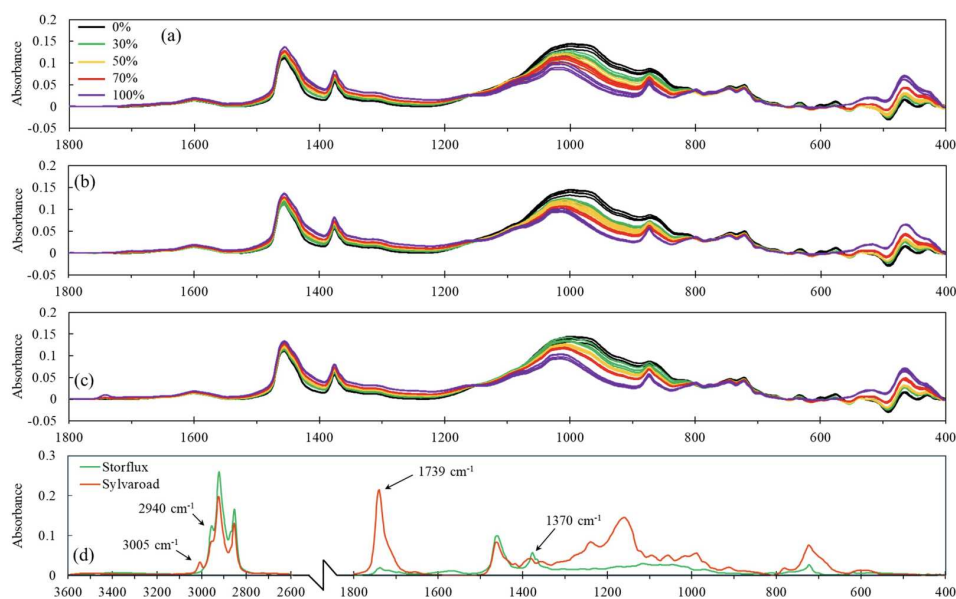
## 5 Tulokset

MVA:n hyödyntäminen elvytetyn bitumin tutkimuksessa on vielä melko uutta, minkä vuoksi kirjallisuudessa esitellyt tutkimukset perustuvat toistaiseksi vain rajallisiin näyttemääriin [7]. Saatu aineisto on kuitenkin lupaavaa, sillä useiden tutkimusten perusteella bitumin kemiallisia muutoksia voidaan tunnistaa monimuuttuja-analyysin avulla huomattavasti tarkemmin kuin perinteisillä, yksittäisiin FTIR-piikkeihin perustuvilla menetelmillä [19]. Tästä huolimatta bitumin elvyttämisen taustalla vaikuttavat mekanismit ovat edelleen puutteellisesti tunnettuja, mikä rajoittaa elvyttimien laajamittaista ja systemaattista hyödyntämistä [19]. Monimuuttuja-analyysi voi olla keino täyttää näitä tiedollisia aukkoja, sillä se kykenee paljastamaan spektrien monimutkaisia riippuvuussuhteita. MVA:n avulla voidaan tarkastella kokonaisia ja suuria spektridatoja, tunnistaa samanaikaisia kemiallisia muutoksia ja erotella elvyttimen tuottamat signaalit bitumin omista rakenteellisista muutoksista [27].

## 5.1 Elvyttimien erottelu

Motevalizadeh toteutti työssään mastiksin elvyttämiskokeilun kahdella eri elvyttimellä, Storfluxilla ja Sylvaroadilla [21]. Vaikka kokeissa käytettiin mastiksia eli bitumin ja kiviaineksen seosta, tulokset ovat hyvin verrattavissa pelkän bitumin käyttäytymiseen. Kiviaineksen kemiallinen vaikutus FTIR-spektrissä on vähäistä ja näkyy aallonlukuilla  $500\text{--}600\text{ cm}^{-1}$ . Näin ollen spektrissä havaitut muutokset kuvastavat pääosin bitumin kemiallisia reaktioita.

Kuvassa 5 esitetään FTIR-spektrit elvyttämättömälle mastiksille, elvytetyille mastiksille sekä elvyttimille. Eri väriset käyrät kuvaavat näytteitä, joissa vanhentuneen mastiksin osuus vaihtelee. Violetti käyrä edustaa näytettä, jossa on vain vanhentunutta mastiksia. Alakuvasta (a) nähdään, että eri mastikspitoisuuksien spektrit ovat yleisesti ottaen samankaltaisia muodoiltaan ja sitä käytetään vertailukohtana elvyttimien vaikutusten arvioinnissa. Spektrit ovat rajattu alueeseen, jossa piikit ovat selkeimmät ja ilmentävät suurimpia muutoksia. Tämä alue kattaa aallonpituudet  $1800\text{ cm}^{-1} - 400\text{ cm}^{-1}$ . Kuvan tulkinnessa täytyy huomioida aaltolukujen eri skaalaus x-akselilla. [21]



Kuva 5: (a) Näytteissä on eri määrä vanhentunutta, mutta elvyttämätöntä mastiksia, (b) Storfluxilla elvytetty mastiksi, (c) Sylvaroadilla elvytetty mastiksi, (d) Elvyttimien spektrit Kuva: [21]

Kuvan 5 kuvaajan (a) spektridatan perusteella vanhentuneen mastiksin osuuden kasvu lisäsi absorptiointensiteettiä alueilla  $600 - 400\text{ cm}^{-1}$  ja  $1700 - 1150\text{ cm}^{-1}$ . Vertaamalla näitä taulukon 1 aaltolukuihin voidaan tunnistaa muutoksia amidi-, karbonyyli- ja aromaattisissa sidoksissa. Näiden funktionaalisten ryhmien kasvaneet pitoisuudet viittaavat hapettumistuotteiden määrän kasvuun [21]. Vastaavasti alueella  $1150 - 800\text{ cm}^{-1}$  havaittiin absorptiointensiteetin laskua, mikä kertoo bitumia pehmentävien alifaattisten  $\text{CH}_3$ -ryhmien vähenemisestä [21].

Silmämääräisesti Storfluxin ja Sylvaroadin lisääminen aiheutti elvytettyjen näytteiden spektreissä (b) ja (c) hyvin pieniä muutoksia verrattuna elvyttämättömään näytteeseen. Tämä tekee elvytyksen vaikutusten visuaalisesta tulkinnasta vaikeaa ja epäluotettavaa. Spektrissä (d) esitetyt elvyttimien omat FTIR-profiilit kuitenkin auttavat tunnistamaan elvyttimien koostumuksia. Sylvaroadin selkeä piikki kohdassa  $1739\text{ cm}^{-1}$  viittaa esterien suureen pitoisuuteen. Ne reagoivat herkästi hapettuneiden yhdisteiden kanssa vähentäen niiden pitoisuutta [21]. Storfluxin piikki kohdassa  $1370\text{ cm}^{-1}$  puolestaan viittaa metyyliryhmiin ( $\text{CH}_3$ ), jotka korvaavat vanhenemisessa menetettyjä hiilivetyjä tasapainottaen komponenttien polaarisuutta [21].

Koska pelkkä visuaalinen tarkastelu ei riitä havaitsemaan merkittäviä eroja, Motivalizadeh sovelsi monimuuttuja-analyysiä näytteiden luokitteluun. Analyysissä hyödynnettiin kaikkia mitattuja aaltolukuja eli pystyttiin hyödyntämään laajempaa datajoukkoa, kuin visuaalisessa spektritulkinnassa [28]. Tavoitteena oli erottaa elvyttämättömät, Sylvaroadilla elvytettyt ja Storfluxilla elvytettyt mastiksit toisistaan. Motivalizadeh ei erittele tarkasti, minkälaisia MVA menetelmiä käytti juuri tässä tutkimuksessa, vaan viittaa soveltaneensa aiemmin toteuttamaansa tutkimusta mastikisien vanhenemisesta [28].

Motivalizadehin viittaaman aiemman tutkimuksen tulokset osoittivat, että PCA:n yhteydessä SNV- ja SG-esikäsitteilyt paransivat merkittävästi luokittelutarkkuutta [28]. Motivalizadeh on voinut päätyä käyttämään juuri näitä esikäsitteilytyökaluja, koska SNV normalisoi spektriä poistamalla näytteen pinnan epätasaisuudesta aiheutuvaa vaihtelua [29]. SG puolestaan tasoittaa datan kohinaa ja korostaa pieniä piikkejä [29]. Näiden yhdistelmä auttaa PCA:ta keskittymään datan suurimpiin vaihteluihin. Motivalizadeh huomasi, että FA:n kanssa vaikutus oli vähäisempi, sillä FA keskittyy muuttujien välisiin korrelaatioihin, joita SNV ja SG saattavat muuttaa epäedulliseksi [28]. Tämä tulos on yhteneväinen teorian kanssa, sillä FA pyrkii mallintamaan muuttujien välisiä korrelaatioita, eikä pelkästään varianssia kuten PCA [24]. Osa kemiallisesti merkittävistä muutoksista saattaa näkyä FA:n näkökulmasta kohinana tai ei-informatiivisina osina.

Motivalizadeh tutki myös datan ensimmäisen ja toisen derivaatan vaikutusta luokitteluun. Ensimmäisen derivaatan käyttö paransi tuloksia kaikilla menetelmillä, sillä se poisti spektrin tasaisen taustan ja korosti eroja näytteiden välillä. Toisen derivaatan käyttö sen sijaan lisäsi kohinaa ja pienensi tarkkuutta. [28]

Paras luokittelutulos saavutettiin valitsemalla PLSR, jonka avulla tunnistettiin komponentit, jotka korreloivat vahvimmin elvytyksen vaikutuksiin. Elvytinerotteeluun Motivalizadeh valitsi lisäksi LDA-luokittelijan. Se mahdollisti elvyttämättömän, Sylvaroad elvytetyn ja Storflux elvytetyn näytteen lineaarisen erottelun. Mallin luokittelutarkkuus oli erinomainen, keskimääräinen tarkkuus oli 0,94. Ainoa ykistäinen luokitteluvirhe ilmeni Storfluxin ja Sylvaroadin erottamisessa. [21]

Motivalizadeh hyödynsi PLSR-mallin sekä LDA:n lisäksi myös kolmea muuta luokittelualgoritmia arvoidakseen datan muuttujien merkittävyyttä monipuolisesti [21]. Tämä on ollut järkevä valinta, sillä yhden algoritmin varaan rakentuva johtopäätös voi olla virheellinen. Kaikki neljä menetelmää tuottivat lopulta yhteneväisen kuvan tärkeimmistä aaltolukualueista [21]. Muuttujien tärkeysjärjestys määritettiin VIP-pisteiden avulla, jotka laskettiin PLSR-mallista [21].

Tärkeimmät aaltolukualueet, jotka erottivat elvytetyt ja elvyttämättömät mastiksit, liittyivät Sylvaroadin vaikutuksiin. Ensimmäinen merkittävä alue oli 1758–1715  $\text{cm}^{-1}$ , mikä vastaa karbonyyliryhmän ( $\text{C}=\text{O}$ ) absorptiota [21]. Karboksyylihapot ja esterit sisältävät tällaisia sidoksia. Kuvassa 5 esitetyn Sylvaroadin spektrin mukaan, se sisältää runsaasti estereitä, jotka voivat reagoida hapettuneiden yhdisteiden kanssa, vähentäen niiden vaikutusta.

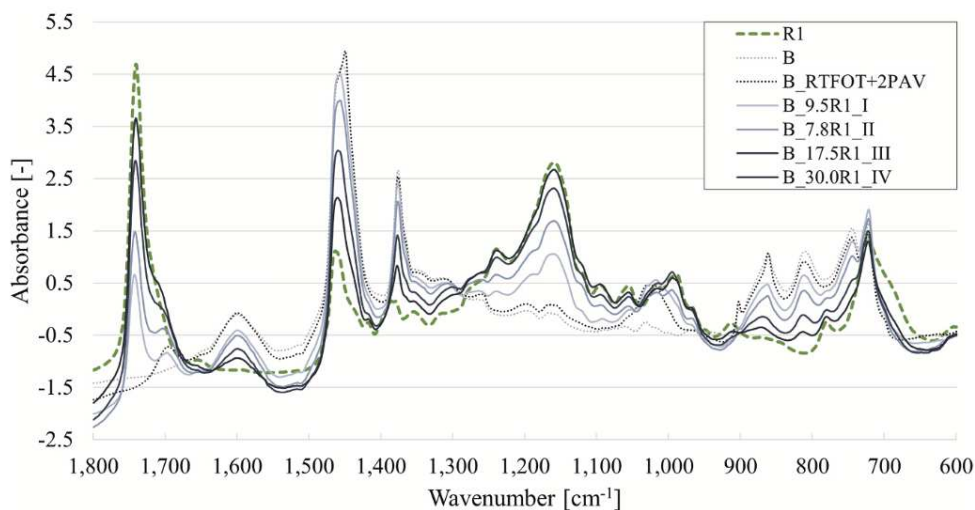
Toinen tärkeä alue havaittiin erityisesti Storflux-elvytetyillä näytteillä alueella 1370–1360  $\text{cm}^{-1}$ , mikä liittyy metyyli- ( $\text{CH}_3$ ) ja metyleeni- ( $\text{CH}_2$ ) ryhmiin [21]. Metyyli-ryhmät esiintyvät sekä kyllästyneissä että aromaattisissa hiilivedyissä, kun taas metyleeniryhmät ovat yleisiä alifaattisissa hiilivedyissä. Motivalizadehin havaitsema lisääntynyt absorptiovoimakkuus tällä alueella viittaa alifaattisten komponenttien palautumiseen, mikä parantaa bitumin elastisuutta. Erottelu tunnisti myös, että aaltolukuvälillä 1330–1275  $\text{cm}^{-1}$  näkyi vielä elvyttämisen jälkeen jäänteitä hapettumisesta [21].

Motivalizadehin tulokset osoittavat, että täydellistä elvytystä ei saavutettu [21]. Tulos oli kuitenkin odotettavissa, sillä osa vanhenemisen reaktioista ei ole palautuvia. Osittain epäonnistuneen elvytyksen lisäksi VIP-pisteet paljastivat alueet, joissa elvyttimet tuottavat spektriin omia karakteristisiä piikkejään. Piikit osuivat samoille alueille, kuin vanhanemiseen liittyvät muutokset [21]. Tämä vuoksi absorptiointensiteetin kasvu tai väheneminen yksittäisellä aaltoluvulla ei kerro todellisesta muutoksesta. Aaltolukualueet, joissa elvyttimet vaikuttivat funktionaalisiin ryhmiin saatiin eroteltua MVA:n avulla ja siten niiden vaikutus pystyttiin huomioimaan ja vähentämään lopullisissa tuloksissa [21]. Näin laadukkaiden ja tarkkojen tulosten saaminen olisi ollut mahdotonta vertailemalla absorptiospektrejä visuaalisesti, koska piikkien muutokset olivat osittain päällekkäisiä ja niiden suhteelliset muutokset pieniä. Voidaan siis todeta, että esikäsittely ja MVA-analyysi paransi FTIR-spektroskopian tuloksia, koska erojen tunnistamisen lisäksi saatiin myös selitys niiden kemialliselle taustalle. Visuaaliseen tulkintaan verrattuna tulokset ovat myös luotettavampia, koska MVA hyödynsi koko spektrialuetta. Käsiteltävä data-joukko on siis huomattavasti suurempi ja spektrissä hapettumistuotteiden ja elvyttimien päällekkäiset ja peittyneet piikit pystyttiin huomioimaan tuloksissa.

## 5.2 Elvyttämisen tehokkuus

Schewettmann-Lui käytti tutkimuksessaan kolmea bituminäytettä sekä kolmea elvytintä [30]. Tutkimuksen tarkoituksena oli lisätä vanhentuneeseen näytteeseen elvytintä asteittain ja mitata muuttuvaa vanhentumisen tasoa. Mittausdataa kerättiin jokaisen elvytintä käytön jälkeen, yhteensä viidestä eri elvytystasosta. Mittaukset toteutettiin ATR-FTIR-spektroskopiolla ja tuloksista saatiin kuva 6 mukainen absorptiospektri. Spektri normalisoitiin ja rajattiin selkeiden piikkien alueelle, aaltolukuvälille 1800 – 600  $\text{cm}^{-1}$ .

Vanhentuneessa bitumissa hapellisten sulfoksidien ja karbonyyli-erien määrän kasvu on tyypillinen merkki vanhenemisestä. Elvytetyissä seoksissa hapen ryhmien määrä ei enää kerro yksiselitteisesti totuutta vanhenemisestä [6]. Schwettmann-Lui havaitsi kyseisen ilmiön, koska elvyttimet toivat mukanaan omia hapellisia ja aromaattisia



Kuva 6: Vanhentuneen bituminäytteen B normalisoitu absorptiospektri verrattuna neljään elvytyvaiheeseen I-IV ja elvyttimeen R1. Kuva: [30]

yhdisteit. Kuvan 6 spektrissä on oleellista nähdä, kuinka seoksen spektri liikkuu lähemmäs elvyttimen R1 spektriä, kun elvytinkierroksia lisättiin. Neljännen (IV) elvytinkierroksen spektri mukailee elvyttimen spektriä jo niin tarkasti, että tulos vaikuttaa epäluotettavalta.

Schwettmann-Luin tutkimuksessa [30] saatu data normalisoitiin SNV:n ja ensimmäisen derivaatan avulla. Esikäsitelty data pienennettiin FA:lla ja näytteet eroteltiin LDA:lla. Koska LDA pystyy erottamaan kerrallaan vain kahden ryhmän välisiä eroja yhdellä diskriminanttifunktiolla, tutkimuksessa täytyi käyttää iteratiivista menetelmää. Jokaisella kierroksella LDA tunnisti sen ryhmän, joka oli kaikkein erottuvin, esimerkiksi ensimmäisellä kierroksella se oli puhtaan, elvyttämättömän bitumin näyte. Koska yksittäisiä iteraatioita ei voida verrata suoraan keskenään, laskettiin keskimääräinen osuutarkkuus. Kokonaisuudessaan kalibroituilla näytteillä keskimääräinen osuutarkkuus oli 94%, kun taas mallin ulkopuolisilla validointinäytteillä osuutarkkuus laski kolmannella kierroksella 80%:iin. Tämä voi tarkoittaa, että malli kykenee erottamaan sille tuttuun näytteiden elvytysvaiheet hyvin toisistaan, mutta sen yleistettävyyys uusille näytteille heikkenee merkittävästi elvytyskierrosten edetessä.

FA- ja LDA-yhdistelmä mahdollisti siis elvytyksen eri vaiheiden tunnistamisen, mutta paljasti myös ongelman. Kun elvyttimen konsentraatio kasvaa näytteessä, sen oma signaali alkaa dominoida spektriä ja peittää alleen bitumin vanhentumiseen liittyviä piirteitä. Tämä näkyi luokittelutarkkuuden selkeänä heikkenemisenä myöhemmissä elvytysvaiheissa, mikä viittaa siihen, että malli ei enää pysty erottamaan, liittyykö spektrin muutos elvyttämiseen vai elvyttimen määrän määrän lisääntymisestä. Toisin sanoen, tutkimus nosti esiin bitumin elvytyksen mallintamisen rajoitteen. Mitä pidemmälle elvytys etenee, sitä vaikeampaa on erottaa, onko havaittu kemiallinen muutos merkki todellisesta vanhenemisen palautumisesta vai elvyttimen vaikutuksesta mittaussignaaliin [30].

Tästä huolimatta saavutettu 94 % luokittelutarkkuus osoittaa, että monimuuttuja-analyysiin perustuva lähestymistapa pystyy mallintamaan elvytyksen etenemistä luotettavasti ainakin alkuvaiheissa. Tämän tuloksen ansioista voisi olla mahdollista vähentää fysikaalisia laboratoriotestejä ja tukeutua enemmän laskennallisiin tapoihin, mikäli MVA-mallit kalibroidaan huolellisesti ja sidotaan tunnettuun elvytysprosessiin. Lisäksi FA-LDA-menetelmän iteratiivinen käyttö osoitti potentiaalia elvytysmallien kehittämiseen, joissa elvytetyn bitumin tilaa voidaan seurata useissa elvytysvaiheissa. Tämän hyödyntäminen voisi olla mahdollista esimerkiksi reaaliaikaisessa laadunvarmistuksessa, sillä menetelmä on nopea ja helposti toistettavissa.

### 5.3 Kriittinen tarkastelu

Motivalizadehin tutkimus osoittaa [21], että luotettavan menetelmän löytäminen vaatii useiden mallien testaamista. Hänen käyttämien neljän luokittelualgoritimin tuottamat VIP-pisteet osoittivat samoihin merkittäviin aaltolukualueisiin. Tästä voidaan päätellä, että VIP-pisteet todella vastaavat tapahtuneita kemiallisia muutoksia ja algoritmit olivat toimivia. Riittävä vertailu todistaa, että kyseessä on todellinen muutos ja sattuman todennäköisyyttä voidaan pitää pienenä. Motivalizadehin tekemässä tutkimuksessa ei kuitenkaan oteta kantaa mahdollisiin epäonnistuneisiin analyysimenetelmiin. Mikäli epäoptimaalisia tapoja löytyi, niiden läpinäkyvä esittäminen olisi tuoreen tutkimusalan kehittämisen kannalta oleellista.

Schwettmann-Luin toteuttaman tutkimuksen raportoinnissa ei käy ilmi Motivalizadehin tutkimuksen kaltaista systemaattista menetelmien testaamista [21, 30]. Schwettmann-Lui ei esitä selvästi numeerista dataa tai sanallista selitystä, miksi juuri FA yhdistettynä LDA:han oli luotettavin vaihtoehto. Tämä on oleellinen huomio, sillä Schwettmann-Luin saama tulos heikkeni elvytyskierrosten edetessä. Mikäli juuri kyseiselle datalle ja tutkimuskysymykselle valittua MVA-menetelmää ei ole tutkittu kriittisesti, on mahdollista, että toisenlaisella analyysimenetelmien yhdistelmällä olisi voitu saavuttaa tarkempi tulos.

Khaligin toteuttama laaja vertailututkimus osoittaa [8], että kokonaiseen ATR-FTIR-spektriin tai sen ensimmäiseen derivaattaan perustuvat mallit tuottavat systemaattisesti paremman ja vakaamman luokittelutarkkuuden, kuin indeksipohjaiset tai yksittäisiin huippuihin perustuvat lähestymistavat. Khaligin mukaan väite perustuu siihen, että koko spektrin käyttö analyysin alusta asti säilyttää myös heikommat, mutta oleelliset toistuvat kemialliset muutokset. Tämä korostaa entisestään Motivalizadehin tulosten uskottavuutta, sillä useiden luokittelualgoritmien tuottamat yhteneväiset VIP-pisteet viittaavat koko spektrissä esiintyviin todellisiin kemiallisiin muutoksiin. Samalla Schwettmann-Luin käyttämä FA-menetelmä mahdollisesti rajoittuu. Faktorianalyysi pienentää dimensioita etsimällä spektrimuutokset, jotka muuttuvat eniten, ilman suoraa yhteyttä luokittelutavoitteeseen. Kun elvyttimen osuus bituminäytteessä kasvaa, se aiheuttaa voimakkaan ja systemaattisen muutoksen spektrissä. Tämä kasvattaa spektrin kokonaisvarianssia ja FA:n muodostamat faktorit alkavat mahdollisesti kuvata ensisijaisesti elvyttimen määrän vaihtelua, eikä bitumin elvyttymiseen liittyviä muutoksia. On mahdollista, että osa elvyttymiseen liittyvästä informaatiosta hukkuu kokonaan ja siksi tulos vääristyy.

Khalighi esittikin tutkimuksessaan väitteen, että osa tutkijoista valitsee MVA-menetelmän olettaen sen parantavan tuloksia ilman lisätutkimusta [8]. Tällä voi olla vaikutus tulosten toistettavuuteen ja se saattaa väristää luokittelumallien todellista suorituskykyä. Koska Schwettmann-Lui ei esitä selkeitä perusteluja FA-menetelmän valitsemiselle, on syytä olla kriittinen MVA-menetelmän valintaa ja saatuja tuloksia kohtaan. Khalighin mukaan on mahdollista, että julkaistuissa tutkimuksissa myös esikäsitelymenetelmän valinta perustuu vain muiden tutkimusten käytäntöihin, eikä systemaattiseen vertailuun tai datan arviointiin. Hänen mukaansa useimmat tutkimukset näyttävät valitsevan menetelmät kirjallisuuden perusteella, sillä vain muutamia menetelmiä tuodaan johdonmukaisesti esiin. Tämä käytäntö voi johtaa siihen, että tutkijat ohittavat juuri heidän datalleen parhaiten sopivat menetelmät. Lisäksi kirjallisuuden perusteella suurin osa tutkijoista ei ole selkeästi verrannut valitun menetelmän suorituskykyä raakadataan eli käsittelemättömään FTIR-spektriin. [8]

Khalighi totesi esimerkiksi, että omassa tutkimuksessaan baseline-korjausmenetelmät eivät parantaneet merkittävästi koko spektrin laatua, mikä voi johtua baseline-väristymien vähäisestä vaikutuksesta spektriin [8]. Khalighin havainto osoittaa, että kaikki datan muokkaus ei välttämättä ole parannusta. Tietyissä tapauksissa esikäsitely voi heikentää mallin kykyä tunnistaa olennaisia riippuvuuksia, jos se poistaa samalla näytteiden todellisiin eroihin liittyvää informaatiota. Tämä osoittaa jälleen, että FTIR-spektroskopian yhteydessä käytettävien MVA-menetelmien tulee perustua tietoisesti valittuun datan käsittelymenetelmään.

Monet aiemmat mallit perustuvat pieniin ja kapeasti rajattuihin aineistoihin, jotka eivät välttämättä edusta eri bitumilaatujen ja elvyttimien välistä vaihtelua. Tämä rajoittaa mallien yleistettävyyttä ja tekee niiden soveltamisesta uusiin näytteisiin epäluotettavaa. Tulevissa tutkimuksissa tulisi selvittää, voidaanko olemassa olevat ja toimivat mallit mukauttaa eri elvytinlaaduille sen sijaan, että jokaiselle kehitetään oma mallinsa. Seuraava askel olisi myös laajentaa mallien tietokantaa, jotta se kattaisi yleisimmät markkinoilla olevat elvyttimet. Jos monimuuttuja-analyysimalli opetetaan vain yhdellä bitumilaadulla, se oppii tunnistamaan kyseisen materiaalin ominaispiirteet, eikä välttämättä yleisiä vanhenemis- tai elvytysmekanismeja. Siksi mallien tulisi perustua laajoihin ja erilaisiin datasetteihin, jotta ne tunnistavat bitumin käyttäytymistä yleisellä tasolla. Laajempien aineistojen avulla olisi mahdollista kehittää yleisiä, standardoitavia MVA-malleja, jotka voisivat tunnistaa eri elvytystasot ja vanhenemiseet riippumatta käytetystä elvyttimestä. Tämä voisi kehittää automatisoituneempia tutkimuskäytäntöjä ja parantaisi elvytyskokeiden vertailukelpoisuutta.

## 6 Johtopäätökset

Bitumi on kemiallisesti erittäin monimutkainen asfalttimassojen sideaine, jonka koostumus ja rakenne muuttuvat jatkuvasti vanhenemisen ja elvyttämisen seurauksena. Sen sisältämien funktionaalisten ryhmien, kuten karbonyyli-, sulfoksidi-, hydroksyyli- ja alifaattisten hiilivetyryhmien väliset reaktiomekanismit ovat edelleen puutteellisesti

ymmärrettyjä. Näiden ryhmien muutokset vaikuttavat suoraan bitumin fysikaalisiin ominaisuuksiin, kuten jäykkyyteen ja viskositeettiin.

FTIR-spektroskopia on tehokas menetelmä bitumin kemiallisten muutosten tarkasteluun, sillä se esittää yhteyden molekyyli-rakenteen ja havaittujen absorptiospektrien välillä. Menetelmän avulla voidaan tunnistaa vanhenemistuotteiden, kuten karbonyyli- ja sulfoksidiryhmien pitoisuuksien kasvua, mikä kertoo hapettumisreaktioiden etenemisestä. Nämä muutokset antavat kuitenkin vain rajallisen kuvan elvytyksen vaikutuksesta, sillä ne eivät paljasta piilossa olevia kemiallisia vuorovaiikutuksia.

Tässä yhteydessä monimuuttuja-analyysi on osoittautunut lupaavaksi työkaluksi. Motevalizadehin tutkimus osoitti, että yhdistämällä PLSR- ja LDA-menetelmät sekä käyttämällä sopivia esikäsitteilyjä, voitiin erottaa elvytetyt ja elvyttämättömät mastiksinäytteet lähes täydellisellä tarkkuudella [21]. Analyysin avulla tunnistettiin myös, mitkä aaltolukualueet korreloivat voimakkaimmin elvyttimien vaikutusten kanssa. Tämä vahvistaa sen, että FTIR-spektrin koko informaatiopotentiali tulee esiin vasta, kun dataa käsitellään matemaattisesti useiden muuttujien samanaikaisena kokonaisuutena.

Schwettmann-Luin tutkimus puolestaan osoitti, että MVA-menetelmät ovat välttämättömiä, kun bitumin seoksessa on useita elvytyskierruksia ja elvytin vaikuttaa suoraan spektriin [30]. Tässäkään tapauksessa pelkkä visuaalinen tarkastelu ei enää riittänyt, koska elvyttimen ja bitumin päällekkäiset piikit hämärsivät vanhenemisen merkkejä. MVA-menetelmien avulla saatiin lupaavia tuloksia, mutta tutkimuksen raportoinnissa oli oleellisia puutteita.

Khalighin analyysi paljastaa monimuuttuja-analyysin tehokkuuden riippuvan olennaisesti käytetyistä esikäsitteilymenetelmistä ja MVA-mallista [8]. Hänen havaintonsa korostavat, ettei esikäsitteilyä tule tehdä mekaanisesti kirjallisuuden perusteella, vaan sen vaikutus tulisi aina miettiä kyseessä olevaan dataan sopivaksi. Kokonaisuutena tarkasteltuna MVA:n käyttö bitumitutkimuksessa antaa etuja verrattuna perinteiseen FTIR-spektrin visuaaliseen tulkintaan. Se mahdollistaa piilossa olevien kemiallisten muutosten ja korrelaatioiden huomaamisen, tunnistaa elvyttämisen muutoksia sekä parantaa analyysin tarkkuutta. Samalla se kuitenkin edellyttää huolellista esikäsitteilyn valintaa ja riittävän laajoja aineistoja, jotta mallien yleistettävyyttä säilyy.

Tulevaisuudessa tutkimusten keskeisenä tavoitteena tulisi olla monimuuttuja-analyysien standardointi ja tietokantojen kehittäminen, jotka kattavat erilaiset bitumilaadut ja elvyttimet. Tällöin olisi mahdollista luoda yleisiä malleja, jotka pystyvät tunnistamaan vanhenemisen ja elvytyksen tason riippumatta näytteen alkuperästä. Tutkimuksissa tulisi myös esittää selkeä ja perusteltu erittely monimuuttuja-analyysimenetelmän valintasyistä sekä kriittinen tarkastelu valitun mallin toimivuudesta kyseisessä aineistossa. Tämä edistäisi automaattisten ja reaaliaikaisten analyysimenetelmien käyttöönottoa.

## Viitteet

- [1] A. S. Dr Robert N. Hunter and P. J. Read, "The shell bitumen handbook," no. 6, 2015.
- [2] J. C. Petersen, "A review of the fundamentals of asphalt oxidation: Chemical, physicochemical, physical property, and durability relationships," *Transportation Research Circular*, no. E-C140, 2009.
- [3] E. Prosperi and E. Bocci, "A review on bitumen aging and rejuvenation chemistry: Processes, materials and analyses," *Sustainability.*, vol. 13, no. 12, 2021-6-08.
- [4] Y. Li, J. Feng, S. Wu, A. Chen, D. Kuang, T. Bai, Y. Gao, J. Zhang, L. Li, L. Wan, Q. Liu, Z. Chen, and D. Gu, "Review of ultraviolet ageing mechanisms and anti-ageing methods for asphalt binders," *Journal of Road Engineering*, vol. 2, no. 2, 2022-06.
- [5] X. Hou, S. Lv, Z. Chen, and F. Xiao, "Applications of fourier transform infrared spectroscopy technologies on asphalt materials," *Measurement.*, vol. 121, 2018-06.
- [6] P. Nayak and U. C. Sahoo, "Rheological, chemical and thermal investigations on an aged binder rejuvenated with two non-edible oils," *Road materials and pavement design : an international journal.*, vol. 18, no. 3, 2017-05-04.
- [7] L. Ma, A. Varveri, R. Jing, and S. Erkens, "Chemical characterisation of bitumen type and ageing state based on ftir spectroscopy and discriminant analysis integrated with variable selection methods," *Road materials and pavement design : an international journal.*, vol. 24, no. sup1, 2023-04-25.
- [8] S. Khalighi, L. Ma, S. Ren, and A. Varveri, "Evaluating the impact of data pre-processing methods on classification of atr-ftir spectra of bituminous binders," *Fuel.*, vol. 376, 2024-11.
- [9] D. N. Little, D. H. Allen, and A. Bhasin, *Modeling and Design of Flexible Pavements and Materials*. Cham: Springer International Publishing AG, 2018. [Online]. Available: <https://doi.org/10.1007/978-3-319-58443-0>
- [10] Z. Chen, H. Zhang, H. Duan, C. Wu, and S. Zhang, "Long-term photo oxidation aging investigation of temperature-regulating bitumen based on thermochromic principle," *Fuel*, vol. 286, p. 119403, 2021. [Online]. Available: <https://doi.org/10.1016/j.fuel.2020.119403>
- [11] V. Loise, P. Caputo, M. Porto, P. Calandra, R. Angelico, and C. Oliviero Rossi, "A review on bitumen rejuvenation: Mechanisms, materials, methods and perspectives," *Applied sciences.*, vol. 9, no. 20, 2019-10-14.
- [12] A. A. Abe, P. Caputo, S. Eskandarsefat, V. Loise, M. Porto, E. Giorno, L. Venturini, and C. O. Rossi, "Rejuvenating agents vs. fluxing agents: Their respective mechanisms of action on bitumen subjected to multiple aging

- cycles,” *Applied Sciences*, vol. 13, no. 2, p. 698, 2023. [Online]. Available: <https://doi.org/10.3390/app13020698>
- [13] C.-P. S. Hsu *et al.*, “Infrared spectroscopy,” *Handbook of instrumental techniques for analytical chemistry*, vol. 249, 1997.
- [14] A. Ben-David and A. Ifarraguerri, “Computation of a spectrum from a single-beam fourier-transform infrared interferogram,” *Applied optics-OT.*, vol. 41, no. 6, 2002-02-20.
- [15] T. Wenzel. (2023) 4.1: Introduction to infrared spectroscopy. Chemistry LibreTexts. CC BY-NC 4.0 licence (LibreTexts), accessed 2025-12-22. [Online]. Available: <https://chem.libretexts.org/@go/page/111678>
- [16] ScienceDirect Topics. (n.d.) Fourier transform infrared spectrometer. Accessed: 2025-12-22. [Online]. Available: [https://www.sciencedirect.com/topics/engineering/fourier-transform-infrared-spectrometer?utm\\_source=chatgpt.com](https://www.sciencedirect.com/topics/engineering/fourier-transform-infrared-spectrometer?utm_source=chatgpt.com)
- [17] J. Van Gompel, “The fundamentals of infrared spectroscopy,” Midac Corporation, —, White Paper / Technical Note TN-100, —, PhD. [Online]. Available: <https://www.midac.com/files/Tn-100.PDF>
- [18] J. Gradadolnik, “Atr-ftir spectroscopy: Its advantage and limitations,” *Acta chimica Slovenica.*, vol. 49, no. 3, 2002.
- [19] S. Weigel and D. Stephan, “The prediction of bitumen properties based on ftir and multivariate analysis methods,” *Fuel.*, vol. 208, 2017-11.
- [20] Amit, S. Kumari, and R. Jamwal, “Use of ftir spectroscopy integrated with multivariate chemometrics as a swift, and non-destructive technique to detect various adulterants in virgin coconut oil: A comprehensive review,” *Food Chemistry Advances*, vol. 2, p. 100203, 2023. [Online]. Available: <https://doi.org/10.1016/j.focha.2023.100203>
- [21] M. Motevalizadeh and K. Mollenhauer, “Rejuvenation mechanisms in bituminous rap mastics: insights from ftir spectroscopy and multivariate discriminant analysis,” *Research Square*, 2025. [Online]. Available: <https://doi.org/10.21203/rs.3.rs-6342505/v1>
- [22] S. M. Motevalizadeh and K. Mollenhauer, “Exploration of chemical changes in bituminous mastics induced by aging: insights from ftir spectroscopy, dsr measurements, and machine learning,” *International Journal of Pavement Engineering*, vol. 25, no. 1, 2024. [Online]. Available: <https://doi.org/10.1080/10298436.2024.2418927>
- [23] N. B. Gallagher, “Savitzky-golay smoothing and differentiation filter,” Eigenvector, Tech. Rep., 2020.

- [24] S. M. Motevalizadeh and K. Mollenhauer, “Exploration of chemical changes in bituminous mastics induced by aging: insights from ftir spectroscopy, dsr measurements, and machine learning,” *International journal of pavement engineering.*, vol. 25, no. 1, 2024-12-31.
- [25] D. M. Selegato, T. R. Freitas, M. Pivatto, A. D. Pivatto, A. C. Pilon, and I. Castro-Gamboa, “Time-related multivariate strategy for the comprehensive evaluation of microbial chemical data,” *Metabolomics*, vol. 18, no. 6, p. 33, 2022. [Online]. Available: <https://doi.org/10.1007/s11306-022-01896-6>
- [26] N. M. Faber and R. Rajkó, “How to avoid over-fitting in multivariate calibration - the conventional validation approach and an alternative,” *Analytica Chimica Acta*, vol. 595, no. 1-2, pp. 98–106, 2007. [Online]. Available: <https://doi.org/10.1016/j.aca.2007.05.030>
- [27] P. de Peinder, *Characterization and classification of crude oils using a combination of spectroscopy and chemometrics*, 2009.
- [28] S. M. Motevalizadeh, K. Mollenhauer, and J. Wetekam, “Ftir spectroscopy and multivariate discriminant analysis for classifying bituminous mastics: Exploring aging states and mastic composition,” *Construction and Building Materials*, vol. 438, p. 137188, 2024, open access under CC BY license.
- [29] M. Martínez-Julca, R. Nazario-Naveda, M. Gallozzo-Cárdenas, S. Rojas-Flores, H. Chinchay-Espino, A. Alvarez-Escobedo, and E. Murga-Torres, “Classification of peruvian flours via nir spectroscopy combined with chemometrics,” *Applied Sciences*, vol. 13, no. 20, p. 11534, 2023. [Online]. Available: <https://doi.org/10.3390/app132011534>
- [30] K. Schwettmann-Lui, N. Nytus, S. Weigel, M. Radenberg, and D. Stephan, “Effect of multiple-repeated aging and rejuvenation processes on bitumen properties evaluated with physical and chemical tests and multivariate analysis methods,” *Construction building materials.*, vol. 472, 2025-04.

## A Monimuuttuja-analyysin matemaattinen tarkastelu

### A.1 Pääkomponenttianalyysi

Kovarianssimatriisin ominaisarvohajotelma tuottaa ominaisvektorit  $V_{(p \times p)}$  ja ominaisarvot  $\Lambda_{(p \times p)}$ , joissa ominaisvektorit edustavat pääkomponentteja ja ominaisarvot ilmaisevat kunkin komponentin selittämän vaihtelun määrän. Yhtälön (A1) ominaisarvohajotelman tuloksena saadaan siis ominaisvektromatriisi  $V_{(p \times p)}$  ja ominaisarvojen diagonaalimatriisi  $\Lambda_{(p \times p)}$ . [24]

$$\Sigma_{(p \times p)} V_{(p \times p)} = V_{(p \times p)} \Lambda_{(p \times p)} \quad (\text{A1})$$

### A.2 Faktorianalyysi

Jokainen muuttuja, eli aaltoluku järjestetään sen selittämän varianssin perusteella. Yhtälön (A2) esittämän ominaisarvojen diagonaalimatriisin  $\Lambda_{(p \times p)}$  jokainen diagonaalielementti  $l_i$  ilmaisee sen vastaavan latenttimuuttujan selittämän vaihtelun määrän. Kun ominaisarvot  $(l_1 \dots l_p)$  on saatu, muodostetaan Scree-kuvaaja, missä ominaisarvot esitetään laskevassa järjestyksessä. Kuvaajassa pystyakseli esittää selitetyn varianssin määrää ja vaaka-akseli komponentin numeroa, eli  $\Lambda$  indeksiä.

$$\Lambda_{p \times p} = \begin{bmatrix} l_1 & 0 & \dots & 0 \\ 0 & l_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & l_p \end{bmatrix}, \quad (\text{A2})$$

missä  $l_1 \geq l_2 \geq l_3 \geq \dots \geq l_p$ .

Matriisia  $V_{(p \times p)}$  pyöritetään tulkittavuuden parantamiseksi. Näin saadaan alkuperäinen matriisi  $\Lambda_{(p \times m)}$ , jossa  $p$  on havaittujen muuttujien lukumäärä ja  $m$  on Scree-kuvaajan perusteella tunnistettujen tekijöiden määrä. Pyöritysprosessia varten etsitään rotaatiomatriisi  $T_{(m \times m)}$ , joka saadaan iteratiivisella laskentamenettelyllä. Pyöritetty matriisi  $V_{(p \times m)}^{rot}$  saadaan matriisin  $\Lambda_{(p \times m)}$  avulla, yhtälön (A3) mukaisesti.

$$V_{(p \times m)}^{rot} = \Lambda_{(p \times m)} T_{(m \times m)} \quad (\text{A3})$$

Optimaalisen  $T_{(m \times m)}$  avulla saatua  $V_{(p \times m)}^{rot}$ -matriisia hyödynnetään faktoripistematriisin  $F_{(n \times m)}$  laskemiseen. [24]

### A.3 Osittaisen pienimmän neliösumman regressio

Jokaisen latenttimuuttujan selittämän varianssin laskemiseksi tarvitaan laskettu kokonaisvaihtelu yhtälön (A4) mukaisesti.

$$\text{Kokonaisvarianssi}_X = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2, \quad (\text{A4})$$

missä  $x_{ij}$  on datan  $X$  alkoi, joka vastaa  $i$ :ttä havaintoa ja  $j$ :ttä muuttujaa.  $\bar{x}_j$  on muuttujan  $j$  keskiarvo.

Kunkin latenttimuuttujan selittämä varianssi saadaan laskemalla suhteellinen osuus kokonaisvariانسsista, kuten yhtälössä (A5) esitetään.

$$\text{Selitetty vaihtelu}_X(t_j) = \frac{\text{Var}(t_j p_j^T)}{\text{Kokonaisvariانسsi}_X}, \quad (\text{A5})$$

missä  $p_j$  on matriisiin  $P_{(p \times k)}$   $j$ :s sarake. Selitetty vaihtelu $_X(t_j)$  esittää kyseisen latenttimuuttujan selittämän variانسsin, jota käytetään määrittämään, kuinka monta latenttimuuttujaa malliin otetaan mukaan.

VIP-piste lasketaan ottamalla huomioon kunkin komponentin painotus ja sen selittämä variانسsi.

Jokaisen latenttimuuttujan  $t_j$  selittämä vasteen neliösumma  $SS_j$  määritetään yhtälön (A6) mukaisesti.

$$SS_j = t_j^T t_j q_j^T q_j \quad (\text{A6})$$

missä  $t_j$  on  $j$ :n latenttimuuttujan score-vektori ja  $q_j$  on vasteen loading-matriisin sarake.

Hyödyntämällä  $SS_j$ , voidaan laskea VIP-pisteet kaavalla (A7).

$$VIP_i = \sqrt{p \cdot \frac{\sum_{j=1}^k SS_j w_{ij}^2}{\sum_{j=1}^k SS_j}} \quad (\text{A7})$$

missä  $p$  on alkuperäisten muuttujien lukumäärä,  $k$  on valittujen latenttimuuttujien lukumäärä,  $w_{ij}$  on painokerroin painomatriisissa  $W$ . Painomatriisi sisältää kullekin latenttimuuttujalleen liittyvät painotukset, jotka osallistuvat latenttimuuttujien, eli scores- $T_{(n \times k)}$ -matriisin muodostamiseen. [24]

## A.4 LDA

Tavoitefunktio  $J(w)$  pyrkii löytämään suuntavektorin  $w$ , joka toteuttaa yhtälön (A8) mukaisen ehdon.

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \quad (\text{A8})$$

Optimointia varten lasketaan matriisin  $S_W^{-1} S_B$  ominaisarvot  $\Lambda_{(p \times m)}$  ja ominaisvektorit  $V_{(p \times p)}$ . Suurimmat ominaisarvot kuvaavat merkittävimmät latenttimuuttujat ja niiden vastaavat ominaisvektorit muodostavat projektionmatriisin  $W_{(p \times k)}$ , missä  $k$  on pienennetyn avaruuden dimensioiden määrä. Yleensä  $k = c - 1$ , kun  $c$  on luokkien lukumäärä.

Yleistetty ominaisarvotehtävä on esitetty yhtälössä (A9)

$$S_w^{-1} S_B v = \lambda v \quad (\text{A9})$$

Koko datan keskiarvovektori  $m$  lasketaan yhtälöllä (A10).

$$\mu = \frac{1}{N} \sum_{i=1}^c \sum_{x \in X_i} x \quad (\text{A10})$$

missä  $N$  on havaintojen lukumäärä.

Kun kokonaiskeskiarvovektori  $\mu$  ja matriisi  $S_W^{-1}S_B$  on laskettu, valitaan ominaisvektorit, jotka vastaavat suurimpia ominaisarvoja. [24]