HYPERSPEKTRISEN KAUKOKARTOITUSDATAN ANALYSOINTIALGORITMEJÄ

Juha Pikkarainen

Pro gradu –tutkielma

Jyväskylän yliopisto

Fysiikan laitos

2003

Esipuhe

Tämä työ on tehty MATINE (Maanpuolustuksen tieteellinen neuvottelukunta) tutkimuksena yhteistyössä Ilmavoimien Esikunnan Geografisen tiedon vastuualueen (GTV) kanssa. Työ on osa Ilmavoimien soveltavan kuvausalan HYPE (Hyperspektristen kuva-aineistojen hyödyntäminen) tutkimushanketta.

Kiitän työhön liittyvistä kommenteista ja huomioista GTV:n FM Henry Ojansivua ja FM Jari Kemppiä sekä työn tarkastamisesta Jyväskylän yliopiston fysiikan laitoksen professori Matti Mannista. Lisäksi kiitän Novosat Oy:n DI Arto Vuorelaa ja VTT:n DI Eija Parmesta.

Tikkakoskella 5.2.2003

Juha Pikkarainen

Tiivistelmä

Hyperspektrinen kaukokartoitus on passiivinen havainnointijärjestelmä, jonka avulla maanpinnan kohteita pyritään tunnistamaan ja luokittelemaan käyttämällä hyväksi kohteiden heijastamaan (auringon-) säteilyyn kätkeytynyttä informaatiota. Verrattuna perinteisiin multispektrisiin järjestelmiin, hyperspektrisen datan tarkka spektrinen resoluutio mahdollistaa kohteiden tarkemman luokittelun ja lisäksi hyperspektriset järjestelmät eivät ole niin sovellusriippuvaisia.

Tutkimuksen tavoitteena oli esitellä algoritmejä, joita on käytetty tai voidaan käyttää hyperspektrisen kaukokartoitusdatan analysointiin. Tutkimuksessa suurimpana vaikeutena oli se, että analysointia käsitteleviä kirjoja on hyvin vähän. Aiheesta löytyi enimmäkseen erillisiä artikkeleja ja tutkimusraportteja, joissa kussakin käsiteltiin vain muutamaa algoritmiä kerrallaan. Yksi tavoite olikin antaa yleiskuva eri menetelmistä ja algoritmeistä, jolloin sovelluskohtaisten algoritmien valinta helpottuu. Pääkohtina olivat spektripiirteiden sisältämän redundanssin vähennysmenetelmät ja epähomogeenisten sekä puhtaiden pikseleiden tutkimiseen tarkoitetut menetelmät. Motivaationa tutkimukselle oli se, että hyperspektrisillä järjestelmillä on mahdollista passiivisesti havaita keinotekoisia naamiomateriaaleja tai ympäristöstään poikkeavia kohteita, silloinkin kun kohteiden spatiaalinen symmetria taustaa vasten on rikottu.

Tutkimuksen yhteydessä ei analysoitu omaa tai muualta hankittua dataa. Työn tuloksia tullaan jatkossa käyttämään Ilmavoimien Geografisen tiedon vastuualueen spektriaineistojen tutkimiseen.

Avainsanat: kaukokartoitus, hyperspektrinen, ilmakuvaus, passiivinen havainnointi, hahmontunnistus

Termit ja lyhenteet:

Anomalia	poikkeama, epänormaalisuus
Data	numeeriset mittausarvot
Dimensio	ulottuvuus, piirteiden määrä
ENVI	kaukokartoitusdatan käsittelyyn tarkoitettu ohjelma
Hyperspektrinen	satoja vierekkäisiä kanavia sisältävä spektri
Multispektrinen	alle parikymmentä kanavaa sisältävä spektri, jossa kanavat eivät välttämättä sijaitse vierekkäin
Redundanssi	datan ylijäämä, joka voidaan poistaa datan sisältämän informaation siitä kärsimättä
Reflektanssi	kohteen ominaisuus heijastaa sähkömagneettista säteilyä
Piirre	objektia karakterisoiva muuttuja
Piirrevektori	vektori jonka komponentteinä ovat piirteet
Spektri	sähkömagneettisen säteilyn intensiteetin tai reflektanssin esitysmuoto aallonpituuden funktiona
Unmixing	epähomogeenisten pikseleiden tutkimiseen tarkoitettu prosessi, jossa pyritään selvittämään eri materiaalien osuudet havaituissa spektreissä

Sisällysluettelo:

Esipuhe ii

Tiivistelmä iii

Termit ja lyhenteet: iv

1. Johdanto 1

2. Spektrien syntymekanismit 3

3. Spektrin esittäminen piirreavaruudessa 4

3.1 Moniulotteisen piirreavaruuden visualisoinnista 5

3.2 Endmember 5

3.3 Todennäköisyystiheys- ja diskriminanttifunktiot 6

4. Datan dimension pienentäminen 7

4.1 Piirteiden valinta vs. ekstraktointi 8

4.2 Suuridimensioisen datan ominaisuuksista 8

4.2.1 Geometriset ja statistiset ominaisuudet 8

4.3 Perinteisten menetelmien soveltaminen hyperspektriseen dataan 15

4.4 Piirteiden valintamenetelmiä 16

4.4.1 Huomioita valintamenetelmistä 17

4.5 Piirteiden ekstraktointimenetelmiä 18

4.5.1 PCA 18

4.5.2 Minimum noise fraction (MNF) 21

4.5.3 Independent Component Analysis (ICA) 21

4.5.4 Empiirisiin havaintoihin perustuvat menetelmät 22

4.5.5 Discriminate Analysis Feature Extraction (DAFE) 23

4.5.6 Decision Boundary Feature Extraction (DBFE) 25

4.5.7 Discriminant Feature Extraction (DFE) 27

4.5.8 Projection Pursuit (PP) 27

4.5.9 Wavelets 33

5. Spektrin unmixing menetelmät 35

5.1 Lineaarinen vs. epälineaarinen malli 35

5.2 Lineaariset unmixing menetelmät 37

5.2.1 Multiple Endmember Spectral Mixture Analysis (MESMA) 37

5.2.2 Matched Filtering (MF) / Constrained Energy Minimization (CEM) 38

5.2.3 Target-constrained Interference-minimized Filter (TCIMF) 41

5.2.4 Mixture Tuned Matched Filtering (MTMF) 41

5.2.5 Foreground–Background Analysis (FBA) 42

5.2.6 Orthogonal Subspace Projection (OSP) 43

5.2.7 OSP vs. CEM 46
5.2.8 RX 48
5.3 Epälineaariset unmixing menetelmät 48

6. Puhtaan pikselin luokittelumenetelmät 51
6.1 Tilastolliset luokittelijat 51
6.1.1 Maximum likelihood (ML) 51
6.1.2 Statistiset etäisyysmitat 52
6.1.2 Ohjaamattomia luokittelumenetelmiä 55
6.1.3 Ei-parametrilliset tiheysfunktiot 58
6.1.4 Neuroverkot 60
6.2 Kahden spektrin vertailuun perustuvat menetelmät 62
6.2.1 Spectral Angle Mapper (SAM) 62
6.2.2 Spectral Correlation Mapper (SCM) 63
6.2.3 Chi-toiseen testi 67
6.2.4 SES 67
6.2.5 Esimerkki MSAM, SCM, χ^2 ja SES luokittelusta. 68
6.2.6 Spectral Feature Fitting (SFF) 68
6.2.7 Binary Encoding (BE) 71

7. Virheanalyysi 72
7.1 Virhematriisi 72
7.1.1 Tuottajan tarkkuus vs. käyttäjän tarkkuus 72
7.2 Kappa kerroin 73
7.3 Testijoukon valinta 73
7.4 ROC käyrä 74

8. Yhteenveto 76

Lähteet: 79

1. Johdanto

Kaukokartoituksella pyritään selvittämään mitä materiaalia tai tyyppiä (esim. metsä, pelto, panssarivaunu, ympäristöstään poikkeava kohde, saastunut metsä, jne.) havainnoitava kohde on, siitä saatavan sähkömagneettisen säteilyn (fotonien) avulla. Säteily voi olla peräisin ulkopuolisesta säteilylähteestä (esim. aurinko), jonka kohde heijastaa tai se on kohteen emittoimaa termistä IP-säteilyä.

Spektriksi kutsutaan esitystä, jossa säteilyn intensiteetti tai kohteen heijastuskyky on ilmaistu säteilyn aallonpituuden (tai aaltoluvun, energian, taajuuden) funktiona. Vaikka kohteesta tuleva säteilyn spektri olisikin jatkuvaa, ei sitä voida mitata rajattoman tarkalla spektrisellä resoluutiolla. Mittauksessa tietyn levyiseen aallonpituusalueeseen kuuluvat fotonit rekisteröidään samaan spektrikanavaan, vaikka niiden aallonpituudet poikkeaisivatkin toisistaan. Perinteisesti kaukokartoitusta on tehty multispektrisillä sensoreilla, joille on tyypillistä että ne rekisteröivät säteilyä muutamalla laajalla kanavalla, jotka eivät ole toisissaan kiinni. Mittaustekniikan kehittyminen on tuonut kaukokartoitukseen hyperspektrisiä sensoreita, joilla on muutama sata kapeaa kanavaa, jotka ovat lisäksi toisissaan kiinni. Hyperspektristen sensoreiden parempi spektrinen resoluutio mahdollistaa sen, että mitattuja kohteita pystytään identifioimaan tarkemmin spektristen ominaisuuksien perusteella, suhteessa multispektriseen dataan. Termi ”hyperspektrinen” ei sinällään kerro käytettävästä aallonpituusalueesta mitään. Useimmiten termillä viitataan auringon spektrin (380-2500 nm) alueella tapahtuvaan kaukokartoitukseen. Hyperspektrisissä sensoreissa on usein myös mukana muutamia kanavia termisen IP-säteilyn alueelta.

Hyperspektrinen data ei automaattisesti kerro mitä materiaalia mitatut kohteet ovat. On luotava algoritmeistä koostuva analysointiketju, joka ottaa sisään mittausdataa ja ulostulona kertoo mitä on mitattu. Ketjun alkupäähän kuuluu mm. raakadatan (volttiarvojen) muuttaminen säteilysuureeksi, häiriöiden (geometriset vääristymät, ilmakehän säteilyä absorboiva ja sirottava vaikutus) poistaminen datasta ja säteilyn normittaminen reflektanssiksi. Tässä tutkimuksessa esitellään algoritmejä, joilla analysointiketjua on mahdollista jatkaa eteenpäin eli kuinka esikäsitellystä spektristä selvitetään että mitä on mitattu. Ensisijainen tarkoitus on kertoa algoritmeistä, joita on yleisesti käytetty tai voidaan käyttää datan analysoinnissa. Taustalla on ajatus siitä mitä

algoritmejä voitaisiin käyttää sotilaallisesti mielenkiintoisten kohteiden havainnointiin ja luokitteluun.

Nykyisillä pankromaattisilla tiedustelukuvausjärjestelmillä on tarkka spatiaalinen resoluutio, mikä mahdollistaa naamioimattomien kohteiden havaitsemisen. Pankromaattisilta järjestelmiltä on kuitenkin helppo suojautua rikkomalla kohteiden symmetria naamioverkoilla tai vastaavilla. Sotilassovelluksissa hyperspektrisen datan mahdollisuudet liittyvät erityisesti naamioitujen kohteiden havainnointiin, koska havainnointi perustuu kohteen spektrisiin ominaisuuksiin, ei niinkään spatiaalisiin muotoihin.

Luvussa 4 käydään läpi menetelmiä, joilla hyperspektrisestä datasta voidaan poimia parhaimmat piirteet itse luokittelua varten. Luvussa 5 käsitellään epähomogeenisten, ei vain yhtä materiaalia sisältävien pikseleiden analysointiin käytettyjä menetelmiä. Nämä menetelmät ovat erityisen käyttökelpoisia pienten kohteiden havaitsemissovelluksissa. Luvussa 6 käsitellään homogeenisten eli ns. puhtaiden pikseleiden identifioimiseen soveltuvia luokittelumenetelmiä ja luvussa 7 luokittelutulosten tarkkuuden arviointia. Tutkimuksessa ei käsitellä algoritmejä, jotka huomioisivat naapuripikseleiden välisen korrelaation. Tälläiset algoritmit olettavat tyypillisesti, että kaksi vierekkäistä pikseliä ovat todennäköisesti samoja materiaaleja. Oletus toiminee hyvin tilanteissa, jossa luokitellaan maastoa esim. eri metsätyyppeihin ja peltoihin, koska nämä luokat muodostavat normaalisti laajoja alueita. Sotilassovelluksissa tilanne on toisenlainen, koska ei voida olettaa, että esim. panssarivaunun vieressä on todennäköisesti toinen panssarivaunu.

Monet hyperpektriset sovellukset ovat keskittyneet kasvillisuuden kaukokartoitukseen. Kasvillisuuden luokittelusta kertovista raporteista on paljon omaksuttavaa myös sotilassovelluksiin. Tästä voi herätä kysymys ”miksi”. Ensinnäkin kasvillisuus on usein maastoon sijoitettujen sotilaskohteiden tausta. Toiseksi kasvillisuuden ja keinotekoisten rakennusmateriaalien heijastusspektreistä puuttuvat voimakkaat absorptio-ominaisuuksien aiheuttamat piikit, jolloin kohteiden luokittelu perustuu etupäässä samoihin algoritmeihin. Ja kolmanneksi, kasvillisuuden spektrit eri kasvilajien välillä ovat hyvin samanlaisia suhteessa keinotekoisiin materiaaleihin (esim. naamiomaaleihin on hankala saada kasvillisuuden spektripiirteitä [Igarss 2002]). Tästä syystä voidaan olettaa, että jos eri kasvilajit onnistutaan erottamaan toisistaan, on hyvin todennäköistä että kasvillisuus voidaan erottaa myös keinotekoisista naamiomateriaaleista.

2. Spektrien syntymekanismit

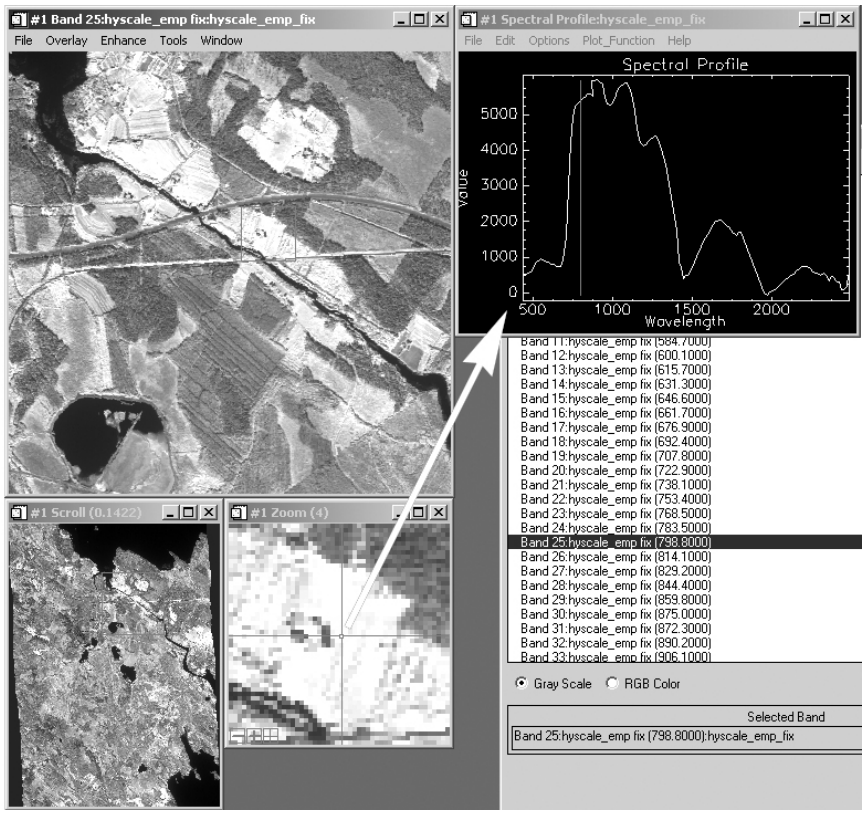
Materiaalien erottelu toisistaan kaukokartoitusmenetelmin perustuu siihen, että kullakin materiaalilla on oma uniikki spektri, ns. spektrinen sormenjälki. Se miksi eri materiaalit heijastavat ja absorboivat eri tavoin sähkömagneettista säteilyä, johtuu materiaalien erilaisista atomi- ja molekyylirakenteista. Atomeissa (elektronitilat) ja molekyyleissä (elektroni-, värähtely- ja rotaatiotilat) on joukko erilaisia kvanttitiloja, joiden väliset energiaerot ovat diskreettejä. Mikäli säteilyn fotonin energia on sama kuin kahden kvanttitilan energiaero, voi fotoni virittää systeemin ylemmälle energiatilalle, jolloin alkuperäinen fotoni häviää. Kahden eri materiaalin spektrinen erilaisuus ei perustu pelkästään siihen, että niistä löytyisi energeettisesti eri kvanttitilaerot. Tiloilla on myös tietty virittymistodennäköisyys eli toinen aine tarvitsee enemmän fotoneita yhtä viritystä kohden kuin toinen, vaikka vaadittavilla fotoneilla olisikin sama energia.

Mitä yksinkertaisimmista molekyyleistä materiaali on rakentunut sitä yksinkertaisempia ovat niiden spektrit. Tämä johtuu siitä, että molekyylien symmetria lisää kvanttitilojen degeneraatioastetta eli eri tilat asettuvat samoille energiatasoille. Esimerkiksi kaasutilaisen hiilidioksidin absorptiospektri (alueella 2000 - 15000 nm) on hieman monimutkaisempi kuin vastaava hiilimonoksidin spektri. Syynä on se, että hiilidioksidi on rakentunut kolmesta atomista, mikä mahdollistaa monipuolisimmat atomien sidosten väliset värähtelymoodit kuin mitä kaksiatomisella hiilimonoksidilla on.

Tyypillinen kaukokartoituksella mitattava kohde on esim. kasvillisuus, joka on rakentunut monista erilaisista orgaanisista molekyyleista, kuten mm. selluloosasta, sokereista ja vahoista. Vaikka edellämainitut molekyylit ovat rakentuneet vain muutamista eri alkuaineista, ovat molekyylit hyvin suuria ja kompleksisen muotoisia (epäsymmetrisiä), jolloin värähtelytilojen määrä kasvaa hyvin suureksi. Tällöin värähtelytilojen välisiä siirtymiä vastaavat absorptiopiikit sijaitsevat niin tiheästi, että spektreistä tulee jatkuvia, toisin sanoen niistä puuttuvat voimakkaat, havaittavat absorptiopiikit. (Myös useimpien keinotekoisten materiaalien spektrit ovat jatkuvia.) Tämän vuoksi kaukokartoituksella saatavien spektrien analysointi eroaa oleellisesti siitä mihin fysiikan ja kemia alaan liittyvissä spektroskooppisissa sovelluksissa on totuttu. Kaukokartoitusdatan tehokas ja luotettava analysointi vaatii matemaattisia menetelmiä, jotka perustuvat etupäässä spektrikäyrien muotojen hyödyntämiseen.

3. Spektrin esittäminen piirreavaruudessa

Spektrinen kuvadata voidaan esittää kolmessa eri avaruudessa: kuva-, spektri- ja piirreavaruudessa, joista on esimerkit kuvissa 3.1 ja 3.2. Hyperspektrisen datan kohdalla, kuva-avaruus on tarkoitettu lähinnä kaukokartoitetun alueen kvalitatiiviseen visualisointiin ja analysointitulosten esittämiseen. Kuva-avaruuden haittapuoli on se, että data voidaan maksimissaan esittää vain kolmen komponentin (punainen, vihreä ja sininen) avulla, jolloin datan spektriominaisuuksista käytetään hyväksi vain murto-osa. Spektriavaruus on käyttökelpoinen esim. silloin, kun halutaan visualisoida yksittäisten pikselien spektrejä.



Kuva 3.1: Kuva- ja spektriavaruus.

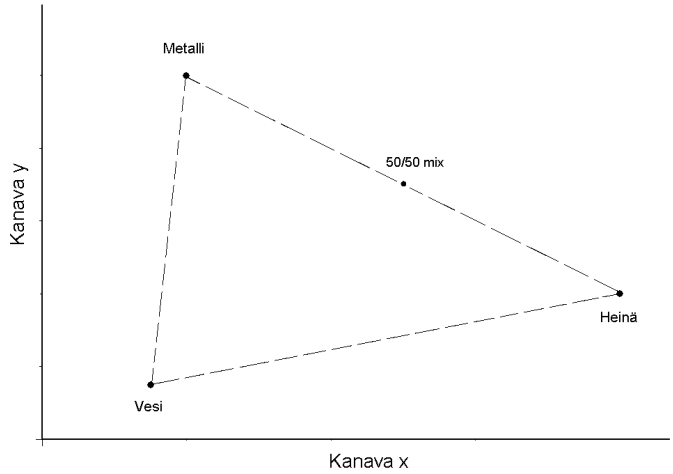
Piirre on kohdetta karakterisoiva muuttuja. Suuri osa kvantitatiivisista analysointimenetelmistä perustuvat piirreavaruuteen. Piirreavaruus on suorakulmainen koordinaatisto, jonka akselit vastaavat spektrikanavia tai spektreistä ekstraktoituja piirteitä. Spektri kuvautuu piirreavaruuteen pisteeksi, jonka koordinaatit riippuvat kunkin kanavan amplitudista. Se miksi on parempi puhua piirteistä kuin kanavista (piirre- / kanava-avaruus), johtuu siitä että datan käsittelyssä on usein hyödyllistä tehdä piirreavaruuteen koordinaattimuunnoksia, jolloin alkuperäisten kanavien merkitys häviää. Nämä koordinaattimuunnokset liittyvät datan oleellisen ja epäoleellisen informaation erottamiseen, jota käsitellään luvussa 4.

3.1 Moniulotteisen piirreavaruuden visualisoinnista

Kolmiulotteisuuteen asti piirreavaruus voidaan visualisoida perinteisin menetelmin, mutta useampiulotteisen avaruuden eli ns. hyperavaruuden kuvaamiseen tarvitaan dynaamisempia menetelmiä. Yksi tälläinen on kaukokartoitusdatan käsittelyyn tarkoitetun ENVI ohjelman ”n-D Visualizer”, jolla voidaan interaktiivisesti pyöritellä moniulotteista avaruutta ja tarkastella miltä datapilvi näyttää eri suunnista katsottuna. [ENVI]

3.2 Endmember

Kuvan 3.2 piirreavaruuteen on piirretty kolmen eri materiaalin kuvitteelliset datapisteet. Mikäli käytössä on spektrejä kohteista, jotka sisältävät vain näitä kolmea materiaalia (vettä, heinää ja metallia), niin datapisteet tulevat sijaitsemaan kolmion (kuva 3.2) sisäpuolella. Esim. puolet metallia ja puolet heinää sisältävän kohteen spektri kuvautuu metalli-heinä janan keskipisteeseen.



Kuva 3.2: Kaksiulotteinen piirreavaruus.

N-ulotteisen datapilven kulmapisteistä käytetään nimitystä ”endmember”. Nämä endmemberit edustavat eri materiaaleista mitattuja puhtaita spektrejä, joiden sekoituksia ovat muut kuvadatan epähomogeeniset pikselit. Näitä sekoittuneita pikseleitä kutsutaan mikseleiksi. Endmember termillä voidaan myös viitata itse fyysiseen maakomponenttiin (metalli, heinä, vesi, jne.). [Igarss 2002]

Pixel Purity Index (PPI)

Pixel Purity Index on tarkoitettu spektrisesti puhtaimpien (vähiten sekoittuneimpien) pikseleiden etsimiseksi. PPI:ssä piirreavaruuden datapisteet projisoidaan satunnaisille vektoreille, jotka kulkevat datapilven keskiarvon kautta. Kullakin projisoinnilla vektoreiden ääripäihin projisoituneet pikselit merkitään ylös. Eniten merkintöjä (suurimmat indeksit) saaneita pikseleitä pidetään puhtaimpina, joita sovelluksesta riippuen voidaan käyttää luokittelun opetusnäytteinä tai laskettaessa endmembereiden osuuksia mikseleissä. [ENVI], [XING]

3.3 Todennäköisyystiheys- ja diskriminanttifunktiot

Tilastollisessa hahmontunnistuksessa tehtävänä on jakaa piirreavaruus osiin, jotka kukin edustavat omaa luokkaa. Avaruuden osiin jakavia rajoja kutsutaan päätösrajoiksi tai päätöspinnoiksi (eng. decision boundary, decision surface), jotka voidaan määrittää todennäköisyystiheys- tai diskriminanttifunktioilla. Molemmilla funktiotyypeillä piirreavaruuden piste luokittuu siihen luokkaan, jonka luokkakohtainen todennäköisyystiheyden tai diskriminanttifunktion arvo on suurin. Keskeisin ero funktiotyypeillä on se, että toinen tekee luokituspäätöksen todennäköisyystiheyden perusteella, toinen diskriminanttifunktiolla. [Webb]

4. Datan dimension pienentäminen

Hyperspektrisen datan suuri kanavamäärä mahdollistaa sen, että mitattuja kohteita voidaan luokitella tarkemmin ja alipikselimenetelmien (unmixing) käyttömahdollisuudet paranevat suhteessa multispektriseen dataan. Lisäksi suuri spektrinen resoluutio mahdollistaa ilmakehäparametrien, esim. vesihöyryn määrän, laskemisen suoraan datasta pikselikohtaisesti (ei käsitellä tässä tutkimuksessa). Suuri dimensionaalisuus (kanavamäärä) tuo kuitenkin mukanaan ongelmia, joiden vuoksi dimensiota täytyy vähentää ennen varsinaista luokittelua. Tärkeimmät syyt dimension pienentämiseen on pitää analysoinnin laskenta-aika kurissa ja nk. Hughesin ilmiö, joka liittyy ohjatun luokittelun luokittelutarkkuuteen.

Dimension pienentäminen ei automaattisesti merkitse luokittelutarkkuuden huonontumista, koska hyperspektrinen data sisältää spektristä redundanssia eli jonkin kanavan sisältämä tieto voidaan täysin tai osittain ennustaa toisen kanavan sisältämällä tiedolla [Richards, pp. 315-316]. Alkuperäinen data voidaan, ainakin käsitteellisesti, jakaa informatiivisiin ja redundantteihin piirrevektoreihin. Jos merkitään informatiivisten piirrevektoreiden määrää symbolilla I ja redundanttisten piirrevektoreiden määrää symbolilla R ja lisäksi vaaditaan että nämä vektorit ovat lineaarisesti riippumattomia, niin pätee että

$$A = I + R,$$

jossa A on datan alkuperäinen dimensio. Tällöin datan ns. sisäinen dimensionaalisuus vastaa lukua I. [Landgrebe 1/1993, pp. 36-37]

Pääongelma dimension pienentämisessä on se, että kuinka informatiiviset piirrevektorit kaivetaan datasta ulos. Lisäksi on päätettävä, että otetaanko luokitteluun mukaan kaikki informatiiviset piirrevektorit vai riittääkö jokin alijoukko niistä.

4.1 Piirteiden valinta vs. ekstraktointi

Dimension pienentäminen voidaan tehdä kahdella tavalla: piirteiden valinnalla ja piirteiden ekstraktoinnilla. Valinnassa tehtävä on seuraava: valitse annetusta piirrejoukosta D alijoukko d_v, joka johtaa suurimpaan luokittelutarkkuuteen. Piirteiden ekstraktoinnissa piirrejoukko D muutetaan (lineaarisilla tai epälineaarisilla) funktioilla alijoukoksi d_e. Näiden menetelmien eräs vaikutus käytäntöön on se, että valinnassa piirteiden fysikaalinen merkitys säilyy, kun taas ekstraktoinnista saaduilla uusilla piirteillä ei välttämättä ole selvää fysikaalista merkitystä. Fyysisen merkittävyyden säilytys on tärkeää esim. silloin kun eri sensoreiden tuottamaa dataa halutaan yhdistellä. Ekstraktoinnin etuna on että menetelmällä voidaan saada aikaan parempia piirteitä. [Jain 2000]

4.2 Suuridimensioisen datan ominaisuuksista

Intuitiomme dimensioista perustuu pitkälti kaksi- tai kolmiulotteisen euklidisen avaruuden ominaisuuksiin. Esimerkiksi piirreavaruuden analysointimenetelmien graafinen havainnollistaminen tehdään usein kaksi- tai kolmiulotteisessa avaruudessa, vaikka hyperspektrisessä datassa dimensioita on huomattavasti enemmän. Seuraavissa kappaleissa esitellään suuridimensioisen avaruuden ominaisuuksia ja niiden vaikutuksia luokitteluun. Suurin osa näistä ominaisuuksista ei sovi yhteen niitten kokemusten kanssa, joita on opittu perinteisestä kolmidimensioisesta avaruudesta. [Landgrebe 4/1995, pp. 5]

4.2.1 Geometriset ja statistiset ominaisuudet

Yksi suuridimensioisen avaruuden ominaisuus on se, että dimension kasvaessa hyperavaruuden suhteellinen tilavuus keskittyy yhä enemmän avaruuden laidoille [Landgrebe 4/1995, pp. 6]. Tätä on havainnollistettu seuraavassa.

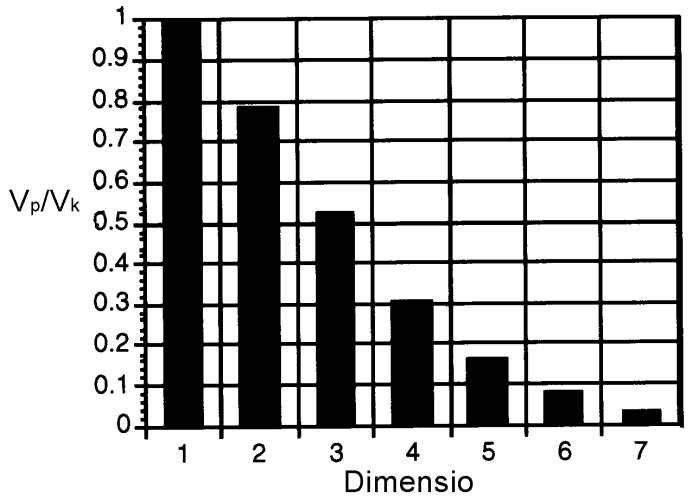
Hyperpallon tilavuus saadaan yhtälöllä [Kendall]:

$$V_p(r) = \frac{2r^d}{d} \frac{\pi^{\frac{1}{2}d}}{\Gamma\left(\frac{1}{2}d\right)} \tag{4.1}$$

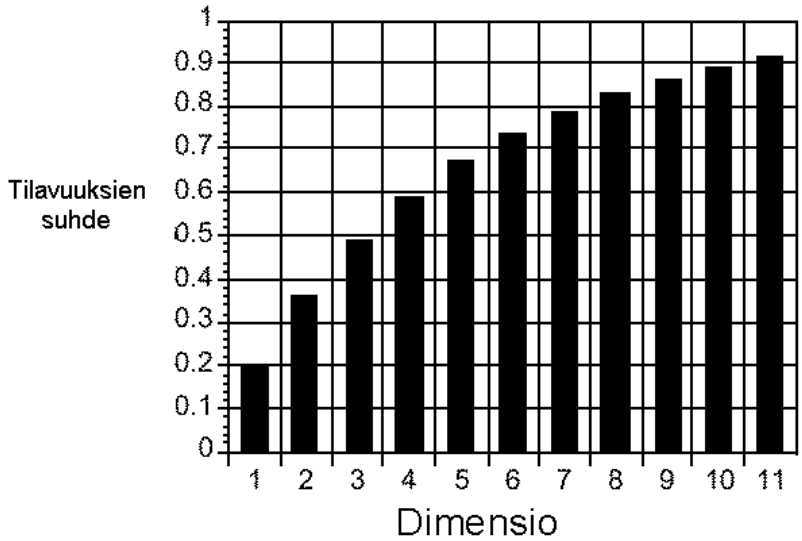
ja hyperkuution tilavuus yhtälöllä:

$$V_k(r) = (2r)^d \,, \tag{4.2}$$

joissa d on dimensio, r pallon säde tai kuution tapauksessa keskipisteen ja reunan välinen etäisyys. Γ on gammafunktio. Yhtälöjä (4.1) ja (4.2) käyttäen on kuvaan 4.1 piirretty hyperpallon ja -kuution tilavuuksien suhde tapauksessa, jossa hyperpallo juuri ja juuri mahtuu hyperkuution sisään. Kuvaan 4.2 on piirretty hyperpallon ja sen ulkokuoren (paksuus viidesosa pallon säteestä) tilavuuksien suhde.



Kuva 4.1 [Landgrebe 4/1995, pp. 6]: Hyperpallon ja -kuution tilavuuksien suhde dimension funktiona. Tilavuuksien suhde on riippumaton suureesta r.



Kuva 4.2 [Landgrebe 4/1995, pp. 7]: Hyperpallon ulkokuoren (paksuus viidesosa pallon säteestä) ja itse pallon tilavuuksien suhde dimension funktiona.

Se että suuridimensioisen avaruuden tilavuus on keskittynyt laidoille, vaikuttaa kahdella tavalla dataan. Ensinnäkin suuridimensioinen avaruus on enimmäkseen tyhjä, mikä merkitsee sitä että datalla on yleensä mataladimensioisempi sisäinen rakenne. Tämän vuoksi dataa voidaan projisoida mataladimensioisempaan aliavaruuteen, ilman että samalla menetetään merkittävää tietoa luokittelu varten. Eräs ongelma tässä on se, että parhaan aliavaruuden löytäminen riippuu sovelluksesta. Toinen seuraamus on se, että normaali- ja tasaisesti jakautuneella datalla on taipumus kerääntymä avaruuden laidoille, mikä vaikeuttaa tiheysfunktioiden arviointia.

Havainnollistaakseen datan käyttäytymistä, Landgrebe et al. [Landgrebe 4/1995, pp. 8-11] generoivat normaali- ja tasaisesti jakautuneet monimuuttuja-aineistot. Näiden aineistojen muuttujat ovat riippumattomasti ja identtisesti jakautuneita otoksia vastaavista jakautumista N(0,1) ja Tas(-1,1). Aineistoista laskettiin satunnaisvektoreiden pituus (datapisteen etäisyys origosta) ja sen neliö yhtälöillä:

$$r = \sqrt{\sum_{i=1}^{d} x_i^2} \tag{4.3}$$

$$R = r^2 \tag{4.4}$$

Pituudesta ja sen neliöstä laskettiin edelleen odotusarvo ja keskihajonta. Tulokset on esitetty taulukoissa 4.1a ja 4.2b.

Taulukko 4.1a: Normaalijakautuneiden (N(0,1)) satunnaisvektoreiden statistiikkaa. [Landgrebe 4/1995, pp. 8-11]

Normaalijakautuneet satunnaisvektorit				
Dimensio	Pituuden odotusarvo	Pituuden keskihajonta	Pituuden neliön odotusarvo	Pituuden neliön keskihajonta
1	0.7736	0.5737	1.0168	1.4017
10	3.1026	0.7042	9.8697	4.5328
220	14.8497	0.7129	220.3732	20.7862

Taulukko 4.1b: Tasaisesti jakautuneiden ((Tas(-1,1)) satunnaisvektoreiden statistiikkaa. [Landgrebe 4/1995, pp. 8-11]

Tasajakautuneet satunnaisvektorit				
Dimensio	Pituuden odotusarvo	Pituuden keskihajonta	Pituuden neliön odotusarvo	Pituuden neliön keskihajonta
1	0.5041	0.2887	0.3277	0.2883
10	1.8010	0.2678	3.3444	0.9390
220	8.5488	0.2505	73.3698	4.3854

Tuloksista näkee, että dimension kasvaessa satunnaisvektoreiden pituus kasvaa. Tämä kertoo siitä, että datalla on taipumus kerääntyä avaruuden ulkokuoreen ja kuinka kuoren etäisyys origosta kasvaa dimension kasvaessa.

<u>Lävistäjät ja koordinaattiakselit</u>

Piirreavaruuden diagonaalivektorin ja euklidisen koordinaattiakselin välinen kulma saadaan yhtälöllä [Landgrebe 4/1995, pp. 11]:

$$\cos(\theta_d) = \pm\frac{1}{\sqrt{d}}, \tag{4.5}$$

josta näkee että dimension kasvaessa kulma lähestyy 90º. Tulos on merkittävä, koska datajoukon projisointi diagonaalille, ts. keskiarvoistamalla piirteet, voi tuhota datan sisältämän informaation. Kvantitatiivisesti tämän voi selittää seuraavasti. Olkoon a_{diag} mielivaltainen diagonaali ja ac_i i:s koordinaattiakseli. Datapiste P d-avaruudessa voidaan esittää muodossa:

$$P = \sum_{i=1}^{d} \alpha_i ac_i \tag{4.6}$$

ja P:n projektio diagonaalille a_{diag}, P_{diag}, yhtälöllä:

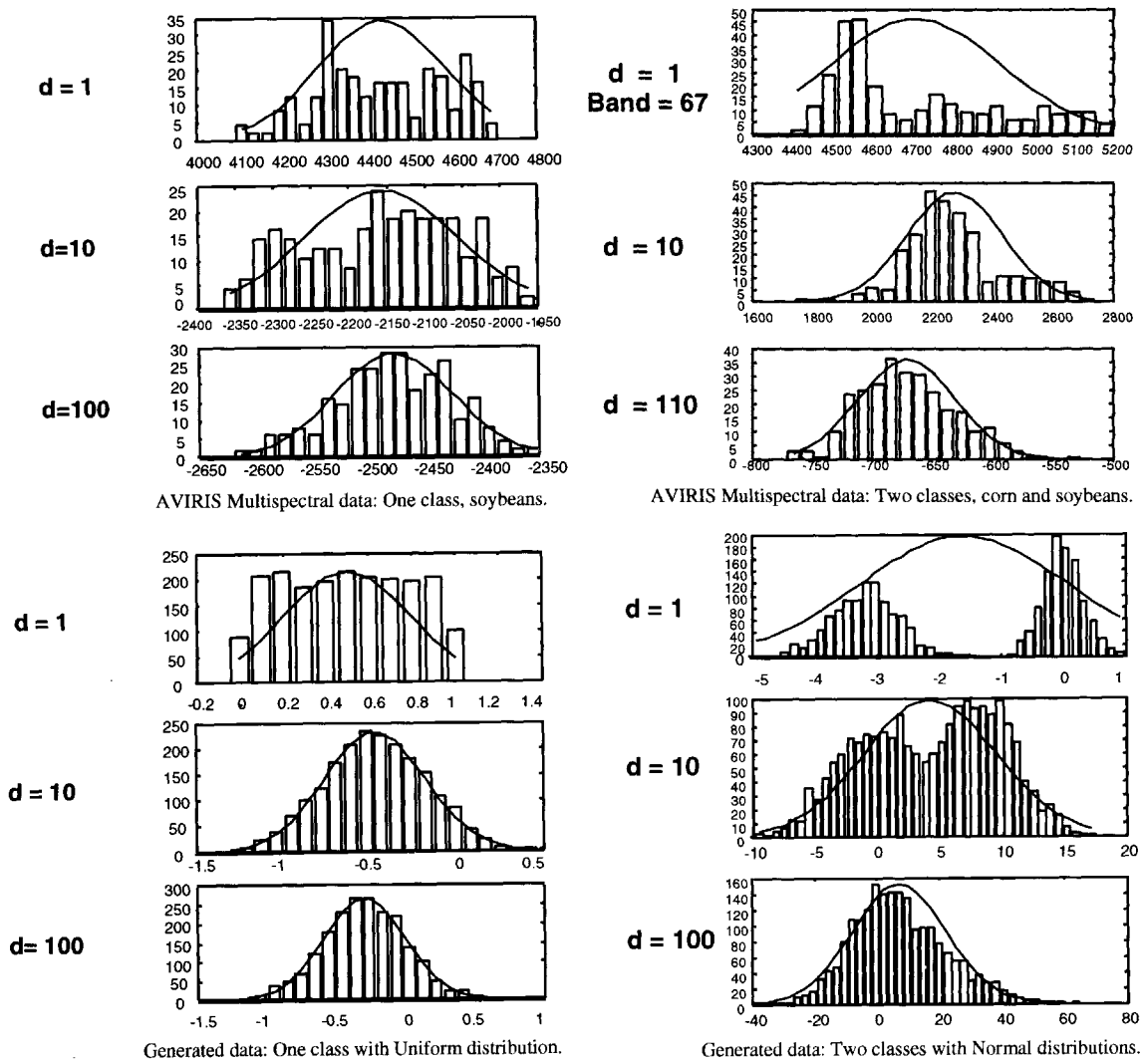
$$P_{diag} = (P^T a_{diag}) a_{diag} = \sum_{i=1}^{d} \alpha_i (ac_i^T a_d) a_d \tag{4.7}$$

Kun d kasvaa, niin $ac_i^T a_{diag} \rightarrow 0$, jolloin $P_{diag} \rightarrow 0$. Eli suuridimensioisessa avaruudessa P projisoituu nolla koordinaattiin, mikä merkitsee alkuperäisen sijaintitiedon menettämistä d dimensioisessa avaruudessa. [Landgrebe 4/1995, pp. 11-12]

Lineaariset projektiot suuri- ja mataladimensioisten avaruuksien välillä

On todistettu [Diaconis] [Hall], että dimension lähestyessä äärettömyyttä, mataladimensioiset lineaariset projektiot lähestyvät normaalimallia. Tässä tapauksessa normaalimalli tarkoittaa normaalijakautumaa tai kombinaatiota normaalijakautumista. Useat kokeet sekä generoidulla, että reaalisella datalla havainnollistavat tätä [Landgrebe 4/1995, pp. 13]. Idea kokeissa on projisoida suuridimensioinen data yksidimensioiseen aliavaruuteen.

Eräässä koejärjestelyssä [Landgrebe 4/1995, pp. 13-16] asiaa tutkittiin generoidulla (normaali- ja tasajakautuneella) ja hyperspektrisellä AVIRIS datalla. Dataa projisoitiin satunnaisella normaalivektorilla yksi-, kymmenen- ja satadimensioisesta lähtöavaruudesta yksidimensioiseen aliavaruuteen. Aineistot olivat: 1) generoitu data, yksi tasajakautunut luokka, 2) AVIRIS data, yksi luokka (soija) 3) generoitu data, kaksi tasajakautunutta luokkaa ja 4) AVIRIS data, kaksi luokkaa (maissi ja soija). Kokeen tulokset ovat kuvassa 4.3.



Kuva 4.3 [Landgrebe 4/1995, pp. 13-16]: Erityyppisten lähtöaineistojen lineaarinen projisointi d-dimensioisesta avaruudesta yksidimensioiseen aliavaruuteen. Histogrammit kertovat kuinka projisoidut datat ovat jakautuneet yksidimensioiselle piirrevektorille (x-akselille). Histogrammeihin on sovitettu normaalijakautuneet tiheysfunktiot.

Yhden luokan tapauksista nähdään että alkuperäisen datan dimension kasvaessa projisoitu data lähestyy normaalijakaumaa. Tulos osoittaa, että projisoidun mataladimensioisen datan voidaan olettaa olevan kombinaatio normaalijakautumista. Kahta eri luokkaa sisältävien aineistojen tulokset osoittavat kuinka projisointi suuridimensioisesta lähtöavaruudesta luhistaa luokat yhdeksi normaalijakautumaksi. Tämä merkitsee informaatiohukkaa, koska eri luokkien erottaminen toisistaan tulee mahdottomaksi. [Landgrebe 4/1995, pp. 13-16]

On huomioitava että edellä tehdyissä kahta eri luokkaa sisältävien aineistojen projisointi perustui satunnaiseen normaalivektoriin, jonka käyttö johti kahden luokan erottuvuuden menettämiseen. Tämä ei tarkoita sitä, että kaikissa projektioissa kävisi samoin. Olennaista on se, että piirteiden projisointi mataladimensioiseen avaruuteen tehdään niin, että erotuskyky luokkien välillä säilyy.

Tärkeä seikka yhtä luokkaa sisältävän aineiston projisointituloksista liittyy oletukseen jakautumasta. Tietyt datan ekstraktointi- ja luokittelumenetelmät nimittäin olettavat, että aineisto on normaalijakautunutta. Kuitenkaan mikään ei takaa, että data olisi alkuperäisessä dimensiossa normaalijakautunutta [Landgrebe 4/1995, pp. 29]. Koska lineaarisesti projisoitu mataladimensioinen data on lähempänä normaalijakautumaa, kun alkuperäisen avaruuden dimensio kasvaa, niin oletettavasti tällöin normaalijakautumaan perustuvat luokittelu- tai ekstraktointimenetelmät toimivat paremmin mataladimensioisessa avaruudessa.

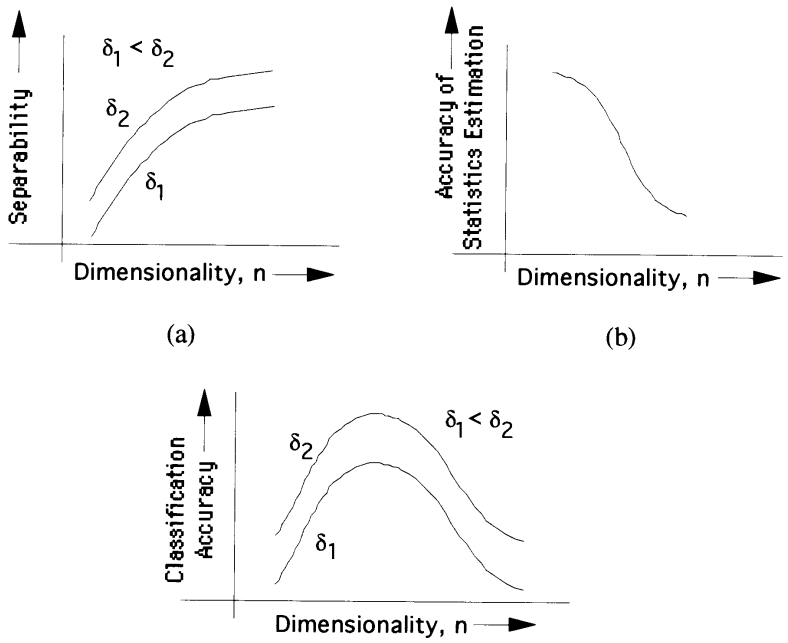
Hughesin ilmiö

Edellä perusteltiin sitä miksi datan dimensionaalisuutta voidaan vähentää ilman merkittävää informaatiohukkaa. Lisäksi kerrottiin siitä kuinka tietyt projisoinnit suuri- ja mataladimensioisten avaruuksien välillä voivat toisaalta hukata informaatiota. Hughesin ilmiö liittyy negatiivisellä tavalla suuridimensioisen datan ohjattuun piirteiden ekstraktointi- ja luokittelumenetelmiin. Ilmiön vaikutus on myös verrattavissa ohjaamattomien menetelmien kohdalla.

Kun piirreavaruuden dimensionaalisuutta lisätään, eri spektriluokkien väliset erot pääsääntöisesti lisääntyvät, jolloin mahdollisuus tarkempiin luokitteluihin kasvaa. Käytännössä ohjatussa luokittelussa opetusnäytteiden rajallinen määrä vaikuttaa negatiivisesti luokittelutarkkuuteen. Kun opetusnäytteiden määrä pidetään vakiona, aluksi dimensionaalisuuden kasvattaminen lisää luokittelutarkkuutta, mutta dimension edelleen kasvaessa luokittelutarkkuus kääntyy laskuun (kts. kuva 4.4). Tulos on yleisesti tunnettu tilastollisessa monimuuttuja-analyysissa ja se tunnetaan nimellä Hughesin ilmiö.

Selitys Hughesin ilmiöön on seuraava. Opetusnäytteiden tarkoituksena on kertoa luokittelijalle luokan tilastollinen käyttäytyminen. Otetaan esimerkiksi luokittelijat, jotka olettavat että aineisto on normaalijakautunutta. Tämä jakauma on täysin tunnettu, kun keskiarvovektori ja kovarianssimatriisi tunnetaan. Näitä parametrejä ei kuitenkaan tunneta yleisesti, vaan ne täytyy arvioida opetusnäytteiden perusteella. Dimension kasvaessa parametrien tai muiden luokkaa karakterisoivien tekijöiden arviointi tulee epätarkemmaksi mikäli opetusnäytteiden määrä pidetään vakiona. Ei-parametrillisten menetelmien kohdalla pienen opetusjoukon vaikutus on vielä suurempi arvioitaessa luokan statistiikkaa verrattuna parametrillisiin menetelmiin. Kuvaan 4.4 on havainnollistettu Hughesin ilmiö. [Landgrebe 4/1995], [Landgrebe 1998a], [Landgrebe 1998b], [Jain 2000] Ilmiö vaikuttaa

myös suoraan tilastollisiin piirteiden vähennysmenetelmiin, ei vain luokittelumenetelmiin. Käytännössä ei ole mahdollista kerätä riittävän paljon opetusnäytteitä, joilla vältettäisiin Hughesin ilmiö.



Kuva 4.4 [Landgrebe 5/1998, pp. 9]: Hughesin ilmiön kvalitatiivinen havainnollistaminen.

4.3 Perinteisten menetelmien soveltaminen hyperspektriseen dataan

Perinteiset multispektriseen dataan sovelletut piirteiden vähennys- ja luokittelumenetelmät ovat usein käyttökelpoisia myös hyperspektrisessä datassa. Mikäli datan käsittelyssä ei huomioida suuridimensioisen datan ominaisuuksia ja pienen opetusjoukon ongelmaa, voi syntyä harhakäsitys, että perinteiset menetelmät eivät soveltuisi hyperspektrisiin aineistoihin. Uusia menetelmiä tarvitaan mm. siihen, että piirteiden määrä voidaan vähentää niin alhaiseksi, ilman merkittävää informaatiohukkaa, että Hughesin ilmiö ei vaikuttaisi perinteisten piirteiden ekstraktointi- ja luokittelumenetelmien antamiin tuloksiin. Toisena esimerkkinä uusista menetelmistä ovat algoritmit, joilla luokkien tilastollista käyttäytymistä voidaan arvioida paremmin, kun opetusnäytteitä ei ole riittävästi käytössä [Landgrebe 5/1998]. Näitä viimeksi mainittuja menetelmiä ei käsitellä tässä tutkimuksessa.

4.4 Piirteiden valintamenetelmiä

Piirteiden valinnassa tehtävänä on valita piirrejoukosta halutun kokoinen alijoukko, jolla eri luokat erottuvat parhaiten toisistaan. Tällöin on päätettävä millä kriteerillä määritellään alijoukon hyvyys ja millä algoritmilla eri piirrekombinaatiot käydään läpi. Usein käytetyt kriteerit ovat luokittelutarkkuus tai luokkien väliset statistiset etäisyydet. Käyttäjän pitää myös päättää haluamansa alijoukon koko. Ainut perusteellinen algoritmi käy läpi kaikki piirrekombinaatiot. Käytännössä tämä ei ole tietenkään mahdollista, koska kombinaatioiden määrä kasvaa räjähdysmäisesti piirteiden määrän lisääntyessä. Seuraavissa kappaleissa esitellään yleisimpiä algoritmejä [Jain 2000], joilla vältetään kaikkien kombinaatioiden läpikäynti, mutta samalla yritetään löytää paras mahdollinen piirrejoukko.

Käytännössä hyperspektrisen datan kanavamäärä vähenee automaattisesti kun poistetaan ilmakehän aiheuttamat voimakkaimmat absorptiokaistat, mikäli ei tutkita nimenomaan ilmakehän ominaisuuksia.

<u>Brach-and-Bound (haaraudu ja rajoita)</u>

Branch-and-bound on ainut ”optimaalinen” menetelmä, joka ei käy kaikkia piirrekombinaatioita läpi [Narendra]. (Tässä yhteydessä ei käydä algoritmin yksityiskohtia läpi.) Menetelmän avainkohtana on, että kriteerifunktio J(X) on monotoninen; kahdelle alijoukolle X_1 ja X_2, jos $X_1 \subset X_2$, niin $J(X_1) < J(X_2)$. Toisin sanoen alijoukon hyvyys pitäisi parantua aina kun uusi piirre lisätään siihen. Välttämättä kaikki kriteerifunktiot eivät täytä tätä vaatimusta. [Jain 2002], [Bacher]

<u>Best Individual Features (BIF)</u>

Menetelmässä arvioidaan kunkin piirteen informaatiosisältö erikseen, minkä jälkeen valitaan haluttu määrä parhaimpia piirteitä. Tapa on laskennallisesti yksinkertainen, mutta voi epäonnistua dramaattisesti. Voi kuitenkin olla käyttökelpoinen ensivaiheen tapa kun piirrejoukko on hyvin suuri (esim. satoja piirteitä). BIF:n jälkeen voidaan käyttää muita menetelmiä, jotka huomioivat eri piirteiden yhteisvaikutuksen.

Sequential Forward Selection (SFS)

Aluksi valitaan paras piirre. Tämän jälkeen lisätään yksi piirre kerrallaan sen mukaan minkä piirteen lisääminen aiheutti parhaan tuloksen. Silmukkaa jatketaan niin kauan, kun halutun kokoinen alijoukko on kasassa. Menetelmässä kerran alijoukkoon kiinnitettyä piirrettä ei voi myöhemmin poistaa.

Sequential Backward Selection (SBS)

Päinvastainen kuin SFS eli aluksi tutkitaan minkä piirteen poistaminen hävittää vähiten informaatiota. Tämän jälkeen poistetaan vastaavalla tavalla piirre kerrallaan, kunnes haluttu dimensio on saavutettu. Vastaava ominaisuus kuin SFS:ssä: kun piirre on poistettu, ei sitä oteta enää myöhemmin mukaan. Laskennallisesti kompleksisempi kuin SFS.

”Plus l - take away r” Selection

Ensin piirrejoukko kasvatetaan SFS algoritmillä kokoon l, jonka jälkeen poistetaan r kappaletta piirteitä SBS:llä. Käyttäjän täytyy itse määrittää arvot l ja r.

Sequental Forward/Backward Floating Search (SFFS, SBFS)

Nämä ovat yleistyksiä ”Plus l - take away r” algoritmista. Menetelmät kulkevat takaisinpäin niin kauan, kunnes löydetään parempi alijoukko suhteessa samankokoiseen joukkoon, joka löydettiin edellisellä kerralla. Menetelmissä luvut l ja r määräytyvät automaattisesti ja päivittyvät dynaamisesti.

4.4.1 Huomioita valintamenetelmistä

Edellä esitellyt BIF, SFS ja SBS menetelmien heikkous on siinä, että alijoukon hyvyys ei ole yksittäisten piirteiden summa, vaan piirteiden yhteisvaikutuksen tulos. Eli yksinkertaistettuna kahteen dimensioon: paras piirrepari ei välttämättä sisällä parasta yksittäistä piirrettä. Tämän vuoksi ”Plus l – take away r”, SFFS ja SBFS menetelmät toimivat melkein missä tahansa suuressa piirrejoukossa paremmin kuin suorat lähestymistavat (BIF, SFS ja SBS). Ferri et al. [Ferri] ja Jain et al. [Jain] ovat vertailleet valintamenetelmiä suhteessa luokittelutarkkuuteen ja laskenta-aikaan. Yleinen johtopäätös heillä oli se, että SFFS antoi melkein yhtä hyviä tuloksia kuin branch-and-bound ja laskentaprosessi vei vähemmän aikaa. [Jain 2000]

On muistettava, että mikäli alijoukkojen hyvyyden arviointi (kriteerifunktio) perustuu ohjattuihin luokittelumenetelmiin, suuren dimensionaalisuuden ongelmat täytyy tarvittaessa huomioida.

4.5 Piirteiden ekstraktointimenetelmiä

4.5.1 PCA

Yksi tunnetuimmista piirteiden ekstraktointimenetelmistä on PC-analyysi (principal components) eli pääkomponenttianalyysi, joka tunnetaan myös nimillä Karhunen-Loéve, Hotelling tai PC-muunnos. Tehtävänä on määrittää ortogonaalinen koordinaatisto, jossa data on korreloimaton minkä tahansa kohden koordinaattiakselin suhteen. PCA on ohjaamaton algoritmi.

Korrelaatiomatriisin elementit saadaan yhtälöllä

$$r_{ij} = k_{ij} / \sqrt{k_{ii} k_{jj}} \,, \tag{4.8}$$

jossa k_{ij}:t ovat kovarianssimatriisin elementtejä. Kovarianssimatriisin k_{ii} elementti on piirrevektorikomponentin i varianssi. Luku r_{ij} on i ja j komponenttien välinen korrelaatio, joka PCA komponenteilla on nolla. Itse kovarianssimatriisi saadaan yhtälöllä

$$\sum_x = \frac{1}{S-1} \sum_{s=1}^{S} (x_s - m)(x_s - m)^T \,, \tag{4.9}$$

jossa m on datan keskiarvovektori, x_s yksittäisen spektrin vektori ja T on matriisin transponointioperaattori. Luku S on x_s vektoreiden lukumäärä.

PCA:ssa idea on esittää data alkuperäisen x avaruuden sijasta y avaruudessa, tekemällä lineaarinen koordinaattimuunnos

$$y = Gx\,, \tag{4.10}$$

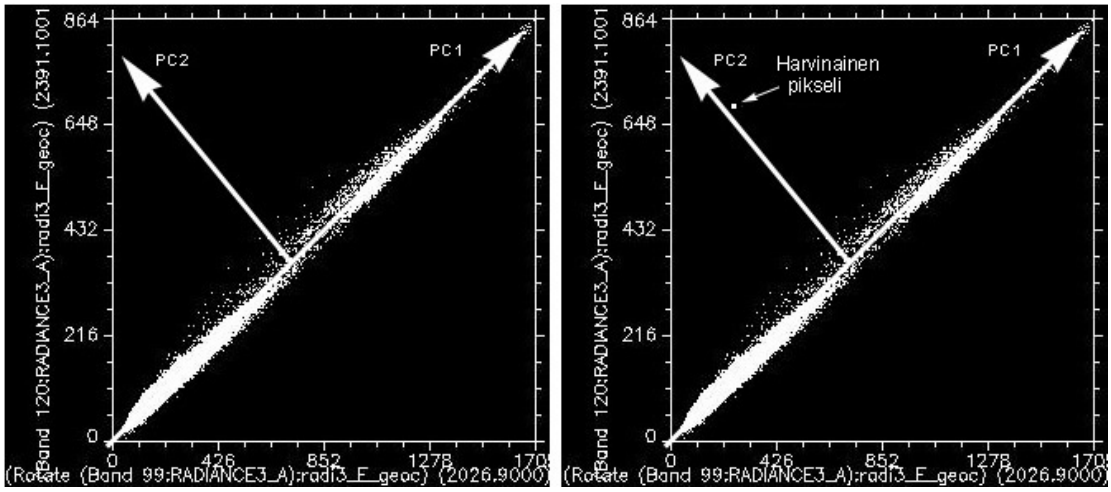
siten että y avaruuden korrelaatio- ja samalla kovarianssimatriisi $\sum_y$ ovat diagonaalimatriiseja (muut paitsi diagonaalielementit ovat nollia). Yhdistämällä kovarianssimatriisin yhtälö (4.9) ja yhtälö (4.10) saadaan välivaiheiden kautta (kts. esim. [Richards, pp. 137]), vanhan ja uuden koordinaatiston kovarianssimatriiseille matemaattinen yhteys:

$$\sum_y = G\sum_x G^T \tag{4.11}$$

Koska neliömatriisilla $\sum_x$ on dimensionsa verran lineaarisesti riippumattomia ominaisvektoreita, on se diagonalisoituva. Yhtälöstä (4.11) huomataan, että $\sum_x$ matriisin diagonalisointi antaa G matriisin, jolla haluttu koordinaatistomuunnos saadaan laskettua. [Richards, pp. 133-142]

Käytännössä kovarianssimatriisin $\sum_y$ elementit järjestetään laskevaan järjestykseen, ja samalla matriisi G niin että yhtälö (4.11) toteutuu, jolloin muunnoksen y=Gx jälkeen datalla on y_1 akselilla suurin varianssi, y_2 akselilla toiseksi suurin jne.. Tällöin ensimmäisillä pääkomponenteilla (y_1, y_2,…) on periaatteessa suurin informaatio ja viimeiset komponentit sisältävät vähiten informaatiota, jolloin ne voidaan unohtaa ja näin laskea dimensionaalisuutta.

PCA:n ongelma on siinä, että dataa kohdellaan yhtenä jakautumana. Tilastollisen analysoinnin tavoitehan on jakaa data erilaisiin jakautumiin, jotka edustavat eri luokkia. PCA voi johtaa siihen, että eri luokat sulautuvat yhteen uudessa koordinaatistossa, mikä huonontaa luokittelutarkkuutta [Landgrebe 4/1995, pp. 32]. Menetelmä voi olla erityisesti huono silloin, kun tehtävänä on etsiä datasta harvinaisia spektrejä.



Kuva 4.5: Kaksidimensioiseen piirreavaruuteen tehty PC-muunnos. Kuva on esimerkki tapauksesta, jossa harvinaisen pikseli sulautuu muuhun dataan, mikäli PC2 komponentti jätetään pois dimension vähentämisen yhteydessä.

Kuvassa 4.5 on havainnollistettu PCA:ta kaksidimensionaalisessa piirreavaruudessa. Vasemmanpuoleisessa kuvaajassa alkuperäinen kaksidimensionaalinen data voidaan ilman ongelmia esittää ensimmäisen pääkomponentin (PC1) avulla. Oikeanpuoleiseen kuvaajaan on lisätty yksi anomalinen datapikseli, joka esittää harvinaista spektriä. Koska muita datapisteitä on lukematon määrä, ei anomaliapiste vaikuta käytännössä mitenkään PC-muunnokseen. Mikäli data halutaan esittää ensimmäisen pääkomponentin avulla, anomaliapiste hukkuu muuhun dataan ja sen luokitteleminen yksidimensioisessa avaruudessa on lähes mahdotonta.

4.5.2 Minimum noise fraction (MNF)

MNF perustuu kahteen PC-muunnokseen, joiden välissä kohina valkaistaan. MNF tehdään seuraavasti [ENVI], [Igarss 2002].

1) Tehdään PC-muunnos käyttäen kohinakovarianssimatriisia, jonka määrittämistä varten kohina täytyy arvioida. Periaatteita kohinan arvioimiseen on esim.: a) olettaa että vierekkäiset kanavat sisältävät saman datan, mutta eri kohinan, b) mittalaitteen ns. pimeiden kuvien mittaaminen.
2) Valkaistaan datan kohina eli tehdään koordinaattimuunnos, jonka jälkeen kohinakovarianssimatriisi on diagonaalinen. Graafisessa mielessä valkaiseva koordinaattimuunnos venyttää tai kutistaa akseleita.
3) Tehdään PC-muunnos käyttäen datasta laskettua kovarianssimatriisia.

Yksi MNF:n variaatio on tehdä vaiheiden 1 ja 2 jälkeen käänteinen PC-muunnos eli muuntaa data takaisin alkuperäiseen koordinaatistoon. Ennen muunnosta viimeisten PC-akseleiden sisältämää dataa voidaan tasoittaa. [ENVI]

Koska MNF perustuu PCA:han, ei menetelmällä voida välttämättä ekstraktoida harvinaisten spektrien piirteitä toivotulla tavalla. Se mitä PCA pitää merkityksettömänä, voi olla arvokasta tietoa luokittelua varten. Tämä on olennaista erityisesti sotilassovelluksissa, koska yleensä sotilaskohteet ovat harvinaisia muiden pikseleiden (esim. metsä, pelto, suo, avohakkuu) joukossa.

4.5.3 Independent Component Analysis (ICA)

Mikäli komponentit ovat riippumattomia, ne ovat myös aina korreloimattomia [Holopainen]. Käänteisyys (korreloimattomat ovat riippumattomia) pätee vain mikäli data on gaussista [Hyvärinen].

PCA on menetelmä joka etsii korreloimattomia komponentteja, ei välttämättä mitään muuta. ICA:n tehtävä on löytää statistisesti riippumattomia (tai mahdollisimman riippumattomia) komponentteja, silloinkin kun data ei ole gaussista. Tämä on keskeisin ero ICA:n ja PCA:n välillä. Lisäksi ICA komponentit eivät välttämättä ole ortogonaalisia [Alparone] ja on olemassa monia modifikaatioita ICA algoritmeistä. [Chen], [Muhammed]

Piirteiden ekstraktoinnin lisäksi ICAa on käytetty myös spektrien unmixing ongelmiin [Chang 2002a] ja kohteen havaitsemiseen [Robila]. Esim. Robila et al. [Robila] käyttämä ICA sovellus perustuu siihen, että anomalisia spektrejä voidaan pitää riippumattomina komponentteina muun aineiston joukossa.

4.5.4 Empiirisiin havaintoihin perustuvat menetelmät

Analyyttisten ekstraktointimenetelmien sijasta tehokkaita keinoja voivat olla yrityksen ja erehdyksen periaatteella löydetyt tai tunnettuihin absorptio-ominaisuuksiin perustuvat ekstraktoinnit. Perusidea näissä on laskea, vähentää, kertoa ja/tai jakaa kanavia keskenään. Menetelmien yksi heikkous on siinä, että ne ovat hyvin sovelluskohtaisia. Kasvillisuuden tutkimiseen käytettyjä menetelmiä ovat mm. ”Two-Band Vegetation Indices” (TBVI), ”Optimum Multiple-Band Vegetation Indices” (OMBVI), ”NIR- and Red-Based Normalized Difference Vegetation Indices” (NDVI), Transformed Soil Adjusted Vegetation Indices” (TSAVI), ”Atmospherically Resistant Vegetation Indices” (ARVI), Middle Infrared-Based Vegetation Indices” (MIVI), ”Principal Component Vegetation Indices” (PCVI) [Thenkabail].

Historiallisesti tunnetuin indeksi lienee Tasseled Cap muunnos, joka kehitettiin 1970-luvulla multispektristen Landsat satelliittikuvien analysointiin. Ideana on tehdä lineaarinen kombinaatio alkuperäisistä kanavista. Tuloksena saadaan piirteet, jotka indikoivat ”maaperän kirkkautta”, ”vehreyttä”, ”keltaista ainetta” ja ”ei mitään edellisistä”. [Richards]

4.5.5 Discriminate Analysis Feature Extraction (DAFE)

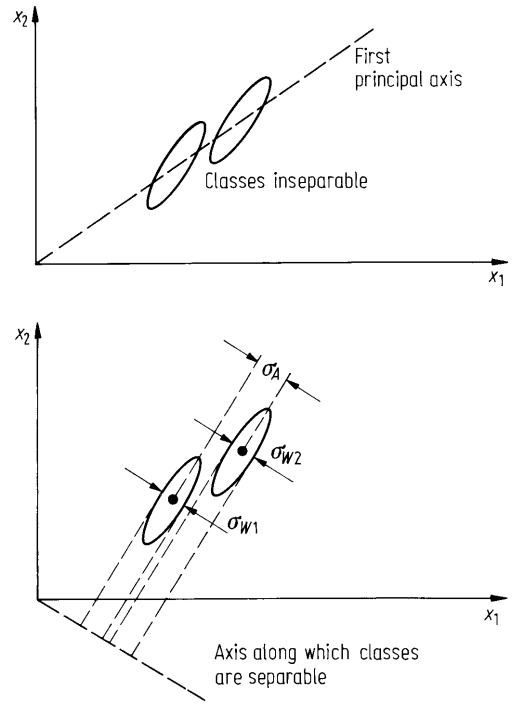
DAFE, joka pienen modifikaation jälkeen tunnetaan myös nimellä Canonical Analysis, on lineaarinen koordinaattimuunnos, jossa tehtävänä on esittää data sellaisten akselien avulla, jotka maksimoivat Fisherin suhteen:

$$\lambda = \frac{\sigma_A^2}{\sigma_w^2} = \frac{d^t \sum_A^x d}{d^t \sum_w^x d} . \qquad (4.12)$$

Yhtälössä $\sum_w$ on (opetusnäytteiden määrällä painotettu) keskiarvo eri luokkien kovarianssimatriiseista ja $\sum_A$ on eri luokkien keskiarvovektoreista laskettu kovarianssimatriisi. (DAFE on siis ohjattu menetelmä.) d:t ovat matriisin D komponentit, jotka määrittelet koordinaattimuunnokset y=Dx (x on alkuperäinen ja y uusi avaruus). Vastaavasti σ_A^2 on eri luokkien keskiarvojen varianssi ja σ_w^2 on keskiarvo luokkien variansseista kyseessä olevalla uudella akselilla. (Huom. varianssin neliöjuuri on keskihajonta.) Maksimointiehdon seurauksena koordinaattimuunnoksen määrittelevä D matriisi löydetään yhtälöllä

$$(\sum_A^x - \Lambda \sum_w^x) D = 0 , \qquad (4.13)$$

jossa Λ on diagonaalinen matriisi, jonka komponentteina ovat luvut λ_i. Suurin informaatio on ominaisarvoa λ_1 ja -vektoria d_1 vastaavalla akselilla, toiseksi suurin λ_2 vastaavalla akselilla jne.. DAFE:a kutsutaan kanoniseksi analyysiksi, mikäli lisäksi vaaditaan että uudessa avaruudessa $\sum_w$ on yksikkömatriisi. [Richards, pp. 249-255]



Kuva 4.6 [Richards]: PC-muunnoksen (ylempi) ja DAFE:n (alempi) eroavaisuudet piirteiden ektraktoinnissa. Kyseiset luokat ovat normaalijakautuneita.

Kuvassa 4.6 on havainnollistettu DAFE:a hypoteettisessä kahden luokan ja dimension tilanteessa sekä tehty samaan aineistoon PC-muunnos. Sekä PC-, että alkuperäisillä komponenteillä luokkia ei voi täysin erottaa toisistaan.

Vaikka DAFE toimisikin joissain tapauksissa hyvin, on sillä useita haittapuolia. Ensinnäkin ekstraktoituja piirteitä voidaan tuottaa vain luokkien määrä miinus yksi kappaletta. Toisaalta kahden luokan ongelmassa DAFE toimii tällöin luokittelumenetelmänä. Toiseksi, jos keskiarvot ovat lähellä toisiaan tai samat, ekstraktoidut piirteet eivät ole luotettavia. Ja kolmanneksi, mikäli monen luokan tilanteessa jollain luokalla on hyvin erilainen keskiarvo, $\sum_w$ painottuu enemmän kyseistä luokkaa kohti, jolloin tuloksena on tehottomia piirteitä. [Landgrebe 1998b, pp. 94]

<u>DAFE:n modifikaatiot</u>

Foley-Sammon menetelmällä voidaan tuottaa alkuperäisen (n) dimension verran uusia piirteitä. Siinä DAFE:en eli Fisherin suhteen maksimointiin liitettävä lisäehto on ortogonaalisuus eli $d_i^T d_n=0$, i=1,…,n-1. [Landgrebe 1998a, pp. 76] Mikäli d_i vektoreiden halutaan lisäksi olevan ortonormaaleja, niin Foley-Sammon nimen sijasta on käytetty nimeä ”Orthonormal discriminant vectors” (ODV) [Landgerebe 4/1995, pp. 34].

Sekä DAFE:ssa, että Foley-Sammon ja ODV laajennuksissa ektraktoinnit perustuvat pelkästään luokkien keskiarvojen eroihin. Parempia tuloksia saadaan kun lisäksi huomioidaan myös luokkien kovarianssien erot. (DAFE:ssa lasketaan luokkien kovarianssien keskiarvo ja kovarianssi luokkien keskiarvoista; ei vertailla eri luokkien kovariansseja toisiinsa.)

Fukunaga-Koontz ekstraktointi menetelmä perustuu kahden luokan kovarianssien erojen vertailuun. Se ei toimi hyvin tilanteissa, joissa keskiarvojen erot ovat hallitsevat, joten menetelmän käyttö on sopivaa vain samansuuruisten keskiarvojen tilanteissa. Fukunaga-Koontz DAFE:n laajennuksena huomioi sekä luokkien keskiarvojen, että kovarianssien erot. Kuitenkaan Fukunaga-Koontz menetelmästä ei ole olemassa/raportoitu laajennusta yli kahta luokkaa sisältävän aineiston piirteiden ekstraktointiin [Landgrebe 1998a, pp. 77-78]. [Landgrebe 1/1993, pp. 28-29]

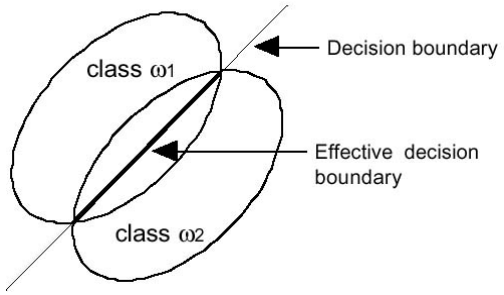
4.5.6 Decision Boundary Feature Extraction (DBFE)

Edellä käsitellyissä menetelmissä piirteiden ekstraktointi perustui koordinaattimuunnokseen, joka toteutti tietyn kriteerifunktion. DBFE menetelmässä lähestymistapa on erilainen. Siinä piirteiden ekstraktointi perustuu suoraan luokittimien päätösrajoihin, mistä DBFE on saanut nimensä. Samalla menetelmä ennustaa sisäisen dimensionaalisuuden päätösrajojen perusteella. DBFE:stä on todettu [Landgrebe 1998b, pp. 95] että se ei ole optimaalinen yli kahden luokan tapauksissa.

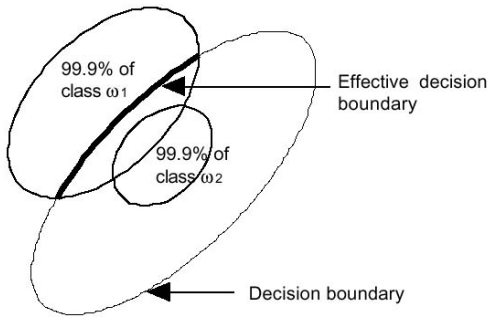
Koska päätösrajat riippuvat luokittimesta, ei ole olemassa yhtä ainoaa tapaa määritellä uudet piirrevektorit. Päätösrajat voidaan johtaa esim. parametrillisin, ei-parametrillisin tai neuroverkkojen avulla. Yleisesti ottaen vain parametrillisten luokittimien kohdalla rajat voidaan löytää analyyttisesti. Landgrebe et al. [Landgrebe 1/1993] ovat käsitelleet laajasti DBFE:tä, mm. kertoneet menetelmän matemaattisen taustan ja vertailleet sitä muihin ekstraktointimenetelmiin.

Kuvassa 4.7a on kuvitteellinen kahden luokan tapaus, jossa luokilla on eri keskiarvo, mutta sama kovarianssi ja kuvassa 4.7b tapaus, jossa luokilla sekä keskiarvo, että kovarianssi ovat erilaiset. Kuvien ellipsit (kuvassa 4.7b kaksi pienintä) kuvaavat luokkien todennäköisyysrajoja, joilla sijaitsevat datapisteet kuuluvat todennäköisyydellä P kyseiseen luokkaan. Vastaavasti rajojen sisäpuolella olevat datapisteet kuuluvat suuremmalla todennäköisyydellä kuin P kyseiseen luokkaan. Kuvien 4.7a-b todennäköisyysrajat ovat säännöllisen muotoisia, koska ne perustuvat normaalijakautumaan. Esim. ei-parametrillisiin menetelmiin perustuvat rajat voivat olla mielivaltaisen muotoisia.

Vaikka päätösraja kuvassa 4.7a on äärettömän pitkä viiva ja kuvassa 4.7b isoin ellipsi, niin käytännössä piirteiden tehokkaan vähentämisen kannalta tarvitsee ottaa huomioon vain ns. tehokkaat päätösrajat. Tehokas päätösraja määrittyy päätösrajan ja todennäköisyysrajojen leikkauskohdista (kts. kuvat 4.7a-b).



Kuva 4.7a: Kaksi luokkaa, joilla on eri keskiarvo, mutta sama kovarianssi.



Kuva 4.7b: Kaksi luokkaa, joilla on eri keskiarvo ja kovarianssi.

Kun päätösrajat on määritelty, voidaan niistä johtaa piirrevektorit. Redundanttinen piirrevektori on sellainen, joka on päätösrajan tangentin suuntainen ja informatiivinen piirrevektori on ainakin jossain päätösrajan pisteessä rajan normaalin suuntainen. DBFE:n piirrevektoreiden määrittäminen perustuu siis päätösrajojen normaalivektoreihin. Tarkemmin sanottuna tehtävänä on määrittää ns. päätösraja piirrematriisi, joka määrittelee ekstraktoidut piirrevektorit. Piirrematriisin lineaarisesti riippumattomien vaaka- tai pystyvektorien määrä (rank) vastaa sisäistä dimensionaalisuutta. Käyttämällä tehokkaita

päätösrajoja piirrematriisin määrittämiseen, saadaan ekstraktoitujen piirteiden määrää laskettua. Toisaalta tällöin informaatio teoriassa vähenee. Esim. mikäli käytetään 99.9% todennäköisyyteen perustuvaa rajaa (kuva 4.7b), niin korkeintaan 99.9% aineistosta on teoriassa mahdollista luokitella oikein. Kuitenkin käytännössä tehokkaan rajan käyttämisellä saavutettava hyöty (pienempi dimensio) on haittaa (teoreettinen informaatiohukka) suurempi. Kuvasta 4.7b ei suoraan näy ilmi tehokkaan päätösrajan hyöty, koska kyseinen tehokas päätösraja on kaareva ja kaksidimensioinen. Mutta mikäli (kuvan 4.7b) tehokasta päätösrajaa approksimoitaisiin suoralla viivalla, voidaan data kuvata DBFE:llä suoraan yhteen dimensioon. Tehokkaiden päätösrajojen lisäksi dimensiota voidaan vähentää hylkäämällä pienen ominaisarvon omaavat piirrevektorit. [Landgrebe 1/1993]

4.5.7 Discriminant Feature Extraction (DFE)

Discriminant Feature Extraction on yleistetty versio DBFE:stä. DBFE:n piirrevektorit perustuivat päätösrajan normaalivektoreihin, kun taas DFE:n piirrevektorit perustuvat diskriminanttifunktion normaalivektoreihin. Tarkemmin sanottuna ne ilmaisevat diskriminanttifunktion muutosta.

Landgrebe et al. [Landgrebe 1/1993] ovat todenneet, että DBFE toimii DFE:tä paremmin kun päätösrajat on voitu määrittää hyvin. Kuitenkin mikäli päätösrajojen määrittäminen on hankalaa ja aikaa vievää, kuten joissakin ei-parametrillisissä tapauksissa, DFE antaa vastaavanlaisia tuloksia kuin DBFE.

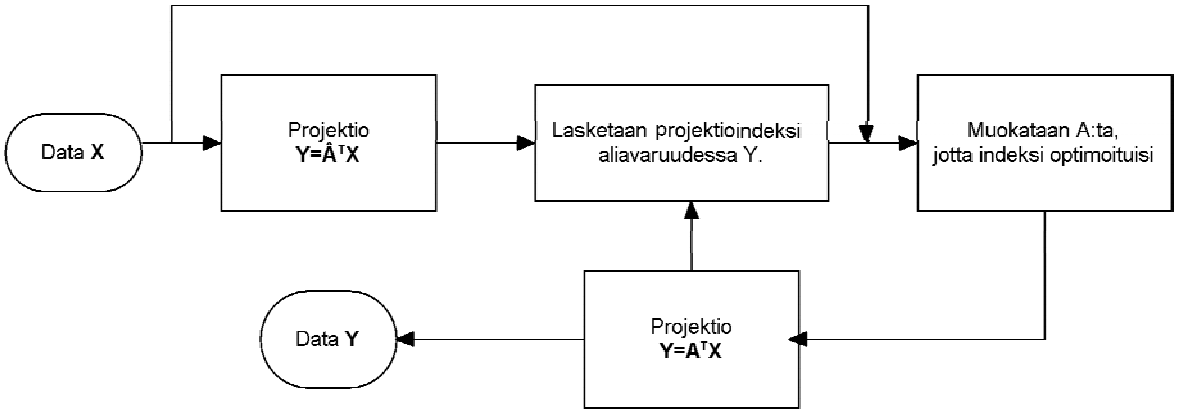
4.5.8 Projection Pursuit (PP)

Kappaleessa 4.2.1 käsiteltiin lisääntyneen kanavamäärän tuomia ongelmia kaukokartoitusdatan analysoinnissa, joista merkittävin on Hughesin ilmiö. Edellä käsitellyt ekstraktointimenetelmät, mm. DAFE ja DBFE, kärsivät siitä että datan muuntaminen aliavaruuteen perustuu parametrien laskemiseen alkuperäisessä suuridimensioisessa avaruudessa, jolloin parametrien tms. arvioiminen on epätarkkaa. Landgrebe et al. [Landgrebe 4/1995] ovat esittäneet ratkaisuksi Projection Pursuit (PP) nimisen optimointimenetelmän.

Yksi tärkeimmistä tekijöistä PP:ssä on projektioindeksin valinta. Indeksi voi olla parametrillinen, ei-parametrillinen, ohjattu tai ohjaamaton. Tavoite PP:ssä on löytää sellainen aliavaruus, jossa valittu projektioindeksi saavuttaa globaalin ääriarvonsa.

Kuvassa 4.8 on kaavio Projection Pursuitista. Käytetyt symbolit ovat seuraavat:

- X on dxN matriisi, joka määrittyy alkuperäisen d-dimensionaalisen avaruuden sisältämän datan perusteella. Matriisin rivit vastaavat aallonpituuksia ja sarakkeet eri spektrejä. Ohjatussa luokittelussa matriisi X rakentuu opetusnäytteistä.
- Y on mxN matriisi, joka määrittelee aliavaruuteen projisoidun datajoukon.
- A on ortonormaali dxm muunnosmatriisi, joka määrittää koordinaattimuunnoksen $Y=A^TX$.
- $I(A^TX)$ on projektioindeksi eli mitta projektion hyvyydelle.



Kuva 4.8: Kaavio Projection Pursuitista.

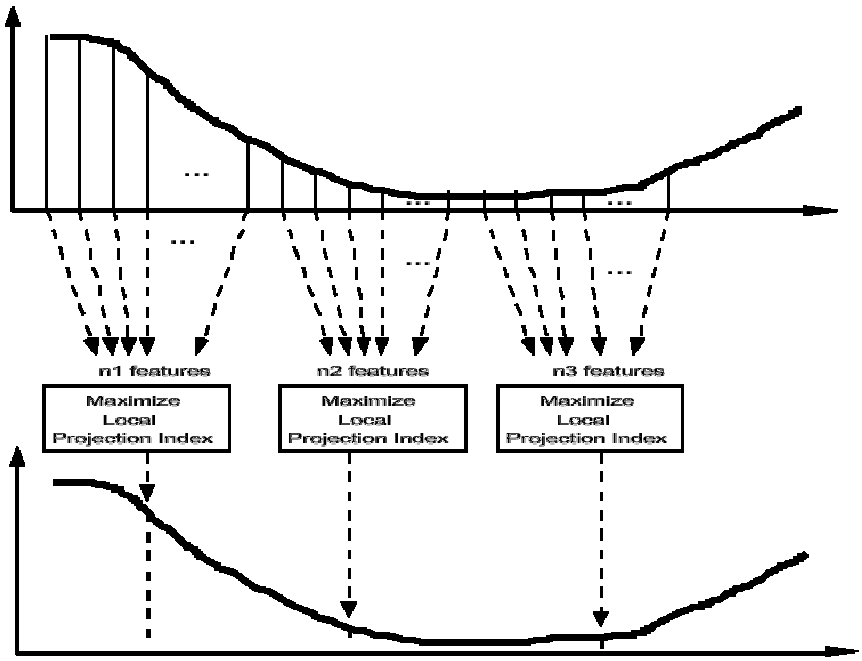
PP:ssä data aluksi projisoidaan Â matriisilla, jonka jälkeen syntyneessä aliavaruudessa lasketaan projektioindeksi. Tämän jälkeen matriisia A muokataan, jotta projektioindeksi optimoituisi. Tällaista iteraatioprosessia jatketaan niin kauan kunnes indeksi ei enää muutu merkittävästi, jolloin (toivottavasti) ollaan minimoitu tai maksimoitu indeksi.

Itseasiassa DAFE on eräänlainen Projection Pursuit, koska DAFE:ssa maksimoidaan Fisherin suhde (4.12). Ratkaiseva ero on siinä, että DAFE:ssa muunnosmatriisi voidaan suoraan määrittää (analyyttisesti) estimoimalla parametrit lähtöavaruudessa. PP:ssä muunnosmatriisin (A) optimoinnin tarve tutkitaan projisoidussa aliavaruudessa, jolloin lähes aina tarvitaan numeerisia iteraatioprosesseja parhaan projektion löytämiseksi.

A matriisin optimoinnin suorittaminen järkevässä laskenta-ajassa voi törmätä ongelmaan, jossa löydetään projektioindeksin paikallinen ääriarvo ja globaali ääriarvo jää saavuttamatta. Landgrebe et al. [Landgrebe 4/1995] ovat ehdottaneet ratkaisuksi menetelmää, jossa vierekkäisistä kanavista muodostetaan ryhmiä, jotka projisoidaan yhteen dimensioon. Ehtona A matriisille on diagonaalisuus, jolloin projisoidut piirteet ovat lineaarisesti riippumattomia. A matriisi on siis muotoa: $A=[A_1,A_2,\ldots,A_m]$, missä $A_i=[0,\ldots 0, a_i, 0,\ldots 0]^T$. Matriisielementti a_i on vektori, joka projisoi i:n vierekkäisten kanavien ryhmän yhteen dimensioon. Elementit a_i:t voidaan löytää kahdella tavalla: 1) kukin a_i optimoidaan lokaalisti ja itsenäisesti tai 2) a_i:n optimoinnissa huomioidaan muiden a_j ($j \neq i$) komponenttien vaikutus globaaliin projektioindeksiin. Edellistä menetelmää kutsutaan nimellä ”Parallel PP” ja jälkimmäistä ”Sequential PP”.

Parallel Projection Pursuit

Kuvassa 4.9 on fysikaalinen tulkinta Parellel PP:stä. Siinä vierekkäisistä kanavista muodostetaan ryhmiä, jonka jälkeen kunkin ryhmän kanavat projisoidaan omalla a_i vektorilla yhteen dimensioon. Ideana on löytää sellaiset a_i vektorit, jotka maksimoivat projektioindeksin omassa yksiulotteisessa avaruudessa, välittämättä muiden ryhmien muunnosvektoreista ja siten globaalista indeksistä. Toisin sanoen matriisi A on itsenäisesti löydettyjen a_i vektoreiden lopputulos. Parellel PP haittana on että se ei huomioi muiden kanavaryhmien projektioita. Ilmiö on sama mitä käsiteltiin piirteiden valintamenetelmien yhteydessä eli aliavaruuden hyvyys ei ole yksittäisten piirteiden summa, vaan yhteisvaikutuksen tulos.



Kuva 4.9 [Landgrebe 4/1995]: Parallel Projektion Pursuit.

Sequential Projection Pursuit

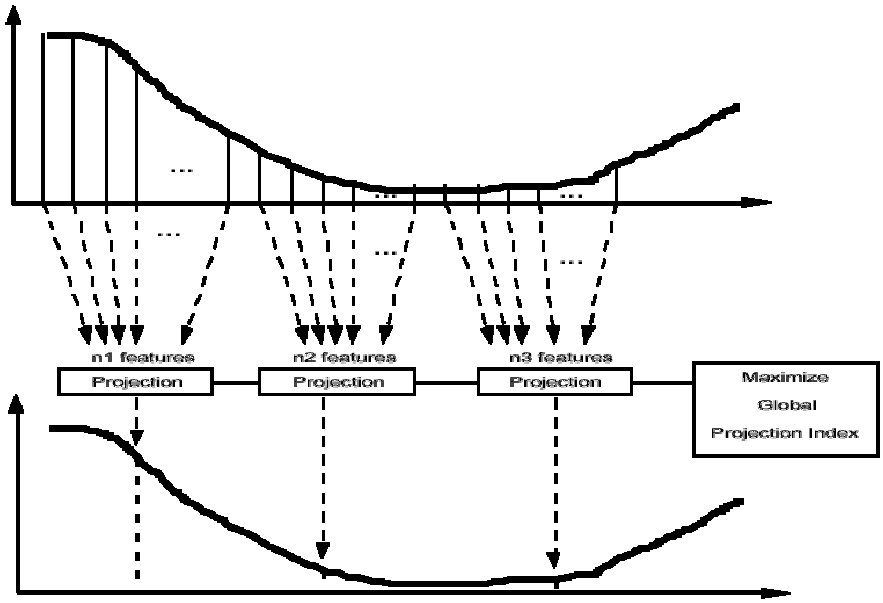
Kuvassa 4.10 on fysikaalinen tulkinta Sequential PP:stä. Siinä kunkin vierekkäisten kanavien ryhmän a_i vektorin löytäminen perustuu globaalin projektioindeksin optimointiin. Iteratiivinen algoritmi vektoreiden löytämiseksi on seuraava:

1) Jokaiselle vierekkäisten kanavien ryhmälle valitaan hypoteettinen a_i.

2) Pidetään muut a_i:t vakiona ja määritetään uusi a_1, jotta globaali projektioindeksi optimoituisi.

3) Siirrytään seuraavan a_i:n määrittämiseen, pitäen muut a_j:t vakiona.

4) Kun kaikki a_i:t on käyty peräkkäin läpi, siirrytään kohtaan 2. Silmukka lopetetaan kun projektioindeksin merkittävä optimoituminen pysähtyy.



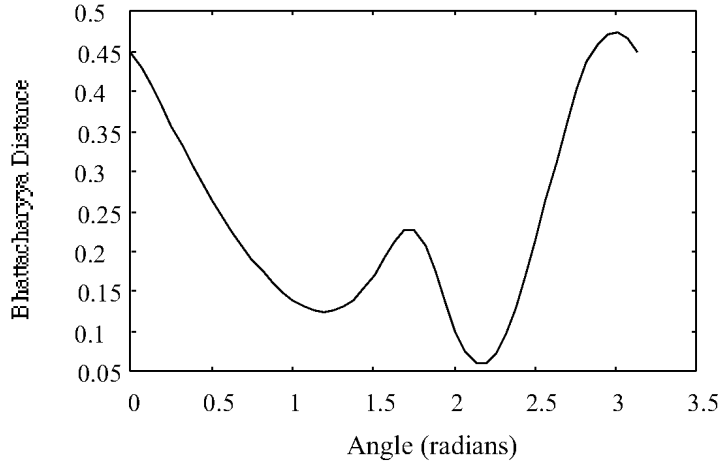
Kuva 4.10 [Landgrebe 4/1995]: Sequential Projektion Pursuit.

Aloitusmatriisin Â valinta

Edellä ei otettu kantaa kuinka PP prosessin aloitusmatriisi Â määritetään. Huono Â:n valinta voi johtaa siihen, että iteraatioprosessi lopetetaan paikalliseen ääriarvoon, jolloin optimaalisin A jää määrittämättä. Seuraavassa on esimerkki, joka havainnollistaa ongelmaa.

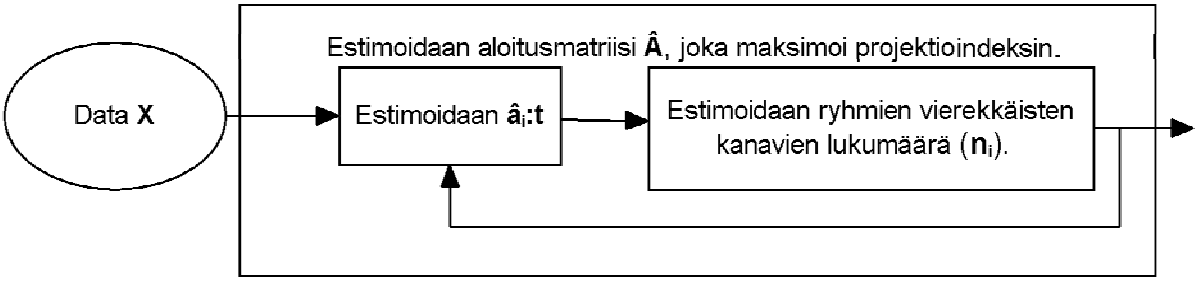
Esimerkissä [Landgrebe 4/1995, pp. 80-81] tehtävänä on projisoida kahden luokan aineisto kaksidimensioisesta avaruudesta yhteen. Käytettävä projektioindeksi on luokkien välinen Bhattacharyya etäisyys, joka määräytyy luokkien keskiarvoista ja kovariansseista. Etsittävä yksidimensioinen avaruus on pisteistä (cos(θ), sin(θ)) muodostuva suora, jossa käytettävä indeksi maksimoituu. Kuvassa 4.11 on Bhattacharyya etäisyys kulman θ funktiona. Paras projektio (indeksin globaali maksimi) löytyy kulmalla 3, mutta kuvaajassa on myös toinen, paikallinen ääriarvo 1,73 radiaanin kohdalla. Tällöin iteraatioprosessi voi mennä seuraavasti:

1) Aloitetaan huonosti valitulla Â matriisilla, joka antaa kulmaksi 1,5 radiaania.
2) Seuraava A antaa arvoksi 1.6 radiaania, jolloin indeksi paranee.
3) 1,73 radiaanin jälkeen indeksi muuttuu huonompaan suuntaan, jolloin iteraatioprosessi lopetetaan ja paras projektio jää löytämättä. Tietysti laskennallisesti ei olisi ongelma jatkaa iteraatiota globaalin maksimin löytämiseksi, mutta todellisissa tapauksissa muuttuvia parametrejä on liikaa, jotta kaikki kombinaatiot voitaisiin käydä läpi.



Kuva 4.11 [Landgrebe 4/1995, pp. 81]: Kahden luokan välinen Bhattacharyya etäisyys kulman funktiona, kun tutkittava data on projisoitu pisteiden (cos(θ), sin(θ)) muodostamalle suoralle.

Ratkaisu pienten lokaalisten ääriarvojen ongelmaan on arvioida aloitusmatriisi Â mahdollisimman hyvin. Landgrebe et al. [Landgrebe 4/1995, pp. 79-92] ovat ehdottaneet menetelmää, joka estimoi $\hat{a}_i$ vektorit ja vierekkäisten kanavien lukumäärän n_i jokaisessa ryhmässä. Menetelmän vaiheita on havainnollistettu kuvassa 4.12.



Kuva 4.12: Menetelmä luotettavan aloitusmatriisin Â löytämiseksi.

Vektoreiden $\hat{a}_i$ estimoinnissa kullekin vierekkäisten kanavien ryhmälle on joukko vektoreita, jotka on löydetty itsenäisesti suhteessa muihin ryhmiin. Kunkin ryhmän vektoreista poimitaan lopulta vektorit, jotka optimoivat globaalin projektioindeksin. Tässä vaiheessa oletuksena on se, että vierekkäisten kanavien lukumäärä (n_i) tunnetaan kussakin ryhmässä, ts. tiedetään kuinka kanavat pitää jakaa ryhmiin.

Se miten vierekkäiset kanavat pitää jakaa ryhmiin, voidaan tehdä binäärisillä puurakenteilla, joita Landgrebe et al. [Landgrebe 4/1995] ovat selittäneet yksityiskohtaisesti.

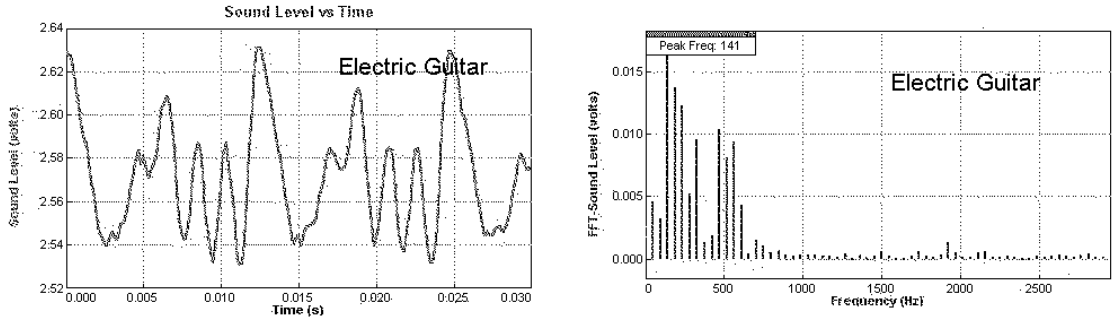
Yhteenveto PP:stä

Projectio Pursuitin vahvuus on siinä, että käsiteltävää dataa tutkitaan projisoidussa aliavaruudessa, ei täydessä dimensiossa, jolloin vältetään Hughesin ilmiön negatiivista vaikutusta. PP:ssä on myös se etu, että lineaarisissa projektioissa projisoidulla datalla on taipumus lähestyä normaalimallia lähtöavaruuden dimension kasvaessa, jolloin datan jatkokäsittely voidaan tehdä luotettavammin normaalijakaumaan perustuvilla menetelmillä. Jatkokäsittely voi olla dimension edelleen pienentämistä, esim. DAFE, DBFE tai muilla ”perinteisillä” menetelmillä, tai luokittelua.

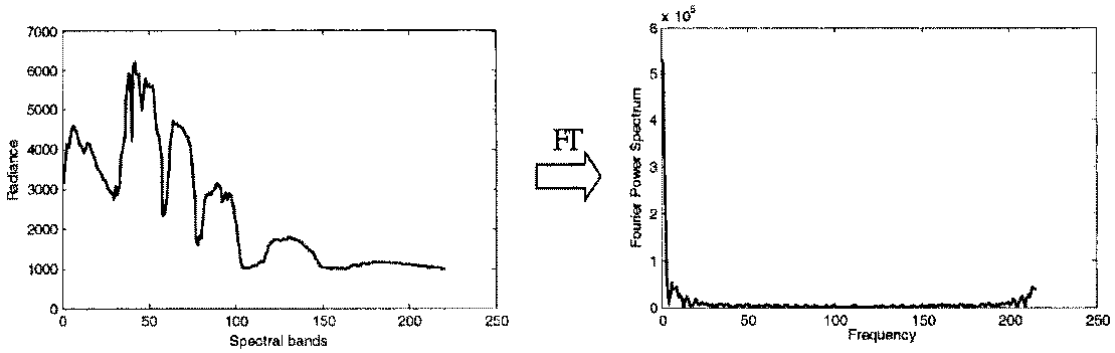
4.5.9 Wavelets

Wavelets (suom. aallokkeet) muunnokset ovat suhteellisen uusia menetelmiä signaalinkäsittelyssä. Kaukokartoitusdatan piirteiden ekstraktointia aallokkeilla ovat tutkineet mm. Hsu et al. [Hsu 1998], [Hsu 1999], [Hsu 2000], [Hsu 2001a], [Hsu 2001b] [Hsu 2002], Li et al. [Li] ja Kaewpijit et al. [Kaewpijit]. Hsu et al. ovat todenneet että ekstraktoinnissa aallokkeet ovat mielekkäämpiä ja vakaampia kuin perinteiset menetelmät (PCA, DAFE ja DBFE). Mielenkiintoiseksi aallokemuunnokset tekee se, että ne lasketaan spektriavaruudessa, ei piirreavaruudessa kuten tavalliset menetelmät. Tämän vuoksi esim. piirteiden ekstraktointiin tarvitaan vain yksi spektri, kun perinteiset menetelmät vaativat pääsääntöisesti monia opetusnäytteitä. Tässä tutkimuksessa ei käsitellä aallokkeiden teoriaa, idea allokkeisiin annetaan fourier muunnoksen kautta, koska niillä on yhtymäkohtia aallokemuunnoksiin.

Fourier muunnoksella voidaan periaatteessa mikä tahansa jaksollinen funktio esittää eritaajuisten sini- ja kosinikomponenttien avulla. Klassinen käytännön esimerkki on soittimen äänen sisältämien taajuuskomponenttien selvittäminen. Tekemällä fourier (sini tai kosini) muunnos äänen intensiteetin ajan funktioon, saadaan selville minkä taajuisista komponenteista (perusäänes ja harmoniset ylä-äänekset) ääni on muodostunut. Tästä on esimerkki kuvassa 4.13. Kuvassa 4.14 on kaukokartoitetulle datalle tehty fourier muunnos.



Kuva 4.13 [http://www.cord.edu/dept/physics/p128/lecture99_34.html#topic4]: Kitaran äänen intensiteetti ajan funktiona (vasemmalla) ja sille tehty fourier muunnos (oikealla). Fourier muunnos toimii hyvin soittimien äänille, koska ne ovat jaksollisia ja erityisesti luonteeltaan superpositio sinifunktioista.



Kuva 4.14: Hyperspektriselle datalle tehty fourier muunnos. Fourier spektrin taajuusakselilla ei ole mitään tekemistä valon taajuuden kanssa. Tässä yhteydessä taajuuskomponentit vastaavat ekstraktoituja piirteitä.

Fourier muunnoksella on muutamia heikkouksia, joiden vuoksi aallokkeet ovat parempia kaukokartoitusdatan piirteiden ekstraktoinnissa. Kun katsotaan kuvan 4.14 spektrille tehtyä muunnosta, niin siitä on mahdoton sanoa millä kanava-alueella tietty spektrin muoto esiintyi. Fourier muunnoksen perusongelma on se, että koko signaalialue yritetään mallintaa sini- ja/tai kosinikomponenttien avulla, jolloin tulokset ovat keskiarvoistamisen tuote. Aallokemuunnokset ovat tässä suhteessa parempia, koska ne pystyvät paloittelemaan spektriä ja siten huomioimaan paremmin spektrin nopeita muutoksia (piikkejä) tai laajojen kanava-alueiden muodostamia loivarakenteisia ominaisuuksia. Aallokemuunnoksissa käytettäviä perusfunktioita, joita vastaa fourierissa sinit ja kosinit, kutsutaan ”äiti-aallokkeiksi” (mother wavelets), joita on olemassa monia eri tyyppiä. Muunnoksissa perusfunktioita voidaan skaalata sen mukaan halutaanko ekstraktoinnissa painottaa pieniä (absorptiopiikkejä) vai laajoja (loivat rakenteet) muotoja spektreissä.

5. Spektrin unmixing menetelmät

Unmixing menetelmien motivaationa on se, että pikselit eivät aina sisällä vain yhtä materiaalia, vaan ovat sekoituksia eri materiaaleista. Tehtävänä on selvittää mistä eri materiaaleista pikselit koostuvat, mistä käytetään nimeä täydellinen unmixing. Tyypillisesti kohteen havaitsemissovelluksissa ei olla kiinnostuneita kaikista materiaaleista, vain yhdestä tietystä. Tällaiseen lähestymistapaan viitataan termillä osittainen unmixing. Kolmas, hieman harvinaisempi lähestymistapa on pitoisuuden arviointi (abundance estimation), jolla halutaan selvittää kuinka paljon pikseli sisältää kiinnostavaa materiaalia.

5.1 Lineaarinen vs. epälineaarinen malli

Mitattu spektri voi olla lineaarinen tai epälineaarinen sekoitus puhtaista spektreistä, joita kutsutaan endmembereiksi. Lineaarisiin malleihin pohjautuvat unmixing menetelmät ovat matemaattisesti yksinkertaisempia, koska mitatun spektrin oletetaan olevan lineaarikombinaatio puhtaista spektreistä. Matemaattisesti tämä kuvataan yhtälöllä:

$$\begin{pmatrix} r_1 \\ r_2 \\ . \\ r_K \end{pmatrix} = \begin{pmatrix} m_{11} & m_{21} & \dots & m_{E1} \\ m_{12} & m_{22} & \dots & m_{E2} \\ . & & & . \\ m_{1K} & m_{2K} & \dots & m_{EK} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ . \\ \alpha_E \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \\ . \\ n_E \end{pmatrix} \tag{5.1a}$$

missä r_i on mittausarvo kanavalla i ja α_i on endmemberin osuus mitatussa spektrissä.

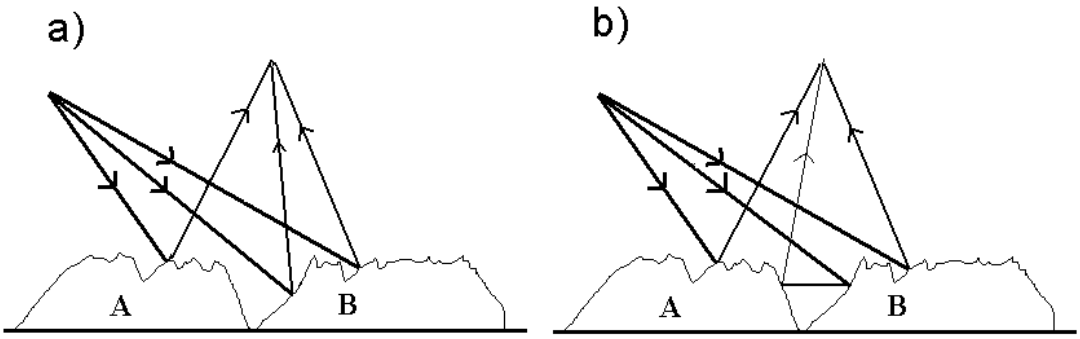
Fysikaalisessa mielessä yhtälö (5.1a) on mielekäs, kun luvut α_i eivät ole negatiivisia ja niiden summa on yksi. Näihin ehtoihin viitataan termillä ”täydellisesti rajoitettu unmixing”. Matriisin M, jonka komponentteina ovat m_{ij}:t, sarakevektoreina ovat endmembereiden spektrit ja kukin rivi liittyy tiettyyn aallonpituuteen. Yhtälön (5.1a) viimeisen pystyvektorin n_i:t ovat jäännöstermejä, joilla voidaan ilmaista mallinnuksesta ja/tai kohinasta aiheutuva virhe. Kun kanavien määrä on suurempi kuin endmembereiden määrä, kuten käytännössä aina hyperspektrisessä datassa, ei ole olemassa yhtä ainoaa ratkaisua löytää matriisi α yhtälöön (5.1a). Optimaalisin ratkaisu voidaan arvioida minimoimalla jäännöstermien RMS-virhe (neliöllisen keskiarvon virhe) eli pienimmän neliösumman (PNS) menetelmällä, jolloin matriisi α saadaan yhtälöllä [Ren] [Wang] [Settle]

$$\alpha_{PNS}=(M^TM)^{-1}M^T r =: M^{\#} r. \qquad (5.2)$$

Jatkossa tähän menetelmään viitataan termillä PNS-unmixing.

Yhtälö (5.2) olettaa että spektrin r kohina n on a) gaussista, b) valkoista eli kohinan kovarianssimatriisi on σI ja c) kohinan keskiarvo on nolla. Lisäksi on huomioitava, että yhtälö (5.2) ei aseta mitään raja-arvoja matriisin α_{PNS} komponenteille eli myös negatiiviset ja ykköstä suuremmat luvut voivat sisältyä ratkaisuun. [Settle]

Lineaarinen unmixing antaa epäluotettavia tuloksia mikäli havaitussa spektrissä on mukana merkittäviä epälineaarisuustekijöitä. Sekoittuminen on lineaarista kun fotonit heijastuvat vain kerran puhtaista materiaaleista. Epälineaarista sekoittumista tapahtuu kun fotonit heijastuvat useammasta materiaalista säteilylähteen ja sensorin välisellä reitillä. Kuvassa 5.1 on havainnollistettu lineaarisen ja epälineaarisen sekoittumisen syntymekanismeja. Kuvassa endmemberit A ja B muodostavat yhden pikselin. Merkitään A:n ja B:n reflektansseja symbolein m_a ja m_b. Lineaarisessa tapauksessa (a) pikselin havaittu reflektanssi on muotoa $R_{lin}=\alpha_a m_a+2\alpha_b m_b$ ja epälineaarisessa (b) muotoa $R_{epälin.}=\alpha_a m_a+\alpha_b m_b+ \alpha_b m_b\cdot \alpha_a m_a$. Lineaarisessa tilanteessa primäärinen sekoittuminen tapahtuu siis itse sensorissa, kun taas epälineaarinen sekoittuminen tapahtuu mittauskohteella.



Kuva 5.1: Lineaarisen (a) ja epälineaarisen (b) sekoittumisen syntymekanismit. A ja B ovat puhteina materiaaleja (endmembereitä). Lineaarisen mallin mukainen reflektanssi on muotoa $R_{lin}=\alpha_a m_a+2\alpha_b m_b$*, kun taas epälineaarisessa tapauksessa reflektanssi on muotoa* $R_{epälin.}=\alpha_a m_a+\alpha_b m_b+ \alpha_b m_b \cdot \alpha_a m_a$.

5.2 Lineaariset unmixing menetelmät

Valtaosa unmixing malleista perustuu siihen, että materiaalit ovat lineaarisesti sekoittuneita. Tyypillistä unmixing algoritmeille on se, että niitä kutsutaan eri nimillä, vaikka ne voidaan teoriassa osoittaa identtisiksi. Käytännössä algoritmien erot liittyvät siihen miten endmemberit valitaan ja millaisia oletuksia tehdään kohinasta.

5.2.1 Multiple Endmember Spectral Mixture Analysis (MESMA)

Yksinkertaisessa lineaarisessa unmixing menetelmässä endmembereiksi valitaan kiinteä ja optimaalisin joukko puhtaita spektrejä. Tämän jälkeen mitatuille kohteilla suoritetaan yhtälön (5.1a) mukainen unmixing prosessi. Mallin etuna on yksinkertaisuus ja painokertoimien fysikaalinen mielekkyys [Adams], [Roberts 1997]. Haittana on kuitenkin se, että mitatulla kuva-alalla eri pikselit eivät välttämättä koostu samoista endmembereistä ja niiden määrä pikseleiden kesken voi vaihdella [Sabol]. Jos pikseliä yritetään hajottaa eri materiaaleihin liian vähällä määrällä endmembereitä, niin mallintamattomien materiaalien spektrit vääristävät valittujen endmembereiden osuuksia ja RMS virhe kasvaa. Toisaalta jos valitaan liian monta endmemberiä, niin unmixing malli on herkkä häiriöille, mm. kohinalle ja ilmakehän vaikutuksille. [Roberts 1998]

MESMA eroaa perinteisestä mallista siten, että endmembereiden määrä unmixingissä voi vaihdella eri pikseleiden kesken. Myös samasta materiaalista voi olla useampia spektrinäytteitä. Se kuinka endmember-joukosta poimitaan parhaat ehdokkaat kullekin pikselille riippuu tapauksesta. Roberts et el. [Roberts 1998] ovat käyttäneet seuraavassa esiteltyä menetelmää.

MESMA:ssa aluksi luodaan joukko kaksi endmemberiä sisältäviä malleja. Tämän jälkeen kunkin mallin hyvyys arvioidaan muutamilla valintakriteereillä ja mikäli tarvetta on, malleja parannetaan ottamalla mukaan enemmän endmembereitä. Tällä vähennetään suuren mallijoukon tuomaa kompleksisuutta. Mallien valintakriteereitä ovat mm.:

1) Endmemberin painokerroin: Malli valitaan vain, jos puhtaiden spektrien painokertoimet ovat fysikaalisesti järkeviä eli väliltä -0.01–1.01 (täydellisesti rajoitettu unmixing). Mittalaitteen kohinasta johtuen painokertoimissa sallitaan 1% virhe.

2) RMS-virhe: Malli valitaan vain, jos jäännöstermien RMS-virhe on alle valitun kynnysarvon. Roberts et al. [Roberts 1998] ovat käyttäneet kynnysarvoa 0.025.

3) Yksittäisten jäännöstermien kynnysarvo. Malli valitaan vain, jos yksittäiset jäännöstermit eivät ylitä kynnysarvoa (voi olla eri kuin RMS-virheen kynnysarvo) liian monta kertaa peräkkäin tietyn levyisellä vierekkäisten kanavien alueella.

5.2.2 Matched Filtering (MF) / Constrained Energy Minimization (CEM)

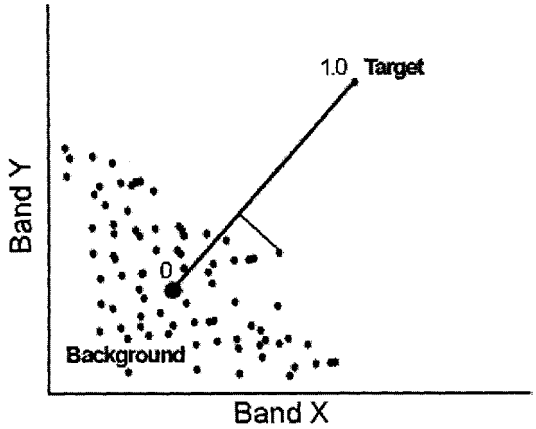
Matched Filtering on ENVI:n käyttämä nimitys algoritmista, joka yleisemmin tunnetaan nimellä Constrained Energy Minimization (CEM) [Farrand]. CEM:n tyypillinen käyttökohde radiotaajuisten signaalien käsittelyssä on tilanne, jossa moniosainen antenniryhmä havainnoi eri puolilta saapuvaa signaalia. Tällöin CEM on ratkaisu ongelmaan, jossa halutaan vahvistaa vain tietystä suunnasta tuleva signaali ja minimoida muista suunnista tuleva ei-toivottu signaali. [Harsanyi]

CEM:iä voi kuvata sanoin: ”anomalisen kohdepikselin etsiminen”. Menetelmässä ei yritetä selvittää kaikkia materiaaleja, joita tutkittavat pikselit voivat sisältää, vaan selvitetään vain jonkin kiinnostavan materiaalin osuutta pikseleissä. CEM:n toimivuuden kannalta keskeisin kysymys kuuluu: ”Onko etsittävä materiaali yleinen vai harvinainen pikseleiden joukossa?” Mitä yleisempi, sitä huonommin CEM toimii.

CEM:ssä käsiteltävää dataa pidetään tilastollisessa mielessä taustana. Idea on verrata tutkittavan pikselin spektriä taustan ja etsittävän materiaalin spektrin kesken. Tässä taustalla tarkoitetaan kaikkia pikseleitä, ei vain tutkittavan pikselin naapureita. Kuvassa 5.2 on kaksiulotteinen esitys CEM:stä. Aluksi tutkittavasta datasta lasketaan suurimman varianssin omaava suunta eli ensimmäinen pääkomponentti sekä keskiarvo. Tämän jälkeen keskiarvosta piirretään yksikkövektori etsittävän kohteen datapistettä kohti.

Yksikkövektori määräytyy optimoimalla kaksi, potentiaalisesti ristikkäistä ehtoa: 1) vektorin pitäisi olla kohti kohdetta ja 2) vektorin pitäisi olla ortogonaalinen pääkomponentin suhteen. Lopuksi datapisteet projisoidaan yksikkövektorille tai sen jatkeille ja pisteille annetaan arvot sen mukaan mihin pisteet ovat projisoituneet. Arvot lähellä ykköstä merkitsevät suurta kohdemateriaalipitoisuutta ja arvot lähellä nollaa päinvastaista tulosta. [Igarss 2002]

Kuvasta 5.2 käy ilmi myös se miksi kohteen pitää olla harvinainen, jotta CEM toimisi. Mikäli datassa on paljon piirteitä etsittävästä kohteesta, niin datapisteiden keskiarvo painottuu enemmän kohdetta kohti, jolloin yksikkövektorin suhteellinen pituus verrattuna datapilveen on lyhyt. Tällöin kohteen ja taustan välinen kontrasti pienenee.



Kuva 5.2: ENVI:n tapa havainnollistaa CEM menetelmää. [Igarss 2002], [ENVI]

Edellinen CEM algoritmin esittelytapa on lähinnä kvalitatiivinen havainnollistus. Matemaattisempi formalismi CEM:iin on seuraava. Tutkittavan pikselin i unmixing tulos, ns. CEM filtteri, saadaan matriisiyhtälöllä

$$y_i = w^T r_i = \frac{d^T R^{-1} r_i}{d^T R^{-1} d} =: \delta_{CEM}(r), \tag{5.3}$$

jossa r_i on pikselin i spektri ja d etsittävän materiaalin spektri. Matriisin w pitää olla sellainen, että lukujen y_i neliösumma (jaettuna pikseleiden määrällä N):

$$E = \frac{1}{N}\sum_{i=1}^{N} y_i^2 \tag{5.4}$$

minimoituu sillä ehdolla, että

$$w^T d = 1. \tag{5.5}$$

Tämä pakotettu minimointiongelma ratkeaa, kun w määritellään yhtälöllä

$$w = \frac{R^{-1}d}{d^T R^{-1} d} \quad , \tag{5.6}$$

jossa R^{-1} on käänteismatriisi datan korrelaatiomatriisista R.

Mikäli R matriisin suurimman ja pienimmän ominaisarvon ero on suuri, voi numeerisesti määritetyn R^{-1}:n tarkkuus alentua huomattavasti. Tällöin R^{-1} voidaan määritellä tarkemmin käyttämällä vain p kappaletta merkittävimpiä ominaisvektoreita laskemalla R yhtälöllä

$$R = V\Lambda V^T, \tag{5.7a}$$

jossa V on korrelaatiomatriisi, joka koostuu merkittävimmistä ominaisvektoreista. Λ on diagonaalimatriisi vastaavista ominaisarvoista. R^{-1} saadaan korvaamalla Λ matriisilla Λ^{-1} yhtälössä (5.7a). [Harsanyi], [Farrand]

Matriisi R voidaan laskea myös autokorrelaatiomatriisin yhtälöllä

$$R = \frac{1}{N}\left[\sum_{i=1}^{N} r_i r_i^T\right] \tag{5.7b}$$

[Du] [Chang 2002b].

Kuinka sitten määritellä arvo p eli kuinka monen merkittävimmän ominaisvektorin mukaan korrelaatiomatriisi pitäisi määritellä. Farrand [Farrand] on todennut, että vastaus on jokseenkin subjektiivinen; paras arvo voidaan löytää kokeellisesti. Valitettavasti ENVI:in (versio 3.5) implementoitu CEM/MF toiminto ei anna käyttäjälle mahdollisuutta valita ominaisvektoreiden määrää. Farrand [Farrand] on vertaillut CEM algoritmia eri p:n arvoilla suhteessa ENVI:n MF toimintoon. Johtopäätöksessään hän toteaa, että tulokset ovat yleisesti samanlaisia, mutta tietyissä tapauksissa ENVI:n MF toimi huonommin ja ehdottaa että ENVI:n puutteellista MF toimintoa voidaan parantaa, mikäli datalle suoritetaan ensin MNF-muunnos.

5.2.3 Target-constrained Interference-minimized Filter (TCIMF)

CEM menetelmällä etsitään kerralla jälkiä vain yhdestä kiinnostavasta endmembereistä, muita endmembereitä ei tarvitse edes tietää. Joissain tilanteissa on hyödyllistä jakaa tutkittavan alueen endmemberit kahteen luokkaan: kiinnostaviin ja ei-kiinnostaviin. Merkitään kiinnostavia endmembereitä matriisilla $D=[d_1,d_2,\ldots,d_{nd}]$ ja ei-kiinnostavia endmembereitä matriisilla $U=[u_1,u_2,\ldots,u_{nu}]$. Tarkoitus on luoda matriisi w, joka operoidessaan D matriisin antaa ulos ykkösiä (vrt. yhtälö (5.5)) ja operoidessaan U matriisin antaa ulos nollia. Samalla yhtälön (5.4) mukainen summa pitää minimoitua.

Optimaalisin matriisi w^{TCIMF} löydetään yhtälöllä

$$w^{TCIMF} = \frac{R^{-1}[DU]}{[DU]^T R^{-1}[DU]} \begin{bmatrix} 1_d \\ 0_u \end{bmatrix}, \tag{5.8}$$

jossa $1_d=(1,1,\ldots,1_{nd})^T$ ja $0_u=(0,0,\ldots,0_{nu})^T$. Sijoittamalla w^{TCIMF} yhtälöön (5.3) saadaan ”Target-constrained Interference-minimized Filter” (TCIMF):

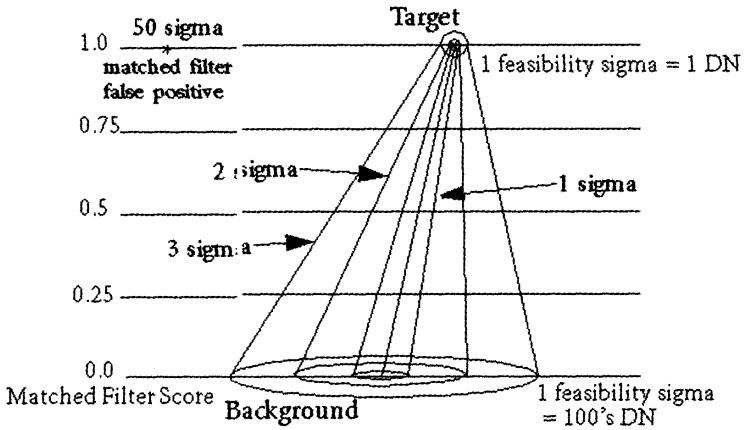
$$\delta_{TCIMF}(r) = (w^{TCIMF})^T r \tag{5.9}$$

[Chang 2002b], [Ren]. Huomaa että menetelmässä korrelaatiomatriisi R määräytyy samoin CEM:ssäkin. R on siis riippumaton matriisien U ja D valinnasta.

Mikäli matriisiin D otetaan mukaan vain yksi endmember eli $D=d_1$ ja matriisista U tehdään tyhjä joukko eli analyysiin ei oteta mukaan yhtään ei-kiinnostavaa endmemberiä, niin TCIMF muuttuu CEM:ksi (ts. $w_{TCIMF} \rightarrow w_{CEM}$). [Chang 2002b]

5.2.4 Mixture Tuned Matched Filtering (MTMF)

CEM:n eräs ongelma on siinä, että se voi tuottaa vääriä positiivisia arvoja eli jotkin pikselit voivat saada korkeita CEM-arvoja, vaikka niiden spektreissä ei olisi jälkiä etsittävästä kohteesta. ”Mixture Tuned Matched Filtering” (MTMF) on yksi tapa saada CEM tuloksiin lisää luotettavuutta. MTMF:ssä on kaksi vaihetta: ensin suoritetaan normaali CEM prosessi, jonka jälkeen datapisteille lasketaan erillinen ns. ”mahdottomuusarvo”.



Kuva 5.3 [ENVI]: Graafinen esitys "Mixture Tuned Matched Filtering" nimisestä unmixing menetelmästä. Menetelmä on sama kuin MF/CEM, sillä lisäyksellä että tutkittavalle datapisteelle lasketaan lisäksi ns. "mahdottomuusarvo" sen mukaan minkä kokoiseen hyperkartioon piste juuri ja juuri mahtuu.

Kuvassa 5.3 on graafinen esitys MTMF:stä. Ylävasemmalla oleva datapiste kuvaa spektriä, jonka pelkkä CEM analysoisi väärin. MTMF menetelmä antaa CEM arvon lisäksi datapisteille toisen arvon "mahdottomuusarvon", jolla voidaan eliminoida virheellisesti analysoitujen pikseleiden määrää suhteessa pelkkään CEM menetelmään.

"Mahdottomuusarvot" määräytyvät sisäkkäisten hyperkartioiden perusteella. Kuvassa 5.3 on kolme kartiota kolmiulotteisessa avaruudessa. Pienimmän kartion pohjan ääriviiva määräytyvät datan keskihajonnan (sigma) perusteella ja kartion huipun ääriviiva kohinan keskihajonnasta. Isompien kartioiden vastaavat ääriviivat ovat keskihajontojen (sigmojen) moninkertoja. Kun kartiot on määritelty, niin datapisteen "mahdottomuusarvo" määräytyy pienimmän sigman omaavan kartion mukaan, johon datapiste juuri mahtuu. Jotta kartion huipun ääriviiva voitaisiin määrittää oikein, niin datan kohina pitää olla isotrooppista jokaiseen suuntaan eli datan kohina pitää valkaista.

5.2.5 Foreground–Background Analysis (FBA)

FBA on käytännössä sama algoritmi kuin CEM, erona on, että yhtälön (5.3) R^{-1} määritetään ohjatusti. CEM:ssä R^{-1} määräytyy tutkittavasta datasta suoraan, mutta FBA:ssa käyttäjän pitää itse määritellä se data (background data), josta R^{-1} lasketaan. [Harsanyi]

Koska käyttäjä voi itse määritellä taustaspektrit FBA:ssa, niin silloin tulosten luotettavuuden kannalta etsittävää materiaalia sisältäviä pikseleitä voi olla mukana hyvinkin paljon, toisin kuin CEM:ssä. Toisaalta FBA:n käyttäminen vaatii jonkinlaista taustatietoa tutkittavasta alueesta, jotta background data määriteltäisiin oikein.

5.2.6 Orthogonal Subspace Projection (OSP)

Orthogonal Subspace Projection on unmixing menetelmä, jonka idea on samanlainen kuin yhtälön (5.1a) eli pikselit koostuvat äärellisestä joukosta endmembereitä, jotka pitäisi määrittää ja tietää mahdollisimman hyvin ennen unmixing prosessia. OSP lähtee siitä, että pikseleiden endmemberit jaetaan kahteen ryhmää: kiinnostavaan ja jäljelle jääviin endmembereihin. Sama asia matemaattisesti ilmaistuna, eli kirjoitetaan yhtälö $r = M\alpha+n$ (5.1a) uudessa muodossa:

$$r = d\alpha_p + U\gamma + n, \qquad (5.1b)$$

jossa d on etsittävän materiaalin spektri, α_p on vastaavan spektrin painokerroin havaitussa spektrissä r. U on matriisi ei-etsittävistä endmembereistä ja γ on niiden painokerroinmatriisi. Matriisi n on kohinatermi, se voidaan myös tulkita mittaus- tai mallinnusvirheeksi. Yhtälö (5.1a) on siis jaettu kahteen osaan: kiinnostavaan endmemberiin (d) ja ei-kiinnostaviin endmembereihin (U). Näin siksi, että voidaan muodostaa projektori $P_U^{\perp}$, joka projisoi r:n avaruuteen, joka on ortogonaalinen U:n virittämän avaruuden <U> kanssa. Toisin sanoen $P_U^{\perp}$ tuhoaa r vektorista U:n vaikutuksen. $P_U^{\perp}$ saadaan yhtälöllä

$$P_U^{\perp} = I - UU^{\#}, \qquad (5.10)$$

jossa I on identiteettimatriisi ja $U^{\#}=(U^TU)^{-1}U^T$ on U:n pseudo-käänteismatriisi. Kertomalla yhtälö (5.1b) puolittain $P_U^{\perp}$:lla saadaan ei-kiinnostavat endmemberit projisoitua pois:

$$P_U^{\perp} r = P_U^{\perp} d\alpha_p + P_U^{\perp} n. \qquad (5.11)$$

Yhtälö (5.11) on yleinen signaalin havaitsemisongelma. Mikäli signaali-kohina suhde valitaan kriteeriksi, optimaalisin ratkaisu löydetään ”matched filter” operaattorilla M_d,

$$M_d (P_U^{\perp} r) = \kappa\, d^T (P_U^{\perp} r), \qquad (5.12)$$

Yhtälössä esiintyvä vakio κ voidaan asettaa ykköseksi eli $\kappa=1$, jolloin OSP-operaattori on

$$P_{OSP} = d^T P_U^{\perp}. \qquad (5.13)$$

Sama asia kirjoitettuna OSP-filtteriksi:

$$\delta_{OSP}(r) = P_{OSP}\, r = d^T P_U^{\perp} r = (d^T P_U^{\perp} d)\, \alpha_p + d^T P_U^{\perp} n. \qquad (5.14)$$

Edellinen tiivistettynä: OSP on kaksivaiheinen prosessi, jossa ensin operaattorilla $P_U^{\perp}$ ”puhdistetaan” tutkittavasta spektristä U:n vaikutus pois. Tämän jälkeen puhdistetusta signaalista eli yhtälöstä (5.11) projisoidaan kohdespektrillä d irti unmixing tulos. [Chang 2002b], [Du], [Ren], [Brumbley]

Kirjallisuudessa (mm. [Brumbley], [Wang], [Tu], [Chang 1998]) esitetään, että lineaarisesta mixing mallista (5.1a-b) pitäisi tietää matriisin M lisäksi endmembereiden painokertoimet α_i:t etukäteen, jotta OSP:tä voitaisiin käyttää. Tästä syystä sanotaan että OSP on priorinen menetelmä. Käytännössä painokertoimien prioritieto on kuitenkin mahdotonta, minkä vuoksi OSP (yhtälö (5.13)) ”pitää” laajentaa postprioriseksi menetelmäksi, jossa painokertoimet arvioidaan suoraan tutkittavasta datasta. Seuraavissa kappaleissa esitellään kolme postpriorista OPS menetelmää: SSC, OBC ja TSC.

Signature Subspace Classifier (SSC)

SSC menetelmän operaattori ja filtteri ovat:

$$P_{SSC} = P_{OSP}\, P_M = P_{OSP}\, M(M^T M)^{-1} M^T = P_{OSP}\, MM^{\#}, \qquad (5.15a)$$

$$\delta_{SSC}(r) = P_{SSC}\, r = P_{OSP}\, P_M\, r = P_{OSP}\, M\alpha_{PNS}. \qquad (5.15b)$$

Operaattori P_M projisoi r vektorin avaruuteen <M> ja samalla vaimentaa kohinan n muotoon $P_M n$ [Wang], [Brumbley]. Toisaalta operaatio $P_M r$ tuottaa tuloksen $M\alpha_{PNS}$ eli on sama kuin yhtälön (5.2) kautta saatu PNS-unmixing ratkaisu täydellisestä unmixing prosessista.

Oblique Subspace Classifier (OBC)

OBC operaattori on

$$P_{OBC} = (d_T\, P_U^{\perp}\, d)^{-1}\, P_{OSP}, \qquad (5.16)$$

joten P_{OSP} ja P_{OBC} eroavat vain luvulla $(d_T\, P_U^{\perp}\, d)^{-1}$ [Wang] [Tu], eli operaattoreiden ero liittyy vain unmixing tulosten normittamiseen.

Target Subspace Classifier (TSC)

Sen sijaan, että projisoitaisiin aluksi vektori r aliavaruuteen <M>, kuten SSC:ssä, ”Target Subspace Classifier” projisoi spektrin r suoraan etsittävän endmemberin aliavaruuteen <d> [Brumbley]. Eli yhtälöjen (5.15a) ja (5.15b) P_M korvataan operaattorilla P_d:

$$P_{TSC} = P_{OSP}\, P_d = P_{OSP}\, d(d^T d)^{-1} d^T, \qquad (5.17a)$$

$$\delta_{TSC}\,(r) = P_{TSC}\, r = P_{OSP}\, P_d\, r. \qquad (5.17b)$$

Mikäli etsittävä spektri d ja ei-etsittävien spektrien avaruus <U> ovat keskenään ortogonaalisia, niin P_{OSP}-operaattorissa ei ole tarvetta käyttää $P_U^{\perp}$ operaattoria [Brumbley], jolloin

$$P'_{TSC} = d^T P_d = d^T d(d^T d)^{-1} d^T = I\, d^T = d^T, \qquad (5.18a)$$

$$\delta'_{TSC}\,(r) = d^T r\,. \qquad (5.18b)$$

Kuitenkin <d> ja <U> eivät ole yleisesti ortogonaalisia. Tämä on myös TSC:n haitta, koska spektrin suora projisoiminen <d> avaruuteen ottaa mukaan komponentin <U> avaruudesta [Chang 1998].

Yhteenveto SSC, OBC, TSC ja OSP menetelmistä

Chang et al. [Chang 1998] ovat vertailleet SSC, OBC ja TSC menetelmiä toisiinsa, ja todenneet että SSC ja OBS ovat samoja algoritmejä, vaikka ne on johdettu eri tavoin. TSC:stä he ovat todenneet, että se toimii teoriassa paremmin kuin SSC/OBC, mutta käytännössä huonommin, koska <d> avaruus ei ole yleisesti ortogonaalinen ei-toivottujen signaaliavaruuksien kanssa.

Settle on todistanut [Settle], että OSP (ja siten myös OBC ja SSC) ja PNS-unmixing (5.2) ovat matemaattisesti identtisiä, mikäli kohina on valkoista, gaussista ja keskiarvoltaan nolla. Myös mm. Chang et al. [Chang 1998] toteavat, että PNS-unmixing, OSP, SSC ja OBC ovat samoja luokittelijoita. OSP operaattorin esittelyn yhteydessä (yhtälöt (5.12-5.13)) esiintyi vakio κ, joka silloin asetettiin ykköseksi. OBC esittelyn yhteydessä todettiin, että OBC:n ja OSP:n ero liittyi normitustekijään $(d^T P_U^{\perp} d)^{-1}$, joka on käsittelemämme vakio κ:n teoreettinen arvo. Kuten aiemmin on todettu, OSP on periaatteessa priorinen menetelmä, jossa kiinnostavan endmemberin painokerroin (α_p) täytyisi tietää/arvata etukäteen. Tästä näkökulmasta katsottuna κ liittyy priorisen α_p arvauksen virhearvioon [Chang 2002b].

OSP menetelmää voidaan tarkastella kahdesta näkökulmasta: kiinnostavan endmemberin pitoisuuden arviointi tai kiinnostavan endmemberin havaitseminen mitatusta spektristä, ero liittyy normitusvakioon κ. Koska vakiolla κ ei ole mitään vaikutusta OSP:n havainnointikykyyn, voidaan asettaa että $\kappa=1$. Tällöin myös priorisen OSP menetelmän yhteydessä voidaan unohtaa vaatimus, että painokertoimet (α_p, γ) tulisi tietää/arvata etukäteen. Jälkikäteen unmixing tulokset voidaan aina normittaa analysoijan itsensä määrittämällä luvulla esim. välille 0-1.

Koska OSP ja PNS-unmixing ovat teoriassa samoja menetelmiä, on syytä kysyä mitä etuja OSP:llä on suhteessa PNS-unmixingiin? Chang [Chang 2002b] on esittänyt, että koska kohina ei ole yleisesti gaussista, niin OSP on tässä suhteessa parempi kuin PNS-unmixing.

5.2.7 OSP vs. CEM

Matemaattisesti on todistettu [Du] [Chang 2002b], että OSP ja CEM ovat samat menetelmät, kun kohina on riippumatonta ja identtisesti jakautunutta (valkoista) kanavien suhteen sekä kohinan varianssi verrattuna signaaliin on pieni, ts. signaali-kohina suhde on tarpeeksi suuri. OSP algoritmissä lähtötietoina tarvitaan kiinnostava spektri d ja muut, ei-kiinnostavat endmemberit U. Matriisin U avulla saadaan operaattori $P_U^{\perp}$, joka projisoi mitatun spektrin avaruuteen <d> eli poistaa (annihiloi) mitatusta spektristä ei-kiinnostavien endmembereiden vaikutuksen. CEM menetelmä toimii samalla tavalla, mutta koska matriisia U ei (tarvitse) tunneta, niin $P_U^{\perp}$ arvioidaan tutkittavasta datasta suoraan. Tämä annihiloiva operaattori ($P_U^{\perp}$) saadaan (auto)korrelaatiomatriisin R käänteismatriisina (yhtälöt (5.7a-b)). Eli kun unohdetaan normitustekijä $(d_T P_U^{\perp} d)^{-1}$ ja muistetaan edellä esitetyt vaatimukset kohinalle, voimme teoreettisessa mielessä kirjoittaa, että $P_U^{\perp} \equiv R^{-1}$. Vaikka kohina olisikin olettamusten mukaista, niin käytännössä CEM menetelmässä ohjaamattomasti arvioidut U_{CEM} endmemberit eivät ole samoja kuin OSP:ssä itse määritellyt U_{OSP} endmemberit.

OSP on johdettu olettamuksella, että kohina on valkeaa, kun taas CEM ei aseta mitään ehtoja kohinalle. Du et al. [Du] ovat todenneet, että CEM:n havainnointikyky on yleisesti parempi kuin OSP:n, mikäli kohina ei ole valkeaa ja kun analysoitavissa spektreissä on jälkiä sellaisista endmembereistä, joita ei ole otettu mukaan annihilaatio-operaattorin $P_U^{\perp}$ (5.10) muodostamiseen. Tästä syystä CEM on parempi vaimentamaan kohinaa ja toimii paremmin tilanteessa, jossa ei-etsittävät endmemberit ovat hankala määrittää tai varsinkin jos ne ovat tuntemattomia.

Koska CEM:ssä spektrinen tausta arvioidaan suoraan pikseleiden spektreistä, ei ennalta määritellyistä puhtaiden materiaalien spektreistä, kuten OSP:ssä, niin CEM ei vaadi taustan spektreiltä lineaarista mixing mallia, toisin kuin OSP [Harsanyi] [Chang 2002b]. Asian voi ymmärtää niin, että CEM:ssä pikseleistä määritetty spektrinen tausta on yksi endmember ja etsittävä kohde toinen endmember. CEM vaatii kuitenkin sen, että tausta ja kohde ovat lineaarisesti sekoittuneita.

CEM:n haitta on siinä, että sillä on huonompi yleistämiskyky: hieman erilainen spektrijälki suhteessa etsittävään signaaliin d, voidaan helpommin tulkita ei-etsittäväksi tai tunnistamattomaksi. Näissä tilanteissa voi olla hyödyllistä käyttää OSP:tä CEM:n sijasta, kunhan ensin parantaa OSP:n suorituskykyä. Tämä voidaan tehdä valkaisemalla kohina tai arvioimalla datan ei-etsittävät endmemberit ohjaamattomasti tai semi-ohjaamattomasti [Du]. Seuraavassa yhteenveto CEM:n vahvuuksista ja heikkouksista suhteessa OSP:hen.

CEM:

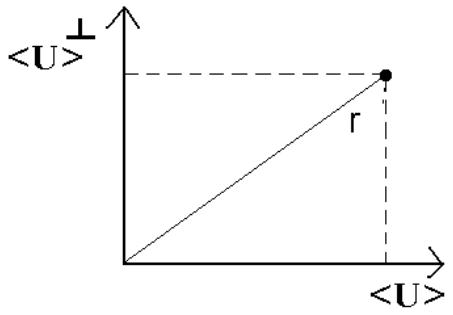
+ ei vaadi prioritietoa taustasta, vain kohteen spektri riittää prioritiedoksi

+ toimii paremmin kun kohina ei ole valkoista

+ tausta voi olla myös epälineaarinen sekoitus puhtaista materiaaleista

– CEM:llä on huonompi yleistämiskyky kuin OSP:llä

– ei toimi kun etsittävä kohde on yleinen kuva-alalla (kohde muuttuu taustaksi)



Kuva 5.4: OSP graafisesti havainnollistettuna. Ei-etsittävistä endmembereistä U muodostetaan annihiloiva operaattori $P_U^{\perp}$, joka projisoi analysoitavan spektrin (r) avaruuteen, joka on ortogonaalinen U:n virittämän avaruuden <U> kanssa. Tämän jälkeen kohdespektrillä d projisoidaan <U>⊥ avaruudesta irti unmixing tulos. CEM toimii myös samalla tavalla. Siinä <U> korvataan koko datasta määritettävällä avaruudella. Kuvassa 5.2 CEM havainnollistettiin kanava-avaruudessa, kun tässä käytetään endmembereiden virittämää piirreavaruutta.

5.2.8 RX

RX algoritmi on tarkoitettu anomalisten spektrien etsimiseen. Menetelmä on ohjaamaton, koska ennalta ei tarvitse määrittää tai tietää yhtään spektriä. Oikeastaan RX ei ole unmixing algoritmi, koska tarkoituksena ei ole selvittää minkään materiaalin osuutta spektrissä, vaan etsiä poikkeavia pikseleitä. RX filtteri on

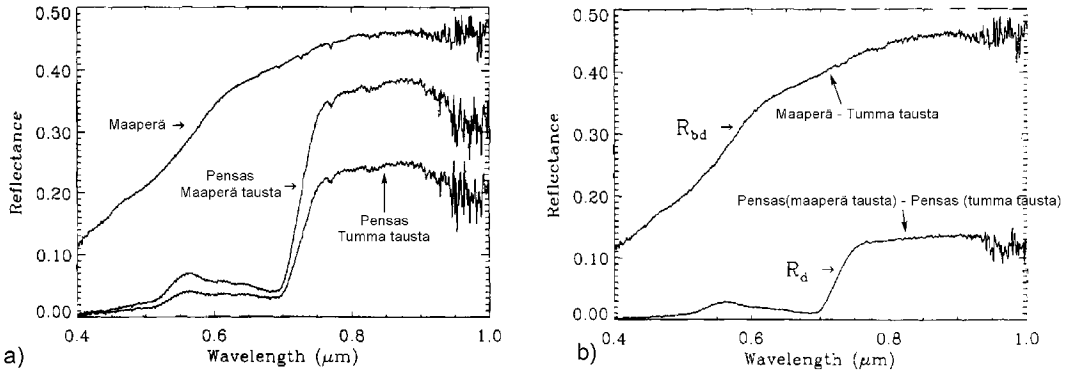
$$\delta_{RX}(r) = r^T R^{-1} r, \qquad (5.19)$$

joka on muuten sama kuin CEM filtteri (yhtälö (5.3)), paitsi että etsittävän kohdespektrin transpoosi d^T on vaihtunut tutkittavan spektrin transpoosiin r^T. Ja koska anomalian etsintä ei ole ensisijaisesti pitoisuuden arviointia, niin CEM ja RX algoritmien vertailussa CEM filtterin normitustekijä $(d_T R^{-1} d)^{-1}$ voidaan unohtaa. [Chang 2002b]

5.3 Epälineaariset unmixing menetelmät

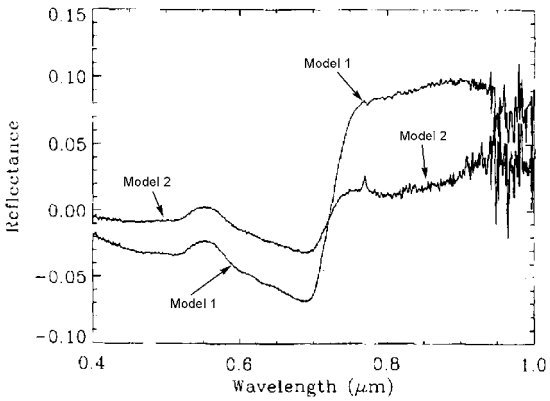
Vaikka mitatuissa spektreissä olisikin mukana epälineaarisia sekoituksia puhtaista materiaaleista, lineaaristen mallien suosio perustuu pitkälti siihen että ne näyttävät toimivan hyvin monissa tapauksissa [Tu]. Tässä tutkimuksessa ei tutustuta analyyttisiin epälineaarisiin tekniikoihin, vaan esitellään yksi tutkimus [Ray], jossa on havainnollistettu epälineaarisuustekijöiden vaikutuksia mittaustulosten ja lineaarisen mallinnuksen välillä sekä esitelty korjausehdotus normaaliin lineaariseen unmixing tekniikkaan.

Ray et al. [Ray] ovat tutkineet autiomaan (Manix Basin of the Mojave Desert, USA) creosote pensaan ja maaperän epälineaarista sekoittumista. Tutkimuksissa mitattiin kannettavalla spektroradiometrillä (ASD 2, spektrialue 0.338-1.064 nm, 512 kanavaa) kolmen eri kohteen reflektanssi: 1) pensas maaperää vasten, 2) maaperä ja 3) pensas kolmen prosentin heijastusstandardia (lähes heijastamaton) vasten. Nämä kolme reflektanssia ovat kuvassa 5.5a.



Kuva 5.5 [Ray]: a) Maaperän reflektanssi, creosote pensaan reflektanssi maaperää vasten mitattuna ja pensaan reflektanssi tummaa taustaa vasten mitattuna. b) Mittaustuloksista lasketut reflektanssien erotukset. Ylempi kuvaaja on maaperän reflektanssi miinus tumman taustan reflektanssi ja alempi kahdella eri tavalla mitatun pensaan reflektanssien erotus.

Tarkastellaan kuvaa 5.5b. Mikäli pensas ja maaperä olisivat pelkästään lineaarisesti sekoittuneita, niin kuvaajien (R_d ja R_{bd}) pitäisi olla toistensa moninkertoja. Kuvaajasta voi selvästi huomata, että näin ei ole (ts. vakio•$R_d \neq R_{bd}$). Ray et al. ovat myös vertailleet kahden lineaarisen mallin soveltuvuutta mitatun pensas-maaperä vuorovaikutteisen reflektanssin unmixingiin. Malli 1 on perinteinen tapa, jossa endmembereinä ovat pensas tummaa taustaa vasten ja maaperä. Malli 2 on myös kaksiendmemberinen, jossa ensimmäisenä komponenttina on pensas tummaa taustaa vasten ja toisena maaperä plus yhdistetty pensas-maaperä komponentti. Malli 2:n toinen endmember on siis muotoa ”maaperä+maaperä•pensas”, joka on määritetty kuvan 5.5b tiedoista. Kuvaan 5.6 on piirretty unmixing mallien virheet aallonpituuden funktiona suhteessa havaittuun pensas-maaperä vuorovaikutteiseen spektriin. Tulokset osoittavat että malli joka sisältää epälineaarisen maaperä-pensas vuorovaikutuksen (malli 2) toimii paremmin.



Kuva 5.6 [Ray et. al]: Kahden lineaarisen unmixing mallin jäännöstermit suhteessa havaittuun spektriin. Malli 1:n endmembereinä ovat pensas tummaa taustaa vasten ja maaperä. Mallin 2:n endmembereinä ovat maaperä ja kuvan 5.5b tiedoista johdettu endmember, joka sisältää epälineaarisen pensas-maaperä vuorovaikutuksen.

Jäännöstermien lisäksi Ray et al. ovat vertailleet unmixing mallien painokertoimia, jotka ovat taulukosta 5.1. Tässä yhteydessä niiden merkitystä ei kuitenkaan käsitellä syvällisemmin. Yhteenvedossa he ovat todenneet, että epälineaarisia termejä sisältävien endmembereiden käyttö tekee kasvillisuuden havaitsemisen helpommaksi, mutta toisaalta tällöin komponenttien pinta-alojen (pitoisuuksien) arviointi painokertoimien perusteella vaikeutuu. Toisaalta kohteen havaitsemisessa komponenttien pinta-alojen arviointi ei ole oleellista, tärkeämpää on lähestyä unmixing ongelmaa siten, että etsittävä kohde tulisi mahdollisimman havaittavaksi.

Taulukko 5.1[Ray]: Unmixing tuloksista saadut komponenttien pitoisuusarvot. Malli 1:ssä arvot ovat absoluuttisia, kun taas malli 2:ssa painokertoimien perusteella saadut arvot on normitettu sataan prosenttiin.

Malli	Pensas	Maaperä	pensas-maaperä vv.
pensas ja maaperä	79.6 %	20.3 %	-
pensas ja pensas-maaperä vuorovaikutus (vv.)	22.5 %	3.3 %	74.1 %

Edellä esitellyn unmixing tekniikan soveltaminen eri sovelluksiin voi olla vaikeaa. Ray et al. ovat mittaustuloksiin pohjautuen todenneet, että pensaiden rakenne on tärkeässä osassa epälineaarisessa unmixing mallinnuksessa. Toisin sanoen epälineaaristen mallien käyttäminen on tapauskohtaista, jopa samojen komponenttien (pensas, maaperä) kesken. Tämä on odotettu tulos, koska valon heijastusten määrä säteilylähteen (aurinko) ja sensorin välillä, ja sitä kautta epälineaaristen vaikutusten syntymekanismit (kuva 5.1) riippuvat materiaalien makroskooppisista rakenteista ja keskinäisistä suhteista.

6. Puhtaan pikselin luokittelumenetelmät

Tässä luvussa käsitellään puhtaan pikselin luokitteluun tarkoitettuja algoritmejä. Ne on jaettu tilastollisiin ja kahta spektriä vertaileviin menetelmiin. Tilastollisille menetelmille tunnusomaista on se, että luokittelija opetetaan ohjaamattomasti, semi-ohjatusti tai ohjatusti käyttämällä opetusnäytteitä tai tutkittavaa dataa suoraan.

Menetelmien kategorisoiminen esim. luokittelu-, kohteen havainnointi-, ohjaamattomiin/ohjattuihin ja dimension vähennysmenetelmiin on käytännössä keinotekoista. Analysointialgoritmien valinta ei saisi perustua sokeaan valintaan eri kategorioiden väliltä. Algoritmien soveltaminen ja yhdisteleminen on tapauskohtaista, mitään yleispäteviä sääntöjä on vaikea luoda.

6.1 Tilastolliset luokittelijat

Tilastolliset luokittelijat on tapana jakaa ohjattuihin ja ohjaamattomiin menetelmiin. Ohjaamattomia menetelmiä kutsutaan usein myös klusteroinniksi.

6.1.1 Maximum likelihood (ML)

Maximum likelihood (suurin uskottavuus) on yksi tunnetuimmista kaukokartoitusdatan luokittelijoista. Varsinaisesti termi “maximum likelihood” ei ole minkään algoritmin nimi, vaan kyseessä on enemmänkin yleisnimitys tilastollisesta luokittelijasta. ML perustuu olettamukseen, että eri luokkien käyttäytyminen piirreavaruudessa voidaan mallintaa johonkin todennäköisyysjakautumaan (parametrilliseen tai ei-parametrilliseen) perustuvilla tiheysfunktioilla, jotka määritellään opetusnäytteiden avulla. Kun tiheysfunktiot on määritelty, tutkittava spektri luokitellaan siihen luokkaan, jossa suurin uskottavuus toteutuu. Suurin uskottavuus ei ole suoranaisesti sama kuin suurin todennäköisyys, ero liittyy kynnysarvon (μ) valintaan. Yleisin kynnysarvon määrittelijä on Bayesin luokitin, muita menetelmiä on mm. Neyman-Pearson ja minimax-sääntö.

Opetusnäytteistä arvioituja luokkakohtaisia tiheysfunktioita on tapana merkitä symbolilla $P(x|\omega_i)$, missä ω_i tarkoittaa luokkaa i ja x on spektri. Bayesian luokittimen mukaan x luokittuu siihen luokkaan, jonka posterioritodennäköisyys

$$P(\omega_i \mid x) = \frac{P(\omega_i)\, P(x \mid \omega_i)}{P(x)} = \frac{P(\omega_i)\, P(x \mid \omega_i)}{\sum_{i=1}^{M} P(\omega_i)\, P(x \mid \omega_i)} \tag{6.1}$$

on suurin. $P(\omega_i)$ on luokan ω_i prioritodennäköisyys ja P(x) on normitustekijä. Bayesian luokittelusääntö voidaan kirjoittaa myös muodossa

$$x \in \omega_i \quad jos \quad \frac{P(x \mid \omega_i)}{P(x \mid \omega_j)} > \frac{P(\omega_j)}{P(\omega_i)} =: \mu \text{, kaikille } j \neq i. \tag{6.2}$$

(Normitustekijä P(x) on supistettu pois.) Prioritodennäköisyyksien $P(\omega_i)$ käyttäminen on tapauskohtaista, koska ne liittyvät luokittelupäätöksen kynnysarvoihin ja niillä ei ole vaikutusta tiheysfunktioiden arvoihin. Tyypillinen esimerkki prioritodennäköisyyksien käytöstä on seuraava esimerkki metsän luokittelusta. Tutkittavasta alueesta tiedetään etukäteen, että 80% pinta-alasta on havumetsää ja loppu 20% lehtimetsää. Tällöin luokkien prioritodennäköisyydet ovat: P(havu)=0.8 ja P(lehti)=0.2.

Kohteen havaitsemissovellus on prioritodennäköisyyden kannalta hyvin erilainen tilanne kuin esim. metsien luokittelu. Kohteen havaitsemisongelmissa pinta-aloihin perustuva prioritodennäköisyyksien käyttö saisi aikaan paljon vääriä negatiivisiä havaintoja. Mikäli kynnysarvo μ (yhtälö (6.2)) asetetaan ykköseksi, luokittelutulos määräytyy suoraan suurimman todennäköisyystiheyden $P(x|\omega_i)$ mukaan.

Usein puhekielessä maximum likelihood luokittelijalla, ilman eri mainintaa, tarkoitetaan normaalijakautuneiden tiheysfunktioiden käyttöä. Normaalijakautuneesta tiheysfunktiosta on tarkemmin kerrottu kappaleessa 6.1.2. Muita tiheysfunktion $P(x|\omega_i)$ arviointimenetelmiä ovat mm. histogram-metodi, Parzen ja k-lähimmän naapurin sääntö, jotka kaikki ovat ei-parametrillisiä menetelmiä.

6.1.2 Statistiset etäisyysmitat

Tässä kappaleessa kerrotaan muutamasta ns. statistisesta etäisyydestä, joilla voidaan mitata piirreavaruuden pisteiden etäisyyttä toisistaan. Etäisyyksien avulla voidaan mitata mm. piirteiden valinnan ja ekstraktoinnin hyvyyttä tai luokitella dataa.

Normaalijakautuman tiheysfunktio

Normaali- eli gaussijakautunutta tiheysfunktiota ei mielletä statistiseksi etäisyydeksi, mutta koska sillä on yhteyksiä muihin jatkossa esiteltäviin etäisyyksiin, käsitellään se tässä. Monimuuttujainen normaalijakautunut tiheysfunktio on

$$P(x|\omega_i) = (2\pi)^{-N/2}|\Sigma_i|^{-½} \exp \{ -½(x-m_i)^T \Sigma_i^{-1} (x-m_i) \}, \qquad (6.3)$$

jossa m_i on luokan i keskiarvo, Σ_i on i:n kovarianssimatriisi ja N on dimensio [Richards]. Ongelma tiheysfunktion käytössä on se, että parametrien (m_i ja Σ_i^{-1}) määrittely tulee epätarkaksi (Hughesin ilmiö) ja luokittelun laskenta-aika kasvaa nopeasti piirteiden määrän kasvaessa [Chen 1996].

Ottamalla yhtälöstä (6.3) luonnollinen logaritmi, poistamalla vakiotermi $(2\pi)^{-N/2}$ ja kertomalla tulos kahdella, saadaan

$$g_i(x)= -\ln |\Sigma_i| - (x-m_i)^T \Sigma_i^{-1} (x-m_i). \qquad (6.4)$$

Yhtälö (6.4) on normaalijakautuneeseen tiheysfunktioon perustuva diskriminanttifunktio, joka on siis sama kuin gaussinen maximum likelihood luokittelija, kun luokkien prioritodennäköisyydet asetetaan kaikille samoiksi. Valintasääntö luokittelulle voidaan nyt kirjoittaa, että

$$x \in \omega_i \quad jos \quad g_i(x) > g_j(x) \quad \text{kaikille } j \neq i. \qquad (6.5)$$

Mahalanobis etäisyys

Mikäli eri luokkien kovarianssimatriisit ovat samoja eli $\Sigma_i = \Sigma$ kaikille i, niin normaalijakautumaan perustuva diskriminanttifunktio (6.4) on sama kuin Mahalanobis etäisyyden neliö m_i:n suhteen, joka kirjoitetaan muodossa [Richards]

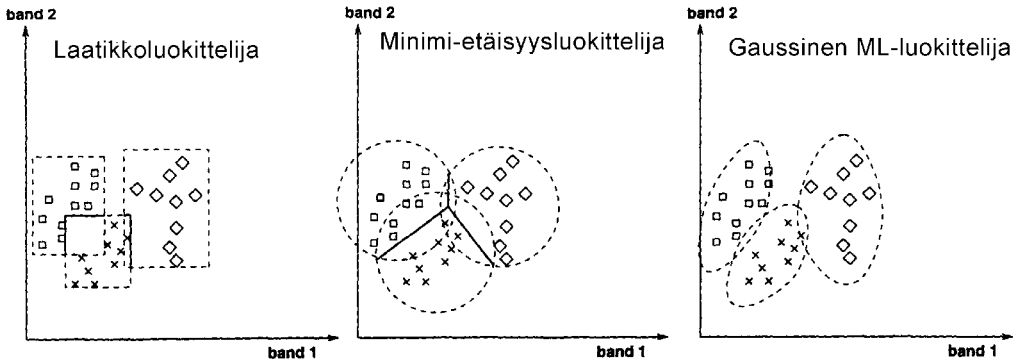
$$d(x,m_i)^2 = (x-m_i)^T \Sigma^{-1} (x-m_i). \qquad (6.6)$$

Euklidinen etäisyys

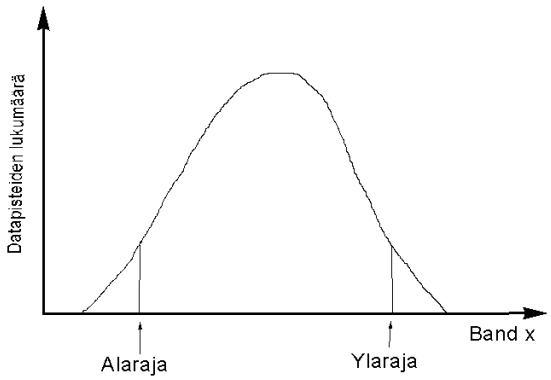
Jos Mahalanobis etäisyyden neliön yhtälöstä (6.5) poistetaan kovarianssimatriisi Σ^{-1}, niin tällöin kyseessä on arkielämästä tutun euklidisen etäisyyden neliö.

Euklidiseen etäisyyteen perustuvia luokittelumenetelmiä on mm. minimi-etäisyys- ja laatikkoluokittelija. Toisaalta esim. Mahalanobis etäisyyden käyttäminen minimi-etäisyysluokittelijassa on mahdollista. Kuvassa 6.1 on havainnollistettu minimi-etäisyys-, laatikko- ja gaussinen ML luokittelija. Minimi-etäisyysluokittelijassa datapiste luokittuu siihen luokkaan, jonka keskiarvo on lähimpänä.

Laatikkoluokittelijassa laatikon sijainti ja rajat voidaan määrittää joko automaattisesti keskiarvon ja keskihajonnan mukaan tai antaa omat rajat histogrammikuvaajan (kuva 6.2) perusteella. [Richards]



Kuva 6.1: Kolme ohjattua luokittelumenetelmää.



Kuva 6.2: Histogrammikuvaaja laatikkoluokittelijan määrittelyä varten. Kuvan mielivaltaisesti asetetut ala- ja ylärajat määrittelevät laatikon rajat piirrevektorin komponentillä x.

Jeffries-Matusita (JM) etäisyys

JM etäisyys kahden todennäkösyysjakauman välillä on

$$J_{ij} = \int_x \{\sqrt{P(x \mid w_i)} - \sqrt{P(x \mid w_j)}\}^2 dx, \qquad (6.7a)$$

jonka voi ajatella olevan keskimääräinen etäisyys kahden luokan tiheysfunktioiden välillä. Normaalijakaumassa JM etäisyys on

$$J_{ij} = 2(1-e^{-B}), \qquad (6.7b)$$

jossa B on Bhattacharyya etäisyys:

$$B = 1/8(m_i - m_j)^T \{\frac{\Sigma_i - \Sigma_j}{2}\}^{-1}(m_i - m_j) + ½\ln\{\frac{|(\Sigma_i - \Sigma_j)/2|}{|\Sigma_i|^{½}|\Sigma_j|^{½}}\}. \qquad (6.8)$$

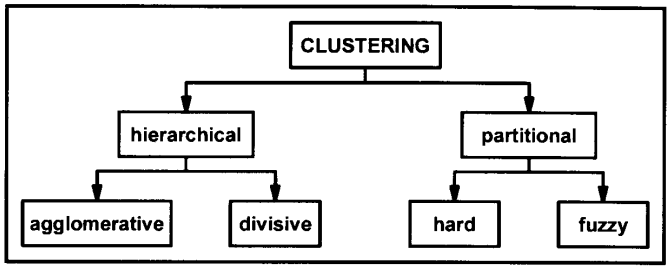
[Richards]

6.1.2 Ohjaamattomia luokittelumenetelmiä

Ohjaamattomien luokittelu- eli klusterointimenetelmien periaate on määrittää piirreavaruuteen ryhmät eli klusterit sen perusteella miten datapisteet ovat jakautuneet piirreavaruuteen. Klusterointialgoritmit voidaan jakaa kuvan 6.3 osoittamalla tavalla hierarkisiin ja ositteleviin menetelmiin [Wilson].

Hierarkisissa menetelmissä kukin pikseli muodostaa aluksi yhden ryhmän, jonka jälkeen naapuriryhmiä yhdisteellään isommiksi ryhmiksi kunnes jäljellä on vain yksi ryhmä (agglomerative) tai päinvastoin (divisive). Oleellista hierarkisissa menetelmissä on esittää prosessin kulku puurakenteen, ns. dendrogrammin avulla. Partitioivat menetelmät luovat datasta iteratiivisesti kiinteän määrän klustereita, niin että pikselit jaotellaan eri klustereihin minimoimalla (tai maksimoimalla) valittu kriteerifunktio.

Hierarkiset menetelmät ovat harvinaisempia ja käyttökelvottomampia kaukokartoitusdatan klusteroinnissa. Ne soveltuvat paremmin muiden tieteenalojen tuottamiin datoihin, jotka ovat luonteeltaan erilaisia kuin spektrinen data. Käytännön syy hierarkisten menetelmien heikkouteen on suuri datamäärä. Dendrogrammit ovat epäkäytännöllisiä suuriin datamääriin, ja suuri pikselimäärä suhteessa odotettavaan klustereiden määrään tekee menetelmästä laskennallisesti raskaan. [Wilson]



Kuva 6.3 [Wilson]: Suurin osa klusterointimenetelmistä voidaan kategorisoida hierarkisiin ja partitioiviin menetelmiin.

Seuraavassa esitellään kolme yleisintä partitioivaa klusterointialgoritmiä: K-Means, sumea K-Means ja ISODATA. Perinteisesti käytetty kriteerifunktio on euklidinen etäisyys, mutta myös muita kahden piirreavaruuden pisteen vertailuun tarkoitettuja funktioita voidaan käyttää. Esim. Sohn et al. [Sohn] ovat käyttäneet kriteerinä SAM (Spectral Angle Mapper) arvoa.

K-Means

K-Means (tai C-Means) on minimi-etäisyysluokittelijasta modifioitu iteratiivinen klusterointimenetelmä, jonka perusidea on seuraava. Alussa määritetään K kappaletta klustereita ja kullekin satunnainen keskiarvo piirreavaruudessa. Iteraatio aloitetaan luokittelemalla datapisteet klusteriin, jonka keskiarvo on lähimpänä (esim. euklidinen etäisyys). Tämän jälkeen klustereille lasketaan uudet keskiarvot, minkä jälkeen datapisteet luokitellaan uudelleen minimietäisyyden perusteella. Iteraatioprosessi lopetetaan, kun keskiarvot eivät muutu enää merkittävästi tai kun käyttäjän määrittelemä maksimi iteraatioiden määrä on saavutettu.

Wong et al. [Wong] ovat todenneet, että alussa valittavilla keskiarvoilla ja arvolla K voi olla suuri vaikutus K-Means klusteroinnin nopeuteen ja lopputuloksiin. Mikäli K on liian pieni, niin iteroidut klusterit eivät välttämättä ole spektrisesti kovin samanlaisia. Toisaalta jos K on liian iso, niin samantyyppiset kohteet voivat luokittua liian moneen klusteriin. Riippuu tietysti sovelluksesta, että mikä on optimaalisin arvo K:lle.

Edellä kerrottu K-Means tunnetaan nimellä Hard K-means (HKM). Menetelmästä on olemassa myös sumeaan logiikkaan perustuva Fuzzy K-means (FKM). FKM:ssä kukin datapiste kuuluu jokaiseen klusteriin tietyllä jäsenyysarvolla (0…1). Myös jokaisella datapisteellä on oma vaikutuksensa laskettaessa klustereiden keskiarvoja. Iteraation päätyttyä datapiste luokitellaan korkeimman jäsenyysarvon perusteella. Jäsenyysarvot saadaan yhtälöllä

$$F_i(x) = \frac{\dfrac{1}{d(x,m_i)}}{\displaystyle\sum_{j=1}^{K} \frac{1}{d(x,m_j)}} \tag{6.9}$$

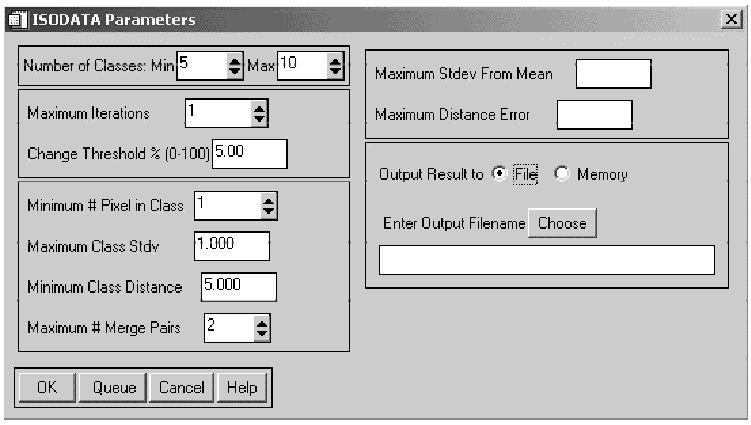
ja klustereiden painotetut keskiarvot yhtälöllä

$$m_i = \frac{\displaystyle\sum_{x=1}^{n} (F_i(x))^2 x}{\displaystyle\sum_{x=1}^{n} (F_i(x))^2} \; . \tag{6.10}$$

Yhtälössä $d(x,m_i)$ on etäisyys datapisteen x ja klusterin i keskiarvon välillä ja n on datapisteiden määrä. Koska nollalla ei voi jakaa, niin tilanteessa jossa x:n ja klusterin i keskiarvon välinen etäisyys on sama, luonnollinen valinta on että $F_i(x)=1$ ja $F_j(x)=0$, kun $j \neq i$. Automaattisesti toteutuva perusehto datapisteen x jäsenyysarvoille on $F_1 + F_2 + \ldots + F_K = 1$. [Törmä]

<u>ISODATA</u>

ISODATA (the Iterative Self-Organizing Data) on muuten sama kuin K-Means, sillä erolla että klustereita voidaan tarpeen mukaan yhdistellä, erotella tai kokonaan poistaa. Määrääviä tekijöitä klustereiden määrän muuttamiselle on esim. klusterin datapisteiden liian suuri hajonta ja liian monta/vähän pisteitä klusterissa. Lisäksi sekä K-Means että ISODATA klusteroinnissa voidaan asettaa ehtoja tilanteille, jolloin pikseleitä ei klusteroida ollenkaan, ts. ne luokitellaan ”jätteiksi”.



Kuva 6.4: ENVI:n ISODATA algoritmin parametrien asetusikkuna.

Kuvassa 6.4 on ENVI:n parametrien asetusikkuna ISODATA klusterointia varten. “Minimum # Pixel in Class” parametrillä voidaan määritellä milloin klusteri poistetaan, jolloin klusterittomat datapisteet (pikselit) luokitellaan lähimpään jäljellä olevaan klusteriin. “Maximum Class Stdv” kentällä määritellään tilanne, jossa klusteri jaetaan kahtia ja “Minimum Class Distance” kentällä tilanne, jossa kaksi klusteria yhdistetään. “Maximum # Merge Pairs” kenttä määrää kuinka monta klusteria voidaan yhdistää. Oikealla olevilla “Maximum Stdev From Mean” ja “Maximum Distance Error” kentillä voidaan asettaa vapaaehtoiset kynnysarvot sille, että datapiste jätetään klustereiden ulkopuolelle. [ENVI]

6.1.3 Ei-parametrilliset tiheysfunktiot

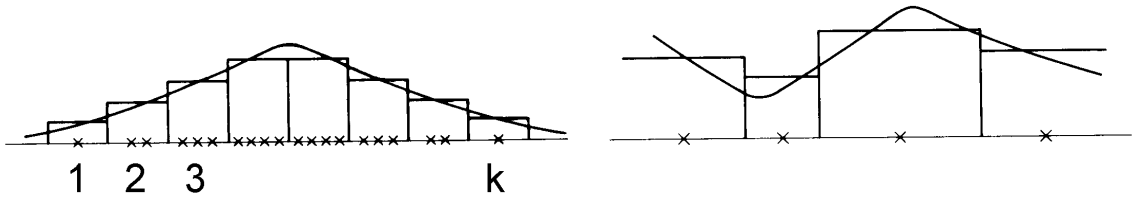
Parametrillisien tiheysfunktioiden käyttäminen perustuu siihen, että datan oletetaan toteuttavan jonkin todennäköisyysjakauman muodon. Kaukokartoituksessa parametrillinen menetelmä viittaa käytännössä normaalijakaumaan. Normaalijakautunut tiheysfunktio määräytyy datasta laskettavien parametrien (keskiarvo ja kovarianssimatriisi) perusteella. (Tilastotieteilijät käyttävät laskemisen sijasta termiä arvioiminen.) Normaalijakauman suosio kaukokartoitusdatan analysoinnissa ei ole osoitus siitä, että data olisi luonteeltaan normaalijakautunutta, vaan pikemminkin osoitus siitä, että menetelmä tuottaa hyväksyttäviä tuloksia.

Parametrillisten menetelmien sijasta voidaan käyttää ei-parametrillisiä menetelmiä, joissa datan ei oleteta noudattavan mitään ennalta tunnetun tiheysfunktion muotoa. Sana “ei-parametrillinen” on jokseenkin harhaanjohtava ja huono nimitys, koska menetelmissä käytetään parametreihin verrattavia karakteerejä, kuten esim. mediaania tai kernelin leveyttä. [Mosteller]

Seuraavassa käsitellään muutamia yleisimpiä kaukokartoituksessa käytettyjä ei-parametrillisiä tiheysfunktion arviointimenetelmiä. Ei-parametrillisten menetelmien haitta on siinä, että ne vaativat enemmän opetusnäytteitä kuin parametrilliset menetelmät, jotka nekin kärsivät Hughesin ilmiöstä. Ei-parametrillisiä menetelmiä voidaan soveltaa klusterointimenetelmiin.

Histogram metodi

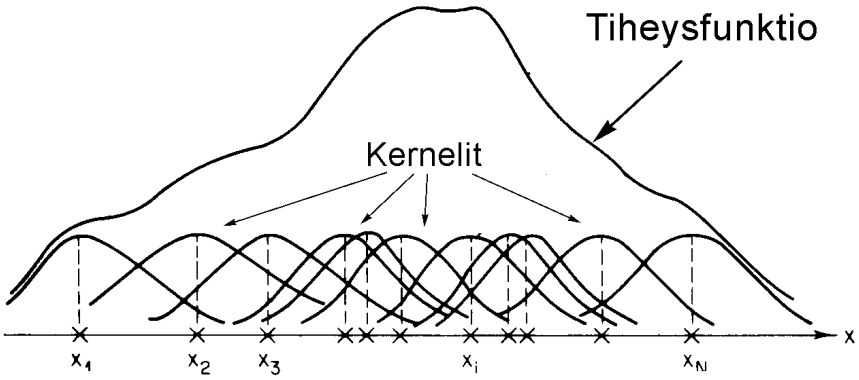
Yksinkertainen tapa arvioida ei-parametrillinen tiheysfunktio perustuu histogrammien käyttöön. Idea on jakaa piirreavaruus soluihin, jonka jälkeen lasketaan kuinka monta opetusnäytettä kussakin solussa on, josta saadaan arvio tiheysfunktiolle. Menetelmä on yksinkertainen laskea, mutta haittana on se, että solujen määräksi tulee k^d kappaletta, kun d on dimensio ja jokainen avaruuden akseli on jaettu k:hon osaan. Solujen määrää voi vähentää käyttämällä erikokoisia soluja, vakiokokoisten sijasta. Kuvassa 6.5 on esimerkki tiheysfunktion määrittämisestä yhdessä dimensiosta sekä kiinteällä että muuttuvalla solukoolla. [Fukunaga 1972]



Kuva 6.5 [Fukunaga 1972]: Histogrammiin perustuva menetelmä tiheysfunktion määrittämistä varten. Kuvaajien pystyakseli on opetusnäytteiden lukumäärä. Vasemmalla on kiinteä ja oikealla muuttuva solukoko.

Parzen

Parzen estimaatiossa tiheysfunktio muodostetaan kernelfunktioiden summana (kts. kuva 6.6). Kerneleitä on olemassa montaa eri tyyppiä, yhteistä niille kaikille on se, että niiden keskipiste ja maksimiarvo on x_i:ssä ja ne laskevat monotonisesti kuljettaessa x_i:stä poispäin. Yleinen ongelma Parzen estimaatiossa on se, että kerneleiden leveysparametrille, ts. etäisyydelle $|x_i - x_{i+1}|$ ei ole olemassa teoreettistä määritystapaa [Landgrebe 1/93 pp. 95-96]. Landgrebe et al. [Landgrebe 1/93] ovat määritelleet tutkimalleen datalle optimaalisen leveyden kokeellisesti, mutta sitä (leveyttä) ei voi yleistää kaikkiin tapauksiin.



Kuva 6.6 [Fukunaga 1972]: Parzen tiheysestimaattorin idea. Tiheysfunktio on summa kernelfunktioista, joiden keskipiste on x_i. Kuvan kernelit ovat normaalijakaumaan perustuvia.

K-lähimmän naapurin menetelmä (KNN)

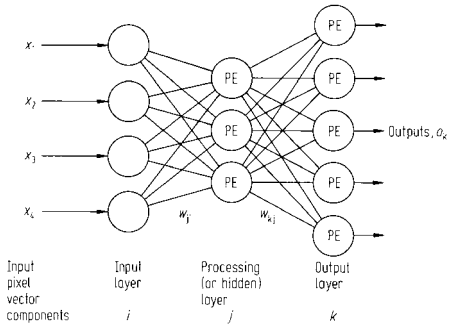
Tilastotieteellisesti KNN (K nearest neighbour) menetelmällä voidaan arvioida tiheysfunktioita. Analysoinnissa tämä ei kuitenkaan ole välttämätöntä. KNN:n idea on luokitella tutkittava datapiste lähimmän tai K lähimpien opetusnäytteiden perusteella eli datapiste kuuluu todennäköisimmin siihen luokkaan, jonka opetusnäytteitä on K lähimpien opetusnäytteiden joukossa. Menetelmään voidaan asettaa ehto, että K lähimpien pisteiden joukossa täytyy olla yksi tai useampi opetusnäyte tietystä luokasta, jotta tutkittava voidaan luokitella kyseiseen luokkaan. [Fukunaga 1972]

KNN on periaatteeltaan sama kuin K-Means klusterointi. KNN:ssä datapiste luokitellaan lähimmän/lähimpien opetusnäytteiden perusteella, kun K-Means luokittelee (iteraatiokierroksessa) datapisteen lähimpään klusteriin.

6.1.4 Neuroverkot

Neuroverkot ovat nykyisin suosittu tutkimusala data-analyysisovelluksissa ja signaalinkäsittelyssä. Neuroverkkoja tutkittiin jo 1960-luvulla, mutta varsinainen innostus verkkoihin on lisääntynyt 1990-luvulta lähtien tietokoneiden laskentatehon kasvun myötä. Kaukokartoitusdatan käsittelyssä verkkoja voidaan käyttää esim. piirteiden ekstraktointiin (mm. [Landgrebe 1/93]) ja luokitteluun. Neurolaskenta on sukua muille statistisille menetelmille, koska idea on löytää annettujen spektrien avulla piirreavaruudesta päätösalueet kullekin luokalle (ohjattu luokittelu) tai klusteroida piirreavaruus osiin (ohjaamaton luokittelu). Uutuutena on tapa, jolla muuttujien väliset (epälineaarisetkin) suhteet opitaan. [NNRC]

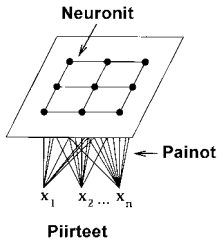
Neurolaskenta jäljittelee ihmisaivojen toimintaa, joka perustuu lukemattomien yksinkertaisten neuronien kytkeytymistä toisiinsa synapsien (painokertoimien) välityksellä. Kuvassa 6.7 on ohjattuun oppimiseen perustuva monikerroksinen perceptron verkko, jossa PE:t kuvaavat verkon laskentayksiköitä eli neuroneita ja viivat painokertoimia. Verkko toimii niin että sisään syötetään piirrevektoreiden komponentit (x_i:t). Neuroneissa piirteistä (syötteistä) lasketaan painokertoimilla painotettu summa, josta neuronin aktivaatiofunktio laskee ulostulon seuraavan kerroksen neuronien syötteiksi. Verkon opettaminen on iteratiivinen prosessi, jossa verkon painokertoimia muutetaan opetusalgoritmin perusteella sellaisiksi, että verkon antamat ulostulot vastaisivat syötteinä annettuja opetusnäytteitä. [Richards]



Kuva 6.7: Monikerroksisen perceptron verkon (multilayer perceptron network, MLP) arkkitehtuuri.

Ohjaamattomassa menetelmässä verkko mukautuu siten, että neuronit kilpailevat keskenään siitä, mikä edustaa kutakin syötettä parhaiten. Tämän vuoksi ohjaamattomasta menetelmästä käytetään myös termiä kilpailuoppiminen [Tekes]. Tunnetuin ohjaamaton verkko on SOM (self-organizing map). Toinen tunnettu ohjaamaton verkko on oppiva vektorikvantisaatio eli LQV. Kuvassa 6.8 on havainnollistus SOM verkon rakenteesta. SOM eroaa muista ohjaamattomista verkoista mm. siten, että myös voittajaneuronin naapuruston painoja muutetaan. Opetusprosessin jälkeen verkko toimii niin, että samaan ryhmään kuuluvat spektrit kuvautuvat samaan neuroniin tai vierekkäisten neuroneiden muodostamaan naapuristoon. [Törmä 1997]

SOM:ia voidaan käyttää myös ohjattuun luokitteluun. Tällöin verkko opetetaan normaalisti käyttäen tunnettuja näytteitä. Verkkoa voidaan hienosäätää LQV:llä. [Törmä 1997] Mm. Martínez et al. [Martìnez 2000], [Martìnez 2001a], [Martìnez 2001b], [Martìnez 2001c] ovat käyttäneet SOM:ia hyperspektrisen datan käsittelyssä.



Kuva 6.8: SOM verkon rakenteen idea.

6.2 Kahden spektrin vertailuun perustuvat menetelmät

Edellisissä kappaleissa käsitellyt tilastolliset luokittelumenetelmät vaativat pääsääntöisesti suurehkon joukon opetusnäytteitä. Tässä kappaleessa käsitellään algoritmejä, jotka vertailevat kahta datapistettä keskenään (edellä käsitelty euklidinen etäisyys on tässä suhteessa samanlainen).

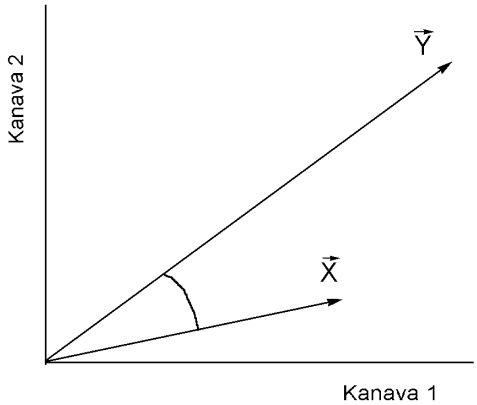
6.2.1 Spectral Angle Mapper (SAM)

Spectral Angle Mapper luokittelija perustuu kahden piirrevektorin pistetulon ja vektoreiden pituuksien tulon suhteeseen, joka on sama kuin vektoreiden kulman kosini:

$$Cos\alpha = \frac{\vec{X} \bullet \vec{Y}}{\left|\vec{X}\right|\left|\vec{Y}\right|}. \qquad (6.10)$$

Yhtälöllä lasketut tulokset voivat olla väliltä 0-1, siten että mitä suurempi luku, sitä enemmän SAM tulkitsee spektrit kuuluvaksi samaan luokkaan.

Graafisessa mielessä SAM arvioi kahden spektrin samankaltaisuutta niiden välisen kulman (0-90º) perusteella. SAM algoritmin etuna on, että tulokset ovat riippumattomia piirrevektoreiden pituuksista, minkä vuoksi spektrejä voi kertoa mielivaltaisilla positiivisilla luvuilla, ilman että SAM tulokset muuttuisivat.



Kuva 6.9: Spectral Angle Mapper (SAM) luokittelu perustuu kahden piirrevektorin väliseen kulmaan. Vektoreiden pituuksilla ei ole vaikutusta SAM tuloksiin.

6.2.2 Spectral Correlation Mapper (SCM)

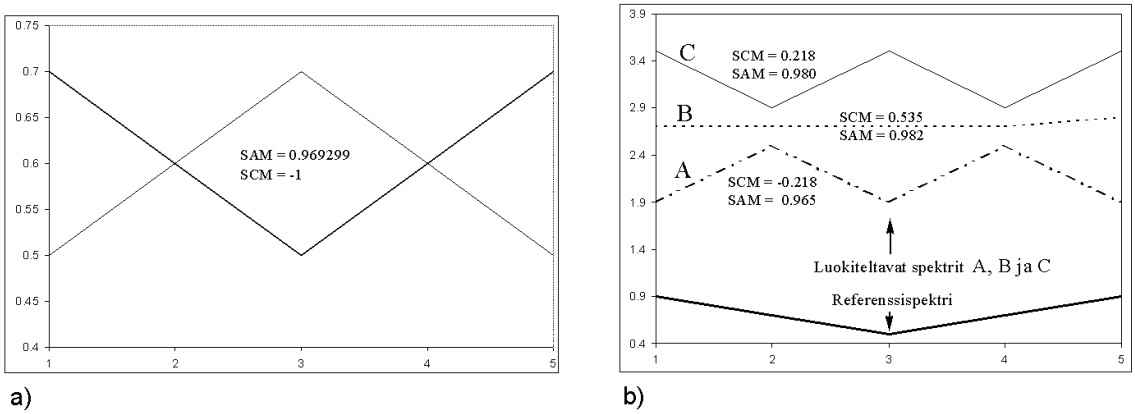
Spectral Correlation Mapper on periaatteeltaan samanlainen kuin SAM. Ero on siinä, että SCM:ssä spektristä vähennetään ensin kaikkien kanavien yli laskettu keskiarvo μ, jonka jälkeen vähennettyä "spektri" luokitellaan SAM algoritmin mukaisesti. SCM algoritmi on siis

$$SCM = \frac{(\vec{X} - \vec{\mu}_x) \bullet (\vec{Y} - \vec{\mu}_y)}{\left|\vec{X} - \vec{\mu}_x\right| \left|(\vec{Y} - \vec{\mu}_y)\right|}, \qquad (6.11)$$

jossa $\vec{X}$ on alkuperäisen tutkittavan objektin (spektrin) piirrevektori ja $\vec{\mu}_x$ on vektorin $\vec{X}$ keskiarvo. SCM tulokset voivat olla väliltä -1…1, siten että mitä suurempi luku, sitä enemmän spektrit tulkitaan kuuluvaksi samaan luokkaan. Kuten SAM:ssa, myös SCM:ssä spektrejä voi kertoa mielivaltaisella positiivisella luvulla, ilman että tulokset muuttuisivat. Mutta toisin kuin SAM:ssa, SCM:ssä tulokset eivät muutu vaikka spektrin jokaiseen piirteeseen lisäisikin saman positiivisen tai negatiivisen luvun [Holopainen].

Koska spektrien piirrevektoreiden komponentit ovat kaikki positiivisiä lukuja, niin SAM algoritmin haitta on siinä, että se ei huomioi piirteiden välistä negatiivistä ja positiivistä korrelaatiota. Carvalho et al. [Carvalho] on ehdottanut ongelman ratkaisuksi SCM:ää SAM:n sijasta. Kuvassa 6.10 on havainnollistettu SAM algoritmin heikkoutta SCM:ään nähden, analysoimalla [Carvalho] viisikanavaisia hypoteettisiä spektrejä. Kuvassa 6.10a on kaksi peilikuvaspektriä, jotka SAM tulkitsee lähes identtiseksi, kun taas SCM tulkitsee ne täysin erilaisiksi spektreiksi. Kuvassa 6.10b on vertailtu kolmea eri spektriä (A, B ja C)

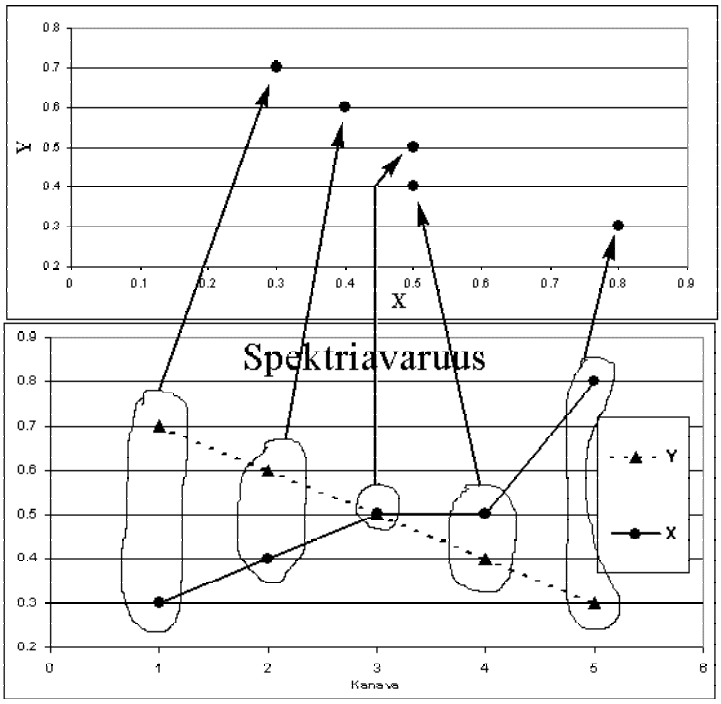
referenssispektriin. Vaikka A, B ja C ovat erimuotoisia, niin SAM pystyy huonosti erottelemaan ne referenssispektrin suhteen, toisin kuin SCM.



Kuva 6.10: SAM ja SCM algoritmien tulosten vertailua. SAM tulokset voivat olla väliltä 0…1 ja SCM tulokset väliltä –1…1, siten että mitä lähempänä arvot ovat ykköstä, sitä enemmän kyseinen algoritmi tulkitsee spektrit samoiksi.

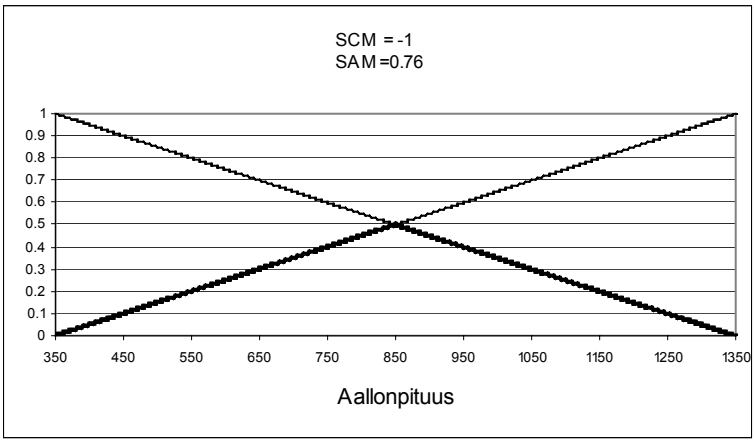
Tilastotieteellisestä näkökulmasta katsottuna, SCM on sama kuin Pearsonin korrelaatiokerroin, jota käytetään kahden muuttujan lineaarisen riippuvuuden mittana. Korrelaation kannalta SCM:ssä muuttujina ovat spektrin X ja Y amplitudit (kts. kuva 6.11).

McFee et al. ovat käyttäneet LCC (linear correlation coefficient) ja OSP algoritmejä maamiinojen havaitsemiseen [McFee 1997], [McFee 1999]. LCC vaikuttaisi olevan sama algoritmi kuin SCM.

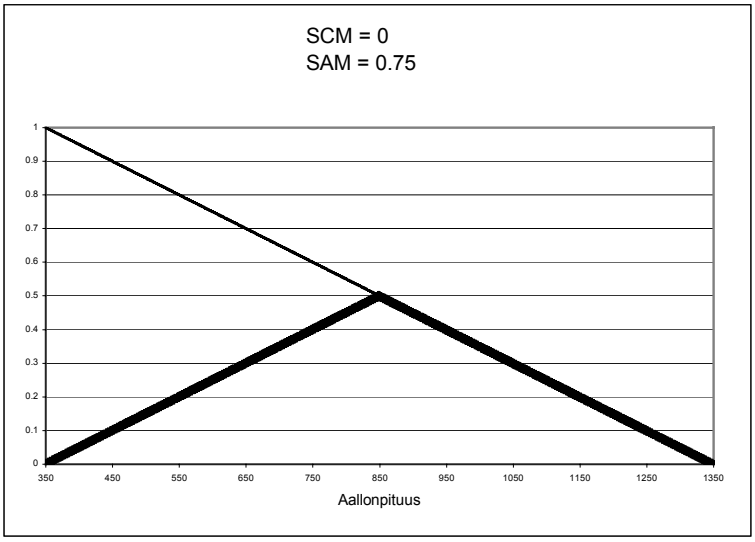


Kuva 6.11: Spectral Correlation Mapper (SCM) algoritmin idea Pearsonin korrelaatiokertoimen kannalta. Tutkittavat spektrit koodataan kaksiulotteiseen koordinaatistoon kuvan osoittamalla tavalla, jonka jälkeen lasketaan pisteiden korrelaatiokerroin, joka on samalla SCM luokittelutulos. Kuvan spektrien luokittelutulokset SCM ja SAM algoritmeillä ovat: SCM=-0.93 ja SAM=0.83.

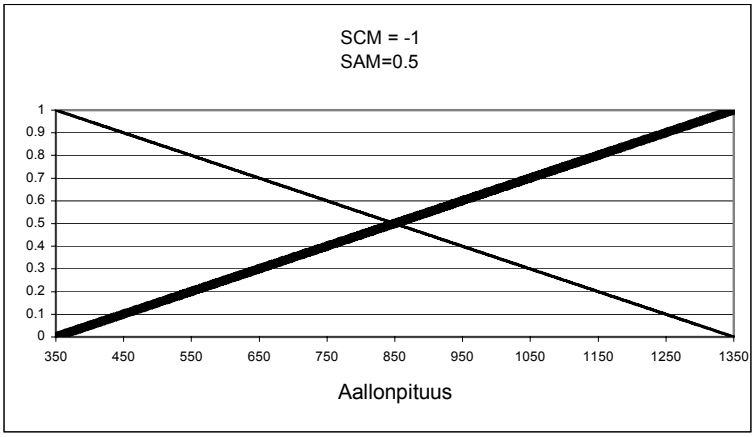
Kuvissa 6.12a-6.12d havainnollistetaan SCM ja SAM algoritmien kykyä vertailla keinotekoisia spektrejä keskenään.



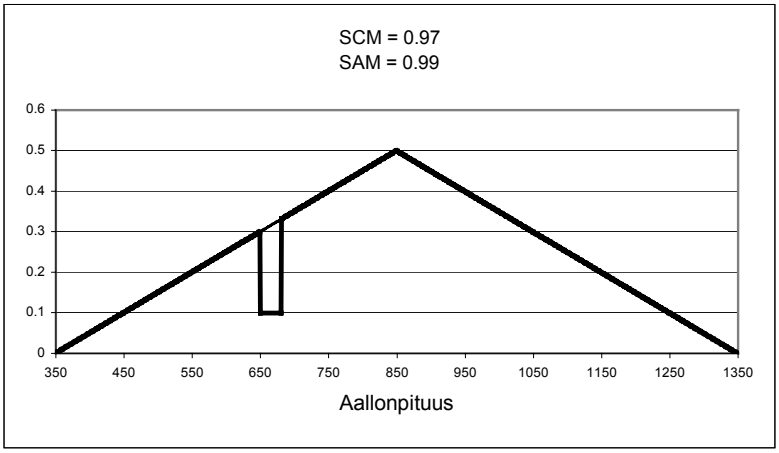
Kuva 6.12a: Kaksi peilikuvamaista spektriä, jotka SAM ja SCM ovat tulkinneet eri tavoin.



Kuva 6.12b: Kahden spektrin vertailu SCM ja SAM algoritmeillä. Huomaa että SAM antaa samansuuruisen tuloksen, kuin kuvan 6.12a tapauksessa.



Kuva 6.12c: Kahden spektrin vertailu SCM ja SAM algoritmeillä.



Kuva 6.12d: Esimerkki siitä kuinka huonosti SCM ja SAM pystyvät erottamaan kaksi spektriä toisistaan, joista toisella on voimakas absorptiopiikki. Syy tähän on se, että vertailuun on otettu mukaan kaikki kanavat (350-1350 nm).

6.2.3 Chi-toiseen testi

Kahden spektrin vertailuun voidaan käyttää tilastotieteestä tuttua χ^2-testiä. χ^2 lasketaan yhtälöllä

$$\chi^2 = \sum_{i=1}^{K} \left(\frac{(x_i - r_i)^2}{r_i} \right), \tag{6.12}$$

jossa x_i on tutkittavan spektrin ja r_i referenssispektrin amplitudi kanavalla i. Jotta tulokset olisivat vertailukelpoisia eri spektriparien välillä, χ^2 voidaan normittaa välille 0..1 seuraavalla tavalla:

$$\chi^2_N = 1-(\chi^2 / \chi^2_{Max}). \text{ [Staenz]} \tag{6.13}$$

6.2.4 SES

SES (square error statistics) menetelmässä kahden spektrin vertailuluku lasketaan yhtälöllä

$$\text{SES} = \sum_{i=1}^{K} \left(\frac{(s x_i - r_i)^2}{\sigma_{ri}} \right) \text{ ja } s = \sum_{i=1}^{K} r_i \Big/ \sum_{i=1}^{K} x_i \tag{6.14}$$

jossa σ_{ri} on referenssispektrin r keskihajonta kanavalla i. Keskihajonnasta johtuen SES:ssä tarvitaan joukko referenssispektrejä, toisin kuin SAM, SCM ja χ^2 menetelmissä. Luokittelua varten SES (yhtälö 6.14) normitetaan välille 0-1 samoin kuin χ^2 menetelmässä:

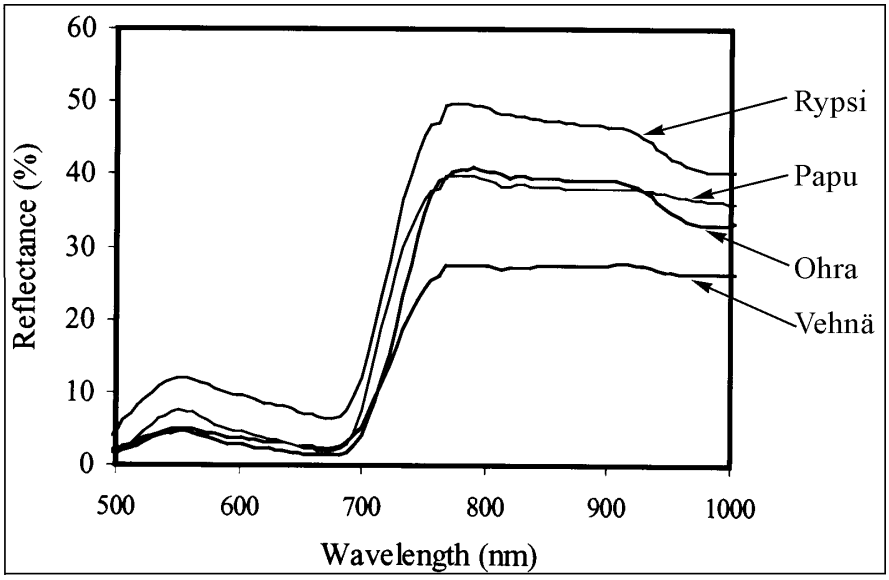
$$\text{SES}_N = 1-(\text{SES}/\text{SES}_{Max}). \tag{6.15}$$

SES_N, kuten myös χ^2 ja SAM, tulokset eivät muutu vaikka spektrejä kertoisi positiivisellä luvulla. Mikäli skaalausparametrin s (yhtälössä (6.14)) asettaa ykköseksi, niin edellinen ei päde. [Staenz]

Sekä χ^2, että SES algoritmien käyttö spektrien luokittelussa on harvinaista. Tämän tutkimuksen yhteydessä niiden käytöstä löydettiin vain yksi raportti (kappale 6.2.5).

6.2.5 Esimerkki MSAM, SCM, χ^2 ja SES luokittelusta.

Staenz et al. [Staenz] ovat vertailleet kuinka MSAM, SCM, χ^2 ja SES algoritmit luokittelevat ohra-, vehnä-, papu- ja rypsipellon pikseleitä reflektanssin perusteella. MSAM on modifioitu SAM algoritmi, siten että MSAM = $1-(2\bullet\alpha_{SAM}/\pi)$, jossa α_{SAM} on SAM:lla laskettu kulma. Tutkimuksissa pellot kuvattiin hyperspektrisellä casi-sensorilla 2,5 km korkeudesta, aallonpituusalueella 458-1000 nm. Kanavien lukumäärä oli 96. Tässä yhteydessä ei tarkemmin käydä läpi sitä kuinka sensorin keräämä data muutettiin reflektanssiksi. Kuvassa 6.13 on eri luokkien opetusnäytteistä keskiarvoistetut referenssispektrit, joihin luokiteltavia pikseleitä verrattiin. Luokkien opetusnäytteiden määrä vaihteli väliltä 286-986 pikseliä.



Kuva 6.13 [Staenz]: Keskiarvoistetut referenssispektrit eri luokista.

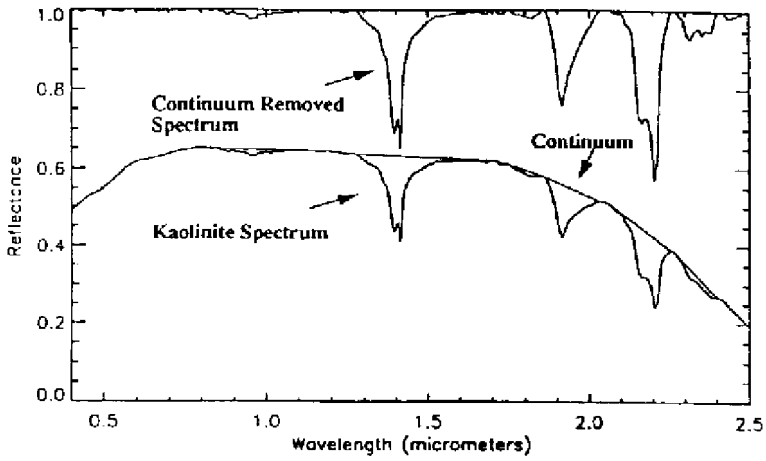
Staenz et al. saivat seuraavat luokittelutarkkuudet eri algoritmeille: SCM 84%, MSAM 78%, χ^2 72% ja SES 68%. Paras yksittäinen luokittelutulos saatiin rypsille, jonka MSAM ja SCM luokittelivat yli 97% oikein.

6.2.6 Spectral Feature Fitting (SFF)

Spectral Feature Fitting (SFF) on ENVI:n nimeämä luokittelumenetelmä, jonka suorituskyky perustuu absorptio-ominaisuuksien hyödyntämiseen. Aluksi SFF:ssä vertailtavista spektreistä poistetaan jatkumo, jonka jälkeen spektrit skaalataan vastaamaan toisiaan pienimmän neliösumman menetelmällä. SFF tulokset ovat skaalaustekijä ja RMS virhe tai niiden yhdistelmä. SFF:n ideaan perustuvat menetelmät (esim. Tetracorder) voivat hieman vaihdella, mutta yhteistä on jatkumon poisto ennen tulosparametrien laskemista.

Continuum Removal

Continuum removal eli jatkumon poisto on toimenpide jolla spektrit saadaan normalisoitua, jotta spektrien absorptiopiirteet olisivat paremmin vertailukelpoisia keskenään. Jatkumo on kupera käyrä, joka sovitetaan kulkevaksi spektrin paikallisten maksimiarvojen kautta, siten että maksimit yhdistetään suorilla viivasegmenteillä. Jatkumopoistettu spektri on alkuperäinen spektri jaettuna jatkumolla. Kuvassa 6.14 on alkuperäinen, jatkumo ja jatkumo-poistettu spektri.



Kuva 6.14: Kaoliitin reflektanssi ja siitä määritetty jatkumo (continuum) sekä jatkumo-poistettu spektri.

SFF tulokset

SFF tulokset määritetään kahden jatkumo-poistetun spektrin (X = tutkittava ja R = referenssi) välillä siten, että X:stä ja R:stä lasketaan käyrät X' ja R' yhtälöillä

$$X' = 1 - X \qquad ja \quad R' = 1 - R \quad , \tag{6.16}$$

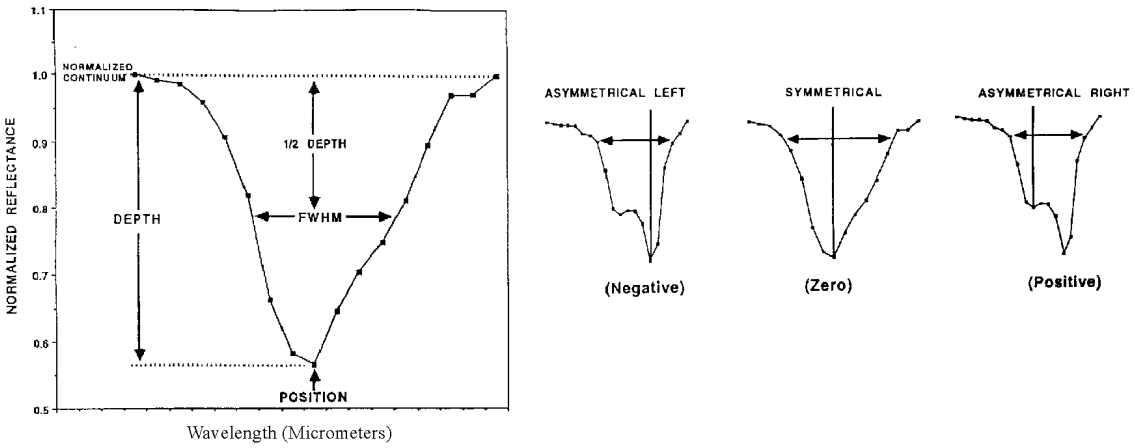
joissa 1 on yksikkövektori. X' ja R' kertovat kuinka syviä ovat jatkumo-poistettujen spektrien absorptiot kullakin kanavalla suhteessa ykköseen. Lopuksi X' ja R' skaalataan vakiolla k vastaamaan toisiaan niin, että yhtälön

$$X' = k \bullet R' + n \tag{6.17}$$

n vektorin komponenteistä (jäännöstermeistä) laskettu RMS virhe minimoituu. Saadut skaalaustekijä k ja RMS arvo ovat SFF tulokset. Suuri skaalaustekijä ja pieni RMS arvo merkitsevät parempaa vastaavuutta vertailtaville spektreille. Arvot voidaan yhdistää esittämällä luokittelutulokset skaalaustekijä-RMS suhteena tai skaalaustekijä-RMS koordinaatistossa.

Koska SFF on absorptio-ominaisuuksiin perustuva menetelmä, on oleellista tiedostaa onko etsittävällä materiaalilla riittävän karakteristisiä absorptiopiirteitä, jotta SFF:n käyttäminen olisi tulosten luotettavuuden kannalta järkevää. Toinen tärkeä seikka on valita sopivin aallonpituusalue, jolla karakteristisimmat absorptiopiirteet esiintyvät. Vaikka SFF ei sinällään vaadi piirteiden vähentämistä, huolimaton aallonpituusalueen valinta tai koko spektrin huomioiminen SFF:ssä voi helposti hukata absorptiopiirteisiin kätkeytyneen informaation. Luonnollisesti aallonpituusalueen valintaan ei vaikuta pelkästään etsittävän materiaalin absorptiopiikit, vaan myös kohde-tausta (spektrinen) kontrasti. [ENVI]

Absorptio-ominaisuuksiin perustuvaa analysointia voidaan laajentaa huomioimalla muitakin kanavaparametrejä kuin pelkästään absorptiopiikin syvyys. Jatkumo-poistetuista spektreistä voidaan ekstraktoida parametrejä kuten absorptiopiikin sijainti, puoliarvonleveys (FWHM) ja symmetrisyys (kuva 6.15) [Kruse].



Kuva 6.15 [Kruse]: Spektrin absorptiovyöstä määritettäviä parametrejä: sijainti, syvyys, puoliarvonleveys (FWHM) ja symmetrisyys.

Tetracorder

USGS (U.S. Geological Survey) on kehittänyt Tetracorder nimisen tietokoneohjelman [Clark], joka perustuu jatkumo-poistettujen spektrien sovittamiseen eli on periaatteeltaan samanlainen kuin SFF. Erona SFF:ään on mm. se, että ohjelma käsittelee sovituksesta saatuja parametrejä hienostuneemmin määrittäessään lopullista tulosta.

6.2.7 Binary Encoding (BE)

Yksinkertaisin BE on yhden bitin versio, jossa spektri R(n) esitetään muodossa

$$B(n) = \begin{cases} 0 & \text{jos } R(n) \leq T \\ 1 & \text{muulloin.} \end{cases} \qquad n = 1,\ldots,N$$

Kynnysarvo T voi periaatteessa olla mikä tahansa luku järkevyyden rajoissa, mutta usein T on spektrin keskiarvo. Menetelmä voidaan laajentaa kahden bitin järjestelmäksi, jossa kolmen kynnysarvon (T_1, T_2 ja T_3) perusteella spektrin kanava n koodataan muotoon

$$B(n) = \begin{cases} 00 & \text{jos } R(n) \leq T_1 \\ 01 & \text{jos } T_1 < R(n) \leq T_2 \\ 11 & \text{jos } T_2 < R(n) \leq T_3 \\ 10 & \text{jos } T_3 < R(n). \end{cases}$$

(Kolmas taso koodataan mielummin muotoon 11 kuin 10, jotta vierekkäisten tasojen erona olisi vain yksi bitti.) B(n) arvojen sijasta voidaan myös käyttää spektrin kulmakertoimiin perustuvaa koodausta. [Richards]

Vertailutulokset binäärijonoista määritetään suorittamalle niille XOR operaatio ja laskemalla saadun binäärijonon alkioiden keskiarvo. (Esim. 0101 (XOR) 0001 → 0100 → keskiarvo=0.25).

7. Virheanalyysi

Tässä luvussa käsitellään yleisimpiä käytössä olevia menetelmiä, joilla analysoinnin luotettavuutta voidaan arvioida. Perusidea on verrata analysoituja pikseleitä testijoukkoon eli tunnettuihin pikseleihin. On huomioitava että tuloksien tarkkuus ei ole suoranainen mittari luokittelualgoritmin toimivuudelle, koska kaukokartoitusprosessiin liittyy monia muitakin virhelähteitä kuin pelkkä luokittelija (esim. mittalaitteiston kalibrointi, ilmakehäkorjaukset, BRDF tekijät, yms.).

7.1 Virhematriisi

Virhematriisi (error matrix, confusion matrix) ei nimestään huolimatta ole matemaattisessa mielessä matriisi, vaan taulukko luokittelutuloksista suhteessa testijoukon pikseleihin. Taulukossa 7.1 on esimerkkinä virhematriisi kolmen luokan luokittelusta. Ideaalitilanteessa virhematriisin kaikki ei-diagonaalielementit olisivat nollia, mutta koska luokittelija ei ole osannut luokitella kaikkia pikseleitä oikein, myös ei-diagonaalielementit taulukossa 7.1 ovat nollasta poikkeavia.

Taulukko 7.1: Esimerkki virhematriisista

		Testijoukon luokka			Yht.
		A	B	C	
Luokittelijan antama luokka	A	35	2	2	39
	B	10	37	3	50
	C	5	1	41	47
Testijoukon pikselit yht.		50	40	46	136

7.1.1 Tuottajan tarkkuus vs. käyttäjän tarkkuus

Virhematriisista voidaan laskea tuottajan ja käyttäjän tarkkuuksia. Tuottajan tarkkuudella, jota Richards [Richards 1996] myös nimittää luokittelijan suorituskyvyksi, tarkoitetaan todennäköisyyttä että luokittelija on luokitellut testijoukon pikselit oikeaan luokkaan. Esim. taulukon 7.1 tuottajan tarkkuus luokalle A on 35/50 = 70 %. Tuottajan tarkkuutta vastaavasta virheestä käytetään termiä “omission error”, joka luokalle A on (100-70)% = 30 %.

Tuottajan tarkkuuden sijasta voidaan kysyä, että kuinka suurella todennäköisyydellä johonkin luokkaan luokiteltu pikseli oikeasti kuuluu siihen luokkaan. Tätä todennäköisyyttä nimitetään käyttäjän tarkkuudeksi, joka luokalle A (taulukko 7.1) on 35/39 = 89.7 %. Vastaavasta virheestä käytetään termiä “commission error”, joka luokalle A on 10.3 %. [Richards]

Viimekädessä käyttäjän tarkkuus on hyödyllisempi mittari kuin tuottajan tarkkuus, koska se vastaa kysymykseen kuinka hyvin luokittelutulokset vastaavat todellisuutta, kun taas tuottajan tarkkuus vastaa kysymykseen kuinka hyvin tilanne maassa (testijoukon pikselit) vastaa luokittelutuloksia.

7.2 Kappa kerroin

Tulosten kokonaistarkkuus ilmoitetaan usein kappa kertoimella, joka määritellään virhematriisin alkioiden x_{ij} ja kaikkien testipikseleiden lukumäärän P perusteella. Kappa lasketaan yhtälöllä

$$\kappa = \frac{P\sum_k x_{kk} - \sum_k x_{k+}x_{+k}}{P^2 - \sum_k x_{k+}x_{+k}} \quad , \qquad (7.1)$$

missä x_{i+} on rivin i ja x_{+j} sarakkeen j alkioiden summa. [Richards]

7.3 Testijoukon valinta

Yksi luokittelijan toivotuista perusominaisuuksista on yleistämiskyky eli kyky luokitella kohteet oikein, vaikka niiden piirteet hieman eroaisivatkin opetusjoukon näytteistä. Tästä syystä opetus- ja testijoukko eivät saa koostua samoista pikseleistä. Toinen virheanalyysissä oleellinen tekijä on tapa jolla testijoukko valitaan.

Ehkä helpoin tapa testipikseleiden määrittämiseen on käyttää testikenttiä. Menetelmän haittana on se, että testikentän (saman luokan) naapuripikselit ovat todennäköisesti hyvin samanlaisia. Tällöin näennäisesti kerätään paljon testipikseleitä, mutta niiden todellinen määrä virheanalyysin kannalta on pienempi. Mikäli testipikselit valitaan mielivaltaisesti, niin pienet luokat saavat suhteellisesti vähemmän edustajia kuin suuret luokat. Tällä ei tosin ole vaikutusta tuottajan tarkkuuteen, mutta käyttäjän tarkkuuteen kylläkin.

Richards [Richards 1996] on vertaillut tuottajan ja käyttäjän tarkkuutta toisiinsa sekä esittänyt kuinka luokittelutulosten lopullinen tarkkuus (myöh. tuloksen tarkkuus) tulisi määrittää. Käyttäjän tarkkuutta voidaan pitää samana kuin tuloksen tarkkuus, mikäli testijoukon pikseleiden suhteelliset lukumäärät ovat vastaavia kuin luokkien todelliset osuudet (prioritodennäköisyydet) maastossa. Tuloksen tarkkuus voidaan myös johtaa matemaattisesti tuottajan tarkkuudesta, mikäli prioritodennäköisyydet tunnetaan. Mutta koska niiden luotettava tunteminen on vaikeaa, niin Richardsin mielestä tulosten tarkkuus on paras määrittää siten, että testipikselit valitaan luokittelutuloksien perusteella.

Testijoukon valintaa ja virheanalyysia ovat käsitelleet mm. Congalton [Congalton], Banko [Banko], Verbyla [Verbyla] ja Jain et al. [Jain 2000].

7.4 ROC käyrä

Tarkastellaan seuraavaksi kohteen havaitsemisongelmaa, joka on erikoistapaus kahden luokan (kohde vs. tausta/kaikki muu) luokittelusta. Havaitsemisjärjestelmän (mittauslaitteisto+datan käsittelyprosessi) tehtävänä on antaa vastaus, että sisältääkö pikseli kohdetta vai ei. Vastauksien vertaaminen testijoukkoon voi tuottaa neljä eri vaihtoehtoa, joiden lukumäärät voidaan koostaa virhematriisiin (kts. taul. 7.2).

Taulukko 7.2: Kohteen havaitsemissovelluksen virhematriisi. Luokittelutuloksien vertaaminen testijoukkoon antaa neljä eri vaihtoehtoa, joiden lukumäärät kirjataan virhematriisiin.

	Onko pikseli kohde	
Luokittuuko pikseli kohteeksi	Kyllä	Ei
Kyllä	Todellinen Positiivinen (TP)	Väärä positiivinen (VP)
Ei	Väärä Negatiivinen (VN)	Todellinen negatiivinen (TN)

Virhematriisin alkioista N_{ij} saadaan määritettyä TP-, TN-, VP- ja VN-suhteet:

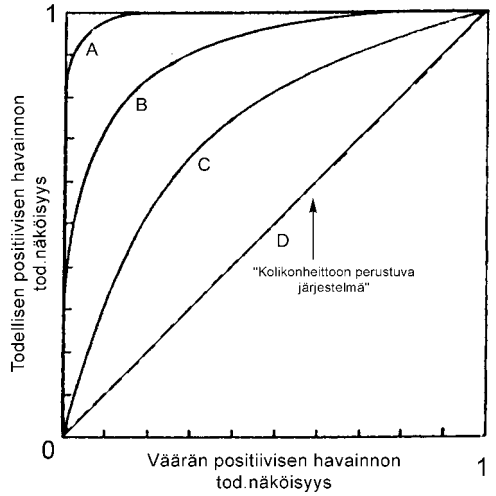
$$\text{TP-suhde} = N_{TP} / (N_{TP} + N_{VN}) \quad \text{(herkkyys)} \qquad (7.2a)$$

$$\text{TN-suhde} = N_{TN} / (N_{TN} + N_{VP}) \quad \text{(spesifistisyys)} \qquad (7.2b)$$

$$\text{VP-suhde} = N_{VP} / (N_{TN} + N_{VP}) \qquad (7.2c)$$

$$\text{VN-suhde} = N_{VN} / (N_{TP} + N_{VN}) \qquad (7.2d)$$

Luokittelualgoritmin antama numeerinen tulos ei sinällään anna vastausta onko kyseessä kohde vai ei, vastaus riippuu kynnysarvon asettamisesta. Väljä kynnysarvo havainnoi kohteet, mutta samalla saa aikaan paljon vääriä positiivisiä hälytyksiä. Tiukka kynnysarvo toimii päinvastoin: ei havainnoi kohteita eikä tee vääriä positiivisiä hälytyksiä. Vaihtelemalla kynnysarvoa saadaan erilaisia virhematriiseja, joista voidaan määrittää ROC (receiver operating characteristic) käyrä, jonka pisteet ovat VP- ja TP-suhteita. Toisin sanoen kynnysarvoja vaihtelemalla saadaan TP havainnon todennäköisyys VP todennäköisyyden funktiona, joka havainnollistaa järjestelmän suorituskykyä. Kuvassa 7.1 on neljän eri järjestelmän ROC käyrät. Paras järjestelmä on A ja huonoin D, joka tuottaa arvotonta tietoa, koska todennäköisyydet oikeaan ja väärään positiiviseen havaintoon ovat samat eli "kohde-vai-ei" kysymykseen vastaaminen on verrattavissa kolikonheittoon. Järjestelmän keskimääräinen suorituskyky, joka on riippumaton kynnysarvon valinnasta, saadaan ROC käyrän alle jäävästä pinta-alasta. [AeroSense 2002]



Kuva 7.1: Neljän kuvitteellisen havainnointijärjestelmän ROC käyrät.

8. Yhteenveto

Hyperspektrisillä kuvausjärjestelmillä maanpinnan kohteista saadaan spektraalisesti tarkkoja sormenjälkiä, joiden perusteella ne on mahdollista luokitella eri materiaalityyppeihin. Hyperspektrinen data ei kuitenkaan suoraan anna vastausta tähän. On käytettävä algoritmejä, jotka muuttavat digitaaliset mittausarvot ihmisen ymmärtämään muotoon.

Spektridatan analysoinnilla on paljon yhteistä muiden tieteenalojen data-analyysisovelluksiin. Yhteistä on se, että tutkittavia objekteja karakterisoidaan piirrejoukolla, jonka perusteella samankaltaisia objekteja pyritään ryhmittelemään tai tunnistamaan. Erityisesti tehokkaiden ja kohtuullisen hintaisten tietokoneiden yleistyminen on lisännyt data-analyysisovelluksia ja tehneet niistä ylipäätään mahdollisia. Nykyisin datan tuottaminen ei ole ongelma, vaan se kuinka dataan kätkeytynyt informaatio pystyttäisiin kaivamaan ulos.

Tämän tutkimuksen päätarkoitus oli selvittää mitä eri algoritmejä hyperspektrisen datan analysointiin on käytetty tai voidaan käyttää. Algoritmien toimivuuden ja luotettavuuden objektiivinen arvioiminen edellyttäisi niiden testaamista samalla kaukokartoitusdatalla. Yleispätevää ”algoritmirankingia” on hankala muodostaa; toisissa sovelluksissa sama algoritmi voi toimia hyvin ja jossain toisissa huonosti. Lähdemateriaalin pohjalta muodostui kuitenkin jonkinlainen käsitys eri algoritmeistä ja siitä miten ne soveltuvat sotilaskohteiden havaitsemissovelluksiin. Sotilaskohteiden havainnoinnissa tyypillistä on mm. se, että kuvadatassa on vain vähän pikseleitä etsittävästä kohteesta suhteessa taustaan, joka Suomen oloissa on useimmiten kasvillisuutta.

Hyperspektristen sensoreiden suuri kanavamäärä tuottaa satoja mittauskohdetta karakterisoivia muuttujia eli piirteitä. Osa näistä piirteistä on kuitenkin hyödyttömiä analysointia varten, koska vierekkäisillä kanavilla on yleensä vahva keskinäinen korrelaatio, ja lisäksi suuridimensioinen piirreavaruus on hyvin ongelmallinen analysointiympäristö. Tämän vuoksi piirteiden määrää on pystyttävä vähentämään niin, että kohdetta karakterisoiva uusi pienempi piirrejoukko olisi mahdollisimman informatiivinen. Piirteiden vähentäminen voidaan tehdä joko valitsemalla tai ekstraktoimalla alkuperäisistä piirteistä uusi mataladimensioisempi piirrejoukko.

Yksi tunnetuin ja käytetyin ekstraktointimenetelmä on PCA (ja sen modifikaatio MNF). PCA:n haitta on että se kohtelee dataa yhtenä jakautumana, jolloin eri luokkien välisiä eroja ei huomioida. Lisäksi erityisesti harvinaisten materiaalien etsintäsovelluksissa PCA on lähtökohdaltaan huono, koska se saattaa kohdella joissain sovelluksissa niiden spektrejä ikään kuin arvottomana kohinana. Mikäli PCA kuitenkin toimisi hyvin harvinaisten materiaalien etsinnässä, on se enemmänkin spektrien luonteesta johtuvaa sattumaa kuin osoitus PCA:n yleisestä tehokkuudesta. DAFE ja DBFE ovat kaksi muuta yleisesti tunnettua ektraktointimenetelmää. Niiden ongelmat liittyvät siihen, että ektraktointioperaattorit määritetään alkuperäisessä suuridimensioisessa avaruudessa, jolloin Hughesin ilmiö vaikeuttaa tehokkaiden piirteiden muodostamista. Wavelets ekstraktoinneista löytyi suhteellisen vähän raportointeja, joten niiden luotettavuudesta ei muodostunut selkeää käsitystä. Kuitenkin ne vaikuttavat idealtaan mielenkiintoisilta, koska ne tehdään spektriavaruudessa piirreavaruuden sijasta. Jatkossa wavelets ekstraktointien toimivuuden tutkimiseen kannattaa panostaa.

Luvussa 5 käsitellyt (lineaariset) unmixing menetelmät voidaan teoriassa todistaa keskenään samoiksi, käytännössä erot liittyvät endmembereiden valintaan ja oletukseen kohinasta. Erityisesti OSP, CEM ja RX algoritmejä on käytetty (tutkimusmielessä) sotilaskohteiden havaitsemissovelluksissa. Unmixing menetelmiä voidaan luonnollisesti käyttää puhtaiden pikseleiden tutkimiseen, mutta epäselvää on se, kuinka hyvin luvun 6 puhtaan pikselin luokittelumenetelmät toimivat epähomogeenisten pikseleiden kohdalla. Periaatteessa unmixing algoritmien antama tulos on verrannollinen kohdemateriaalin pitoisuuteen pikselissä, mutta se kuinka puhtaiden pikseleiden luokittelualgoritmien tulokset korreloivat pitoisuuteen on arvoitus.

Luvussa 6.1 käsitellyt algoritmit (poislukien neuroverkot) ovat olleet hyvin suosittuja multispektristen datojen analysoinnissa. Erityisesti näiden menetelmien käyttäminen hyperspektriseen dataan vaatii dimension vähentämistä. Luvun 6.2 algoritmejä on käytetty paljon hyperspektrisissä datoissa. Tämän tutkimuksen yhteydessä ei löydetty raportointeja niiden soveltamisesta multispektriseen dataan. Ehkä merkittävin luvun 6.2 algoritmeistä on Spectral Correlation Mapper (SCM), joka vaikuttaa kehittyneemmälle kuin sen lähisukulainen Spectral Angle Mapper (SAM). Molemmat algoritmit ovat tunteettomia spektrin skaalaustekijälle (piirrevektorin pituudelle) ja lisäksi SCM on tunteeton spektrin vakiotekijälle (offsetille). Nämä ominaisuudet ovat hyödyllisiä erityisesti auringon varjotekijöiden eliminoimiselle, vaikka ilmakehän sirontaominaisuuksista johtuen

varjotekijät eivät täsmällisesti ottaen käyttäydy skaalaus- ja vakiotekijöiden tavoin. Yleensä klusterointimenetelmissä käytetty vertailuindeksi on euklidinen etäisyys, mutta on muistettava että indeksinä voi käyttää myös SAM tai SCM arvoa. Spectral Feature Fitting (SFF) on suosittu geologisissä sovelluksissa, joissa malmien ja mineraalien spektreissä esiintyy voimakkaita absorptio-ominaisuuksia. Perusidea SFF:ssä on samanlainen kuin lineaarisissa unmixing menetelmissä, koska molemmissa etsitään skaalaustekijää siten että PNS-virhe minimoituisi. Näkyvin ero liittyy SFF:ssä suoritettavaan jatkumon poistoon. Tästä herää kysymys, että voitaisiinko projisoivia unmixing algoritmejä (OSP, CEM, RX) parantaa poistamalla spektreistä ensin jatkumo. Binary Encodingissa on erikoista se, että monissa hyperspektrien analysointia käsittelevissä artikkeleissä se mainitaan, mutta sitä koskevista analysointituloksista ei löytynyt raportointeja.

Neuroverkot ovat nykyisin suosittu tutkimusaihe hyvin monenlaisissa data-analyysisovelluksissa, minkä vuoksi niiden soveltaminen kaukokartoitusdataan tuntuu hyvin tarkoituksenmukaiselle. Verkot pystyvät mm. oppimaan piirteiden välisiä epälineaarisiakin suhteita. Erityisesti luvun 6.1 algoritmeistä neuroverkot vaikuttaisivat järkevimmiltä. Verkkojen käytössä vaikeutena on, että oikeanlaisen verkkoarkkitehtuurin (kerrosten ja neuroneiden määrän) löytäminen ei ole itsestään selvää. Tässä tutkimuksessa neuroverkkoja käsiteltiin esittelynomaisesti.

Tutkimuksessa suurimpana vaikeutena oli se, että hyperspektrisen kaukokartoitusdatan analysointia käsitteleviä kirjoja on hyvin vähän. Aiheesta löytyi enimmäkseen erillisiä artikkeleja ja tutkimusraportteja, joissa kussakin käsiteltiin vain muutamaa algoritmiä kerrallaan. Yksi tavoite olikin antaa yleiskuva eri menetelmistä ja algoritmeistä, jolloin sovelluskohtaisten algoritmien valinta helpottuu.

Tämän tutkimuksen yhteydessä ei analysoitu mitään dataa. Jatkotutkimuksissa tullaan keskittymään Ilmavoimien geografisen tiedon vastuualueen hankkiman ja mittaaman hyperspektrisen datan analysointiin. Esimerkkeinä hankituista datoista ovat vuonna 2000 Kainuusta mitattu Hymap-data ja samalta alueelta 2002 EO-1 -satelliitin keräämä Hyperion-data.

Lähteet:

[Adams]
John B. Adams, Donald E. Sabol, Dar A. Roberts, Milton O. Smith, Alan R. Gillespie, Valerie Kapos and Raimundo Almeida Filho, "Classification of multispectral images based on fractions of endmembers: application to land-cover change in the Brazilian Amazon", Remote Sensing of Environment, 52(2):137-154, May 1995, http://dana.geology.washington.edu/amazon/amazon.htm.

[AeroSense 2002]
Eustace L. Dereniak and Michael R. Descour, "Imaging Spectrometers", short course notes, AeroSense Conference, 3 April 2002, Orlando, USA.

[Bacher]
Steve De Bacher, "Unsupervised Pattern Recognition", PhD Dissertation, University of Antwerpen, 2002, http://143.129.203.3/visielab/theses/debacker/SteveThesis.pdf.

[Banko]
Gebhard Banko, "A Review of Assessing the Accuracy of Classification of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory", Interim Report IP-98-081, November 1998, http://www.iiasa.ac.at/Publications/Documents/IR-98-081.pdf.

[Brumbley]
Clark Brumbley and Chen-I Chang, "An unsupervised vector quantization-based target subspace projection approach to mixed pixel detection and classification in unknown background for remotely sensed imagery", Pattern Recognition 32(7):1161-1174, July 1999.

[Carvalho]
Osmar Abílio de Carvalho Jr. and Paulo Roberto Meneses, "Spectral Correlation Mapper (SCM): An Improvement on the Spectral Angle Mapper (SAM)", AVIRIS Proceedings: 2000, http://popo.jpl.nasa.gov/docs/workshops/00_docs/Osmar_1_carvalho__web.pdf.

[Chang 1998]
Chein-I Chang, Xiao-Li Zhao, Mark L. G. Althouse and Jeng Jong Pan, “Least Squares Subspace Projection Approach to Mixed Pixel Classification for Hyperspectral Images”, IEEE Transactions on Geoscience and Remote Sensing, 36(3):898-912, May 1998.

[Chang 2002a]
C-I Chang, "Linear Spectral Random Mixture Analysis for Hyperspectral Images", IEEE Transactions on Geoscience and Remote Sensing, 40(2):375-392, February 2002.

[Chang 2002b]
Chein-I Chang, "On Relationship Among Orthogonal Subspace Projection, Constrained Energy Minimization and RX-Algorithm", Proceedings of AeroSense 2002, SPIE.

[Chen]
C. H. Chen, "The Use of Independent Component Analysis as a Tool for Data Mining", Proceedings of Igarss 2002

[Clark]
Roger N. Clark, Gregg A. Swayze, K. Eric Livo, Raymond F. Kokaly, Steve J. Sutley, J. Brad Dalton, Robert R. McDougal and Carol A. Gent, ”Imaging Spectroscopy: Earth and Planetary Remote Sensing with the USGS Tetracorder and Expert Systems", Version 12f: 8/03/2002, USGS Spectroscopy Lab, http://speclab.cr.usgs.gov/PAPERS/tetracorder/.

[Chen 1996]
Chin-Hsing Chen and Te-Ming Tu, "Computation Reduction of the Maximum Likelihood Classification Using the Winograd Identity", Pattern Recognition, 29(7):1213-1220, 1996.

[Congalton]
R. Congalton, "A Review of Assessing the Accuracy of Classification of Remotely Sensed Data", Remote Sensing of Environment, 37:35-46 (1991).

[Diaconis]
P. Diaconis and D. Freeman, "Asymptotics of Graphical Projection Pursuit", The Annals of Statistic, Vol. 12, No. 3, pp. 793-815, 1984.

[Du]
Qian Du and Hsuan Ren, "Performance Analysis for CEM and OSP", Proceedings of AeroSense 2002, SPIE.

[Egmont-Petersen]
M. Egmont-Petersen, D. de Ridder and H. Handels, "Image Prosessing with Neural Networks - a Review", Pattern Recognition, 35(10):2279-2301, October 2002.

[Evans]
Fiona Evans, "An investication into the use of maximum likelihood classifier, decision trees, neural networks and conditional probabilistic networks for mapping and predicting salinity", Master Thesis, School of Computing, Curtin University of Technology, Australia, March 1998.

[ENVI]
"ENVI User´s Guide", Research System Inc. (RSI), ENVI Version 3.4, September 2000.

[Farrand]
William H. Farrand, "Analysis of AVIRIS Data: A Comparision of the Performance of Commercial Software with Published Algorithms", AVIRIS Proceedings: 2001, http://popo.jpl.nasa.gov/docs/workshops/01_docs/2001Farrand_web.pdf.

[Ferri]
F. Ferri, P. Pudil, M. Hafer, and J. Kittler, "Comparative Study of Techniques for Large Scale Feature Selection", Pattern Recognition in Practice IV, pp. 403-413, 1994.

[Fukunaga 1972]
K. Fukunaga, "Introduction to Statistical Pattern Recognition", School of Electrical and Computer Engineering, Purdue University, Academic Press, New York, 1972.

[Hall]
P. Hall and K. Li, "On Almost Linearity of Low Dimensional Projection Form High Dimensional Data", The Annals of Statistic, Vol. 21, No. 2, pp. 867-889, 1993.

[Harsanyi]
William H. Farrand and Joseph C. Harsanyi, "Mapping the Distribution of Mine Tailings in the Coeur d´Alene River Valley, Idaho, through the Use of a Constrained Energy Minimization Technique", Remote Sensing of Environment 59(1):64-76, January 1997.

[Holopainen]
Martti Holopainen, "Tilastomatematiikan perusteet", s. 79, 1.-3. painos, 1984.

[Hsu, 1998]
Y.-H, Tseng and P.-H., Hsu, " A Primary Study of Fourier Spectrum Feature Extraction for HyperSpectral Image", Proceedings of the 19th Asia Conference on Remote Sensing, November 16-20, 1998.

[Hsu, 1999]
Y.-H, Tseng and P.-H., Hsu, "Feature Extraction for Hyperspectral Image", Proceedings of the 20th Asia Conference on Remote Sensing, November 22-25, 1999.

[Hsu, 2000]
Y.-H, Tseng and P.-H., Hsu, "Multiscale Analysis of Hyperspectral Data Using Wavelets for Spectral Feature Extraction", Proceedings of the 21th Asia Conference on Remote Sensing, December 4-8, 2000.

[Hsu, 2001(a)]
Y.-H, Tseng and P.-H., Hsu, "Feature Extraction for Hyperspectral Images", Proceedings of the 5th International Airborne Remote Sensing Conference September 17-20, 2001.

[Hsu, 2001(b)]
Y.-H, Tseng and P.-H., Hsu," Multiscale Feature Extraction of Hyperspectral Data Using Wavelet Transform", Department of Surveying Engineering, National Cheng Kung University, Proceedings of the 5th International Airborne Remote Sensing Conference September 17-20, 2001.

[Hsu, 2002]
Y.-H, Tseng and P.-H., Hsu, "Feature Extraction of Hyperspectral Data Using the Best Wavelet Packet Basis", Proceedings of Igarss 2002.

[Hyvärinen]
A. Hyvärinen, J. Karhunen and E. Oja, "Independent Component Analysis", John Wiley & Sons Inc., 2001.

[Igarss 2002]
"Introduction to Hyperspectral Image Analysis", short course notes, Igarss 2002, Toronto, Canada.

[Jain]
A.K. Jain and D. Zongker, "Feature Selection: Evaluation, Application, and Small Sample Performance", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no.2, pp. 153-158, Feb. 1997.

[Jain 2000]
Anil K. Jain, Robert P.W Duin, Jianchang Mao, "Statistical Pattern Recognition: A Review", IEEE Transaction on Pattern Analysis and Machine Intelligence, 22(1), January 2000.

[Jimenez 1999]
Luis Jimenez and David Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical and Asymptotical Properties of Multivariate Data", IEEE Transaction on Geoscience and Remote Sensing, 37(6), November 1999, http://www.ece.purdue.edu/~landgreb/SMC%201%20final.pdf.

[Kaewpijit]
S. Kaewpijit, J. Le Moigne and T. El-Ghazawi, "A Wavelet-based PCA Reduction for Hyperspectral Imagenery", Proceedings of Igarss 2002.

[Kendal]
M.G. Kendall, "A Curse in the Geometry of n-dimensions", Hafner Publishing Co., New York 1961, pp.35.

[Kruse]
F. A. Kruse and A. B. Lefkoff, "Analysis of Spectral Data of Manmade Materials, Military Targets, and Background Using an Expert System Based Approach", Proceedings of ISSSR'99, 31 October – 4 November 1999, Las Vegas, Nevada, http://www.aigllc.com/pdf/fak/Kruse_ISSSR99.pdf.

[Kurkinen]
Ville Kurkinen, "Neuroverkot", Tietoliikennearkkitehtuurit kurssin kotitehtävä 4 (4.10.1998), Teknillinen korkeakoulu, www.tml.hut.fi/Studies/Tik-110.300/1998/Newtech/neuroverkot_1.html.

[Landgrebe 1/1993]
Chulsee Lee and David Landgrebe, "Feature Extraction and Classification Algorithms for High Dimensional Data", TR-EE 93-1, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, January 1993, http://dynamo.ecn.purdue.edu/~landgreb/ChulheeTR.pdf.

[Landgrebe 4/1995]
Luis Jimenez and David Landgrebe, "High Dimensional Feature Reduction Via Projection Pursuit", TR-EE 96-5, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, April 1995 http://dynamo.ecn.purdue.edu/~landgreb/JimenezTR.pdf.

[Landgrebe 1998a]
Pi-Fuei Hsieh and David Landgrebe, "Classification of High Dimensional Data", TR-ECE 98-4, School of Electrical and Computer Engineering, Purdue University, May 1998, http://dynamo.ecn.purdue.edu/~landgreb/Hsieh.TRECE4.pdf.

[Landgrebe 1998b]
Saldju Tadjudin and David Landgrebe, "Classification of High Dimensional Data with Limited Training Samples", TR-ECE 98-8, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, May 1998 , http://dynamo.ecn.purdue.edu/~landgreb/Saldju_TR.pdf.

[Li]
J. Li, L. M. Bruce and A. Mathur, "Wavelet Transform for Dimensionality Reduction in Hyperspectral Linear Unmixing", Proceedings of Igarss 2002

[Martinez 2000]
P.L. Aguilar, A. Plaza, P. Martínez and R.M. Pérez, "Endmember Extraction by a Self-Organizing Neural Network on Hyperspectral Images.", VI International Conference on Automation, Robotics, Control and Vision (ICARCV), Singapore 2000.

[Martìnez 2001a]
P. Martínez, P.L. Aguilar, R.M. Pérez and A. Plaza, "Self-Organizing Map for Hyperspectral Image Analysis", VI International Work-conference on Artificial and Natural Neural Networks (IWANN), Granada, Spain, 2001.

[Martìnez 2001b]
P. Martínez, P. Aguilar, A. Plaza, M. Linaje, J.C. Preciado and R.M. Pérez. "Unsupervised Classification of Hyperspectral Images Using a Self-Organizing Neural Network", IX Spanish Remote Sensing Symposium. Lleida, Spain, 2001.

[Martinez 2001c]
P.L. Aguilar, R.M.Pérez, P.Martínez and A.Plaza. "A Self-Organizing Neural Network for Hyperspectral Unmixing", Revista Internacional Información Tecnológica, 12(6):185-192, December 2001.

[McFee 1997]
John E. McFee and Herb T. Ribley, "Detection of buried landmines using a CASI hyperpectral imager", Proceedings of SPIE Conference on Detection and Remediation Technologies for Mines and Mine-like Targets II. Vol. 3079, Orlando, FL, USA, 21-25 April, 1997, http://www.ccmat.gc.ca/TechReports/ReportsPDF/spie97_casiagi.pdf

[McFee 1999]
Steven B. Achal, Clifford D. Anger, John E. McFee and Robert W. Herring, "Detection of surface-laid mine fields in VNIR hyperspectral high spatial resolution data", Proceedings of SPIE Conference on Detection and Remediation Technologies for Mines and Mine-like Targets IV. Vol.3710, Orlando, FL, USA, 5-9 April, 1999, http://www.ccmat.gc.ca/TechReports/ReportsPDF/spie99_casi.pdf.

[Mosteller]
F. Mosteller, S. E. Fienberg and R. E. K. Rourke, "Beginning Statistics with Data Analysis", Addison-Wesley Publishing Company Inc., 1983.

[Muhammed]
H. H. Muhammed, P. Ammenberg, E. Bengtsson, "Using Feature-Vector Based Analysis, Based on Principal Component Analysis and Independent Component Analysis, for Analysing Hyperspectral Images", Centre for Image Analysis, Uppsala University, http://www.cb.uu.se/~hamed/iciap2001.html.

[Narendra]
N. Narendra and K. Fukunaga, "A Branch and Bound Algorithm for Feature Subset Selection", IEEE Trans. Computers, 26(9):917-922, September 1977.

[NNRC]
Neural Networks Research Center, "Neuroverkkojen esittely", Helsinki University of Technology, Press Release, www.cis.hut.fi/research/lehdisto/, viitattu 1.12.2002

[Ray]
Terrill W. Ray and Bruce C. Murray, "Nonlinear Spectral Mixing in Desert Vegetation", Remote Sensing of Environment, 55(1):59-64, January 1996.

[Ren]
Hsuan Ren, Quian Du and James Jensen, "Constrained Weighted Least Squares Approaches for Target Detection ans Classification in Hyperspectral Imagery", Proceedings of Igarss 2002.

[Richards]
John A. Richards and Xiuping Jia, "Remote Sensing Digital Image Analysis: an introduction", 3rd edition, Springer-Verlag, 1999.

[Richards 1996]
John A. Richards, "Classifier Performance and Map Accuracy", Remote Sensing of Environment 57(3):161-166, September 1996.

[Roberts 1997]
D. A. Roberts, R. O. Green and J. B. Adam, "Temporal and spatial patterns in vegetation and atmospheric properties from AVIRIS", Remote Sensing of Environment, 62(3):201-312, December 1997.

[Roberts 1998]
D.A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer and R.O. Green, "Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models", Remote Sensing of Environment, 65(3):227-375, September 1998.

[Robila]
S. A. Robila and P. K. Varshney, "Target Detection in Hyperspectral Images Based on Independent Component Analysis", Proceedings of AeroSense 2002, SPIE.

[Sabol]
Sabol, D. E., Adams, J. B., and Smith, M. O. (1992), "Quantitative sub-pixel spectral detection of targets in multispectral images", Journal of Geophysical Research, 97:2659-2672.

[Settle]
J. J. Settle, "On the Relationship Between Spectral Unmixing and Subspace Projection", IEEE Transactions on Geoscience and Remote Sensing, 34(4):1045-1046, July 1996.

[Sohn]
Youngsinn Sohn and N. Sanjay Rebello, "Supervised and Unsupervised Spectral Angle Classifier", Photogrammetric Engineering & Remote Sensing, 68(12):1271-1280, December 2002.

[Staenz]
K. Staenz, J. Schwarz, L. Vernaccio, F. Vachon and C. Nadeau, "Classification of Hyperspectral Agricultural Data with Spectral Matching Techniques", Proceedings of the International Symposium on Spectral Sensing Research (ISSSR 99) Las Vegas, Nevada, October 31 - November 4 , 1999, http://www.ccrs.nrcan.gc.ca/ccrs/rd/sci_pub/bibpdf/4736.pdf.

[Tekes]
Tekes, "Neurolaskennan mahdollisuudet", Tekes-raportti 43/94, www.tekes.fi/julkaisut/raportit/43_94, päivitetty 10/1999, viitattu 1.12.2002.

[Thenkabail]
P. S. Thenkabail, R. B. Smith and E. De Pauw, "Evaluation of Narrowband and Broadband Vegetation Indices for Determining Optimal Hyperspectral Wavebands for Agricultural Crop Characterization", Photogrammetric Engineering & Remote Sensing, 68(6): 607-621, June 2002.

[Tu]
Te-Ming Tu, Hsuen-Chyun Shyu, Ching-Hai Lee and Chein-I Chang, "An oblique projection approach for mixed pixel classification in hyperspectral images", Pattern Recognition 32(8):1399-1408, August 1999.

[Törmä]
Markus Törmä, "Comparison Between Three Different Clustering Algorithms", Institute of Photogrammetry and Remote Sensing, Helsinki University of Technology.

[Törmä 1997]
Markus Törmä, "Neuraaliverkot ja niiden käyttö kuvien analysoinnissa", Maankäyttö 1/1997, s. 6-9.

[Verbyla]
Dave Verbyla and Tim Hammond, "How to Lie With an Error Matrix", Department of Forest Sciences, University of Alaska Fairbanks, http://nrm.salrm.uaf.edu/~dverbyla/online/errormatrix.html.

[Wang]
C.-M. Wang, S.-C¨. Yang, P.-C. Chung, C.-I Chang, C.-S. Lo, C.-C. Chen, C.-W. Yang, C.-H. Wen, "Orthogonal subspace projection-based approaches to classification of MR image sequences", Computerized Medical Imaging and Graphics 25(6):449-540, November - December 2001.

[Webb]
Andrew R. Webb, "Statistical Pattern Recognition", 2nd Edition, John Wiley & Sons, 2002.

[Wilson]
H. G. Wilson, B. Boots and A. A. Millward, "A Comparison of Hierarchical and Partitional Clusterin Techniques for Multispectral Image Classification", Proceedings of Igarss 2002.

[XING]
Y. Xing and R. B. Gomez, "Hyperspectral Image Analysis using ENVI (ENvironment for Visualizing Images)", Proceeding of AeroSense 2001.