

Bilag 10. Korrelationsmetoder og resultater

GEUS: Jim Rasmussen, Per Rosenberg

I dette bilag beskrives først de to metoder anvendt ved korrelationsanalyserne, efterfølgende gennemgås fremgangsmåde ved datafortolkningen og til slut gives en uddybende dokumentation for resultaterne i hovedteksten.

Korrelationsmetoder

I korrelationsanalyser undersøges det om der eksisterer sammenhænge mellem forskellige variable, således at oplysninger om den ene type variable kan bruges til at forudsige værdierne af den anden type variable (kovarians), for eksempel mellem simple jorddata og bindingen af et bestemt pesticid til den pågældende jordtype. Det kan dreje sig om enkle situationer, hvor værdierne af to målte variable testes mod hinanden (bivariat sammenhænge) eller komplicerede undersøgelser, hvor mange variable sammenholdes på én gang (multivariat sammenhænge). I naturlige systemer er det yderst sjældent der findes generelt anvendelige bivariater sammenhænge mellem variable der ikke er direkte afledt af hinanden. Korrelationsanalyserne er i projektet brugt dels til at vurdere, om der er indbyrdes sammenhænge mellem de variable som direkte karakteriserer pesticidudvaskningen (modelparametrene), dels til at afklare om modelparametrenes værdier kan fastlægges ved en regressionsanalyse ved hjælp af simple jorddata. Disse variable kan efterfølgende bruges til at kortlægge udvaskningsforholdene, og dermed være med til at pege på et areals følsomhed. En anden meget vigtig anvendelse af korrelationsanalyser er til studier af sammenhænge mellem variable. Gennem en undersøgelse af strukturerne i et datasæt kan de mest betydende variable udpeges og der kan opnås forståelse af hvilke variable der kan undværes. Brugt proaktivt kan dette anvendes til optimering af analysestrategier i et projekt.

En række variable er bestemt gennem analyser udført på delprøver af samme jordprøve. Hermed er der opbygget et samlet sæt af oplysninger om jordprøvens simple jordparametre (bl.a. pH, indhold af organisk stof, sand og ler), jordprøvens evne til at tilbageholde vand, potentialitet for nedbrydning og binding af pesticid. To metoder til undersøgelse af sammenhænge i data er anvendt i projektet, neurale netværk og partial least square regression (PLS).

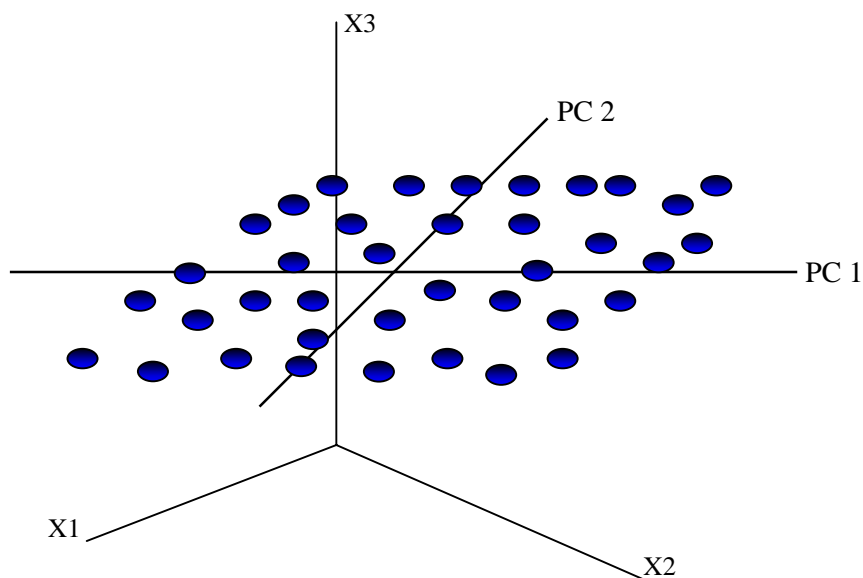
Neurale netværk. Neurale netværk er karakteriseret ved en algoritme der ved hjælp af et læredatasæt, det vil sige et datasæt hvor udfaldsrummet er kendt, kan opbygge en sammenhæng mellem input data og output data (resultatet). Neural netværksteknik er her brugt til at beskrive de hydrauliske modelparametre.

Styrken i neurale netværk er at der ikke behøver at eksistere eller forudskikkes en lineær sammenhæng mellem input variable og resultat. Til gengæld fordres der normalt et stort læresæt, og det er ikke muligt direkte at beskrive datasammenhængen og dermed, hvordan netværket er kommet frem til resultatet.

Anvendelsen af neurale netværk er veletableret, og almindeligt anerkendt, i forbindelse med prædiktions af hydrauliske inputdata til modeller som f.eks. MACRO.

Partial Least Squares Regression (PLS). I analysen bruges et princip om at finde kombinationer af variable (latente variable) til at beskrive tendenserne i et datasæt. Normalt er variable ikke indbyrdes uafhængige, hvilket gør en metode som Multiple Lineær Regression (MLR) uegnet. For at komme ud over dette problem, projiceres variablerne ind på retninger der er indbyrdes uafhængige, såkaldte latente variable. Latente variable er således linearkombinationer af de oprindelige variable, med det krav at de skal være vinkelrette på hinanden (uafhængige). I et XY-plot mellem to variable kan der tegnes en regressionslinie, som beskriver sammenhængen mellem de to variable. Ved at gøre dette, er det muligt at bruge regressionslinien eller regressionsvektoren som latent variabel i beregningerne i stedet for de oprindelige

to variable. Hermed er et todimensionelt problem reduceret til et endimensionelt. Fejlen ved at gøre dette svarer til afstanden mellem de enkelte punkter og regressionslinien. Denne fejl minimeres ved en mindste kvadraters metode. Hvis der er mange variable er fremgangsmåden mere kompliceret, figur 10.1. Med analogi til XY-plottet angives først den mest udtalte retning i datastrukturen. Herefter angives en ny retning som beskriver datasættet bedst, samtidigt med at den er vinkelret på retningen af den første latente variabel. Denne proces fortsættes indtil antallet af frihedsgrader er opbrugt eller at beskrivelsen af datasættet ikke øges yderligere. Ved PLS optimeres udvælgelsen af latente variable således at der både tages hensyn til beskrivelsen af de variable, der indgår i korrelationen (de uafhængige variable), og evnen til at beskrive den afhængige y variabel efter en veldefineret og standardiseret algoritme. Der henvises til Brereton (1990) eller Esbensen (2002) for en mere uddybende beskrivelse af metoden.



Figur 10.1. Et datasæt med tre variable kan illustreres perspektivistisk; princippet i illustrationen gælder også for et vilkårligt højere antal variable (i denne rapport arbejdes der med op til 20 variable af gangen). Ved multivariat korrelation beregnes et lille antal lineære kombinationer af de oprindelige variable (kaldet PC1, PC2, ...), som tillader afbildning ned i et overskueligt, lavt antal dimensioner (ofte kun to eller tre). Figuren viser hvordan et tredimensionalt problem enkelt kan udtrykkes i bare to dimensioner (et plan). Projektionsfejlen i forhold til det oprindelige datasæt, er den vinkelrette afstand mellem punkterne og planet udspændt af de to principal komponenter. Illustration efter Esbensen (2002, figur 3.11).

Figur 10.1 viser udspændingen af et datasæt i et tre dimensionalt rum, hvor PC1 og PC2 (principal komponent) betegner de to vigtigste retninger i 3D-rummet. PC1 og PC2 fremkommer gennem en projektion af det oprindelige datasæt, hvor de indre sammenhænge i et datasæt bestående af en række prøver (objekter) og en række parametre (variable) undersøges. I projektet er anvendt multivariat regression, specielt PLS, som giver mulighed for at se på sammenhænge mellem to datasæt. De to datasæt kan f.eks. bestå af vanskeligt tilgængelige DT50 værdier i det ene sæt, og simple billige analyseparametre i det andet sæt. .. Ved at reducere antallet af dimensioner i det ene datasæt ved hjælp af latente variable, kan der etableres en simpel regressionsrelation mellem det ene og det andet datasæt. Denne korrelation kan så anvendes til at prædiktion af f.eks. DT50 værdier ud fra simple jorddata, og altså medføre at dyre og vanskelige analyser kan erstattes med simple og billige.

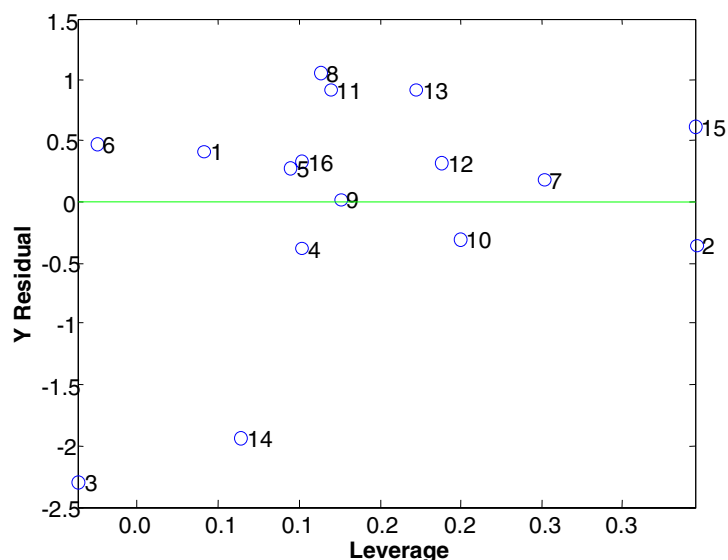
Regressions metoden anvendt i dette projekt kaldes for Partial Least Squares Regression og forkortes PLS-R eller blot PLS. Metoden udmærker sig ved at optimere udvælgelsen af latente variable, sådan at de bedst beskriver både det uafhængige datasæt og det afhængige datasæt.. Det betyder at korrelationerne ofte kommer til at bestå af færre latente variable end ved andre metoder og dermed bliver mere simple og lette at fortolke. De latente variable i denne metode kaldes for "PLS-komponenter" analogt med "principal komponenter" ved principal component regression og analyse. PLS-regressionsmetoden er begrænset til at beskrive lineære sammenhænge. Dette betyder at rådata i situationer med ikke lineære sammenhænge må transformeres, f.eks. ved en log transformation, for at kunne anvende metoden. Er det ikke muligt af finde en passende transformation, kan det være nødvendigt f.eks. at anvende neurale netværk til at beskrive sammenhængen..

Korrelationerne er i projektet valideret ved krydsvalidering enten efter "leave one out" eller segment principperne. Validerings principperne går alle ud på, iterativt og ved forskellige udvælgelses metoder at fjerne variable eller objekter for regressionskorrelationen. Herefter laves ny korrelation og virkningen af udeladelsen undersøges. Krydsvalidering anvendes når det ikke er muligt eller er for dyrt at designe undersøgelserne således at der kan etableres både et læredatasæt og et valideringsdatasæt.

Identifikation af outliers er sket ved at vurdere den enkelte prøves indflydelse (leverage) på korrelationen i sammenhæng med korrelationens (manglende) evne til at prædiktere værdien for prøven (Y residual, fejl på de enkelte y prædiktioner). På figur 10.2 er vist et plot af "leverage mod Y residual", hvor det ses at prøve 14 er en outlier. Prøve 14 udviser en relativ stor fejl på prædiktionen af prøvens værdi samtidigt med at prøven har høj indflydelse på korrelationen (leverage). Prøve 3 derimod er ikke en outlier, idet prøven på trods af den dårlige prædiktion (høj Y residual) ikke har indflydelse på korrelationen. I sidstnævnte tilfælde ville det være forkert at fjerne prøven, idet denne er med til at beskrive usikkerheden på prædiktionerne ved hjælp af korrelationen.

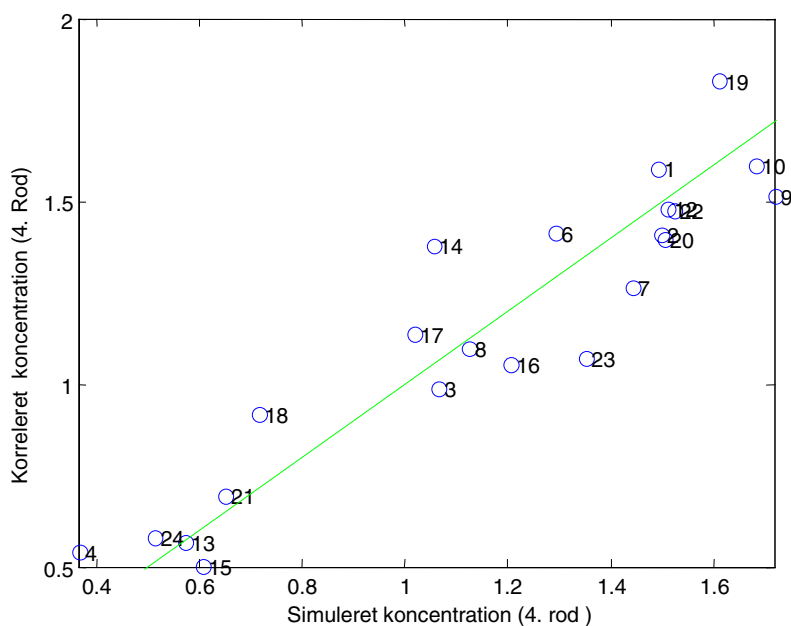
Outliers opstår sædvanligvis ved at nogle prøver udviser en unormal "adfærd" for eller imellem de variable der har betydning for korrelationen. Er det f.eks. sædvanligt at prøver med højt c-total indhold også har et højt lerindhold, og har disse to variable stor betydning i korrelationen, vil prøver med f.eks. højt indhold af c-total og lavt lerindhold typisk være outliers. Outliers kan også opstå hvis de ligger i yderkanten af den variationsbredde der er defineret i gennem korrelationen, hvis det er en eller få prøver der giver en ekstra udstrækning af variationsbredden for korrelationen. Endeligt kan outliers opstå ved analysefejl eller prøveudtagningsfejl. I sådanne tilfælde skal outliers fjernes helt fra datasættet. Det er meget væsentligt at bemærke at outliers i en korrelation, ikke nødvendigvis er outliers i en anden korrelation. Korrelationerne vil udvise forskellig afhængighed af de samme variable, og en "skæv" variationsbredde i en korrelation vil måske slet ikke slå igennem i en anden.

Selv om man naturligvis kan få nogle ideer til årsagen til at pågældende prøver er outliers ved at studere rådata, f.eks. tabel 10.2, er det ikke altid muligt at få en entydig årsag, idet outliers netop defineres i en multivariat sammenhæng. I stedet kan man projicere outliers ind på korrelationen, se hvordan de plotter i forhold til de prøver der indgår i korrelationen, samt forsøge at tolke deres egenskaber i det reducerede variabel rum ved hjælp af en forståelse af de latente variables betydning. Dette er illustreret i figur 10.7 og figur 10.8 .



Figur 10.2. Prøvernes indflydelse (Leverage) på korrelationen mod fejlen i korrelationsevnen ved prædiktion af prøvens værdi (Y Residual). Plottet bruges ved udpegnung af outliers i forbindelse med en PLS korrelationsanalyse

y: korreleret vær-

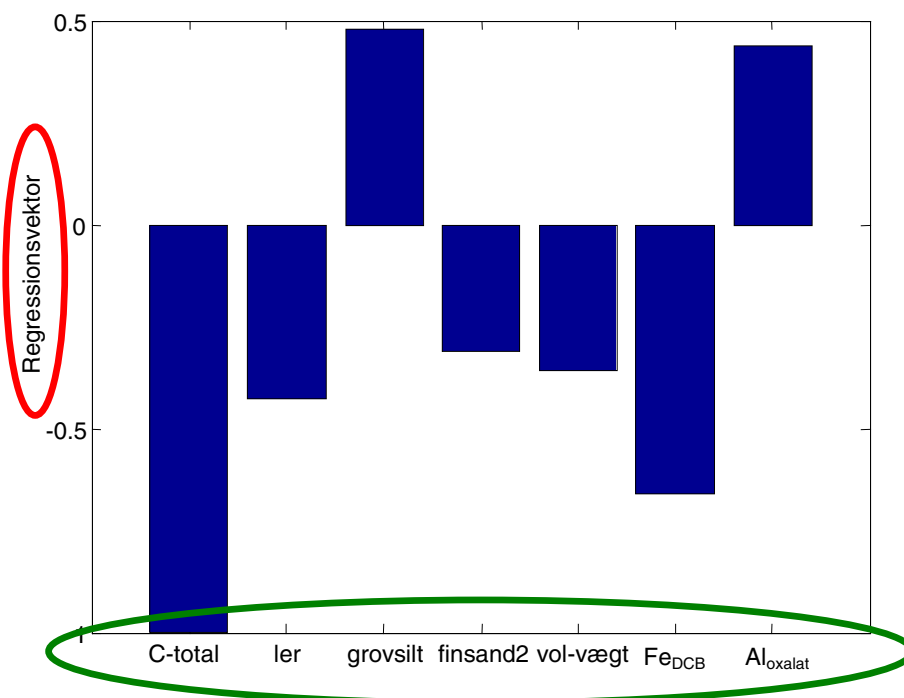


Figur 20a. Simuleret koncentration i 2 meters dybde (4.rod) mod korreleret koncentration i 2 meters dybde (4.rod) for MTB fra de 24 fuldprofiler. Baseret på målt binding og forsvinding. **Varians forklaret: X: 51%; Y: 87%; 3 PLS-komponenter.** **Outliers: 2/24. Outlier identifikation:** se tabel 10.10 i bilag 10.

Figur 10.3. Eksempel på figur i hovedteksten der viser korrelationens prædiktionsevne. De tilføjede fremhævelser er forklaret yderligere i teksten herunder.

Resultaterne fra korrelationsanalyserne illustreres i figurer med korrelationens evne til at prædiktere prøvernes værdi (figur 10.3), og de uafhængige variables betydning regnet med fortegn for korrelationen (regressionen), se figur 10.4. På de to figurer er der taget udgangspunkt i et konkret eksempel fra rapporten, og vist korrelationen af de vigtigste jorddata til den simulerede udvaskning af MTB.

Figur 10.3 gengiver prædiktionsplottet for korrelationen af centrale jorddata til udvaskningen af MTB. På x-aksen vises den målte værdi (i dette tilfælde en simuleret koncentration), mens der på y-aksen vises den korrelerede værdi. I figurteksten angives det hvor godt korrelationen forklarer variationen (hvor stor en del af variationen der er forklaret i modellen) i henholdsvis uafhængige (X) og afhængige (Y) variable (rødt). Det er vigtigt at notere at jo højere/betere forklaringen er af Y, jo bedre er korrelationen. På denne måde gives også et billede af usikkerheden ved prediktionen. Det angives hvor mange PLS-komponenter korrelationen benytter (grønt) og endelig oplyses antallet af outliers (2 ud af 24 prøver), samt hvor der kan indhentes yderligere information om outliers. Den grønne linie på figuren angiver den ideelle regressionslinie for korrelationen; jo tættere prøverne ligger på denne linie, jo større overensstemmelse mellem den målte og den korrelerede værdi.



Figur 20b. Den relative betydning af jorddata (regressionsvektorer for PLS-model) for simuleret koncentration i 2 meters dybde af MTB baseret på målt binding og nedbrydning.

Figur 10.4. Eksempel på figur i hovedteksten, der viser de uafhængige variables indflydelse på korrelationens evne til prædiktion af de afhængige variable. De tilføjede fremhævelser er forklaret yderligere i teksten herunder.

Figur 10.4 viser det andet plot, regressionsplottet, der altid angives som resultat fra korrelationsanalyserne. Af regressionsplottet fremgår de uafhængige variables (markeret med grøn cirkel) indflydelse på korrelationen. Regressionsvektorerne (rød cirkel) viser størrelsen af indflydelsen, samt om der er tale om negativ eller positiv regression til den afhængige variabel. I eksemplet på figur 10.4 har C-total en negativ regression til den simulerede koncentration af MTB i 2 meter. Det betyder at en stigning i C-total vil medføre et fald i den simulerede kon-

centration. Modsat har Al_{oxalat} en positiv korrelation til den simulerede koncentration af MTB. I dette tilfælde vil en stigning i Al_{oxalat} medføre en stigning i den simulerede koncentration af MTB. Figuren viser også at indflydelsen fra C-total er større numerisk set end indflydelsen fra Al_{oxalat} , hvilket er angivet ved barens længde.

PLS korrelationsanalyserne er udført med enten programmet Unscrambler version 8.0.5 fra CAMO eller med programmerne Matlab version 6.5 (release 13) fra Mathworks inc. med tilføjesprogrammet PLS_Toolbox version 3.0 fra Eigenvector Research inc.

Cluster analyse. Clusteranalyse kan anvendes på samme datastrukturer som f.eks. PLS, det vil sige der skal være en række prøver, objekter, med en tilhørende række variable, og disse skal være bestemt for alle de objekter der skal indgå i cluster analysen.

Cluster analyse er et velegnet værktøj til på simpel og visuel måde at finde strukturer i datasæt baseret på de indgåede variable, det vil sige at afgøre hvor godt prøverne ligner hinanden, om der er grupperinger og om der er prøver der er afvigende i forhold til disse grupperinger.

I PLS regressionen er afhængigheden til f.eks. jorddata altid vist i et plot, fx. som det der er vist i figur 10.4, også kaldet regressionsvektoren. Plottypen kan fx. vise hvilken afhængighed der er mellem sorption og jorddata. For at undersøge om denne afhængighed mellem sorption og jorddata er vidt forskellig for forskellige pesticider, eller om der findes similariteter og grupperinger, kan der laves en cluster analyse på de regressionsvektorer der er opnået i PLS regressionen. På den måde afdækkes det i hvor høj grad de samme jorde vil være følsomme/ikke følsomme m.h.t. simuleret udvaskning overfor hovedparten af pesticiderne, og hvilke pesticider man må undersøge separat.

Clusteranalysen foregår ved en parvis beregning af hvor stor forskel der er på (afstand der er mellem) forskellige prøver (pesticider). Identiske prøver har afstanden 0. Fremgangsmåden er iterativt efter følgende princip:

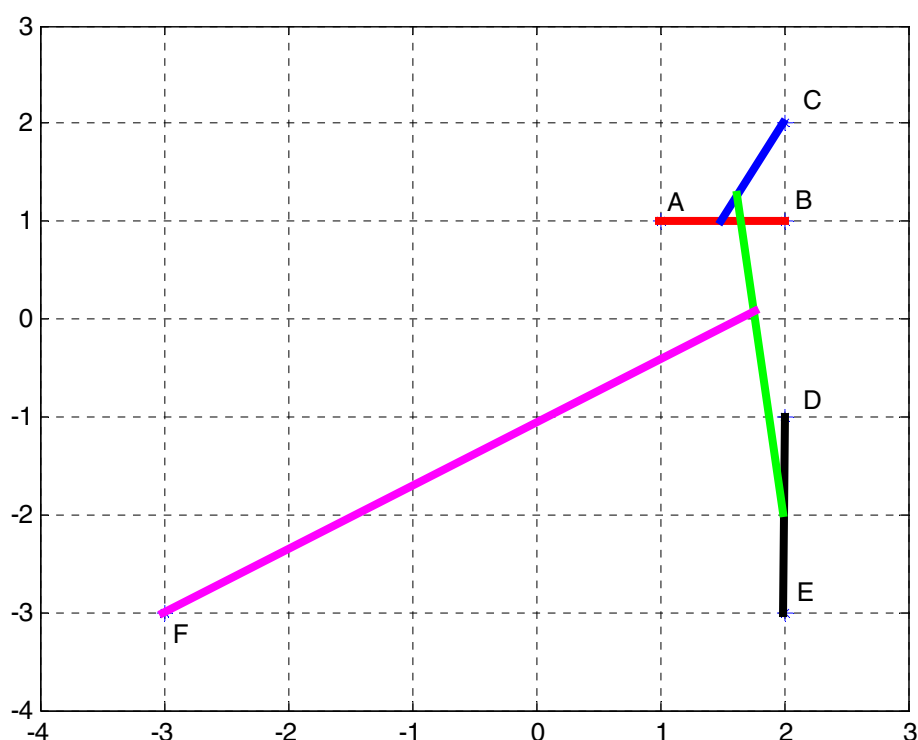
1. Analysen starter således med afstanden 0, som øges indtil man når afstanden mellem regressionsvektorerne for de to pesticider der ligner hinanden mest. For disse to pesticider, som udgør den første gruppe, udregnes gennemsnitssammensætningen (gennemsnitsregressionsvektoren) som indgår i den videre analyse i stedet for de to grupperede pesticiders regressionsvektorer, figur 10.5.
2. Dernæst øges afstanden indtil en af følgende hændelser indtræffer:
 - Et pesticid ligner den udskilte gruppes gennemsnitssammensætning, men i større afstand end for den første gruppe. Alle andre afstande er større. Pesticidet tilknyttes den første gruppe og en ny gennemsnitssammensætning for gruppen beregnes.eller
 - To pesticider som ikke er afstandsmæssigt knyttet til den første gruppe ligner hinanden med den korteste afstand mellem pesticiderne. Disse to pesticider udgør nu en ny gruppe, hvis gennemsnitssammensætning udregnes.eller
 - Afstanden mellem to grupper (to grupperes gennemsnit) er nu den korteste indenfor datasættet. De to grupper danner derfor en ny gruppe hvis gennemsnitssammensætning beregnes.
3. Processen fortsætter indtil alle pesticider og grupper er tilknyttet hinanden med et afstandskriterie.

Fremgangsmåden, der lyder kompliceret, kan illustreres med følgende simple konstruerede eksempel for et todimensionalt datasæt, tabel 10.1:

Tabel 10.1. Tabel over konstrueret simpelt datasæt.

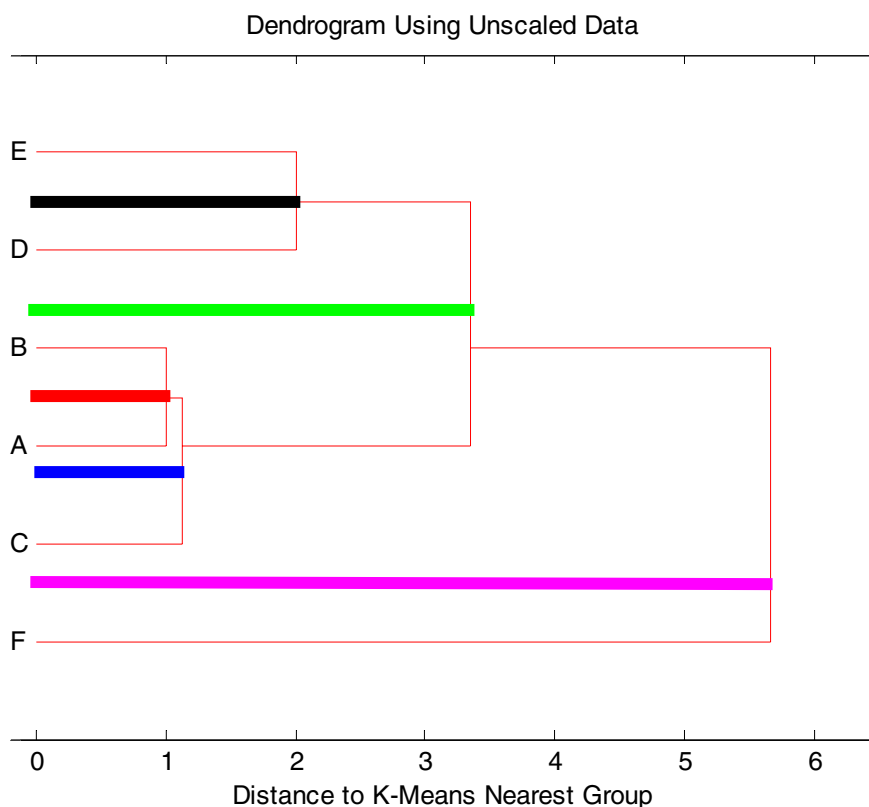
Pesticid	X (egenskab)	Y (egenskab)
A	1	1
B	2	1
C	2	2
D	2	-1
E	2	-3
F	-3	-3

Afstandene mellem pesticidernes egenskaber (regressionsvektorer) kan illustreres i et XY-plot, figur 10.5.



Figur 10.5. XY plot der viser afstandsberegning for datasættet i tabel 10.1. De enkelte pesticiders egenskaber er angivet ved gridnettets krydspunkter. De farvede linier forbinder de pesticider eller beregnede gennemsnit af pesticiders egenskaber som successivt gennem analysen ligger hinanden nærmest, startende med den røde linie og afsluttende med den lilla. Yderligere forklaring i teksten.

Fremstillingen af resultatet af clusteranalysen i figur 10.5 bliver uanvendelig ved multivariate sammenhænge, da man ikke kan visualisere afstande i et multidimensionalt rum.. Derfor illustreres resultatet (af datasættet i tabel 10.2) i praksis i et såkaldt dendogram, figur 10.6. De farvede linier i XY plottet figur 10.5 svarer til de farvede linier i dendogrammet figur 10.6 og illustrerer identiske afstande.



Figur 10.6. Dendrogram. Output fra clusteranalysen. Navnene på pesticiderne A-F (objekterne) fremgår i venstre side af dendrogrammet. Afstandene mellem (ligheder og forskelle mellem) pesticiderne er illustreret ved afstanden på x-aksen til den lodrette forbindelseslinie mellem to pesticider eller pesticidgrupper. Dimensionerne er bestemt af længden af de farvede linier som repræsenterer de samme afstande som i figur 10.5. Yderligere forklaring i teksten.

Pesticiderne (punkterne) A og B ligger (jævnfør ovenstående indledende forklaring) tættest og udgør den første gruppe, afstanden mellem A og B er markeret med rød linie. C ligger tættere på middelværdien af AB end de øvrige punkter og tilknyttes derfor denne gruppe, der nu består af AB og C, markeret med blå linie. Punktet D ligger tættere på E end på middelværdien af ABC, hvorfor DE udgør sin egen gruppe, markeret med sort linie. Gruppen DE ligger tættere på gruppen ABC end på punktet F, der har den største afstand både fra de øvrige pesticider (punkter) og grupper. Grupperne samles derfor i en storgruppe bestående af ABCDE. Afstanden mellem ABC og DE er markeret med grøn linie. Slutteligt findes afstanden fra F til gruppen ABCDE (markeret med magenta linie).

Grafisk kan cluster dendrogrammet tolkes som at AB og C udgør en gruppe der ligner hinanden meget. D og E ligner hinanden noget og ligner gruppen ABC mere end de ligner F, der må betegnes som forskellig fra alle de øvrige.

Metoden kan anvendes med et vilkårligt antal dimensioner af datasæt.

Fremgangsmåde ved datafortolkning

Af dette kapitel fremgår hvilken korrelationsmetode og hvilke variable, der er anvendt ved de enkelte korrelationsanalyser.

De mest betydningsfulde jorddata. De jordegenskaber der kan beskrive henholdsvis hydrauliske modelparametre, binding og nedbrydning er tilsammen de jordegenskaber, der skal anvendes som grundlag for udpegning af særligt pesticidfølsomme arealer.

Hydrauliske egenskaber

De hydrauliske egenskaber (vandretention samt umættet og mættet hydraulisk ledningsevne) beskrives ved funktionsudtryk foreslået af Brooks & Corey (1964) med Jarvis' (1991) udvidelse til to domæner til brug i modellen MACRO. Funktionsparametrene K_b , K_s , θ_b , θ_s , θ_r , λ (se signaturforklaring i hovedteksten figur 18). er bestemt for data fra fuld profiler (sandede jorde) ved en kombination af statistisk optimering og "eye-fitting". For K_s og θ_s , bruges de målte værdier. Ved neural netværksprocedure er sammenhængen mellem modelparametrene og jordegenskaber som tekstur, humus, og landskabelement bestemt. Der er udviklet forskellige korrelationer under anvendelse af forskellige inputparametre efter proceduren beskrevet i Børgesen (2003). Resultaterne er rangeret efter prædiktionssevner. Usikkerheden af de prædikterede hydrauliske egenskaber er bestemt ved bootstrapping (hvor det neurale netværk gentagne gange er optimeret på tilfældig udvalgte 2/3 af data og testet på den sidste 1/3 af data).

Det optimerede neurale netværk er brugt til at beregne de hydrauliske egenskaber for kvadratnetprofilen på sandjord.

Binding

Korrelationsanalysen af simple jorddata til bindingen af MTB, MCPA og MTA er udført med PLS-regression. Som forklaret i hovedteksten (kapitel 4.3) har det kun været muligt at lave korrelation til tre af de fire hovedstoffer i projektet. Til brug for simuleringen af stoffernes udvaskning blev K_d værdierne omregnet til K_f værdier (se bilag 8). I korrelationsanalysen for de tre hovedstoffer blev de omregnede værdier for bindingen anvendt. Udgangsmatrix af simple jorddata bestod af kornstørrelsesfordelingen, pH_{Kd} (pH målt i forbindelse med K_d bestemmelsen), C-total, Fe_{DCB} , $\text{Fe}_{\text{oxalat}}$, Al_{DCB} , $\text{Al}_{\text{oxalat}}$, CEC, H^+ ombyttelig, samt jordens specifikke overflade. Termerne er forklaret i bilag 4.

Nedbrydning

Korrelationsanalysen af simple jorddata til nedbrydningen af MTB, MCPA og MTA er udført med PLS-regression. Korrelationsanalysen er lavet til nedbrydningsraten (k) som er den målte størrelse og som anvendes til beregning af nedbrydningen (DT50). Udgangsmatrix af simple jorddata bestod af kornstørrelsesfordelingen, pH_{Kd} (pH målt i forbindelse med K_d bestemmelsen), C-total, Fe_{DCB} , $\text{Fe}_{\text{oxalat}}$, Al_{DCB} , $\text{Al}_{\text{oxalat}}$, CEC, H^+ ombyttelig, jordens specifikke overflade, samt mål for den biologiske aktivitet SIR, ASA, FDA, CFU_{TSA} og $\text{CFU}_{\text{Goulds}}$. Termerne er forklaret i bilag 4 og 3A (side 1).

Vægtning af de vigtigste jordegenskabers betydning. Formålet med disse korrelationsanalyser er at undersøge hvilke jordegenskaber der har indflydelse på udvaskningen, og hvilken betydning de har, når modelparametre for hydraulik, binding og nedbrydning inte-

greres i udvaskningssimuleringer med MACRO modellen. Korrelationsanalyserne er udført med PLS-regression.

Simuleringerne med MACRO modellen består i tre trin, hvor første trin udelukkende beskriver de hydrauliske forhold (9 K_f -sæt). I det andet trin anvendes den målte bindings-evne på de pågældende jorde af pesticidet og de hydrauliske forhold (udvaskning af MTB med fastholdt nedbrydning), mens det tredje trin giver en integreret beskrivelse af hydraulik, binding og nedbrydning, idet alle målte modelparametre anvendes. Korrelationsanalyserne til hver af de tre trin er gennemført med de relevante jordegenskaber. Hvilket for eksempel betyder at korrelationen til den simulerede koncentration af MTB, hvor nedbrydning er holdt konstant, har taget udgangspunkt i de jordegenskaber der beskriver henholdsvis de hydrauliske forhold og bindingen af MTB.

Følsomhed af A-horisontens tykkelse. Der er lavet en følsomhedsvurdering af betydningen af A-horisontens tykkelse. Denne er foregået ved at simulere udvaskningen for aktuelle iboende egenskaber, men lade a-horisonten variere på bekostning af b-horisonten. På den måde beskrives hvorledes den simulerede udvaskning varierer med a-horisontens tykkelse og det giver en indikation af med hvilken sikkerhed A-horisontens tykkelse skal bestemmes. Simuleringen er som i de øvrige simuleringer foretaget med fast nedbrydning, men med målt K_f værdi.

Resultater af datafortolkningen

I dette afsnit gives en uddybende forklaring vedr. de resultater der præsenteres i hovedteksten.

Vedr. hydrauliske modelparametre

Der henvises til hovedteksten, figur 17 kapitel 4.4.

Vedr. modelparametre for binding

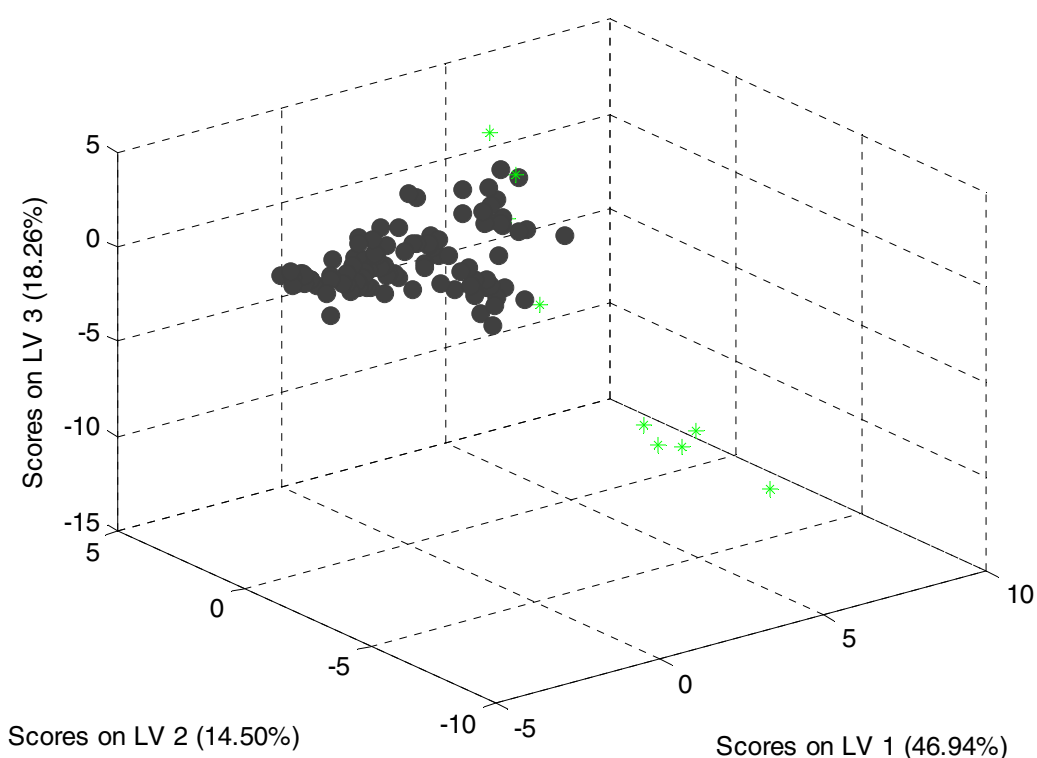
Korrelationsanalyser mellem jordegenskaber og binding (K_f) for stofferne: MTB, MCPA og MTA:

MTB

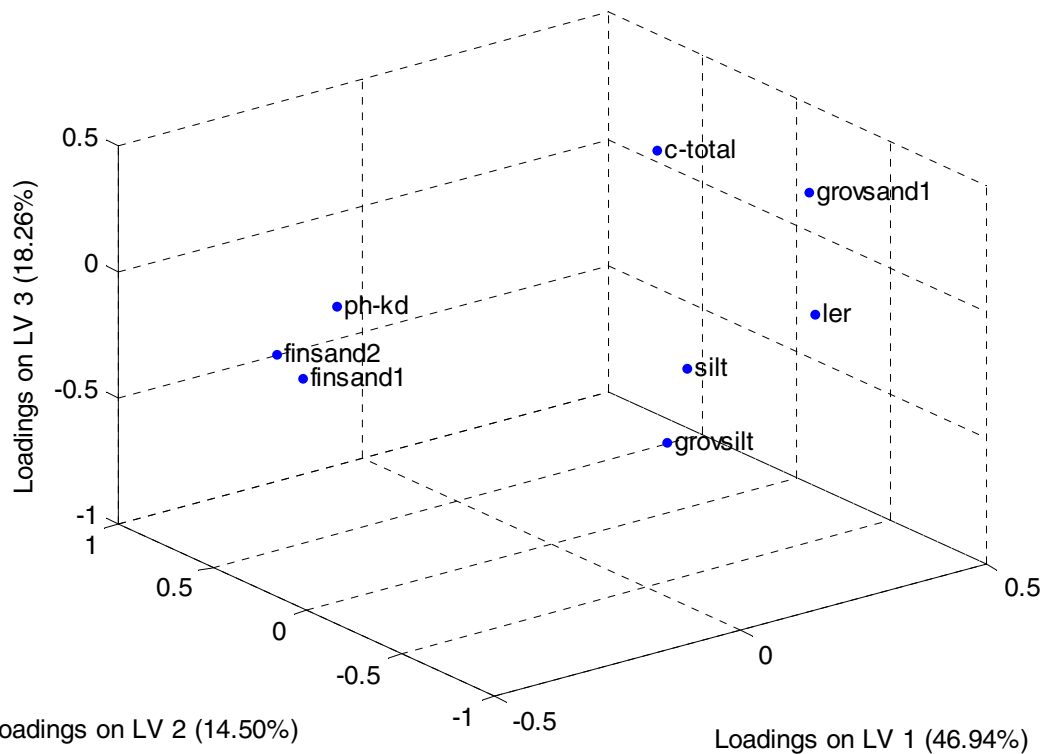
Der er identificeret outliere for korrelationen mellem jordegenskaber og binding for MTB, tabel 10.2 og figur 10.7. Bemærk at det kun drejer sig om outlier detektion i den population af prøver der har K_f værdier over 0.5. For de øvrige resultater henvises der til hovedteksten, figur 26a og 26b.

Tabel 10.2. Outliers i korrelation til K_f for MTB. Tabellen giver oversigt over målte data for de objekter der er identificeret som outliers. Generelt set er ourlierne karakteriseret af høje indhold af ler og de to siltfractioner.

Kupa ID	K_f	ph-kd	c-total	ler	silt	grovsilt	finsand1	finsand2	grovsand1
04-2-1-260	4.03	4.19	2.53	3.60	2.90	1.00	4.80	7.80	58.70
13-2-5-660	0.76	4.20	0.06	12.90	6.60	10.40	13.70	18.00	28.40
13-3-2-612	0.52	4.88	1.44	7.40	9.10	8.90	17.80	21.60	25.60
14-1-3-678	0.50	4.11	0.08	12.80	8.70	13.80	14.80	11.60	31.10
14-1-5-680	1.17	4.01	0.08	12.30	8.20	13.40	11.20	4.90	41.90
15-1-1-681	1.77	6.53	2.88	6.20	8.80	11.90	13.80	13.20	33.80
15-1-4-684	4.38	4.26	0.29	17.80	8.20	11.20	9.50	10.80	28.80
15-1-5-685	2.20	6.99	0.40	15.40	8.10	10.00	10.80	11.30	31.50
21-1-1-833	1.34	5.64	4.58	5.30	7.70	7.00	19.80	17.60	26.20



Figur 10.7. Score plot der viser projektionen af korrelationens objekter i det reducerede variabelrum beskrevet ved de tre latente variable (i tilfældet MTB findes der kun tre latente variable). De sorte punkter viser korrelationens objekter, mens de grønne stjerner viser outliersnes projektion ind på korrelationen. Bemærk at procenterne angivet ved de tre score vektorer angiver forklaringen på X, ikke forklaringen på Y.



Figur 10.8. Loading plot der viser projektionen af variable på de tre loading vektorer. Plottet kan bruges til at vurdere betydningen (forstået som variable eller egenskaber) af de tre latente variable.

Figur 10.7 viser at de fleste outliers ligger separeret fra de prøver der indgår i korrelationen. Disse outliers har en normal første komponent, det vil sige det er ikke på første latente variabel de udskiller sig, men derimod på anden og tredje.

Figur 10.8. viser at første latente variabel primært udspænder variationen pH, de to finsand på den ene side og c-total, ler og de to silt på den anden side. Latent variabel 2 udspænder grovsand og ler på den ene side og c-total på den anden side. Tredje latente variabel, der er den mindst betydende, udspænder de to silt på den ene side og grovsand på den anden side. Det er ikke helt nemt at overskue et 3 D plot, og i praksis vendes og drejes plottet, for at anskue det i forskellige vinkler, før der drages konklusioner. Tildeling af egenskaber til de latente variable kan bruges til en kvalitativ vurdering af årsagen til at visse prøver er outliers. I andre sammenhænge kan det desuden bruges til at vurdere datastrukturen, hvilket også er gjort i dette projekt, men ikke rapporteret.

Kombineres de to plot kan det forsøges at forklare årsagen til outliers. De outlier der udskiller sig klart fra hovedpopulationen, er det primært prøver der er rige i silt/grovsilt og ler fraktionen og samtidigt fattige i c-total fraktionen. Der er desuden en tendens til at de er rige i grovsand. For en to prøvers vedkommende er der tale om et ekstremt højt lerindhold.

For prøverne 260, 612 og 681 kan det diskuteres om disse er outliers. Fjernelse af dem har forbedret korrelationen en smule, men det kan sagtens vælges at lade prøverne indgå i korrelationen.

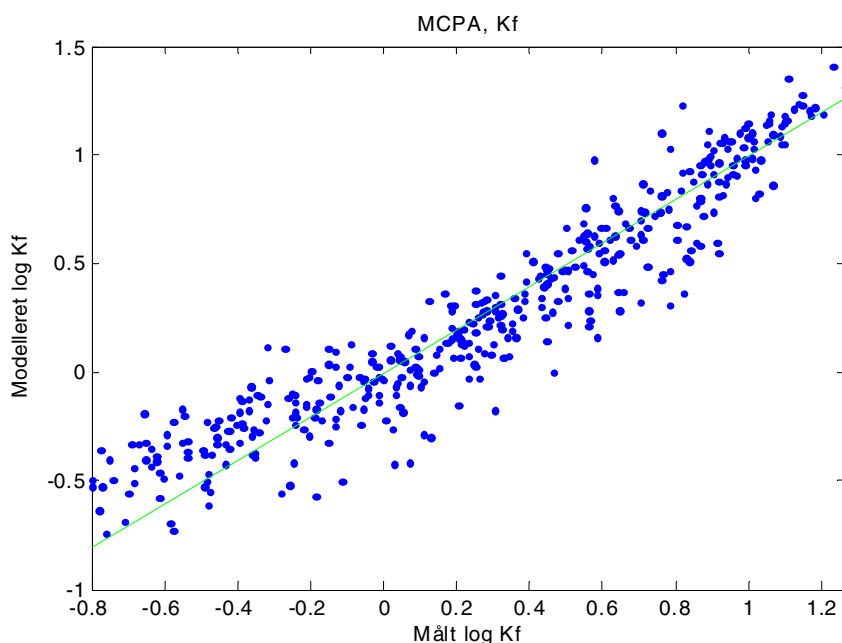
Prøve 833 er meget høj i c-total, og ligger på det punkt udenfor korrelationens variationsområde, men udviser i øvrigt normal sammenhæng mellem iboende egenskaber og bindingsevnen.

I tilfældet MTB har outliers ingen betydning for vurderingen af udvaskningsrisikoen, idet de alle har en K_f værdi der placerer dem i den ikke sårbare population i forhold til udvaskning. Bruges

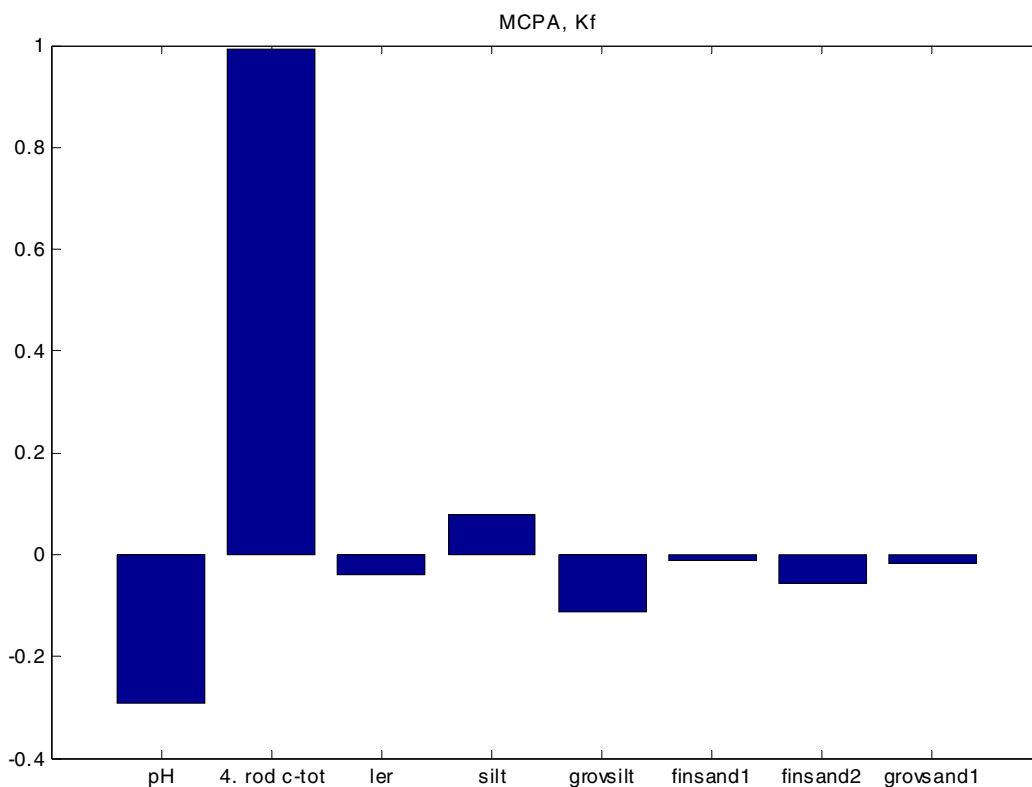
C-total alene som afskæringskriterie er der 4 prøver der fejlagtigt vil blive placeret i gruppen potentielt sårbare arealer. Ønskes det at gøre afskæringen mere sikker ved at inddrage de øvrige variable, vil deres usædvanlige sammensætning træde frem, og prøverne må behandles særskilt.

MCPA

Der er korreleret og identificeret outliers for korrelation mellem jordegenskaber og binding for MCPA, figur 10.9 og 10.10, samt tabel 10.3.



Figur 10.9. Målt binding (K_f) mod korreleret binding (K_f) for MCPA. Varians forklaret: X: 53%; Y: 89%. 3 PLS komponenter. Outliers: 4/435. Outlier identifikation: se tabel 6.2.



Figur 10.10. Den relative betydning af jorddata (regressionsvektorer for PLS-korrelationen) for binding (K_f) af MCPA.

Det er C-total der har størst indflydelse på bindingen af MCPA, hvor et stigende indhold af organisk kulstof medfører en øget binding. pH er negativt korreleret til MCPA, hvilket betyder at et fald i pH betyder en stigende binding. Kornstørrelsesfordelingen er ikke af væsentlig betydning for beskrivelsen af MCPAs binding.

Tablet 10.3. Outliers i korrelationsanalysen til K_f for MCPA

	Kf	ph-kd	4.rod c-total	ler	silt	grovstilt	finsand1	finsand2	grovсанд1
07-3-1-293	8.60	5.45	0.73	3.60	0.90	1.90	1.50	3.20	29.20
13-2-2-662	0.36	5.70	0.80	11.30	7.20	9.30	13.80	16.30	30.40
15-1-4-684	11.42	4.09	0.73	17.80	8.20	11.20	9.50	10.80	28.80
21-1-1-833	5.33	5.55	1.46	5.30	7.70	7.00	19.80	17.60	26.20

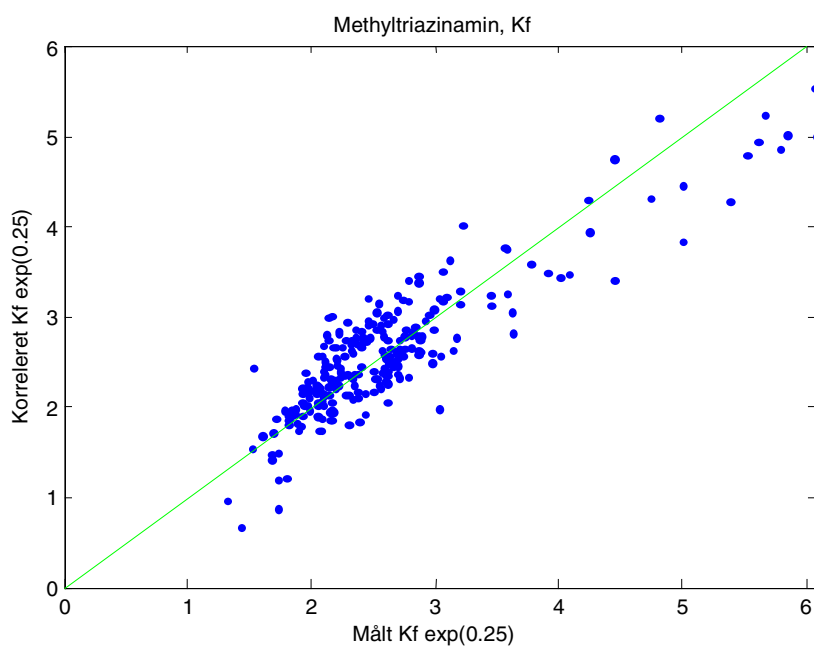
Laves en lignende projektion af outliers ind på korrelationen for MCPA som det skete for MTB, er det ikke muligt at pege på unormale egenskaber ved outlierens variabelsammensætning. Forsøges det at prædiktere K_f værdien for outlierne fås følgende værdier:

07-3-1-293: $K_f = 1.1342$
 13-2-2-662: $K_f = 1.1813$
 15-1-4-684: $K_f = 2.0025$
 21-1-1-833: $K_f = 28.2392$

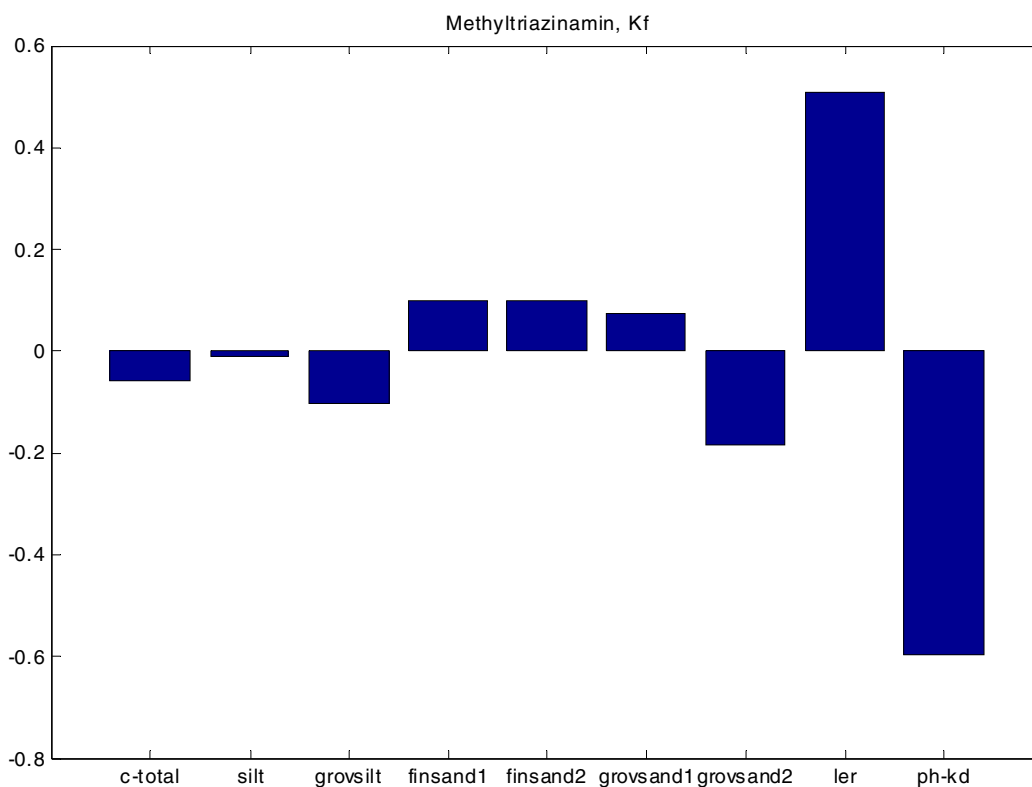
Afvigelsen mellem prædikeret K_f og målt K_f er tydeligvis årsagen til disse outliers. Denne afvigelse kan opstå som følge af at K_f kan være forkert, eller det kan skyldes at de pågældende prøver har afvigende egenskaber af betydning for K_f som der ikke er målt for. Det er ikke muligt at afgøre om den ene eller den anden forklaring er den rigtige. Imidlertid er antallet af outliers for K_f på mcpa meget lille, og kun i tilfældet med prøve 662 er der reelt en mulighed for at dømme en prøve ikke sårbar, selv om den kan være potentielt sårbar.

MTA

Der er korreleret og identificeret outliers for korrelationen mellem jordegenskaber og binding for MTA, figur 10.11 og 10.12, samt tabel 10.4.



Figur 10.11. Målt binding (K_f) mod korreleret binding (K_f) for MTA. Varians forklaret: X: 67%; Y: 79%. 3 PLS komponenter. Outliers: 4/253. Outlier identifikation: se tabel 10.4.



Figur 10.12. Den relative betydning af jorddata (regressionsvektorer for PLS-korrelationen) for binding (K_f) af MTA.

Bindingen af MTA er stærkt negativ afhængig af pH og positivt afhængig af ler. C-total spiller ikke nogen rolle for bindingen og hvis der er en effekt, så er den svagt negativ.

Det skal bemærkes at korrelationen ikke er særlig god, Den vil kunne anvendes til at vurdere om en prøves K_f værdi er høj eller lav, altså det er ikke muligt at komme med et godt estimat for K_f . I en vurdering af sårbarheden overfor MTA er det imidlertid fuldt tilstrækkelig at afgøre om en K_f værdi er høj eller lav, ligesom korrelationen kan bruges til at undersøge hvilke variable der har indflydelse på K_f .

Det skal desuden bemærkes at figur 10.11 indikerer at der er tale om to populationer der følger hver sin udvikling. Tages den ene population der hovedsageligt består af underjorde, kan der etableres en meget mere sikker korrelation. Denne er ikke vist.

Tabel 10.4. Outliers i korrelationen til K_f for MTA

id-kupa	K_f -middel	c-total	silt	grov-silt	finsand1	finsand2	grov-sand 1	grov-sand 2	ler	ph-kd
01-2-3-110	89.19206	0.17	14.4	18.1	21.8	10.2	12.4	2.2	20.6	5.61
15-1-3-683	68.12055	0.14	6.2	9.1	13.2	12.1	35.3	9.6	14.3	5.09
13-1-4-654	2.180527	0.06	0.9	1.8	25.3	18	47.3	4.5	2.1	4.72
15-1-5-685	767.3062	0.4	8.1	10	10.8	11.3	31.5	12.2	15.4	6.79

Igen er der fundet relativt få outliers ud af en temmelig stor population. Ingen outliers udviser lave K_f værdier og der er derfor ikke risiko fejlagtigt at bedømme disse ikke sårbare. For prøverne 110, 683 og 685 bærer den meget høje lerprocent årsagen til at de er outliers. For prøve 654 skyldes grovsand indholdet at prøven er outlier. Der er anvendt samme plots som for MTB, men med så mange punkter er det vanskeligt at illustrere disse sammenhænge, hvorfor figurerne er undladt her.

Det skal bemærkes at outlierne kun er "kosmetiske", forstået på den måde at fjernelse kun gør korrelationen lidt bedre, inkludering ødelægger den ikke. Prædiktionen på de fire outliers ville kunne gøres med nogenlunde samme sikkerhed som på de øvrige objekter, altså man kan med sikkerhed afgøre om der er tale om en høj eller en lav Kf værdi.

Vedr. korrelation til nedbrydningsparametre. Korrelationsanalysen er lavet til nedbrydningsraten (k) for de tre stoffer: MTB, MCPA og MTA.

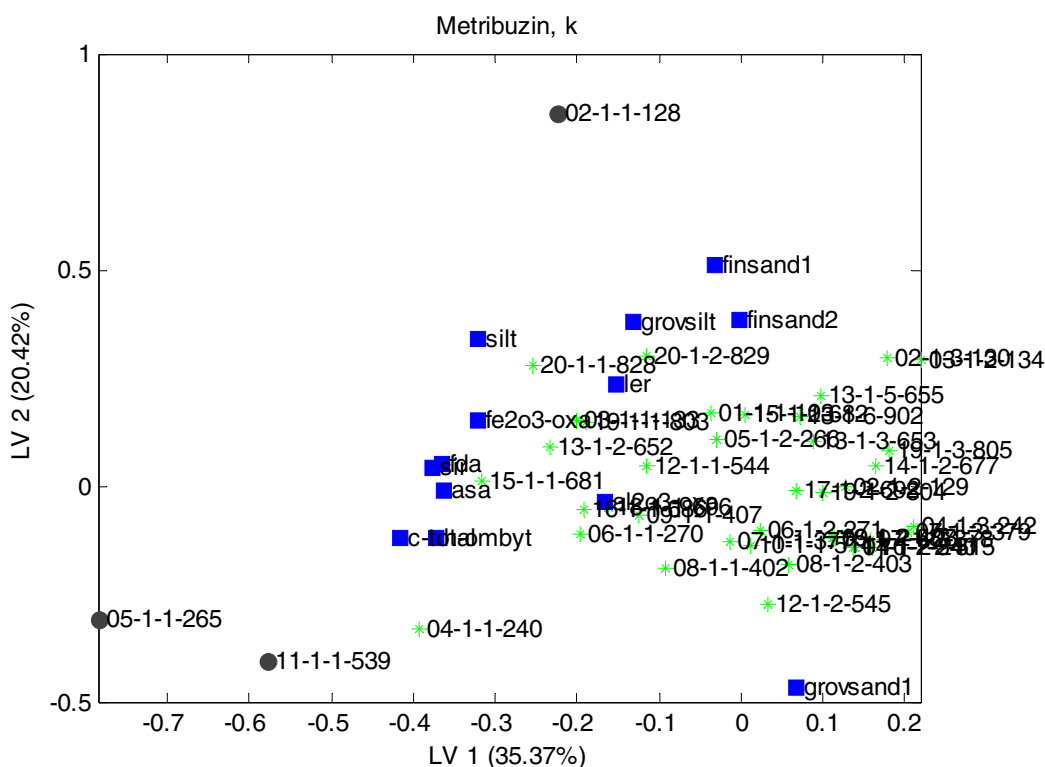
MTB

Der henvises til resultater i hovedteksten, figur 17a og 17b. Her gives dog en outlier identifikation, tabel 10.5.

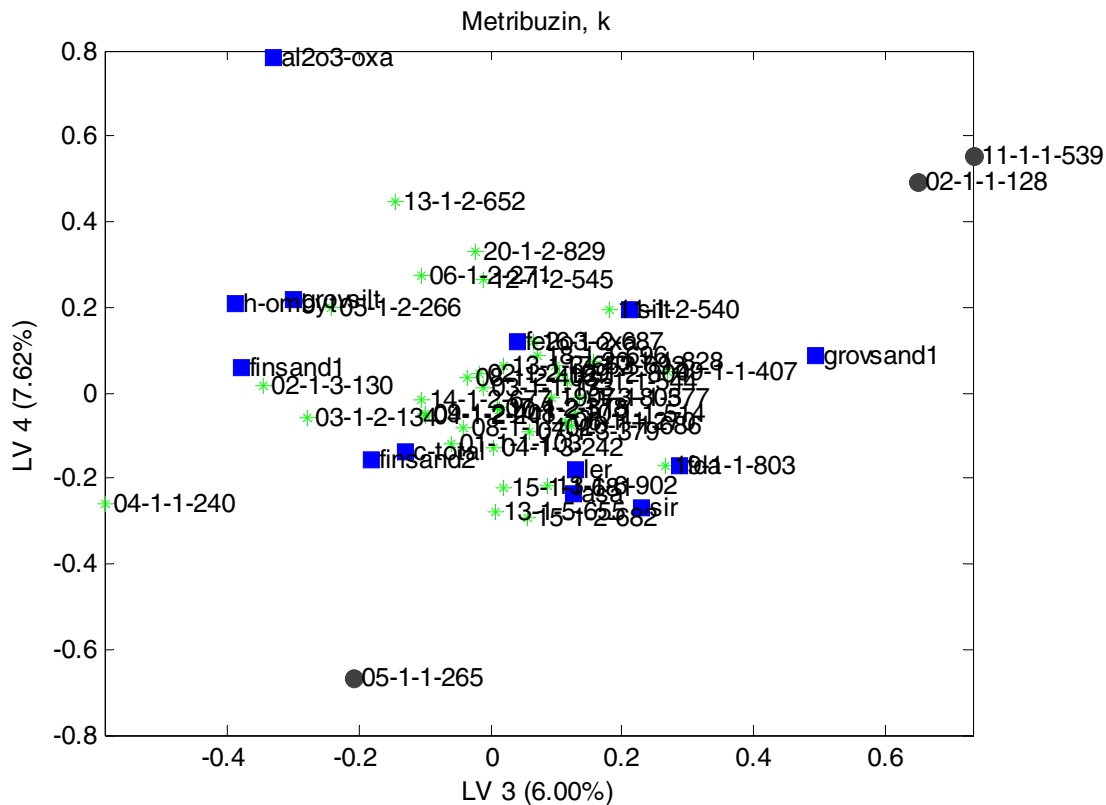
Tabel 10.5. Outliers i korrelationsanalysen til nedbrydningsraten (k) af MTB

	K	c-total	h-ombygt	fe2o3-oxa	al2o3-oxa	ler	silt	grovsilt	fin-sand1	fin-sand2	grov-sand1	asa	fda	sir
02-1-1-128	-0.0068	1.10	5.07	2454.19	1426.70	5.80	8.20	13.60	22.60	20.50	24.70	15.93	47.44	3.68
05-1-1-265	-0.0162	2.79	10.02	3367.28	1425.84	6.30	7.20	10.40	9.20	7.20	34.40	32.98	43.88	4.67
11-1-1-539	0.0001	2.38	6.72	3306.95	2097.23	4.30	6.20	3.10	5.10	13.80	59.50	32.28	41.23	4.32

I dette tilfælde er der 4 principal komponenter hvilket umuliggør 3 D plottene vist i figur 10.7 og 10.8. Imidlertid er det muligt at kombinere de to plot så både objekt strukturen og variabel strukturen kan studeres i et såkaldt bi-plot. For at afbilde de fire dimensioner er brugt to plots, figur 6.13 og 10.14.



Figur 10.13. Biplot der samtidigt viser objekt og variabel strukturen i MTB korrelationen med outlierne projiserede ind. Figuren viser de to første latente variable. Signaturer: blå firkanter er variablene, grønne stjerner er deobjekter der indgår i korrelationen og sorte pletter er outliers.

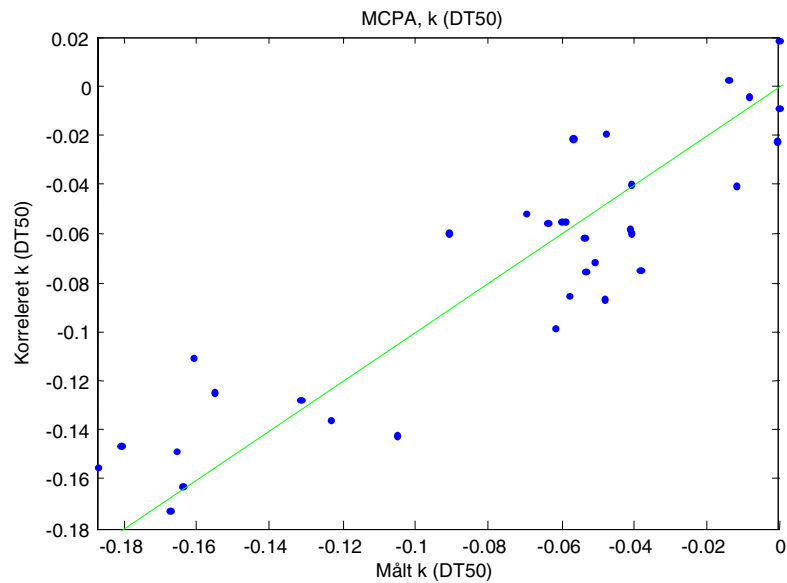


Figur 10.14. Biplot for MTB korrelationen med outliers projiseret ind. Figuren viser tredje og fjerde latente variabel. Signaturer: blå firkanter er variablene, grønne stjerner er de objekter der indgår i korrelationen og sorte pletter er outliers.

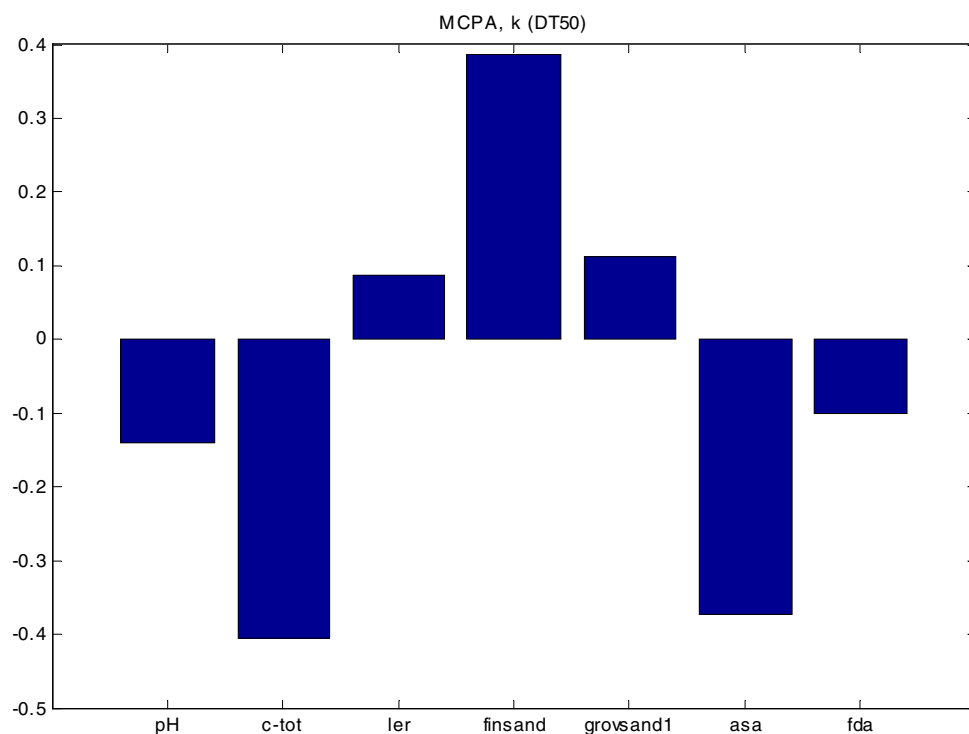
Biplottene i figur 10.13 og 10.14 viser korrelationens objekter markeret med grøn stjerne, variablene markeret med blå firkant og outliers markeret med sort cirkel. Især er figur 10.13 vigtig da den repræsenterer den største varians i datasættet. Umiddelbart er det uhyre vanskeligt at knytte entydige årsager til outlierne. Formentligt har det noget med den grove fraktion at gøre, ikke mindst for prøve 539 og det omvendte for 128, men også C-total og H⁺ ombyt syntes at spille en rolle. Årsagen til outlierne skal formentlig findes i at de alle bevæger sig i yderkanterne for korrelationens variationsbredde for en række variable, og dermed er vanskelige at korrelere.

MCPA

Der er korreleret og identificeret outliers for korrelationen mellem jordegenskaber og nedbrydning for MCPA, idet der er analyseret for komplette profiler, så vel som for over- og underjorde separat, figur 10.15.

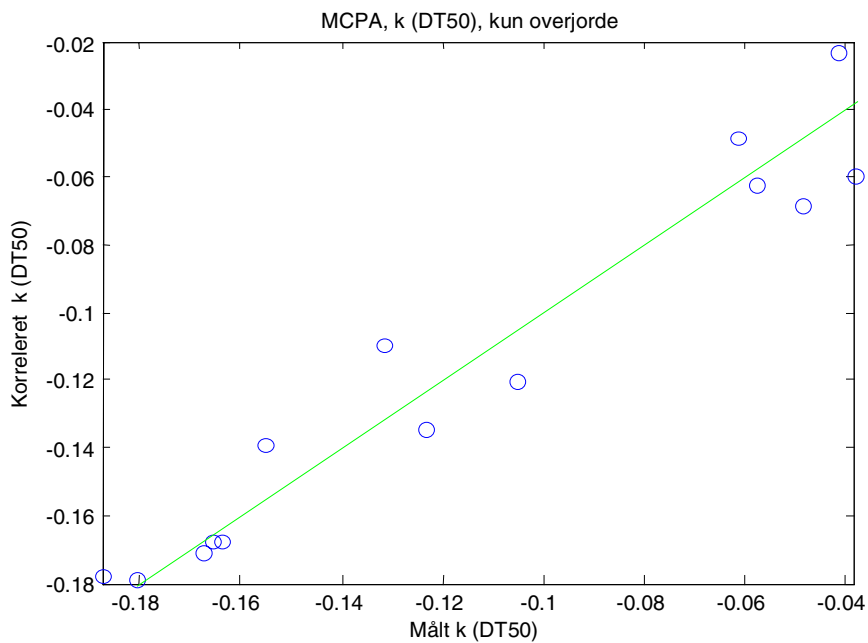


Figur 10.15. Målt nedbrydningsrate (k) mod korreleret nedbrydningsrate (k) for MCPA ved brug af hele datasættet. Varians forklaret: X: 70%; Y: 81%. 3 PLS komponenter. Outliers: 10/44. Outlier identifikation: se tabel 10.6

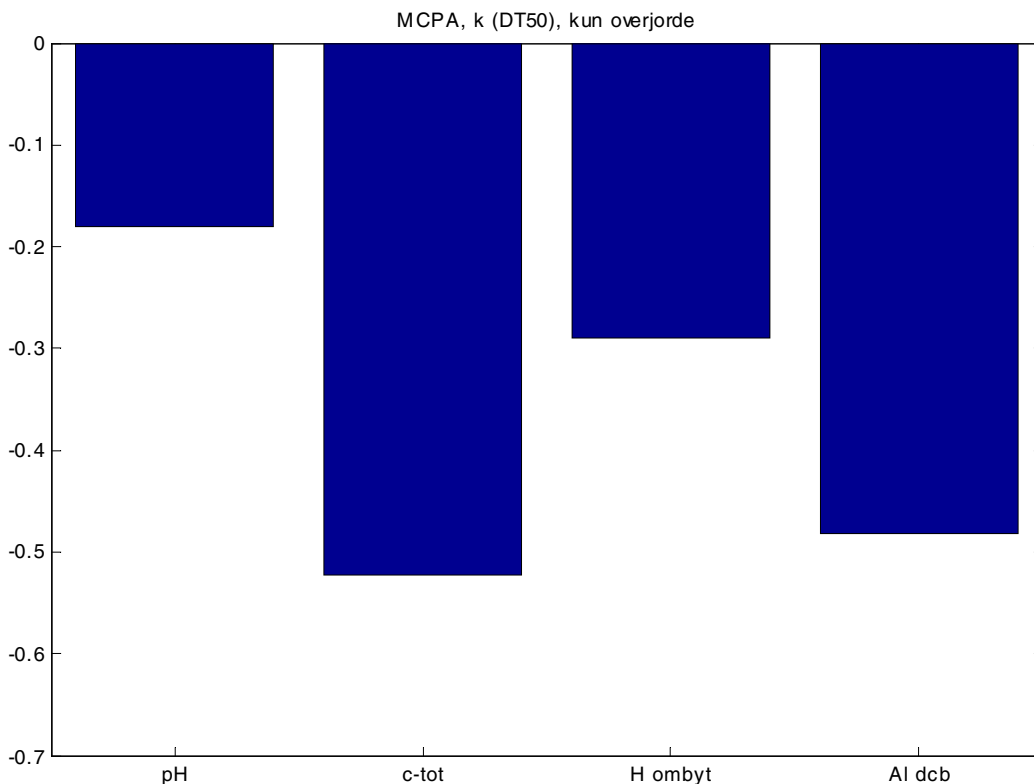


Figur 10.16. Den relative betydning af jorddata (regressionsvektorer for PLS-korrelationen) for nedbrydningen af MCPA ved brug af hele datasættet.

Når hele datasættet anvendes i korrelationsanalysen er det især C-total samme med ASA der har betydning for korrelationen, figur 10.16. Der er tale om en negativ afhængighed, hvilket betyder at en stigning i indholdet af organisk kulstof vil medføre en højere k -værdi (numerisk) og dermed en kortere DT50.



Figur 10.17. Målt nedbrydningsrate (k) mod korreleret nedbrydningsrate (k) for MCPA med prøver fra overjorde. Varians forklaret: X: 76%; Y: 94%. 2 PLS komponenter. Outliers: 3/17. Outlier identifikation: se tabel 10.8.



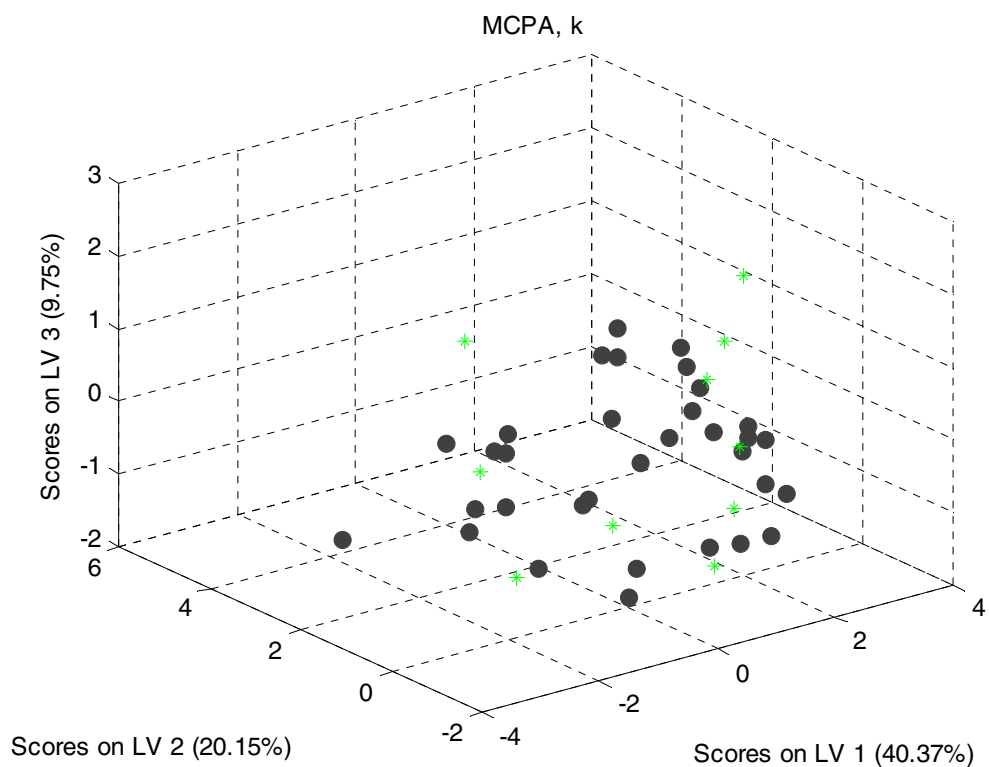
Figur 10.18. Den relative betydning af jorddata (regressionsvektorer for PLS-korrelationen) for nedbrydningen af MCPA med prøver fra overjorde.

I en korrelation, hvor det kun er prøverne fra overjorden, figur 10.17 og 10.18, der anvendes som datagrundlag, viser det sig igen at C-total er den vigtigste parameter, suppleret med et mål for indholdet af aluminium.

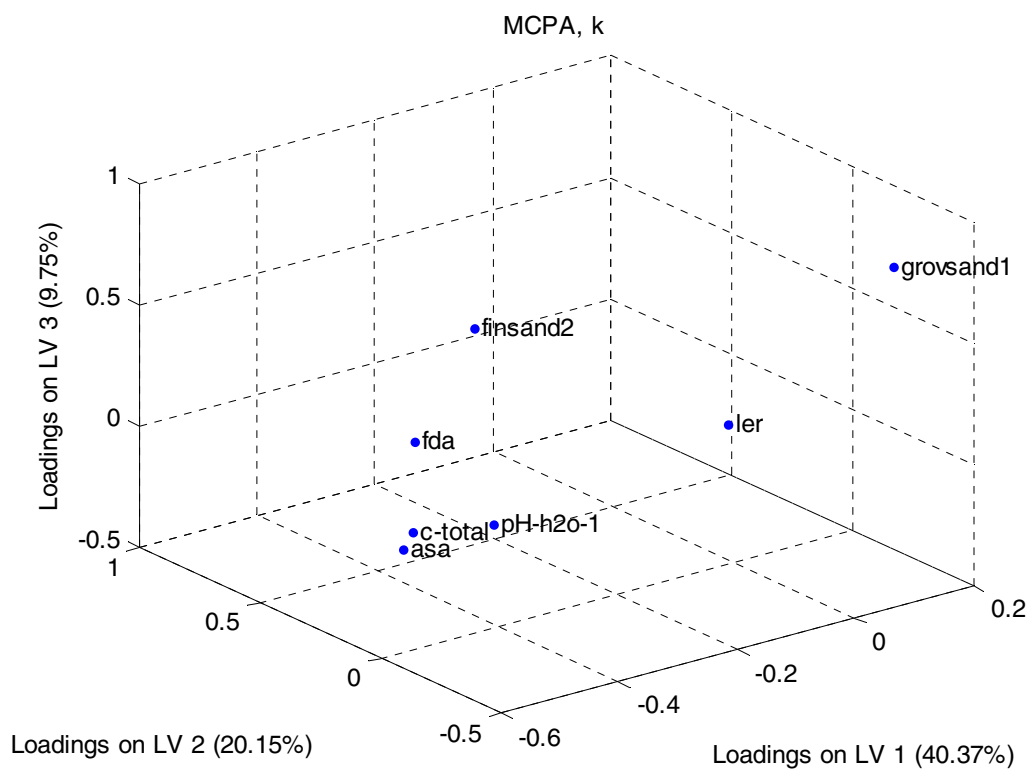
Tabel 10.6. Outliers i korrelationen til nedbrydningsraten (k) af MCPA for alle prøver

	k	pH-h2o-1	c-total	ler	finsand2	grov sand 1	asa	fda
03-1-2-134	- 0.1113	7.05	0.15	2.10	43.90	4.10	0.50	12.88
04-1-1-240	- 0.0266	5.82	3.86	5.20	8.60	37.70	25.73	40.22
05-1-2-266	- 0.1233	5.63	0.77	6.70	4.80	20.90	2.87	7.59
10-1-3-516	- 0.0706	5.44	0.05	1.00	16.00	72.00	0.00	0.25
11-1-2-540	- 0.1731	6.27	0.40	3.10	10.80	76.10	1.11	2.25
12-1-2-545	- 0.1727	6.67	0.77	3.50	1.00	71.70	0.68	6.50
08-1-2-403	- 0.1402	5.63	0.77	4.10	10.70	57.60	1.10	13.93
09-1-2-408	- 0.1530	6.12	0.29	5.00	5.40	37.40	0.00	1.08
16-1-1-686	- 0.2113	6.61	2.03	5.60	15.30	47.20	33.87	46.71
20-1-2-829	- 0.0208	6.01	0.95	6.90	11.90	10.40	1.44	33.62

Som det fremgår af ovenstående er korrelationen "alle prøver" temmelig dårlig, hvilket også udtrykkes i at der er mange outliers, figur 10.19. Denne korrelation tjener i virkeligheden kun til at afgøre om vi har en situation med høj nedbrydning eller en situation med lav nedbrydning. I tilfældet lav nedbrydning kan der tildeles en lav værdi, f.eks. ingen nedbrydning, mens der i tilfældet høj nedbrydning korreleres efter korrelationen for overjorde, der giver et meget mere præcist estimat for nedbrydningen.



Figur 10.19. Score plot for korrelationen "alle prøver" (sorte pletter) med outliers projiceret ind (grønne stjerner)



Figur 10.20. Loading plot for korrelationen "alle prøver".

Figur 10.19 og 10.20 viser korrelationen “alle prøver” med de 10 outliers fra tabel 10.7 projiseret ind, sammen med variablenes betydning for de tre latente variable. Figur 10.19 viser at outliers generelt ligger i yderkanten af sværmen med korrelationens prøver, og altså repræsenterer ekstremer i datasættet. For enkelte prøver er det muligt entydigt at pege på en årsag til at de er outliers. For prøverne 516, 540 og 545 skyldes forholdet grovsand, hvilket tildels også gælder for prøve 403 (de fire prøver i yderste højre del af plottet). For prøve 134 skyldes forholdet højt indhold af finsand (øverste prøve i venstre del af plottet). Prøve 686 og 240 har begge højt ASA og SIR værdi.

Hvis k-værdien prædikteres for outliers fås følgende, tabel 10.7:

Tabel 10.7. Prædikterede k-værdier for outliers. K-pred angiver den korrelerede (forudsagte) k værdi, mens k angiver den målte k værdi. Tabellen angiver prediktionsfejlen på k ved anvendes af korrelationsmodellen på outlierne.

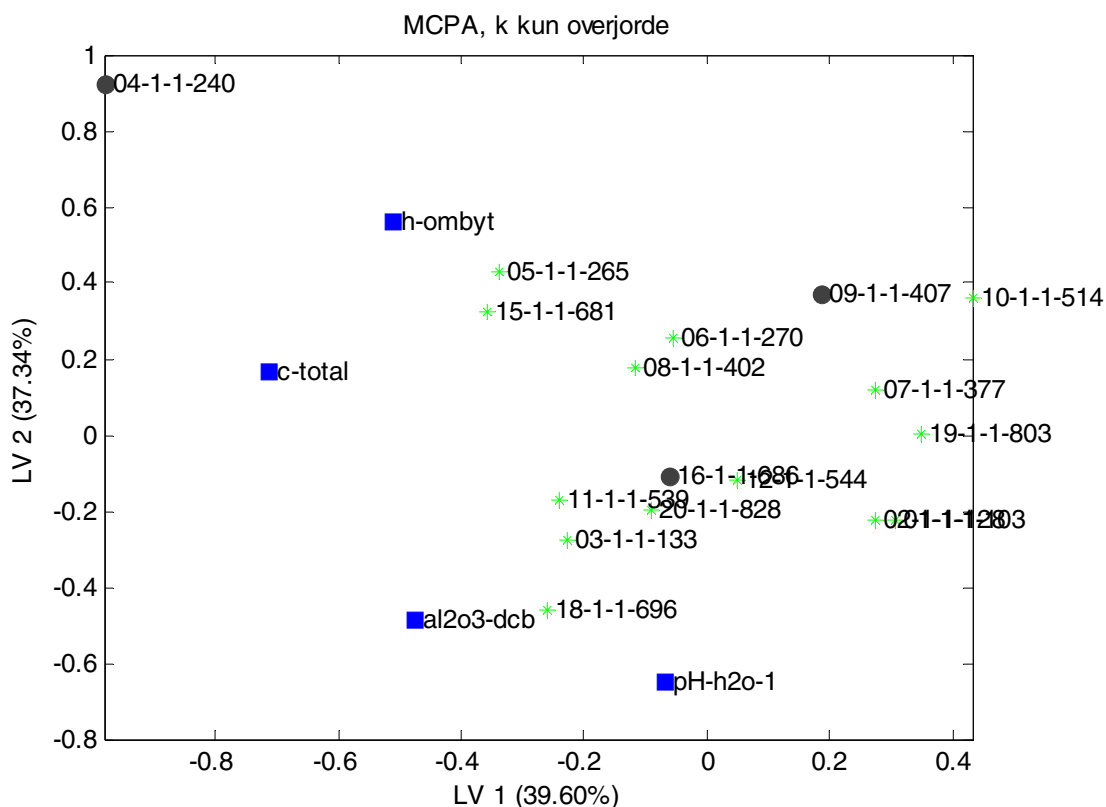
	k_pred	k
03-1-2-134	0	-0.1113
04-1-1-240	-0.1868	-0.0266
05-1-2-266	-0.0739	-0.1233
10-1-3-516	-0.0024	-0.0706
11-1-2-540	-0.0349	-0.1731
12-1-2-545	-0.0794	-0.1727
08-1-2-403	-0.0444	-0.1402
09-1-2-408	-0.0566	-0.1530
16-1-1-686	-0.1442	-0.2113
20-1-2-829	-0.0722	-0.0208

Med en afskæringsværdi på -0.04 mellem den gruppe med lav nedbrydning og den gruppe med høj nedbrydning, vil prøve 240 og prøve 829 blive vurderet mere “sikker” end de er målt til, mens de øvrige enten forbliver i gruppen eller vil blive vurderet mere “usikker”.

Korrelationen der beskriver overjordene alene fremgår af tabel 10.8.

Tabel 10.8. Outliers for korrelation hvor kun overjorde er anvendt. Den forudsagte værdi af k er medtaget k_pred. Ikke forståeligt, for at illustrere prediktionsfejls størrelse på outlierne.

	kpred	k	pH-h2o-1	c-total	h-ombyt	al2o3-dcb
04-1-1-240	-0.3249	-0.0266	5.82	3.86	16.92	2670.00
09-1-1-407	-0.0681	-0.1609	5.53	1.31	6.78	1405.50
16-1-1-686	-0.1324	-0.2113	6.61	2.03	7.08	1618.00

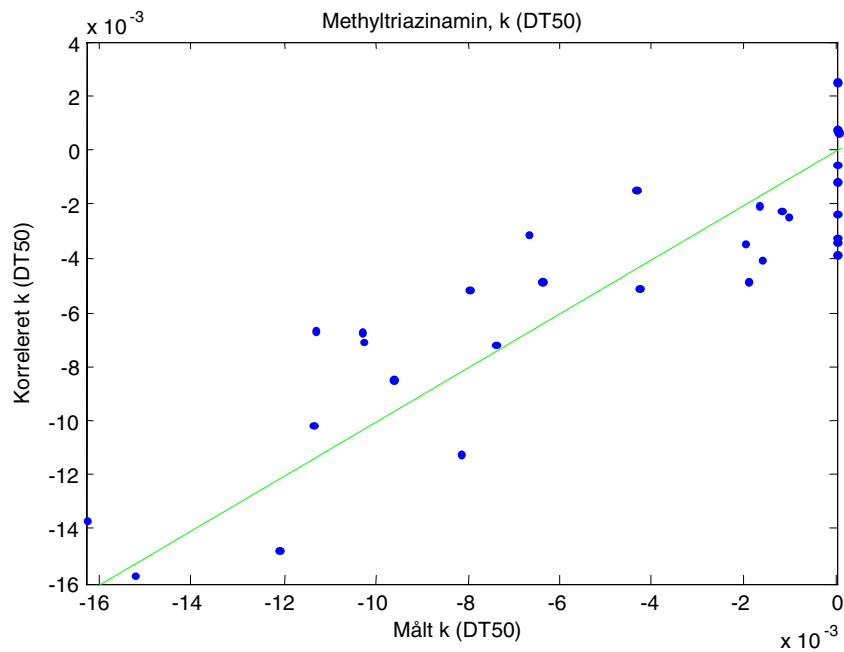


Figur 10.21. Biplot af korrelation for overjorde med outliers projiceret ind. Signaturer: blå firkanter er variablene, grønne stjerner er korrelationsobjekterne og sorte pletter er outliers.

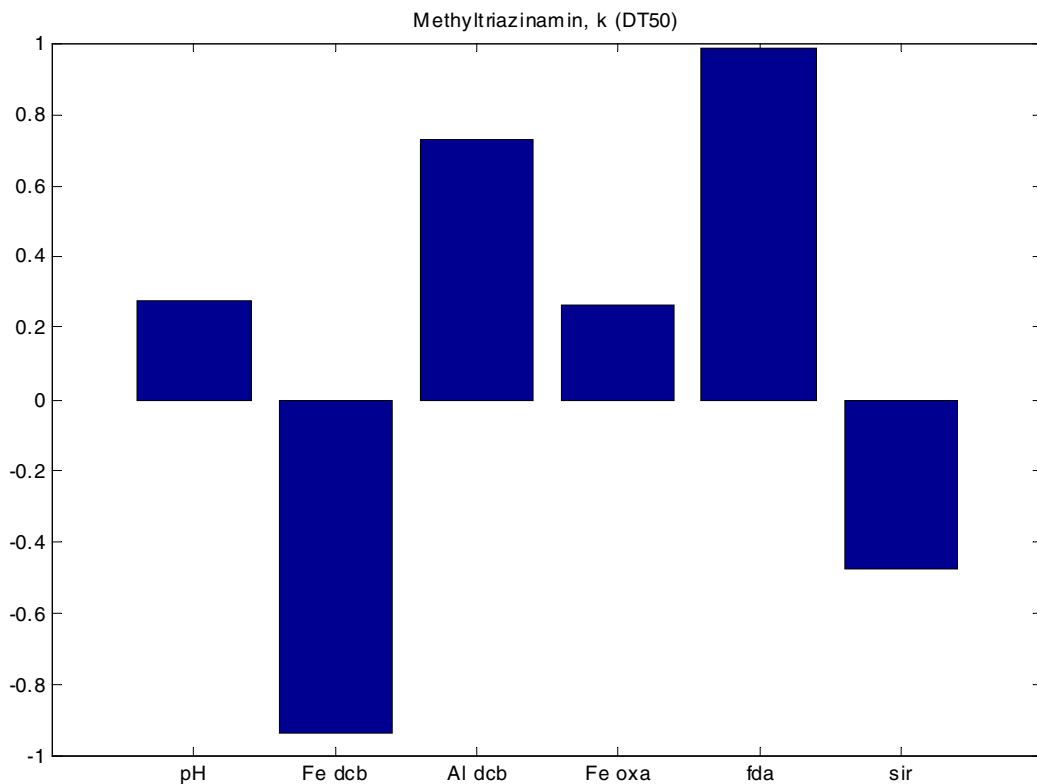
I figur 10.21 fremgår det at der ikke er noget i datastrukturen der indikerer at der skulle være noget unormalt ved outliers. Årsagen skal findes i at korrelationen er meget dårlig til at prædiktere værdien af k for disse prøver, hvilket kan skyldes fejl på bestemmelse af k værdi eller en af variablene, eller det kan skyldes at der findes afhængigheder for disse prøver der ikke er beskrevet i datasættet. Prøve 240 og 686 er gengangere fra korrelationen med alle prøverne, mens 407 er en ny outlier som jo vil blive prædikeret rimeligt med den tidligere korrelation.

MTA

Der er korreleret og identificeret outliers for korrelationen mellem jordegenskaber og nedbrydning for MTA, figur 10.22.



Figur 10.22. Målt nedbrydningsrate (k) mod korreleret nedbrydningsrate (k) for MTA. Varians forklaret: X: 94%; Y: 77%. 4 PLS komponenter. Outliers: 6/36. Outlier identifikation: se tabel 10.9.



Figur 10.23. Den relative betydning af jorddata (regressionsvektorer for PLS-korrelationen) for nedbrydningen af MTA.

Som det fremgår af figur 10.22 er der ikke tale om nogen særlig god korrelation. At der med 94% af variationen i X matricen kun kan forklares 77% af variationen i k, kan skyldes at enten k er meget usikkert bestemt, hvilket formentligt er tilfældet da alle er meget små, eller at der findes egenskaber af betydning for k der ikke er målt i projektet.

De to vigtigste variable for k er Fe_{dcb} og FDA, figur 10.23. I modsætning til de øvrige korreleringer betyder C-total intet. Dette hænger formentligt sammen med at k jo også er en funktion af sorptionen, og at sorptionen for MTA ikke er afhængig af C-total.

Tabel 10.9. Outliers i korrelationen til nedbrydningsraten (k) af MTA. Den prædikterede k værdi er desuden anført (k_pred).

	k_pred	K	pH-h2o-1	fe2o3-dcb	al2o3-dcb	fe2o3-oxa	fda	sir
12-1-1-544	0.0016	-0.0032	6.83	2898.74	1288.33	2150.00	43.21	2.79
12-1-2-545	0.0013	-0.0165	6.67	2913.61	3193.43	1570.00	6.50	0.35
07-1-2-378	-0.0159	-0.0093	5.83	4947.46	853.77	1840.02	0.21	0.00
08-1-2-403	-0.002	-0.0084	5.63	363.96	1460.21	84.00	13.93	0.71
09-1-2-408	-0.0059	0.0000	6.12	2321.51	1677.62	1454.95	1.08	0.21
18-1-1-696	-0.0097	-0.0161	6.42	5652.94	3435.58	3066.63	0.35	5.97

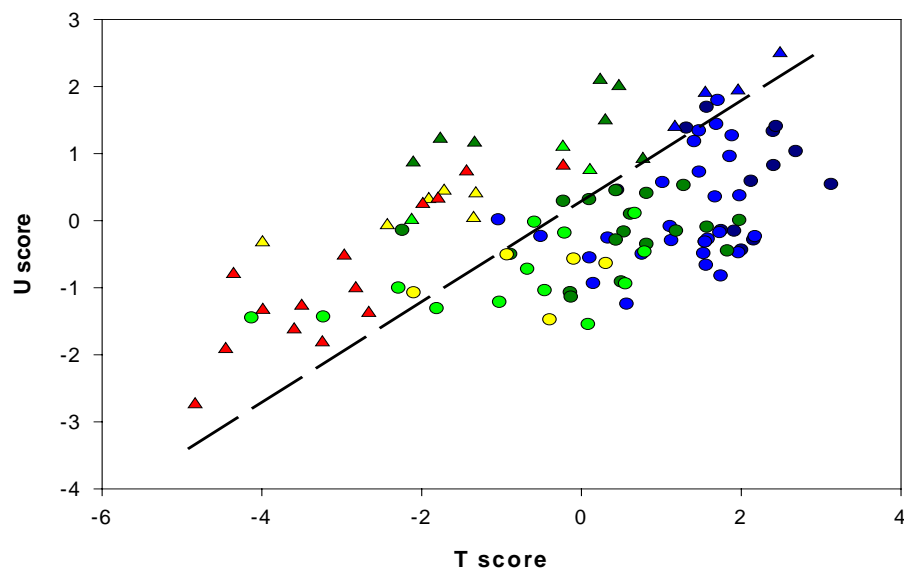
En gennemgang af score og loading plots hvor outliers er projiceret ind, giver ingen entydige forklaringer til outliers. Måske vil det heller ikke være vigtigt at efterfølge årsagen, idet alle målte k værdier små og følgelig behæftet med stor usikkerhed. At de alle er små værdier betyder samtidigt at der ikke er nævneværdig omsætning af pesticidet og man ville kunne tildele stoffet en konstant lille k værdi uden at lave nogen nævneværdig fejl.

Vægtning af jordegenskaber efter betydning

Hydraulik

Det neurale netværk kan i modsætning til PLS-regression lave korrelationer, hvor der er ikke-lineære sammenhænge mellem variable. Til gengæld giver neurale netværk ikke umiddelbart indsigt i de enkelte variables betydning for korrelationen. I forbindelse med vægtning af de jordegenskaber, som indgår i beskrivelsen af de hydrauliske forhold, er denne indsigt nødvendig. Ved at opdele datasættet med jordegenskaber for hydraulik og den simulerede koncentration for kK_f -sæt nr. 2 kan der imidlertid muligvis opnås en opdeling af datasættet i mindre dele, hvor det er muligt at anvende PLS-regression. Dette er forsøgt i det følgende.

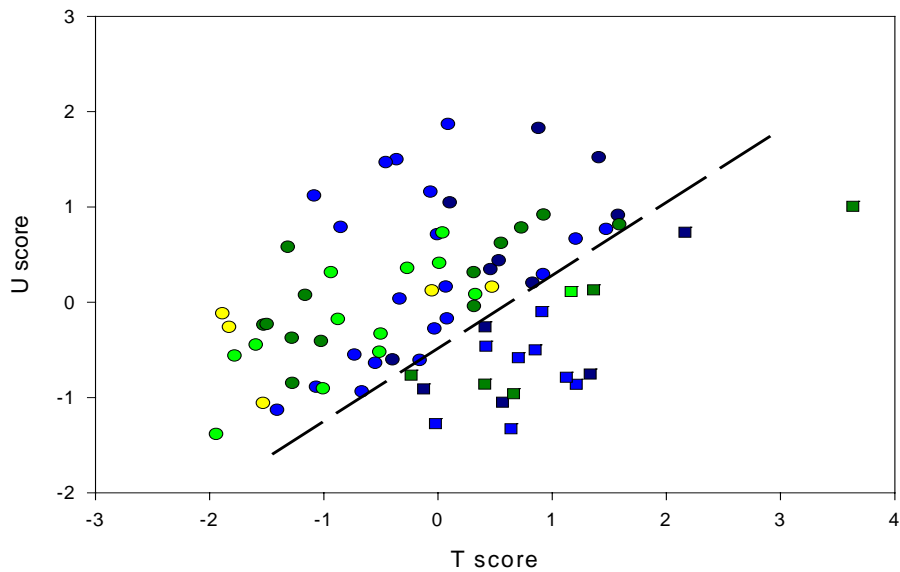
Den første opdeling sker på baggrund af kapitel 4.4, figur 20 (i den sammenfattende rapport), som viser at korrelationsanalysen skal ske for profiler med samme tykkelse af A-horisonten. Det betyder at følgende korrelationsanalyse anvender de 111 profiler, hvor tykkelsen af A-horisonten er 30cm. Opdelingen sker herefter i første omgang på baggrund af TU-plottet for første PLS-komponent, som farvet efter lerindholdet viser to grupper (figur 10.24); en "høj gruppe" (trekanter) og en "lav gruppe" (cirkler). Den stiplede linie på figuren angiver hovedadskillelsen, som giver en høj gruppe på 34 profiler og en lav gruppe med 77 profiler.



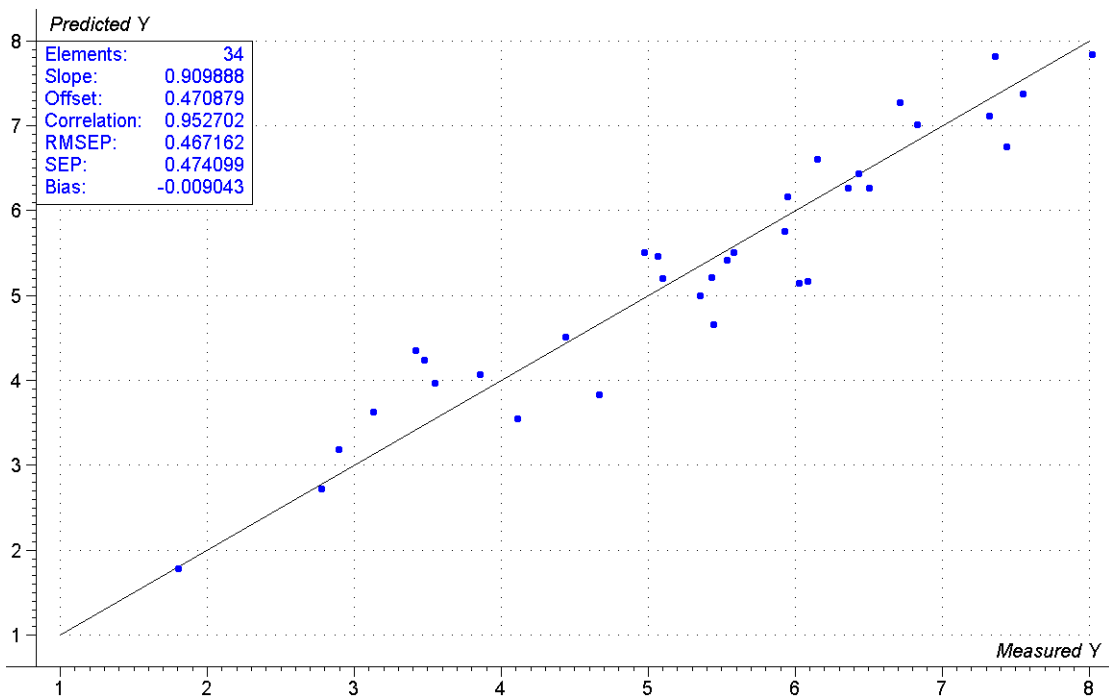
Figur 10.24. TU-scoreplot for første PLS-komponent fra korrelationsanalyse af jordegenskaber for hydraulik til simuleret koncentration af kK_f -sæt nr. 2 for 111 profiler fra profildatabasen med en tykkelse af A-horisonten på 30 cm. Profilerne kan, på basis af lerindholdet, opdeles i en "høj gruppe" (trekanter) og en "lav gruppe" (cirkler). Farverne angiver det summerede lerindhold i den øverste meter: mørkeblå <30; blå: 30-40; mørkegrøn: 40-50; grøn: 50-60; gul: 60-70 og rød: >70 Kg/m^2

Den "lave gruppe" (lavt lerindhold) opdeles yderligere på baggrund af TU-plottet for anden PLS-komponent fra korrelationsanalyse på de 77 profiler i gruppen. Opdelingen ses på figur 10.25, hvor den stiplede linie viser grænsen mellem "overdelen" (relativt højt lerindhold, små cirkler) og "underdelen" (relativt lavt lerindhold, store firkanter) af den "lave gruppe". "Overdelen" består af 58 profiler, mens "underdelen" består af 19 profiler, figur 10.25.

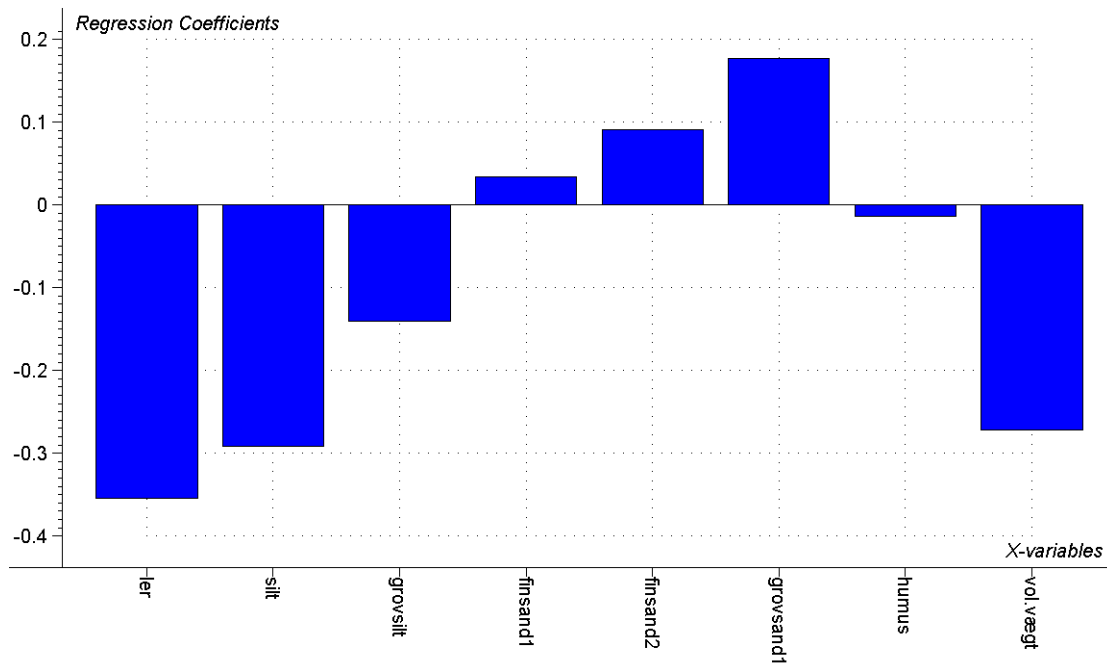
Efter denne opdeling efter lerindhold af de 111 profiler i tre datasæt på henholdsvis 34 ("høj gruppe"), 58 ("højeste fraktion af lav gruppe") og 19 profiler ("laveste fraktion af lav gruppe") gennemføres der korrelationer for hvert subsæt af de jordegenskaber der er vigtigst for fastlæggelsen af de hydrauliske parametre (ved neural netværksteknik) til den simulerede koncentration for kK_f -sæt nr. 2, figur 10.26 – 10.28.



Figur 10.25. TU-scoreplot for anden PLS-komponent fra korrelationsanalyse af de jordegenskaber der er vigtigst for fastlæggelsen af de hydrauliske parametre ved neural netværksteknik til simuleret koncentration af kK_r -sæt nr. 2 for 77 profiler i den "lave gruppe" (figur 10.24). Profilerne kan opdeles i to grupper, hvor cirkler angiver "overdelen" og firkanter "underdelen". Farverne angiver det summerede lerindhold i den øverste meter: mørkeblå < 30 ; blå: 30-40; mørkegrøn: 40-50; grøn: 50-60 og gul: 60-70 Kg/m^2

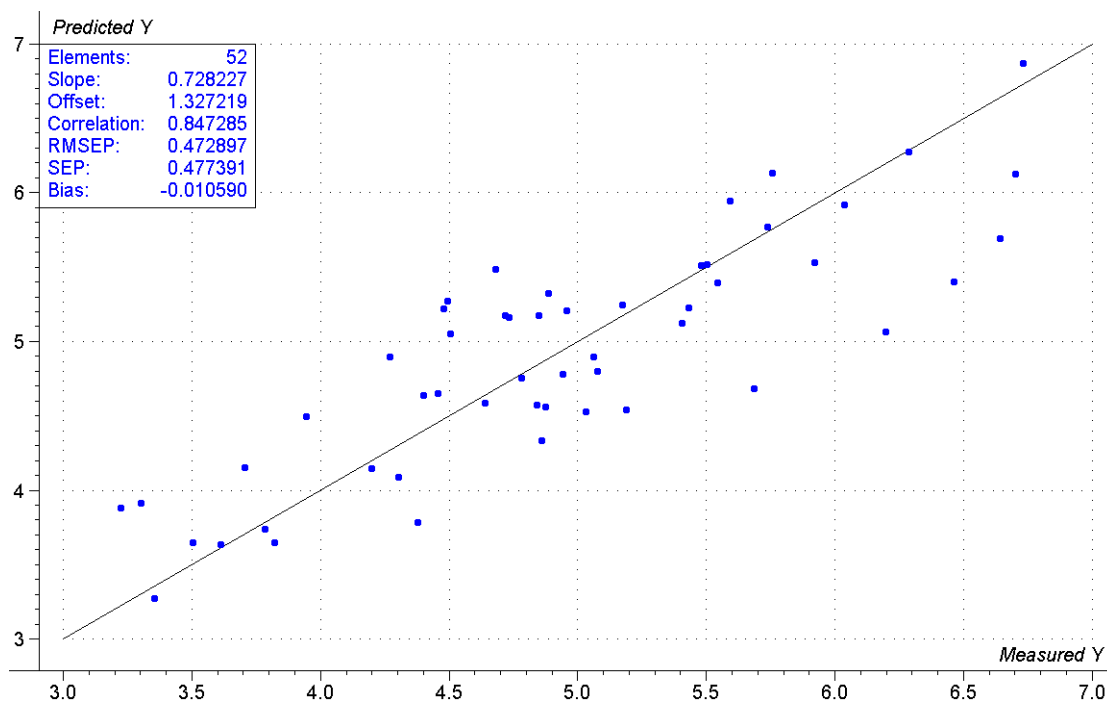


Figur 10.26a. Simuleret koncentration (measured Y) mod korreleret koncentration (predicted Y) kK_r -sæt nr. 2 for "den høje gruppe" (figur 10.24). Varians forklaret: X: 70%; Y: 93%. 2 PLS komponenter. Outliers: 0/34.

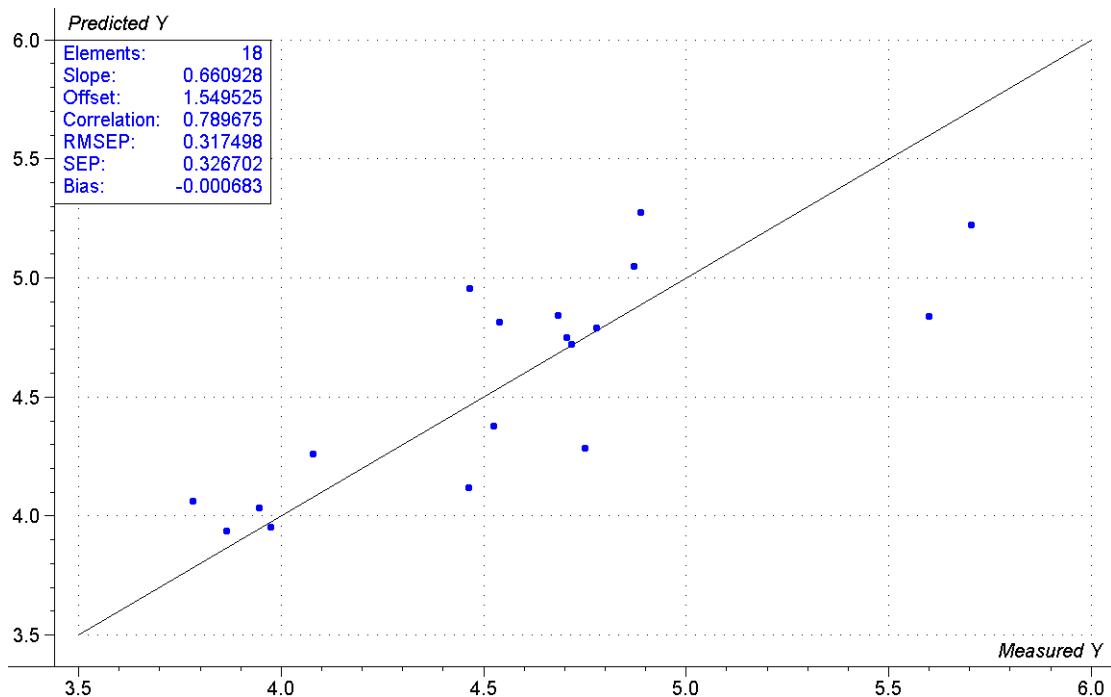


30cm høj lergru..., (Y-var, PC): (TS/TF_2,2) B0W = 8.665479

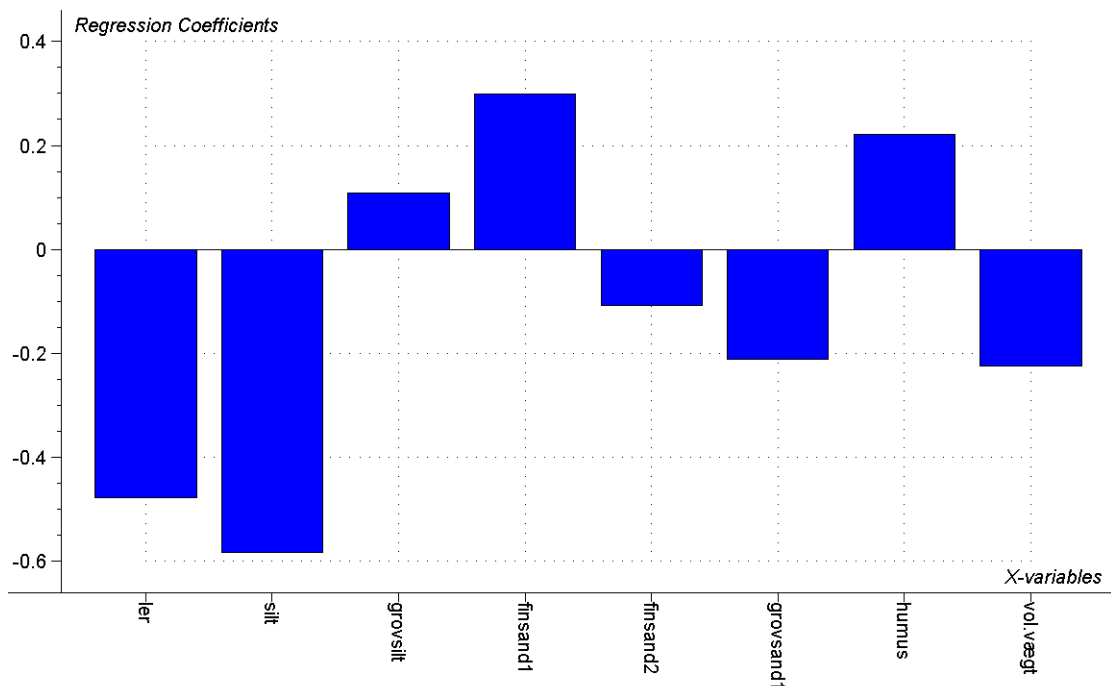
Figur 10.26b. Den relative betydning af jorddata (regressionsvektorer for PLS-korrelation) for simuleret koncentration af kK_r -sæt nr. 2 for den "høje gruppe" (figur 10.24).



Figur 10.27. Simuleret koncentration (measured Y) mod korreleret koncentration (predicted Y) kK_r -sæt nr. 2 for "overdelen af den lave gruppe". Varians forklaret: X: 73%; Y: 78%. 3 PLS komponenter. Outliers: 6/58..



Figur 10.28a. Simuleret koncentration (measured Y) mod korreleret koncentration (predicted Y) kK_f -sæt nr. 2 for "underdelen af den lave gruppe". Varians forklaret: X: 85%; Y: 82%. 3 PLS komponenter. Outliers: 1/19.



Figur 10.28b. Den relative betydning af jorddata (regressionsvektorer for PLS-korrelation) for simuleret koncentration af kK_f -sæt nr. 2 for "underdelen af den lave gruppe". Tolkningen af korrelationsanalysen er beskrevet i hovedteksten, kapitel 4.4.

ydraulik og binding

Korrelationsanalysen er gennemgået i kapitel 4.5, figur 23a og 23b, i hovedteksten.

Hydraulik, binding og nedbrydning

Korrelationsanalysen er gennemgået i kapitel 4.6, figur 26a og 26b, i hovedteksten.

Følsomhed af A-horisontens tykkelse. Der henvises til gennemgangen i hovedteksten under kapitel 4.7.

Korrelering til udvaskning under hensyn til volumenvægt. I denne korrelering er de enkelte variable divideret med volumenvægten for den pågældende profil, inden de er summeret i forhold til profilets dybde, f.eks. for C-total:

$C\text{-tot}(i \text{ alt}) = \sum C\text{-tot}(i \text{ j-profil}) \times \text{dybde } j\text{-profil} / \text{volumenvægt}(j\text{-profil})$ hvor j er jordhorisonterne A, B og C.

Undtaget fra denne summation er pH der summeres uden division med volumenvægten. Følgelig er volumenvægten udtaget som selvstændig variabel i korreleringen, da den i modsat fald ville optræde flere gange.

Referencer

Brereton, 1990 – 1992, Multivariate pattern recognition in chemometrics, Elsevier isbn 0-444-89784-4

Brooks, R. H. and A. T. Corey. 1964. Hydraulic properties of porous media. Hydraulic Paper #3, Colorado State University, Ft. Collins, Co.,

Børgesen, C.D. 2003. On models Predicting Soil Hydraulic Properties and their Use in Agro Hydrological Modelling. Ph.D-thesis. Institut for Jordbrugsvidenskab. KVL.

Esbensen, K.H. 2002. Multivariate data analysis – in practice. 5th edition. CAMO Process AS, Oslo, Norway.

Jarvis, N. 1991. MACRO - A Model of Water Movement and Solute Transport in Macroporous Soils. Swedish University of Agricultural Sciences. Department of Soil Sciences. Reports and Dissertations 9.

