

3. Fremgangsmåde ved fortolkning af data

For at finde de jordegenskaber som rummer de nødvendige oplysninger til udpegning af særligt pesticidfølsomme områder og som kan fremskaffes med den mindste ressource er nogle problemstillinger undersøgt ved flere metoder:

- jordegenskaber der bedst kan beskrive modelparametre
- afklare om relativt dyre modelparametre kan erstattes med lettere tilgængelige jordegenskaber.

Der er fire hovedelementer i fortolkningerne af de indsamlede data:

- Sammenlignende undersøgelser med tilsvarende data af anden oprindelse
- parvis fortolkning (bivariat dataanalyse) m.h.p. dataforståelse
- simulering af udvaskning på basis af værdier for binding, nedbrydning og hydrauliske forhold som er målte eller beregnede (multivariat korrelationsanalyse og neural netværksteknik)
- statistisk undersøgelse af sammenhænge mellem jordegenskaber og både modelparametre og simulerede udvaskninger (multivariat kalibreringsmodellering).

Sammenhænge mellem jordegenskaber er undersøgt gennem parvise sammenligninger (bilag 7) eller ved multivariat korrelationsanalyse mellem jordegenskaber, som pH og kornstørrelsefordeling, og mere komplekse jordegenskaber, som kationombytningskapacitet (CEC) og jordens specifikke overflade. De hydrauliske forhold er vurderet ved hjælp af neural netværksteknik. Valget mellem enten multivariat korrelationsanalyse eller neural netværksteknik er baseret på at neural netværksteknik bruges, hvor de ikke-lineære relationer mellem jordegenskaberne ikke kan lineariseres ved en transformation.

3.1 Sammenlignende undersøgelser

Disse undersøgelser og analyser skal dokumentere om de valgte undersøgelseslokaliteters egenskaber svarer til de umiddelbare omgivers, landskabselementernes eller sandjordens egenskaber i almindelighed. Undersøgelserne har elementerne:

- repræsentativitet
- variabilitet

Ved repræsentativitetsundersøgelserne sammenlignes de indsamlede data for jordegenskaber og marker, som projektet har produceret, med oplysninger i forskellige baser og kort. Sammenligningerne er foregået ved at plote projektets kornstørrelsesdata i trekantsdiagrammer sammen med eksisterende data for de enkelte landskabselementer. Der er især lagt vægt på oplysninger fra geologiske jordartskort, teksturdatabasen og profildatabasen.

Den skalamæssige variabilitet af de jordegenskaber og inddata for jordegenskaber, fra de undersøgte marker, som projektet har produceret, er dels undersøgt ved hjælp af variogrammer, dels gjort sammenlignelige ved box-plot for de enkelte marker og typer af parametre. Ud over de konkret undersøgte jordegenskaber er også resultaterne af de geoelektriske målinger (EM38) undersøgt og Georadar profilerne vurderet.

3.2 Rammer for simuleringer

Resultaterne af undersøgelserne er baseret på simulering af udvaskning og korrelation mellem parametre og jordegenskaber. Dette kræver i et vist omfang omregning mellem forskellige datatyper og sammenlægning af lag i de undersøgte jordprofiler.

Binding omregnet fra K_d til K_f (Freundlich isotermen)

Bindingen af pesticiderne er bestemt ved fordelingskoefficienten K_d . Til brug for simulering af udvaskning (her med MACRO modellen) beskrives bindingen med Freundlich isotermen, som indeholder koefficienten K_f , der fortæller om størrelsen af bindingen, og eksponenten n_f , som beskriver i hvor høj grad bindingen stiger konstant ved stigende koncentration af pesticid.

Bindingen (målt som K_d) er blevet omregnet til en K_f -værdi for stofferne MCPA, MTA og MTB. Omregningen er sket ved hjælp af Freundlich eksponenter bestemt for henholdsvis over- og underjord fra tre af projektets lokaliteter for hver af de tre stoffer. Freundlich isoternerne og forudsætningerne for omregningen er beskrevet i bilag 8.

Summering af data indenfor den øverste meter af jordprofilerne

For at kunne give en samlet beskrivelse af forholdene i de øverste jordlag er jordegenskaber blevet summeret med henblik på at korrelere til simuleret pesticidudvaskning til jordlagene under den øverste meter af profilet. Summeringen er foretaget ved at tillægge jordegenskaber en vægt som står i forhold til tykkelsen af det jordlag hvor data er målt. Et eksempel på summeringerne af organisk stof (C-total) og finsand for fuldprofilet på Ulsted marken er vist i tabel 5.

Tabel 5. Lagtykkelse, finsand1 og C-total for fuldprofilet på Ulsted marken.

	A-horisont	B-horisont	C-horisont	Sum
Interval	0-30 cm	30-80 cm	80-260 cm	-
Lagtykkelse (meter)	30 cm	50 cm	20 cm	100 cm
C-total (g/100g)>kg/m ²	1,19	0,13	0,08	43,8 kg/m ²
Finsand1 (%)>kg/m ²	34,9	35,8	63,9	411,5 kg/m ²
Volumen vægt	1,44	1,48	1,54	-

Summering af jordegenskaberne i den øverste meter skal sættes i forhold til jordens volumen vægt (v). Dette gøres ved at summere jordmængderne i de enkelte horisonter. Enheden bliver for C-total: $(g \text{ kulstof}/100 \text{ g jord}) \times (cm) \times (g \text{ jord}/cm^3) = g \text{ kulstof} / 100 \text{ cm}^2$ eller kg/m^2 .

Summen af C-total i A-horisonten er: $1,19 \times 30 \times 1,44$, hvortil lægges tilsvarende bidrag fra B- og C-horisonterne.

Alternativt kan summen udregnes som gennemsnit af indholdet gennem horisonten, hvor der laves en antagelse af at volumenvægten er relativt ens i den øverste meter. Denne antagelse er afprøvet ved korrelationen til den simulerede udvaskning af MTB, bilag 10.

De summerede jordegenskaber er brugt ved den statistiske sammenligning mellem jordegenskaber og simuleret udvaskning af pesticid, se yderligere i afsnit 3.4.

3.3. Simulering af pesticidudvaskning

For at kunne vurdere og sammenligne sårbarheden overfor pesticidudvaskning fra arealer med forskellige egenskaber er der gennemført simuleringer af udvaskning med den hydrologiske stoftransportmodel MACRO4.3, som er den model der anvendes i forbindelse med godkendelsesordningen for pesticider. I modellen simuleres udvaskning af pesticid på baggrund af vandbalancen og stofs specifikke parametre. Som vandbalance scenarier er der brugt to nedbørs tidsserier på 12 eller 24 år (Tylstrup og Estrup), og hydrauliske parametre som er bestemt for de undersøgte profiler. De stofs specifikke modelparametre beskriver bindingen og nedbrydningen af pesticid. Opsætningen af MACRO modellen og simuleringerne er gennemgået i bilag 9.

Udvaskningen er blevet simuleret for tre af de fire modelstoffer pesticider. Det sidste pesticid, glyphosat, er ikke blevet simuleret fordi laboratorieresultaterne viser, at det bindes meget stærkt til jordens faste bestanddele og at der ikke ville kunne simuleres nogen udvaskning med de forudsætninger, som er gjort. For de tre øvrige pesticider er den målte binding (K_d) blevet omregnet til en K_f -værdi (beskrevet i afsnit 3.2). For hver af de tre stoffer er det den Freundlich eksponent (n_f), som er benyttet til omregning af K_d til K_f for prøver fra overjord, der er brugt som inputparameter til modellen (bilag 8).

Simuleringen af udvaskning er gennemført for både de 24 fuldprofiler på sandjord og for 170 sandjordsprofiler fra profildatabasen. For at beskrive sammenhængen mellem modelparametre (hydraulik, binding og nedbrydning) og de mest betydende jordegenskaber er der gennemført tre simuleringer.

- Først er hydraulikkens indflydelse på simuleret udvaskning undersøgt ved at holde binding og nedbrydning konstant.
 - Dernæst er bindingen undersøgt i samspil med hydraulikken ved at holde nedbrydningen konstant.
 - Endelig er nedbrydningen undersøgt i samspil med både binding og hydraulik.
- Herudover er der lavet gentagne simuleringer hvor alle ovenstående parametre er varieret for at undersøge hvilke modelparametre der har størst indflydelse på simuleret udvaskning (Monte Carlo type simuleringer), bilag 9.

Udvaskningens følsomhed overfor variationer i tykkelsen af pløjelaget, nedbørsmængden (nettonedbøren) og dybden til grundvandsspejlet er også blevet undersøgt ved simulering med MACRO modellen.

Hydraulik

Til simulering af udvaskning med MACRO modellen er der behov for hydrauliske inputparametre. Derfor er der beregnet hydrauliske modelparametre til simulering af modelstoffernes udvaskning på baggrund af målinger af retentionskurver og korrelation med neural netværksteknik, bilag 5.

For (isoleret betragtet) at afklare de hydrauliske forholds betydning for udvaskning af pesticid er der simuleret med 9 faste kombinationer af binding (K_f) og nedbrydning (k) for 170 sandjordsprofiler fra profildatabasen. Værdierne af de 9 kombinationer af binding og nedbrydning for muldlaget (A-horisonten) og de dybere jordlag er fastlagt med udgangspunkt i resultaterne af målte og simulerede værdier for MTB. Værdierne og kombinationerne af værdier er valgt så de medfører at de simulerede udvaskninger dækker intervallet fra det ganske lave til det meget høje udvaskningsniveau, tabel 6.

Tabel 6. Input værdier for de 9 kombinationer af k og K_f

k K_f -kombination	k (øvre) [d-1]	k (nedre) [d-1]	K_f (øvre) [cm ³ g ⁻¹]	K_f (nedre) [cm ³ g ⁻¹]
1	$1 \cdot 10^{-3}$	0	5	1
2	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	5	0
3	$1 \cdot 10^{-3}$	0	5	0
4	$1 \cdot 10^{-2}$	0	1	1
5	$1 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	1	0
6	$1 \cdot 10^{-2}$	0	1	0
7	$1 \cdot 10^{-3}$	0	1	1
8	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	1	0
9	$1 \cdot 10^{-3}$	0	1	0

Binding

Udvaskningen af MTB er blevet simuleret med de målte hydrauliske egenskaber og bindingsværdier, men med fast nedbrydningsværdi for de 24 fuldprofiler. Formålet er at undersøge bindingens betydning for udvaskning. Nedbrydningen er i disse simuleringer fastsat til 80% fraktilen af de forsvindinger som er målt i de 24 fuldprofiler mhp. at tage udgangspunkt i relativt dårlige nedbrydningsforhold.

Med udgangspunkt i MCPA's egenskaber, hvor bindingen er blevet mindsket således at det vil medføre gradueret simuleret udvaskning, det vil sige at der kan opnås sikre værdier

for udvaskning både i det høje område og i det lave område, er udvaskningen simuleret med de målte hydrauliske egenskaber og fast nedbrydningsværdi for knapt 170 af profildatabasens sandjordsprofiler. Bindingsværdierne er målt for fuldprofilerne og tillagt profildatabaseprofilerne ved korrelation ud fra disses målte egenskaber. Formålet er at isolere effekten af binding på simuleret udvaskning.

Nedbrydning

Nedbrydningens (forsvindingens) betydning for simuleret udvaskning er undersøgt for de aktuelle hydrauliske parametre, og de målte værdier for binding og nedbrydning af MTB, MTA og MCPA på de 24 fuldprofiler.

Betydningen af A-horisontens tykkelse

A-horisonten (herunder især A_p -horisonten, pløjelaget, med det i reglen højeste indhold af humus) er den horisont i den umættede zone hvor der, for langt de fleste pesticider, sker den største binding (og dermed nedbrydning). Dette er vist gennem at hovedparten af de undersøgte pesticider viser størst bindingsevne overfor jorde med højt humusindhold. Derfor må pesticider normalt forventes at have længere opholdstid i denne horisont jo tykkere den er. Der er gennemført simuleringer af udvaskning for 10 udvalgte profiler for at finde en kritisk tykkelse af A-horisonten, og opnå viden om den nøjagtighed, hvormed denne tykkelse skal fastlægges.

De 10 profiler repræsenterer de fire landskabstyper: Yoldiaflade, Hedeslette, Bakkeø og Weichsel moræneflade. Samtidig dækker de 10 profiler hele variationsområdet i simuleret udvasket koncentration i 2 meters dybde. For de 10 profiler varieres tykkelsen af A-horisonten beregningsmæssigt fra 5 cm til 80 cm og kompenseres ved at ændre B-horisontens tykkelse modsvarende. Den beregningsmæssige variation i tykkelsen af A-horisonten strækker sig ind i ekstreme værdier som er urealistiske i praksis. Dette er gjort for at tydeliggøre effekten af variationer i tykkelsen. Disse simuleringer er udført med aktuelle hydrauliske modelparametre og målte værdier for binding af MTB, mens nedbrydningen er holdt konstant.

Betydningen af nedbørsmængden (nettonedbøren)

De fleste simuleringer er foretaget med udgangspunkt i klimaet i Tylstrup som har en gennemsnitlig årsnedbør for Danmark. For at undersøge betydningen af variationer i nedbørsmængden er der også simuleret med udgangspunkt i nedbørsforholdene med høj årlig nedbør i Estrup for de omtalte 9 kombinationer af k og K_f i 170 af profildatabasens sandjordsprofiler. Estrup og Tylstrup klima er beskrevet i bilag 9.

Betydningen af dybden til grundvandet

For at vurdere betydningen af grundvandsspejlets beliggenhed for udvaskningen er der, på basis af målte trykniveauer i sandområder, udført simuleringer med fastholdt grundvandsspejl. Grundvandsspejlet varieres fra 5 m, som er den fastholdte standarddybde, til 3 m, 2.5 m og 2.1 m (dvs. hhv. 3 m, 1 m, 0.5 m og 0.10 m under bunden af de 2 m dybe modelprofiler. Derfor ligger grundvandsspejlet altid under det 2 m profil/modeldybde hvorfra modellens output tages. For de undersøgte sandjorde har denne variation kun medført beskedne ændringer i simuleret udvaskning, hvorfor der ikke i de videre undersøgelser er taget højde for variationer i grundvandsspejlet, bilag 9.

3.4. Korrelationsanalyser

To vigtige problemstillinger som søges afklaret er om der kan etableres en sammenhæng mellem modelparametre med betydning for udvaskning og simple jordegenskaber, som allerede er tilgængelige eller relativt let kan fremskaffes (kortlægges), og i hvilken grad resultaterne gælder for alle pesticider, incl. de der vil blive anvendt i fremtiden. Sammenhængen mellem jordegenskaber og stofspecifikke modelparametre er undersøgt med multivariat statistik (korrelation), bilag 10, mens sammenhængen mellem jordegenskaber og hydrauliske modelparametre er beskrevet ved brug af neural netværksteknik, bilag 5.

I korrelationsanalysen til binding og nedbrydning er der brugt målte værdier for jordegenskaber mens der til multivariat korrelationsanalyse til udvaskningsresultaterne er brugt summerede jordegenskaber for den øverste meter af hvert profil (afsnit 3.2). Ved tolkning af variabelnes betydning i korrelationsanalysen, må man være opmærksom på at nogle data udgør såkaldte lukkede datasæt. Lukkede datasæt er f.eks. tekstur data, hvor variablene (de enkelte kornstørrelser) er normerede til 100%. Ved tolkning af sådanne data er der altid en indbygget tendens til indbyrdes negativ korrelation for variable som indgår i relativt mindre (men betydningsfulde) koncentrationer. Dette betyder at hvis korrelationsanalysen peger på at en variabel har stor positiv betydning, kan dette evt. alternativt afspejle at en anden eller andre af variablene indenfor det lukkede datasæt, har en stor negativ betydning. Det vil derfor være vanskeligt i praksis at afgøre i et lukket datasæt om fx. en udvaskning stiger med faldende andel af fine bestanddele (ler og silt), eller om den stiger som funktion af stigende andel af grove bestanddele (sand og grus). Dette forhold har kun betydning for forståelsen af sammenhængene og er uden kortlægningsmæssig betydning.

Korrelationerne er foretaget til simuleret udvaskning i 2 meters dybde frem for intervallet ned til 1 meter, hvorfra data stammer. Dette skyldes at et enkelt simuleret udvaskningsresultat fra MACRO modellen i 1 meters dybde er tvivlsomt, om end der generelt er god overensstemmelse mellem de beregnede udvaskninger i 1 og 2 meters dybde (Bilag 9).

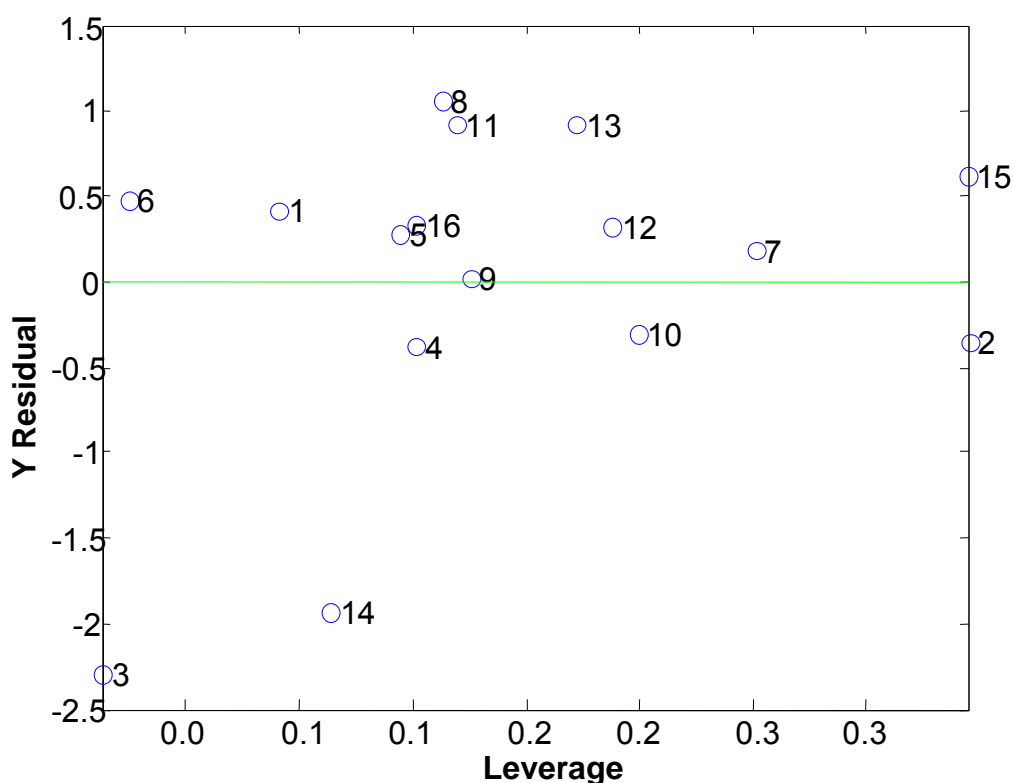
Alle de målte jordegenskaber indgår i korrelationen mellem jordegenskaber og modelparametre for at afklare hvilke jordegenskaber, der er de mest betydende. Et væsentligt udkomme af korrelationsanalysen er at afgøre hvilke egenskaber der er vigtige og hvilke der er uden eller af ringe betydning, og altså ikke behøver at blive målt. Derfor er korrelationen mellem jordegenskaber og simuleret udvaskning foretaget med de jordegenskaber som har vist sig at være mest betydende ved korrelationen til modelparametrene, netop for at demonstrere, netop for at demonstrere, hvordan få datatyper kan anvendes til prædiktion af en relativt kompliceret størrelse: den simulerede udvaskning.

Validering og outliers

Ved den regressions metode som er anvendt her (Partial Least Squares Regression, forkortet: PLS-R eller blot PLS) optimeres udvælgelsen af "latente variable", sådan at de bedst muligt beskriver både den variabel der skal prædikteres, og det datasæt der skal bruges til prædiktionen (i dette tilfælde de iboende jordegenskaber). Det betyder at korrelationerne ofte kommer til at bestå af færre latente variable end ved andre metoder og dermed bliver enklere og lettere at fortolke. Det betyder også at der ikke stilles krav til de variable der indgår, fx. Om de er indbyrdes uafhængige eller normalfordelte, som det kan være tilfældet ved andre multivariate metoder, så som multibel lineær korrelation. De latente variable kaldes ved denne metode "PLS-komponenter" (analogt med "principal komponenter" ved principal component regression og -analyse).

Regressionerne er blevet krydsvalideret enten efter "leave one out" eller "segment" principperne. Validerings principperne går alle ud på, iterativt og ved forskellige udvælgelses metoder at fjerne objekter fra korrelationen. Herefter laves en ny korrelation og virkningen af udeladelsen undersøges. Krydsvalidering anvendes når det ikke er muligt eller er for dyrt at tilrettelægge undersøgelserne, således at der kan etableres både et læredatasæt og et valideringsdatasæt.

Outliers kan identificeres som prøver, der har stor indflydelse på korrelationen, men som dårligt lader sig prædiktere ved hjælp af korrelationen. Andre former for outliers kan være prøver, hvis egenskaber ligger udenfor det interval hvor flertallet af prøver befinder sig, eller prøver der har unormale egenskaber. Outliers kan være opstået som følge af bl.a. målefejl, og det er vigtigt altid at undersøge, hvorvidt outlierne tilhører en gruppe man netop vil finde (fx. prøver med ekstrem lav bindingsevne). Outliers fjernes fra korrelationen for at forbedre prædiktionssevnen. Ved prædiktion af "ukendte" prøver gennemføres en test for om prøverne udviser "normal" adfærd eller tilhører gruppen af outliers, således at prædiktionen ikke kan anvendes på disse. 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), bilag 10. Figur 5 viser et plot af "leverage mod Y residual" hvor prøve 14 er en outlier. Prøve 14 udviser en relativ stor fejl på prædiktionen af prøvens værdi samtidig 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 korrelationens prædiktioner.



Figur 5. Prøvernes indflydelse (Leverage) på korrelationen mod fejlen ved korrelationens prædiktion af prøvens værdi (Y Residual). Plottet bruges ved udpegning af outliers i forbindelse med en PLS korrelationsanalyse.

Outliers opstår sædvanligvis ved at nogle prøver udviser en unormal "adfærd" for eller imellem de variable der har betydning for korrelationen. Hvis det fx. var sædvanligt at prøver med højt c-total indhold også havde 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 data ligger i yderkanten af den variationsbredde der er defineret for korrelationen, dvs. hvis det er en eller få prøver der giver en ekstra udstrækning af variationsbredden for korrelationen. Endelig kan outliers opstå ved analysefejl eller prøveudtagningsfejl. I sidstnævnte tilfælde skal outliers fjernes helt fra datasættet. Det er meget væsentligt at outliers i én korrelation, ikke nødvendigvis er outliers i en anden. De enkelte korrelationer vil udvise forskellig afhængighed af de samme variable, og en "skæv" variationsbredde i én model vil måske slet ikke slå igennem i en anden.

Selv om man naturligvis, på grundlag af rådata (fx. bilag 4), kan få ideer til hvad der er årsagen til at de pågældende prøver er outliers, kan dette let medføre fejlslutninger, idet det er meget vanskeligt at overskue mere end to variable på en gang. I stedet kan man projicere outliers ind på korrelationsresultatet, se hvordan de plotter i forhold til de øvrige prøver i korrelationen, samt forsøge at tolke deres egenskaber i det reducerede variabel rum ved hjælp af en forståelse af de latente variables betydning, se figur 10.1, 10.8 og figur 10.9 i bilag 10.