



AALBORG UNIVERSITET
STUDENTERRAPPORT

Kandidatspeciale i Geografi v/ BUILD
Institut for Byggeri, By og Miljø,
Aalborg Universitet
Thomas Manns Vej 23
9220 Aalborg Øst
<https://www.build.aau.dk>

Titel:

En konceptuel model til geografisk
overvågning af jordbundssundhed

Projekt:

Specialeprojekt i Geografi

Projektperiode:

01/02/2024 - 21/10/2024

Deltagere:

Asbjørn Birkbak Bendtsen
Gustav Lund Tobiasen

Vejledere:

Hovedvejleder:

Per Møldrup

Bivejleder:

Lene Werner Kristensen

Sidetæl: 148

Bilag: 1

Afsluttet: 21/10/2024

Abstract in English:

This report addresses the need to monitor soil health to ensure food security in the face of a growing global population and climate change. The study utilises remote sensing data, including Sentinel-2 satellite imagery, CORINE land use data, climatic variables, and terrain information. PyCaret in Python is used to develop the machine learning models, enabling efficient model training and comparison of various algorithms. These predictive models are integrated into a broader conceptual framework designed for large-scale soil health monitoring. By combining multiple data sources, this framework ensures a comprehensive assessment of soil conditions across diverse regions, making it a valuable tool for evaluating soil health at a large geographical scale. A key component of this approach involves using pedotransfer functions to estimate important soil properties, such as hydraulic conductivity and plant-available water, which are not directly measured. These estimates are included in the multi-criteria analysis, ensuring a holistic evaluation that incorporates both direct and inferred data. The models are calibrated with soil samples to validate their accuracy, ensuring robust predictions across broad geographic regions. Various machine learning models are compared based on accuracy, using metrics such as RMSE, to minimize deviations in the subsequent multi-criteria analysis. This process integrates multiple factors, to make a comprehensive assessment of soil health and supporting large-scale monitoring efforts. Although machine learning proves effective for predicting soil conditions, the complexity of soil composition due to interacting factors remains a challenge. The study concludes that large parts of the Iberian Peninsula show poor soil health, while smaller areas, particularly in the east, exhibit acceptable to good conditions. This model and methodology thus offers a conceptual framework for future large-scale soil health monitoring.

Forord

Dette specialeprojekt er udarbejdet af to geografistuderende på 10. semester på Aalborg Universitet i perioden februar til oktober 2024. Projektet fokuserer på at udvikle et koncept, der med høj præcision kan estimere sundhedsindikatorer og give en samlet vurdering af jordens sundhedstilstand på stor skala. Projektet anvender PyCaret i Python til maskinlæringsmodeller for at skabe præcise regressionsmodeller til prædiktation af jordparametre, og resultaterne af disse modeller integreres med geografiske informationssystemer (GIS) til statistisk analyse og kortlægning. Yderligere anvendes en multikriterieanalyse, som kombinerer de forskellige indikatorer for at gennemføre en holistisk vurdering af jordens sundhed.

En stor tak rettes til vejleder Per Møldrup og bivejleder Lene Werner Kristensen for deres vejledning og støtte i løbet af projektet. Deres sparring har især bidraget til projektet i forbindelse med at finde relevant litteratur, dataindsamling samt udvælgelsen af relevante sundhedsindikatorerne som anvendes i analyserne til vurdering af jordens sundhed.

Læsevejledning

Højopløsningsbilleder for de mange producerede kort i forbindelse med GIS kan findes i vedlagte bilag A, hvor navnene på billederne svarer til figurnummeret i teksten som *Fig. X.Y (abc)*.



Asbjørn Birkbak Bendtsen



Gustav Lund Tobiasen

Indholdsfortegnelse

1	Indledning	1
1.1	Problemformulering	2
1.2	Litteraturstudie	6
2	Udvælgelse af caseområde	11
2.1	Samplingsområdet	14
2.2	Caseområdet	14
3	Geografisk undersøgelse af samplingområdet	17
3.1	Kortlægning af LUCAS jordprøver	17
3.2	Köppens klimaklassifikation	20
3.3	CORINE land cover/land use	23
3.4	Terræn og højdemodel for samplingsområdet	26
4	Jordbundens sundhedsparmetre	31
4.1	Jordens grundlæggende fysio- og hydrologiske egenskaber	31
4.2	Aggregatdannelse og Dexter Ratio	32
4.3	Jordens plantetilgængelige vand	33
4.4	Geniltning af jorden	38
4.5	Hydraulisk ledningsevne	39
4.6	Toleranceintervaller for planter i forhold til jordens surhedsgrad	41
4.7	Elektrisk konduktivitet	44
5	Indsamling af inputparametre til regressionsmodellen	49
5.1	Spektrale signaturer i jorden	49
5.2	Sentinel-2 spektrale bånd	52
5.3	Klimatiske data	56
5.4	Topografiske og geomorfologi	58
6	Maskinlæring og opstilling af regressionsmodellen	61
6.1	Sammenligning og bestemmelse af regressionsmodeller	67
7	Statistisk analyse af regressionsmodellen	75
7.1	Interkorrelationsmatricer	75
8	Kortlægning af data og Multikriterieanalyse	105
8.1	Prædiktioner	105
8.2	Sundhedsmæssige jordbundsparmetre	115
8.3	Multikriterieanalyse af faktorer for jordbundssundhed	125
9	Diskussion	141

9.1 Kritik over projektets tilgang til vurdering af jordens sundhed	144
10 Konklusion	147
Litteraturliste	149

Jordens sundhedstilstand er en kritisk faktor for at sikre fødevarer sikkerhed, bevare økosystemer og bekæmpe jordnedbrydning. I 2024 er der over 8,1 milliarder mennesker på Jorden, og FN's fødevare- og landbrugsorganisation (FAO) har forudsagt, at verdens befolkning vil nå 9,7 milliarder i 2050 og 11 milliarder i 2100 [United Nations, 2024; FAO, 2018]. Med en stigende verdensbefolkning og øget efterspørgsel på fødevarer bliver det derfor fortsat mere presserende at optimere landbrugsproduktionen på en bæredygtig måde [Ritchie og Roser, 2024].

Dette sætter et pres på jordressourcerne, især i EU, hvor det anslås, at 60-70% af jorden er usund eller truet af nedbrydning på grund af ubæredygtig anvendelse, overudnyttelse, forurening samt effekten af klimaforandringer og ekstreme vejrforhold [EU-Kommissionen, 2021, s. 1]. EU-Kommissionen har derfor fastsat mål for at forbedre sundheden i europæiske jorde inden 2030 og at sikre, at alle jorde er sunde i 2050 [EU-Kommissionen, u.d.].

Jordsundhed referer til jordens evne til at understøtte plantevækst, bevare sin struktur og opretholde en balance mellem de fysiske, kemiske og biologiske processer, der foregår i jorden. Desuden udgør sund jord grundlaget for produktive økosystemer, idet den understøtter biodiversiteten af mikroorganismer, som lever i jorden og fremmer plantevækst, hvilket bidrager til opretholdelsen af hele fødekæden. Mikroorganismene spiller en central rolle i nedbrydning af organisk materiale og frigivelse af næringsstoffer, som planter har brug for [FAO et al., 2020].

Desuden hjælper en sund jord med at opretholde økosystemtjenester som vandoplagring og regulering af kulstofkredsløbet, hvilket er vigtigt for både naturens økosystemer og landbruget. Desuden er jordens sundhed tæt forbundet med fødevarer sikkerhed, da en sund jord opretholder den nødvendige næringsstofbalance og vandholdningsevne nødvendig for planter [FAO et al., 2020].

Jordnedbrydning udgør en trussel mod jordens sundhed og kan opstå som følge af flere faktorer, herunder erosion, tab af organisk stof, forurening og dårlig jordforvaltning. Klimaforandringer forværrer nedbrydningen gennem ekstreme vejrforhold som tørke og oversvømmelser, der kan være med til at nedbryde eller gøre jordbunden ubrugelig til landbrug [FAO et al., 2020; Ros et al., 2022b]. Overdreven brug af pesticider og gødning skader også jordens mikroorganismer, hvilket svækker dens frugtbarhed og reducerer dens evne til at understøtte plantevækst [FAO et al., 2020]. En sund jord er mere modstandsdygtig over for erosion og klimatiske påvirkninger, og den spiller en central rolle i vigtige økosystemfunktioner som kulstoflagring og opretholdelse af rent grundvand til drikkevand [European Environment Agency, 2022; Ros et al., 2022a]. Hvis

jordnedbrydningen ikke håndteres, kan det derfor føre til faldende produktivitet og udgøre en alvorlig risiko for økosystemer og den globale fødevareforsyning [FAO et al., 2020].

I takt med at klimaændringer og jordnedbrydning skaber stigende pres på jordens produktivitet, er der derfor et voksende behov for at vurdere og overvåge jordens sundhed på en systematisk og omfattende måde. Traditionelt har in-situ dataindsamling, hvor jordprøver tages in-situ, været den mest anvendte metode til at levere præcise målinger af jordens fysiske, kemiske og biologiske egenskaber [Maaz et al., 2023]. Denne metode har dog betydelige begrænsninger, da den er tidskrævende, geografisk begrænset og ofte dyr at anvende på større skala, især i utilgængelige områder, hvilket gør det svært at dække store områder effektivt [Das et al., 2024; Gatea et al., 2018].

1.1 Problemformulering

For at adressere udfordringerne ved traditionelle in-situ dataindsamlingsmetoder er det blevet mere relevant at supplere eller erstatte dem med fjernmålingsteknologier (remote sensing), der kan indsamle information over store geografiske områder hurtigt og effektivt [Gholizadeh et al., 2018]. Fjernmåling, især med satellitter som Sentinel-2, gør det muligt at overvåge flere af jordens fysiske og kemiske egenskaber. Disse satellitter leverer spektrale data, som kan bruges til at estimere jordparametre som tekstur og organisk kulstofindhold med høj præcision [Gholizadeh et al., 2018].

LUCAS-datasættet er centralt i denne proces, da det giver en omfattende mængde in-situ jordmålinger fra hele Europa. Dette er også blevet demonstreret i studiet af Ballabio et al. [2016], hvor LUCAS-data kombineres med satellitbilleder for at etablere korrelation mellem de målte jordprøver og fjernmålte data. Ved at anvende denne tilgang opnås en hybrid model, hvor LUCAS-datasættet bruges som grundlag for træning og kalibrering af modellerne, hvilket forbedrer præcisionen af estimeringerne [Ballabio et al., 2016].

Ved at kombinere in-situ fra LUCAS med fjernmåling eller *remote sensing*, kan forskere dermed skabe avancerede modeller, der præcist kortlægger jordens sundhedstilstand. Fjernmålte data bidrager til en mere holistisk vurdering af jordens egenskaber, især når flere datakilder integreres i en hybrid tilgang [Ballabio et al., 2016; Gholizadeh et al., 2018]. Denne metode gør det muligt at forbedre præcisionen af prædiktioner af relevante jord- og sundhedsparemetre, hvilket er afgørende for at vurdere jordens sundhed over store geografiske områder.

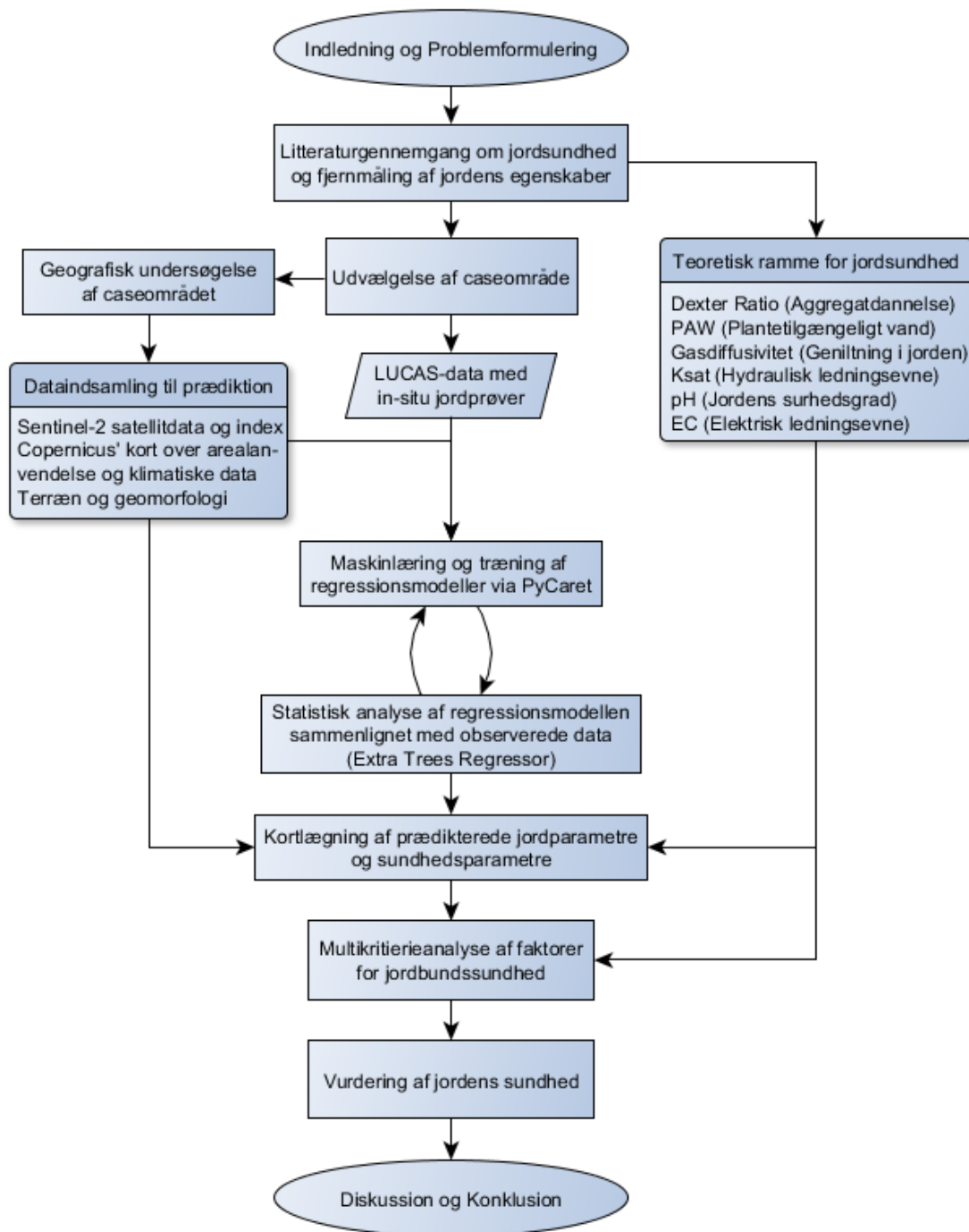
Maskinlæring spiller derfor en central rolle i udviklingen af den konceptuelle model i dette projekt, hvor målet er at kombinere fjernmålte data med in-situ målinger. Ifølge Gholizadeh et al. [2018] har brugen af maskinlæringsalgoritmer i regressionsanalyser vist sig som en effektiv metode til at analysere komplekse datasæt, der integrerer fjernmåling med jordprøver. Ved brug af algoritmer som beslutningstræer kan modeller udvikles til præcist at forudsige jordparametre som jordtekstur og organisk kulstof. Denne tilgang anvendes i projektet for at udvikle en model, der kan vurdere jordens sundhed over store områder, baseret på både fjernmålte data og in-situ prøver. Dermed lyder problemformuleringen som følgende:

"Hvor anvendelig og præcis er remote sensing til at estimere relevante jordparametre til vurdering af jordens sundhedstilstand?"

Denne problemformulering, som projektet søger at besvare, vil være central for analysen og tæt forbundet med de anvendte metoder og data, der benyttes i projektet. Problemformuleringen styrer særligt projektets metodiske tilgang i konceptmodellen til vurdering af jordsundhed, ved at sætte fokus på de dataindsamlingsmetoder, der anvendes, herunder kombinationen af in-situ målinger og fjernmålte data. Derudover vil den maskinbaserede modellering spille en vigtig rolle i at besvare problemformuleringen, da maskinlæring kan bruges til at analysere store mængder data og udlede mønstre, der kan forbedre prædiktionsnøjagtigheden for de relevante jordparametre. Diagrammet i figur 1.1 illustrerer faserne og den metodiske fremgangsmåde, der anvendes i projektet til vurdering af jordsundheden. Det er vigtigt at bemærke, at de illustrerede faser ikke nødvendigvis følger rapportens struktur kronologisk, men de repræsenterer de centrale trin, der blev gennemført i løbet af projektet. De vigtigste komponenter i rapporten omfatter følgende faser:

- **Introduktion og problemformulering:** Dette afsnit præsenterer baggrunden for projektet, hvor jordsundheden beskrives i forhold til dens vigtighed for bevarelsen af økosystemer og fødevarer sikkerheden. Desuden formuleres det centrale forskningsspørgsmål i form af problemformuleringen, der danner grundlag for rapportens analyse.
- **Udvælgelse og geografisk undersøgelse af caseområde:** Her beskrives kravene til afgrænsningen af caseområdet, med en gennemgang og kortlægning af relevante jordprøver, samt områdets klimaklassifikationer og arealanvendelser for at få indblik i dets kontekst.
- **Litteraturstudie og teoretisk ramme for jordsundhed:** I disse dele af projektet gennemgås relevant forskning omkring jordsundhed, og danner desuden grundlaget for valget af den metodiske tilgang, såsom anvendelse af fjernmåling i den konceptuelle model, og at modelleringen skal foregå med en maskinlæring algoritme. Derudover danner disse afsnit en teoretisk ramme, med henblik på at kortlægge relevante sundhedsindikatorer, der udgør de relevante faktorer for at lave en multikriterieanalyse og en samlet vurdering af jordens sundhed.
- **Dataindsamling og -behandling:** Dette afsnit beskriver de data, der er blevet indsamlet og anvendt til prædiktions som inputparametre i regressionsmodellen. Det omfatter satellitdata og de tilhørende indekser, samt oplysninger om arealanvendelse, klimatiske forhold, terrænkarakteristika og geomorfologiske data. Disse data er essentielle for at træne modellen og sikre en præcis estimering af jordbundsparametrene.
- **Modellering med maskinlæring:** For at forbedre prædiktionserne og sikre en mere præcis vurdering anvendes maskinlæring til at udvikle modeller. Heraf udvælges modellen, der bedst kan anvende de indsamlede inputparametre til at forudsige jordbundsparametre baseret på de indsamlede data (Extra Trees Regressor). Derefter præsenteres den statistiske analyse af resultaterne fra regressionsmodellen.
- **Resultater og kortlægning:** I denne fase præsenteres de prædikterede jordparametre og sundhedsparametre fra modelleringen i form af kortlægning af resultaterne, for at give en visualisering af modellens evne til at prædiktere de rumlige variationer af jordens egenskaber.

- **Multikriterieanalysen:** Her præsenteres den centrale metode til at vurdere de forskellige jordparametre, der bidrager til jordens sundhedstilstand. I multikriterieanalysen integreres flere parametre for at give en holistisk tilgang til vurdering af jordens sundhed. Her klassificeres hver sundhedsparameter ud fra kriterier om planters optimale plantevækst, og tildeles indbyrdes vægtninger. Det endelige kort vil udgøre modellens samlede sundhedsvurdering af hele caseområdet.
- **Diskussion og konklusion:** Afslutningsvis diskuteres de anvendte teorier, metoder, samt resultaterne i forhold til den eksisterende litteratur og de praktiske implikationer af korelationen. Her vurderes modellens præcision og relevans, og der gives en samlet konklusion på projektets bidrag til forståelsen af jordens sundhed.



Figur 1.1. Diagrammet illustrerer rapportens faser og hovedelementer, herunder gennemgangen af relevant litteratur, opbygning af den teoretiske ramme for jordens sundhed, dataindsamling samt analyse. Faserne er ikke nødvendigvis opstillet i en kronologisk rækkefølge i forhold til projektstrukturen, men de viser de centrale metoder og værktøjer, der anvendes i projektet. Dette inkluderer in-situ jordprøver, fjernmålte data, regressionsmodeller baseret på maskinlæring til prædiktion og en opsamling med vurdering af jordsundheden i en multikriterieanalyse. Samlet danner disse elementer grundlaget for et koncept, der på stor skala kan vurdere og overvåge jordens sundhed på en struktureret og omfattende måde.

1.2 Litteraturstudie

I studiet af [Hughes et al., 2023] fremhæves der et stigende ønske om at definere jordsundhed ud fra et holistisk perspektiv, der integrerer forskellige fysiske, kemiske og biologiske indikatorer for at få en mere omfattende forståelse af jordens tilstand. Dette perspektiv understøtter behovet for at overvåge jordens sundhed regelmæssigt og effektivt, især i forbindelse med bæredygtig landbrugspraksis og forvaltning af naturressourcer. Da jordsundheden ikke kan måles direkte, er en vurdering af jordens sundhed derfor afhængig af en kombination af målbare indikatorer for at estimere, hvor godt jorden fungerer. Studiet identificerer flere nøglefunktioner og metoder, der er essentielle for vurderingen af jordsundhed, for henholdsvis EU's LANDMARK 2020 og USA's USDA, og betegner henholdsvis jordens kvalitet og sundhed. De primære indikatorer for jordsundhed i EU og USA inkluderer: Primær produktion (af mad, foder, fibre, energi osv.), opretholdelse af økosystemer for planter og dyr, rent drikkevand, jordens fysiske stabilitet og struktur, klimaregulering (kulstofsekvistrering), samt tilgængeligheden af vand og næringsstoffer [Hughes et al., 2023, s. 2]. Det er dermed nødvendigt at dække en bred vifte af jordfunktioner for at afspejle den overordnede sundhedstilstand i jorden samt eventuelle ændringer heri.

I dette projekt er der i kapitel 4 en liste over og beskrivelse af anvendte sundhedsindikatorer til vurdering af jordens sundhedstilstand, for at afspejle den brede vifte af indikatorer i artiklen. Disse jordparametre samles i multikriterieanalysen i kapitel 8.3, for at lave en samlet vurdering af jordens sundhed. For at kortlægge de nødvendige sundhedsindikatorer og jordbundsparametre, er det derfor nødvendigt at lave en omfattende kortlægning med høj nøjagtighed.

I en forskningsartikel af Gholizadeh et al. [2018] demonstreres det, hvordan jordprøver kombineret med spektrale data fra Sentinel-2 satellitter effektivt kan anvendes til at prædiktere lerindhold og kulstofindhold i jorde med rimelig nøjagtighed i bare jorde, mens sand og silt er mere udfordrende at estimere. I studiet blev der indsamlet 50 jordprøver fra hver af fire forskellige landbrugsområder i Tjekkiet for at undersøge potentialet i at anvende spektrale signaturer til at forudsige jordens karakteristika.

I studiet af Gholizadeh et al. [2018] fremhæves det, at in-situ data er afgørende for at træne og kalibrere de prædiktive modeller baseret på de fjernmålte data, da det sikrer validiteten og nøjagtigheden af prædiktionerne. Studiet anvender en hybrid tilgang, der kombinerer direkte målinger med proksimale målinger foretaget med et spektroradiometer, luftbårne spektralfotos fra droneoverflyvninger samt satellitbilleder indsamlet af Sentinel-2. Kombinationen af disse metoder gør det muligt at integrere spektrale data indsamlet på forskellige skalaer, hvilket er velegnet til både meget lokale, såvel som regionale og nationale vurderinger af jordparametrene.

Til maskinlæring og regressionsmodelleringen anvendes 75% af de indsamlede jordprøver til at træne modellerne, hvilket understøtter udviklingen af en robust model til prædiktion. Modelpræstationen vurderes gennem 5-fold krydsvalidering, der forbedrer modellens pålidelighed ved at tilpasse den til forskellige delmængder af træningsdataene. Denne proces hjælper med at minimere potentielle afvigelser i datasættet, og hjælper desuden med at tilpasse og validere modellen. I tabel 1.1 sammenlignes opgørelser fra studiet af

Gholizadeh et al. [2018] over modellens præcision ved brug af satellitbilleder, for alle de fire jorde i Tjekkiet. Bemærk særligt RMSE-værdierne opdeles i *Root Mean Square of Cross-validation* ($RMSE_{cv}$) og *Root Mean Square of Prediction* ($RMSE_p$), hvor lavere RMSE indikerer bedre forudsigelser.

Tabel 1.1. Statistik over jordegenskaberne i de fire jorde fra studiet af Gholizadeh et al. [2018] sammenlignet med modellens evne til at prædiktere værdierne udelukkende med satellitbilleder.

Opgørelsen består af jordprøvernes gennemsnit, standardafvigelse og variationskoefficient for jordteksturen og kulstofindholdet for de fire jorde i Tjekkiet, samt tilhørende RMSE- og RPD-værdier for modellens evne til at forudsige jordens kulstof-, ler-, silt- og sandindhold [%w/w] [Gholizadeh et al., 2018].

Přestavlky	Kulstof	Ler	Silt	Sand
Gennemsnit	1,19	12,26	49,63	38,12
SD	0,25	3,22	7,17	8,37
CV (%)	21	26	14	22
RMSE cv	0,14	2,87	5,71	5,93
RMSE p	0,14	3,05	6,28	8,22
RPD	1,60	1,27	1,13	1,02
Šardice	Kulstof	Ler	Silt	Sand
Gennemsnit	1,44	22,6	38,49	38,91
SD	0,39	6,80	4,67	8,34
CV (%)	27	30	12	21
RMSE cv	0,23	3,37	4,27	5,63
RMSE p	0,24	3,47	4,31	5,96
RPD	1,70	1,05	1,14	0,89
Nová Ves	Kulstof	Ler	Silt	Sand
Gennemsnit	1,07	11,16	37,59	51,24
SD	0,17	3,57	9,95	12,32
CV (%)	16	32	26	24
RMSE cv	0,08	2,51	7,01	7,28
RMSE p	0,09	2,83	8,46	9,15
RPD	1,73	1,55	1,10	1,14
Jičín	Kulstof	Ler	Silt	Sand
Gennemsnit	1,03	21,64	66,12	12,24
SD	0,15	5,14	5,56	2,41
CV (%)	15	24	8	20
RMSE cv	0,08	1,96	3,68	5,33
RMSE p	0,08	2,57	3,90	7,54
RPD	1,92	1,89	1,43	0,91

RPD-værdierne (*Ratio of Performance to Deviation*) anvendes til at vurdere modellens præcision ved at beregne forskellen mellem de observerede og estimerede værdier ved sammenligning mellem standardafvigelse og RMSE. Ifølge Gholizadeh et al. [2018] anvendes følgende klassifikationssystem for RPD-værdier:

- $RPD < 1,0$: En meget dårlig model, og ikke anbefalsesværdig.
- $RPD < 1,4$: En ringe model, hvor de mindste og største værdier er detekterbare.
- $1,4 < RPD < 1,8$: Indikerer en rimelig model, der kan bruges til evaluering og korrelationer.
- RPD mellem 1,8 og 2,0: God model, der muliggør kvantificerbare prædiktioner.
- $RPD < 2,0$: Meget god model til kvantificering.
- $RPD < 2,5$: Perfekt model til kvantificerbare. forudsigelser

Ud fra jordenes RPD-værdier tyder det dermed på, at jordenes lerindhold og organiske kulstofindhold kunne forudsiges med en god/rimelig model, mens modellerne for silt var mere tilbøjelig til at være ringe, mens modellen prædikterede sand i ringe eller dårlig grad. Resultaterne indikerer, at modellerne for lerindhold og organisk kulstof kunne forudsiges med en rimelig grad af nøjagtighed, mens silt og sand blev prædikteret med lavere præcision. Denne analyse fungerer som en vigtig reference for brugen af fjernmåling i prædiktionsmodeller for jordvariable ved hjælp af maskinlæring.

I en anden artikel af Ballabio et al. [2016], har de udført et omfangsrigt studie til formål at kortlægge indholdet af ler, silt og sand i jorde i hele EU. Studiet anvender data fra LUCAS (*Land Use/Cover Area frame statistical Survey*) databasen, der indeholder en omfattende mængde jordprøver, bestående af 20.000 observationer taget i 20 cm dybde, indsamlet på tværs af EU. I studiet supplerer de jordprøverne med yderligere variable såsom klimatiske data, arealdække, topografiske data og spektrale data i form af satellitbilleder, der sammen inkorporeres i en hybridtilgang sammen med LUCAS-jordprøver for at udarbejde en model til at prædiktere jordegenskaberne over et større område, også hvor der ikke nødvendigvis er taget jordprøver. Arealdække skaffede de fra CORINE Land Cover databasen, hvor den rummelige udbredelse af arealdække og arealanvendelse er kortlagt i raster format og med 44 forskellige klassificeringer. Deres spektrale data blev opsamlet med spektrometere, der målte reflektansen fra jordens overflade i blå, røde og nærinfrarøde farvebånd, herunder for at beregne indekserne NDVI og EVI, blandt andet for at måle graden af klorofyl og plantedække, for at vurdere andelen af vegetation. De anvendte også satellitdata til at kortlægge topografien, med højdedata og terrænhældning i 100 m opløsning. Typisk måles topografi ved brug af laser-scanninger (LIDAR), hvor sensoren måler hvor lang tid det tager laseren at reflektere tilbage, og dermed bestemmes en afstand til et objekt. For at samle alle disse data, anvendes der i studiet en multivariat regressionsmetode MARS (*Multivariate Adaptive Regression Spline*). MARS er en machine learning algoritme, som automatisk kan fremskrive en model, der kan estimere og interpolere på bestemte variable. I studiet anvendte de MARS til at prædiktere jordteksturen ud fra alle inputvariable, herunder de indsamlede data om arealdække, samt de topografiske, klimatiske og spektrale data. Modellen blev anvendt til at kortlægge jordteksturen for hele EU, i raster format og med 500 m opløsning per pixel. Krydsvalidering mellem de in situ indsamlede LUCAS-jordbundsprøver og den kortlagte jordtekstur prædikteret ud fra modellen, viser at

modellen giver et rimeligt resultat i forhold til henholdsvis ler ($R^2 = 0,65$), silt ($R^2 = 0,62$) og sand ($R^2 = 0,60$) [Ballabio et al., 2016].

Remote sensing-metodologien bygger på fjernindsamling af data gennem satellit- og luftbilleder og udgør et vigtigt redskab til at supplere observationer og data fra målestationer, ved at bidrage med at udfylde videnshuller. Dataindsamling ved hjælp af remote sensing kan for eksempel udføres med spektroradiometre monteret på satellitter, som kan identificere forskellige fænomener ved jordoverfladen baseret på deres spektrale signaturer. Der findes også andre metoder, såsom overflyvninger med LIDAR-teknologi, hvor laserbaserede afstandsmålinger mellem flyet og jordoverfladen muliggør præcise højdemålinger og produktion af højdekort over store udstrakte områder [Tempfli et al., 2009]. Jordprøver, der anvendes som såkaldt *ground truthing*, er essentielle både for træning af machine learning-modeller og for validering af modellernes resultater.

De undersøgte artikler i litteraturstudiet demonstrerer vigtigheden i at overvåge jordens sundhed, og hvordan fjernmålte data effektivt kan bruges til at estimere flere jordparametre og præsentere kort over deres rumlige fordeling gennem en hybrid tilgang, der kombinerer jordprøver og remote sensing-data [Gholizadeh et al., 2018; Ballabio et al., 2016]. Denne tilgang vil derfor blive benyttet i dette projekt til at estimere forskellige jordparametre og vurdere jordens sundhedstilstand.

Udvælgelse af caseområde 2

Som udgangspunkt er en høj kvantitet af data afgørende for at sikre en præcis undersøgelse og analyse af caseområdets jordforhold. Kravet om en stor datamængde bliver især vigtigt, når jordparametre som ler-, silt-, og sandindhold skal estimeres ved brug af regressionsmodeller. En høj datatæthed tillader desuden bedre modellering og mere præcise forudsigelser, idet flere datapunkter sikrer, at modellerne kan dække forskellige klimatiske og topografiske forhold, hvilket desuden vil gøre resultaterne mere anvendelige og robuste på tværs af forskellige geografiske områder og terræntyper.

Ydermere afgrænses det potentielle område også af datadiversiteten, hvilket vil sige, at jo flere parametre dataene dækker, såsom lerindhold, organisk kulstof, osv., desto bredere en vifte af parametre kan der laves estimationer på. Det forventes samtidigt at flere datapunkter, kan være med til at højne præcisionen (eks. R^2 og $RMSE$) af regressionerne for de enkelte jordparametre, i forbindelse med, at den større mængde forventeligt kan hjælpe med at dække flere typer af bl.a. klima og terræn, som kan hjælpe med at gøre resultaterne af regressionsanalyserne mere anvendelige på en større geografisk skala.

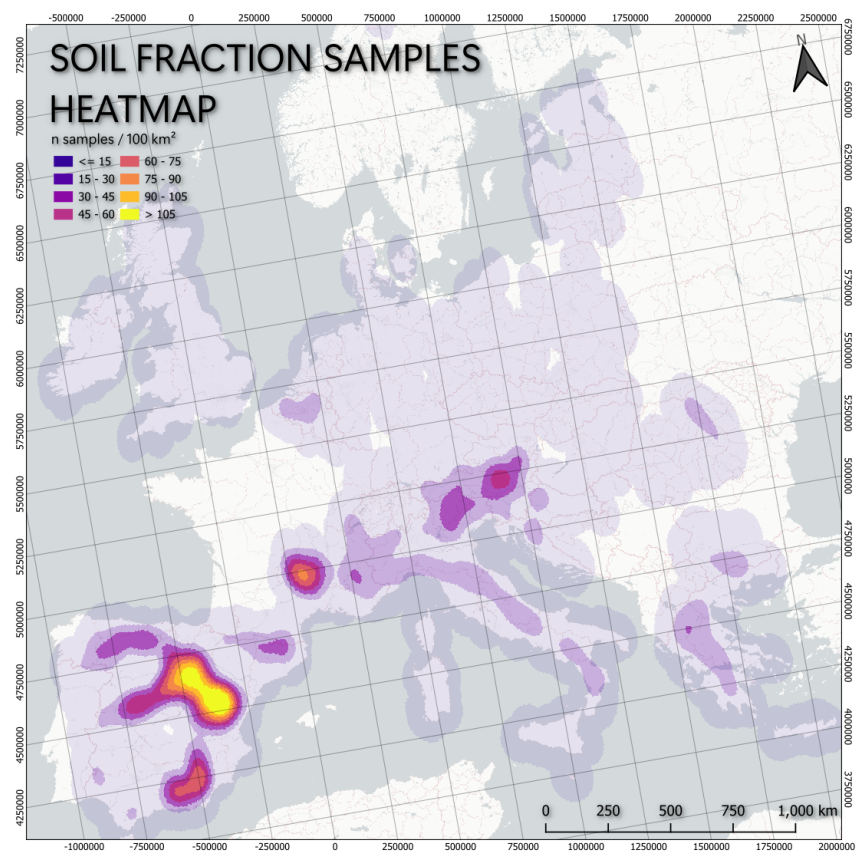
Datamæssigt og i forhold til de jordparametre der anvendes som de afhængige variabler i analyserne, kommer de i forbindelse med de mange jordprøver taget ved LUCAS-undersøgelserne, hvor LUCAS er akronymet for *Land Use/Cover Area frame statistical Survey*. For analyserne og undersøgelserne i dette projekt, kombineres data fra både 2015- og 2018-undersøgelserne, hvor studiet fra 2015 af [European Commission et al., 2020, s. 15] inkluderer essentielle parametre som ler-, silt- og sandindhold i jorden, mens studiet fra 2018 af [European Commission et al., 2022, s. 17-18] dækker opdaterede parametre af organisk kulstof, pH, elektrisk konduktivitet, volumenvægt med videre. I tabel 2.1 er samtlige jordparametre oplistet for begge LUCAS-datasæt.

Tabel 2.1. Oversigt over målte jordparametre i LUCAS-datasættene fra 2015 og 2018, der indeholder målinger fra de øverste jordlag (0-20 cm) i størstedelen af EU. Baseret på deres offentlige LUCAS topsoil survey rapporter udgivet af European Commission et al. [2020, 2022], og hvor partikelstørrelserne for ler, silt og sand er defineret jf. [European Commission et al., 2020, s. 10].

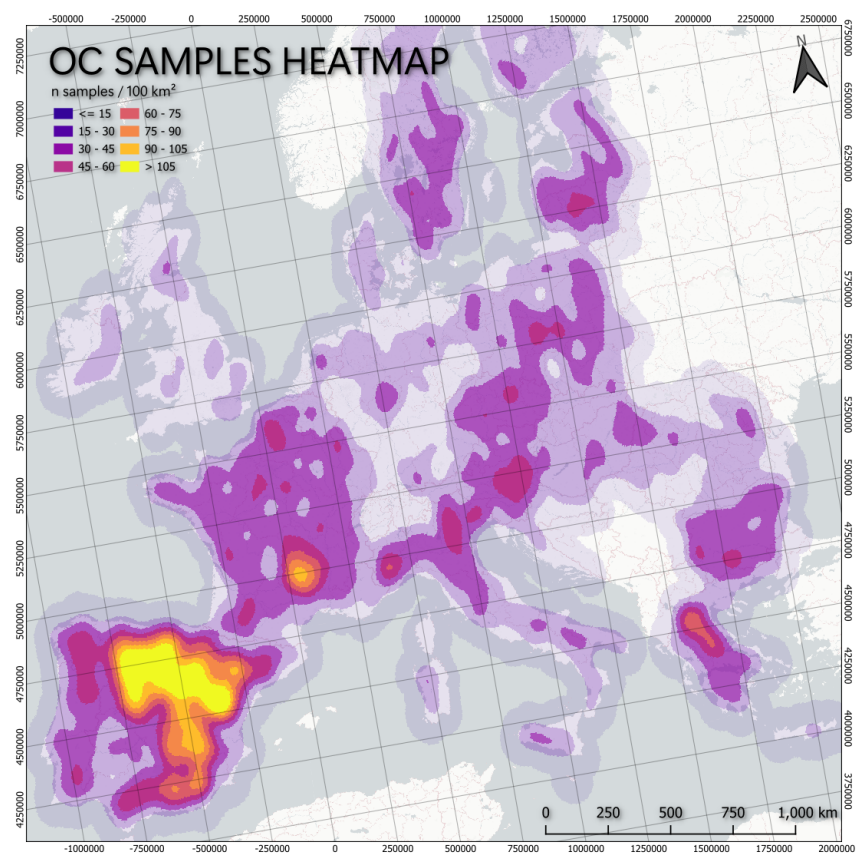
Jordparameter	Beskrivelse	Enheder
Lerindhold	Mineralske partikler $< 2\mu m$	[w/w%]
Siltindhold	Mineralske partikler $2 - 50\mu m$	[w/w%]
Sandindhold	Mineralske partikler $50 - 2000\mu m$	[w/w%]
Grove partikler (Coarse)	Mineralske partikler $> 2000\mu m$	[w/w%]
OC	Organisk kulstofindhold	[g/kg]
BD	Jordens volumenvægt	[g/cm ³]
CaCO ₃	Karbonatindhold	[g/kg]
P	Jordens fosforindhold	[mg/kg]
N	Jordens kvælstofindhold	[g/kg]
K	Ekstraherbart kaliumindhold (med ammoniumacetat)	[mg/kg]
EC	Elektrisk konduktivitet	[mS/m]
pH i H ₂ O og CaCl ₂	pH målt ved henholdsvis en suspension af jord i vand og i en CaCl ₂ opløsning	-
Ox Al og Fe	Ekstraherbart jern- og aluminiumindhold (med ammoniumoxalat ved pH 3)	[mg/kg]

I følgende figur 2.1 er der lavet en spatial analyse på LUCAS punktdata på tværs af EU. Her er der derved lavet en densitetanalyse på de mest essentielle jordparametre, der skal kendes for at kunne vurdere jordbundens sundhed; jordfraktionerne og det organisk kulstofindhold. Subfigurerne 2.1 *a* analyserer på LUCAS 2015 datasættet i forhold til punktdensiteten af jordprøver, for punkter hvor der er undersøgt lerindhold. Det samme er gjort for 2018-datasættet, som vist i subfigurerne 2.1 *b*, dog med fokus på organisk kulstof. Begge kort undersøger for antal jordprøver per $100km^2$. For alle punkter, hvor der er målt for eksempelvis lerindhold, er der også målt for dets korresponderende silt- og sandindhold, hvorfor der ikke behøves at lave separate analyse deres rumlige fordeling. For begge subfigurer ses det, at der på den Iberiske Halvø er foretaget et stort antal jordprøver jævnfør den højt målte punktdensitet relativt til andre steder indenfor EU. Dog er der også enkelte hotspots i Østrig, langs Appenninerne i Italien, Cevennerne i det sydlige Frankrig og Alperne.

(a)
LUCAS 2015
jordfraktioner
(ler, silt, sand)



(b)
LUCAS 2018
organisk kulstof

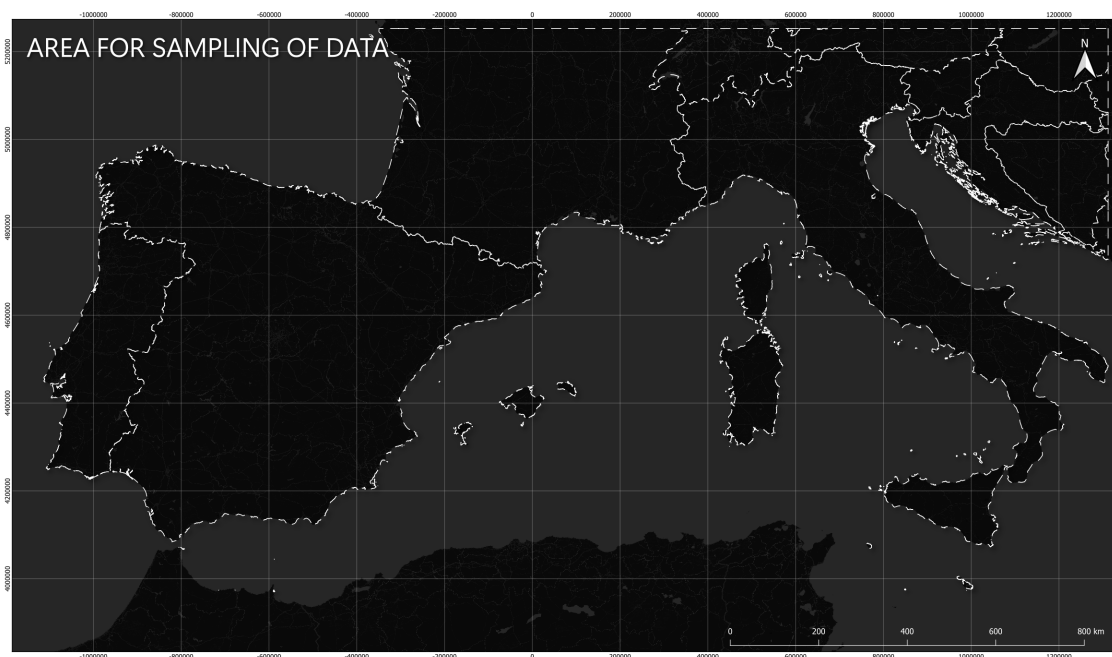


Figur 2.1. Punktdensitetanalyse for jordprøvepunkter der indeholder data for ler, silt og sand fraktioner fra LUCAS 2015 datasættet, og ligeledes for organisk kulstof fra 2018 datasættet.

Da forudsætningen for at kunne besvare problemformuleringen og ved brug af Remote Sensing, formodes kendskabet til blandt andet jordens lerindhold, at være af en afgørende rolle, for at kunne opnå bedst mulige resultater til forudsigelse af jordparametre og fremtidige pedotransfer funktioner, som kommer til udtryk i analyse af jordens tilgængelige vandindhold mv., som kan være relevant at vide, til at kunne vurdere netop jordens sundhed.

2.1 Samplingsområdet

I følgende figur 2.2 kan ses det udvalgte område, for hvori der indsamles data til brug til produktion af multivariate regressionsanalyser for jordparametre. Dette område kaldes for samplingsområde, og afgrænses af den definerede hvid-stiplede streg, som blandt andet indeholder lande som Italien, Slovenien, dele af Østrig med videre. Området udgør et større område end selve caseområdet, som er den Iberiske Halvø, som her i projektet vil inkludere landene Spanien, Portugal, Andorra, det helt sydlige område af Frankrig, samt halvøen Gibraltar. Det er indenfor dette område, hvor der laves de egentlige prædiktioner for jordparametrene, ud fra regressionerne, som bliver udarbejdet på baggrund af hele samplingsområdet som defineret i figur 2.2.

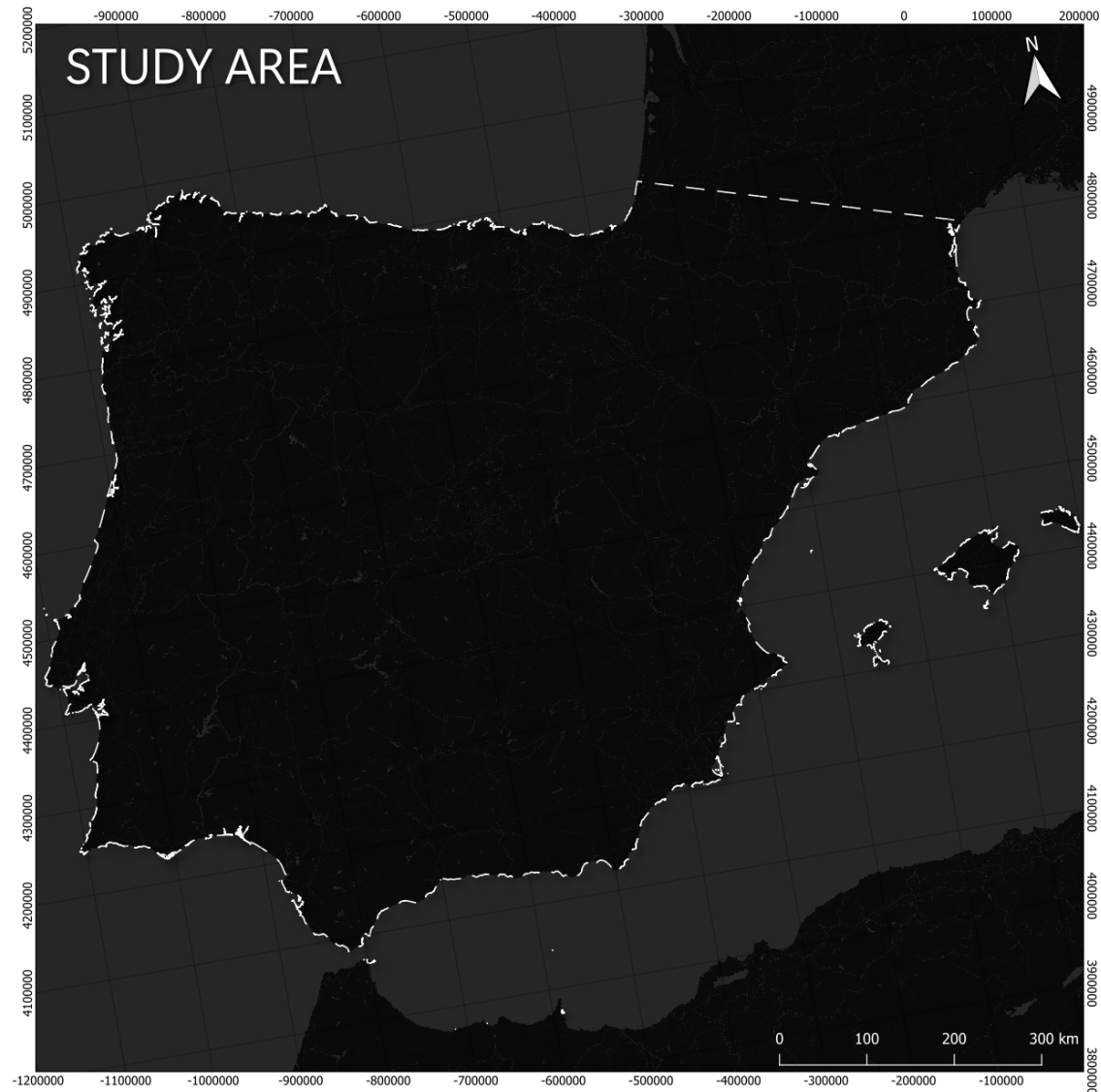


Figur 2.2. Afbildning af område hvor indenfor der anvendes jordprøver til brug i udvikling af modeller, som bruges til prædiktering af jordparametre. Der bruges kun LUCAS-punkter indenfor området som er defineret af de hvid-stiplede grænser. Landegrænser er inkluderet for kontekstfortælse.

2.2 Caseområdet

Det egentlige caseområde hvori der prædikteres for jordparametre er afbilledet i følgende figur 2.3. Området dækker ikke et lige så stort område som samplingsområdet, da det er for stort et område til at analysere på med henblik på computermæssige kapaciteter. Iberien som område er yderligere et fint område at arbejde med, eftersom det netop er her der er

foretaget mange jordprøver, hvilket kan bruges til statistik på analyserne. Caseområdet dækker et samlet areal på omtrentligt $641.585,23 \text{ km}^2$, og kortene som udarbejdes i forbindelse med prædiktion af jordparametre, vil have en opløsning på $250 \times 250 \text{ m}$. Det vil sige der prædikteres værdier for 10.270.328 pixels.



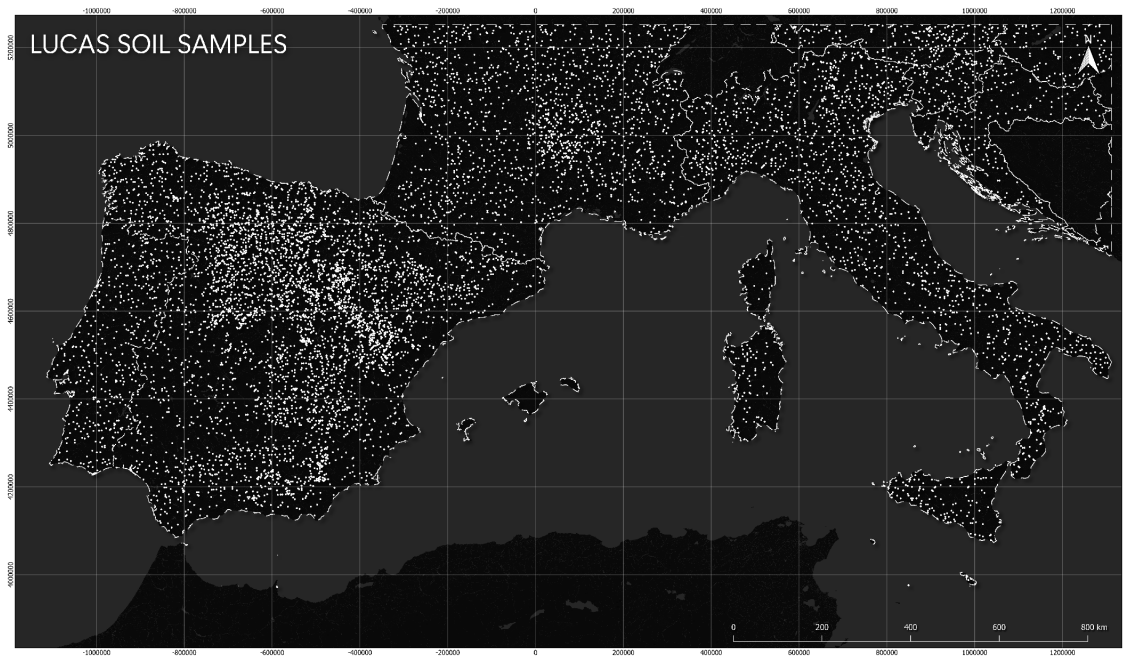
Figur 2.3. Afbildning af caseområdet hvori der prædikteres for jordparametre. Området inkluderer landene Portugal, Spanien, Andorra og det sydlige Frankrig. Gibraltar er ikke inkluderet i caseområdet.

Geografisk undersøgelse af samplingområdet 3

I dette afsnit gives en geografisk redegørelse for det sampling-område, der bruges for bedre at kunne estimere indenfor det mindre caseområde, hvilke begge blev præsenteret og defineret i afsnit 2, nærmere bestemt figur 2.2 og 2.3. Her beskrives den rumlige distribution af de forskellige anvendte jordparametre, der anvendes i LUCAS-undersøgelserne fra 2015 og 2018, som præsenteret i tabel 2.1. Endvidere gennemgås områdets Köppen-klimaklassifikation for 2018, arealanvendelsen som defineret via CORINE Land Cover/Land Use undersøgelsen for 2018, samt et højdekort. Kortlægningen og præsentationen af disse datatyper bidrager til en bedre forståelse af området og dens kontekst, hvori jordprøverne fra LUCAS-undersøgelserne er taget. Disse lag kan eksempelvis fortælle om hvilken terræntype prøverne er taget fra, fra hvilken højde, om de er samlet i en skov, på en mark, i et vådområde osv., samt om der er tale om et varmt steppeområde eller tempereret regnskov.

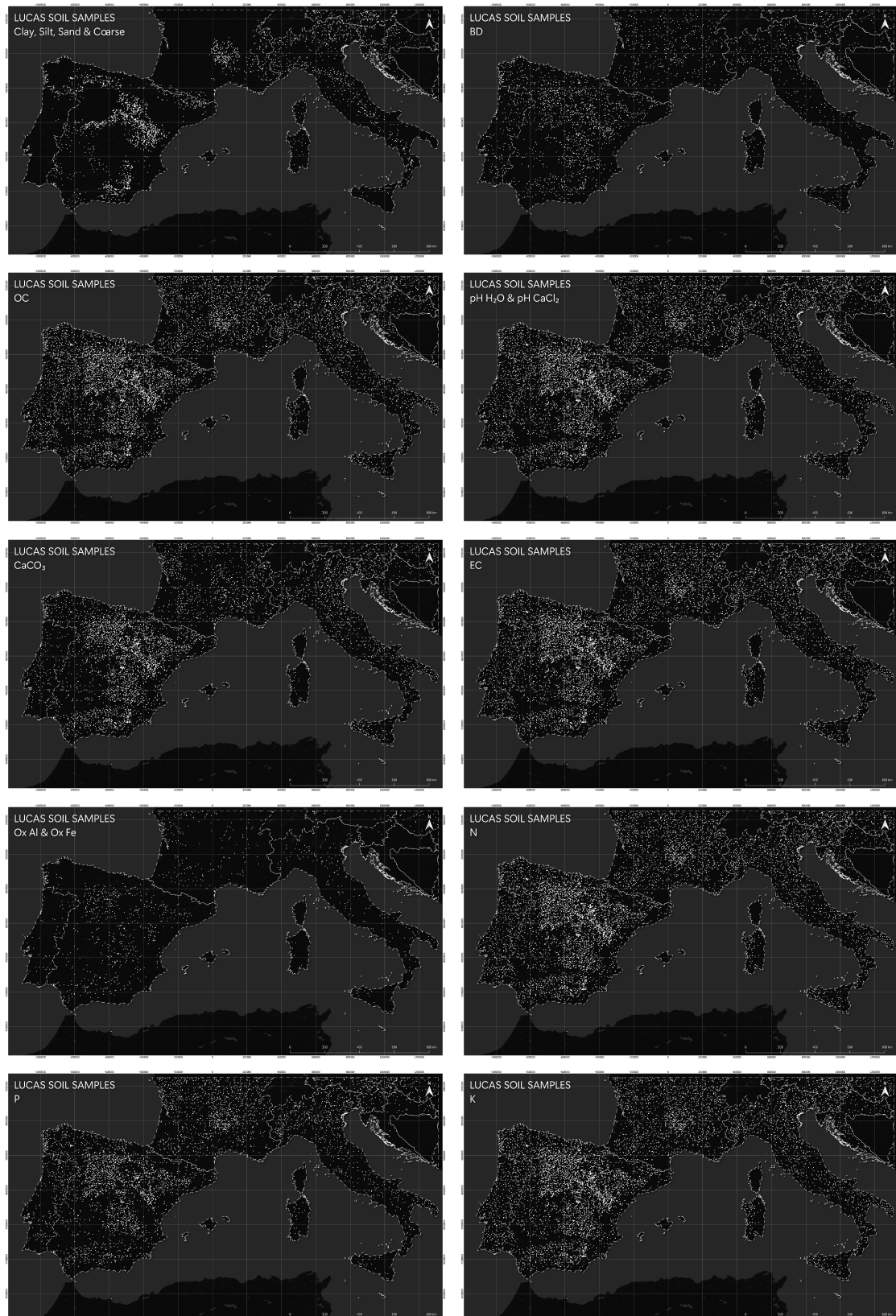
3.1 Kortlægning af LUCAS jordprøver

LUCAS-datasættene er en central komponent i udviklingen af det maskinlæringsbaserede koncept for jordprædiktering i dette projekt, idet datasættet tilbyder en bred dækning af forskellige jordparametre på tværs af hele EU. LUCAS-datasættene er fra henholdsvis 2015 og 2018, og tilbyder hver med information om de undersøgte jordparametre. Fra 2015-undersøgelsen er der data om jordens tekstur, herunder ler, silt, sand og grove partikler, mens 2018-undersøgelsen indeholder opdaterede data om organisk kulstof, pH, bulk density (BD) og andre kemiske parametre som Ox Al og Fe. I den følgende figur 3.1 ses den overordnede rumlige fordeling af jordprøver indsamlet gennem LUCAS-undersøgelserne fra 2015 og 2018.



Figur 3.1. Rumlig distribution af jordprøverne taget i forbindelse med LUCAS 2015 og 2018 undersøgelserne.

LUCAS-undersøgelserne fra 2015 og 2018 indeholder en bred vifte af jordparametre, som oplistet i tabel 2.1, men i mange af undersøgelsernes jordprøver mangler der målinger af samtlige parametre. For at præcisere datatilgængeligheden og distributionen af jordparametrene på tværs af disse prøver, opdeles de individuelle jordparametre i flere kort i figur 3.2. Her bliver det tydeligt, at der for de individuelle parametre er en stor variation i antal, fordeling og geografisk udbredelse afhængigt af den pågældende jordparameter i jordprøverne.



Figur 3.2. Rumlig distribution af de individuelle jordparametre taget i forbindelse med LUCAS 2015 og 2018 undersøgelsene. Kun parametrene for ler, silt, sand og coarse kommer af 2015 undersøgelsen, imens de andre er fra 2018.

I figur 3.2 kan det observeres, at jordprøverne for de fleste parametre er jævnt fordelt, selvom enkelte områder har en lavere tæthed af prøvetagninger. Der ses dog en undtagelse for parametrene ler, silt sand og coarse, hvor prøverne er koncentrerede i specifikke geografiske områder. Dette er særligt tydeligt i visse områder i Spanien og det sydlige Frankrig.

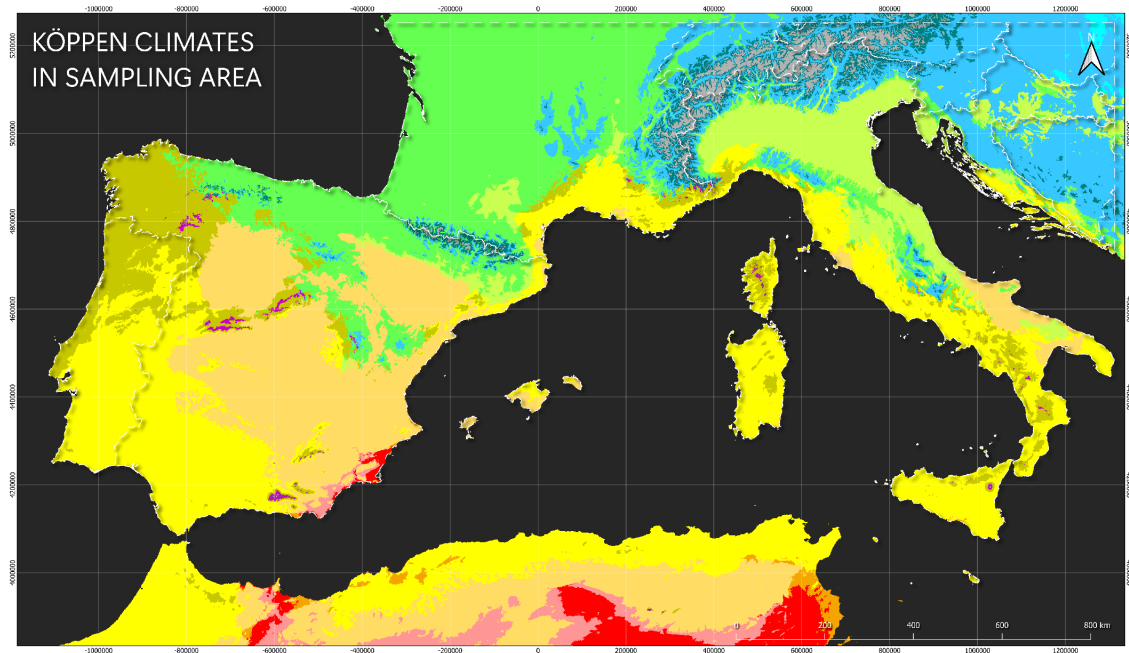
For at sætte densiteten af LUCAS jordprøver i relief til konceptets omfang, vides det, at én LUCAS jordprøve dækker et areal på hvad der svarer til $4 \times 4 = 16 \text{ m}^2$ antaget en kvadrat [European Commission et al., 2020, s. 2], [European Commission et al., 2022, s. 10]. Dette betyder, at der cirka er $\frac{1}{16} = 0,0625/\text{m}^2$ LUCAS jordprøver per kvadratmeter, da en hel jordprøve repræsenterer hvad der svarer til 16 m^2 . Da dette koncept anvender rasters med en pixelstørrelse på i alt $250 \times 250 = 62.500 \text{ m}^2$, betyder det at sådan en pixel i bedste scenario, vil repræsentere et antal af jordprøver fra kvadratmeter på $\frac{4 \times 4 \text{ m}^2}{250 \times 250 \text{ m}^2} = 0,000256/\text{m}^2$, som overført betyder, at ét LUCAS jordprøve, der reelt er gennemsnittet for 5 "subsamples", dvs. små jordprøver der smmenlagt giver et gennemsnit af et 16 m^2 areal, skal repræsentere et areal på op til 62.500 m^2 . Der er derfor et stort element i dette koncept, som handler om skala. Ved at forstå disse potentielle implikationer der er gældende i forbindelse med udvikling af dette Maskin Læringsbaserede jordprædikterings konceptmodel, hjælper med at tydeliggøre, at dette er et koncept og derfor ikke isoleret kan bruges til at prædiktere præcist hvad der findes i en jord i mikroskala, uden at ændre på den overordnede skala, så anvendt data stemmer helt overens med hvad eksempelvis LUCAS-jordprøverne i bedste fald repræsenterer.

3.2 Köppens klimaklassifikation

For at få en bedre forståelse af områdets forskellige klima, kan der i figur 3.3 ses hvordan området er delt op i i forhold til Köppen Klima Klassifikationerne. Følgende tabel 3.1 giver et overblik over de forskellige indekser og farverne allokeret til de forskellige klimaer, som er afbilledet i figur 3.3 derefter.

Tabel 3.1. Liste over Köppen Klima Klassifikation og hvad det enkelte indeks betyder.
Köppen Klima Klassifikationsdefinitionerne som brugt her er fra [Beck et al., 2018].

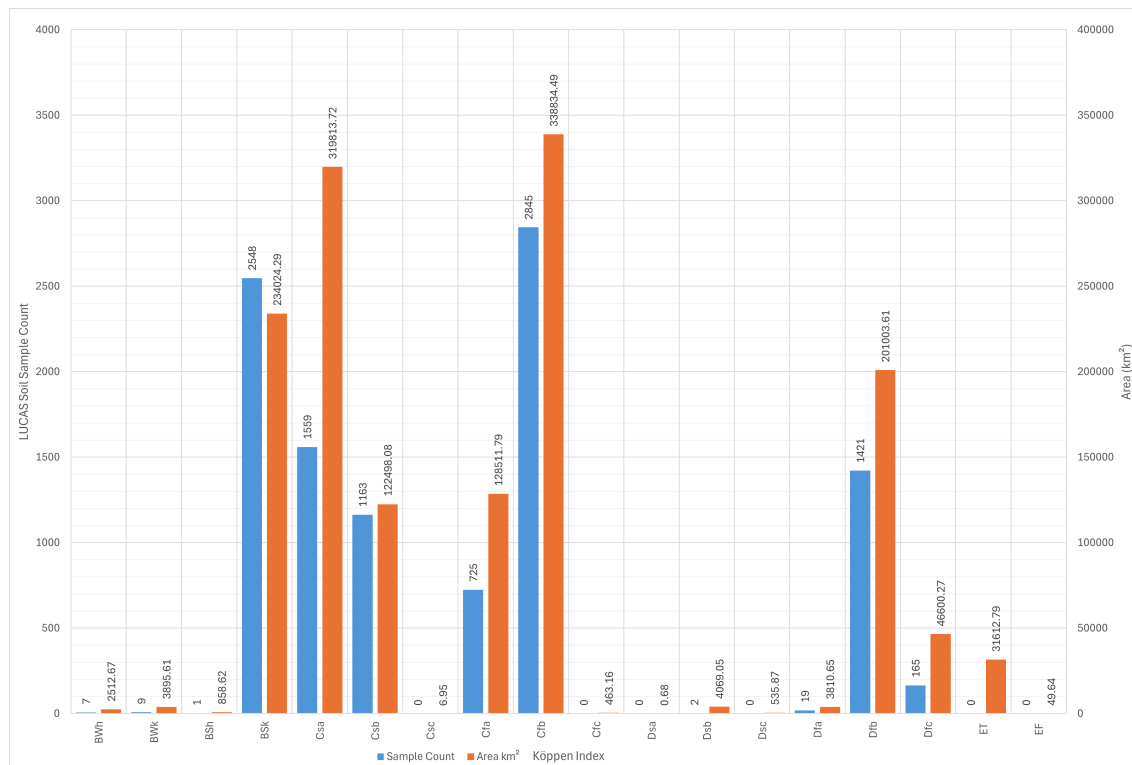
Indeks	
Af	Tropical, rainforest
Am	Tropical, monsoon
Aw	Tropical, savannah
BWh	Arid, desert, hot
BWk	Arid, desert, cold
BSh	Arid, steppe, hot
BSk	Arid, steppe, cold
Csa	Temperate, dry summer, hot summer
Csb	Temperate, dry summer, warm summer
Csc	Temperate, dry summer, cold summer
Cwa	Temperate, dry winter, hot summer
Cwb	Temperate, dry winter, warm summer
Cwc	Temperate, dry winter, cold summer
Cfa	Temperate, no dry season, hot summer
Cfb	Temperate, no dry season, warm summer
Cfc	Temperate, no dry season, cold summer
Dsa	Cold, dry summer, hot summer
Dsb	Cold, dry summer, warm summer
Dsc	Cold, dry summer, cold summer
Dsd	Cold, dry summer, very cold winter
Dwa	Cold, dry winter, hot summer
Dwb	Cold, dry winter, warm summer
Dwc	Cold, dry winter, cold summer
Dwd	Cold, dry winter, very cold winter
Dfa	Cold, no dry season, hot summer
Dfb	Cold, no dry season, warm summer
Dfc	Cold, no dry season, cold summer
Dfd	Cold, no dry season, very cold winter
ET	Polar, tundra
EF	Polar, frost



Figur 3.3. Kortlægning af Köppen Klima Klassifikationerne for det større samplingsområde. Data fra [Beck et al., 2018].

I følgende figur 3.4 er der beregnet størrelsen af samplingsområdet som de forskellige klimaer udgør, samt hvor mange af punkterne der er taget ved hver af dem. Grafen inkluderer alle klimatyperne selvom der hverken er jordprøver taget i området, eller at klimaet ikke eksisterer i samplingsområdet. Dette er netop for at belyse, at disse klima ikke er eksisterende deri.

Ud fra kalkulationerne kan det ses, at klimatyperne *Csc*, *Cfc*, *Dsa*, *Dsc* og *EF* findes indenfor samplingsområdet, men at der ikke er taget nogle jordprøver herved. I modsætning er de mest repræsenterede klimatyper i forhold til mængden af jordprøver, blandt andet *BSk* og *Cfb*. Dette giver også mening, taget den rumlige fordeling af jordprøverne i betragtning, som kan ses i figur 3.1, samt de isolerede jordprøvetyper i den efterfølgende figur 3.2, hvor mange af jordprøverne for organisk kulstof er at finde i Spanien og Frankrig, samt hvor der er høj mængde af jordprøver, der undersøger parametrene ler, silt, sand og coarse i Spanien, hvor netop store dele af landet er udgjort af klimatypen *BSk*. Derudover ses det også af histogrammet i figur 3.4, at flere af de klimatyper som kun udgør en meget lille del af det totale samplingsområde, generelt har lav mængde af tilhørende LUCAS prøver og vis a vis har de områder, der udgør større dele af samplingsområdet, stor mængde af jordprøver.



Figur 3.4. Histogramanalyse af distributionen af Köppen Klima Klassifikationerne for og ved de samlede jordprøver per LUCAS 2015 og 2018 undersøgelserne. Plottet beskriver antallet af jordprøver for hvert klimaindeks, samt hvor stort et areal hvert klima repræsenterer indenfor området. Forskellen mellem akserne er en faktor 100.

I følgende tabel 3.2 er antallet af jordprøver per klima indsat samt størrelsen hver klimaklassifikation udgør af samplingsområdet. Summen af arealerne for alle Köppen Klimaklassifikationerne for samplingsområdet kan i dette scenario ikke udgøre 100% af samplingsområdets polygon, som brugt til at afgrænse området, fordi klimalaget, som er at se i figur 3.3, er et raster, hvilket medføre til enkelte komplikationer. Denne hindring har dog ingen effekt på hvad der er aflæst for hvert LUCAS-punkter, da de alle overlapper klimalaget. I tabel 3.2 er parametrene ler, silt, sand og coarse fra LUCAS-datasættet grupperet til én kolonne her navngivet *Soil Fractions*, i og med at hver parameter er målt ved samme punkter, hvorfor individuelle kolonner vil være 1:1 kopier af hinanden, og derved ikke bringer ny information. Det samme gælder og er gjort for både *pH H₂O* og *pH CaCl₂*, samt *Ox Al* og *Ox Fe* parametrene. Rød farve i kolonnerne indikerer større interne relative mængdefordeling, hvor blå markerer det modsatte. Dvs. at eks. den mest røde celle i kolonnen for OC, betyder at dette klima, i denne kontekst, er hvor der er flest jordprøver. De klimaer, som bla. fremgår af tabel 3.1, som ikke er benævnt i denne tabel 3.2, hverken udgør noget areal af samplingsområdet eller har jordprøver. Dvs at selv klimaer, som udgør et areal, ikke er inkluderet i tabellen, hvis der for ingen af parametrene er foretaget en jordprøve i det. Størrelsen af de forskellige klimaer, som ingen jordprøver har, kan fortsat aflæses i figur 3.4.

Tabel 3.2. Matrice over fordelingen i antal punkter per parameter fra LUCAS datasættene i forhold til typen af Köppen Klimaklassifikation. Den rød/blå farveskala er indikator for det relative antal, hvor rød er højest.

Köppen Category	Area km ²	Soil Fractions	OC	BD	CaCO ₃	EC	pH CaCl ₂ & H ₂ O	N	P	K	Ox Al & Fe
BWh	2512.67	0	6	3	7	7	7	7	7	7	1
BWk	3895.61	0	9	5	9	9	9	9	9	9	2
BSh	858.62	0	1	1	1	1	1	1	0	1	1
BSk	234024.29	946	2485	801	2194	2502	2502	2502	1605	2502	452
Csa	319813.72	573	1463	399	1078	1461	1463	1463	709	1463	254
Csb	122498.08	732	1071	353	614	1073	1073	1069	569	1071	62
Cfa	128511.79	258	675	161	552	675	675	675	468	675	145
Cfb	338834.49	1447	2684	789	1887	2690	2690	2688	1657	2690	287
Dsb	4069.05	2	2	2	2	2	2	2	2	2	0
Dfa	3810.65	1	18	4	13	18	18	18	9	18	2
Dfb	201003.61	1086	1265	395	824	1265	1265	1265	863	1265	57
Dfc	46600.27	157	134	48	79	134	134	134	114	132	0

Som det fremgår af tabel 3.2 bekræftes det yderligere, at majoriteten af punkterne, også for hver parameter, er samplet i klimatyperne *BSk* og *Cfb*. Derudover kan det dog også aflæses, at den arealmæssige fordeling af Köppen klimatyperne, følger et andet mønster; at typen *Cfb* repræsenterer det største areal af samplingsområde, hvor *Csa*, *BSk* og *Dfb* er følger tæt.

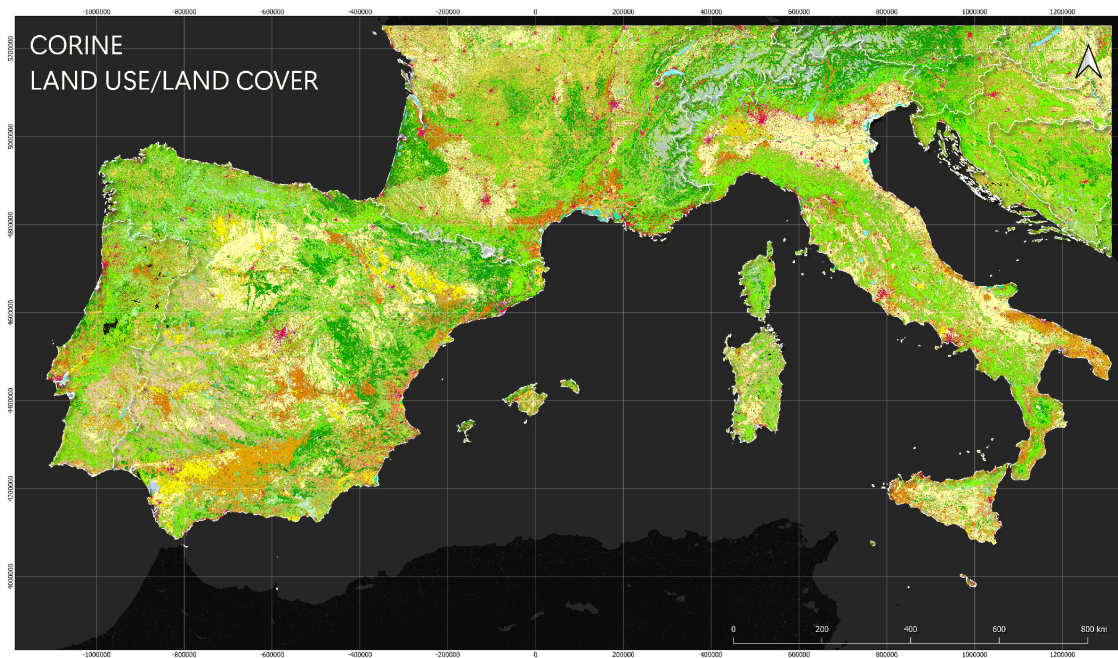
3.3 CORINE land cover/land use

Dernæst er der i figur 3.5 kortlagt de forskellige typer arealanvendelser der eksisterer i caseområdet, jf. CORINE Land Cover/Land Use undersøgelsen for 2018 [Copernicus, 2020]. Dette er baseret på CORINE 2018 datasættet af Copernicus [2020], der via Remote Sensing analyserer terrænet og indekserer områderne i forskellige arealanvendelsestyper på baggrund af hvad der ses. Alle arealanvendelser som er taget højde for via CORINE, kan ses i følgende tabel 3.3. I figur 3.6 oplyses der om antallet af jordprøver, som taget i forbindelse med LUCAS undersøgelserne, der er taget inden for hver CORINE arealtype, samt det totale areal [km^2] hver enkelte arealtype udgør af samplingsområdet.

Tabel 3.3. Beskrivelse af hver CORINE Land Cover 2018 (CLC) type, og hvad hver farve repræsenterer af arealanvendelse eller arealtype. Indeks, navne og farveklassifikationer er fra [Copernicus, 2020].

Indeks	Beskrivelse
111	Continuous urban fabric
112	Discontinuous urban fabric
121	Industrial or commercial units
122	Road and rail network and associated land
123	Port areas
124	Airports
131	Mineral extraction sites
132	Dump sites
133	Construction sites
141	Green urban areas
142	Sport and leisure facilities
211	Non-irrigated arable land
212	Permanently irrigated land
213	Rice fields
221	Vineyards
222	Fruit trees and berry plantations
223	Olive groves
231	Pastures
241	Annual crops associated with permanent crops
242	Complex cultivation patterns
243	Land principally occupied by agriculture, with significant areas of natural vegetation
244	Agro-forestry areas
311	Broad-leaved forest
312	Coniferous forest
313	Mixed forest
321	Natural grasslands
322	Moors and heathland
323	Sclerophyllous vegetation
324	Transitional woodland-shrub
331	Beeches, dunes, sands
332	Bare rocks
333	Sparsely vegetated areas
334	Burnt areas
335	Glaciers and perpetual snow
411	Inland marshes
412	Peat bogs
421	Salt marshes
422	Salines
423	Intertidal flats
511	Water courses
512	Water bodies
521	Coastal lagoons
522	Estuaries
523	Sea and ocean

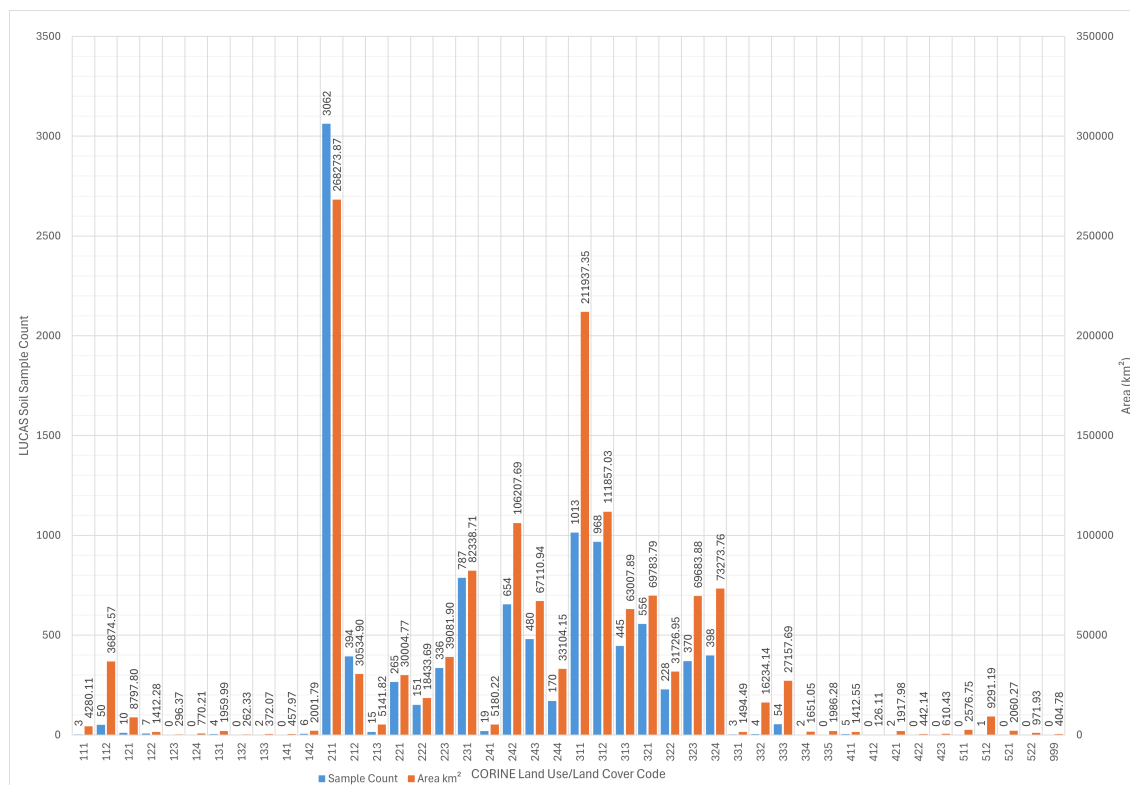
Følgende figur 3.5 formidler den rumlige fordeling, altså en kortlægning, af diverse arealtyper som defineret i forbindelse med CORINE Land Use/Land Cover.



Figur 3.5. Kortlægning af CORINE Land Cover 2018 (CLC) i og omkring caseområdet. Data fra [Copernicus, 2020].

I figur 3.6 herunder er der lavet en optælling af LUCAS punkterne samt en analyse og beregning af hvor store dele af samplingsområdet de forskellige CORINE arealtyper udgør. Størrelsesforskellen på akserne er en faktor 100, med arealet formidlet i km^2 og orange farve, imens blå farve er antallet af jordprøver. Antallet af jordprøver er for alle jordprøvetyper, dvs. at det inkluderer alle punkterne i datasættet, og at der derfor ikke

differentieres mellem om jordprøven eksempelvis har lerindhold eller ej. CORINE-koden 999 repræsenterer arealer af ukendt karakter.



Figur 3.6. Histogramanalyse af distributionen af CORINE Land Use/Land Cover for og ved de samlede jordprøver per LUCAS 2015 og 2018 undersøgelserne. Plottet beskriver antallet af jordprøver for hver type arealanvendelse, samt hvor stort et areal hver arealanvendelse repræsenterer indenfor området. Forskellen mellem akserne er en faktor 100.

Ud fra histogramundersøgelsen som præsenteret i figur 3.6, kan det ses, at store dele af samplingsområdet består af arealtypen af indeks 211, ækvivalent til "*Non-irrigated arable land*", med andre ord; dyrkede marker, som ikke bliver kunstigt vandet, men samtidigt drives landbrug på [Copernicus, 2020]. Dette er samtidigt den arealtype hvorindenfor en stor andel, på 3062 styk, af alle LUCAS jordprøverne er taget i.

Som gjort ved tabel 3.2 er punktaflæsningerne for jordparametrene ler, silt, sand og coarse fra LUCAS-datasættet samt for $pH\ H_2O$ og $pH\ CaCl_2$, samt $Ox\ Al$ og $Ox\ Fe$ grupperet af samme årsag i følgende tabel 3.4. Rød farve i kolonnerne indikerer større interne relative mængdefordeling, hvor blå markerer det modsatte. Dvs. at eks. den mest røde celle i kolonnen for OC, betyder at dette klima, i denne kontekst, er hvor der er flest jordprøver. Derudover er der frafilteret for alle CORINE Land Use/Land Cover arealerne, som enten ingen jordprøver har eller intet areal udgør af samplingsområdet.

Tabel 3.4. Matrice over fordelingen i antal punkter per parameter fra LUCAS datasættene i forhold til CORINE Land Use/Land Cover. Den rød/blå farveskala er indikator for det relative antal, hvor rød er højest.

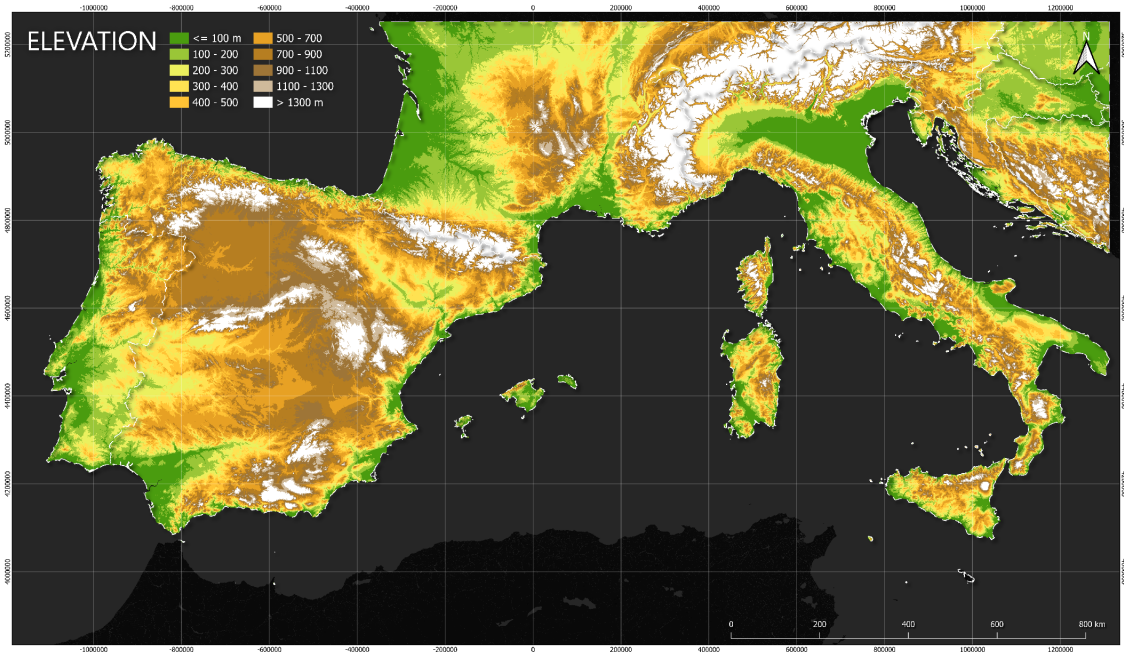
CORINE Code	Area km ²	Soil Fractions	OC	BD	CaCO ₃	EC	pH CaCl ₂ & H ₂ O	N	P	K	Ox Al & Fe
111	4280.11	0	3	1	3	3	3	3	1	3	0
112	36874.57	23	45	12	31	45	45	45	34	45	2
121	8797.80	4	8	2	7	8	8	8	5	8	0
122	1412.28	6	7	2	4	7	7	7	6	7	0
131	1959.99	1	3	0	2	3	3	3	1	3	1
133	372.07	0	2	1	2	2	2	2	1	2	0
142	2001.79	6	6	2	6	6	6	6	4	6	0
211	268273.87	914	2990	958	2394	3000	3000	2999	2213	3000	691
212	30534.90	50	388	136	345	392	392	391	328	392	83
213	5141.82	1	14	7	5	14	14	14	13	14	5
221	30004.77	37	262	78	222	263	264	264	217	264	73
222	18433.69	61	146	47	136	146	146	146	102	146	35
223	39081.90	92	325	105	284	328	328	328	192	328	102
231	82338.71	395	750	211	324	750	750	750	557	750	39
241	5180.22	4	19	10	10	19	19	19	11	19	3
242	106207.69	200	637	175	407	638	638	638	466	638	115
243	67110.94	253	447	118	303	447	447	447	254	447	41
244	33104.15	97	159	42	98	159	159	159	60	159	5
311	211937.35	656	890	238	680	890	890	890	393	890	24
312	111857.03	829	863	260	663	863	863	861	423	859	6
313	63007.89	321	390	110	321	390	390	390	173	390	2
321	69783.79	456	508	150	342	509	510	508	213	510	14
322	31726.95	172	210	66	94	210	210	210	106	210	4
323	69683.88	304	330	107	283	334	334	334	105	334	4
324	73273.76	267	353	103	256	353	353	353	114	353	11
331	1494.49	2	3	0	1	3	3	3	2	3	0
332	16234.14	3	3	0	2	3	3	3	0	3	0
333	27157.69	45	43	17	27	43	43	43	12	43	2
334	1651.05	0	2	1	1	2	2	2	0	2	0
411	1412.55	2	5	2	5	5	5	5	5	5	0
421	1917.98	0	2	0	2	2	2	2	1	2	1
512	9291.19	1	0	0	0	0	0	0	0	0	0

Som det kan ses af resultaterne fra tabel 3.3, er størstedelen af jordprøverne, for hver parameter, taget i områdekoden 211, altså *Non-irrigated arable land*. Dette er også arealet der udgør den største relative størrelse af samplingsområdet på omtrentligt 268273,87 km², efterfulgt af 311, *Broad-leaved forest*, med et areal på 211937,35, som trods størrelsen, rummer tildels samme mængde af jordprøver som 312, *Coniferous forest*, som udgør knap halvt så stort et område af 111857,03 km². Område 312 har dog flere jordprøver, med information om ler, silt, sand og coarse fraktionerne i jorden end område 311.

3.4 Terræn og højdemodel for samplingsområdet

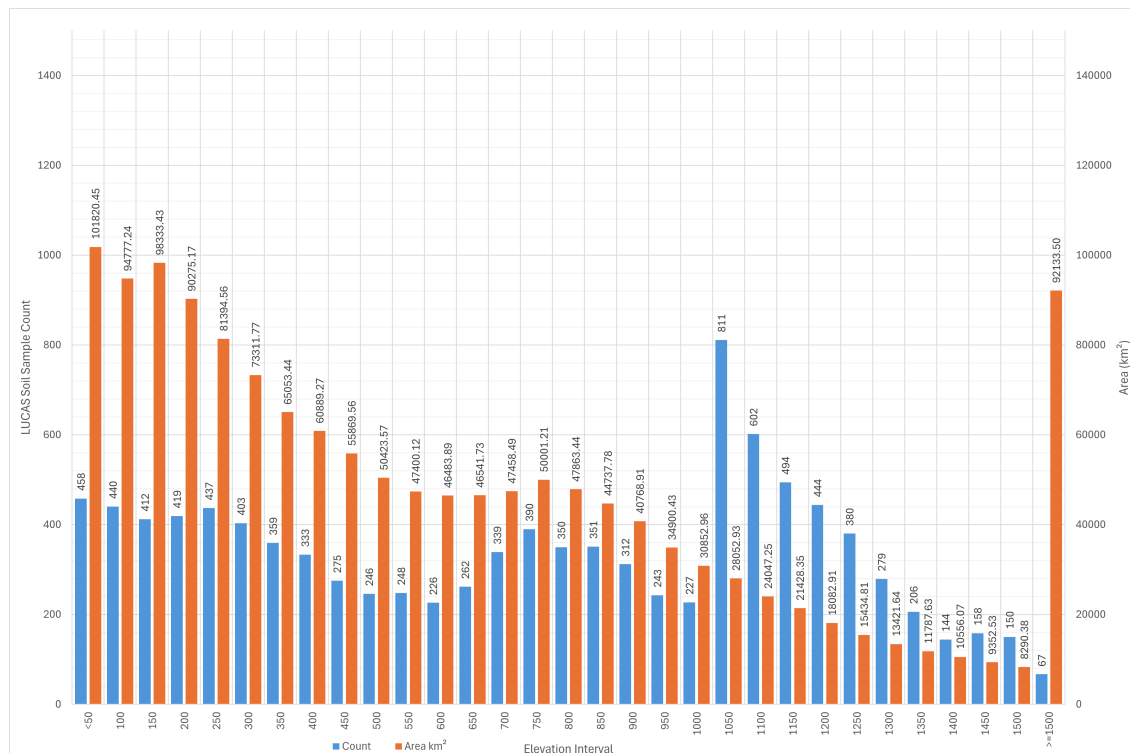
Dernæst, og som det sidste element i redegørelsen af samplingsområdet, er der for figur 3.7 produceret og præsenteret et højdekort over caseområdet. Dette højdekort viser terræts højde i enheden meter. Højderne er opdelt i intervaller af 100 meter i højden til og med 500 meter, hvorefter intervallet lyder på 200. Dette fremgår også af signaturforklaringen på kortet. Dataen der udgør højdekortet, også kendt som en *Digital Elevation Model (DEM)*, er distribueret af Geographic Information System of the Commission [2024]. Det højeste område i caseområdet er estimeret til at være 4379,87 m og laveste punkt på cirka -11,8 m. Eksempelvis er flere områder nær udmundingen af Po-floden i Italien under havniveau, samt områder i og omkring Venedig, Marseille i det sydlige Frankrig, udmundingen ved Ebro-floden i det nordlige Spanien, samt enkelte andre steder. Da hver

højde repræsenterer et areal på $30 \times 30 \text{ m}$ kan det derfor ikke vides med sikkerhed, om der i realiteten er højere eller lavere lokationer i samplingsområdet. Derudover er RMSE for højderne $2,9 \text{ m}$, hvorfor den egentligt højde kan varierer yderligere [Geographic Information System of the Commission, 2024].



Figur 3.7. Højdekort over samplingsområdet med enheden meter i intervaller som oplyst i signaturforklaringen. Højdekortet ekskluderer objekter som skove, bygninger og lignende, og formidler derfor kun selve terrænet. Højdekortet er endvidere en approksimation af terrænhøjden med netop de elementer ekskluderet. Data fra [Geographic Information System of the Commission, 2024].

Histogrammet i figur 3.8 er en undersøgelse på LUCAS punkterne for samplingsområdet, hvori grafen der sammenlignes mellem antallet af jordprøver per højdeinterval og det samlede areal per højdeinterval. Ud fra grafen kan det udledes, sammenlignet med alle de andre højdeintervaller, at højdeintervallet ved $[1000 - 1050 \text{ m})$ står i kontrast. Denne kontrast forstærkes i og med at dette højdeinterval udgør en væsentligt mindre del af det samlede samplingsområde. Udledt fra denne substantielle repræsentation af denne højde, kan netop denne højde fungerer overrepræsenterende for datasættet og dermed hvad der kommer til at indgå, og deraf beregnet på, i modellerne til prædiktions af de forskellige jordparametre.



Figur 3.8. Analyse af fordelingen af antal jordprøver per højdeinterval af inkrementer på 50 meter, hvor op til 50 meter inkluderer alt derunder, og 1500 alt derfra og over.

Følgende tabel 3.5 undersøger distributionen af jordprøver i højere detalje for hver undersøgte jordparameter. Et stort område af det totale samplingsområde er under 500 meters højde, og derover er arealerne gradvist mindre. Ved 1500 m og opefter ses der igen et stort areal, som dog har lille jordprøvemæssig repræsentation.

Tabel 3.5. Distribution af jordprøver taget per højdeintervaller af 50 meter. Tabellen viser distributionen for den enkelte jordparameter, men hvor *ler*, *silt*, *sand* og *coarse* er grupperet i kategorien *Soil Fractions* kolonnen.

Elevation	Area km ²	Soil Fractions	OC	BD	CaCO ₃	EC	pH CaCl ₂ & H ₂ O	N	P	K	Ox Al & Fe
<50	101820.45	86	441	124	361	442	442	439	331	440	111
100	94777.24	68	152	133	323	427	427	427	288	427	101
150	98333.43	87	233	110	283	395	395	395	264	395	79
200	90275.17	64	284	123	265	398	399	399	250	399	72
250	81394.56	74	346	104	273	425	425	425	269	425	73
300	73311.77	90	412	117	267	383	383	383	237	383	56
350	65053.44	89	473	88	234	350	350	350	210	350	45
400	60889.27	80	509	77	228	323	323	323	192	323	56
450	55869.56	72	540	70	168	256	257	257	157	257	35
500	50423.57	79	585	59	159	235	235	235	145	235	34
550	47400.12	83	234	69	150	235	235	235	139	235	32
600	46483.89	71	214	56	147	215	215	215	129	215	24
650	46541.73	61	253	83	199	255	255	255	154	255	39
700	47458.49	90	325	84	254	325	325	325	211	325	49
750	50001.21	76	374	104	304	378	378	378	250	378	56
800	47863.44	66	342	124	261	346	346	346	252	346	62
850	44737.78	70	343	98	275	345	345	345	253	345	61
900	40768.91	82	302	89	216	302	302	302	205	302	42
950	34900.43	75	235	49	173	236	236	236	161	236	37
1000	30852.96	105	217	44	173	218	218	217	149	218	26
1050	28052.93	782	747	214	587	749	749	747	406	749	42
1100	24047.25	584	546	221	433	546	546	546	278	546	42
1150	21428.35	484	442	143	329	442	442	442	236	442	27
1200	18082.91	434	408	133	313	412	412	412	228	412	28
1250	15434.81	374	338	142	250	338	338	338	182	338	16
1300	13421.64	275	246	87	182	246	246	246	124	244	8
1350	11787.63	202	170	71	128	170	170	170	81	170	2
1400	10556.07	136	124	48	89	124	124	124	64	124	4
1450	9352.53	154	140	37	101	140	140	140	67	140	2
1500	8290.38	146	126	36	101	126	126	126	62	126	2
>=1500	92133.50	64	55	24	34	55	55	55	39	55	0

I tabel 3.5 er jordparametrene *ler*, *silt*, *sand* og *coarse* parametrene af LUCAS datasættet grupperet i samme kolonne ved navn *Soil Fractions*, og ligeledes er *pH H₂O* og *pH CaCl₂*, samt *Ox Al* og *Ox Fe* for de respektive kolonner. I forlængelse af distributionen af prøver som vist i figur 3.8 kan det ses for jordparametrene *ler*, *silt*, *sand* og *coarse*, at i højden under 1000 m er tildels underrepræsenteret relativt til mængden og distributionen af prøver foretaget for de andre jordparametre. Af datasættet kan det derfor udledes, at der kan være en særligt højere detaljeringsgrad for *ler*, *silt*, *sand* og *coarse* i højderne over 1000 m. Hvad denne distribution netop kan påvirke, er eventuelle mønstre der kan udledes mellem forskellige kombinationer af uafhængige input parametre, herunder klima, CORINE, satellit mv., og målte indhold af de nævnte fire jordparametre.

Jordbundens sundhedsparametre 4

I projektets koncept er jordens sundhed baseret på en række sundhedsparametre eller sundhedsindikatorer. En sundhedsindikator udgør en væsentlig delmængde af faktorer, der er med til at kvantificere jordens sundhed. Hver indikator repræsenterer specifikke egenskaber ved jorden, der giver indsigt i dens evne til at opretholde sunde økosystemer, understøtte plantevækst, og bevare vigtige jordfunktioner såsom afdræning, iltning og næringsstofcyklusser. Sundhedsindikatorerne er derfor afgørende for at forstå de forskellige aspekter af jordens sundhedstilstand, og når de analyseres samlet, kan de danne et samlet billede af jorden som helhed.

I projektets koncept er disse sundhedsindikatorer beregnet på baggrund af jordens fysisk-kemiske egenskaber, som kan kvantificeres. De kan bl.a. omfatte blandt andet jordens tekstur (f.eks. ler-, silt- og sandindhold), plantetilgængelig porøsitet, surhedsgrad, gasdiffusivitet og hydraulisk ledningsevne. En samlet analyse af disse parametre kan give en indikation af, hvordan jorden vil reagere under forskellige forhold, såsom nedbør og tørring, samt hvordan den understøtter biologisk aktivitet. Det er derfor vigtigt at overveje blandt andet afdræningsevne og iltningsevne som centrale sundhedsindikatorer, da disse er afgørende for mikroorganismer og planterødderne, da disse har brug for adgang til ilt for at kunne fungere optimalt [Buckingham, 1904, 1907; Ashman og Puri, 2002].

Desuden spiller jordens organiske indhold en væsentlig rolle i lagringen af kulstof og opbygningen af jordaggregater. De organiske materialer, der nedbrydes til humus, fremmer dannelsen af stabile aggregater, som forbedrer jordens struktur. Disse aggregater øger jordens evne til at modstå erosion, forbedrer vandinfiltrationen og gør jorden mere produktiv. En velstruktureret jord med stabile aggregater er også mere modstandsdygtig over for menneskelige aktiviteter som intensiv dyrkning og klimaændringer [Schjønning et al., 2009].

I dette kapitel vil der blive fokuseret på disse aspekter, for at vurdere hvordan jordbundens sundhedsparametre er kvantificerbare og kan beregnes for en given jordbund, samt hvordan sundhedsindikatorerne hver især udgør en delmængde af jordens samlede sundhedstilstand.

4.1 Jordens grundlæggende fysio- og hydrologiske egenskaber

Jordens sundhed bestemmes af mange indikatorer, som er vigtige at kende for at kvantificere eller give en vurdering af, for at kunne give en vurdering af sundhedstilstanden i jordbunden. Der findes forskellige typer indikatorer, herunder fysiske, biologiske og

kemiske [Rakshit et al., 2020, s. 221, 244]. Blandt de fysiske indikatorer er volumenvægt, porøsitet, aggregatstabilitet og temperatur [Minnesota Pollution Control Agency, 2022b]. De biologiske og kemiske indikatorer inkluderer organisk indhold, samt toksiske stoffer og tilgængelige næringsstoffer, som desuden afhænger af jordens surhedsgrad (pH) [Minnesota Pollution Control Agency, 2022a].

Til at bestemme jordens sundhed, er en af de særlige indikatorer der bør undersøges jordens volumenvægt ρ_b [$g \text{ tørstof}/cm^3 \text{ jord}$], som netop beskriver det proportionelle forhold mellem jordens vægt og dets volumen, og er endvidere en sigende parameter af jordens volumenindhold af partikler V_p [db/ds]. Formlen for at beregne volumenvægt lyder generelt som følgende ligning 4.1. Heri er M_s tørvægten af en jordprøve, og V_t er jordenprøvens samlede volumen [Loll og Moldrup, 2000, s. 11].

$$\rho_b = \frac{M_s}{V_t} \quad (4.1)$$

Jordens partikeldensitet ρ_s [$g \text{ tørstof}/cm^3 \text{ tørstof}$] defineres ofte til at være mellem 2,65 - 2,67 g/cm^3 , da de fleste jorde typisk har en partikeldensitet mellem 2,6 og 2,7 afhængigt af jordtekstur-klassificeringen. Jordens partikeldensitet beregnes som via ligning 4.2, hvor V_s er volumen af jordens partikler [Loll og Moldrup, 2000, s. 11].

$$\rho_s = \frac{M_s}{V_s} \quad (4.2)$$

Endvidere påvirker indholdet af organisk materiale partikeldensiteten ved at formindske den [Loll og Moldrup, 2000, s. 11], [Minnesota Pollution Control Agency, 2022b]. Partikeldensiteten er relevant i forhold til bestemmelse af jordens partikelvolumen [$cm^3 \text{ partikler}/cm^3 \text{ jord}$], som blandt andet kan beregnes ved kendskabet til både jordens volumenvægt og partikeldensitet, og lyder som vist i ligning 4.3 [Loll og Moldrup, 2000, s. 11].

$$V_p = \frac{\rho_b}{\rho_s} \quad (4.3)$$

Hvor jordens partikelvolumen kan bruges til at beregne jordens totalporøsitet, ϕ [$cm^3 \text{ porer}/cm^3 \text{ jord}$], som anvist i ligning 4.4 herunder [Loll og Moldrup, 2000, s. 11].

$$\phi = 1 - V_p \quad (4.4)$$

4.2 Aggregatdannelse og Dexter Ratio

Yderligere afhænger jordens sundhed af dens modstandssygtighed overfor klimaændringer og menneskelig aktivitet, særligt i forhold til at vedligeholde det organiske stofindhold (OM) i jorden, der sammen med små mineralske partikler som ler, er med til at danne jordaggregater og opretholde jordens struktur. Dyrkning af marker er med til at mindske OM-indholdet i jorden, og kan i længden lede til en nedbrydning af jordaggregaterne og dermed en løsrivelse af lerpartiklerne. Dette kan potentielt lede til, at lerpartiklerne herefter udvaskes eller eroderes fra jordbunden, og manglen på ler kan samtidig mindske

tærsklen for aggregatdannelse og potentialet for lagring af organisk kulstofindhold (OC) i jorden [Schjønning et al., 2009].

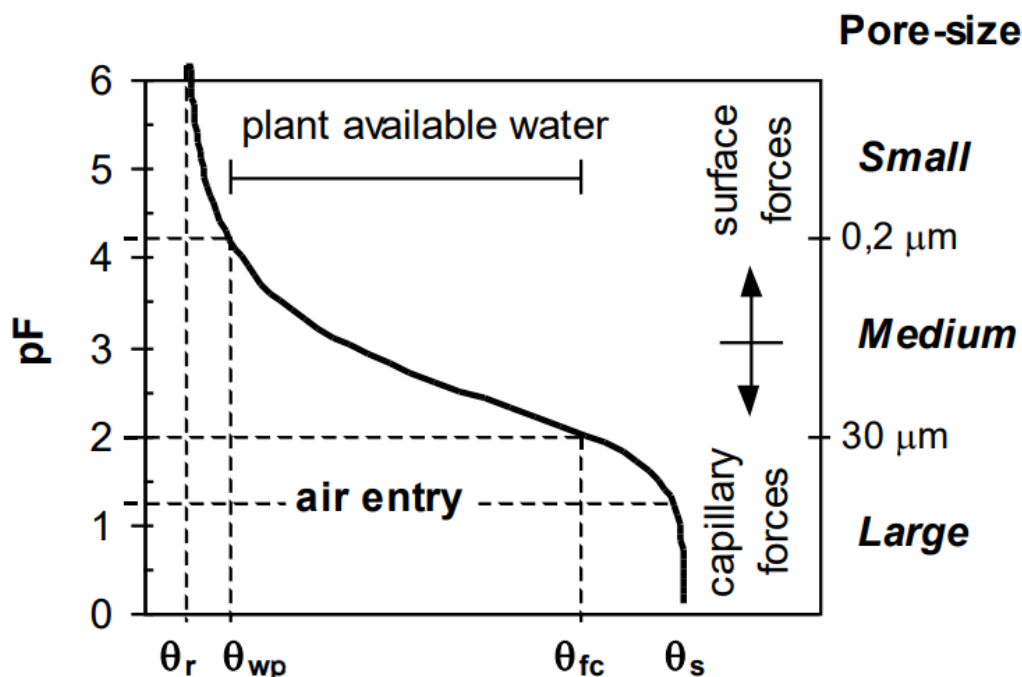
Dexter et al. [2008] undersøgte hvordan OC-indholdet i en jord påvirker dets fysiske egenskaber, særligt aggregatdannelsen mellem ler ($< 20 \mu m$), og organisk kulstof, OC . Heraf har de vist, at mange dyrkbare jorde har en ratio, n , mellem jordens lerindhold i vægtprocent, L [$g \text{ ler } g \text{ jord}$] og OC [$g \text{ C } g \text{ jord}$] på over 10 [Schjønning et al., 2010].

Det antages at der er et forhold mellem organisk stof (OM) og organisk kulstof (OC) med en ratio på $(OM/OC) = 1,724 \frac{g}{g}$ [Dexter et al., 2008]. Det svarer til 1 kg OM for hver $10/1,724 = 5,8 \text{ g}$ ler.

Baseret på resultaterne fra studiet af Dexter et al. [2008], er jorde med $L/OC > 10$ umættet med OC, hvilket betyder, at der er risiko for at leret derfor ikke kan indgå i komplekse aggregater. Jorde med en Dexter ratio $n = L/OC$ på over 10 (eller $L/OM > 5,8$), indikerer dermed, at jorden har for lidt organisk kulstofindhold i forhold til lerindholdet [Schjønning et al., 2009], [Dexter et al., 2008].

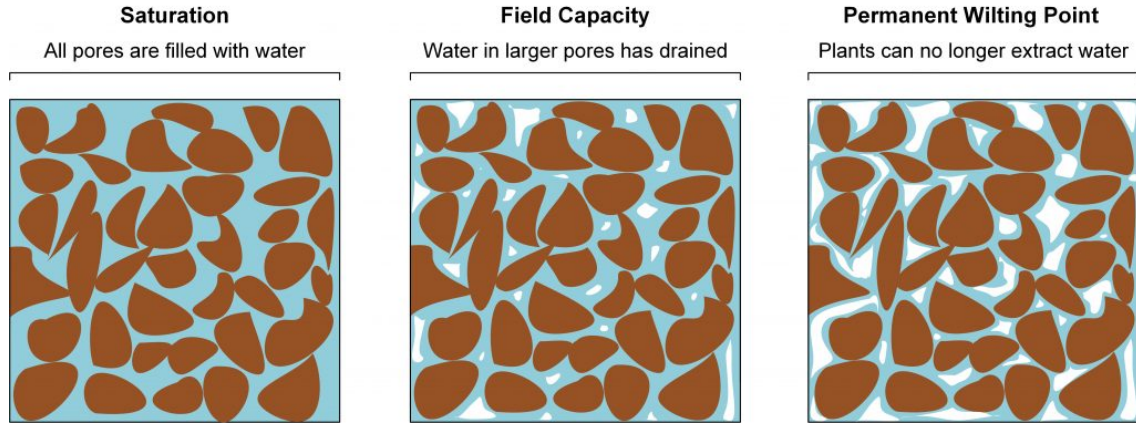
4.3 Jordens plantetilgængelige vand

Porehulrummene varierer i størrelse, og deres størrelse er med til at bestemme både vandretentionsevnen og ledningsevnen af luft i jorden. I de *små porehulrum* holdes vandet tæt ind på jordpartiklerne, hvor overfladespændingen og adhæsionen er for stærk til, at planterne kan opsuge vandet. I modsætning er kapillærkræfterne svage nok til at vandet drænes hurtigt i de *store porer*, hvorefter de fyldes med luft, og bidrager til jordens iltning [Ashman og Puri, 2002, s. 38-39]. De *mellemstore porer* på mellem 30 og $0,2 \mu m$ (se figur 4.1) er små nok til at kapillærkræfterne holder på vandet, og er samtidig store nok til, at planter kan suge vand ud af dem. Derfor anses volumen i de mellemstore porer som plantetilgængelige vand (PAW), og beskriver mængden af vand mellem *markkapacitet* θ_{fc} ($pF \ 2$) når jorden er meget våd, hvilket gør det let optageligt for planter, og *visnegrænse* θ_{wp} ($pF \ 4,2$) hvor planterødderne visner på grund af vandmangel [Ashman og Puri, 2002, s. 121]. På figur 4.1 er der opstillet et eksempel på en vandretentionskurve, som viser hvordan vandets trykpotentiale pF påvirker jordens vandindhold, heraf at det falder i takt med at pF stiger.



Figur 4.1. Vandretentionskurven viser forholdet mellem jordens vandindhold og sugetrykket pF. Jorden er vandmættet, θ_s ved pF 0; markkapacitet θ_{fc} er ved pF 2; visnegrænsen θ_{wp} er ved pF 4,2; og jorden er luft-tør θ_r ved pF 6. Jordens evne til at tilbageholde vand afhænger primært af porernes størrelse. Air entry beskriver sugetrykket, hvor de første porer begynder at dræne [Loll og Moldrup, 2000, s. 23].

Det vil sige, at hvis en jord er vandmættet θ_s (pF 0), vil de store porehulrum være de første, der udtømmes for vand gennem dræning ved hjælp af gravitationskraften, indtil jorden når markkapacitet θ_{fc} (pF 2), svarende til 10 kPa (1 m vandsøjle). Planterødderne kan yderligere suge vandet op fra jorden ned til visnegrænsen θ_{wp} , tilsvarende til et porevandstryk på pF 4,2 eller 1500 kPa (150 m vandsøjle). Herefter ($>$ pF 4,2) vil vandet kun være tilbage i de små porehulrum, hvor planterøddernes sugesevne er for svag til at optage vandet og derfor er uoptageligt for planterne. Ved pF 6 klassificeres jorden som værende luft-tør θ_r , og hvor vandindholdet ikke kan mindskes yderligere, og hvor det resterende vand sidder helt tæt på jordpartiklerne grundet adhæsion [Ashman og Puri, 2002, s. 44], [Loll og Moldrup, 2000, s. 23]. I følgende figur 4.2 ses en skitsering over vandmolekylernes evne til at adsorbere til jordpartikler og hinanden via intermolekylære effekter, ved tre forskellige grader af vandindhold; herunder vandindholdet ved fuldkommen mætning (pF 0), indholdet ved markkapacitet (pF 2) og indholdet ved visnegrænsen (pF 4,2).



Figur 4.2. Skitse af vandets absorptionsevne til jordpartikler ved forskellige niveauer af vandindhold; vandmættet, markkapacitet og visnegrænsen, svarende til sugetryk på henholdsvis pF 0, pF 2 og pF 4,2. Kilde: [Datta et al., 2018].

Jordens plantetilgængelige vand (PAW) [$\frac{cm^3 \text{ vand}}{cm^3 \text{ jord}}$] kan som udgangspunkt beregnes som anvist i ligning 4.5, hvor der kendes til vandindholdet ved både visnegrænsen (θ_{wp} , pF 4,2) [$\frac{cm^3 \text{ vand}}{cm^3 \text{ jord}}$] samt vandindholdet ved markkapacitet (θ_{fc} , pF 2) [$\frac{cm^3 \text{ vand}}{cm^3 \text{ jord}}$]. Denne formel er givet, eftersom θ_{wp} indikerer vandindholdet for den nedre grænse, hvilket vil sige, hvor planterødderne typisk ikke længere kan opsuge vandet i jorden. θ_{fc} markerer den øvre grænse, da overskydende vand påvirkes for nemt af tyngdekraften og derved drænes relativt uhindret afhængigt af jordteksturen. θ_{fc} beskriver med andre ord den øvre grænse, for hvor planterødderne kan nå at optage vand [Minnesota Pollution Control Agency, 2023].

$$PAW = \theta_{fc} - \theta_{wp} \quad (4.5)$$

Da der skal kendes til både θ_{fc} og θ_{wp} for at estimere plantetilgængeligt vand (PAW), kan ligning 4.6 anvendes, hvortil koefficienterne for 12 specifikke pF er defineret i følgende tabel 4.1. Koefficienter til estimering af volumetrisk vandindhold [cm^3/cm^3] for de specifikke pF er estimeret på baggrund af 43 jordprøver taget i både Missouri og South Carolina i USA, ved brug af parametrene ler [%], silt [%], sand [%], organisk materiale (OM [%]) og volumenvægt (ρ_b , [g/cm^3]) [Gupta og Larson, 1979, s. 1633].

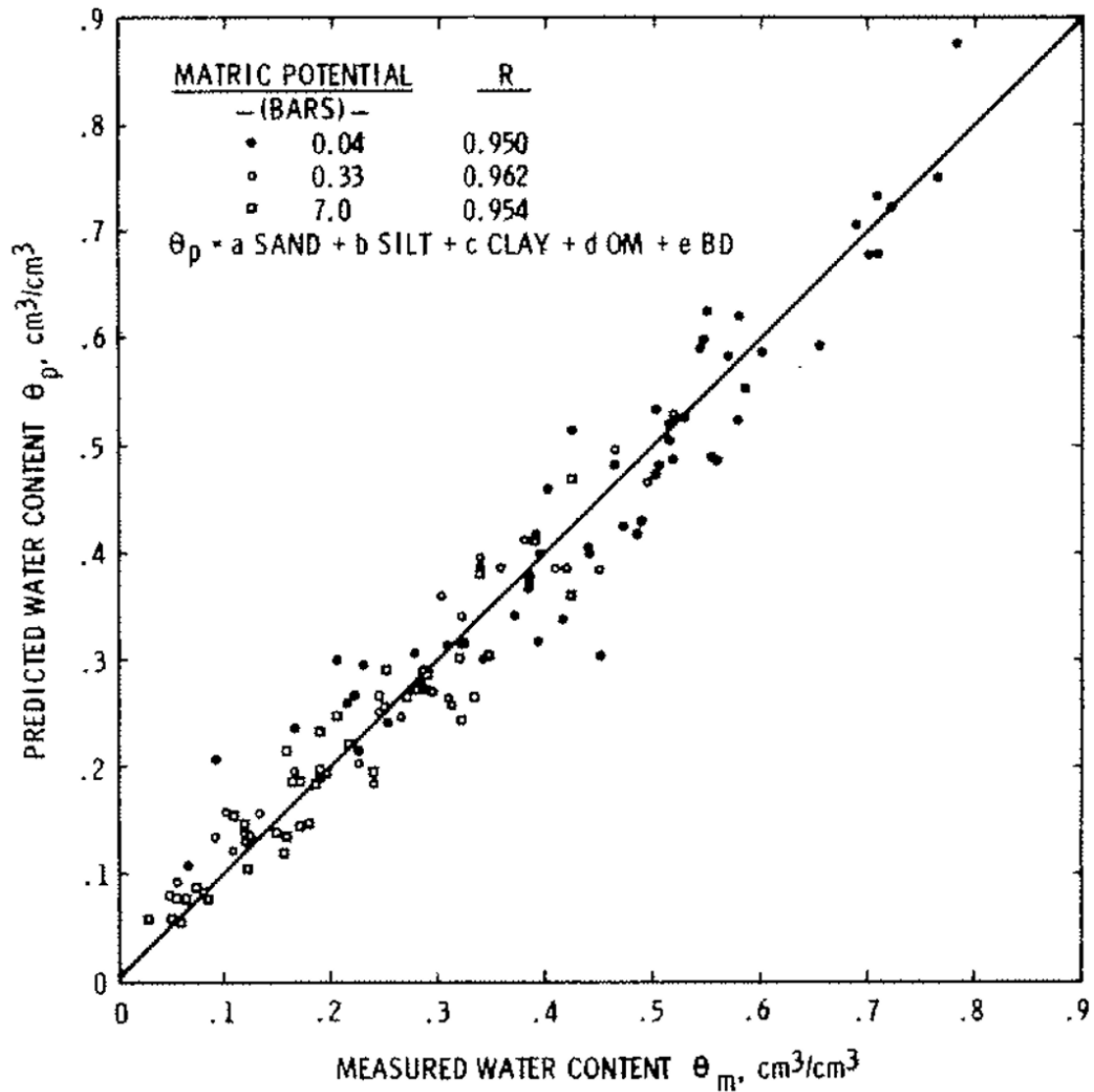
Tabel 4.1. Oversigt over de forskellige koefficienter som indgår i ligningen for prædiktion af volumetrisk vandindhold [cm^3/cm^3] i ligning 4.6 ved specifikke vandtryk, og hvor $pF = \log_{10}(|kPa| * 10)$. Ligningerne er udarbejdet med udgangspunkt i følgende partikelstørrelser; ler = $< 0,002 mm$, silt = $0,002 - 0,05 mm$ og sand = $0,05 - 2,0 mm$, med udgangspunkt i jordprøver foretaget i South Carolina, USA [Gupta og Larson, 1979, s. 1634].

Matric Potential		Regression Coefficients					Correlation Coefficient
pF	kPa	$a \times 10^3$	$b \times 10^3$	$c \times 10^3$	$d \times 10^3$	$e \times 10^2$	R
1,6	-4	7,053	10,242	10,070	6,333	-32,120	0,950
1,85	-7	5,678	9,228	9,135	6,103	-26,960	0,959
2	-10	5,018	8,548	8,833	4,966	-24,230	0,961
2,3	-20	3,890	7,066	8,408	2,817	-18,780	0,962
2,52	-33	3,075	5,886	8,039	2,208	-14,340	0,962
2,78	-60	2,181	4,557	7,557	2,191	-9,276	0,964
3	-100	1,563	3,620	7,154	2,388	-5,759	0,966
3,3	-200	0,932	2,643	6,636	2,717	-2,214	0,967
3,6	-400	0,483	1,943	6,128	2,925	-0,204	0,962
3,85	-700	0,214	1,538	5,908	2,855	1,530	0,954
4	-1000	0,076	1,334	5,802	2,653	2,145	0,951
4,18	-1500	-0,059	1,142	5,766	2,228	2,671	0,947

Følgende figur 4.3 viser et 1:1 diagram for præcisionen af tre specifikke ligninger for pF 1,6, 2,52 og 3,85, hvor der for dem alle opnås en R på mindst 0,95. Heri er anvendt netop koefficienterne som anvist i tabel 4.1, og testet på 61 jordprøver taget i Missouri [Gupta og Larson, 1979, s. 1633-1635].

Som Gupta og Larson [1979] konkluderer på resultaterne regressionerne giver, kan de udviklede regressioner være brugbare til at estimere plantetilgængeligt vand (PAW) eller hydraulisk ledningsevne for jorde med en fornuftig præcision alene ud fra de anvendte jordparametre, som blandt andet vist i ligning 4.6. Dog bemærkes det også, at der

på baggrund af statistik udført på regressionsanalyserne som vist i figur 4.3, for alle regressioner, eksisterer et potentielt konstant bias, eftersom skæringspunkterne (α) ikke er nul for nogle af regressionerne. Denne bias bemærkes dog til at kunne være på grund af måden jordprøverne er behandlet på [Gupta og Larson, 1979, s. 1634-1635].



Figur 4.3. Et 1:1 diagram der viser forholdet mellem prædikeret og målt vandindhold, for pF 1,6 (0,04 bar), pF 2,52 (0,33 bar) og pF 3,85 (7 bar), for 61 jordprøver taget i Missouri i USA [Gupta og Larson, 1979, s. 1633-1635].

Med koefficienterne som oplyst i tabel 4.1 kan de anvendes til at prædiktere for vandindholdet [cm^3/cm^3] ved det enkelte potentielle vandtryk, ved kendskab til jordparametrene ler, silt, sand, organisk materiale (OM) og volumenvægt (ρ_b), og derefter blot at indsætte koefficienterne på deres respektive pladser i følgende ligning 4.6 [Gupta og Larson, 1979, s. 1633].

$$\theta_p = a \cdot \text{Sand} + b \cdot \text{Silt} + c \cdot \text{Ler} + d \cdot \text{OM} + e \cdot \rho_b \quad (4.6)$$

Følgende ligninger 4.7 og 4.8 er ligninger til estimering af jordens vandindhold ved

henholdvist 10 kPa og 1500 kPa (*pF 2 og 4,2*), med udgangspunkt i koefficienterne som benævnt i tabel 4.1 og dem indsat på de respektive pladser som anvist i ligning 4.6 [Gupta og Larson, 1979, s. 1633-1634].

$$\theta_{wp} = -0,000059 \cdot Sand + 0,001142 \cdot Silt + 0,005766 \cdot Ler + 0,002228 \cdot OM + 0,02671 \cdot \rho_b \quad (4.7)$$

$$\theta_{fc} = 0,005018 \cdot Sand + 0,008548 \cdot Silt + 0,008833 \cdot Ler + 0,004966 \cdot OM + -0,2423 \cdot \rho_b \quad (4.8)$$

Hvor OM [g/100g] bestemmes ved at omregne OC [g/kg] til procent via følgende ligning, således både θ_{wp} og θ_{fc} kan beregnes. Det antages at $\frac{OM}{OC}$ forholdet er 1,724 [Dexter et al., 2008].

$$OM[\%] = \frac{OC[g/kg] \cdot 1,724}{10} \quad (4.9)$$

4.4 Geniltning af jorden

Udover det planteoptagelige vand er transporten af luft og ilt i jorden afgørende for både planters vækst og jordens sundhed. Dette blev tidligt i 1900-tallet undersøgt af Edgar Buckingham, som væsentligt øgede forståelsen af vand- og luftbevægelser i jorden. I sine undersøgelser fandt Buckingham [1904], at gasdiffusiviteten i jord primært afhænger af jordens tekstur, struktur og kompakthed. Buckingham konkluderede i øvrigt, at luftens strømningshastighed kan bestemmes ud fra jordens volumetriske luftindhold, med følgende ligning:

$$\epsilon \left[\frac{cm^3 \text{ luft}}{cm^3 \text{ jord}} \right] = \theta_s - \theta_{fc} \quad (4.10)$$

Her er θ_s den mættede vandindholdskapacitet, og θ_{fc} er jordens markkapacitet. Buckingham opdagede, at gasdiffusiviteten i jord er proportional med jordens luftindhold i anden potens, og formulerede derfor geniltningsfunktionen som:

$$\frac{D}{D_o} = \epsilon^2 \quad (4.11)$$

Hvor $D[m^2/s]$ er diffusionskoefficienten i jorden, og $D_o[m^2/s]$ er diffusionskoefficienten i fri luft. Da luftindholdet ved θ_{fc} eller *pF 2* svarer til $-100 \text{ cm } H_2O$ matrixpotentiale, kan gasdiffusiviteten ved dette matrixpotentiale også udtrykkes som $D(p, 100)/D_o$.

Denne ligning kan relateres til Adolf Fick's lov om diffusion, der beskriver, hvordan gasser bevæger sig fra områder med høj koncentration til områder med lav koncentration. Fick's første lov beskriver gasdiffusion gennem et medie som:

$$J = -D \cdot \frac{dc}{dx} \quad (4.12)$$

Hvor J er diffusionsfluksen (mængden af gas per tidsenhed), D er diffusionskoefficienten (gasdiffusiviteten) og $\frac{dc}{dx}$ er koncentrationsgradienten, eller hvordan ændringen i en koncentration afhænger af afstanden.

I forhold til gasdiffusion i en jord betyder det, at jo højere gasdiffusivitet (D) i jorden, desto hurtigere kan ilt bevæge sig gennem jordens porer og nå planternes rødder. Buckingham's ligning, $\frac{D}{D_o} = \epsilon^2$, beskriver at der er et simpelt forhold mellem gasdiffusivitetskoefficienten og en jord og jordens volumetriske luftindhold, hvilket indikerer en stærk afhængighed af jordens luftfyldte porer for effektiv gasdiffusion. Dette understreger vigtigheden af porøsitet og strukturel integritet i jorden for at sikre tilstrækkelig ilttilførsel [Buckingham, 1904; Loll og Moldrup, 2000].

Flere observationer har vist, at gasdiffusiviteten påvirker koncentrationerne af ilt, CO_2 , metan og nitrat i jorden, og at jorden generelt bør have en gasdiffusionskoefficient tilsvarende til over 2% ($\frac{D}{D_o} > 0,02$) i forhold til fri luft for at sikre en tilstrækkelig geniltning af jorden, hvilket svarer til et volumetrisk luftindhold på ca. $\epsilon > 0,15$.

Dette påpeger, at jordens strukturelle egenskaber og gastransport er afgørende for jordens sundhed og planters vækst, og hvordan gasdiffusion påvirkes af ændringer i jordens porøsitet.

I senere studier opdagede Buckingham [1907] desuden, hvordan kapillærkræfter påvirker vandets bevægelse i jorden, og hvordan porevandstrykket (pF) bestemmer vandindholdet og afdræningen af vand. Buckingham definerede kapillærvandet som den mængde vand, der drænes ud af en vandmættet jord udelukkende på grund af tyngdekraften, uden fordampning.

Han sammenlignede jordens porer med kapillærrør, der tillader vand at bevæge sig opad, ligesom vand kan bevæge sig opad i et stykke papir, hvilket betyder, at jorden lige over grundvandsspejlet kan have et vandindhold over markkapacitet. Kapillærvandet svarer til de porer, der fyldes med luft ved markkapacitet, θ_{fc} (se figur 4.1, 4.2). Buckingham observerede også, at når jordens vandindhold falder, mindskes den hydrauliske ledningsevne tilsvarende. Jordens afdræning afhænger derfor også af det volumetriske luftindhold, ϵ , som kaldes den effektive porøsitet.

For både gasdiffusivitet og afdræning er det derfor ønskeligt, at den luftfyldte porøsitet ved markkapacitet, θ_{fc} , er $\epsilon > 0,15$ for at sikre en hensigtsmæssig afdræning og ilttilførsel i jorden [Loll og Moldrup, 2000; Buckingham, 1907].

4.5 Hydraulisk ledningsevne

Derudover har jordens hydrauliske ledningsevne (K) også en rolle i forhold til jordens sundhed og stabilitet. Elementer det har indflydelse på eller beretter om, er blandt andet hvor nemt vandet bevæger sig igennem jordsøjlen og derved hvor frit det løber. Eksempelvis, hvis vandet ikke kan infiltrere ned og igennem jordsøjlen, kan det blandt andet forøge risikoen for at jorden i rodzonen, dvs. jorden omkring planterødderne i variende dybder, bliver vandmættet [Bradford et al., 2008, s. 103]. Afhængigt af jordens vandmættethed, kan det potentielt forårsage skadelige niveauer af stress for planter, hvoraf tolerance-tærsklen dog kan variere afhængigt af plantens tilpassethed. En vandmættet jord kan netop over tid udvikle hypoxi, mangel på ilt, hvilket kan være skadeligt for planterødder, og i forlængelse hele planten, da nogle plantearter absorberer ilt fra jorden via deres rødder. Derudover kan en vandmættet jord også have implikationer for

planternes evne til at absorbere nødvendige næringsstoffer [Xu et al., 2024, s. 2877-2880]. Ligeledes kan manglen på tilgængeligt vand i rodzonen også være skadeligt for planternes levedygtighed. Derudover kan for lav hydraulisk ledningsevne i jorden også medføre til udvaskning af nødvendige næringsstoffer for planter, da majoriteten af vandet kan blive i overfladen, opløse næringsstoffer, og fjerne dem fra jorden igennem overfladeafstrømning eller lignende. I forlængelse af overfladeafstrømning, kan lav hydraulisk ledningsevne også medføre til forøget erosion af jordoverfladen, hvorfor også jordens fysiske helbred kan påvirkes [Meter Group, ud.]. I følgende tabel 4.2 er specifikke intervaller for K_{sat} givet en klassifikation i forhold til dets dræningsevne og fungerer som tærskelværdier i forhold til netop at kunne give en vurdering af jordbundens sundhed i forhold til dræning.

Tabel 4.2. Definitioner af klassifikationer for specifikke intervaller af mættet hydraulisk ledningsevne (K_{sat}), hvor klassifikationerne beskriver jordens dræningsevne [Juliá et al., 2021, s. 2].

Saturated Hydraulic Conductivity Range ($\mu m/s$)	Range of Log K_{sat} Values	Description of Saturated Hydraulic Conductivity Class
<0,01	<-2	Very low
0,01 - 0,1	-2 to -1	Low
0,1 - 1,0	-1 to 0	Moderately low
1,0 - 10	0 to 1	Moderately high
10 - 100	1 to 2	High
>100	>2	Very high

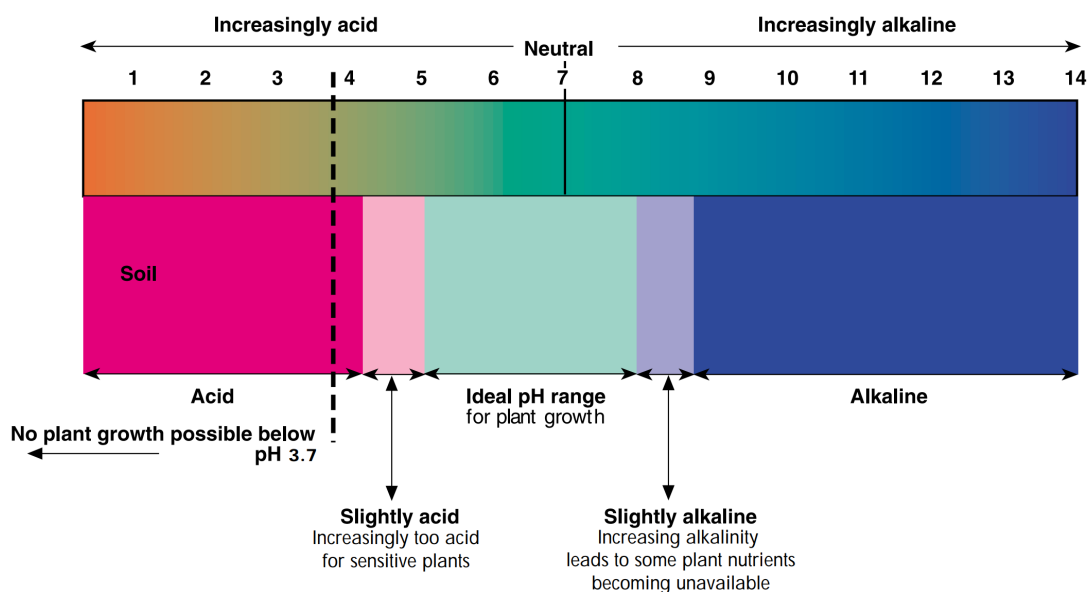
Af tabel 4.2 fremgår det eksempelvis, at ved en K_{sat} i spændet 10 – 100 $\mu m/s$ (*High*) klassificeres som værende af høj konduktivitet. En ledningsevne i denne skala indikerer en veldrænende og typisk hovedsageligt sandet jord. Ved en ledningsevne i på 10 – 100 $\mu m/s$, kan vandet passere tildels problemfrit igennem jordsøjlen, hvorved risikoen for vandmætning af jorden er minimal og samtidigt med god geniltning, hvilket kan være fordelagtigt for flere planter. I klassen under for intervallet 1,0 – 10 $\mu m/s$ (*Moderately high*) ses typisk jorde i "loam" kategorien, hvilket typisk anses for værende en optimal jordkomposition for planter. Disse jorde holder både bedre på vandet, men har fortsat en god dræningsevne til at undgå lange perioder med for høj grad af vandmætning over for lange perioder. Jorde med K_{sat} af klasserne *High* og *Moderately high* kan derfor anses som typisk værende sunde jorde. Dårlige jorde ligger udenfor dette interval, da dræningsevnen enten er for høj, eller for lav, hvor problemer som vandmætning, hypoxi, erosion eller lignende kan udvikle sig [Juliá et al., 2021, s. 2], [Meter Group, ud.].

Følgende pedotransfer ligning 4.13 anvendes til at estimere K_{sat} [mm/h], hvori organisk materiale (OM) omregnes fra organisk kulstof (OC) via ligning 4.9. Ligning 4.13 er udviklet på baggrund af i alt 3372 jordprøver, hvoraf 3193 også har undersøgt for organisk materiale, på den Iberiske Halvø, der dækker et bredt spænd af forskellige jord arter. Ligningen er endvidere en udvikling på pedotransfer funktioner udviklet for specifikke undersøgte jordtyper, hvor denne er en generaliseret ligning, der opnår en R^2 på 0,715 på de testede jordprøver, med en RMSE på 0,7484 hvoraf standardafvigelsen ligger på 0,356 [Ferrer Julià et al., 2004, s. 260-271].

$$K_{sat} = -4,994 + 0,56728 \cdot sand - 0,131 \cdot clay - 0,0127 \cdot OM \quad (4.13)$$

4.6 Toleranceintervaller for planter i forhold til jordens surhedsgrad

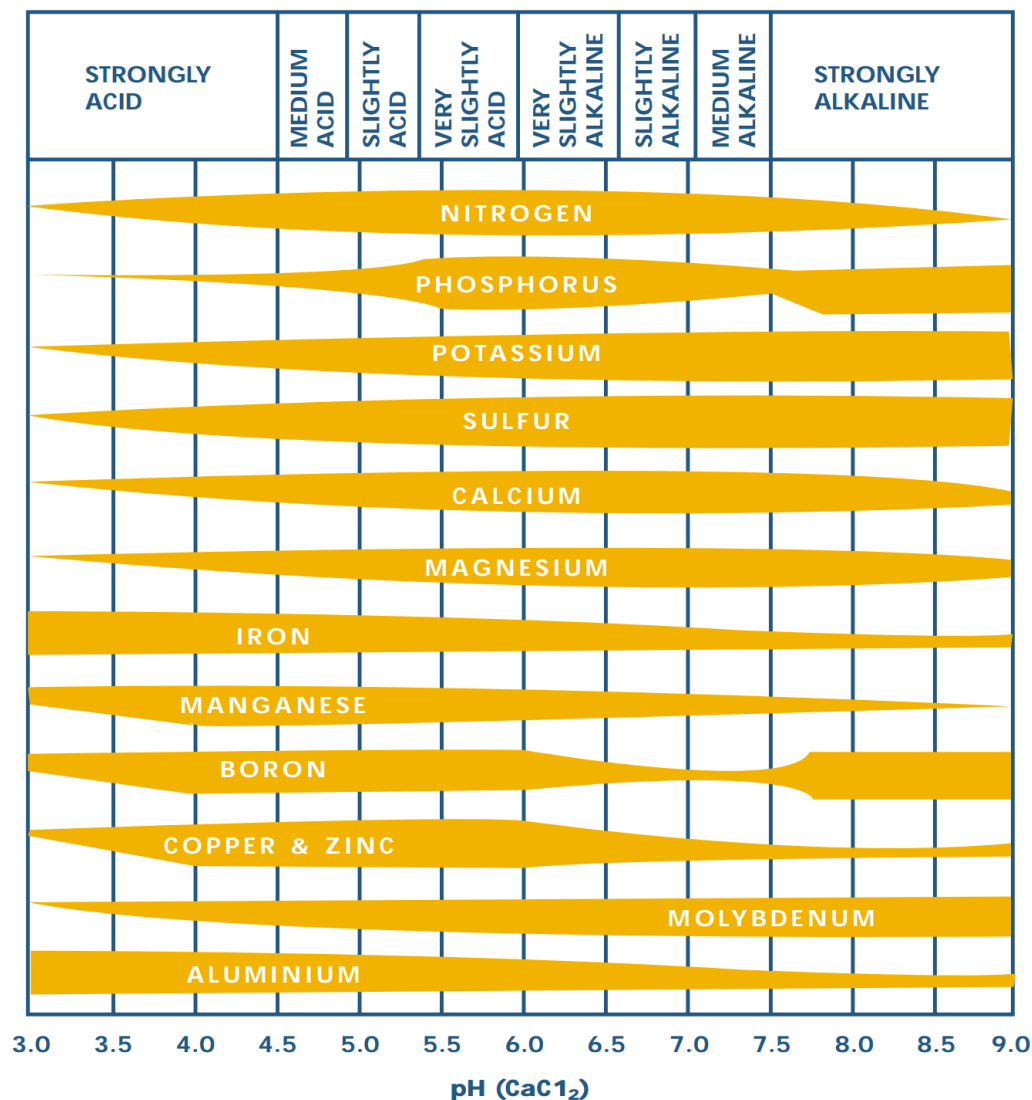
Jordens surhedsgrad, der måles som pH-værdi, er en væsentlig faktor for jordens sundhed, da den påvirker både de kemiske og biologiske processer i jorden. pH-værdien, som måles i jordvandet, reflekterer koncentrationen af hydrogenioner (H^+), hvor jo større en koncentration af H^+ -ioner, des stærkere er syren. Dermed kan surhedsgraden i en syre måles ved brug af pH, der er en logaritmisk skala baseret på koncentrationen af H^+ -ioner i en opløsning [Ashman og Puri, 2002, s. 58]. Surhedsgraden målt ud fra pH-skalaen spænder fra 0 (meget sur) til 14 (meget basisk), mens en pH-værdi på 7 betegnes som neutral, se figur 4.4.



Figur 4.4. Generel påvirkning af pH på planter for hele pH skalaen, eksempelvis hæmmer både for høj og for lav pH plantevæksten. Figur fra [NSW Department of Primary Industries, 2000, s. 1].

Jordens surhedsgrad kan også bruges som indikator for hvilke kationer, der er bundet til jordens partikler eller opløst i jordvæsken, hvor de er tilgængelige for planter. Kationer er også kendetegnet som positivt ladede ioner i jorden, som tiltrækkes af de generelt negativt

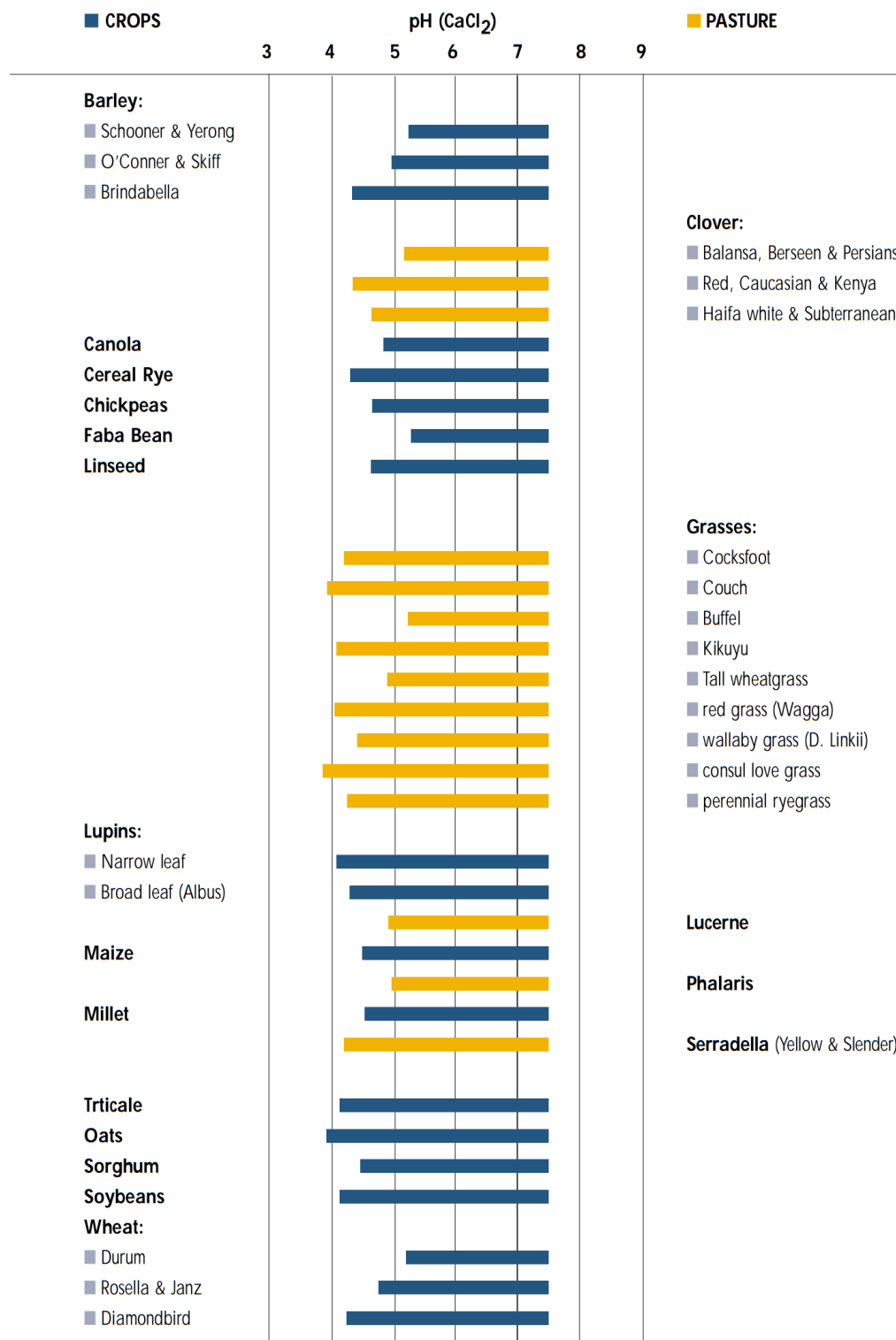
ladede overflader i jordens partikler. Heraf vil forskellige mineraler og næringssalte have varierende tiltrækningskraft, afhængigt af jordvæskens pH-værdi [Ashman og Puri, 2002, s. 58-59]. En generalisering af næringsstofferne tilgængelighed i jordvæsken for planter er illustreret i figur 4.5



Figur 4.5. Effekten af pH på specifikke næringsstoffer i jorden og hvordan de reagerer i forhold til pH. Tykkere linjer indikerer højere tilgængelighed. Figur fra [NSW Department of Primary Industries, 2000, s. 3].

Surhedsgraden har direkte indflydelse på plantevækst, og forskellige plantearter har specifikke pH-krav for effektiv næringsstopoptagelse, se figur 4.6. En pH-værdi uden for det optimale interval kan hæmme planternes evne til at optage føde, bestående af dens essentielle mineraler og næringssalte, hvilket kan resultere i vækstproblemer og i forbindelse med landbrug kan dette give et reduceret udbytte. De fleste afgrøder kræver en pH-værdi tæt på neutral for optimal vækst, mens visse planter foretrækker en mere sur jord, hvor der er mindre konkurrence for føden. For at optimere plantevæksten er det derfor nødvendigt at overvåge og justere jordens pH-værdi efter behov. Justering kan foretages ved at tilsætte kalk for at hæve pH-værdien i sur jord eller svovl for at sænke pH-værdien i basisk jord

[NSW Department of Primary Industries, 2000; Ashman og Puri, 2002].



Figur 4.6. Toleranceintervaller for mellem planter og pH-niveau. Den generelle konsensus er, at $pH > 7,5$ gør essentielle næringsstoffer utilgængelige, der de kan binde sig til andre ioner i jorden, og derved bliver inerte og uoptagelige. Figur fra [NSW Department of Primary Industries, 2000, s. 2].

I LUCAS-datasættet er der udvalgt kun at måle indholdet af specifikke næringsstoffer

i jorden, herunder de mest afgørende næringsstoffer for plantevækst, i form af makronæringsstofferne nitrogen, fosfor og kalium, som helst skal være til rådighed i store nok mængder for optimal plantevækst. Derudover fokuseres også på mikronæringsstoffet jern, som planter kun skal optage i små mængder, samt aluminium der er toksisk overfor jordens biologi. I de kommende afsnit vil der komme flere detaljer for disse stoffers plantetilgængelighed og påvirkning på plantevækst.

4.7 Elektrisk konduktivitet

Jordens elektriske konduktivitet (EC) er en væsentlig indikator for jordens evne til at lede en elektrisk ladning. EC påvirkes af faktorer som jordfugtighed, mineralindhold, temperatur og mængden af ioner i jorden. Ionerne omfatter opløste salte i jorden, der fungerer som næringsstoffer for planterne, hvoraf de mest almindelige er calcium (Ca^{++}), magnesium (Mg^{++}), kalium (K^+), ammonium (NH_4^+) og hydrogen (H^+) [Atlas Scientific, 2023].

EC kan derfor agere som en proxy eller indikator for koncentrationen af opløselige salte i jorden, der netop kan påvirke næringsstoffetilgængeligheden for planter. Ved at måle EC -niveauet i jorden kan det dermed hjælpe med at vurdere jordens saltindhold og give indsigt i potentielle problemer såsom jordkomprimering og dræning. Dermed påvirker EC også plantesundheden indirekte, og kan derfor fungere som en proxy for symptomer på næringsstofmangel eller saltstress, hvoraf for høje niveauer kan føre til osmotisk stress, mens for lave niveauer kan begrænse tilgængeligheden af essentielle næringsstoffer. [Atlas Scientific, 2023].

Tabel 4.3. Liste af forskellige afgrøder og deres optimale EC-intervaller (målt i $[mS/cm]$) for plantevækst, eller hvor planten bedst kan tilpasse sig uden det hæmmer dens vækst [Gong, 2022].

Crop	Min	Max	Crop	Min	Max
African Violet	1,0	1,2	Lavender	1,0	1,4
Asparagus	1,4	1,8	Leek	1,6	2,0
Anthurium	2,0	2,5	Lettuce Fancy	0,3	0,8
Balm	1,0	1,4	Lettuce Head	0,6	1,4
Banana	1,8	2,2	Melons	1,0	2,2
Basil	1,0	1,4	Mint	1,0	1,4
Beans	1,8	2,5	Mustard/Cress	1,2	2,4
Beetroot	1,4	2,2	Onion	1,8	2,2
Blueberry	1,8	2,0	Parsley	0,8	1,8
Broccoli	1,4	2,2	Passion Fruit	1,6	2,4
Bromeliad	0,8	1,0	Pea	1,4	1,8
Brussel Sprout	1,8	2,4	Pumpkin	1,4	2,4
Cabbage	1,4	2,4	Radish	1,2	2,2
Capsicum	2,0	2,7	Rhubarb	1,6	2,0
Carrot	1,4	2,2	Roses	1,8	2,6
Cauliflower	1,4	2,4	Sage	1,0	1,6
Celery	1,5	2,4	Spinach	1,8	3,5
Chives	1,2	2,2	Silver-beet	1,8	2,4
Cucumber	1,6	2,4	Squash	1,8	2,4
Roses	1,8	2,2	Strawberry	1,8	2,5
Eggplant	1,8	2,2	Thyme	1,2	1,6
Endive	0,8	1,5	Tomato	2,2	2,8
Fennel	1,0	1,4	Turnip/Parsnip	1,8	2,4
Garlic	1,4	1,8	Watercress	0,4	1,8

De værdier, der er præsenteret i tabel 4.3, illustrerer variationen i optimale EC-niveauer for forskellige afgrøder, og er dermed et godt udgangspunkt for at sikre en god plantevækst. Hvis EC-niveauet er for lavt eller for højt, kan det dermed indikere om planten ikke får nok eller får for meget næring. Eksempelvis kræver planter som asparges, porre, gulerødder og løg EC-niveauer i området 1,4 til 2,2 mS/cm for at opnå god vækst, mens afgrøder som roser, tomater og spinat kan klare højere EC-niveauer, op til henholdsvis 2,6, 2,8 og 3,5 mS/cm. Det er dog tydeligt, at ekstremt høje eller lave EC-niveauer generelt har negative effekter for planter [Gong, 2022].

På tværs af alle de i tabellen nævnte afgrøder, giver deres gennemsnitlige minimum og maksimum EC-værdier et interval på mellem 1,39 og 2,05 mS/cm, hvilket indikerer, at dette spænd sandsynligvis er det mest optimale for plantevækst generelt. Derfor antages det, at de fleste planter kan trives inden for dette interval uden at opleve næringsstofmangel eller saltstress [Gong, 2022].

Makronæringsstofferne Nitrogen, fosfor og kalium

Optagelsen af næringsstofferne nitrogen (N), fosfor (P) og kalium (K) eller på engelsk, *potassium* er alle afgørende for jordens sundhed og optimal plantevækst. N, P og K kaldes også makronæringsstoffer, idet de udgør de mest essentielle næringsstoffer for planter fotosyntese, vækst og helbred. Heraf påvirker jordens pH i væsentlig grad næringsstoffernes opløselighed, og dermed deres tilgængelighed for planter [Ashman og Puri, 2002, s. 59-61].

Nitrogen (N) er et af de vigtigste makronæringsstoffer for planter, og udgør ofte den begrænsende faktor for plantevækst. Nitrogen er særlig vigtig for syntesen af aminosyrer, proteiner og nukleinsyrer, hvilke er vigtige for alt liv. I jorden er nitrogen primært planteoptageligt i sine tre uorganiske former: ammonium (NH_4^+), nitrat (NO_3^-), og i mindre grad nitrit (NO_2^-). Ammonium dannes gennem nedbrydning af organisk materiale og er en positivt ladet ion, der kan optages direkte af planternes rødder. På trods af dette er ammonium i sig selv mindre tilgængeligt for planter, og derfor er nitrifikationsprocessen, hvor (NH_4^+) omdannes til nitrat (NO_3^-), afgørende for at øge kvælstoftilgængeligheden. Nitrifikationen udføres af jordbakterier, som trives bedst ved pH-niveauer mellem 6,0 og 7,5 hvor omdannelsen af ammonium til nitrat er mest effektiv. Nitrat (NO_3^-) er den mest optagelige form for kvælstof, da det er meget mobilt i jordvandet og nemmere kan absorberes af planterødderne [Ashman og Puri, 2002; Aciego Pietri og Brookes, 2008, s. 61-86].

Fosfor (P) er essentiel for planter fotosyntese, nitrogenfiksering og udgør en vigtig byggesten for plantevækst. Det er både et komponent i plantens DNA, og dermed dens genetiske arvemateriale, samt ATP der driver plantens energioverførsel og stofskifte. Under mere sure forhold ($\text{pH} < 6,0$) vil fosfat-ionerne i større grad danne uopløselige forbindelser med eksempelvis jern og aluminium og kan fikseres af lerminerallerne i jorden, hvilket yderligere mindsker deres tilgængelighed. Omvendt kan fosfat under basiske forhold danne uopløselige forbindelser med calcium ved $\text{pH} > 7,5$ der også er mindre planteoptagelige. Derfor er fosfor generelt mest planteoptagelig som et fosfat med formen H_2PO_4^- og med en pH mellem 6,0 og 7,5 [Ashman og Puri, 2002, s. 63].

Kalium (K) er et vigtigt grundstof for planter spalteåbninger for indtagelse og frigivelse af CO_2 og O_2 i forbindelse med fotosyntese og respiration, er byggesten i mange enzymer, og gør planter mere resistente mod sygdomme. Generelt er det kun omkring 1-2% af jordens lager af kalium, der er planteoptagelig, men kan som regel udskiftes med fikserede kalium-ioner, dog meget langsomt ved de fleste pH-værdier. Kalium er generelt tilgængeligt over et bredere pH-interval end nitrogen og fosfor, men jordens pH påvirker stadig dets tilgængelighed. Særligt i sure jorde kan plantetilgængeligheden være mangelfuld, ved $\text{pH} < 6$ [Ashman og Puri, 2002, s. 62].

Jern og aluminium

Jernindholdet i jorden kan også spille en afgørende rolle i forhold til jordens sundhed med henblik på dets plantehospitalitet. Som udgangspunkt er det kun oxideret jern, såsom Fe^{2+} og Fe^{3+} , der er planteoptageligt via absorption igennem planterødderne, men netop tilstedeværelsen af Fe -ioner påvirkes af høj grad af jordens pH, samt indholdet af organisk materiale, jordfugtighed og jordtemperaturen [ST Biologicals, 2024], [Mahender et al., 2019, s. 2-4].

Den generelle dynamik mellem tilgængeligt jern og temperatur er, at temperaturen blot påvirker plantens stofskifte, og derved udvikling af rødder. Men temperaturen kan også have et samspil med jordfugtigheden, hvoraf eksempelvis en tør jord har svækket evne til at transportere de nødvendige næringsstoffer og ioner igennem jordsøjlen, og i modsatte ende, ved for høj jordfugtighed, kan ugunstige koncentrationer af jern opnås. Generelt opstår problemer i forhold til tilgængeligt jern i forbindelse med pH -niveauet [ST Biologicals, 2024], [Mahender et al., 2019, s. 2-4].

Problemet der kan opstå i jorden kemiske komposition ved eksempelvis $\text{pH} > 7$, et basisk miljø er, at de frie Fe^{2+} ioner kan binde sig til andre frie ioner i jorden, som blandt andet kan være andre nødvendige næringsstoffer. Det vidnes, at blot ved ét trin over neutral pH kan reducere op til 95% af alt plantetilgængeligt jern [Mahender et al., 2019, s. 2-4]. Disse nye molekulære forbindelser kan typisk være optagelige for planterødderne, og kan samtidigt være svært nedbrydelige. Syrlige forhold i jorden, ved $\text{pH} < 5$ kan medføre til forhøjede koncentrationer af opløst jern, hvilket kan kulminere i giftige koncentrationer for planter, hvor for høje koncentrationer kan medføre til skade af plantevæv, celler, hæmmet rodvækst med videre. Niveauet for hvad der kan klassificeres for sund mængde af tilgængeligt jern for planterødderne, spiller også i større grad sammen med jordens fugtighed, og derfor også jordens hydrauliske ledningsevne (eks. K_{sat}) men også jordens pH [ST Biologicals, 2024], [Mahender et al., 2019, s. 2-4].

I en syrlig jord, $\text{pH} < 5$, der er vandmættet, samtidigt med at der er høj mængde af organisk materiale, øges risikoen for udvikling af anaerobiske forhold i jorden, hvilket videre fører til forøgede koncentrationer af tilgængeligt jern, hvilket i for høje koncentrationer kan være af skadelig karakter for planter. For høje koncentrationer af jern i jorden kan endvidere påvirke tilgængeligheden af fosfor og kalium [ST Biologicals, 2024], [Mahender et al., 2019, s. 2-4].

Som med jern påvirker pH i større grad også tilgængeligheden af aluminiumioner såsom Al^{3+} . Typisk indgår aluminium i molekulære forbindelser såsom aluminiumsilikater som

er inaktive i jorden. Dog nedbrydes de forskellige aluminium-forbindelser ved $pH < 5,5$, under syrlige forhold, hvorved ioner af aluminium som Al^{3+} frigives i jorden, som planter kan optage. Netop aluminium i denne form er giftig for planter, hvor det kan svække plantens evne til at udføre fotosyntese, samt ødelægger klorofyl i plantevæv, planternes evne til at optage nødvendige næringsstoffer med videre [Munyaneza et al., 2024, s. 1-3].

Indsamling af inputparametre til regressionsmodellen

5

I dette projekt vil flere jordparametre fra LUCAS-datasættet blive estimeret ved hjælp af en omfattende regressionsmodel, der bygger på tilgangen beskrevet af Ballabio et al. [2016]. Dette omfatter en model, der vil inddrage en række miljømæssige faktorer og fjernmålte data for at udvikle en robust og pålidelig metode til prædiktion af jordparametre. De multivariate inputdata, der anvendes i modellen, vil bestå af Sentinel-2 satellitbaserede bånd og indekser, klimadata, samt forskellige terrænegenskaber. Ved at kombinere disse forskellige input, forventes det at modellen kan udnytte styrkerne fra hver af de respektive datakilder for at lave mere præcise og pålidelige prædiktioner af jordegenskaberne på tværs af caseområdet. Dette kapitel vil introducere de relevante inputdata og forklare, hvordan de indgår i modellen, med henblik på at skabe en dybere forståelse af de variable, der vil blive brugt til at estimere jordparametrene. Når de forskellige inputdata introduceres i regressionsmodellen, anvendes en induktiv tilgang for at finde sammenhænge mellem de udvalgte inputdata og observationer. Denne tilgang for at undersøge relevansen af hver inputparameter, beskrives og analyseres i afsnit 7.1.

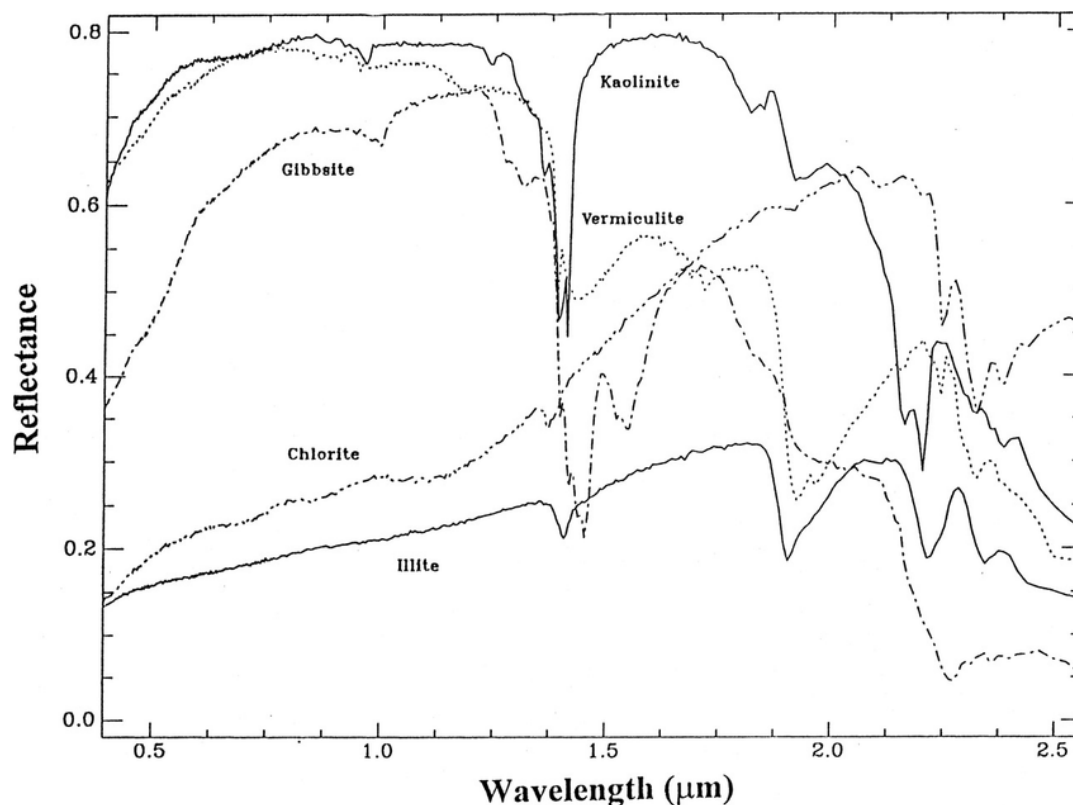
5.1 Spektrale signaturer i jorden

I forbindelse med remote sensing er det vigtigt at kende til den spektrale signatur af de elementer, der ønskes undersøgt, da det er nødvendigt for at studere de mulige forhold mellem jordens egenskaber og hvad der måles med de spektrale bånd. Jordens spektrale signatur er hovedsageligt påvirket af faktorer såsom jordens mineralsammensætning og organiske indhold, fugtighed og tekstur [Gholizadeh et al., 2018].

Særligt jordens indhold af organisk stof og mineralske overflader som ler, der har kromoforer, hvilket er atom-grupper eller molekyler, der absorberer og reflekterer bestemte dele af det elektromagnetiske spektrum. En jordprøves spektrale signatur afhænger derfor særligt af hvilke mineraler og dermed de kromoforer, der er tilstede i jorden og deres indbyrdes sammensættelse. Samtidig kan kromoforerne have overlap i deres spektrale signaturer, hvilket potentielt kan skabe støj i de spektrale målinger og derfor gøre det vanskeligt at vurdere jordens sammensætning [Ben-Dor, 2002, s. 184-185].

Særligt de mineralske partikler i jorden kan variere i sammensætning og struktur, hvoraf leret typisk fremkommer efter partiklers forvitring ned til mineralets mindste bestanddele. Der er fundet visse sammenhæng mellem typen af forvitring og de dominerende lermineraler

i en given jord. Eksempelvis kan mineralet kaolinit, der er en af de mest udbredte lermineraller, opstå gennem kemisk forvitring af mineraler, med et 1:1 forhold af silicium- og aluminiumoxider, såsom feldspat eller bauxit og kvarts [Ben-Dor, 2002, s. 186]. I en jord er lerpartikler med til at holde på vand og næringsstoffer, da disse små partikler har en meget stor overflade, som de kan hænge fast til. Derudover kan ler sammen med organisk stof indgå i jordaggregater, der kan være med til at skabe en god jordstruktur og porefordeling i jorden for plantevækst [Loll og Moldrup, 2000]. I figur 5.1 fremgår forskellige lermineralers spektrale signaturer fra og med det synlige lys, til og med den kortbølgede infrarøde del af det elektromagnetiske spektrum.

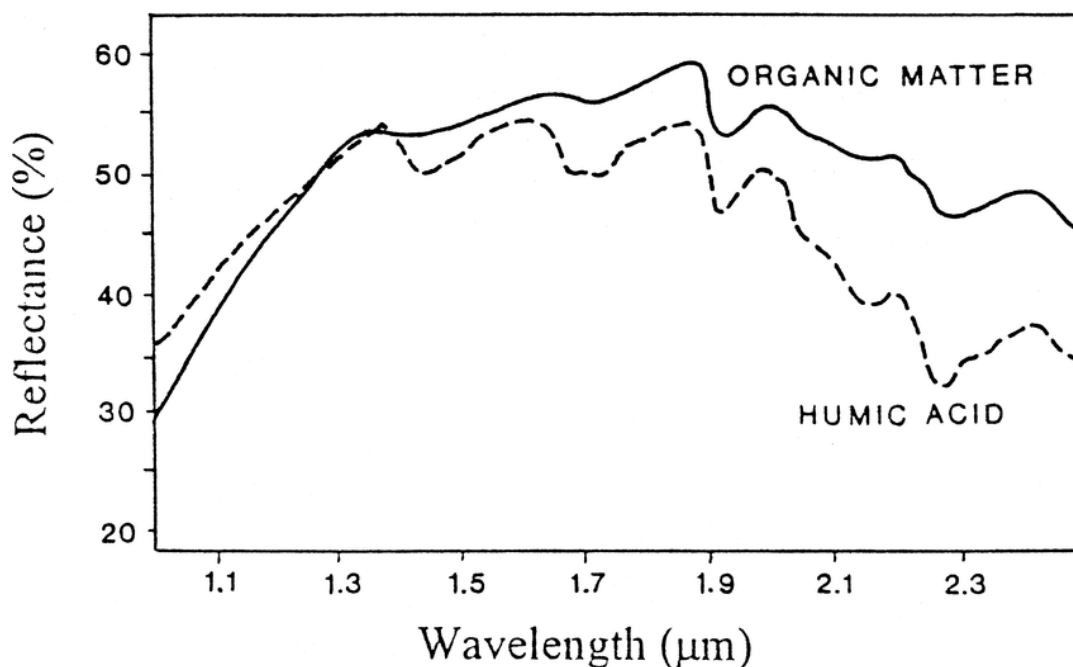


Figur 5.1. Spektrale reflektans signaturer af fem lermineraller [Ben-Dor, 2002, s. 187].

Ud fra figur 5.1 kan det ses, at samtlige lermineraller har en meget lav spektral reflektans i regioner på cirka $1,3-1,4 \mu m$, $1,8-1,9 \mu m$ og $2,2-2,5 \mu m$. Derimod har flere mineraler, særligt kaolinit, en høj reflektans i intervallerne omkring $0,6-1,2 \mu m$ og $1,5-1,7 \mu m$, hvilket passer godt sammen med det røde, nærinfrarøde og kortbølget infrarøde (SWIR-1) spektrale bånd i Landsat og lignende SWIR bånd på Sentinel-2 satellitterne, hvoraf Sentinel-båndene er opstillet i tabel 5.1 [Ben-Dor, 2002, s. 188-189]. Derfor antages det, at disse spektrale bånd også har en vis grad af sammenhæng med lermineraller i samtlige jorde, så længe der ikke er for meget spektralt støj eller uformodede overlap med øvrige spektrale signaturer.

Udover de mineralske partikler, har det organiske indhold (OM) i en jord også en større betydning for flere fysiske og kemiske parametre deri, eksempelvis volumenvægt, vandretentionsevne og jordaggregering. Derfor kan evnen til at estimere OM-indholdet være nyttigt i forhold til netop estimering af disse OM-afhængige jordkarakteristikker,

som er brugbare til vurdering af jordens sundhed. En problematik i forhold til remote sensing af OM-indholdet i jorden, er at OM og humussyre i en jord, dækker et meget bredt spektrum. Nedbrydningsgraden af humus/OM i jorden har dog en betydning for jordens reflektans, hvor højere grad af nedbrydning medføre til lavere reflektion [Ben-Dor et al., 1997, s. 5-7]. I følgende figur 5.2 kan spektrale profiler for OM og humussyre aflæses, som dækkes af båndene NIR og SWIR ($0,9 - 2,5 \mu\text{m}$) [Ben-Dor, 2002, s. 192-195].



Figur 5.2. Spektrale reflektansprofiler af organisk materiale (OM) og humussyre fra en *alfisol* jordart i spektrummet NIR-SWIR ($0,9 - 2,5 \mu\text{m}$) [Ben-Dor, 2002, s. 194].

Som det fremstår af figur 5.2, har både OM og humussyre relativt høj reflektans i intervallet $1,3 - 2,0 \mu\text{m}$. Begge af disse er isoleret fra *alfisol* jordarten, som er rig på jern og aluminium. Studier der har arbejdet med estimering af OM-indholdet i jorde, konkluderer at indholdet heraf viser en korrelation med bølgelængderne ved $0,9 - 1,22 \mu\text{m}$, samt at der for eksempelvis Landsat 5's spektrale bånd 4, 5 og 6, som henholdvist dækker NIR ($0,77 - 0,9 \mu\text{m}$), SWIR-1 ($1,55 - 1,75 \mu\text{m}$), samt TIRS ($10,4 - 12,5 \mu\text{m}$), er observeret signifikante korrelationer med jordens OM-indhold [Ben-Dor et al., 1997, s. 12-14], [Ben-Dor, 2002, s. 192-195]. I forhold estimering af OM-indholdet via remote sensing, kan både tørheden af jordens vegetation og alderen af den organiske jord også have implikationer for hvad der opfanges [Ben-Dor, 2002, s. 192-195].

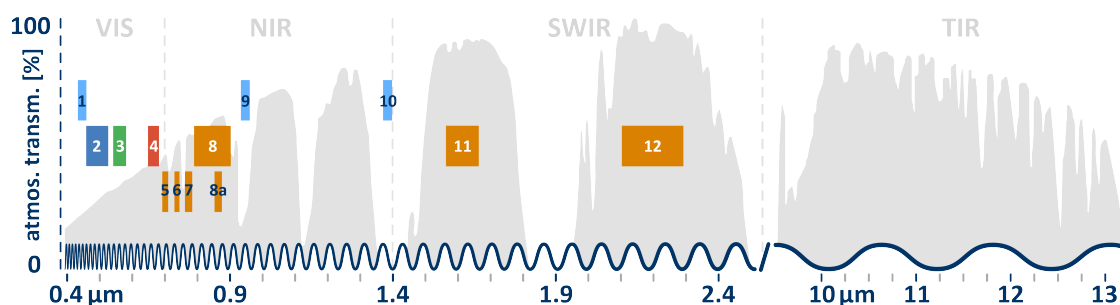
Som præsenteret i figur 5.1 og 5.2, har jordens lerminerale og OM-indholdet nogle særtræk i sine spektrale signaturer. Der kan dog af flere årsager være flere unøjagtigheder i estimeringen af disse med satellit, såsom det overliggende vegetation, antropogene faktorer, og andre eventuelle temporale ændringer i jordoverfladen mellem prøvetagningen og satellitbilledet. Med hensyn til præcisionen vil den metodiske tilgang til remote sensing være, at tage udgangspunkt i jordobservationer ved bare jorde, hvoraf områder med for høj grad af vegetation kan udtages med et NDVI-vegetationsindeks. Derudover kan præcisionen også være påvirket af den iboende imprecision af jordens spektrale signatur og den spektrale generalisering over et større område i forhold til en punktobservation på

grund af Landsat 8/9 og Sentinel-2 satellitternes billedopløsning for de diverse spektrale bånd, se tabel 5.1) [Gallo et al., 2018, s. 7-9].

Nøjagtigheden af blandt andet jordens lerminerale eller OM-indhold, på baggrund af deres spektrale profiler som præsenteret i henholdvist figur 5.1 og 5.2, præges af terrestriske konstituenters såsom det vegetative indeks, antropologisk bebyggelse og temporale ændringer af jordoverfladen fra mellem prøvetagning og fotografering af jordoverfladen med satellit. Forkludrende objekter på terrænoverfladen kan blandt andet udgøres af markarealer, hvorpå der er høj densitet af høstet halm, græs eller andet, som derved præger overfladerefleksionen, og deraf hvad satellitten opfanger. Derfor vil remote sensing som metodisk tilgang til indsamling af jordmæssig data, fungere bedst ved optisk adgang til bar jord. Derudover påvirkes præcisionen også af den iboende vurderingsmæssige imprecision og generalisering af jordens delikate spektra, på grund af satellittens billedopløsning for diverse bånd, som eksempelvis oplyses i tabel 5.1 for Sentinel-2 [Gallo et al., 2018, s. 7-9].

5.2 Sentinel-2 spektrale bånd

Sentinel-2-satellitterne er en central del af ESA's (Den Europæiske Rumorganisations) samarbejde med EU's jordobservationsprogram, Copernicus, som har til formål at indsamle optiske data fra Jordens overflade [Copernicus og ESA, u.d.land12010032]. Sentinel-2-missionen består af to satellitter, Sentinel-2A og Sentinel-2B, der blev opsendt i henholdsvis juni 2015 og marts 2017. Disse satellitter spiller en vigtig rolle i studier af klimaændringer samt i en række miljø- og klimarelaterede analyser. Med sin multispektralscanner kan Sentinel-2 opsamle data i 13 bånd, der dækker det synlige lys, nærinfrarøde (NIR) og kortbølgede infrarøde (SWIR) områder af det elektromagnetiske spektrum, se figur 5.3 [Copernicus Land Monitoring Service, u.d.].



Figur 5.3. Oversigt over Sentinel-2 båndene B1-B12. Den bagvedliggende graf viser andelen af transmitteret lys (%) gennem atmosfæren i forhold til størrelsen på bølgelængderne (μm) i det elektromagnetiske spektrum [Berlin, u.d.].

Hver af disse 13 bånd har unikke egenskaber, der gør det muligt at udforske specifikke områder af atmosfæren og jordens overflade. Disse data er især værdifulde inden for landbrug og skovbrug, hvor de anvendes til overvågning af vegetationssundhed, analyse af jordfugtighed og ændringer i landskabet [Berlin, u.d.; Copernicus og ESA, u.d.land12010032]. En oversigt over de forskellige spektrale bånd, deres bølgelængder, båndbredder og spatiale opløsninger for Sentinel-2 er præsenteret i tabel 5.1.

Tabel 5.1. Detaljer om båndene i Sentinel-2's multispektralscanner [Berlin, u.d.].

Sentinel-2 bånd	Central bølgelængde (nm)	Båndbredde (nm)	Opløsning (m/pixel)
B1: Aerosoler	443	20	60
B2: Blå	490	65	10
B3: Grøn	560	35	10
B4: Rød	665	30	10
B5: Red Edge	705	15	20
B6: Red Edge	740	15	20
B7: Red Edge	783	20	20
B8: NIR	842	115	10
B8A: Narrow NIR	865	20	20
B9: Vanddampe	945	20	60
B10: Cirrus-skyer	1375	30	60
B11: SWIR 1	1610	90	20
B12: SWIR 2	2190	180	20

Sentinel-2's multispektralscanner registrerer kontinuerligt refleksionen af lys fra jordoverfladen, fra det synlige lys (B1-B4) til den kortbølgede infrarøde stråling (B10-B12). De tre 'Red Edge'-bånd (B5-B7), som ligger mellem rød og nærinfrarød, er særligt vigtige for klassificering af vegetation, da de muliggør differentiering af klorofylindholdet i forskellige plantearter. Båndene B1, B9 og B10 bruges primært til atmosfærisk korrektion, idet de registrerer aerosoler, vanddampe og cirrusskyer. Da disse bånd hovedsageligt bruges til korrektion og ikke som direkte input i mange anvendelser, er det svært at anskaffe data fra disse bånd, hvorfor båndene B9 og B10 ikke vil indgå i den aktuelle analyse eller regressionsmodel [Copernicus og ESA, u.d.land12010032].

Dermed vil regressionsmodellen kun anvende data fra de øvrige 11 spektrale bånd i Sentinel-2, herunder B1-B8 samt B8A, B11 og B12.

5.2.1 Indekseringer af Sentinel-2 bånd

De indsamlede data fra de respektive Sentinel-2 spektrale bånd kan indgå i supplerende analyser, herunder indekseringer baseret på båndene. Disse indekser udnytter specifikke båndkombinationer fra Sentinel-2 for at kvantificere forskellige jordparametre. I analysen anvendes der på baggrund af de indsamlede 11 spektrale bånd fra Sentinel-2 i alt 16

beregnete spektrale indekser som supplerende variable, og forventes at forbedre modellens prædiktionssevne.

Jordens optiske egenskaber påvirkes primært af faktorer såsom jordens fugtighed, organisk stofindhold, jordens tekstur og mineralsammensætning [van der Meer et al., 2014; Zolfaghari Nia et al., 2023]. For at kunne udlede disse jordparametre indirekte ved hjælp af remote sensing, blev der udvalgt forskellige grupper af spektrale indekser. Dette inkluderer vegetationsindekser, som også er følsomme over for ændringer i organisk stofindhold, vandindekser som er følsomme over for jordfugtighed, samt indekser relateret til lysstyrke, der er følsomme over for jordtekstur [Gholizadeh et al., 2018, s. 92]. En detaljeret oversigt over de anvendte indekser samt de afledte formler baseret på Sentinel-2 bånd for at beregne dem, fremgår af tabel 5.2.

Tabel 5.2. Oversigt over forskellige remote sensing indekser baseret på Sentinel-2 spektrale bånd. Indekser fra studier af Zolfaghari Nia et al. [2023]; van der Meer et al. [2014]. I studierne er der fundet korrelationer mellem indekserne og forskellige jordparametre. Sentinel-2 båndene B1 til B12 kan aflæses i tabel 5.1.

Remote Sensing Indeks (Fork.)	Formel
NDVI (Normalized difference vegetation index)	$\frac{B8-B4}{B8+B4}$
EVI (Enhanced vegetation index)	$2,5 \cdot \frac{B8-B4}{B8+6 \cdot B4-7,5 \cdot B2+1}$
SAVI (Soil adjusted vegetation index)	$1,5 \cdot \frac{B8-B4}{B8+B4+0,5}$
NDMI (Normalized difference moisture index)	$\frac{B8-B11}{B8+B11}$
COSRI (Combined spectral response index)	$NDVI \cdot \frac{B2+B3}{B8+B4}$
LSWI (Land surface water index)	$\frac{B8-B12}{B8+B12}$
BI (Brightness Index)	$\sqrt{B8^2 + B4^2}$
Clay Index	$B11/B12$
Salinity Index	$\frac{B4-B8}{B3+B8}$
Carbonate Index	$B4/B3$
Gypsum Index	$\frac{B11-B8}{B11+B8}$
Ferric Iron	$B4/B3$
Ferrous Iron	$\frac{B12}{B8} + \frac{B3}{B4}$
Gossan	$B11/B4$
Ferrous Silicate	$B12/B11$
Ferric Oxide	$B11/B8$

Indekserne er særligt relevante, da de kan supplere med en mere præcis og kvantificerbar vurdering af jordens tilstand over store områder og hvor traditionelle jordobservationer i form af jordprøver kan være begrænsede. Indekserne muliggør kontinuerlig overvågning og opdatering af jordparametrene over tid. Dette er især vigtigt i konteksten af klimaændringer, hvor jordens tilstand kan ændre sig hurtigt og uforudsigeligt. For eksempel kan Normalized Difference Vegetation Index (NDVI) og Enhanced Vegetation

Index (EVI) bruges til at overvåge vegetationsdække, hvilket er en vigtig indikator for jordens sundhed. Et fald i vegetationsdække vil derfor kunne identificeres med disse indekser, hvilket kan være en indikation på jordforringelse, og kan udløse behovet for yderligere undersøgelse og afhjælpning af jordsundheden i ramte områder. Ved at udnytte flere af disse spektrale bånd fra Sentinel-2, som er følsomme over for forskellige mineraler og jordkomponenter, forventes det at forbedre modellens evne til nøjagtigt at bestemme og forudsige flere af jordparametrene [Zolfaghari Nia et al., 2023; Castaldi et al., 2023].

5.3 Klimatiske data

Copernicus tilbyder klimadata gennem deres interaktive klimaatlas, et værktøj udviklet af Copernicus Climate Change Service (C3S) for at overvåge klimaændringer. Denne tjeneste administreres af Det Europæiske Center for Mellemlistede Vejrprognoser (ECMWF) på vegne af EU-Kommissionen [Copernicus Climate Change Service (C3S), u.d.]. Klimadataene er indsamlet af Copernicus gennem adskillige metoder, herunder med satellit, observationer, målinger og klimamodeller, der enten danner grundlag for eller supplerer med de data, der er til rådighed på deres klimaatlas.

I forbindelse med projektet anvendes flere forskellige klimatiske data samt arealanvendelser fra Copernicus som datainput i regressionsmodellen. De anvendte data er opført i tabel 5.3, og er baseret på historiske data fra klimaperioden 1991-2020, hvor de tilgængelige data er en blanding af fjernmålte data og målinger ved jordoverfladen.

Tabel 5.3. Oversigt over anvendte klimatiske datainput fra Copernicus' klimadata fra deres interaktive klimaatslas [Copernicus Climate Change Service (C3S), u.d.], samt Copernicus' opgørelse af Dry Matter Productivity og Corine Land Cover [European Commission Directorate-General Joint Research Centre, 2018; Copernicus, 2020]

Klimatiske data	Detaljer
Consecutive dry days	Max antal dage i træk med under 1 mm nedbør
Frost days	Antal dage med under 0°C ved jordoverfladen
Mean temperature	Gennemsnitlig daglig gennemsnitstemperatur målt ved jordoverfladen
Maximum of daily max temperature	Maksimal daglig maksimum temperatur målt ved jordoverfladen
Minimum of daily min temperature	Minimal daglig minimum temperatur målt ved jordoverfladen
Mean of daily accumulated precipitation	Gennemsnitlig dagligt akkumuleret vandækivalent nedbør fra alle faser
Mean of daily accumulated snowfall	Gennemsnitlig dagligt akkumuleret vandækivalent snefald
Mean of daily evaporation	Gennemsnitlig daglig fordampning til atmosfæren, inkluderer sublimation og transpiration
Mean of daily runoff	Gennemsnitlig daglig afstrømning per arealenhed, inkluderer overfladisk og underjordisk afstrømning
Soil shallow moisture content	Jordens overfladenære fugtindhold ned til en dybde på ca. 10 cm
Surface downwelling shortwave radiation	Indkommende kortbølget (solstråling) (W/m^2), der når en horisontal flade ved jordoverfladen
Surface downwelling longwave radiation	Indkommende langbølget (termisk) stråling (W/m^2) ved jordoverfladen (både under skyfri og overskyede forhold)
Dry matter production (DMP)	Vækst af biomasse opgjort i mængden af produceret tørstof i et givet vegetationsdækket område per dag ($DM/ha/dag$)
Corine Land Cover	Klassifikation af arealtyper og arealanvendelser, som der står forklaret og refereres til i tabel 3.3 og figur 3.5

5.4 Topografiske og geomorfologi

Til at kortlægge caseområdets topografi og geomorfologi, anvendes Copernicus Digital Elevation Model (DEM). Copernicus-programmets DEM er en højt detaljeret, global højdemodel, der anvendes til en bred vifte af applikationer som kortlægning, landforvaltning og miljøovervågning. DEM-dataene er tilgængelige i tre forskellige opløsninger: 10 meter, 30 meter og 90 meter, som hver giver forskellige detaljeringsgrader afhængigt af brugerens behov. Af disse er Copernicus' DEM med 30m og 90m opløsning frit tilgængelige, hvoraf der i projektet anvendes DEM-datasættet med en 30 meter opløsning. DEM'erne er opmålt ved hjælp af remote sensing-teknologi, primært fra radar-satellitterne TanDEM-X og TerraSAR-X. Disse satellitter opsamler data ved brug af Interferometrisk Syntetisk Aperturradar (InSAR), som giver meget præcise målinger af jordens overfladehøjde. Denne remote sensing metode gør det muligt at opnå en global dækning og kortlægge terrænhøjder med høj præcision [Copernicus, u.d.].

Både DEM'en med 30 meter opløsning og de afledte geomorfologiske parametre vil kortlægges for hele caseområdet og derefter omskaleres til en 250 meter opløsning. De afledte inputparametre blev beregnet med SAGA-GIS, i form af et plugin til brug i QGIS, af Conrad et al. [2015], et GIS-software som indeholder flere GIS-værktøjer til geovidenskabelige analyser, herunder terrænanalyse. En række af de tilgængelige værktøjer til terrænanalyse i SAGA-GIS bliver anvendt som inputparametre i forbindelse med modellering. De anvendte metoder er opstillet i tabel 5.4.

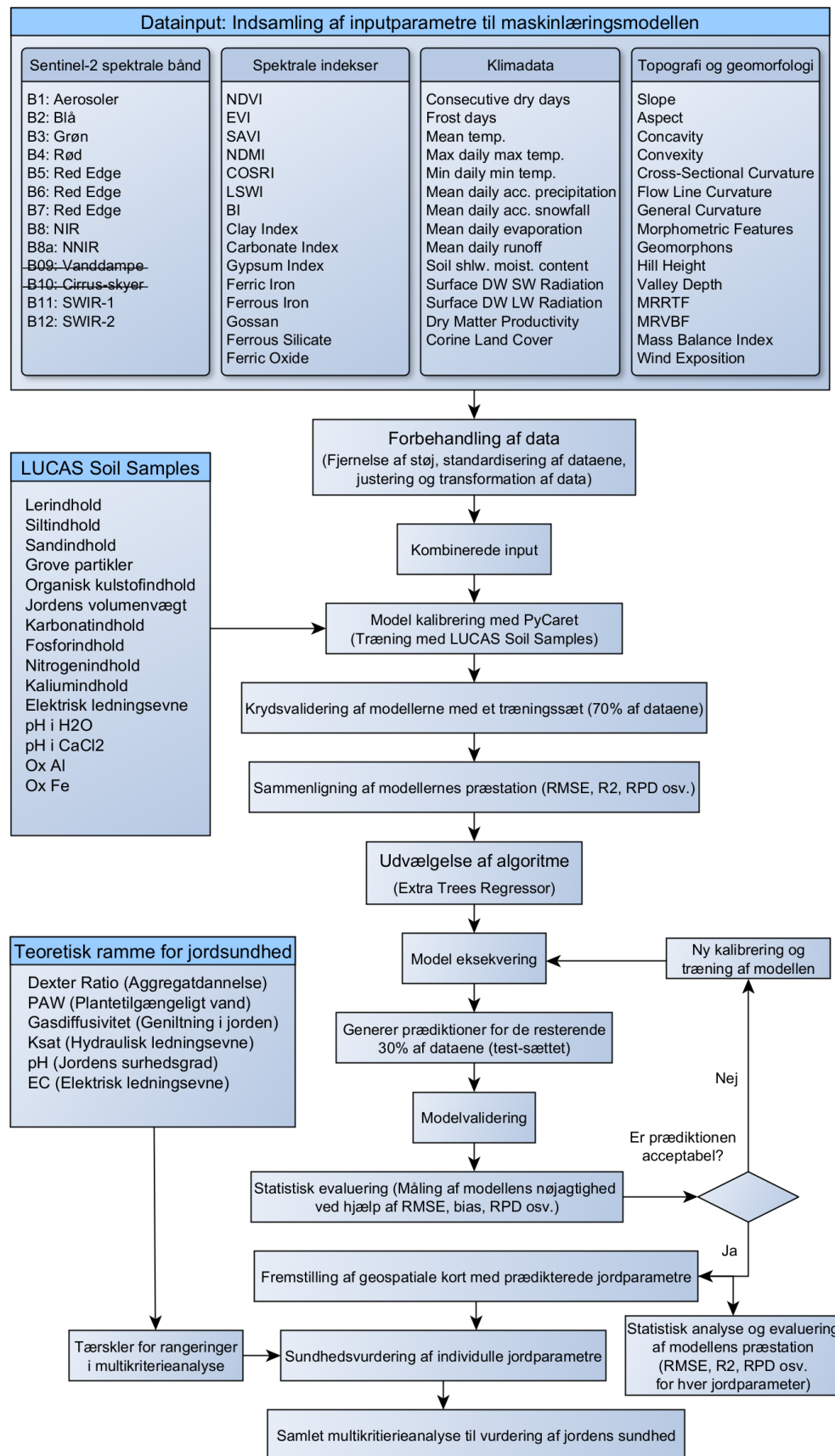
Tabel 5.4. Oversigt over anvendte terrænanalyseværktøjer på DEM'en i SAGA-GIS, der indgår som inputs i regressionsmodellen [Conrad et al., 2024, 2015].

SAGA-GIS metode	Formål og anvendelse
Slope (Hældning)	Beregner terrænets stejlhed, hvilket hjælper med at forstå erosionspotentiale og vandafstrømning.
Aspect (Hældningsretning)	Måler kompasretningen for den stejleste hældning i hvert grid. Kan bruges til at forstå eksponering for sollys og vind, som påvirker plantevækst.
Concavity/Convexity	Identificerer områder med henholdsvis fordybninger og højderygge/bakketoppe i landskabet, der kan hjælpe med at identificere, hvor vand ophober sig.
Cross-Sectional Curvature	Analyserer krumning vinkelret på strømningsretningen. Kan bruges til at vurdere vandstrømmens koncentration.
Flow Line Curvature	Måler krumning langs strømningsretningen, hvilket påvirker vandets hastighed.
General Curvature	Kombinerer flere krumningsmålinger, som er vigtige for at forstå strømningsdynamik i terrænet.
Morphometric Features	Identificerer landskabsformer som dale, højdedrag og rygge for at vurdere terræntypen.
Geomorphons	Klassificerer basale landskabsformer ved hjælp af topografiske data.
Bakkehøjde (Hill Height)	Beregner højden på en bakke i forhold til omgivelserne.
Dybde af dal (Valley Depth)	Måler dybden af dale i forhold til omgivende terræn.
MRRTF og MRVBF (Multi-Resolution Ridge Top Flatness og Multi-Resolution Valley Bottom Flatness)	Identificerer henholdsvis flade toppe på højdedrag og flade dalbunde.
Mass Balance Index	Analyserer balancen mellem erosion og aflejring.
TOBIA Index	Bruges til at vurdere terrænets åbenhed, hvilket kan være relevant i forhold til udsyn, landskabssklassificering og vindeksponering.
Vindeksponering (Wind Exposition)	Analyserer terrænets vindeksponering, hvilket er afgørende for energiproduktion i vindmøller samt forståelse af vindens påvirkning på landskabet.

Maskinlæring og opstilling af regressionsmodellen 6

Som primære og essentielle værktøj til netop at kunne prædiktere for de 15 udvalgte jordparametre som oplyst via LUCAS datasættene, som defineret i tabel 2.1, anvendes biblioteket og værktøjet kaldet PyCaret. Fordelen ved at bruge PyCaret, fremfor at anvende andre eller udvikle egne værktøjer til at automatisere og udføre Machine Learning arbejde, er at PyCaret er et low-level programmeringsbibliotek, hvilket betyder, at man som forbruger af værktøjet, som udgangspunkt ikke behøver at have stort kendskab eller evne til at programmere. Som konkret eksempel, kan PyCaret allerede give resultater blot ved brug af enkelte linjer af kode, fremfor flere tusinde, som kan være tilfældet hvis en selv ønsker at lave et fuldt funktionelt Machine Learnings bibliotek til Python, eller eventuelt ved brug af andre eksisterende biblioteker, som er udviklet og designet til andre brugssager, som hovedsageligt appellerer til mere trænede programmører og videnskabsfolk. PyCaret kan derfor hjælpe med hurtigt at udvikle modeller og resultater, uden det store kendskab til hverken programmering eller opsætning af Machine Learnings programmer [PyCaret, 2024].

For at PyCaret kunne bruges i kontekst af formålet og intentionen med dette projekt, skulle der udvikles et program via programmeringsproget Python. Specifikt blev samlet data fra LUCAS datasættene, dvs de 15 jordparametre, importeret i PyCaret, hvorpå der blev trænet algoritmer til formål at kunne prædiktere for dem. Måden dette blev gjort på, var ved at opsætte en serie af kendte parametre, herunder kategorier såsom klima, geomorfologi og satellit, og sætte dem overfor de observerede jordparametre. Indenfor Machine Learning kendetegnes denne metodiske tilgang også som *supervised learning*, hvorpå Machine Learnings algoritmerne trænes på et set af definerede parametre i træningsdatasættet, hvortil programmet forbinder dem til den observerede værdi, eks. lerindhold, via eksempelvis mønstergenkendelse i datasættet med henblik på at opnå det ønskede produkt [Google, ud.], altså den enkelte jordparameter såsom lerindholdet i jorden. Med andre ord kan det derfor siges, at PyCaret, i kontekst med geografiske informationssystemer (GIS), anvendes ud fra dets funktionalitet til netop at kunne træne algoritmer på et sæt af kendte parametre, til at prædiktere for en ønsket parameter. Heri anvendes modellerne konkret til at prædiktere for på, mellem og udenfor træningspunkterne, ækvivalent til ekstra- og interpolationsværktøj. Kombineret med GIS giver det derfor en højnet evne til både at kunne inter- og ekstrapolere på de supplerede målinger fra LUCAS datasættene med detaljeringsgrad som kun begrænses af opløsningen af de anvendte inputparametre [PyCaret, 2024], [Google, ud.].

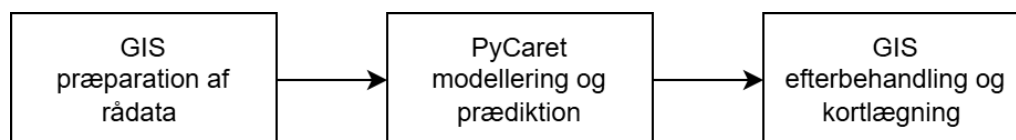


Figur 6.1. Flowchart over maskinlæringsprocessen fra inputparametre til endelige output i form af de geospatiale kort over prædikterede jordparametre, der vil indgå i multikriterieanalysen til vurdering af jordens sundhed.

I projektets maskinlæringskoncept anvendes indsamlede jordparametre som grundlag for at prædiktere jordens tilstand på tværs af store geografiske områder. De rådata, der bruges som input, inkluderer Sentinel-2 spektrale bånd og indekser, klimadata samt information om terræn og geomorfologi. Disse data fungerer som inputparametre, der kalibreres med LUCAS jordprøver, som er essentielle for at træne og validere de maskinlæringsmodeller, der anvendes til prædiktation. Figur 6.1 viser flowchartet over denne proces. Det starter med indsamlingen af inputparametrene, som derefter gennemgår en kalibreringsfase, hvor LUCAS-prøverne anvendes til at tilpasse maskinlæringsmodellerne. Efterfølgende udvælges den bedste model ud fra deres præstation i kalibreringsprocessen, der optimerer modellens præcision i forhold til prædiktationen af jordparametrene. Heraf er Extra Trees Regressor-algoritmen udvalgt, og denne anvendes til at lave de endelige prædikterede jordparametre, der visualiseres som geospatiale kort, og som indgår i multikriterieanalysen, hvor de vurderes i forhold til deres indflydelse på jordens sundhedstilstand.

6.0.1 GIS som det essentielle bindeled mellem tal og visualisering

GIS' funktioner kommer på bane i starten og slutningen af arbejdet, netop som det essentielle bindeled, der indsamler og derefter kortlægger og visualiserer alle de prædikterede jordparametre i en rumlig kontekst. Se eksempelvis figur 6.2, som sætter de tre overordnede processer i kasser. Som det første i processen anvendes GIS som værktøj til at bearbejde de indsamlede satellit, klima, CORINE og geomorfologiske data. Her tilpasses de mange forskellige datalag til det udvalgte samplingsområde, som defineret i figur 2.2. Det vil sige, at lagene eksempelvis klippes til det definerede område, eller behandles i forhold til blandt andet at beregne de flere forskellige undersøgte geomorfologiske parametre, som anvist i tabel 5.4 via algoritmer og funktioner, der er tilgængelige via SAGA-værktøjspakken som anvendt i GIS-miljøet.



Figur 6.2. Flowchart der har til formål at give indsigt i arbejdsprocessen, hvor alt rådata først bearbejdes i GIS, for derefter at eksportere til CSV til at modellere på, for derefter at genimportere til GIS og lave kortlægning deraf.

Når de forskellige lag er klargjort, samples der for hvert af dem, for hvert af de eksisterende og anvendte LUCAS jordprøvepunkter i samplingsområdet. Dette samlede datasæt har til funktion at blive anvendt til at træne og udvikle den enkelte model til prædiktation af den enkelte jordparameter. Når lagene er samlet eksporteres punktdatasættene som CSV-fil (*comma-separated values*), således det kan læses af PyCaret.

I lighed med studiet af Gholizadeh et al. [2018] opdeles jordprøverne fra LUCAS-datasættet i dette projekt i et træningssæt, som anvendes til at træne modellen, og et testsæt til at evaluere modellens præstation. Til denne proces udvælges 70% af datasættets punkter til træning af modellen, hvor krydsvalidering bruges til at vurdere modellens præcision. De resterende 30% af datasættet anvendes herefter som et uafhængigt testsæt, der bruges til at validere den trænedes models evne til at prædiktere de relevante jordparametre, hvilket

hjælper med at vurdere og kalibrere modellens generelle præstation og dermed sikre en pålidelig og robust model til prædiktion af jordparametrene tilgængelige i jordprøverne.

Simultant samples de forskellige datalag via et punktdatasæt, der spænder arealet af caseområdet, som defineret i figur 2.3. Dette datasæt, består af de omtrentligt 10.270.328 pixels, svarende til punkter, fungerer som det egentlige og endelige test-datasæt. Dette datasæt har også til den vigtige funktion af kunne kortlægge de prædikterede jordparametre i caseområdet. De træned modeller, som lavet på punktdatasættet med de observerede LUCAS-jordparametre, anvendes til at prædiktere jordparametrene for hele testdatasættet, ved brug af alle de kendte og målte satellit, klima, CORINE og geomorfologiske datalag.

6.0.2 PyCaret som prædiktionsværktøj

Det følgende stykke kode som vist i figur 6.3 viser det grundlæggende programmerede stykke kode, der har til opgave at træne modeller til prædiktering af ønskede jordparametre ud fra kendte inputparametre. Jordparametrene der prædikteres for, og som først præsenteret i tabel 2.1, er indsat i et array navngivet *yArray*, som kan ses på linje 10 i figuren. Herpå anvendes alle inputparametrene som er indsat i et array kaldet *xArray*, og som kan ses på linje 7. Disse inputparametre er blandt andet dem, som er præsenteret i kapitel 5 og CORINE i tabel 3.3. Programmet kører for hver jordparameter som defineret i *yArray* én gang, hvorefter den producerer en CSV-fil på de prædikterede værdier, som derefter importeres til GIS ved anvendelse af kendte feature-ID'er, således de kan georefereres hvor de oprindeligt hører til i det originale punktdatasæt.

I koden indgår fire helt essentielle og derfor nødvendige funktioner, som gør at arbejdet overhovedet kan gennemføres. Disse funktioner er `setup()`, `compare_models()`, `finalize_model()` og `predict_model()`.

I `setup()` angives inputdata og målet for hvad der skal modelleres for, dvs. *y* i dette tilfælde, som kommer af *yArray* som itereres over, og input-datasættet, `dfTrain`, er dataene fra LUCAS punkterne i CSV-format og de samlede lag for blandt andet satellit, klima osv. Netop `setup()` funktionen er imperativ at bruge i dette scenario for at opsætte det nødvendige miljø til processering af data. Funktionen behandler input data igennem hvad der kaldes for *transformation pipeline*, hvilket har til formål at gøre dataene konsistente og checker for, at dataene er korrekt transformeret, således det kan modelleres på. Hvad denne proces gør, er eksempelvis at sørge for, at ændre kategoriserede data til numerisk format, eller normalisering og standardisering numeriske værdier. Samtidigt er det også denne funktion, som splitter datasættet op til eksempelvis 70% og 30% som i projektets model, til henholdvist træning af model og til test af modellen i et uafhængigt miljø. Funktionen træner dog ikke en model, hvilket er hvor `compare_models()` funktionen kommer i spil [PyCaret, 2024].

Efter at have opsat et miljø via `setup()`, kan `compare_models()` funktionen bruges til at træne modeller. Denne funktion sammenligner forskellige modeller, herunder *Extra Trees Regressor*, *Linear Regressor* med videre, hvor de blandt andet kan sorteres efter, i dette tilfælde, *RMSE*. Denne funktion er derfor den, der er essentiel, til netop at udvikle en Machine Learning model til anvendelse, samtidigt med at den giver brugeren et

indblik i forskellige statistikker og sammenligninger af modellerne. Den laver modellerne på den angivne træningsstørrelse fra `setup()` funktionen, og tester på den resterende teststørrelse [PyCaret, 2024]. Dernæst er `finalize_model()`, som har den simple funktion af at klargøre den trænedede model, hvilket den gør på hele datasættet, netop for at maksimere mængden af tilgængelig data at lære fra og derved træne på. Med andre ord "gentræner" `finalize_model()` funktionen modellen, med henblik på at højne præcisionen og deraf prædiktionssevnen, således at modellen er klar til at blive anvendt til en reel analyse, hvortil netop `predict_model()` funktionen bliver relevant [PyCaret, 2024]. Funktionen `predict_model()` kommer i spil ved netop at bruge den trænedede model på den definerede y -parameter, på baggrund af alle de observerede satellit, klimamæssige, geomorfologiske data osv., som modellen er trænet på. Heri gøres det på et andet datasæt, som her blot er kaldet `testData`, og som initialiseres på linje 17 i kodelykket herefter [PyCaret, 2024].

```

1  import pandas as pd
2  from pycaret.regression import *
3  from pycaret.regression import *
4  from pycaret.utils import *
5
6  # Independent variables
7  xArray = ['B01', 'B02', 'B03', 'B04', 'B05', 'B06', 'B07', 'B08', 'B8A', 'B11', 'B12', 'NDVI',
            'EVI', 'SAVI', 'NDMI', 'COSRI', 'LSWI', 'BI', 'Clay Index', 'Salinity Index', 'Carbonate
            Index', 'Gypsum Index', 'Ferric Iron Index', 'Ferrous Iron Index', 'Gossan Index',
            'Ferrous Silicates Index', 'Ferric Oxides Index', 'Dry days', 'DMP', 'Frost days', 'Max
            temp', 'Precipitation', 'Snowfall', 'Evaporation', 'Runoff', 'Mean temp', 'Min temp',
            'Soil moisture', 'Longwave', 'Shortwave', 'Code', 'DEM', 'Aspect', 'Slope', 'Concavity',
            'Convexity', 'Cross-Sectional Curvature', 'Flow Line Curvature', 'General Curvature',
            'Geomorphons', 'Hill Height', 'Longitudinal Curvature', 'Mass Balance Index', 'Maximal
            Curvature', 'Minimal Curvature', 'Morphometric Features', 'MRRTF', 'MRVBF', 'Plan
            Curvature', 'Profile Curvature', 'Tangential Curvature', 'Terrain Ruggedness Index',
            'TOBIA Index', 'Topographic Position Index', 'Total Curvature', 'Valley Depth', 'Wind
            Exposition', 'Horizontal Distance to River Network']
8
9  # Dependent variables
10 yArray = ['Clay', 'Silt', 'Sand', 'Coarse', 'OC', 'BD 0-20', 'pH CaCl2', 'pH H2O', 'EC',
            'CaCO3', 'P', 'N', 'K', 'Ox Al', 'Ox Fe']
11
12 print("""Reading training data (measured LUCAS soil parameters with sampled X array predictors
            for the entire sampling area)
            and removing rows with empty cells for chosen X parameters from dataset""")
13 trainingData = pd.read_csv('training_data.csv').dropna(subset=xArray)
14
15 print("Reading test data for the entire case area. No need to re-initialize dataset for every
            y")
16 testData = pd.read_csv('test_data.csv')
17 testData = testData.dropna(subset=xArray)
18
19 # Looping over and predicting for all entries in the yArray
20 for y in yArray:
21     print("Making model and predicting for " + f"{y}")
22     arraysToKeep = [y] + xArray
23
24     print("Preparing training data and removing rows with empty cells for chosen Y parameter")
25     dfTrain = trainingData.dropna(subset=y)
26     dfTrain = dfTrain[arraysToKeep].dropna(subset=xArray).dropna(subset=y)
27
28     print("Initializing the training environment")
29     trainingData = setup(dfTrain, target=y, train_size=0.7)
30
31     print("Sanity check by comparing different regression models on the remaining 30% of the
            training data, and selecting the best based on RMSE value")
32     model = trainingData.compare_models(sort='RMSE', turbo=False, n_select=1)
33
34     print("Finalizing model by modelling algorithm on the entire training dataset")
35     finalizedModel = finalize_model(model)
36
37     print("Predicting " + f"{y}" + " for each row present in the supplied test dataset")
38     predictedValues = trainingData.predict_model(finalizedModel, testData)
39     predictedValues = pd.DataFrame(predictedValues)
40
41     print("Setting up new environment with a column with the existing point fid's and the
            predicted values, in order to be processable within GIS to make a map")
42     testData[f"{y}"] = predictedValues['prediction_label']
43     testData = testData[['fid', f"{y}"]]
44
45     print("Writing to CSV")
46     testData.to_csv('Predicted_' + f"{y}", index=False)
47
48     print("Finished predicting values for " + f"{y}")

```

Figur 6.3. Udsnit af Python kode der bruger PyCaret-biblioteket til at tr ne, modellere og pr diktere jordparametrene som kan afl ses p  linje 10.

6.1 Sammenligning og bestemmelse af regressionsmodeller

I tabel 6.2 og 6.3 er der indsat tabeller der viser statistik for jordparametrene ler, silt, sand og organisk kulstof, med henblik på undersøgelse og sammenligning af maskinlæringsmodeller for mest præcise resultater. Dette er gjort via *PyCaret* i Python, med en træningspulje på hhv. 3585 for ler, silt og sand, samt 6743 for organisk kulstof. Statistikken beskriver de 26 forskellige regressionsmodellers evne til at prædiktere den enkelte jordparameter i en uafhængig test. Afhængigt af hvilken jordparameter der undersøges, ændres teststørrelsen grundet mængden af eksisterende data. For eksempelvis ler, er der i alt 5122 punkter indenfor det totale samplingsområde som defineret i figur 2.2, hvoraf 70% anvendes til træning af modellerne med en tilfældigt udvalgt *random state*, og de resterende 30% anvendes som punkter til at lave en uafhængig test af modellerne. Da denne præliminære undersøgelse er til for at finde ud af, hvilken model der vil være bedst at anvende fremadrettet, er valget af *random state* til pseudo-tilfældig udvælgelse af punkter hovedsageligt irrelevant, og vil kun have marginal påvirkning på de potentielle resultater. Antallet af punkter, dvs. jordprøver, som er foretaget og findes indenfor de to områder, kan læses i følgende tabel 6.1, som giver en oversigt over antallet af jordprøver for hver jordparameter, og hvor mange der er indenfor de to områder, som blev belyst i afsnit 2 med figurene 2.3 og 2.2. Caseområdet er mindre, hvorfor der naturligvis er færre punkter her for de forskellige jordparametre. Det ses dog, at trods udvidelsen af et område til hvad der er det totale samplingsområde, er det ikke fordi, at der tilføjes mange flere jordprøver. Dette vidner igen til, at Iberien er det sted indenfor EU, som er foretaget flest jordprøver i, eller som blev godkendt til at komme med i LUCAS databaserne.

Tabel 6.1. Oversigt over antallet af jordprøver for hver jordparameter og område, hvor områderne kan ses i figurene 2.3 og 2.2, for hhv. caseområdet og det totale samplingsområde, som indeholder jordprøver fra yderligere lande som eksempelvis Italien og Østrig.

Parameter Område	Clay	Silt	Sand	Coarse	OC	BD	pH	CaCl ₂	pH	H ₂ O	EC	CaCO ₃	N	P	K	Ox Al	Ox Fe
Caseområde	2972	2972	2972	2972	5727	1898		5753		5753	5753	4482	5747	3243	5751	776	776
Samplingsområde	5122	5122	5122	5122	9633	2924		9659		9659	9657	7147	9653	5871	9655	1240	1240

Tabel 6.2. Sammenligning af ler og silt med henblik p  pr cisionen af 26 forskellige regressionmodeller. For disse to, s vel som de 13 andre jordparametre der unders ges, fremtr der Extra Trees Regressor som modellen med blandt andet bedste *RMSE* og *R* score.

Model Comparisons	[Clay; Total Size 5122, Training Size 3585]						[Silt; Total Size 5122, Training Size 3585]					
Model	MAE	MSE	RMSE	R	RMSLE	MAPE	MAE	MSE	RMSE	R	RMSLE	MAPE
AdaBoost Regressor	7.937	91.188	9.544	0.196	0.533	0.652	8.592	110.379	10.494	0.208	0.293	0.247
Automatic Relevance Determination	7.549	90.006	9.477	0.208	0.491	0.545	8.297	109.320	10.441	0.216	0.296	0.253
Bayesian Ridge	7.538	90.358	9.497	0.205	0.494	0.549	8.341	111.256	10.531	0.203	0.304	0.257
CatBoost Regressor	5.161	47.519	6.885	0.581	0.364	0.355	6.150	66.035	8.114	0.526	0.234	0.185
Decision Tree Regressor	4.080	66.534	8.146	0.413	0.400	0.266	4.896	89.405	9.425	0.354	0.266	0.141
Dummy Regressor	8.581	114.102	10.672	-0.003	0.551	0.644	9.503	139.679	11.803	-0.001	0.332	0.294
Elastic Net	7.528	89.587	9.456	0.212	0.494	0.549	8.319	110.543	10.497	0.208	0.303	0.256
Extra Trees Regressor	2.968	32.747	5.714	0.712	0.297	0.213	3.411	42.185	6.478	0.697	0.180	0.104
Extreme Gradient Boosting	3.517	36.126	5.997	0.682	0.310	0.241	4.201	49.196	6.994	0.647	0.192	0.124
Gradient Boosting Regressor	6.381	65.661	8.096	0.422	0.430	0.458	7.470	89.714	9.459	0.356	0.272	0.226
Huber Regressor	7.596	94.689	9.721	0.167	0.496	0.526	8.515	120.128	10.945	0.139	0.317	0.275
K Neighbors Regressor	7.520	94.193	9.700	0.167	0.500	0.541	8.589	119.241	10.903	0.146	0.308	0.259
Kernel Ridge	7.432	91.329	9.536	0.195	0.484	0.531	8.274	110.550	10.494	0.208	0.298	0.252
Lasso Least Angle Regression	7.574	90.430	9.500	0.205	0.496	0.554	8.339	111.045	10.520	0.205	0.303	0.257
Lasso Regression	7.574	90.425	9.500	0.205	0.496	0.554	8.339	111.031	10.520	0.205	0.303	0.257
Least Angle Regression	5.77E+16	4.97E+34	9.44E+16	-4.23E+32	23.423	3.68E+15	1.92E+11	7.13E+23	2.86E+11	-4.43E+21	16.565	5.27E+09
Light Gradient Boosting Machine	4.762	42.188	6.487	0.628	0.347	0.333	5.668	58.133	7.613	0.583	0.224	0.171
Linear Regression	7.391	91.276	9.528	0.194	0.484	0.530	8.257	112.045	10.554	0.197	0.293	0.250
MLP Regressor	9.936	170.222	12.966	-0.524	0.676	0.696	11.843	220.207	14.642	-0.574	0.417	0.318
Orthogonal Matching Pursuit	7.744	94.972	9.734	0.165	0.509	0.578	8.724	121.237	10.998	0.131	0.315	0.271
Passive Aggressive Regressor	11.437	210.678	14.296	-0.853	0.863	0.570	11.789	226.249	14.458	-0.606	0.390	0.351
Random Forest Regressor	4.365	38.811	6.223	0.658	0.332	0.310	5.127	51.250	7.146	0.632	0.214	0.155
Random Sample Consensus	1.91E+04	4.68E+11	2.72E+05	-4.97E+09	1.147	797.900	4.17E+04	3.89E+12	6.30E+05	-2.41E+10	0.894	196.917
Ridge Regression	7.420	91.391	9.537	0.194	0.484	0.531	8.245	109.000	10.423	0.219	0.293	0.250
Support Vector Regression	7.765	98.236	9.902	0.136	0.504	0.550	8.780	125.135	11.173	0.103	0.322	0.281
TheilSen Regressor	826.542	2.27E+08	1.02E+04	-1.98E+06	1.243	38.245	6.15E+03	1.77E+10	9.75E+04	-1.20E+08	0.905	564.072

Tabel 6.3. Sammenligning af sand og organisk materiale (OC) med henblik på præcisionen af 26 forskellige regressionmodeller. For disse to, såvel som de 13 andre jordparametre der undersøges, fremtræder Extra Trees Regressor som modellen med blandt andet bedste *RMSE* og *R* score.

Model Comparisons	[Sand; Total Size 5122, Training Size 3585]						[OC; Total Size 9633, Training Size 6743]					
Model	MAE	MSE	RMSE	R	RMSLE	MAPE	MAE	MSE	RMSE	R	RMSLE	MAPE
AdaBoost Regressor	14.724	298.670	17.277	0.068	0.550	0.650	49.332	3106.816	55.605	-0.759	1.315	3.208
Automatic Relevance Determination	13.911	284.651	16.862	0.114	0.508	0.545	18.834	1179.614	34.099	0.368	0.735	0.812
Bayesian Ridge	13.805	284.433	16.850	0.116	0.504	0.539	18.913	1185.605	34.203	0.363	0.743	0.816
CatBoost Regressor	9.655	161.581	12.703	0.496	0.390	0.370	14.338	694.040	26.257	0.618	0.559	0.622
Decision Tree Regressor	7.592	216.552	14.692	0.322	0.433	0.275	14.003	1117.744	33.178	0.370	0.639	0.577
Dummy Regressor	14.827	322.846	17.958	-0.005	0.547	0.607	25.766	1854.516	42.868	-0.003	0.939	1.469
Elastic Net	13.848	284.822	16.865	0.114	0.506	0.542	19.062	1194.105	34.334	0.358	0.755	0.833
Extra Trees Regressor	5.832	118.096	10.853	0.632	0.327	0.230	10.074	528.609	22.859	0.707	0.481	0.463
Extreme Gradient Boosting	6.836	131.576	11.459	0.590	0.347	0.262	11.906	610.206	24.559	0.659	0.532	0.535
Gradient Boosting Regressor	12.124	223.183	14.934	0.304	0.454	0.471	15.764	762.050	27.492	0.580	0.593	0.683
Huber Regressor	14.104	310.231	17.598	0.036	0.512	0.513	18.322	1484.449	38.285	0.203	0.681	0.678
K Neighbors Regressor	13.707	300.203	17.318	0.064	0.514	0.524	20.250	1401.185	37.248	0.241	0.707	0.856
Kernel Ridge	13.785	285.949	16.894	0.111	0.506	0.540	18.813	1183.248	34.130	0.366	0.734	0.809
Lasso Least Angle Regression	13.875	286.236	16.908	0.110	0.508	0.544	18.945	1188.545	34.252	0.362	0.748	0.824
Lasso Regression	13.874	286.200	16.907	0.110	0.508	0.544	18.947	1188.492	34.251	0.362	0.749	0.824
Least Angle Regression	5.04E+14	5.07E+30	7.12E+14	-1.66E+28	16.819	2.03E+13	6.67E+16	1.08E+35	1.04E+17	-4.59E+31	20.941	4.97E+15
Light Gradient Boosting Machine	8.973	145.581	12.057	0.546	0.369	0.343	13.768	668.084	25.757	0.633	0.550	0.604
Linear Regression	13.651	280.448	16.729	0.128	0.502	0.532	18.929	1206.148	34.431	0.355	0.734	0.812
MLP Regressor	16.500	453.259	21.037	-0.407	0.689	0.551	22.866	1449.167	37.852	0.215	0.913	1.052
Orthogonal Matching Pursuit	14.295	304.039	17.426	0.054	0.527	0.571	21.423	1432.593	37.646	0.228	0.830	1.068
Passive Aggressive Regressor	20.152	644.450	24.649	-1.049	0.945	0.619	38.591	3246.909	52.156	-0.653	1.094	2.168
Random Forest Regressor	8.377	135.136	11.615	0.579	0.358	0.328	12.889	631.999	25.022	0.652	0.515	0.555
Random Sample Consensus	5.76E+04	4.76E+12	1.07E+06	-1.49E+10	1.256	866.619	3.58E+04	1.11E+12	8.32E+05	-5.95E+08	1.388	2795.211
Ridge Regression	13.662	278.631	16.678	0.134	0.502	0.533	18.813	1183.520	34.134	0.366	0.733	0.809
Support Vector Regression	14.286	317.509	17.804	0.013	0.522	0.537	19.503	1650.037	40.381	0.113	0.700	0.699
TheilSen Regressor	5.63E+03	1.97E+10	9.09E+04	-6.25E+07	0.896	107.708	2467.359	5.34E+09	4.79E+04	-2.96E+06	1.250	189.072

Ud fra de fire tabeller som vist i tabellerne 6.2 og 6.3 herover for ler, silt, sand og organisk kulstof, samt sammenligningsanalyser af de 26 forskellige regressionsmodeller, lavet på de 13 andre jordparametre, ses det at modellen *Extra Trees Regressor* (ET) opnår bedste resultater. Da det ønskes at opnå de bedste resultater, bliver det derfor ET-modellen som anvendes fremadrettet for de fremtidige prædiktioner og modelleringer. Da der for disse sammenligninger kun er anvendt en sampling størrelse på 70% af den totale pulje for hver parameter, kan det forventes at resultaterne for eksempelvis *RMSE* og *R* enten kan blive bedre eller værre, afhængigt af den *random state*, som bruges som *seed* til udvælgelse af pseudo-tilfældige punkter til træning af modellerne. Generelt for sammenligningerne mellem de forskellige regressionsmodeller for de 15 udvalgte jordparametre, er det typisk *ensemble* (grupperinger) af *decision tree*-modeller, som giver bedste resultater.

6.1.1 Decision Trees - Regressionsmodeller

Da de bedste modeller jf. tabel 6.2 og 6.3 typisk er af typen *decision trees*, anses de som hovedsageligt værende de mest relevante at beskrive. Af de bedste tre modeller gengår modellerne *Extra Trees Regressor*, *Extreme Gradient Boosting* og *Random Forest Regressor* ofte.

Et *decision tree*, som er den overordnede kategori de tilhører, skal anses som værende af en træformet struktur. Det vil sige, at den laver prædiktioner ved at kontrollere og checke for features der indgår i modellen. En såkaldt *feature* kan eksempelvis være en inputparameter som et satellitbånd. Ved hver intern node i strukturen, inspicerer en funktion og dirigerer processen videre til et af dens *sub-trees*, afhængigt af en features værdi. For enden af et decision tree findes der blade (leaves), som repræsenterer prædiktionen eller outputtet for et specifikt datasæt. Disse blade udgør slutpunkterne i træet og indeholder delmængder af træningsdata. Outputtet fra hvert blad bestemmes af enten majoritetsklassen eller gennemsnitsværdien for den pågældende delmængde. Enkelte træer kan dog være ustabile og unøjagtige afhængigt af datatypen. Generelt er *decision trees* ikke-lineære modeller, hvor de rekursivt opdeler data i subsets baseret på værdierne af features [Scikit-learn developers, 2024], [Simic, 2024].

Random Forest Regressor er grupper (*ensembles*) af *decision trees*, hvor flere træer trænes ad gangen. Disse træer trænes uafhængigt af hinanden, på forskellige subsets af instancer (*random states*) og features, som indgår i dataene. Ved anvendelse af flere træer til prædiktion, kan netop præcisionen højnes. Lignende er *Extra Trees Regressor*-modellen, som arbejder ved at have en meget højere randomisering i dets træ. Dette gøres ved, at den tilfældigt udvælger feature-opdelinger og accept-tærskler for hver node i træet, uden at optimere for kriterier. Dette gør også, at denne model er mindre præget af mønstre der kan forekomme af datasættet og har fordel i at have en lavere varians. Samtidigt har modellen en fordel i typisk at have en lavere bias i dets prædiktioner. Som den tredje i listen er *Extreme Gradient Boosting*. Denne er hovedsageligt ens med de andre *decision tree*-modeller, men denne skiller sig ud i, at den træner sekventielle træer, til formål at rette eventuelle fejl fra tidligere for hvert stadie i træet. Denne model er også anderledes fra eksempelvis *Random Forest Regressor*, eftersom denne i modsætning ikke opererer og kreerer træer parallelt, hvilket netop udgør den sekventielle udvikling af modellen, der gør at den kan rette sine egne fejl. En ulempe ved *Extreme Gradient Boosting*-modellen kan dog være, at ved netop at rette sine egne fejl, kan risikere at *overfitte*, hvis det anvendte data er støjet, og kan derfor komme til at modellere videre på eventuelle fejltagelser [Scikit-learn developers, 2024], [Simic, 2024].

6.1.2 Potentielle og ubestemte afvigelsesstørrelser af prædiktionsresultater

Herefter er der lavet en dybere undersøgelse på at finde den optimale *random state* til udvælgelse af punkter til træning og sammenligning ved brug af ET-modellen. Denne undersøgelse, ved sammenligning af flere forskellige *random states* til udvælgelse af punkter til træning og test af model for hver parameter, kan bidrage til at belyse, hvor stor forskel der kan være mellem resultaterne, alene på baggrund af hvilke punkter der udvælges. Dette kan ses i følgende tabeller 6.4, 6.5 og 6.6, der hhv. viser bedste, værste og forskellen mellem

random states, hvor der er testet for 50 forskellige *random states*. Denne undersøgelse kan også hjælpe med at belyse, hvor meget den enkelte *random state* kan påvirke modellernes prædiktionssevne ud fra en 70% sampling størrelse. Bedst til værst for de forskellige *random states* er sorteret ud fra *RMSE*-scoren. Værdistørrelsen på de forskellige statistikker, som eksempelvis *RMSE* på 5,599 for ler, afspejler enheden de forskellige jordparametre har, som kan aflæses i tabel 2.1. De teststørrelse på *random states* blot er på 50, kan der være chance eller risiko for, at enkelte states kan byde på betydeligt bedre resultater eller meget værre resultater. Dette kan der dog ikke realistisk testes for, da mængden af potentielle *random states* er til dels uendelig, hvorfor dette ikke kan blive en absolut sammenligning, og derfor kun på 50 forskellige i denne sammenhæng.

Tabel 6.4. Sammenligning af bedste *random states* for de 15 forskellige jordparametre ud fra træning med *Extra Trees Regressor*-modellen. Rød farve indikerer bedste relative resultat for R^2 -scoren, imens blå det modsatte.

Best Results [70% training size, ~50 random states]								
Parameter	Enhed	MAE	MSE	RMSE	R^2	RMSLE	MAPE	State
Coarse	[w/w%]	4.496	77.245	8.771	0.680	0.494	0.499	16
Clay	[w/w%]	2.837	31.512	5.599	0.730	0.297	0.217	42
Silt	[w/w%]	3.234	39.560	6.276	0.711	0.179	0.097	14
Sand	[w/w%]	5.287	99.193	9.933	0.687	0.315	0.221	14
BD	[g/cm ³]	0.126	0.036	0.189	0.623	0.094	0.213	31
OC	[g/kg]	9.436	438.296	20.575	0.761	0.476	0.458	40
$CaCO_3$	[g/kg]	77.352	14595.368	120.647	0.596	1.750	15.550	30
EC	[mS/m]	9.244	532.818	22.858	0.422	0.565	0.598	5
$pH\ CaCl_2$	-	0.420	0.417	0.645	0.734	0.097	0.074	16
$pH\ H_2O$	-	0.407	0.382	0.617	0.738	0.085	0.065	16
N	[g/kg]	0.655	1.592	1.253	0.748	0.270	0.341	23
P	[mg/kg]	12.662	467.592	21.543	0.245	0.546	0.540	27
K	[mg/kg]	90.708	27213.878	163.826	0.407	0.595	0.603	27
Ox Al	[mg/kg]	0.436	0.664	0.774	0.309	0.269	0.578	20
Ox Fe	[mg/kg]	0.816	1.935	1.358	0.532	0.351	0.690	30

Tabel 6.5. Sammenligning af v rste *random states* for de 15 forskellige jordparametre ud fra tr ning med *Extra Trees Regressor*-modellen. R d farve indikerer bedste relative resultat for R^2 -scoren, imens bl  det modsatte.

Worst Results [70% training size, ~50 random states]								
Parameter	Enhed	MAE	MSE	RMSE	R^2	RMSLE	MAPE	State
Coarse	[w/w%]	4.976	88.179	9.377	0.622	0.521	0.566	32
Clay	[w/w%]	3.159	36.148	6.004	0.684	0.317	0.235	20
Silt	[w/w%]	3.670	47.538	6.878	0.650	0.193	0.107	10
Sand	[w/w%]	5.974	118.596	10.868	0.630	0.330	0.237	8
BD	[g/cm ³]	0.144	0.079	0.267	0.461	0.113	0.249	35
OC	[g/kg]	10.321	647.115	25.056	0.664	0.483	0.471	34
$CaCO_3$	[g/kg]	80.919	15712.498	125.206	0.577	1.774	16.019	23
EC	[mS/m]	10.142	873.376	28.167	0.285	0.582	0.626	33
$pH\ CaCl_2$	-	0.444	0.456	0.675	0.713	0.101	0.079	4
$pH\ H_2O$	-	0.425	0.410	0.640	0.724	0.088	0.068	4
N	[g/kg]	0.685	2.413	1.511	0.649	0.274	0.340	34
P	[mg/kg]	12.992	594.713	24.261	0.157	0.549	0.537	49
K	[mg/kg]	94.749	46837.895	205.244	0.316	0.593	0.599	10
Ox Al	[mg/kg]	0.573	2.684	1.394	-0.275	0.313	0.668	37
Ox Fe	[mg/kg]	0.972	3.258	1.766	0.404	0.381	0.738	37

Tabel 6.6. Differencen mellem bedste og værste *random states* for de 15 forskellige jordparametre ud fra træning med *Extra Trees Regressor*-modellen. Rød farve indikerer bedste relative resultat for R^2 -scoren, imens blå det modsatte.

Result Difference [70% training size, ~50 random states]							
Parameter	Enhed	MAE	MSE	RMSE	R^2	RMSLE	MAPE
Coarse	[w/w%]	0.480	10.933	0.606	-0.058	0.027	0.067
Clay	[w/w%]	0.323	4.636	0.405	-0.046	0.020	0.018
Silt	[w/w%]	0.437	7.978	0.601	-0.061	0.014	0.009
Sand	[w/w%]	0.687	19.403	0.935	-0.057	0.015	0.016
BD	[g/cm ³]	0.018	0.043	0.078	-0.162	0.019	0.037
OC	[g/kg]	0.886	208.818	4.481	-0.096	0.007	0.013
$CaCO_3$	[g/kg]	3.567	1117.130	4.559	-0.020	0.024	0.469
EC	[mS/m]	0.898	340.558	5.309	-0.137	0.017	0.028
$pH\ CaCl_2$	-	0.024	0.039	0.030	-0.020	0.005	0.005
$pH\ H_2O$	-	0.019	0.028	0.023	-0.013	0.003	0.004
N	[g/kg]	0.030	0.821	0.258	-0.099	0.005	-0.001
P	[mg/kg]	0.331	127.122	2.718	-0.088	0.003	-0.002
K	[mg/kg]	4.041	19624.017	41.418	-0.091	-0.002	-0.004
Ox Al	[mg/kg]	0.138	2.020	0.620	-0.584	0.044	0.090
Ox Fe	[mg/kg]	0.156	1.323	0.408	-0.129	0.031	0.048

Statistikken som lavet på resultaterne af sammenligningerne mellem omtrent 50 forskellige *random states* for de forskellige jordparametre, ved en træningsstørrelse på 70% af den tilgængelige pulje, er R^2 -scoren gennemsnitligt $-0,111$ point lavere end for de bedste *random states*. For volumenvægt (BD) ses det dog, at R^2 for den værste er bedre end for den bedste *random state* værdi. Dette ses da resultaterne sorteres i forhold til $RMSE$, hvorfor $RMSE$ stadig er lavere for den bedste *random state* trods den værre R^2 -score. For parametrene ler, silt, sand og coarse, er forskellen på $RMSE$ mellem bedst og værst alle under 1 procentpoint, hvilket kan være en indikator for, at der er god geografisk fordeling og repræsentation af jordprøver fra LUCAS for disse, hvor uanset den pseudo-tilfældige udvælgelse af punkter, er der god nok miljømæssig repræsentation iblandt punkterne til, at resultaterne ikke varierer kraftigere fra bedst til værst. Dette ser dog anderledes ud for organisk kulstof, da der her er estimeret en potentiel R^2 forskel på $-0,096$, og en $RMSE$ 4,481 g/kg værre. Disse forskelle indikerer, at der for visse jordparametre kan findes mere optimale metoder til punktudvælgelse i forbindelse med modeltræning, mens andre jordparametre, såsom ler, generelt er godt repræsenteret i caseområdet, uafhængigt af hvordan de 70% træningspunkter udvælges.

Statistisk analyse af regressionsmodellen 7

7.1 Interkorrelationsmatricer

I tabel 7.1 vises en korrelationsmatrice baseret på korrelationskoefficienten r , som måler styrken og retningen af den lineære sammenhæng mellem to variable. Korrelationskoefficienten varierer mellem -1 og 1, hvor værdier tæt på 1 indikerer en stærk positiv sammenhæng mellem parametrene, hvor hvis én parameter stiger, stiger den anden, mens værdier tæt på -1 indikerer en stærk negativ sammenhæng mellem parametrene, dvs. hvis en parameter stiger, falder den anden. Derimod vil værdier omkring 0 være et tegn på, at der er en svag eller ingen lineær sammenhæng.

For samtlige korrelationsmatricer går en diagonal linje fra cellen i øverste venstre hjørne til cellen i nederste højre hjørne, som repræsenterer korrelationen mellem en parameter og sig selv. Da en parameter altid er perfekt korreleret med sig selv, vil værdierne langs diagonalen altid være 1. Dette afhænger af om det er en korrelation mellem observerede versus observerede værdier, prædikterede versus prædikterede værdier, eller mellem observerede og prædikterede værdier.

Cellerne uden for diagonalen viser korrelationerne mellem forskellige parametre, hvilket giver indblik i deres indbyrdes relationer. Dette kan være særligt nyttigt for at forstå, hvordan de parametre som fremtræder i en jord, kan påvirke hinanden, såsom hvordan jordparametre som ler, silt og organisk kulstof (OC) påvirker hinanden i og mellem de observerede og prædikterede datasæt. I tabel 7.1 kan det eksempelvis aflæses, at sandindholdet i de observerede datasæt er negativt korreleret med både ler og silt med korrelationskoefficienterne på henholdsvis $r = -0,803$ og $r = -0,828$, mens organisk kulstof (OC) er positivt korreleret med nitrogen (N) med $r = 0,868$.

En vigtig faktor i interkorrelationsanalyse, er *multikollinearitet*, som opstår, når to eller flere parametre er stærkt korrelerede med hinanden. Dette kan føre til problemer i modellen, da det bliver svært at adskille deres individuelle effekter på prædiktionen. I lineære regressioner kan multikollinearitet medføre, at modellen bliver mindre robust eller *overfitted*, hvis den tilpasser sig for meget til specifikke mønstre i træningsdataene, og dermed kan det påvirke modellens generaliserbarhed i forhold til nye data. Derudover kan stærkt korrelerede parametre øge modellens kompleksitet uden at tilføre ekstra information [Frost, 2023].

Denne effekt, som multikollinearitet kan medføre, kan dog mildnes afhængigt af den anvendte regression, såsom ridge regression eller ved brug af en algoritme [Schreiber-

Gregory, 2017]. Heraf er Extra Trees Regressor en algoritme, der med sin ensemblemetode af flere beslutningstræer, er med til at gøre multikollinearitet mindre problematisk. Algoritmen arbejder med tilfældige splits og træer afhængigt af træningsdataene, hvilket kompenserer for multikollineariteten ved at fordele beslutningerne på tværs af mange træer, og dermed er metoden mindre følsom over for stærkt korrelerede parametre, og mildner effekten af multikollinearitet, selvom det stadig kan påvirke ens endelige resultater [Mohammed og Kora, 2023]. Omvendt kan parametre der har svage korrelationer med andre parametre være særligt værdifulde, da de tilføjer unik information, som kan forbedre modellens evne til at prædiktere mere præcist.

Tabel 7.1. Interkorrelationsanalyse på de observerede jordparametre. Heri ses det eksempelvis, at der er en stærk korrelation mellem *OC* og *N* ($r = 0,868$), mens volumenvægt (*BD*, ρ_b) har en moderat negativ korrelation med organisk kulstofindhold (*OC*) ($r = -0,562$) og som forventet også med nitrogen (*N*) ($r = -0,601$). Begge akser repræsenterer LUCAS jordprøverne.

	Coarse	Clay	Silt	Sand	BD	OC	$CaCO_3$	EC	$pH\ CaCl_2$	$pH\ H_2O$	N	P	K	Ox Al	Ox Fe
Coarse	1.000	-0.035	0.167	-0.084	0.110	0.118	0.053	-0.025	0.015	0.002	0.112	-0.038	0.000	0.063	0.096
Clay	-0.035	1.000	0.332	-0.803	0.054	-0.178	0.373	0.094	0.506	0.495	-0.108	-0.057	0.459	-0.190	-0.235
Silt	0.167	0.332	1.000	-0.828	-0.382	0.344	0.035	0.170	0.073	0.012	0.473	-0.084	0.207	0.057	0.206
Sand	-0.084	-0.803	-0.828	1.000	0.201	-0.111	-0.258	-0.164	-0.348	-0.302	-0.235	0.087	-0.405	0.106	0.054
BD	0.110	0.054	-0.382	0.201	1.000	-0.562	0.009	-0.093	0.274	0.310	-0.601	-0.019	-0.165	-0.221	-0.169
OC	0.118	-0.178	0.344	-0.111	-0.562	1.000	-0.161	0.021	-0.292	-0.328	0.868	0.019	0.058	0.624	0.563
$CaCO_3$	0.053	0.373	0.035	-0.258	0.009	-0.161	1.000	0.117	0.525	0.533	-0.167	0.027	0.172	-0.189	-0.353
EC	-0.025	0.094	0.170	-0.164	-0.093	0.021	0.117	1.000	0.191	0.131	0.067	0.069	0.152	-0.062	-0.055
$pH\ CaCl_2$	0.015	0.506	0.073	-0.348	0.274	-0.292	0.525	0.191	1.000	0.980	-0.301	-0.029	0.444	-0.349	-0.520
$pH\ H_2O$	0.002	0.495	0.012	-0.302	0.310	-0.328	0.533	0.131	0.980	1.000	-0.354	-0.038	0.416	-0.351	-0.535
N	0.112	-0.108	0.473	-0.235	-0.601	0.868	-0.167	0.067	-0.301	-0.354	1.000	0.063	0.132	0.552	0.575
P	-0.038	-0.057	-0.084	0.087	-0.019	0.019	0.027	0.069	-0.029	-0.038	0.063	1.000	0.273	0.095	0.169
K	0.000	0.459	0.207	-0.405	-0.165	0.058	0.172	0.152	0.444	0.416	0.132	0.273	1.000	0.023	-0.060
Ox Al	0.063	-0.190	0.057	0.106	-0.221	0.624	-0.189	-0.062	-0.349	-0.351	0.552	0.095	0.023	1.000	0.670
Ox Fe	0.096	-0.235	0.206	0.054	-0.169	0.563	-0.353	-0.055	-0.520	-0.535	0.575	0.169	-0.060	0.670	1.000

For at opnå en dybere forståelse af resultaterne for de prædikterede jordparametre, er der udviklet en interkorrelationsmatrice for de prædikterede værdier, som præsenteres i tabel 7.2. Generelt viser korrelationerne mellem de prædikterede jordparametre en lignende tendens som mellem de observerede værdier, dog med enkelte undtagelser. Et bemærkelsesværdigt eksempel er korrelationen mellem *Ox Al* og bulk density (*BD*), hvor der for de prædikterede værdier er en svag moderat negativ korrelation på $r = -0,568$,

mens der i de observerede data kun er en svag negativ korrelation på $r = -0,221$.

Forskellen mellem de to datasæt kan skyldes flere faktorer. En af de mulige årsager er, at de prædikterede værdier er påvirket af de inputdata, som blev brugt i modelleringen, hvor der kan introduceres nye mønstre eller relationer, der ikke nødvendigvis afspejler de faktiske fysiske processer i jorden. Derfor kan modellen i visse tilfælde overdrive sammenhængen mellem parametrene, og resultere i en mindre grad af forvrængethed eller forstærket korrelation, som ikke er til stede i virkelige data.

Tabel 7.2. Interkorrelationsanalyse på de prædikterede jordparametre. De prædikterede parameter følger hovedsaget samme korrelationstendenser som for de observerede, hvorfor det kan antages, at de prædikterede jordparametre opfører sig på samme måde, som de observerede. Begge akser repræsenterer prædiktioner.

	Coarse	Clay	Silt	Sand	BD	OC	$CaCO_3$	EC	$pH\ CaCl_2$	$pH\ H_2O$	N	P	K	Ox Al	Ox Fe
Coarse	1.000	-0.079	0.174	-0.068	-0.010	0.155	0.098	-0.042	0.061	0.056	0.124	-0.123	0.012	-0.010	0.001
Clay	-0.079	1.000	0.100	-0.673	0.287	-0.379	0.402	0.120	0.506	0.508	-0.328	0.042	0.372	-0.321	-0.403
Silt	0.174	0.100	1.000	-0.750	-0.396	0.463	-0.058	0.063	-0.115	-0.169	0.540	-0.070	0.030	0.282	0.363
Sand	-0.068	-0.673	-0.750	1.000	0.079	-0.080	-0.215	-0.101	-0.239	-0.202	-0.166	0.032	-0.248	0.006	0.005
BD	-0.010	0.287	-0.396	0.079	1.000	-0.651	0.237	-0.031	0.415	0.454	-0.653	0.057	0.035	-0.568	-0.620
OC	0.155	-0.379	0.463	-0.080	-0.651	1.000	-0.240	-0.005	-0.392	-0.433	0.897	-0.070	-0.030	0.597	0.630
$CaCO_3$	0.098	0.402	-0.058	-0.215	0.237	-0.240	1.000	0.186	0.579	0.588	-0.254	0.015	0.287	-0.352	-0.456
EC	-0.042	0.120	0.063	-0.101	-0.031	-0.005	0.186	1.000	0.204	0.152	0.035	0.067	0.163	-0.063	-0.074
$pH\ CaCl_2$	0.061	0.506	-0.115	-0.239	0.415	-0.392	0.579	0.204	1.000	0.982	-0.391	-0.005	0.490	-0.562	-0.660
$pH\ H_2O$	0.056	0.508	-0.169	-0.202	0.454	-0.433	0.588	0.152	0.982	1.000	-0.446	-0.011	0.464	-0.584	-0.689
N	0.124	-0.328	0.540	-0.166	-0.653	0.897	-0.254	0.035	-0.391	-0.446	1.000	-0.014	0.027	0.612	0.667
P	-0.123	0.042	-0.070	0.032	0.057	-0.070	0.015	0.067	-0.005	-0.011	-0.014	1.000	0.201	0.014	0.018
K	0.012	0.372	0.030	-0.248	0.035	-0.030	0.287	0.163	0.490	0.464	0.027	0.201	1.000	-0.215	-0.268
Ox Al	-0.010	-0.321	0.282	0.006	-0.568	0.597	-0.352	-0.063	-0.562	-0.584	0.612	0.014	-0.215	1.000	0.887
Ox Fe	0.001	-0.403	0.363	0.005	-0.620	0.630	-0.456	-0.074	-0.660	-0.689	0.667	0.018	-0.268	0.887	1.000

Derefter kan der laves en interkorrelationsmatrice mellem de observerede og de prædikterede jordparametre, hvilket kan hjælpe med at belyse, hvor godt resultaterne fra modellen de stemmer overens med de tilsvarende jordprøver fra LUCAS-datasættet, fra samme geografiske placering. En sådan undersøgelse kan ses i følgende tabel 7.3, hvori korrelationskoefficienten r er beregnet på baggrund af lineær korrelation mellem observeret og prædikteret jordparameter for samme punkt, samlet ud fra de foretagne jordprøver i forbindelse med LUCAS 2015 og 2018 undersøgelserne.

Tabel 7.3. Interkorrelationsanalyse mellem de observerede og prædikterede jordparametre. Længs diagonalen, hvor der er beregnet r -værdier mellem de respektive jordparametre, er det tydeligt, at der er et stærkt sammenhæng mellem de prædikterede og observerede jordparametre, og at flere er næsten ækvivalente.

	Coarse	Clay	Silt	Sand	BD	OC	$CaCO_3$	EC	$pH\ CaCl_2$	$pH\ H_2O$	N	P	K	Ox Al	Ox Fe
Coarse	0.918	-0.013	0.189	-0.101	0.015	0.127	0.080	-0.013	0.025	0.013	0.122	-0.061	0.009	0.058	0.051
Clay	-0.015	0.953	0.293	-0.786	0.198	-0.213	0.390	0.085	0.498	0.487	-0.156	-0.012	0.444	-0.269	-0.295
Silt	0.181	0.300	0.945	-0.793	-0.301	0.354	0.013	0.146	0.060	-0.001	0.448	-0.080	0.180	0.167	0.207
Sand	-0.105	-0.756	-0.770	0.968	0.073	-0.096	-0.239	-0.142	-0.334	-0.289	-0.190	0.057	-0.378	0.054	0.044
BD	0.074	0.150	-0.355	0.143	0.949	-0.560	0.107	-0.091	0.284	0.312	-0.584	-0.001	-0.134	-0.394	-0.407
OC	0.107	-0.298	0.398	-0.091	-0.527	0.904	-0.194	0.002	-0.310	-0.345	0.808	-0.040	0.018	0.477	0.501
$CaCO_3$	0.070	0.365	-0.020	-0.236	0.167	-0.193	0.956	0.145	0.537	0.540	-0.198	0.026	0.220	-0.300	-0.391
EC	-0.047	0.107	0.068	-0.102	-0.027	-0.002	0.152	0.966	0.190	0.137	0.043	0.052	0.152	-0.065	-0.074
$pH\ CaCl_2$	0.062	0.482	-0.099	-0.239	0.380	-0.359	0.545	0.202	0.981	0.961	-0.358	-0.015	0.476	-0.529	-0.621
$pH\ H_2O$	0.058	0.482	-0.150	-0.204	0.416	-0.398	0.556	0.145	0.966	0.980	-0.412	-0.022	0.451	-0.548	-0.647
N	0.084	-0.255	0.495	-0.188	-0.555	0.823	-0.213	0.043	-0.322	-0.373	0.931	0.008	0.081	0.511	0.554
P	-0.089	-0.026	-0.037	0.044	0.002	0.000	-0.005	0.065	-0.026	-0.035	0.044	0.887	0.246	0.012	0.030
K	0.001	0.349	0.037	-0.251	0.020	-0.014	0.230	0.144	0.444	0.416	0.048	0.181	0.949	-0.200	-0.246
Ox Al	-0.028	-0.165	0.161	0.004	-0.278	0.620	-0.224	-0.070	-0.377	-0.385	0.558	0.070	-0.005	0.981	0.692
Ox Fe	-0.083	-0.248	0.262	0.011	-0.260	0.585	-0.378	-0.060	-0.550	-0.567	0.601	0.142	-0.088	0.687	0.978

Som det fremgår af resultaterne i tabel 7.3, er der generelt en høj grad af overensstemmelse mellem de prædikterede og observerede jordparametre, hvilket kan aflæses langs diagonalen, og baseret på den generelt høje lineære korrelationskoefficient. Ud af de 15 jordparametre har 8 en korrelationskoefficient på over 0,95, mens seks parametre ligger i intervallet mellem 0,90 og 0,95. Den eneste parameter, der har en korrelationskoefficient under 0,90, er fosfor, som stadig viser en stærk sammenhæng med en værdi på $r = 0,887$. Der er dog ingen af parametrene, der viser en perfekt korrelation ($r = 1$), hvilket kan forventes at skyldes en række faktorer, herunder den rumlige opløsning af de anvendte raster til at udvikle modeller, tidsforskelle i dataindsamlingen, samt potentielle unøjagtigheder og variationer i både de observerede data og prædiktionsprocessen.

7.1.1 For inputparametrene

Som en yderligere undersøgelse af de inputparametre, der indgår i prædiktation af de forskellige jordparametre via Machine Learning-regressionsmodellering, er der udført interkorrelationsanalyser ved hjælp af korrelationskoefficienten r . Ved at analysere

korrelationerne mellem inputparametrene, kan det hjælpe med at afdække, hvordan de enkelte parametre relaterer sig til hinanden. Dette kan hjælpe med at identificere redundante parametre, der potentielt kan udelades fra modelleringen, hvis de viser sig at være for stærkt korrelerede med hinanden. Disse parametre vil højst sandsynligt indeholde overlappende information, hvilket kan reducere deres individuelle vægt i forbindelse med prædiktionen af jordparametrene, hvilket kan føre til mindre præcise resultater.

Derudover kan arbejdet med færre inputparametre også medhjælpe til at gøre arbejdet nemmere og hurtigere at gennemføre, eftersom det vil gøre den mindre kompleks og dermed forkortes mængden af tid, der kræves til forarbejdning og modellering.

Tabel 7.4. Interkorrelationsanalyse på inputparametrene af satellitdata-kategorien. Heri indgår både de individuelle bånd samt de forskellige anvendte indekser som beregnet via båndene fra Sentinel 2. Hvordan de forskellige indekser er beregnet kan læses i tabel 5.2.

	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B11	B12	NDVI	EVI	SAVI	NDMI	COSRI	LSWI	BI	Clay Index	Salinity Index	Carbonate Index	Gypsum Index	Ferric Iron Index	Ferrous Iron Index	Gossan Index	Ferrous Silicates Index	Ferric Oxides Index
B01	1.000	0.902	0.888	0.841	0.852	0.577	0.411	0.386	0.391	0.781	0.810	-0.678	-0.170	-0.678	-0.524	-0.715	-0.577	0.647	-0.555	0.685	0.563	0.524	0.563	0.406	-0.564	0.556	0.512
B02	0.902	1.000	0.983	0.941	0.914	0.618	0.448	0.448	0.433	0.837	0.853	-0.745	-0.145	-0.745	-0.535	-0.783	-0.583	0.736	-0.551	0.751	0.643	0.535	0.643	0.340	-0.678	0.551	0.516
B03	0.888	0.983	1.000	0.969	0.956	0.689	0.516	0.514	0.496	0.872	0.884	-0.738	-0.168	-0.738	-0.524	-0.798	-0.574	0.795	-0.547	0.744	0.679	0.524	0.679	0.309	-0.678	0.550	0.504
B04	0.841	0.941	0.969	1.000	0.978	0.630	0.442	0.440	0.429	0.917	0.926	-0.818	-0.203	-0.818	-0.622	-0.880	-0.662	0.755	-0.585	0.818	0.831	0.622	0.831	0.322	-0.704	0.580	0.605
B05	0.852	0.914	0.956	0.978	1.000	0.717	0.532	0.518	0.521	0.931	0.924	-0.749	-0.208	-0.749	-0.582	-0.829	-0.614	0.800	-0.528	0.748	0.790	0.582	0.790	0.289	-0.620	0.526	0.564
B06	0.577	0.618	0.689	0.630	0.717	1.000	0.955	0.935	0.941	0.616	0.552	-0.111	-0.074	-0.111	0.053	-0.288	0.026	0.955	-0.015	0.113	0.281	-0.053	0.281	-0.187	-0.224	0.053	-0.051
B07	0.411	0.448	0.516	0.442	0.532	0.955	1.000	0.972	0.989	0.449	0.368	0.116	-0.019	0.116	0.254	-0.079	0.235	0.902	0.172	-0.114	0.092	-0.254	0.092	-0.328	-0.055	-0.124	-0.240
B08	0.386	0.448	0.514	0.440	0.518	0.935	0.972	1.000	0.974	0.448	0.355	0.132	-0.013	0.132	0.273	-0.063	0.261	0.920	0.206	-0.132	0.092	-0.273	0.092	-0.373	-0.050	-0.162	-0.260
B8A	0.391	0.433	0.496	0.429	0.521	0.941	0.989	0.974	1.000	0.461	0.357	0.129	-0.016	0.129	0.240	-0.064	0.243	0.896	0.234	-0.130	0.097	-0.240	0.097	-0.352	-0.010	-0.191	-0.226
B11	0.781	0.837	0.872	0.917	0.931	0.616	0.449	0.448	0.461	1.000	0.966	-0.746	-0.223	-0.746	-0.724	-0.822	-0.717	0.720	-0.480	0.740	0.818	0.724	0.818	0.437	-0.403	0.474	0.724
B12	0.810	0.853	0.884	0.926	0.924	0.552	0.368	0.355	0.357	0.966	1.000	-0.809	-0.223	-0.809	-0.751	-0.868	-0.795	0.659	-0.673	0.806	0.826	0.751	0.826	0.559	-0.496	0.676	0.750
NDVI	-0.678	-0.745	-0.738	-0.818	-0.749	-0.111	0.116	0.132	0.129	-0.746	-0.809	1.000	0.229	1.000	0.889	0.959	0.923	-0.258	0.776	-0.999	-0.874	-0.889	-0.874	-0.629	0.720	-0.746	-0.868
EVI	-0.170	-0.145	-0.168	-0.203	-0.208	-0.074	-0.019	-0.013	-0.016	-0.223	-0.223	0.229	1.000	0.229	0.236	0.236	0.232	-0.095	0.159	-0.225	-0.246	-0.236	-0.246	-0.138	0.121	-0.150	-0.231
SAVI	-0.678	-0.745	-0.738	-0.818	-0.749	-0.111	0.116	0.132	0.129	-0.746	-0.809	1.000	0.229	1.000	0.889	0.959	0.923	-0.258	0.776	-0.999	-0.874	-0.889	-0.874	-0.629	0.720	-0.746	-0.868
NDMI	-0.524	-0.535	-0.524	-0.622	-0.582	0.053	0.254	0.273	0.240	-0.724	-0.751	0.889	0.236	0.889	1.000	0.832	0.975	-0.066	0.648	-0.882	-0.790	-1.000	-0.790	-0.776	0.336	-0.610	-0.990
COSRI	-0.715	-0.783	-0.798	-0.880	-0.829	-0.288	-0.079	-0.063	-0.064	-0.822	-0.868	0.959	0.236	0.959	0.832	1.000	0.868	-0.433	0.714	-0.959	-0.903	-0.832	-0.903	-0.580	0.678	-0.704	-0.830
LSWI	-0.577	-0.583	-0.574	-0.662	-0.614	0.026	0.235	0.261	0.243	-0.717	-0.795	0.923	0.232	0.923	0.975	0.868	1.000	-0.095	0.796	-0.919	-0.796	-0.975	-0.796	-0.826	0.439	-0.769	-0.967
BI	0.647	0.736	0.795	0.755	0.800	0.955	0.902	0.920	0.896	0.720	0.659	-0.258	-0.095	-0.258	-0.066	-0.433	-0.095	1.000	-0.105	0.260	0.423	0.066	0.423	-0.123	-0.342	0.138	0.071
Clay Index	-0.555	-0.551	-0.547	-0.585	-0.528	-0.015	0.172	0.206	0.234	-0.480	-0.673	0.776	0.159	0.776	0.648	0.714	0.796	-0.105	1.000	-0.780	-0.599	-0.648	-0.599	-0.698	0.641	-0.992	-0.628
Salinity Index	0.685	0.751	0.744	0.818	0.748	0.113	-0.114	-0.132	-0.130	0.740	0.806	-0.999	-0.225	-0.999	-0.882	-0.959	-0.919	0.260	-0.780	1.000	0.860	0.882	0.860	0.640	-0.725	0.752	0.863
Carbonate Index	0.563	0.643	0.679	0.831	0.790	0.281	0.092	0.092	0.097	0.818	0.826	-0.874	-0.246	-0.874	-0.790	-0.903	-0.796	0.423	-0.599	0.860	1.000	0.790	1.000	0.372	-0.606	0.574	0.777
Gypsum Index	0.524	0.535	0.524	0.622	0.582	-0.053	-0.254	-0.273	-0.240	0.724	0.751	-0.889	-0.236	-0.889	-1.000	-0.832	-0.975	0.066	-0.648	0.882	0.790	1.000	0.790	0.776	-0.336	0.610	0.990
Ferric Iron Index	0.563	0.643	0.679	0.831	0.790	0.281	0.092	0.092	0.097	0.818	0.826	-0.874	-0.246	-0.874	-0.790	-0.903	-0.796	0.423	-0.599	0.860	1.000	0.790	1.000	0.372	-0.606	0.574	0.777
Ferrous Iron Index	0.406	0.340	0.309	0.322	0.289	-0.187	-0.328	-0.373	-0.352	0.437	0.559	-0.629	-0.138	-0.629	-0.776	-0.580	-0.826	-0.123	-0.698	0.640	0.372	0.776	0.372	1.000	-0.106	0.702	0.809
Gossan Index	-0.564	-0.678	-0.678	-0.704	-0.620	-0.224	-0.055	-0.050	-0.010	-0.403	-0.496	0.720	0.121	0.720	0.336	0.678	0.439	-0.342	0.641	-0.725	-0.606	-0.336	-0.606	-0.106	1.000	-0.619	-0.297
Ferrous Silicates Index	0.556	0.551	0.550	0.580	0.526	0.053	-0.124	-0.162	-0.191	0.474	0.676	-0.746	-0.150	-0.746	-0.610	-0.704	-0.769	0.138	-0.992	0.752	0.574	0.610	0.574	0.702	-0.619	1.000	0.596
Ferric Oxides Index	0.512	0.516	0.504	0.605	0.564	-0.051	-0.240	-0.260	-0.226	0.724	0.750	-0.868	-0.231	-0.868	-0.990	-0.830	-0.967	0.071	-0.628	0.863	0.777	0.990	0.777	0.809	-0.297	0.596	1.000

For de 27 satellitmæssige parametre, kan en interkorrelationsanalyse på dem aflæses i følgende tabel 7.4. Heri kan det blandt andet ses, at flere af indekserne, såsom NDVI og SAVI er perfekt korrelerede med en $r = 1,000$, mens NDMI og Gypsum Index er perfekt negativt korrelerede med en $r = -1,000$. Den høje grad af korrelation mellem inputparametrene gælder også for selve råbåndene, hvortil det eksempelvis kan ses, at Sentinel $B03$ har en stærk sammenhæng med $B02$, eller $B8A$ med $B08$. Dette kan også give mening, eftersom der er en bølgedifference på henholdsvis 70 nm og 23 nm til hver centrale bølgelængde, hvor $B08$ samtidigt overlapper større dele af $B8A$. Mere information

om deres bølgelængder kan aflæses i tabel 5.1.

Ud fra interkorrelationsanalysen i tabel 7.5 mellem klimainputparametrene, kan det ses, at der også her er en høj grad af korrelation mellem flere af parametrene. Eksempelvis antallet af sammenhængende tørkedage til dels følger den maksimale temperature for det givne område. Ligeledes ses der også sammenhæng mellem mængden af fordampning og mængden af nedbør, samt at der tildels også er sammenhæng mellem hvor meget vand der fordamper, og hvor stort den lokale planteproduktion er. Ud fra denne matrice, og hvis der skulle skæres ned på antallet af inputparametre, kunne nedbørsparameteren fungere som proxy for flere af de andre parametre. Dog kan resultaterne af interkorrelationsanalysen ikke nødvendigvis konkludere, at de klimamæssige parametre kan erstatte eller fungere som proxy for CORINE Land Cover/Land Use, da netop denne repræsenterer mere unikke elementer, og er samtidigt ikke af kontinuerte data, da den opererer via indekser.

Tabel 7.5. Interkorrelationsanalyse på inputparametrene af klimadata-kategorien. Heri indgår både klimadata samt CORINE Land Cover/Land Use-koden som input, med deres korrelationer i forhold til øvrige klimadata-input. En oversigt over disse input kan aflæses i tabel 5.3.

	Dry days	DMP	Frost days	Max temp	Precipitation	Snowfall	Evaporation	Runoff	Mean temp	Min temp	Soil moisture	Longwave	Shortwave	Code
Dry days	1.000	-0.492	-0.617	0.713	-0.472	-0.503	-0.459	-0.383	0.817	0.707	-0.461	0.531	0.715	-0.241
DMP	-0.492	1.000	0.235	-0.610	0.731	0.379	0.706	0.642	-0.463	-0.292	0.593	0.011	-0.691	0.409
Frost days	-0.617	0.235	1.000	-0.700	0.142	0.757	0.145	0.083	-0.897	-0.942	0.376	-0.868	-0.284	0.282
Max temp	0.713	-0.610	-0.700	1.000	-0.636	-0.713	-0.616	-0.552	0.866	0.703	-0.624	0.486	0.608	-0.424
Precipitation	-0.472	0.731	0.142	-0.636	1.000	0.451	0.859	0.953	-0.410	-0.230	0.725	0.137	-0.716	0.344
Snowfall	-0.503	0.379	0.757	-0.713	0.451	1.000	0.312	0.454	-0.738	-0.780	0.546	-0.503	-0.374	0.327
Evaporation	-0.459	0.706	0.145	-0.616	0.859	0.312	1.000	0.678	-0.404	-0.186	0.631	0.104	-0.687	0.381
Runoff	-0.383	0.642	0.083	-0.552	0.953	0.454	0.678	1.000	-0.320	-0.177	0.675	0.176	-0.622	0.273
Mean temp	0.817	-0.463	-0.897	0.866	-0.410	-0.738	-0.404	-0.320	1.000	0.902	-0.530	0.736	0.593	-0.348
Min temp	0.707	-0.292	-0.942	0.703	-0.230	-0.780	-0.186	-0.177	0.902	1.000	-0.432	0.822	0.349	-0.294
Soil moisture	-0.461	0.593	0.376	-0.624	0.725	0.546	0.631	0.675	-0.530	-0.432	1.000	-0.110	-0.649	0.264
Longwave	0.531	0.011	-0.868	0.486	0.137	-0.503	0.104	0.176	0.736	0.822	-0.110	1.000	0.018	-0.160
Shortwave	0.715	-0.691	-0.284	0.608	-0.716	-0.374	-0.687	-0.622	0.593	0.349	-0.649	0.018	1.000	-0.211
Code	-0.241	0.409	0.282	-0.424	0.344	0.327	0.381	0.273	-0.348	-0.294	0.264	-0.160	-0.211	1.000

Interkorrelationsanalysen på den forskellige geomorfologiske inputparametre, som kan aflæses i følgende tabel 7.6, står til dels i kontrast til både satellitparametrene og klima/CORINE parametrene, idet netop få geomorfologiske parametre er stærkt

korrelerede. De mest korrelerede ses her *Concavity* og *Convexity*, samt *Terrain Ruggedness Index* og *Slope*, samt andre af parametrene der beskriver hældninger. Her er der derfor enkelte der kunne fraskæres, skulle udviklingen af machine learning modellerne simplificeres. De fleste andre geomorfologiske parametre viser lav korrelation, hvilket tyder på, at de måler forskellige aspekter af terrænet og ikke nødvendigvis korrelerer.

Tabel 7.6. Interkorrelationsanalyse på inputparametrene af geomorfologikategorien, som baseret på DEM-data og diverse beregningsfunktioner, som beskrevet i tabel 5.4.

	DEM	Aspect	Slope	Concavity	Convexity	Cross-Sectional Curvature	Flow Line Curvature	General Curvature	Geomorphons	Hill Height	Longitudinal Curvature	Mass Balance Index	Maximal Curvature	Minimal Curvature	Morphometric Features	MRRTF	MRVBF	Plan Curvature	Profile Curvature	Tangential Curvature	Terrain Ruggedness Index	TOBIA Index	Topographic Position Index	Total Curvature	Valley Depth	Wind Exposition	Horizontal Distance to River Network
DEM	1.000	-0.020	0.310	-0.121	0.124	0.020	-0.033	0.020	0.041	0.125	0.012	0.042	0.227	-0.190	0.193	-0.162	-0.356	0.030	0.009	0.022	0.314	-0.045	0.026	0.189	0.110	0.179	0.157
Aspect	-0.020	1.000	-0.009	0.003	-0.001	0.026	-0.001	0.030	-0.022	-0.001	0.022	0.022	-0.002	0.036	-0.022	0.012	0.016	-0.005	0.023	0.016	-0.008	-0.516	0.022	-0.011	-0.001	0.014	0.012
Slope	0.310	-0.009	1.000	-0.026	0.026	0.004	-0.053	-0.011	0.151	0.215	-0.021	0.022	0.476	-0.458	0.382	-0.452	-0.559	0.016	-0.051	0.054	0.989	-0.168	-0.031	0.429	0.277	0.103	0.072
Concavity	-0.121	0.003	-0.026	1.000	-0.996	-0.177	0.031	-0.210	0.377	-0.165	-0.158	-0.151	-0.161	-0.091	-0.103	-0.230	0.121	-0.086	-0.167	-0.114	-0.028	0.019	-0.243	-0.025	0.148	-0.535	-0.280
Convexity	0.124	-0.001	0.026	-0.996	1.000	0.174	-0.032	0.204	-0.375	0.165	0.152	0.146	0.156	0.089	0.101	0.230	-0.120	0.085	0.162	0.111	0.027	-0.019	0.238	0.021	-0.147	0.532	0.280
Cross-Sectional Curvature	0.020	0.026	0.004	-0.177	0.174	1.000	-0.130	0.792	-0.366	0.327	0.275	0.672	0.483	0.461	0.120	0.097	-0.087	0.373	0.143	0.939	0.004	-0.042	0.737	0.040	-0.224	0.488	0.101
Flow Line Curvature	-0.033	-0.001	-0.053	0.031	-0.032	-0.130	1.000	-0.117	0.018	-0.096	-0.059	-0.094	-0.140	-0.004	-0.030	0.011	0.013	-0.012	-0.038	-0.120	-0.064	-0.015	-0.108	-0.144	0.083	-0.051	-0.019
General Curvature	0.020	0.030	-0.011	-0.210	0.204	0.792	-0.117	1.000	-0.453	0.358	0.805	0.830	0.559	0.629	0.172	0.125	-0.099	0.248	0.698	0.612	-0.012	-0.056	0.913	-0.051	-0.305	0.593	0.111
Geomorphons	0.041	-0.022	0.151	0.377	-0.375	-0.366	0.018	-0.453	1.000	-0.134	-0.357	-0.417	-0.131	-0.399	-0.023	-0.497	-0.171	-0.195	-0.336	-0.273	0.155	0.012	-0.464	0.096	0.174	-0.588	-0.201
Hill Height	0.125	-0.001	0.215	-0.165	0.165	0.327	-0.096	0.358	-0.134	1.000	0.246	0.253	0.502	-0.058	0.249	-0.077	-0.129	0.084	0.219	0.254	0.239	-0.062	0.342	0.383	-0.021	0.322	0.138
Longitudinal Curvature	0.012	0.022	-0.021	-0.158	0.152	0.275	-0.059	0.805	-0.357	0.246	1.000	0.654	0.411	0.543	0.155	0.102	-0.072	0.027	0.960	0.051	-0.024	-0.047	0.721	-0.120	-0.261	0.460	0.076
Mass Balance Index	0.042	0.022	0.022	-0.151	0.146	0.672	-0.094	0.830	-0.417	0.253	0.654	1.000	0.480	0.507	0.123	0.090	-0.110	0.224	0.551	0.543	0.021	-0.058	0.909	-0.019	-0.197	0.496	0.084
Maximal Curvature	0.227	-0.002	0.476	-0.161	0.156	0.483	-0.140	0.559	-0.131	0.502	0.411	0.480	1.000	-0.293	0.464	-0.215	-0.425	0.169	0.307	0.436	0.519	-0.124	0.514	0.695	0.131	0.416	0.118
Minimal Curvature	-0.190	0.036	-0.458	-0.091	0.089	0.461	-0.004	0.629	-0.399	-0.058	0.543	0.507	-0.293	1.000	-0.236	0.345	0.284	0.127	0.517	0.297	-0.501	0.052	0.571	-0.711	-0.474	0.294	0.018
Morphometric Features	0.193	-0.022	0.382	-0.103	0.101	0.120	-0.030	0.172	-0.023	0.249	0.155	0.123	0.464	-0.236	1.000	-0.192	-0.354	0.037	0.119	0.115	0.398	-0.058	0.169	0.325	0.099	0.276	0.099
MRRTF	-0.162	0.012	-0.452	-0.230	0.230	0.097	0.011	0.125	-0.497	-0.077	0.102	0.090	-0.215	0.345	-0.192	1.000	0.492	0.072	0.111	0.057	-0.451	0.046	0.126	-0.217	-0.111	0.217	0.109
MRVBF	-0.356	0.016	-0.559	0.121	-0.120	-0.087	0.013	-0.099	-0.171	-0.129	-0.072	-0.110	-0.425	0.284	-0.354	0.492	1.000	-0.094	-0.039	-0.104	-0.559	0.065	-0.099	-0.268	-0.113	-0.344	-0.238
Plan Curvature	0.030	-0.005	0.016	-0.086	0.085	0.373	-0.012	0.248	-0.195	0.084	0.027	0.224	0.169	0.127	0.037	0.072	-0.094	1.000	-0.049	0.405	0.016	-0.007	0.230	0.031	-0.033	0.249	0.052
Profile Curvature	0.009	0.023	-0.051	-0.167	0.162	0.143	-0.038	0.698	-0.336	0.219	0.960	0.551	0.307	0.517	0.119	0.111	-0.039	-0.049	1.000	-0.134	-0.054	-0.041	0.618	-0.159	-0.258	0.430	0.088
Tangential Curvature	0.022	0.016	0.054	-0.114	0.111	0.939	-0.120	0.612	-0.273	0.254	0.051	0.543	0.436	0.297	0.115	0.057	-0.104	0.405	-0.134	1.000	0.056	-0.035	0.579	0.110	-0.124	0.367	0.062
Terrain Ruggedness Index	0.314	-0.008	0.989	-0.028	0.027	0.004	-0.064	-0.012	0.155	0.239	-0.024	0.021	0.519	-0.501	0.398	-0.451	-0.559	0.016	-0.054	0.056	1.000	-0.166	-0.033	0.492	0.301	0.103	0.071
TOBIA Index	-0.045	-0.516	-0.168	0.019	-0.019	-0.042	-0.015	-0.056	0.012	-0.062	-0.047	-0.058	-0.124	0.052	-0.058	0.046	0.065	-0.007	-0.041	-0.035	-0.166	1.000	-0.052	-0.111	-0.062	-0.051	-0.016
Topographic Position Index	0.026	0.022	-0.031	-0.243	0.238	0.737	-0.108	0.913	-0.464	0.342	0.721	0.909	0.514	0.571	0.169	0.126	-0.099	0.230	0.618	0.579	-0.033	-0.052	1.000	-0.044	-0.329	0.599	0.120
Total Curvature	0.189	-0.011	0.429	-0.025	0.021	0.040	-0.144	-0.051	0.096	0.383	-0.120	-0.019	0.695	-0.711	0.325	-0.217	-0.268	0.031	-0.159	0.110	0.492	-0.111	-0.044	1.000	0.433	0.034	0.029
Valley Depth	0.110	-0.001	0.277	0.148	-0.147	-0.224	0.083	-0.305	0.174	-0.021	-0.261	-0.197	0.131	-0.474	0.099	-0.111	-0.113	-0.033	-0.258	-0.124	0.301	-0.062	-0.329	0.433	1.000	-0.235	-0.113
Wind Exposition	0.179	0.014	0.103	-0.535	0.532	0.488	-0.051	0.593	-0.588	0.322	0.460	0.496	0.416	0.294	0.276	0.217	-0.344	0.249	0.430	0.367	0.103	-0.051	0.599	0.034	-0.235	1.000	0.549
Horizontal Distance to River Network	0.157	0.012	0.072	-0.280	0.280	0.101	-0.019	0.111	-0.201	0.138	0.076	0.084	0.118	0.018	0.099	0.109	-0.238	0.052	0.088	0.062	0.071	-0.016	0.120	0.029	-0.113	0.549	1.000

7.1.2 Mellem input-, observerede og prædikterede parametre

I dette afsnit analyseres interkorrelationsmatricer, der sammenligner observerede værdier fra jordprøver, inputparametre fra remote sensing, samt prædikterede værdier fra modellen

ved brug af Extra Trees Regressor modellen. Matricerne giver et mere dybdegående indblik i sammenhængen mellem de forskellige datasæt og bidrager til en forståelse af, hvordan de observerede data relaterer sig til modellens inputparametre og prædikterede værdier. Korrelationskoefficienten r , der varierer mellem -1 og 1, bruges her til at kvantificere styrken og retningen af de lineære relationer mellem variablerne.

Først analyseres korrelationerne mellem de observerede og prædikterede værdier med input fra Sentinel-2 satellitdata, herunder de målte spektralbånd og afledte indekseringer. Dernæst undersøges sammenhængen mellem de observerede og prædikterede jordparametre og klimatiske faktorer samt arealanvendelse. Til sidst sammenlignes parametrene med geomorfologiske input, der beskriver terrænnændringerne i landskabet.

Ved at analysere korrelationskoefficienterne mellem de forskellige input af data i modellen med observerede eller prædikterede output, er det muligt at vurdere hvilke faktorer der har størst betydning for modellens prædiktionsnøjagtighed, samt i hvor høj grad hvert input bidrager til modellens evne til at forudsige de forskellige jordparametre i jordens overflade.

Tabel 7.7. Interkorrelationsmatricer beregnet for observerede værdier på alle satellitinputparametrene. Værdier er korrelationskoefficienten r .

	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B11	B12	NDVI	EVI	SAVI	NDMI	COSRI	LSWI	BI	Clay Index	Salinity Index	Carbonate Index	Gypsum Index	Ferric Iron Index	Ferrous Iron Index	Gossan Index	Ferrous Silicates Index	Ferric Oxides Index
Coarse	0.027	-0.009	-0.021	-0.046	-0.040	-0.037	-0.044	-0.062	-0.059	-0.070	-0.043	0.002	0.043	0.002	0.028	0.019	0.005	-0.062	-0.053	0.005	-0.082	-0.028	-0.082	0.071	-0.035	0.058	-0.024
Clay	0.334	0.321	0.342	0.344	0.358	0.270	0.207	0.186	0.195	0.330	0.345	-0.263	-0.175	-0.263	-0.216	-0.292	-0.239	0.281	-0.226	0.263	0.281	0.216	0.281	0.147	-0.188	0.224	0.213
Silt	-0.082	-0.102	-0.116	-0.163	-0.148	-0.017	0.015	0.009	0.018	-0.143	-0.150	0.186	0.168	0.186	0.158	0.191	0.161	-0.061	0.129	-0.180	-0.233	-0.158	-0.233	-0.038	0.141	-0.121	-0.154
Sand	-0.147	-0.127	-0.130	-0.102	-0.120	-0.150	-0.133	-0.117	-0.128	-0.105	-0.111	0.039	-0.002	0.039	0.028	0.053	0.040	-0.129	0.052	-0.043	-0.019	-0.028	-0.019	-0.063	0.023	-0.057	-0.029
BD	0.216	0.220	0.248	0.274	0.280	0.124	0.052	0.043	0.035	0.280	0.300	-0.308	-0.062	-0.308	-0.278	-0.321	-0.299	0.150	-0.280	0.306	0.296	0.278	0.296	0.200	-0.229	0.265	0.278
OC	-0.349	-0.355	-0.380	-0.412	-0.409	-0.233	-0.159	-0.152	-0.145	-0.412	-0.414	0.392	0.124	0.392	0.342	0.436	0.355	-0.289	0.289	-0.388	-0.416	-0.342	-0.416	-0.196	0.274	-0.281	-0.341
CaCO ₃	0.474	0.420	0.421	0.382	0.402	0.314	0.242	0.223	0.235	0.381	0.409	-0.301	-0.242	-0.301	-0.254	-0.302	-0.293	0.321	-0.317	0.303	0.220	0.254	0.220	0.242	-0.232	0.310	0.243
EC	0.213	0.182	0.174	0.131	0.143	0.162	0.143	0.128	0.136	0.101	0.101	-0.054	-0.015	-0.054	-0.010	-0.075	-0.021	0.153	-0.040	0.060	0.004	0.010	0.004	0.030	-0.074	0.045	0.007
pH CaCl ₂	0.451	0.423	0.459	0.459	0.479	0.311	0.206	0.176	0.181	0.424	0.468	-0.415	-0.130	-0.415	-0.332	-0.441	-0.386	0.325	-0.434	0.416	0.374	0.332	0.374	0.263	-0.344	0.425	0.310
pH H ₂ O	0.465	0.439	0.476	0.482	0.499	0.302	0.190	0.160	0.164	0.449	0.496	-0.452	-0.139	-0.452	-0.371	-0.475	-0.426	0.323	-0.467	0.453	0.408	0.371	0.408	0.294	-0.364	0.456	0.349
N	-0.356	-0.359	-0.383	-0.421	-0.410	-0.178	-0.091	-0.077	-0.068	-0.386	-0.415	0.436	0.127	0.436	0.358	0.458	0.391	-0.238	0.378	-0.433	-0.433	-0.358	-0.433	-0.226	0.356	-0.363	-0.353
P	-0.042	-0.047	-0.048	-0.059	-0.055	0.021	0.050	0.046	0.051	-0.052	-0.053	0.083	0.046	0.083	0.081	0.077	0.077	0.006	0.047	-0.085	-0.068	-0.081	-0.068	-0.057	0.051	-0.043	-0.081
K	0.165	0.153	0.165	0.166	0.176	0.138	0.107	0.097	0.104	0.164	0.175	-0.134	-0.053	-0.134	-0.114	-0.148	-0.132	0.140	-0.150	0.133	0.143	0.114	0.143	0.080	-0.106	0.145	0.104
Ox Al	-0.236	-0.211	-0.222	-0.194	-0.219	-0.220	-0.197	-0.180	-0.180	-0.193	-0.202	0.124	0.155	0.124	0.072	0.159	0.100	-0.216	0.126	-0.127	-0.092	-0.072	-0.092	-0.076	0.139	-0.125	-0.072
Ox Fe	-0.369	-0.352	-0.375	-0.365	-0.393	-0.312	-0.239	-0.206	-0.225	-0.362	-0.388	0.300	0.345	0.300	0.228	0.330	0.271	-0.309	0.287	-0.301	-0.264	-0.228	-0.264	-0.164	0.277	-0.278	-0.216

Tabel 7.8. Interkorrelationsmatricer beregnet for prædikterede værdier på alle satellitinputparametrene. Værdier er korrelationskoefficienten r . Heri kan det ses, at de prædikterede værdier fra Machine Learning-modellen følger samme tendens som interkorrelationerne for de observerede jordparametre, som kan ses i tabel 7.7.

	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B11	B12	NDVI	EVI	SAVI	NDMI	COSI	LSWI	BI	Clay Index	Salinity Index	Carbonate Index	Gypsum Index	Ferric Iron Index	Ferrous Iron Index	Gossan Index	Ferrous Silicates Index	Ferric Oxides Index
Coarse	0.004	-0.019	-0.027	-0.029	-0.034	-0.101	-0.127	-0.139	-0.144	-0.054	-0.020	-0.063	-0.006	-0.063	-0.050	-0.029	-0.069	-0.112	-0.104	0.068	-0.015	0.050	-0.015	0.115	-0.060	0.101	0.048
Clay	0.495	0.492	0.521	0.528	0.544	0.420	0.328	0.310	0.317	0.510	0.514	-0.394	-0.101	-0.394	-0.306	-0.451	-0.336	0.456	-0.314	0.396	0.416	0.306	0.416	0.197	-0.320	0.314	0.304
Silt	-0.213	-0.236	-0.258	-0.318	-0.306	-0.111	-0.045	-0.050	-0.039	-0.304	-0.314	0.331	0.098	0.331	0.283	0.348	0.297	-0.172	0.263	-0.324	-0.389	-0.283	-0.389	-0.121	0.265	-0.247	-0.280
Sand	-0.148	-0.130	-0.134	-0.101	-0.118	-0.171	-0.159	-0.143	-0.155	-0.101	-0.098	0.016	-0.005	0.016	-0.001	0.037	0.008	-0.150	0.018	-0.022	0.006	0.001	0.006	-0.037	0.014	-0.027	0.002
BD	0.364	0.377	0.410	0.452	0.450	0.186	0.081	0.069	0.067	0.470	0.489	-0.496	-0.161	-0.496	-0.467	-0.520	-0.487	0.243	-0.412	0.491	0.494	0.467	0.494	0.317	-0.317	0.391	0.463
OC	-0.449	-0.467	-0.498	-0.542	-0.538	-0.309	-0.211	-0.200	-0.197	-0.545	-0.546	0.510	0.164	0.510	0.450	0.568	0.465	-0.380	0.371	-0.504	-0.551	-0.450	-0.551	-0.250	0.351	-0.362	-0.447
CaCO ₃	0.581	0.526	0.533	0.491	0.513	0.373	0.273	0.245	0.256	0.466	0.507	-0.389	-0.107	-0.389	-0.311	-0.397	-0.364	0.387	-0.414	0.393	0.303	0.311	0.303	0.286	-0.324	0.409	0.293
EC	0.250	0.219	0.210	0.157	0.170	0.194	0.173	0.154	0.161	0.115	0.118	-0.065	-0.012	-0.065	-0.004	-0.094	-0.020	0.185	-0.054	0.072	0.003	0.004	0.003	0.034	-0.096	0.062	0.001
pH CaCl ₂	0.492	0.466	0.505	0.504	0.525	0.340	0.225	0.193	0.198	0.467	0.514	-0.455	-0.144	-0.455	-0.366	-0.484	-0.424	0.357	-0.473	0.456	0.414	0.366	0.414	0.288	-0.374	0.465	0.342
pH H ₂ O	0.509	0.486	0.525	0.532	0.548	0.331	0.209	0.177	0.182	0.496	0.546	-0.496	-0.155	-0.496	-0.408	-0.521	-0.467	0.357	-0.508	0.497	0.453	0.408	0.453	0.319	-0.396	0.497	0.385
N	-0.434	-0.448	-0.475	-0.524	-0.512	-0.230	-0.121	-0.105	-0.098	-0.487	-0.517	0.533	0.160	0.533	0.444	0.561	0.479	-0.303	0.451	-0.529	-0.540	-0.444	-0.540	-0.270	0.424	-0.433	-0.437
P	0.015	0.016	0.021	0.004	0.016	0.109	0.130	0.129	0.128	0.006	-0.002	0.061	0.019	0.061	0.080	0.030	0.072	0.095	0.040	-0.062	-0.031	-0.080	-0.031	-0.073	0.016	-0.031	-0.078
K	0.198	0.181	0.195	0.191	0.204	0.162	0.126	0.113	0.119	0.187	0.204	-0.155	-0.054	-0.155	-0.130	-0.171	-0.155	0.163	-0.185	0.153	0.157	0.130	0.157	0.101	-0.121	0.182	0.118
Ox Al	-0.398	-0.404	-0.436	-0.456	-0.465	-0.263	-0.172	-0.165	-0.159	-0.472	-0.477	0.438	0.153	0.438	0.411	0.480	0.425	-0.315	0.344	-0.434	-0.439	-0.411	-0.439	-0.266	0.269	-0.331	-0.401
Ox Fe	-0.523	-0.522	-0.562	-0.591	-0.603	-0.348	-0.228	-0.214	-0.211	-0.593	-0.611	0.550	0.188	0.550	0.492	0.598	0.522	-0.409	0.464	-0.546	-0.558	-0.492	-0.558	-0.318	0.378	-0.448	-0.479

Tabel 7.9. Interkorrelationsmatricer beregnet for observerede værdier på alle klima- og CORINE-inputparametrene. Værdier er korrelationskoefficienten r .

	Dry days	DMP	Frost days	Max temp	Precipitation	Snowfall	Evaporation	Runoff	Mean temp	Min temp	Soil moisture	Longwave	Shortwave	Code
Coarse	-0.024	-0.019	0.019	-0.064	0.012	0.031	-0.073	0.055	0.001	-0.021	0.027	-0.023	0.054	0.080
Clay	0.124	-0.331	-0.048	0.196	-0.259	-0.142	-0.273	-0.215	0.139	0.115	-0.180	-0.001	0.165	-0.324
Silt	-0.171	0.172	0.146	-0.204	0.210	0.196	0.147	0.204	-0.182	-0.160	0.244	0.017	-0.220	0.175
Sand	0.035	0.089	-0.064	0.013	0.020	-0.040	0.069	-0.002	0.032	0.033	-0.048	-0.010	0.041	0.083
BD	0.141	-0.326	-0.066	0.251	-0.379	-0.206	-0.313	-0.374	0.157	0.099	-0.267	-0.080	0.237	-0.327
OC	-0.283	0.425	0.169	-0.417	0.455	0.279	0.402	0.427	-0.299	-0.219	0.347	-0.040	-0.307	0.480
$CaCO_3$	0.084	-0.356	-0.066	0.221	-0.354	-0.186	-0.336	-0.312	0.162	0.048	-0.248	-0.057	0.293	-0.208
EC	-0.017	-0.070	-0.059	0.076	-0.075	-0.035	-0.119	-0.038	0.065	0.038	-0.091	0.031	0.022	-0.070
$pH\ CaCl_2$	0.124	-0.516	0.052	0.308	-0.593	-0.162	-0.552	-0.549	0.150	-0.030	-0.388	-0.210	0.408	-0.378
$pH\ H_2O$	0.153	-0.554	0.028	0.345	-0.623	-0.194	-0.578	-0.576	0.180	-0.004	-0.425	-0.200	0.438	-0.393
N	-0.347	0.480	0.215	-0.489	0.539	0.360	0.459	0.515	-0.360	-0.262	0.440	-0.037	-0.407	0.436
P	0.000	0.111	-0.091	0.035	0.021	-0.078	0.017	0.022	0.065	0.077	-0.020	0.091	-0.038	-0.117
K	0.128	-0.189	0.013	0.148	-0.241	-0.069	-0.225	-0.220	0.105	0.009	-0.111	-0.054	0.209	-0.142
Ox Al	-0.084	0.276	-0.052	-0.246	0.570	0.095	0.391	0.657	-0.080	0.019	0.273	0.118	-0.212	0.324
Ox Fe	-0.126	0.454	-0.038	-0.300	0.627	0.198	0.470	0.676	-0.123	0.012	0.349	0.164	-0.348	0.355

Tabel 7.10. Interkorrelationsmatricer beregnet for prædikterede værdier på alle klima- og CORINE-inputparametrene. Værdier er korrelationskoefficienten r . Heri kan det ses, at de prædikterede værdier fra Machine Learning-modellen følger lignende tendenser som interkorrelationerne for de observerede jordparametre, som kan ses i tabel 7.9.

	Dry days	DMP	Frost days	Max temp	Precipitation	Snowfall	Evaporation	Runoff	Mean temp	Min temp	Soil moisture	Longwave	Shortwave	Code
Coarse	-0.065	-0.086	0.243	-0.135	-0.082	0.179	-0.145	-0.040	-0.166	-0.238	0.003	-0.285	0.147	0.180
Clay	0.250	-0.424	-0.206	0.353	-0.333	-0.284	-0.349	-0.275	0.300	0.246	-0.247	0.137	0.222	-0.490
Silt	-0.322	0.341	0.189	-0.378	0.399	0.301	0.347	0.368	-0.300	-0.236	0.404	-0.015	-0.386	0.292
Sand	0.086	0.027	-0.022	0.058	-0.072	-0.036	-0.037	-0.078	0.041	0.025	-0.137	-0.050	0.132	0.091
BD	0.252	-0.509	-0.160	0.431	-0.578	-0.387	-0.466	-0.568	0.286	0.222	-0.457	-0.069	0.395	-0.472
OC	-0.364	0.535	0.249	-0.528	0.572	0.385	0.499	0.539	-0.391	-0.307	0.447	-0.075	-0.389	0.573
CaCO ₃	0.117	-0.434	-0.052	0.259	-0.424	-0.195	-0.422	-0.371	0.182	0.046	-0.319	-0.105	0.373	-0.240
EC	-0.008	-0.077	-0.071	0.086	-0.079	-0.032	-0.141	-0.031	0.082	0.047	-0.101	0.043	0.034	-0.079
pH CaCl ₂	0.138	-0.551	0.037	0.331	-0.627	-0.188	-0.593	-0.573	0.169	-0.011	-0.417	-0.216	0.437	-0.398
pH H ₂ O	0.170	-0.592	0.010	0.371	-0.660	-0.223	-0.621	-0.604	0.203	0.019	-0.458	-0.204	0.471	-0.415
N	-0.418	0.576	0.284	-0.589	0.646	0.456	0.545	0.619	-0.443	-0.339	0.534	-0.064	-0.485	0.501
P	0.056	0.081	-0.170	0.075	0.028	-0.122	0.009	0.043	0.135	0.160	-0.047	0.170	-0.022	-0.197
K	0.153	-0.216	-0.004	0.168	-0.265	-0.085	-0.252	-0.238	0.132	0.027	-0.125	-0.045	0.242	-0.162
Ox Al	-0.318	0.582	0.074	-0.529	0.804	0.306	0.690	0.783	-0.280	-0.125	0.533	0.137	-0.498	0.383
Ox Fe	-0.362	0.702	0.145	-0.601	0.851	0.441	0.741	0.812	-0.365	-0.209	0.643	0.136	-0.604	0.453

Tabel 7.11. Interkorrelationsmatricer beregnet for observerede værdier på alle geomorfologiinputparametrene. Værdier er korrelationskoefficienten r .

	DEM	Aspect	Slope	Concavity	Convexity	Cross-Sectional Curvature	Flow Line Curvature	General Curvature	Geomorphons	Hill Height	Longitudinal Curvature	Mass Balance Index	Maximal Curvature	Minimal Curvature	Morphometric Features	MRRTF	MRVBF	Plan Curvature	Profile Curvature	Tangential Curvature	Terrain Ruggedness Index	TOBIA Index	Topographic Position Index	Total Curvature	Valley Depth	Wind Exposition	Horizontal Distance to River Network
Coarse	0.059	-0.016	0.211	-0.088	0.088	0.052	-0.051	0.065	-0.063	0.063	0.052	0.078	0.142	-0.055	0.097	-0.091	-0.207	0.009	0.038	0.053	0.215	-0.061	0.069	0.091	0.069	0.104	0.025
Clay	-0.074	-0.009	-0.167	-0.001	0.001	-0.017	-0.002	-0.007	-0.001	-0.033	0.006	-0.010	-0.102	0.087	-0.106	0.113	0.116	-0.002	0.017	-0.031	-0.168	0.040	0.006	-0.096	-0.067	-0.020	0.007
Silt	0.127	-0.018	0.177	-0.119	0.119	0.032	-0.033	0.041	-0.030	0.127	0.034	0.039	0.205	-0.143	0.105	-0.077	-0.189	0.037	0.035	0.022	0.185	-0.005	0.049	0.177	0.123	0.081	0.035
Sand	-0.036	0.017	-0.013	0.076	-0.076	-0.010	0.023	-0.022	0.020	-0.061	-0.025	-0.019	-0.070	0.038	-0.003	-0.018	0.050	-0.022	-0.032	0.005	-0.017	-0.021	-0.035	-0.055	-0.037	-0.039	-0.027
BD	-0.083	-0.018	-0.288	0.119	-0.118	-0.082	-0.037	-0.065	-0.044	-0.173	-0.024	-0.056	-0.253	0.157	-0.199	0.146	0.263	-0.032	-0.012	-0.078	-0.297	0.078	-0.060	-0.205	-0.132	-0.140	-0.069
OC	0.232	0.016	0.392	-0.118	0.115	0.077	-0.014	0.092	0.032	0.273	0.071	0.092	0.366	-0.245	0.283	-0.210	-0.334	0.036	0.042	0.084	0.406	-0.084	0.094	0.306	0.165	0.179	0.106
CaCO ₃	-0.012	-0.001	-0.086	0.054	-0.053	-0.011	-0.030	-0.004	0.028	-0.084	0.005	-0.009	-0.059	0.052	-0.025	-0.081	-0.048	-0.023	0.007	-0.019	-0.083	0.030	0.009	-0.072	-0.072	0.008	0.015
EC	-0.095	-0.021	-0.036	0.055	-0.054	-0.026	-0.002	-0.024	0.031	-0.004	-0.012	-0.033	-0.027	-0.002	-0.028	-0.037	0.041	-0.019	-0.011	-0.024	-0.034	0.020	-0.035	-0.003	0.018	-0.060	-0.045
pH CaCl ₂	0.063	0.012	-0.231	0.134	-0.130	-0.075	-0.027	-0.069	-0.008	-0.132	-0.036	-0.077	-0.207	0.119	-0.186	0.079	0.197	-0.020	-0.018	-0.081	-0.232	0.041	-0.075	-0.158	-0.071	-0.114	-0.077
pH H ₂ O	0.047	0.011	-0.258	0.133	-0.129	-0.076	-0.027	-0.068	-0.015	-0.146	-0.034	-0.077	-0.231	0.142	-0.201	0.098	0.213	-0.017	-0.015	-0.084	-0.260	0.043	-0.071	-0.185	-0.094	-0.110	-0.072
N	0.209	0.021	0.378	-0.093	0.090	0.067	-0.018	0.070	0.054	0.238	0.046	0.073	0.353	-0.258	0.269	-0.210	-0.312	0.027	0.024	0.073	0.392	-0.069	0.063	0.318	0.216	0.129	0.072
P	-0.126	0.003	-0.093	0.047	-0.046	-0.004	0.018	0.004	-0.027	-0.005	0.010	0.004	-0.038	0.042	-0.060	0.012	0.053	-0.007	0.007	-0.004	-0.089	0.023	-0.004	-0.027	-0.017	-0.049	-0.066
K	0.019	0.003	-0.113	0.049	-0.047	-0.017	-0.020	-0.005	-0.001	-0.036	0.009	-0.012	-0.072	0.064	-0.076	0.026	0.103	-0.014	0.015	-0.028	-0.112	0.006	-0.007	-0.064	-0.043	-0.061	-0.057
Ox Al	-0.033	-0.016	0.146	-0.110	0.103	0.018	-0.086	0.023	-0.036	0.137	0.020	0.043	0.127	-0.106	0.030	0.012	-0.121	-0.004	0.040	-0.008	0.142	-0.045	0.079	0.156	0.120	0.084	0.011
Ox Fe	-0.098	-0.023	0.162	-0.018	0.014	0.057	-0.149	0.050	0.008	0.331	0.025	0.048	0.256	-0.208	0.057	-0.010	-0.082	0.005	0.025	0.044	0.175	-0.006	0.060	0.348	0.258	-0.016	-0.056

Tabel 7.12. Interkorrelationsmatricer beregnet for prædikterede værdier på alle geomorfologiinputparametrene. Værdier er korrelationskoefficienten r . Heri kan det ses, at de prædikterede værdier fra Machine Learning-modellen følger lignende tendenser som interkorrelationerne for de observerede jordparametre, som kan ses i tabel 7.11.

	DEM	Aspect	Slope	Concavity	Convexity	Cross-Sectional Curvature	Flow Line Curvature	General Curvature	Geomorphons	Hill Height	Longitudinal Curvature	Mass Balance Index	Maximal Curvature	Minimal Curvature	Morphometric Features	MRTF	MRVBF	Plan Curvature	Profile Curvature	Tangential Curvature	Terrain Ruggedness Index	TOBIA Index	Topographic Position Index	Total Curvature	Valley Depth	Wind Exposition	Horizontal Distance to River Network
Coarse	0.335	-0.020	0.284	-0.124	0.125	0.058	-0.043	0.064	-0.038	0.092	0.043	0.080	0.220	-0.133	0.148	-0.084	-0.254	0.033	0.027	0.064	0.290	-0.073	0.068	0.162	0.098	0.160	0.084
Clay	-0.236	0.004	-0.305	0.031	-0.031	-0.011	0.002	-0.003	-0.091	-0.087	0.007	-0.019	-0.210	0.194	-0.185	0.226	0.258	-0.013	0.021	-0.030	-0.310	0.046	-0.001	-0.181	-0.106	-0.055	-0.031
Silt	0.145	-0.001	0.319	-0.107	0.106	0.042	-0.024	0.043	0.082	0.160	0.026	0.047	0.302	-0.234	0.211	-0.254	-0.387	0.037	0.016	0.046	0.327	-0.040	0.043	0.256	0.160	0.110	0.060
Sand	0.023	0.006	-0.031	0.044	-0.044	-0.014	0.016	-0.019	-0.033	-0.050	-0.015	-0.013	-0.069	0.043	-0.026	0.066	0.138	-0.011	-0.019	-0.005	-0.032	-0.011	-0.022	-0.057	-0.037	-0.033	-0.027
BD	-0.146	-0.021	-0.456	0.135	-0.133	-0.088	-0.012	-0.077	-0.042	-0.241	-0.036	-0.077	-0.398	0.285	-0.318	0.232	0.386	-0.051	-0.011	-0.102	-0.467	0.086	-0.067	-0.331	-0.220	-0.176	-0.097
OC	0.300	0.014	0.520	-0.142	0.139	0.083	-0.028	0.081	0.049	0.329	0.048	0.093	0.470	-0.346	0.365	-0.279	-0.429	0.045	0.024	0.096	0.537	-0.103	0.086	0.407	0.221	0.212	0.143
CaCO ₃	0.016	0.002	-0.108	0.080	-0.079	-0.020	-0.042	-0.010	0.037	-0.082	0.004	-0.019	-0.073	0.057	-0.048	-0.078	-0.030	-0.012	0.005	-0.025	-0.107	0.023	-0.003	-0.081	-0.072	0.000	0.003
EC	-0.111	-0.019	-0.027	0.067	-0.066	-0.028	-0.003	-0.029	0.036	-0.003	-0.018	-0.032	-0.021	-0.013	-0.029	-0.043	0.042	-0.017	-0.016	-0.024	-0.026	0.014	-0.040	0.010	0.032	-0.068	-0.055
pH CaCl ₂	0.056	0.008	-0.251	0.142	-0.138	-0.075	-0.026	-0.070	-0.007	-0.150	-0.037	-0.075	-0.231	0.136	-0.201	0.090	0.215	-0.030	-0.019	-0.081	-0.254	0.039	-0.074	-0.174	-0.081	-0.121	-0.083
pH H ₂ O	0.038	0.007	-0.283	0.140	-0.137	-0.073	-0.025	-0.067	-0.015	-0.164	-0.034	-0.072	-0.256	0.164	-0.219	0.112	0.236	-0.027	-0.016	-0.081	-0.286	0.043	-0.068	-0.203	-0.106	-0.118	-0.078
N	0.263	0.021	0.477	-0.108	0.105	0.063	-0.031	0.055	0.072	0.293	0.025	0.067	0.432	-0.342	0.331	-0.264	-0.385	0.038	0.002	0.080	0.494	-0.085	0.050	0.403	0.266	0.148	0.095
P	-0.213	0.006	-0.125	0.068	-0.068	-0.010	0.006	-0.019	-0.033	-0.011	-0.020	-0.023	-0.068	0.041	-0.064	0.044	0.122	-0.006	-0.017	-0.009	-0.122	0.030	-0.026	-0.028	-0.014	-0.082	-0.085
K	0.009	0.006	-0.111	0.050	-0.048	-0.013	-0.021	-0.006	-0.002	-0.035	0.004	-0.010	-0.067	0.056	-0.071	0.018	0.102	-0.013	0.009	-0.021	-0.110	0.002	-0.005	-0.055	-0.043	-0.058	-0.056
Ox Al	0.032	0.002	0.388	-0.133	0.125	0.112	-0.023	0.112	0.004	0.293	0.068	0.109	0.378	-0.225	0.292	-0.176	-0.291	0.043	0.039	0.119	0.398	-0.039	0.109	0.309	0.146	0.152	0.070
Ox Fe	0.050	-0.002	0.450	-0.083	0.078	0.091	-0.017	0.074	0.054	0.270	0.028	0.084	0.419	-0.308	0.316	-0.216	-0.315	0.045	-0.005	0.114	0.463	-0.045	0.063	0.374	0.229	0.088	0.038

Overordnet er der lav til insignifikant forskel mellem de udregnede korrelationer mellem jordparametre og inputparametre. Der er dog enkelte tilfælde, hvor der er opstået særlige ændringer i korrelationen fra observeret til prædikteret. Et eksempel på sådan en forskel er for blandt andet parameteren *coarse* relativt til *DEM*, der har en score på $r = 0,059$ for observationen, imens prædiktionen opnår en score på $r = 0,335$. Begge med positive korrelationer, men med en forskel på 0,276 point. Lignende ses også mellem ler, hvor forskellen dog er mindre og på 0,162. Disse forskelle som ses mellem observationer og prædiktioner, kan tyde på, at der potentielt kan mangle nøgle-inputparametre i modellen. Af denne grund kan det ske, at maskinlæringsmodellen vægter andre inputparametre højere for at opnå bedre resultater på bekostning af hvordan virkeligheden måske ser ud, altså at der kompenseres for eventuelle mangler. Det skal dog også tages in mente, at korrelationerne ikke nødvendigvis er lineære, hvilket i sig selv kan påvirke en r -score. Ligeledes kan forskelle også være påvirket af, at der af simple årsager kan være mangel på træningsdata i forbindelse med LUCAS-jordprøverne, og at de ikke dækker en tilstrækkelig bred mængde af eksempelvis miljøer eller andre kombinationer af de 64 inputparametre med videre. Generelt kan der derfor være mange årsager til at der ses forskelle, men at finde årsagen kræver en dybere målrettet undersøgelse af både iterativ og induktiv karakter.

7.1.3 Feature Importance Plots

Feature importance plots er anvendt for at visualisere vigtigheden af de forskellige inputparametre i forhold til prædiktionen af de individuelle jordparametre. Tabellerne viser, hvordan hver inputparameter bidrager til modellens estimering af de forskellige jordparametre, hvor højere værdier indikerer, at inputparameteren har større indflydelse på prædiktionsresultatet. Denne værdi, kaldet en vigtighedsfaktor, kan variere mellem 0 og 1, hvor 0 betyder, at parameteren ikke har nogen indflydelse, og 1 betyder, at parameteren har maksimal indflydelse på modellens prædiktionsoutput. Disse vigtighedsfaktorer giver indsigt i, hvilke inputparametre der er mest afgørende for modellens prædiktioner.

I tabel 7.13, 7.14 og 7.15 er der opgjort hvor vigtige hver inputparametre er i forhold til prædiktionsresultatet af hver jordparameter. Farvekodningen i tabellerne afspejler vægtningen af disse vigtighedsfaktorer for hver enkelte række, hvor blå indikerer de input med laveste vægtninger, mens den røde farve indikerer at dette input vægter højest og dermed har størst indflydelse på prædiktionsresultatet. Hvid repræsenterer midtpunktet mellem den laveste og højeste vægtning. Eksempelvis ses det for satellitinputparametrene, at bånd 1 (B01) og bånd 12 (B12) har størst vægt ved estimering af ler (Clay) i den anvendte maskinlæringsmodel.

Informationen disse *feature importance plots* indeholder kan være yderst værdifulde for fremtidig udvikling og optimering af modeller til prædiktion af jordparametre, eksempelvis hvis det ønskes at forsimple modellen ved at reducere antallet af inputparametre. Her kan ønsket være at eliminere de input med lavere vigtighedsfaktorer, da disse er mindre betydningsfulde og har mindst indflydelse på prædiktionsresultaterne. En anden overvejelse ved reduktion af inputparametrene er multikollinearitet, idet fjernelse af stærkt korrelerede inputparametre kan være med til at reducere deres overflødighed uden det går ud over modellens præcision og potentielt føre til at visse parametre får større betydning for prædiktionen. Dermed kan modellens kompleksitet reduceres uden det vil have en væsentlig effekt på modellens præcision. Fjernelse af inputparametre kan dog ændre vægtningen af de resterende parametre, hvilket vil føre til en justering af deres indbyrdes betydning for modellens prædiktionsoutput.

Tabel 7.13. Vigtighedsfaktorerne for satellit-inputparametrene for de forskellige prædikterede jordparametre. Højere værdi indikerer større indflydelse på prædiktionsresultaterne.

	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B11	B12	NDVI	EVI	SAVI	NDMI	COSRI	LSWI	BI	Clay Index	Salinity Index	Carbonate Index	Gypsum Index	Ferric Iron Index	Ferrous Iron Index	Gossan Index	Ferrous Silicates Index	Ferric Oxides Index
Coarse	0.0144	0.0098	0.0089	0.0076	0.0092	0.0098	0.0102	0.0088	0.0100	0.0103	0.0095	0.0072	0.0097	0.0073	0.0082	0.0075	0.0080	0.0080	0.0136	0.0081	0.0094	0.0081	0.0097	0.0130	0.0119	0.0136	0.0082
Clay	0.0263	0.0115	0.0125	0.0152	0.0131	0.0121	0.0110	0.0089	0.0101	0.0130	0.0237	0.0083	0.0082	0.0079	0.0096	0.0108	0.0080	0.0092	0.0136	0.0109	0.0110	0.0101	0.0117	0.0108	0.0135	0.0124	0.0096
Silt	0.0113	0.0109	0.0086	0.0080	0.0099	0.0118	0.0100	0.0098	0.0097	0.0105	0.0095	0.0108	0.0092	0.0090	0.0094	0.0135	0.0101	0.0099	0.0131	0.0116	0.0172	0.0108	0.0212	0.0123	0.0131	0.0145	0.0100
Sand	0.0193	0.0133	0.0114	0.0089	0.0123	0.0137	0.0141	0.0123	0.0131	0.0128	0.0113	0.0086	0.0084	0.0084	0.0101	0.0095	0.0094	0.0124	0.0159	0.0099	0.0124	0.0114	0.0137	0.0138	0.0146	0.0144	0.0106
BD	0.0106	0.0088	0.0079	0.0110	0.0087	0.0088	0.0089	0.0074	0.0083	0.0098	0.0112	0.0079	0.0113	0.0079	0.0139	0.0113	0.0147	0.0080	0.0113	0.0064	0.0157	0.0134	0.0152	0.0082	0.0118	0.0108	0.0100
OC	0.0102	0.0082	0.0063	0.0110	0.0088	0.0072	0.0083	0.0049	0.0057	0.0087	0.0094	0.0077	0.0096	0.0076	0.0075	0.0091	0.0082	0.0059	0.0079	0.0060	0.0096	0.0062	0.0100	0.0061	0.0100	0.0072	0.0074
CaCO ₃	0.0765	0.0224	0.0261	0.0173	0.0303	0.0165	0.0137	0.0106	0.0148	0.0193	0.0341	0.0082	0.0076	0.0093	0.0075	0.0078	0.0097	0.0113	0.0127	0.0081	0.0103	0.0075	0.0111	0.0116	0.0107	0.0132	0.0073
EC	0.0474	0.0249	0.0151	0.0116	0.0137	0.0101	0.0094	0.0087	0.0113	0.0106	0.0071	0.0119	0.0097	0.0094	0.0096	0.0124	0.0072	0.0102	0.0212	0.0142	0.0118	0.0090	0.0091	0.0107	0.0163	0.0198	0.0084
pH CaCl ₂	0.0277	0.0076	0.0204	0.0137	0.0162	0.0106	0.0073	0.0058	0.0069	0.0117	0.0166	0.0059	0.0057	0.0071	0.0078	0.0064	0.0120	0.0064	0.0181	0.0044	0.0073	0.0058	0.0073	0.0080	0.0091	0.0148	0.0075
pH H ₂ O	0.0282	0.0073	0.0204	0.0123	0.0174	0.0094	0.0070	0.0054	0.0063	0.0102	0.0108	0.0106	0.0055	0.0057	0.0071	0.0113	0.0143	0.0061	0.0165	0.0042	0.0068	0.0057	0.0067	0.0085	0.0086	0.0152	0.0065
N	0.0097	0.0107	0.0081	0.0112	0.0104	0.0057	0.0066	0.0050	0.0054	0.0082	0.0075	0.0059	0.0094	0.0086	0.0061	0.0085	0.0084	0.0054	0.0084	0.0058	0.0070	0.0080	0.0090	0.0061	0.0110	0.0096	0.0060
P	0.0143	0.0111	0.0108	0.0102	0.0132	0.0115	0.0108	0.0116	0.0115	0.0135	0.0119	0.0095	0.0102	0.0102	0.0119	0.0102	0.0123	0.0105	0.0144	0.0094	0.0143	0.0121	0.0115	0.0176	0.0130	0.0142	0.0119
K	0.0170	0.0121	0.0115	0.0135	0.0107	0.0095	0.0102	0.0090	0.0106	0.0138	0.0130	0.0098	0.0166	0.0112	0.0092	0.0140	0.0124	0.0097	0.0138	0.0091	0.0108	0.0133	0.0108	0.0141	0.0128	0.0165	0.0099
Ox Al	0.0100	0.0083	0.0062	0.0055	0.0055	0.0112	0.0103	0.0102	0.0121	0.0052	0.0042	0.0032	0.0257	0.0035	0.0056	0.0030	0.0038	0.0133	0.0096	0.0086	0.0097	0.0038	0.0088	0.0110	0.0065	0.0101	0.0062
Ox Fe	0.0206	0.0084	0.0089	0.0102	0.0139	0.0178	0.0101	0.0161	0.0126	0.0077	0.0076	0.0080	0.0090	0.0083	0.0055	0.0075	0.0061	0.0165	0.0069	0.0057	0.0073	0.0075	0.0074	0.0225	0.0106	0.0070	0.0074

Tabel 7.14. Vigtighedsfaktorerne for klima/CORINE-inputparametrene for de forskellige prædikterede jordparametre. Højere værdi indikerer større indflydelse på prædiktionsresultaterne.

	Dry days	DMP	Frost days	Max temp	Precipitation	Snowfall	Evaporation	Runoff	Mean temp	Min temp	Soil moisture	Longwave	Shortwave	Code
Coarse	0.0222	0.0153	0.0183	0.0245	0.0197	0.0187	0.0248	0.0178	0.0215	0.0352	0.0285	0.0461	0.0551	0.0147
Clay	0.0252	0.0396	0.0147	0.0334	0.0190	0.0177	0.0170	0.0195	0.0314	0.0200	0.0186	0.0275	0.0322	0.0611
Silt	0.0366	0.0227	0.0123	0.0152	0.0366	0.0150	0.0119	0.0322	0.0104	0.0210	0.0444	0.0159	0.0772	0.0266
Sand	0.0302	0.0265	0.0153	0.0191	0.0227	0.0186	0.0172	0.0246	0.0158	0.0213	0.0268	0.0293	0.0378	0.0206
BD	0.0127	0.0243	0.0150	0.0640	0.0583	0.0258	0.0118	0.0545	0.0224	0.0296	0.0133	0.0186	0.0403	0.0712
OC	0.0126	0.0139	0.0165	0.0643	0.0486	0.0121	0.0088	0.0489	0.0354	0.0140	0.0181	0.0101	0.0447	0.0986
$CaCO_3$	0.0157	0.0385	0.0132	0.0152	0.0294	0.0125	0.0137	0.0292	0.0140	0.0107	0.0128	0.0145	0.0280	0.0180
EC	0.0238	0.0144	0.0155	0.0249	0.0188	0.0163	0.0242	0.0166	0.0224	0.0167	0.0173	0.0220	0.0473	0.0574
$pH\ CaCl_2$	0.0282	0.0748	0.0138	0.0236	0.0638	0.0169	0.0177	0.0837	0.0183	0.0194	0.0180	0.0168	0.0360	0.0822
$pH\ H_2O$	0.0240	0.0958	0.0133	0.0247	0.0650	0.0157	0.0182	0.0858	0.0194	0.0178	0.0165	0.0158	0.0360	0.0778
N	0.0158	0.0222	0.0181	0.0894	0.0661	0.0145	0.0127	0.0541	0.0432	0.0128	0.0315	0.0135	0.0299	0.0401
P	0.0215	0.0244	0.0147	0.0171	0.0191	0.0136	0.0213	0.0157	0.0174	0.0175	0.0291	0.0193	0.0191	0.0349
K	0.0172	0.0235	0.0171	0.0190	0.0208	0.0171	0.0158	0.0224	0.0269	0.0162	0.0270	0.0211	0.0367	0.0270
Ox Al	0.0118	0.0122	0.0152	0.0280	0.0781	0.0137	0.1018	0.0561	0.0146	0.0141	0.0250	0.0117	0.0206	0.0282
Ox Fe	0.0171	0.0272	0.0142	0.0263	0.1446	0.0079	0.0378	0.0833	0.0154	0.0106	0.0288	0.0221	0.0450	0.0117

Tabel 7.15. Vigtighedsfaktorerne for geomorfologi-inputparametrene for de forskellige prædikterede jordparametre. Højere værdi indikerer større indflydelse på prædiktionsresultaterne.

	DEM	Aspect	Slope	Concavity	Convexity	Cross-Sectional Curvature	Flow Line Curvature	General Curvature	Geomorphons	Hill Height	Longitudinal Curvature	Mass Balance Index	Maximal Curvature	Minimal Curvature	Morphometric Features	MRRTF	MRVDF	Plan Curvature	Profile Curvature	Tangential Curvature	Terrain Ruggedness Index	TOBIA Index	Topographic Position Index	Total Curvature	Valley Depth	Wind Exposition	Horizontal Distance to River Network
Coarse	0.0406	0.0149	0.0187	0.0149	0.0134	0.0116	0.0155	0.0106	0.0110	0.0061	0.0102	0.0148	0.0116	0.0124	0.0094	0.0116	0.0259	0.0122	0.0112	0.0107	0.0212	0.0146	0.0124	0.0088	0.0063	0.0125	0.0142
Clay	0.0401	0.0126	0.0102	0.0144	0.0150	0.0086	0.0089	0.0079	0.0107	0.0040	0.0081	0.0097	0.0094	0.0090	0.0088	0.0132	0.0142	0.0103	0.0082	0.0090	0.0096	0.0101	0.0081	0.0066	0.0067	0.0120	0.0146
Silt	0.0161	0.0151	0.0117	0.0169	0.0182	0.0085	0.0087	0.0089	0.0106	0.0028	0.0094	0.0105	0.0094	0.0107	0.0106	0.0163	0.0207	0.0127	0.0106	0.0103	0.0106	0.0130	0.0093	0.0070	0.0059	0.0135	0.0184
Sand	0.0349	0.0166	0.0117	0.0160	0.0184	0.0092	0.0089	0.0078	0.0132	0.0038	0.0094	0.0114	0.0099	0.0111	0.0108	0.0172	0.0249	0.0137	0.0103	0.0093	0.0117	0.0124	0.0090	0.0069	0.0067	0.0127	0.0203
BD	0.0227	0.0109	0.0093	0.0165	0.0156	0.0074	0.0068	0.0071	0.0112	0.0022	0.0070	0.0073	0.0090	0.0067	0.0060	0.0132	0.0175	0.0071	0.0069	0.0130	0.0097	0.0108	0.0064	0.0043	0.0013	0.0119	0.0109
OC	0.0310	0.0100	0.0123	0.0143	0.0151	0.0106	0.0143	0.0132	0.0075	0.0059	0.0137	0.0100	0.0092	0.0178	0.0074	0.0130	0.0126	0.0079	0.0085	0.0091	0.0147	0.0119	0.0209	0.0067	0.0086	0.0154	0.0170
CaCO ₃	0.0167	0.0115	0.0154	0.0106	0.0106	0.0093	0.0104	0.0084	0.0104	0.0016	0.0102	0.0102	0.0108	0.0098	0.0085	0.0185	0.0255	0.0097	0.0102	0.0096	0.0179	0.0120	0.0085	0.0066	0.0029	0.0110	0.0119
EC	0.0280	0.0172	0.0115	0.0091	0.0105	0.0130	0.0120	0.0075	0.0092	0.0015	0.0107	0.0138	0.0150	0.0103	0.0068	0.0143	0.0126	0.0110	0.0096	0.0101	0.0093	0.0137	0.0090	0.0033	0.0017	0.0149	0.0164
pH CaCl ₂	0.0269	0.0086	0.0098	0.0111	0.0119	0.0057	0.0058	0.0051	0.0074	0.0021	0.0056	0.0063	0.0059	0.0063	0.0058	0.0096	0.0112	0.0067	0.0057	0.0057	0.0079	0.0076	0.0060	0.0042	0.0022	0.0075	0.0101
pH H ₂ O	0.0285	0.0082	0.0083	0.0107	0.0120	0.0053	0.0050	0.0050	0.0072	0.0018	0.0051	0.0062	0.0059	0.0061	0.0052	0.0092	0.0104	0.0064	0.0055	0.0056	0.0077	0.0070	0.0056	0.0040	0.0017	0.0075	0.0090
N	0.0390	0.0106	0.0116	0.0124	0.0115	0.0098	0.0124	0.0097	0.0088	0.0061	0.0101	0.0084	0.0086	0.0113	0.0074	0.0132	0.0134	0.0105	0.0102	0.0085	0.0116	0.0138	0.0134	0.0058	0.0093	0.0171	0.0197
P	0.0233	0.0176	0.0177	0.0180	0.0194	0.0130	0.0117	0.0127	0.0137	0.0007	0.0172	0.0155	0.0139	0.0152	0.0079	0.0189	0.0170	0.0145	0.0150	0.0126	0.0162	0.0202	0.0132	0.0065	0.0053	0.0163	0.0182
K	0.0203	0.0143	0.0127	0.0175	0.0187	0.0115	0.0064	0.0142	0.0132	0.0012	0.0119	0.0131	0.0120	0.0149	0.0068	0.0163	0.0263	0.0223	0.0116	0.0118	0.0111	0.0115	0.0123	0.0047	0.0033	0.0199	0.0279
Ox Al	0.0295	0.0405	0.0067	0.0379	0.0440	0.0099	0.0049	0.0096	0.0100	0.0024	0.0139	0.0120	0.0066	0.0144	0.0021	0.0064	0.0055	0.0076	0.0153	0.0094	0.0063	0.0122	0.0120	0.0034	0.0002	0.0117	0.0132
Ox Fe	0.0180	0.0101	0.0078	0.0150	0.0194	0.0088	0.0068	0.0065	0.0091	0.0014	0.0069	0.0096	0.0085	0.0090	0.0040	0.0080	0.0081	0.0100	0.0091	0.0087	0.0077	0.0073	0.0085	0.0041	0.0013	0.0092	0.0086

Blandt de inputparametre, der i tabellerne 7.13, 7.14 og 7.15 viste sig at have den største indflydelse på modellens prædiktation, fremtræder B01, Shortwave radiation, *Corine Land Cover Code* og DEM (højemodellen). Disse parametre havde generelt en høj vigtighedsfaktor, hvilket betyder, at de i modellen indeholder den mest væsentlige information for at kunne forudsige jordens egenskaber med højest præcision.

DEM er især vigtig i forhold til jordens fysiske karakteristika såsom erosion og vandstrømning, mens Shortwave og B01 sandsynligvis afspejler spektrale egenskaber, der er tæt forbundet med jordens overfladeforhold. Der kunne potentielt, som i figur 5.1, være en sammenhæng mellem disse spektrale egenskaber og specifikke jordkarakteristika, hvilket øger deres indflydelse på prædiktionsresultaterne.

Precipitation og Runoff havde generelt højere vigtighedsfaktor i forhold til organisk kulstof og BD, men også flere af de kemiske jordparametre, hvilket understøtter, at disse hydrologiske faktorer spiller en central rolle i jordens evne til at transportere næringsstoffer, særligt N (kvælstof) og opretholde en bestemt kemisk balance. Dette kan skyldes deres direkte indflydelse på jordens fugtighed, men også evnen til at udvaske eller koncentrere kemiske forbindelser i jorden.

På den anden side viste flere parametre generelt en lavere vigtighedsfaktor i modellen. Dette inkluderede spektrale bånd som B08 og B8A, samt vegetationsindekser som NDVI,

SAVI, NDMI, og øvrige indekser såsom Gypsum Index, BI (Brightness Index), Sanility Index og Ferric Oxide Index. Disse parametre har muligvis mindre indflydelse på modellens prædiktion, fordi de i begrænset omfang er med til at forklare variationen i de forskellige jordparametre. For eksempel kan nogle indekser, der er designet til at måle bestemte jordtyper eller betingelser, være mindre relevante eller mindre generaliserbare for jordparametre, der er spredt på tværs af mange forskellige jordtyper, klimatyper og arealanvendelser.

Derudover har klimatiske input som Frost Days, Min. Temp, Snowfall og Evaporation også generelt lave vigtighedsfaktorer i forhold til de prædikterede jordbundsparametre. Dette kan skyldes, at disse parametre ikke direkte påvirker de fysiske eller kemiske egenskaber i jordbunden i det undersøgte område, eller fordi deres effekt allerede er afspejlet i andre relaterede inputparametre grundet deres interkorrelation.

Interkorrelationer mellem inputparametrene kan have en betydelig indflydelse på vigtighedsfaktoren, idet at når to eller flere parametre er interkorrelerede, deler de den samme information, og dermed reduceres deres individuelle vigtighed i modellen. For eksempel er der i tabel 7.13 en stærk positiv korrelation mellem NDVI og SAVI, samt mellem B08 og B8A. Dette betyder, at disse par af parametre måler de samme aspekter af vegetationsdækning eller spektralsignaturer fra jordbunden, hvilket kan være med til at forklare deres lavere vigtighedsfaktor i prædiktionen, idet modellen allerede har adgang til den nødvendige information fra blot én af parametrene, og dermed reduceres den relative betydning af den anden. Derudover viser parametrene Precipitation og Runoff i tabel 7.14 en stærk positiv korrelation, hvilket kan være en fordel i forhold til prædiktionen af jordens kemiske egenskaber, og være en indikator for at de tilsammen bidrager væsentligt til modellens prædiktionspræcision, i stedet for at skabe overflødig information.

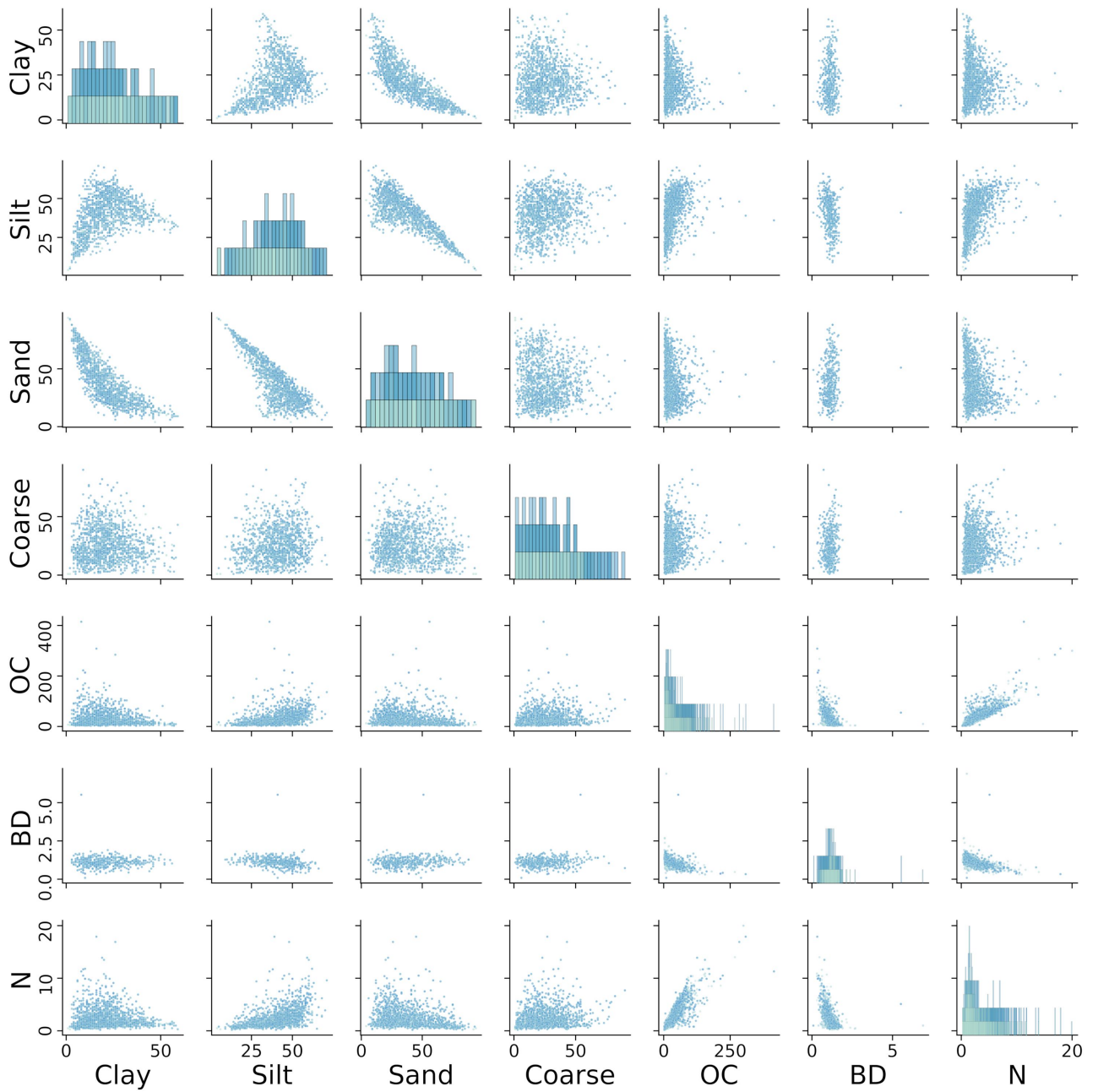
Der er også flere af de geomorfologiske inputparametre såsom Hill Height, Morphometric Features, Total Curvature og Valley Depth, der generelt havde lave vigtighedsfaktorer i modellen, som samtidig viste lav interkorrelation med andre parametre i tabel 7.15. Dette indikerer, at de tilføjer modellen unik information, men siden de har en lav vigtighed, er deres betydning for prædiktionen af jordens egenskaber begrænset. Den lave korrelation med andre parametre kan dog betyde, at de stadig kan have en vigtig rolle i andre modelleringskontekster, især hvor specifikke terrænforhold spiller en vigtig rolle for forholdet mellem jordbundsparametrene.

7.1.4 Scatterplots og identitetslinjer

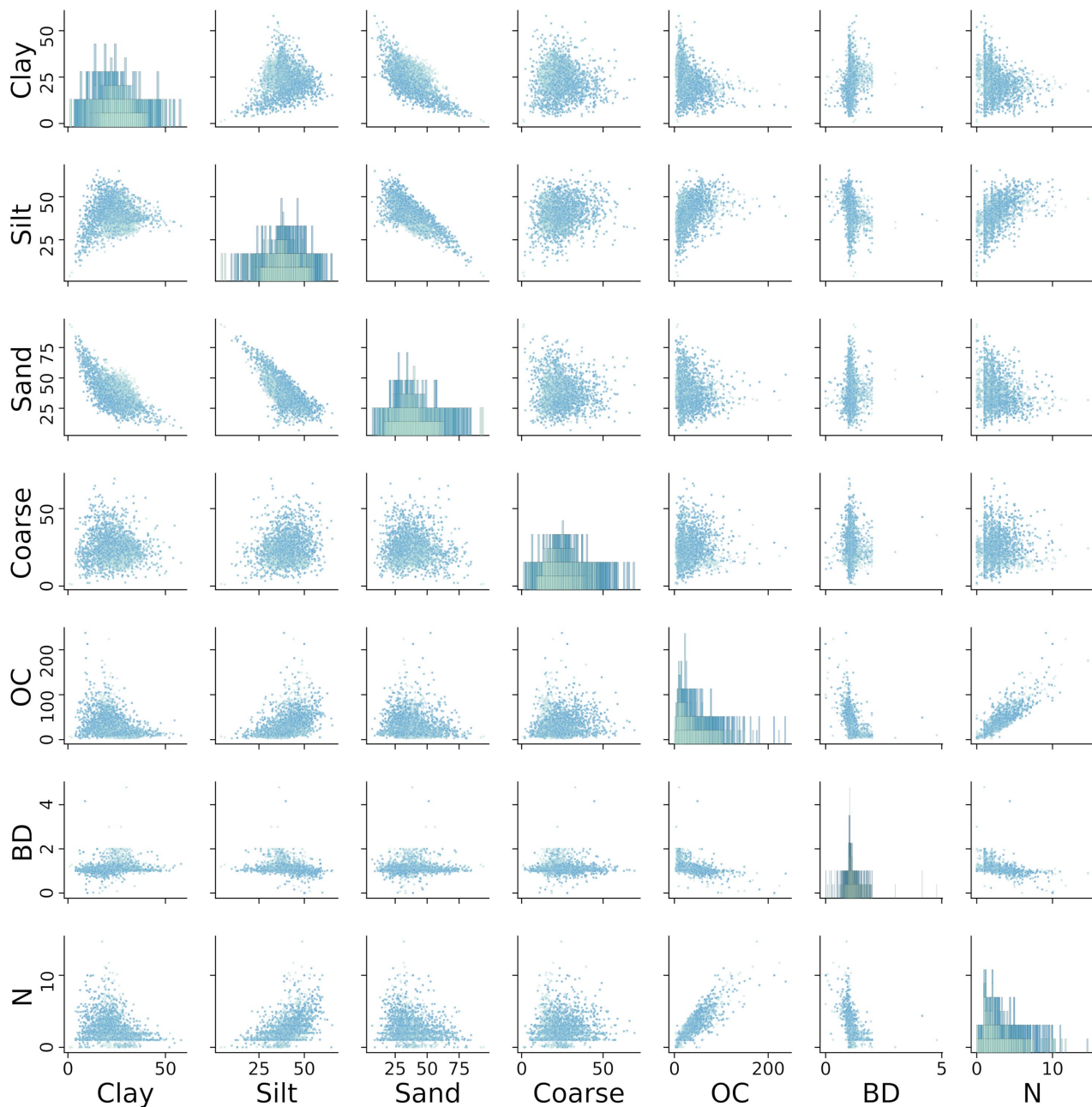
For at få en bedre forståelse af præcisionen af modellen på prædiktering af de forskellige jordparametre, er der lavet scatterplotsmatricer for både de observerede samt de prædikterede data. Herudfra kan der derfor med visuel analyse undersøges for, om de følger samme tendenser. Denne visuelle repræsentation af dataene kan hjælpe til at forklare om tendenser og korrelationer i datasættet end blot determinationskoefficienten, R , isoleret. På følgende figurer 7.1 og 7.2 ses et udsnit på syv af parametrene fra henholdsvis LUCAS datasættet samt datasættet for de prædikterede i scatterplots. Disse syv parametre er med til at formidle, om maskinlæringsmodellen formår at producere værdier for jordparametrene, der giver samme fortælling, hertil eksempelvis forhold og interkorrelationer, som hvad observationerne fortæller. Heri kan det dermed ses,

at de præsikterede data i større træk fortæller samme historie som de observerede, hvilket beretter om at modellen giver resultater der er virkelighedsnære med samme grundlæggende tendenser.

Ud fra scatterplotsene i figurene, kan det blandt andet ses at for jordparametrene ler, silt, sand og coarse, at de præsikterede resultater tyder på at følge hvad observationerne beretter. Der ses dog, som også gives til udtryk i blandt andet jordteksturtrekanterne i figur 7.7, samt 1:1 graferne i figurene 7.3, 7.4, 7.5 og 7.6, at modellen har til tendens at underestimere ved høje observationsværdier, hvorfor resultaterne for de præsikterede værdier er mere grupperede for nogle af parametrene.



Figur 7.1. Scatterplotmatrice analyse for syv af de observerede parametre fra LUCAS datasættet indenfor case området.



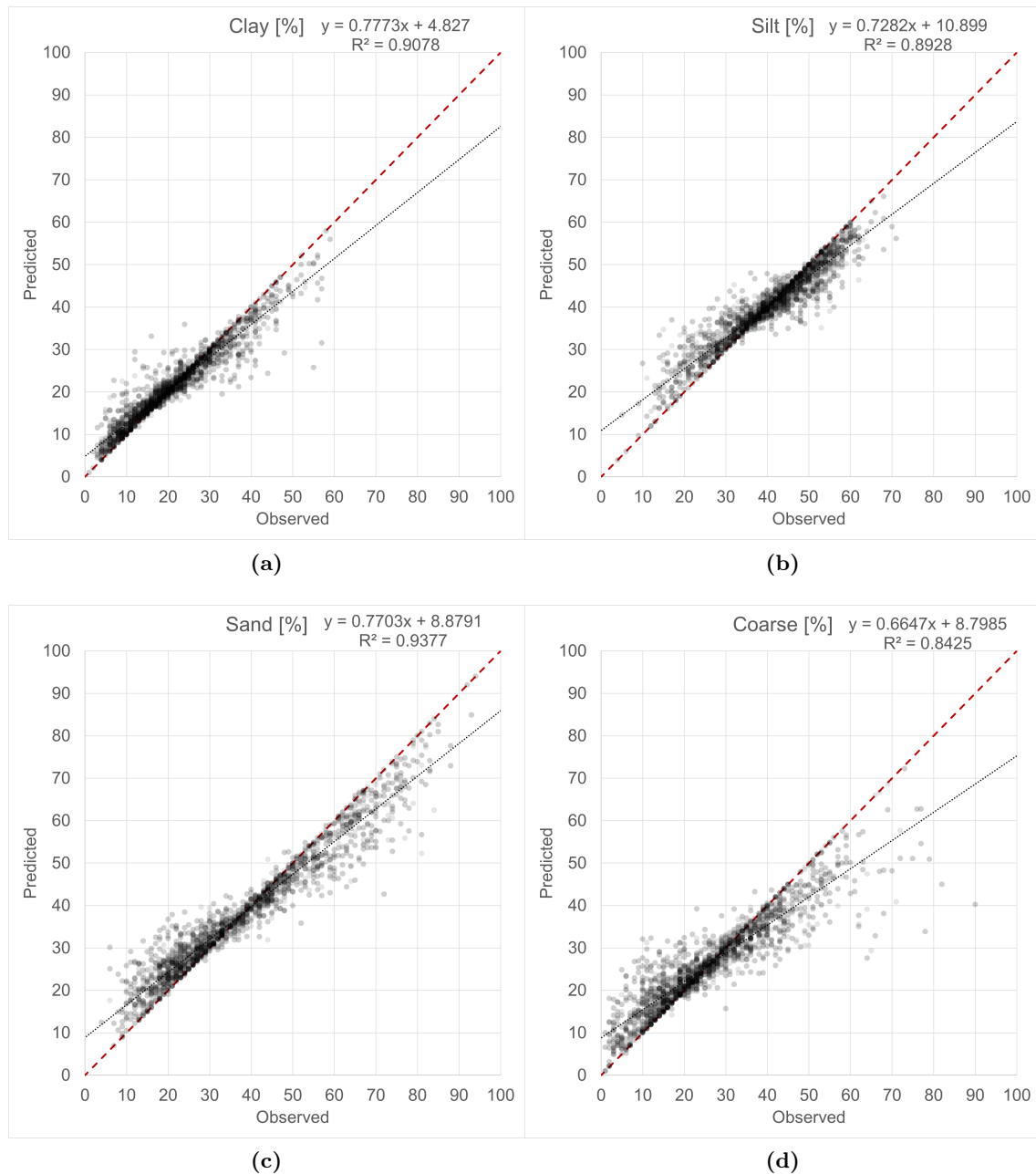
Figur 7.2. Scatterplotmatrice analyse for syv de præsikterede parametre, hvor der er samlet for alle punktlokationerne fra LUCAS datasættet for bedre sammenlignelighed.

Overordnet følger resultaterne fra maskinlæringsmodellen, det vil sige de tendenser og eventuelle korrelationer der er mellem parametre, med i forhold til hvad observationerne viser, som burde fungere som den faktiske repræsentation af virkeligheden. Umiddelbart ud fra disse scatterplots mellem parametrene, tyder det på, at anvendelse af maskinlæring til at prædiktere for en bred vifte af forskellige jordmæssige parametre, kan lade sig gøre. Præcisionen heraf kan dog forventes at variere til det værre eller bedre, alt afhængigt af input-data samt tid og sted for både træningsdataene såvel som de uafhængige data, som her eksempelvis er udgjort af blandt andet satellitbånd med mere.

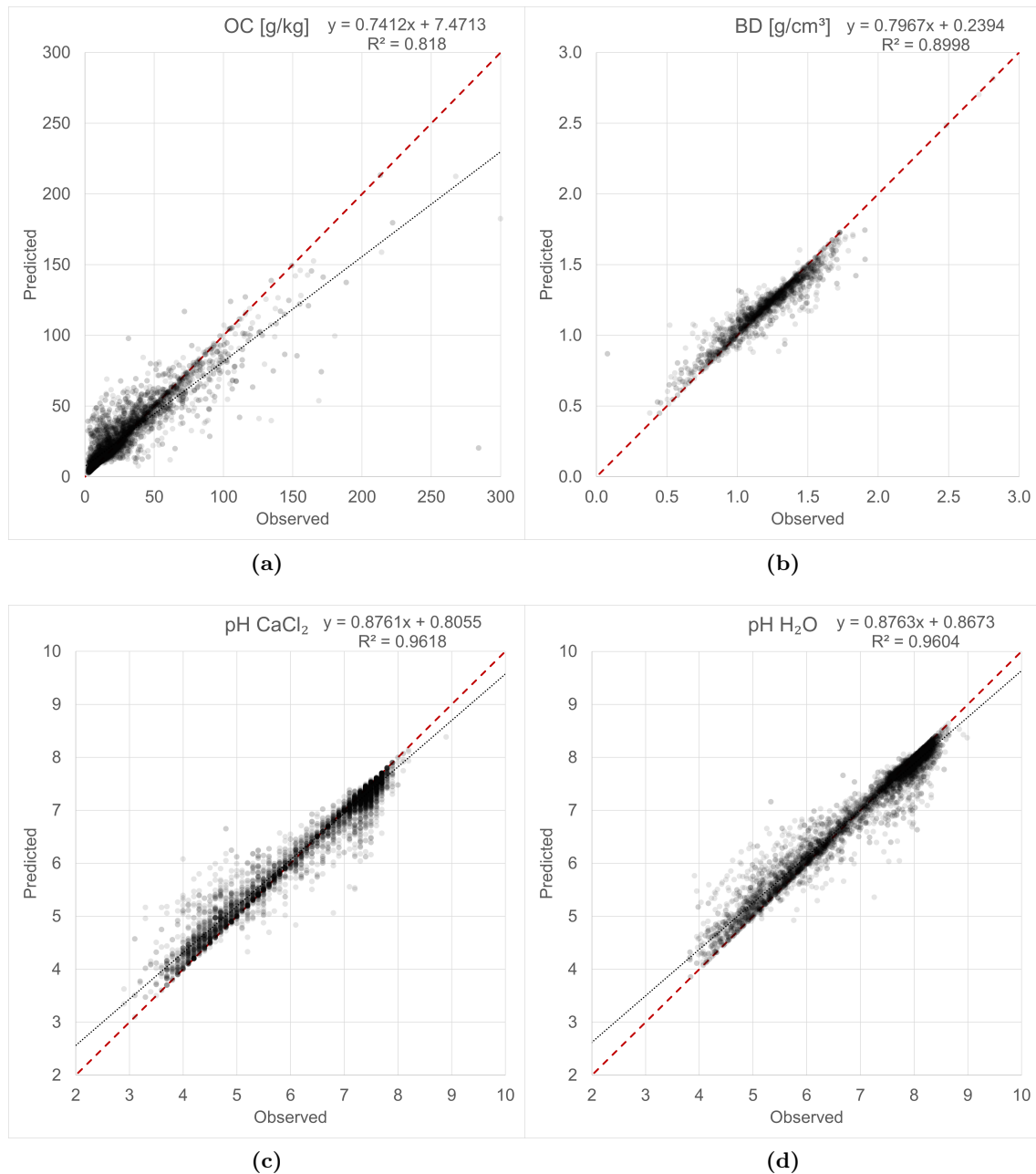
Herefter kan der i følgende figurer 7.3, 7.4, 7.5 og 7.6 lavet 1:1 grafer for de undersøgte jordparametre. Herved kan det aflæses hvor godt de prædikterede data følger observationerne, og om de eksempelvis kronisk overvurderer eller undervurderer og lignende. Det vil sige, at hvis de prædikterede data stemmer overens med hvad der er set, kan det netop bevidne til, hvor god modellen er til netop at prædiktere de undersøgte jordparametre. For hver plot er der beregnet R samt givet en lineær regression, der både beskriver præcisionen af modellen, samt elementer som bias der kan opstå. Kulminationen af statistikkerne på prædiktionerne sammenlignet fra de observerede, alle indenfor caseområdet, kan aflæses i tabel 7.16, hvortil der er beregnet og oplyst $RMSE$, RPD , $Bias$ og STD . Heri er RPD beregnet $\frac{STD_{obs}}{RMSE}$, og hvor det værste resultat lyder for prædiktionen af fosfor (P), der opnår en RPD på 1,997, svarende til knap 2. Sammenlignes dette eksempelvis med tærskelværdier for RPD , som opgivet i afsnit 1.2, opnår denne derved anden bedste score fra lige under $< 2,0$ lige på grænsen til bedste score.

For prædiktionerne for ler, silt, sand og coarse som vist i figur 7.3, ses der en generel tendens for, at der sker en overvurdering af prædiktionerne, hvor observationerne har lave værdier, imens der undervurderes ved de højere observationsværdier. Denne under og overvurdering ved de ekstreme værdier går igen ved alle de undersøgte jordparametre, og kan derfor anses som værende en kronisk fejl der går igen i maskinlæringsmodellen, hvorfor der er et sted der kan optimeres med henblik på at udrette disse afvigelse. Dette vil dog være en iterativ og induktiv proces, da det kan være svært at diagnosticere hvad den konkrete årsag er til problemet. Som udgangspunkt accepteres resultaterne af maskinlæringsmodellen dog på baggrund af de beregnede $RMSE$ for jordparametrene, som kan ses i tabel 7.16. Eksempelvis lyder $RMSE$ for ler på omtrentlige 3,6%, 4,3% for silt, 5,6% for sand, og 12,2 g/kg for OC.

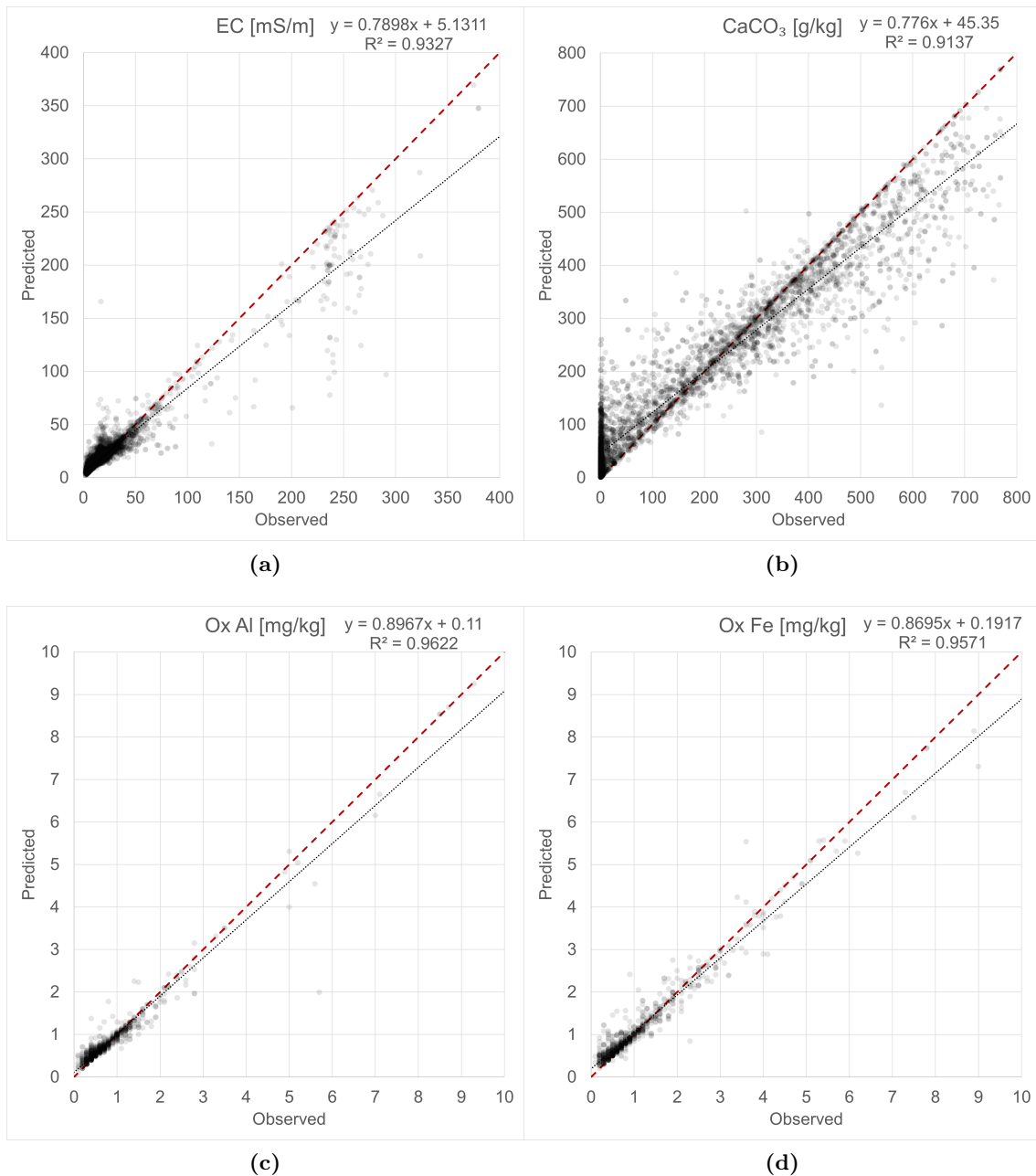
Som nævnt men som også kan ses af 1:1 graferne i følgende figurer 7.3, 7.4, 7.5 og 7.6, kan "grupperingen" af prædiktionsdataene også ses i forbindelse med de udregnede værdier for standardafvigelsen (STD) for de respektive parametre, ved at sammenligne dem med standardafvigelsen for deres respektive observerede data. Et visuelt eksempel på hvad der menes med "gruppering" kan blandt andet ses for ler/ler scatterplottet i scatterplotmatricen i figur 7.2, hvor der ved sammenligning på de observerede, fremstår en tydelig cirkulær "gruppering" i den centrale del af distributionen. Her lyder der en generelt lavere standardafvigelse for alle de prædikterede data relativt til de observerede. Beregningerne kan ses i tabel 7.16. "Datagrupperingen" er også til dels tydelig i forbindelse med opsætning af teksturtrekanter for både observerede og prædikterede data, som ses i henholdvist figur 7.7 a og b.



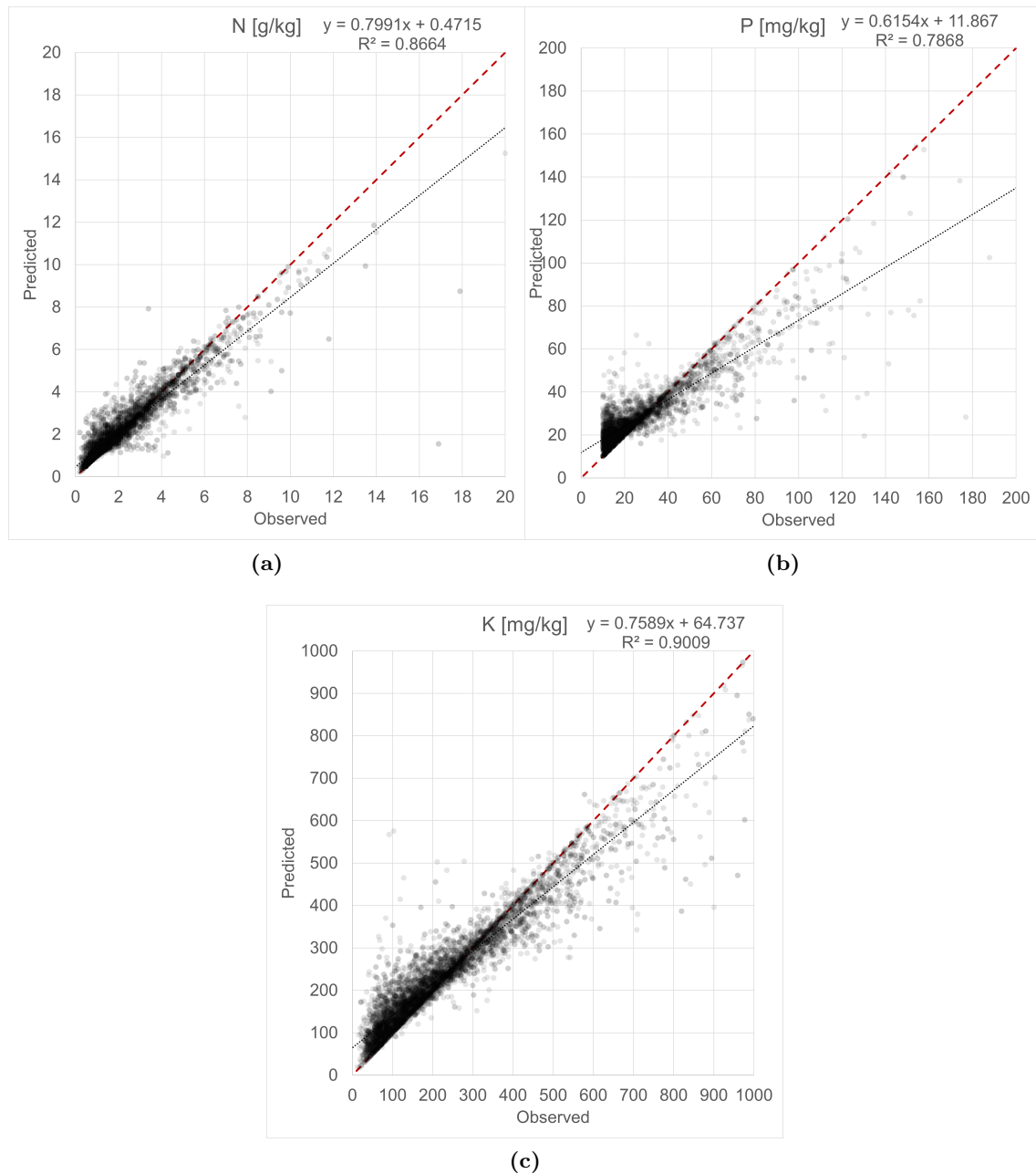
Figur 7.3. 1:1 grafer mellem prædiktionerne for *ler* (a), *silt* (b), *sand* (c) og *grovpartikler* (d) samt de respektive observerede værdier, samlet ved hver LUCAS-jordprøve lokation.



Figur 7.4. 1:1 grafer mellem prædiktionerne for *organiskkulstof* (a), *volumenvægt* (b), *pH*CaCl₂ (c) og *pH*H₂O (d) samt de respektive observerede værdier, samlet ved hver LUCAS-jordprøve lokation.



Figur 7.5. 1:1 grafer mellem prædiktionerne for *EC* (a), *CaCO₃* (b), *OxAl* (c) og *OxFe* (d) samt de respektive observerede værdier, samlet ved hver LUCAS-jordprøve lokation.



Figur 7.6. 1:1 grafer mellem prædiktionerne for *nitrogen* (a), *phosphor* (b) og *kalium* (c) samt de respektive observerede værdier, samlet ved hver LUCAS-jordprøve lokation.

Tabel 7.16. Statistik for RMSE, RPD, Bias og standardafvigelsen (STD) på differencen mellem de prædikterede og observerede jordparametre. Standardafvigelsen er beregnet for både observerede (*obs*) og prædikterede (*pred*) værdier. Jordparametrenes enheder kan læses i tabel 2.1.

	%	%	%	%	g/cm^3	g/kg	g/kg	mS/m	-	-	mg/kg	g/kg	mg/kg	mg/kg	mg/kg
	<i>Coarse</i>	<i>Clay</i>	<i>Silt</i>	<i>Sand</i>	<i>BD</i>	<i>OC</i>	<i>CaCO₃</i>	<i>EC</i>	<i>pH CaCl₂</i>	<i>pH H₂O</i>	<i>N</i>	<i>P</i>	<i>K</i>	<i>Ox Al</i>	<i>Ox Fe</i>
<i>RMSE</i>	6,383	3,623	4,281	5,561	0,086	12,216	67,785	12,084	0,270	0,262	0,648	13,459	64,397	0,229	0,302
<i>RPD</i>	2,265	3,002	2,697	3,294	2,986	2,295	3,056	3,346	4,670	4,614	2,679	1,997	2,851	4,849	4,397
<i>BIAS</i>	0,014	0,044	-0,064	0,109	-0,004	0,723	1,652	0,435	0,003	0,000	0,040	0,391	7,288	0,019	0,038
<i>STD_{obs}</i>	14,456	10,877	11,543	18,314	0,258	28,037	207,162	40,435	1,259	1,208	1,735	26,884	183,575	1,112	1,326
<i>STD_{pred}</i>	10,469	8,874	8,895	14,569	0,217	22,977	168,178	33,066	1,125	1,080	1,490	18,651	146,780	1,017	1,179

I tabel 7.16 kan det ses, at standardafvigelsen for de prædikterede jordparametre konsekvent er lavere end standardafvigelsen for de observerede jordparametre. Dette indikerer, at de prædikterede værdier har en tendens til at klumpe sig mere sammen omkring deres gennemsnit, hvilket tyder på en mindre variation i de prædikterede værdier sammenlignet med de observerede data. Dette fænomen kan også observeres i tidligere scatterplots og 1:1-regressionslinjer, hvor prædiktionerne er tættere på regressionslinjen, hvilket også understøtter ideen om, at modellen genererer mere homogene forudsigelser. En lavere standardafvigelse i de prædikterede værdier kan være en konsekvens af, at modellen er tilbøjelig til at udjævne ekstreme variationer i data, hvilket betyder, at modellen ikke altid fanger alle nuancer og variationer, der er til stede i de faktiske observerede jordparametre.

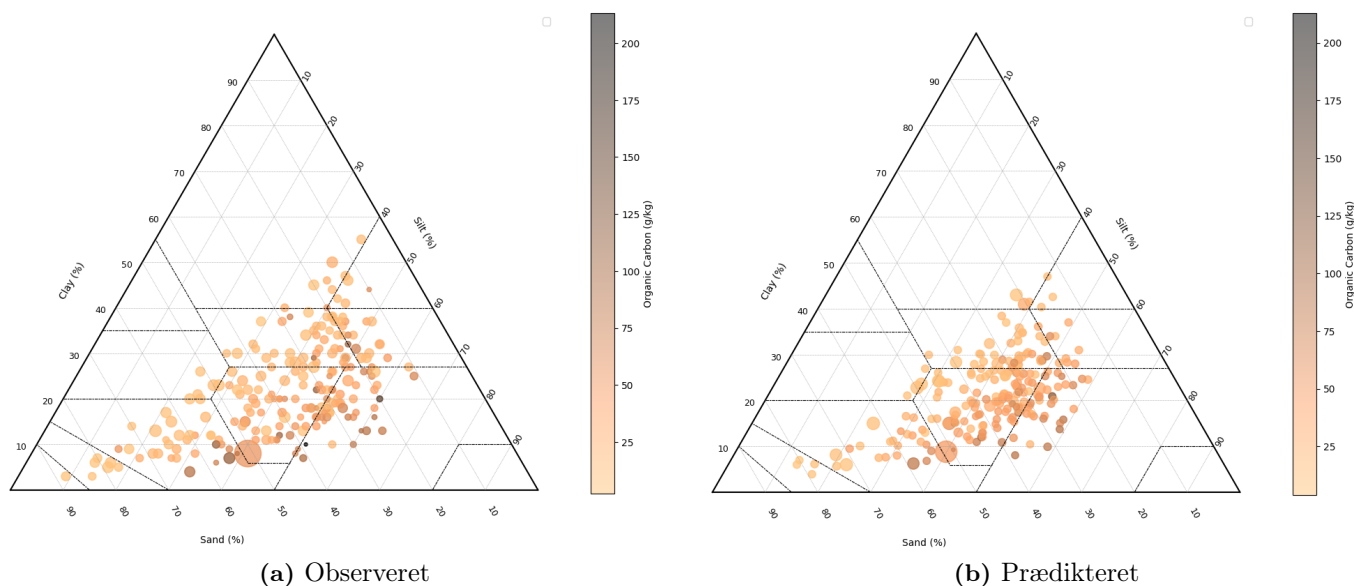
7.1.5 Teksturtrekanter

Yderligere analyse på de prædikterede og observerede jordprøver, især med fokus på ler-, silt-, og sandindholdet, bringer følgende jordteksturtrekantsresultater, som kan aflæses i følgende figur 7.7.

I subfigur *a* er der indsat de observerede fraktioner for LUCAS-jordprøverne i caseområdet jævnfør afgrænsningerne som vist på kortet i figur 2.3, og subfigur *b* præsenterer fraktionerne for de prædikterede data for selv samme afgrænsede område. Jordteksturen for de prædikterede jorde er samlet for samme punktlokationer for hvor der er foretaget LUCAS-jordprøver. Jordteksturtrekanterne i figur 7.7 herefter anvender *USDA* definitionerne for jordklasser, da det også stemmer overens i forhold til hvordan partikelstørrelsesafgrænsningerne er defineret for jordfraktionerne, ler, silt og sand som anvendes. Derudover formidles også jordens volumenvægt [*BD*] samt organisk kulstof [*OC*] i teksturtrekanterne, som henholdvist er præsenteret ved punkternes størrelse samt

farve. Jo større det enkelte punkt er, des højere er BD . Indholdet af OC følger farverne som givent i kolonnen til højre, hvor mørkere farve betyder mere OC imens lysere farve er lig med lavere OC .

Kortlægningen af jordklassifikationerne for hele caseområdet på baggrund af de prædikterede data fra maskinlæringsmodellen kan ses i figur 8.2, hvor langt størstedelen af området for tildelt klassifikationen *loam*, hvilket vil sige, at jordene består et sted mellem omtrentlige 7 – 27% ler, 23 – 52% sand, og 28 – 50% silt, som også kan ses i følgende figur 7.7.



Figur 7.7. Teksturtrekanter for hhv. observerede jordprøver fra LUCAS-jordprøverne og de prædikterede værdier for samme lokation. Størrelsen af det enkelte punkt afspejler volumenvægten (hhv. den observerede og prædikterede), og farven af det enkelte punkt afspejler mængden af organisk kulstof (hhv. den observerede og prædikterede). Teksturene følger USDA definitionen.

På baggrund af hvad der kan aflæses fra jordteksturerne observerede og prædikterede jordteksturerer i subfigurerne i figur 7.7, ses det, at de prædikterede i større træk er ækvivalent til de observerede jordteksturer. Forskelle mellem dem, ud fra hvad der kan ses i teksturtrekanterne, er dog, at maskinlæringsmodellen ikke formår at opnå, og derved prædiktere, jordteksturer, som ligger i samme ekstremer som dem der ses for de observerede jordprøver. Hvad der menes med ekstremer er eksempelvis, at der for de prædikterede ikke er nogle jorde, hvori der er sandindhold på over 85%, hvilket et tilfældet for de observerede.

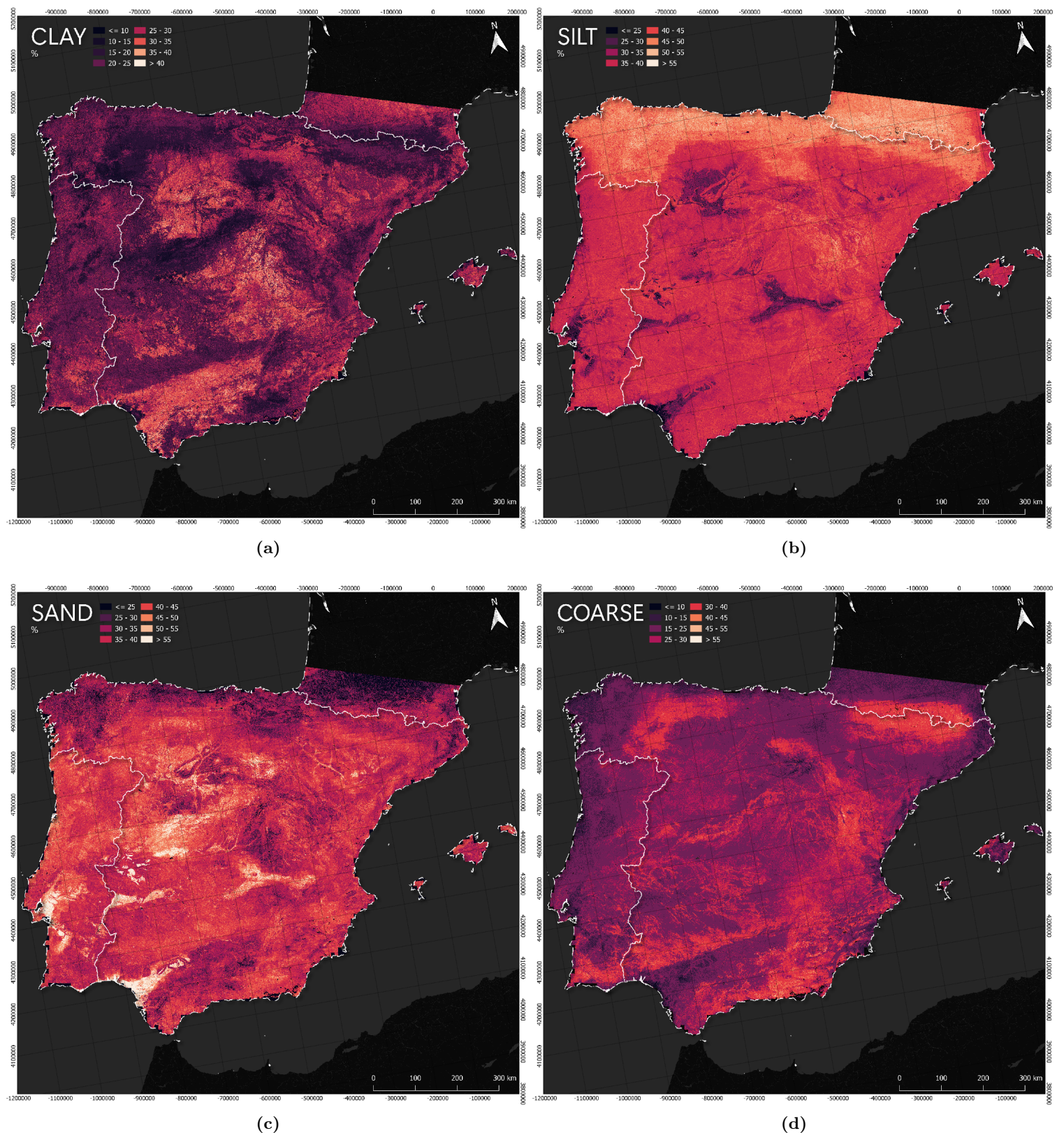
Ligeledes er der andre steder observeret jorde med et lerindhold på og over 50%, hvilket maskinlæringsmodellen heller ikke formår at opnå. Dette kan potentielt tyde på, at den udviklede model generelt underestimerer ved høje jordfraktioner, og modsat ved de lave fraktioner. Dette er evident ud fra de udarbejdede 1:1-plots for ler, silt og sand som kan aflæses i figur 7.3, som netop belyser at dette problem eksisterer. Netop det samme problem ses for det prædikterede BD og OC , hvor begge 1:1-plots herfor kan ses i figur 7.4. Dog følger både BD og OC , ligesom jordfraktionerne, grundlæggende hvad der er observeret, og afviger hovedsageligt kun ved ekstremerne.

Kortlægning af data og Multikriterieanalyse

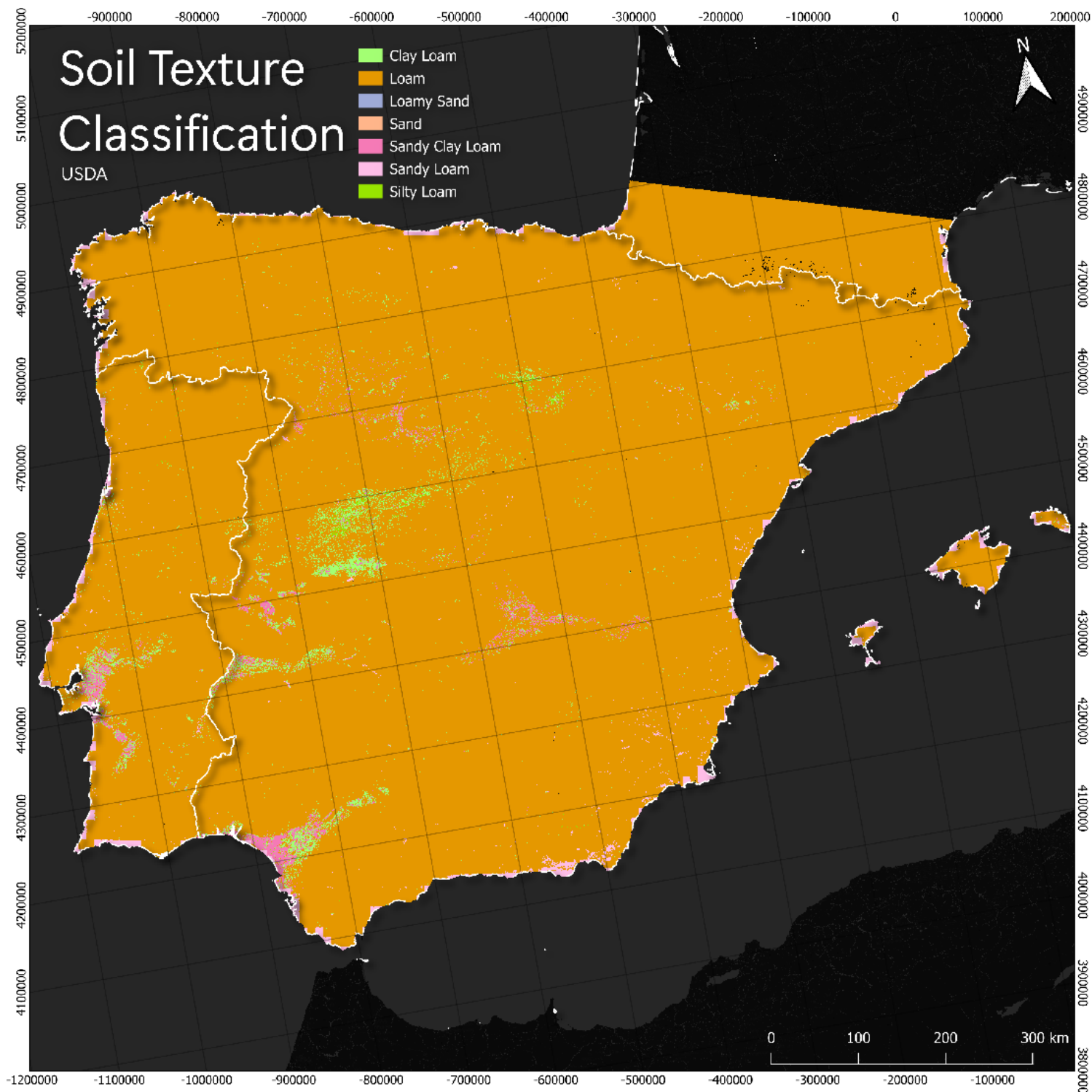
8

8.1 Prædiktioner

Alle kortene som præsenteres heri, er produktet af modellerne, som blev udviklet i forbindelse med Machine Learning, via bibliotekspakken kaldet *PyCaret* til Python, ved anvendelse af Extra Trees Regressor algoritmen i et Python programmerings miljø. Her blev der lavet en model for hver parameter, herunder ler, silt, sand, coarse, OC mv., med træning på den respektive observerede parameter, y , samlet fra de forskellige LUCAS punkter for det totale *samplings område*, som defineret i figur 2.2. For alle modellerne til hver jordparameter der præsikteres for, er der anvendt *random state* 0. Hvert kalibreringspunkt fra LUCAS datasættene er kombineret med et sæt af kendte uafhængige variabler, x_n , i en størrelsesorden af 64 forskellige lag, i kategorierne satellit, klima, land use/land cover, geomorfologi og hydromorfologi, også samlet fra samme lokation. Disse inputparametre kan blandt andet findes i tabellerne 5.1, 5.2, 5.3, 5.4 og 3.3. Herefter blev hver model anvendt til at prædiktere den enkelte parameter for hver enkelt celle på 250×250 m, hvoraf der er i alt 10.270.328 styk, som dækker hele caseområdet.



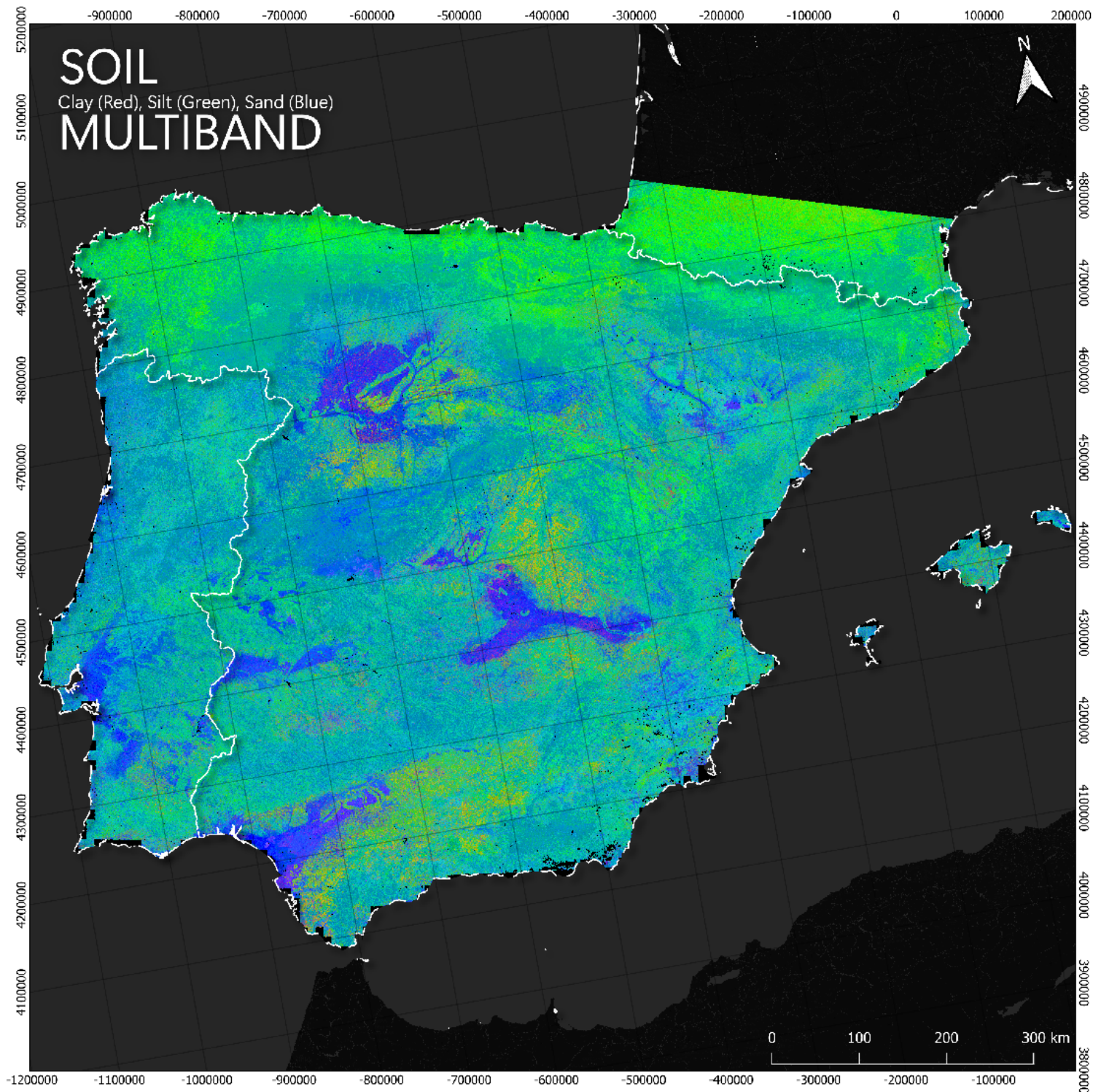
Figur 8.1. Kortlægning af prædikeret *ler* (a), *silt* (b), *sand* (c) og *grovpartikler* (d) som produktet af modeller udviklet igennem Machine Learning.



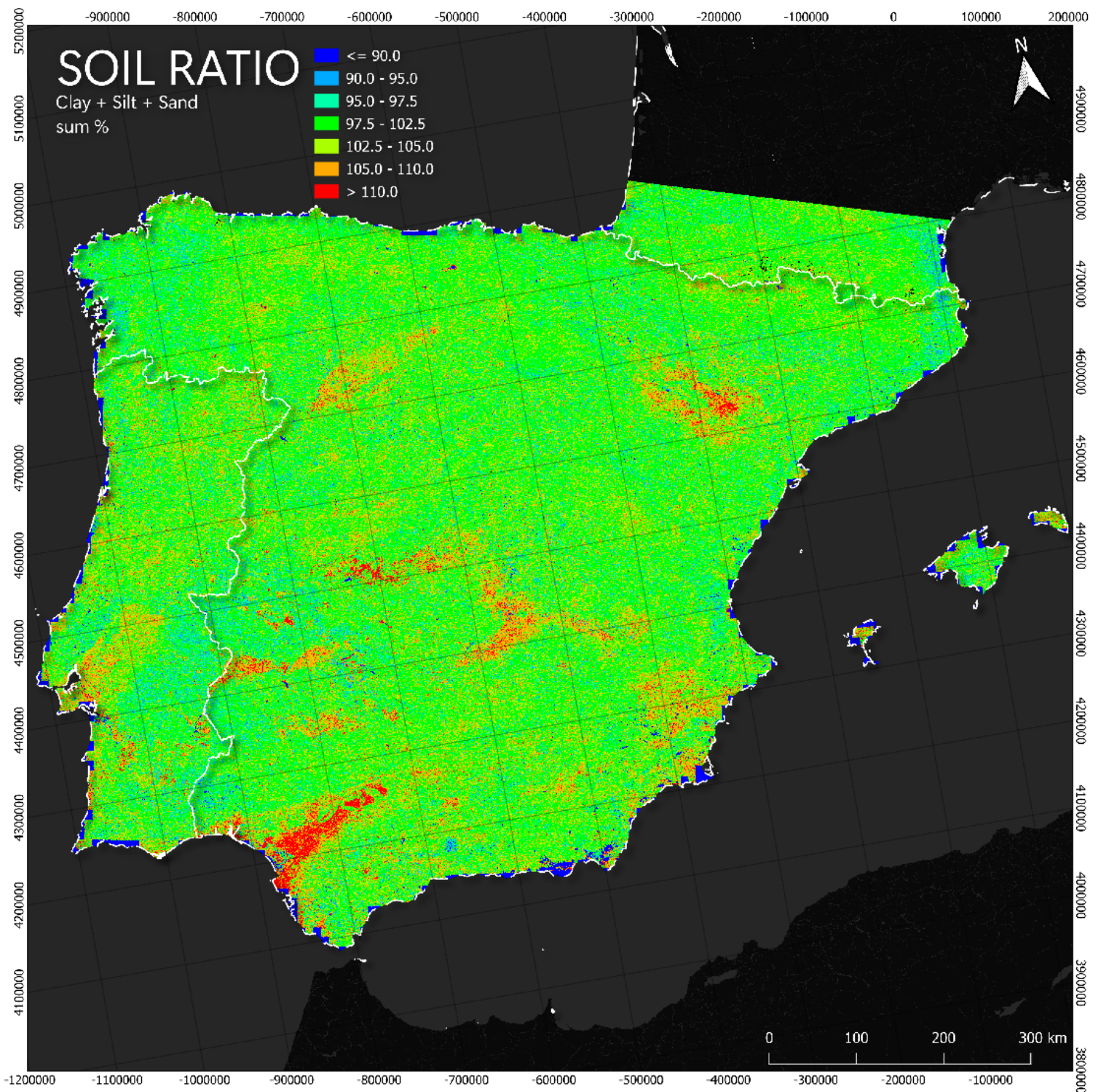
Figur 8.2. Kortlægning af den estimerede jordteksturklassifikation jf. USDA definition på baggrund af de prædikterede ler-, silt- og sandfraktioner som vist i figur 8.1. Overordnet er caseområdet, og afledt deraf den Iberiske Halvø, estimeret til at være særligt domineret af jordklassen *loam*.

Ved at kombinere kortene for ler, silt og sand i ét, hvor de hver repræsenterer rødt, grønt og blåt bånd, opnås følgende kort i figur 8.3. Med dette kan det nemt ses hvilke områder der er domineret af hvilken type. Dette kort fungerer til dels ligesom jordklassifikationskortet,

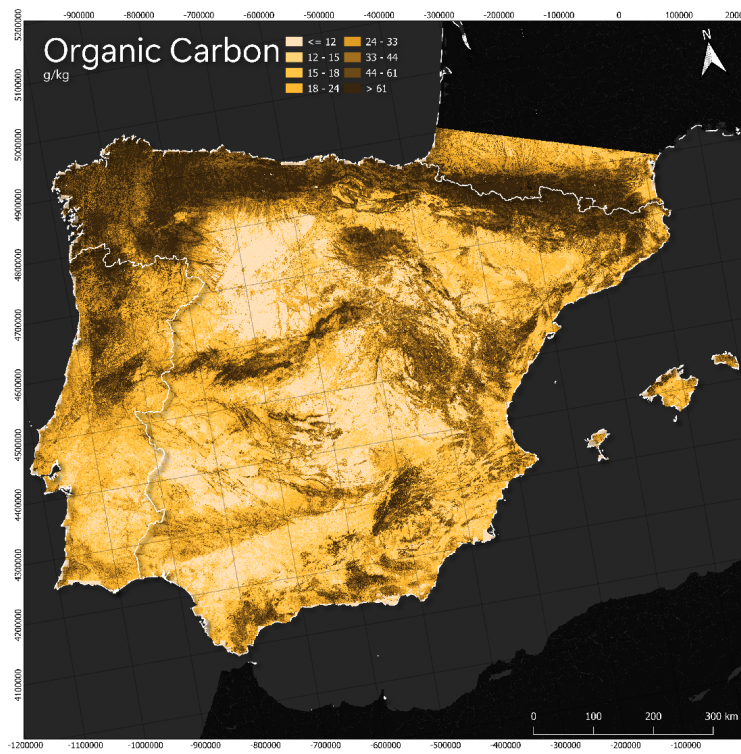
som vist i figur 8.2, men i stedet for indeksering og navngivning deraf, bidrager en opsætning i *RGB-multibånd*-format til at se i højere detaljegrad, samt de finere gradienter mellem hver jordtype. Eksempelvis, hvis et område hovedsageligt er ren blå, betyder det at området består udelukkende af sand, hvorimod grøn fortæller at stedet udelukkende er silt, og rød for ler. På kortet i figuren kan ler-dominerede områder blandt andet ses ved XY-koordinaterne $-700000, 470000$ og i det omkringliggende område, i områder der kan tyde på at være ved foden af de nærliggende bjerge, plateauer og bakketoppe, som kan ses i figur 3.7. Denne type af områder er, ud fra figuren, typisk hvor der er at finde lerede jorde. Generelt er størstedelen af den Iberiske Halvø, jf. prædiktionerne for ler, silt og sand, domineret af sand og silt, imens den nordliggende bjergterræn imod Biscayabugten og Atlanterhavet, samt Pyrenæerne er silt-dominerede. I figur 8.4 er summen af ler, silt og sand udregnet for hver pixel i caseområdet. Heri er der farvelagt i forhold til en sum af 100%, hvor 100% er det optimale og efterstræbte resultat for prædiktionerne på ler, silt og sand.



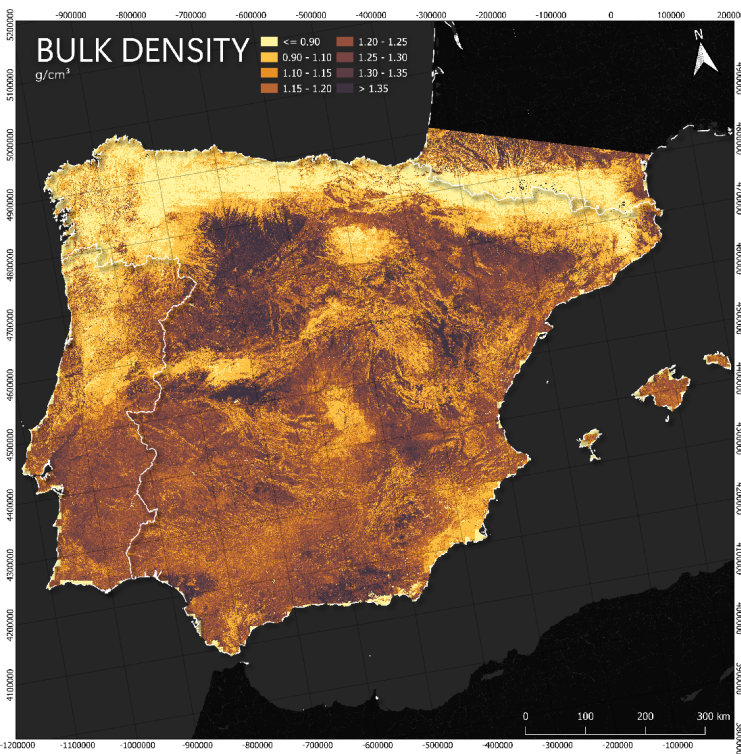
Figur 8.3. Formidling af ler-, silt- og sandfraktionen i jorde i caseområdet i et multibånd-raster format normaliseret ift. hinanden. Heri repræsenteres ler som den røde farve, silt som den grønne og sand som den blå. Dominans af den enkelte type vil repræsenteres ved at netop den farve vil fremstå som den primære, og ellers vil den viste farve være en kombination af de tilstedeværende jorde i området.



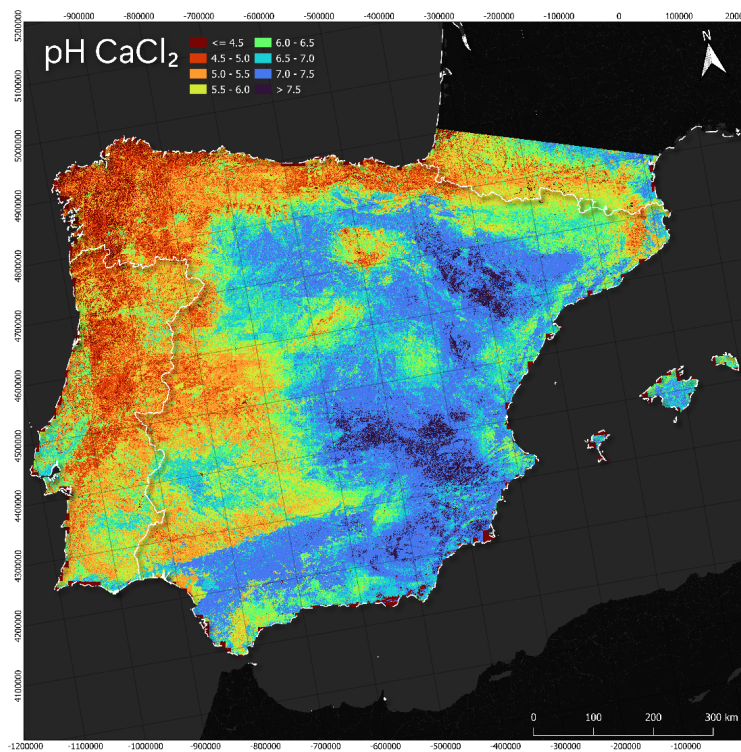
Figur 8.4. Beregning af jordfraktions-ratioen, for at kunne vurdere hvor og hvor meget modellen over og eller undervurderer det samlede indhold af ler, silt og sand. Størstedelen af caseområdet afviger med under 5% fra en samlet jordfraktion af 100%.



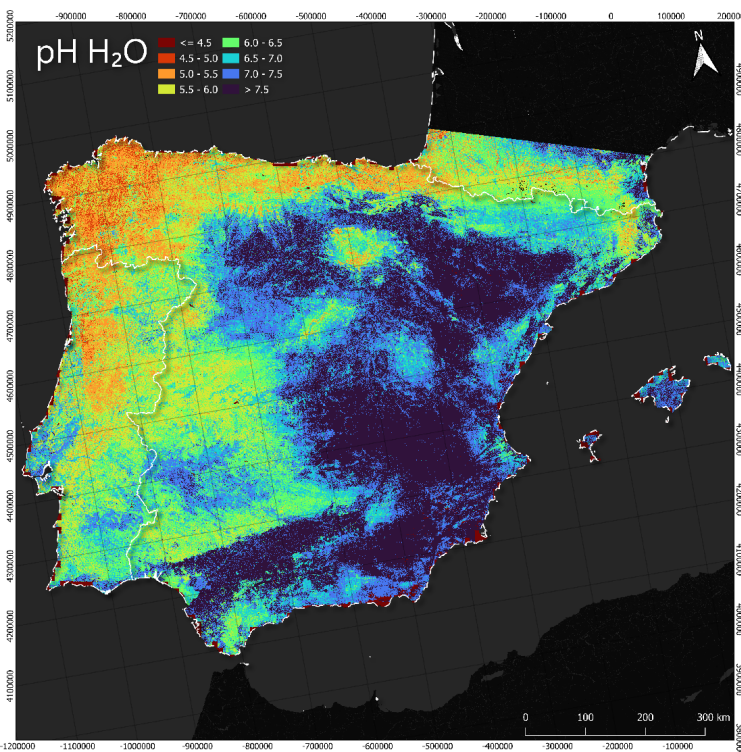
(a)



(b)

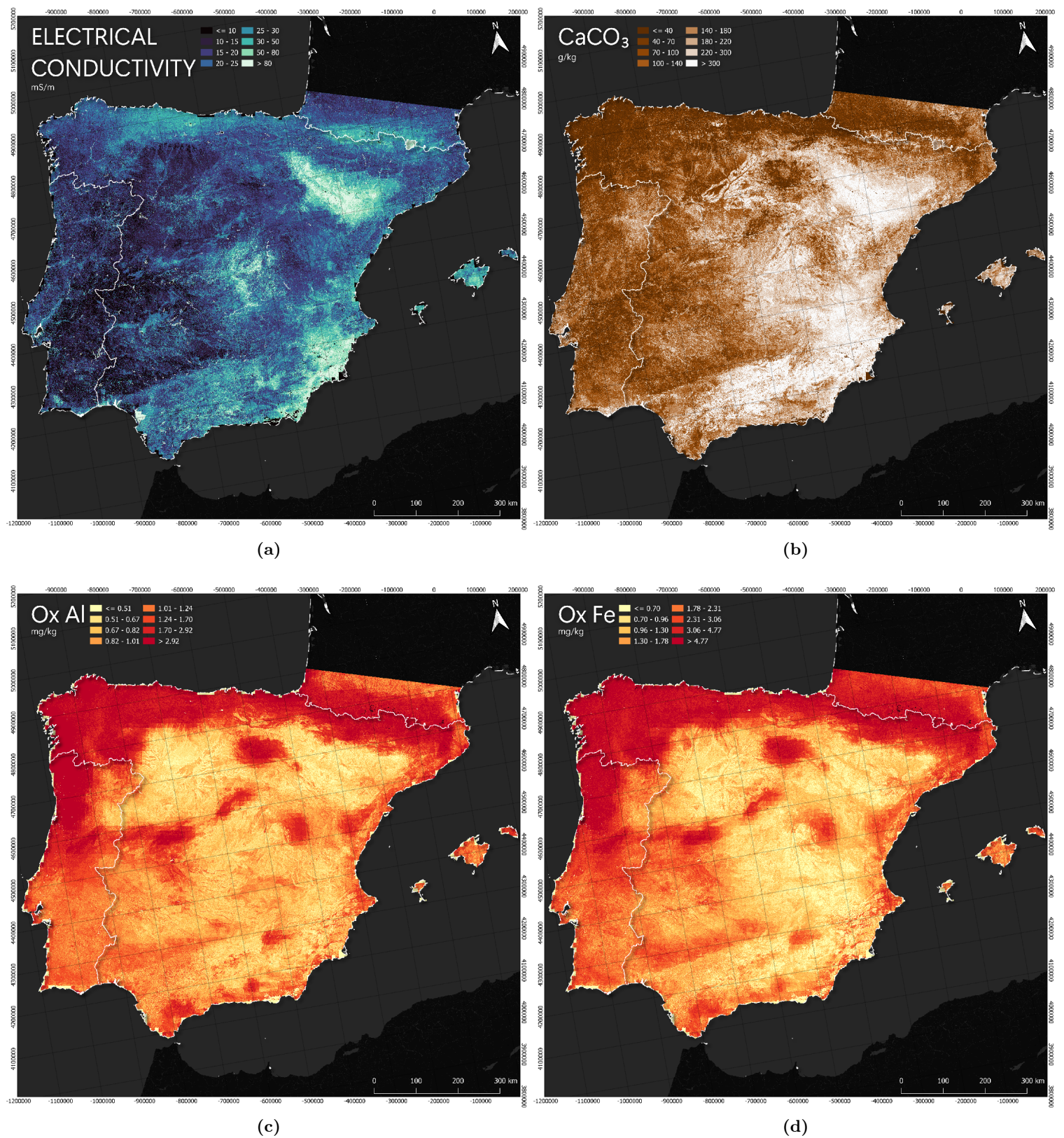


(c)

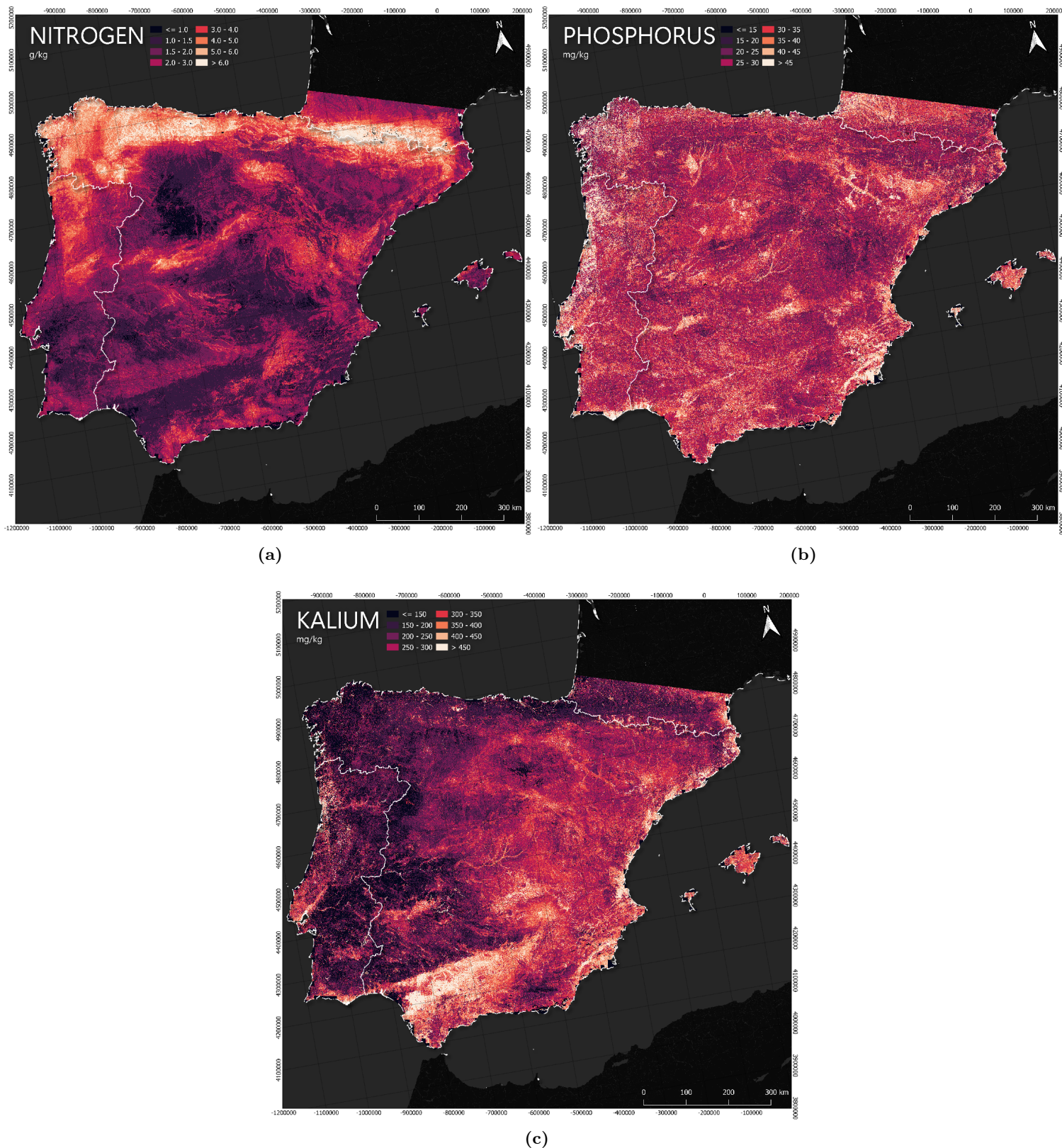


(d)

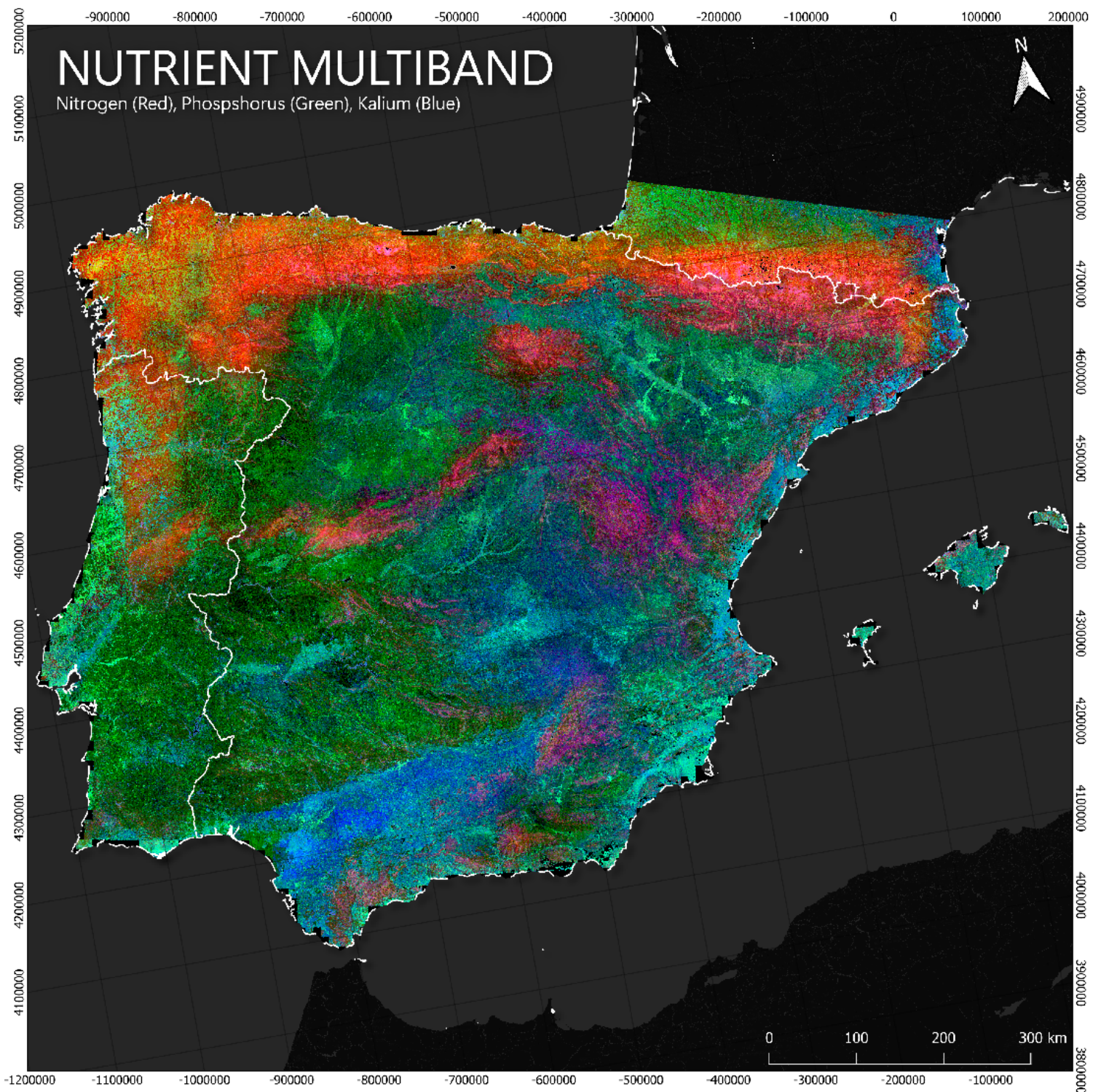
Figur 8.5. Kortlægning af prædikeret *organiskkulstof* (a), *volumenvægt* (b), *pH* CaCl_2 (c) og *pH* H_2O (d).



Figur 8.6. Kortlægning af prædikeret EC (a), $CaCO_3$ (b), $OxAl$ (c) og $OxFe$ (d).



Figur 8.7. Kortlægning af prædikteret *Nitrogen* (a), *Fosfor* (b) og *Kalium* (c). En kombination af nitrogen, fosfor og kalium sat overfor hinanden i et multispektralt bånd, dvs. RGB (rødt, grønt, blå) kan ses i figur 8.8.

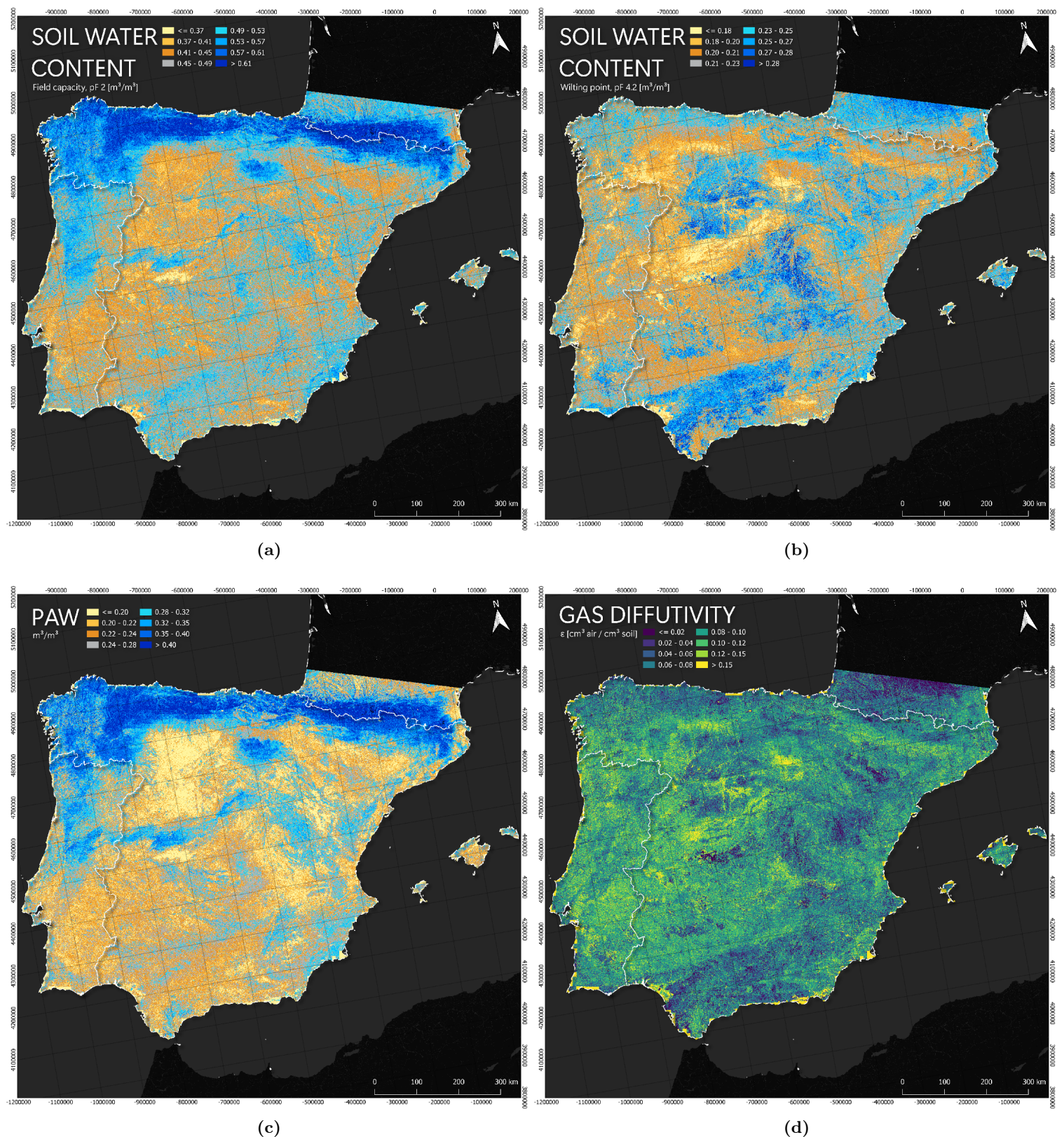


Figur 8.8. Formidling af nitrogen, fosfor og kalium i jorde i caseområdet i et multibånd-raster format justeret ift. hinanden. Heri repræsenteres nitrogen som det røde bånd, fosfor som det grønne og kalium som det blå. Dominans af den enkelte type vil repræsenteres ved at netop den farve vil fremstå med højeste styrke, og ellers vil den viste farve være en kombination af de tilstedeværende jorde i området. Minimum/maksimum skalaene for rød er 0 – 10 *g/kg*, 0 – 100 *mg/kg* for grøn, og 0 – 1000 *mg/kg* for blå.

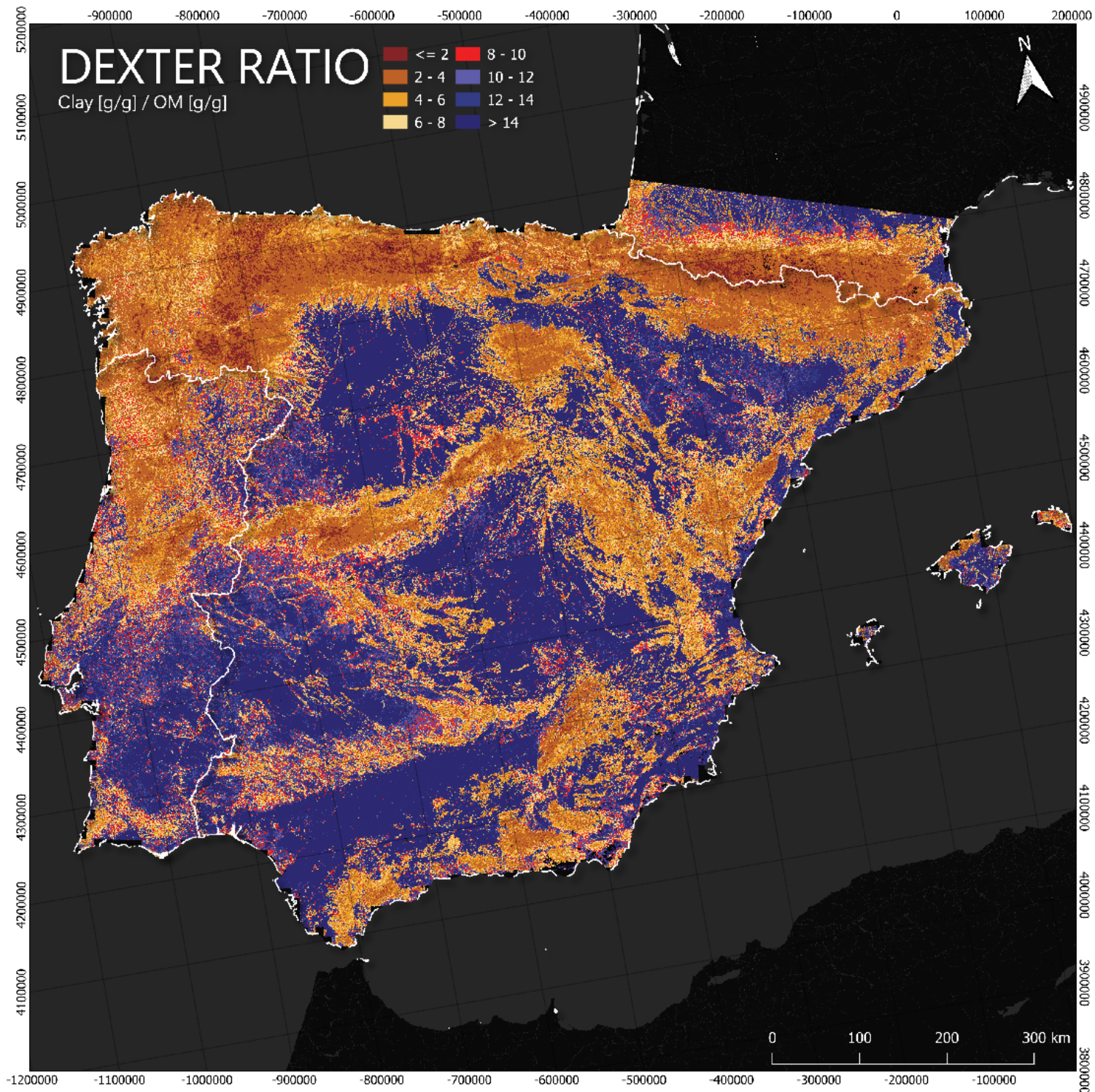
8.2 Sundhedsmæssige jordbundsparametre

Ved anvendelse af data som formidlet i afsnit 8.1 og i kombination med pedotransfer-funktioner og ligninger fra kapitel 4, er der udarbejdet kort for forskellige jordparametre. Disse kort inkluderer vandindholdet ($\frac{m^3 \text{ vand}}{m^3 \text{ jord}}$) ved pF 2 og 4,2, plantetilgængeligt vand ($\frac{cm^3 \text{ vand}}{cm^3 \text{ jord}}$), Dexter-ratioen ($\frac{L}{OC}$), jordens gasdiffusivitet ved $-100 \text{ cm } H_2O$ ($\frac{D_{F,100}}{D_o}$), hydraulisk ledningsevne ved vandmætning (mm/h) samt totalporøsiteten af jorden ($\frac{cm^3 \text{ porer}}{cm^3 \text{ jord}}$). Disse parametre giver et overblik over jorden funktionalitet og bidrager til en helhedsforståelse af jordens sundhedstilstand.

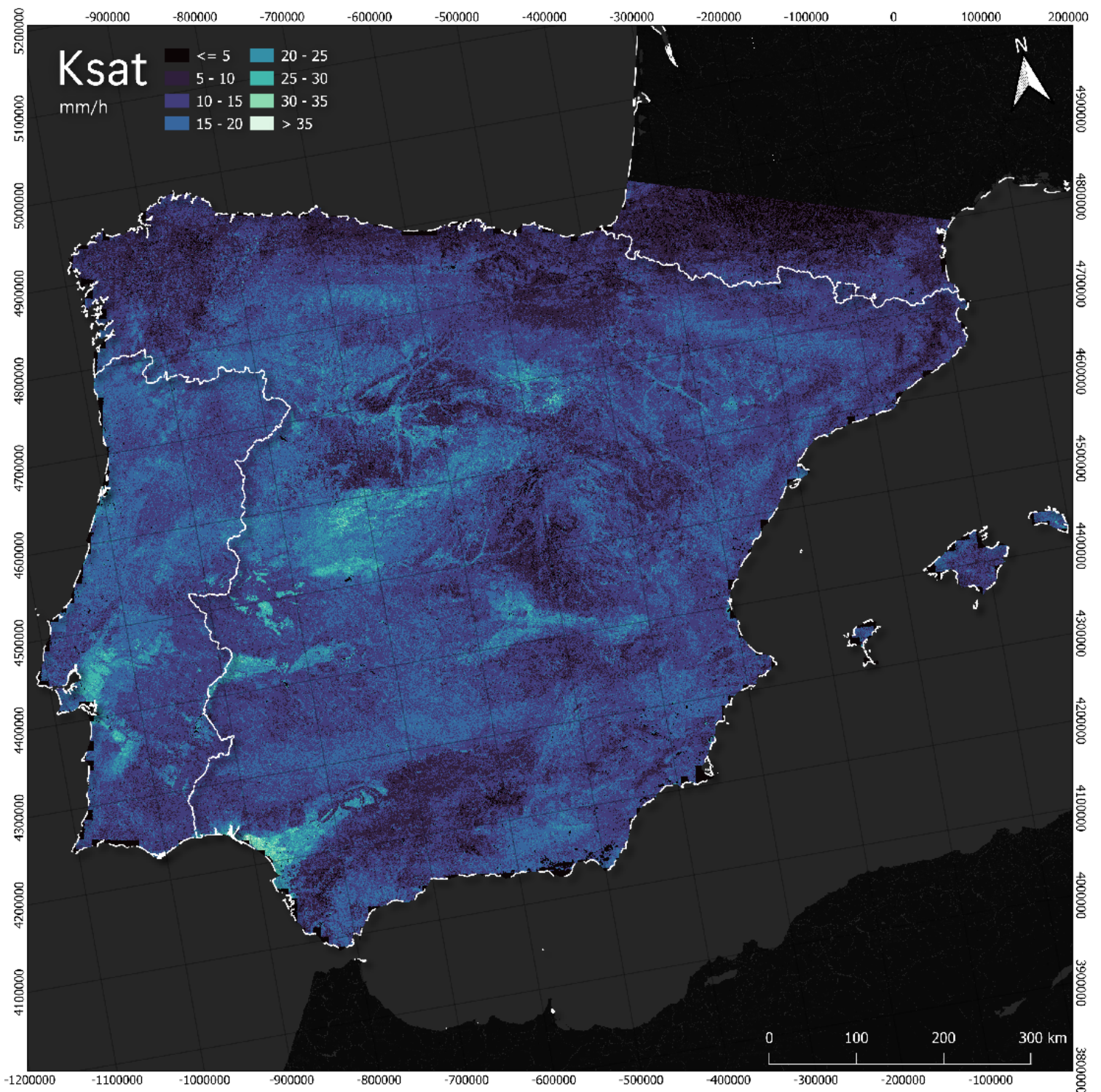
De syv pedotransfer-baserede sundhedsparametre vil, ligesom de 15 andre prædikterede jordparametre, blive præsenteret i geospatiale kort. Kortlægningen giver mulighed for at visualisere den rumlige fordeling af parametrene, hvilket er afgørende for at vurdere jordens sundhed på tværs af forskellige regioner. Som tidligere nævnt i afsnit 8.1, ses der igen tendenser til både under- og overvurderinger for disse parametre, som også kan observeres i figurerne 8.13 og 8.14. Dette er forventeligt, da de sundhedsmæssige parametre baserer sig på de prædikterede jordparametre. Det betyder også, at disse under- og overvurderinger kan forstærkes ved ekstreme værdier, hvilket bør tages i betragtning ved tolkning af resultaterne.



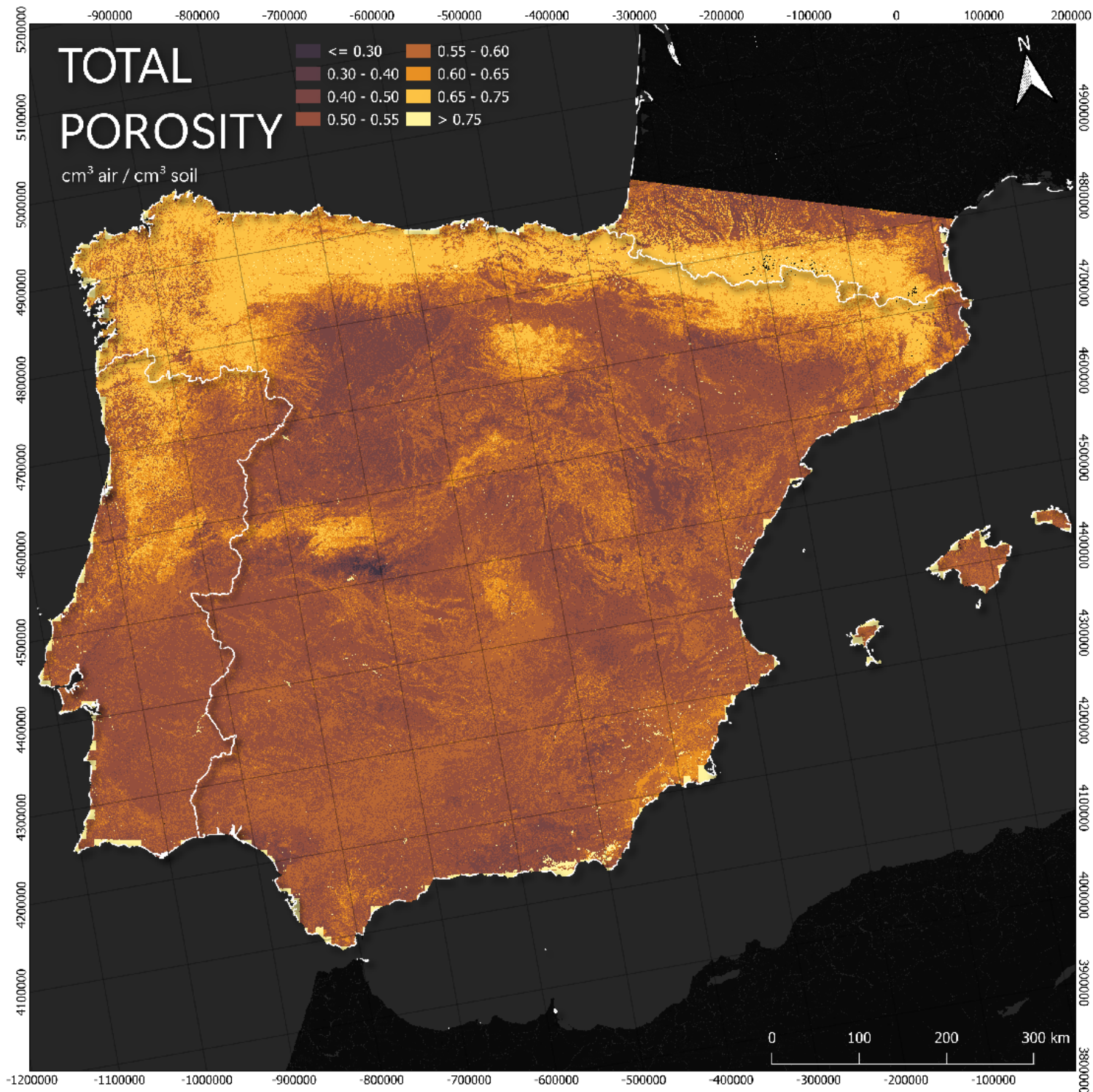
Figur 8.9. Kortlægning af estimeret vandindhold ved $pF\ 2$ og $4,2$, samt kortlægning af plantetilgængeligt vand og den estimerede gasdiffusivitet for området.



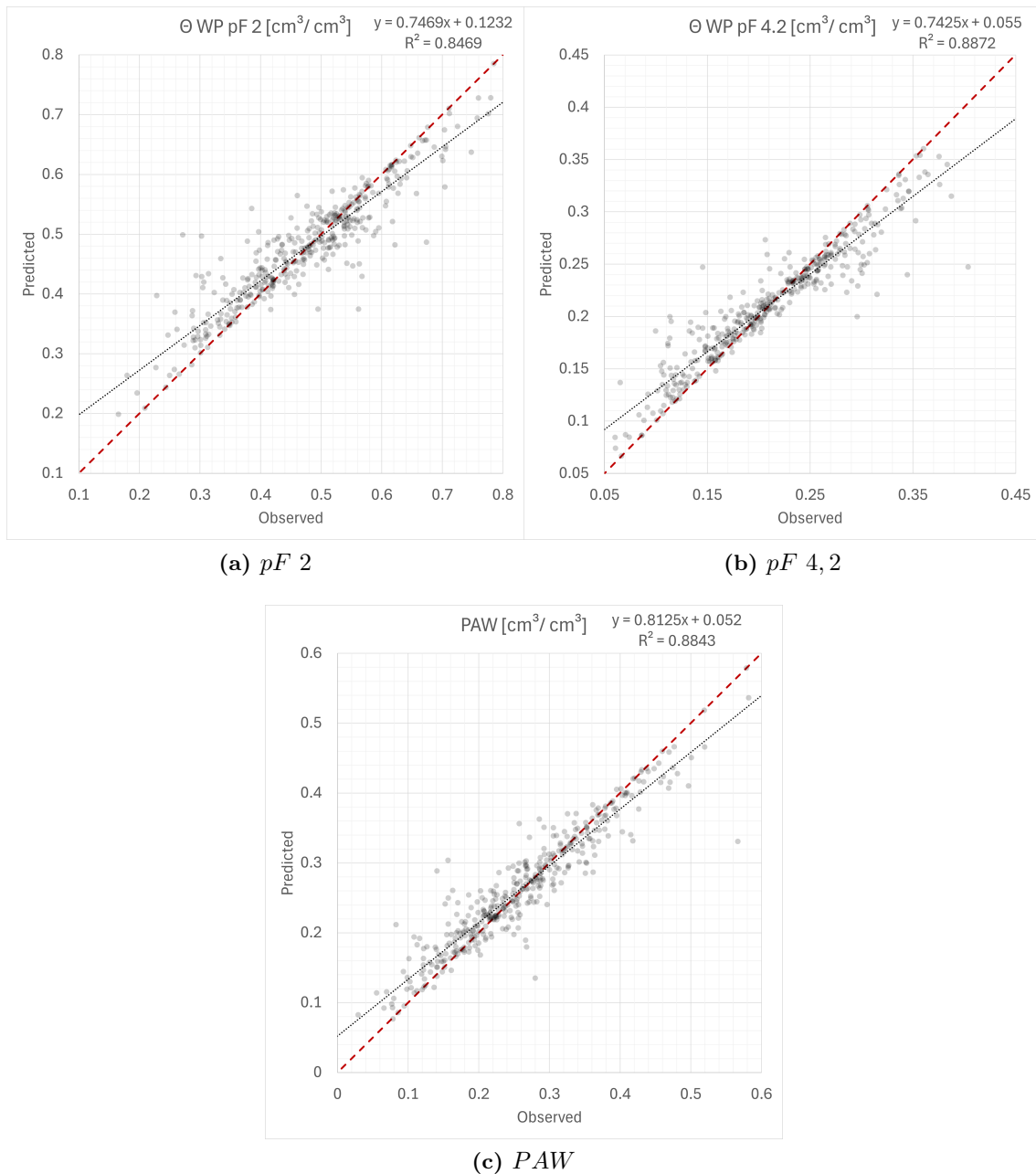
Figur 8.10. Kortlægning af estimeret dexter ratio for hele case området via de prædiktioner for lerindhold og organisk kulstof.



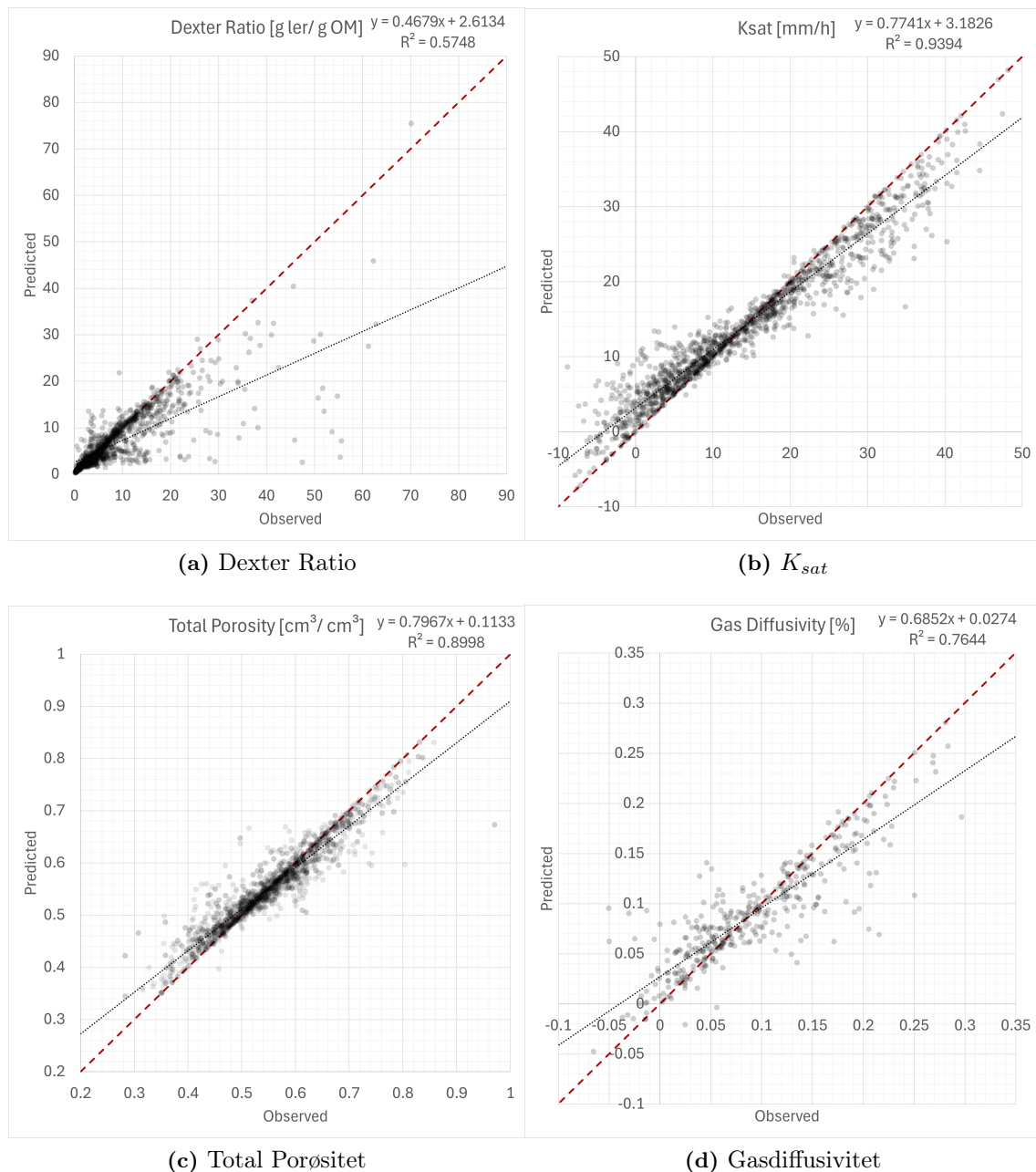
Figur 8.11. Kortlægning af jordens mættede hydrauliske ledningsevne (K_{sat}) ud fra de prædikterede værdier for hele case området.



Figur 8.12. Kortlægning af jordens totalporøsitet (ϕ) ud fra de prædikterede værdier for hele case området. Da total porøsiteten er beregnet ud fra volumenvægten (ρ_b), og at ρ_b er korreleret med indholdet af OC , er det som forventet, at ϕ følger samme trend. Heri ses det derfor, at for blandt andet de nordliggende regioner, hvor OC også er prædikteret til at være høj, at ϕ også er høj.



Figur 8.13. 1:1 grafer til sammenligning af resultaterne af de anvendte pedotransfer funktioner, fra kapitel 4, på observerede og prædikterede data for henholdsvis $pF\ 2$, $pF\ 4,2$ og PAW . For alle tre grafer ses en R på mindst 0,84, men en kronisk overvurdering af dataene ved lave værdier og omvendt ved høje.



Figur 8.14. 1:1 grafer til sammenligning af de fire andre parametre, som blev estimeret ud fra pedotransfer funktioner som oplyst i kapitel 4. De følger lignende tendenser som den direkte præsikterede jordparametre, i at der typisk overvurderes ved meget lave værdier og undervurderes ved højere værdier. Resultaterne for 1:1 sammenligningen af Dexter Ratioen skiller sig ud ved at give ikke favorable resultater, idet blandt andet R er under 0,6. Dette må formodes at være forskyldt af imprecisioner i OC/ler-prædiktionsresultaterne.

Tabel 8.1. Korrelationsmatrice mellem resultaterne opnået via pedotransfer funktioner på de observerede og prædikterede data.

	Dexter Ratio	$\theta_{FC\ pF\ 2}$	$\theta_{WP\ pF\ 4.2}$	PAW	Gas Diffusivity	K_{sat}	Total Porosity
Dexter Ratio	0.761	-0.140	0.423	-0.477	-0.208	-0.289	-0.346
$\theta_{FC\ pF\ 2}$	-0.139	0.920	0.462	0.714	-0.402	-0.642	0.778
$\theta_{WP\ pF\ 4.2}$	0.469	0.422	0.942	-0.145	-0.768	-0.903	-0.019
PAW	-0.495	0.779	-0.127	0.940	0.074	-0.113	0.926
Gas Diffusivity	-0.262	-0.423	-0.819	0.064	0.874	0.867	0.086
K_{sat}	-0.307	-0.613	-0.891	-0.016	0.842	0.969	-0.049
Total Porosity	-0.403	0.795	-0.088	0.872	0.264	-0.111	0.949

På baggrund af statistikkerne beregnet for de parametre, der er estimeret ud fra pedotransferfunktioner, ses det, at parameteren Dexter Ratio opnår tildels ringe resultater. Her opnår Dexter Ratio blandt andet en *RMSE* på 11,7, hvilket betyder at den egentlige dexter ratio for det givne punkt, kan variere stort, hvorfor hvis der estimeres en dexter ratio på eksempelvis 15, så kan den reelle værdi ligge et sted mellem 3,3 og 26,7, hvilket har stor betydning. Det skal dog tages in mente, at denne statistik igen kun er for de steder, hvor der eksisterer LUCAS jordprøver, hvilket kan spille en betydelig rolle. Der er derfor særligt høj usikkerhed på dexter ratio estimeringerne. De øvrige parametre opnår meget små *RMSE*, og vurderes som acceptable i forhold til repræsentation for hvordan de virkelige forhold kan se ud. Derudover er der generelt meget lav bias, RPD og standardafvigelse for de øvrige seks parametre relativt til deres enheder.

Udover statistik for blandt andet RMSE, RPD med videre som vist i tabel 8.2, er der også lavet boksplots for de seks sundhedsparametre, som anvendes til multikriterieanalysen. Disse kan ses og aflæses fra figur 8.15 og de respektive subfigurerer. Til boksplotsene for parametrene er også de konkrete tal for flere af elementerne som indgår i boksplots, såsom minimum, maksimum, kvartiler med videre, hvilket kan aflæses i tabel 8.3 for de respektive parametre. De producerede boksplots kan blandt andet hjælpe til at forklare datene, såsom ved at visualisere outliers i datasættet, samt hvor kvartilerne ligger, samt middel og median værdierne. Yderligere for boksplotsene som lavet hertil, er der inkluderet rangeringstærsklerne, i form af farver til baggrund der afspejler rangeringsfarven, som defineret i tabel 8.4, hvilket kan hjælpe med at visualisere hvor værdierne i datasættet ligger for den respektive parameter. Som eksempel kan det ses i figur 8.15 *f*, at kun punkter over det 75. kvartil for EC opnår over acceptable rangeringer, men at dette felt hovedsageligt domineres af outliers. Dette afspejles også i forbindelse med kortlægningen

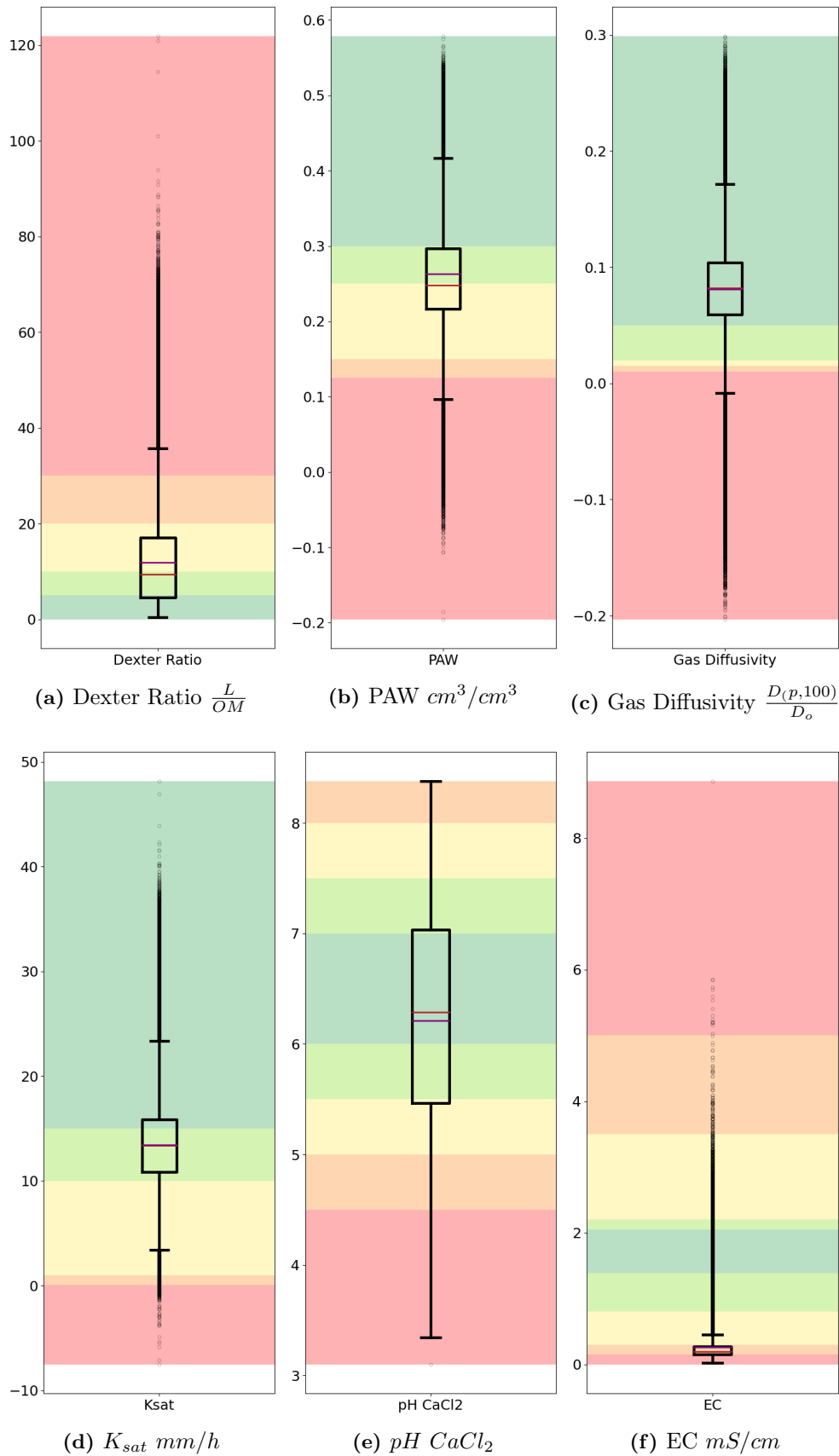
af rangeringerne, som for EC finder sted i figur 8.21.

Tabel 8.2. Statistik for RMSE, RPD, Bias og standardafvigelsen (STD) på de prædikterede parametre fra pedotransferfunktionerne.

	$\frac{L}{OM}$	cm^3/cm^3	cm^3/cm^3	cm^3/cm^3	$\frac{D(p,100)}{D_o}$	mm/h	cm^3/cm^3
	<i>Dexter Ratio</i>	θ_{FC}	θ_{WP}	θ_{PAW}	<i>Gas Diffusivity</i>	K_{sat}	<i>Total Porosity</i>
<i>RMSE</i>	6,775	0,047	0,026	0,034	0,034	3,459	0,033
<i>RPD</i>	1,455	2,458	2,709	2,860	2,025	3,338	2,986
<i>BIAS</i>	6,347	0,003	0,000	0,003	-0,001	0,088	0,001
<i>STD_{obs}</i>	9,858	0,116	0,070	0,099	0,068	11,548	0,097
<i>STD_{pred}</i>	6,084	0,094	0,055	0,085	0,054	9,223	0,082

Tabel 8.3. Statistik for de endelige sundhedsparametre som anvendes til multikriterieanalysen på blandt andet minimum og maksimum værdierne, samt øvre og nedre "whiskers" med videre, og hvor *IQR* står for *interquartile range* der beskriver selve boksens bredde.

	Min	Max	Mean	Median	Q1	Q3	IQR	Lower whisker	Upper whisker
Dexter Ratio $\frac{L}{OM}$	0.384	121.893	11.823	9.410	4.608	17.038	12.430	0.384	35.683
PAW cm^3/cm^3	-0.196	0.579	0.263	0.248	0.217	0.297	0.080	0.097	0.417
Gas Diffusivity $\frac{D(p,100)}{D_o}$	-0.203	0.299	0.081	0.082	0.059	0.104	0.045	-0.008	0.172
K_{sat} mm/h	-7.506	48.183	13.446	13.373	10.882	15.872	4.990	3.397	23.357
pH $CaCl_2$	3.100	8.382	6.213	6.288	5.464	7.033	1.569	3.345	8.382
EC mS/m	2.235	886.932	25.949	19.593	15.202	27.120	2.235	2.235	44.997



Figur 8.15. Boksplot grafer for hver af sundhedsparametrene som bliver anvendt i forbindelse med multikriterieanalysen. EC er omregnet fra mS/m til mS/cm for at passe til rangeringerne. For konkrete tal for de forskellige elementer i boksplotsene henvises der til tabel 8.3.

8.3 Multikriterieanalyse af faktorer for jordbundssundhed

Nedenfor er opsat den generelle ligning, der beskriver, hvordan en multikriterieanalyse beregnes ud fra ét eller flere inputparametre, eller rasterlag i denne kontekst. Hver inputparameter tildeles en rangering samt en vægtningsværdi. Rangeringen af inputparametrene, eller rasterlagene i denne kontekst, kan variere afhængigt af deres placering på kortet.

Dette skyldes, at rasterlagene består af et gitter af celler, hvor hver celle indeholder en specifik værdi, der repræsenterer en egenskab i den pågældende del af området. Dette kan også blot kendetegnes som et raster. Værdierne af cellerne er uafhængige af hinanden, og repræsenterer udelukkende sin egen værdi. Værdierne for cellerne kan derfor være forskellige for diverse steder eller være ens.

Den generelle ligning for *Weighted Sum* lyder derfor således, som opskrevet i ligning 8.1, hvor der blot indsættes parameter, det vil sige raster, og den udvalgte vægt hertil.

$$MKA_{ws} = \sum P_R \cdot P_W \quad (8.1)$$

Ligningen viser hvordan inputparameter indgår med en rangering P_R , der ganges med sin respektive vægt P_W , og at disse produkter summeres ved hjælp af Weighted Sum.

Weighted Sum er en metode, der bruges til at kombinere flere rasterlag ved at vægte hvert lag i forhold til dets betydning for analysen. I en multikriterieanalysen anvendes Weighted Sum metoden til at aggregere værdierne fra forskellige inputlag og skabe ét samlet resultat, hvor hvert lag påvirker den endelige værdi [ArcGIS, u.d.]. Dette gør værktøjet særligt nyttigt hvor flere faktorer skal tages i betragtning, hvor hver parameter kan tildeles forskellig vægtning, altså betydning for jordens sundhed i denne kontekst, hvorfor denne metode anvendes til at lave en samlet vurdering af jordbundssundheden.

8.3.1 Rangering af sundhedsparametrene

Multikriterieanalysen anvender centrale parametre såsom Dexter ratio, plantetilgængeligt vand (PAW), gasdiffusivitet, vandmættet hydraulisk ledningsevne (K_{sat}), surhedsgraden (pH) og elektrisk ledningsevne (EC). Disse parametre er udvalgt for at illustrere, hvordan numeriske sundhedsindikatorer kan benyttes til at vurdere jordens tilstand.

Jordens sundhed og plantevækst vurderes dermed ud fra disse centrale jordparametre, som rangeres på en skala fra 0 til 4. Hver sundhedsparameter er uddybet i teoriafsnittet (se kapitel 4), mens rangeringen af hver parameter er baseret på deres respektive tærskelværdier for sund plantevækst. Disse tærskelværdier danner grundlaget for rangeringen fra 0 til 4. Rangeringen 0 indikerer optimale forhold for plantevækst, mens en rangering på 4 indikerer mere kritiske tilstande, der kan hæmme jordens evne til at understøtte sund plantevækst og andre vigtige funktioner i jorden. Derudover kan sundhedsindikatorerne vurderes forskelligt i forhold til hvor vigtige de er for den samlede vurdering af jordens tilstand. For eksempel vægtes pH som en af de mest betydningsfulde faktorer, da det påvirker jordens mikrobiologiske processer og har en betydelig indvirkning

på næringsstofoptagelsen i planter. Derfor får pH en vægtning på 2, mens de øvrige parametre vægtes med en faktor på 1.

De anvendte tærskelværdier og intervaller til rangeringerne er baseret på subjektive vurderinger ud fra den eksisterende teori. Disse rangeringer og vægtninger skal derfor ikke betragtes som en præcis eller absolut måling af jordens sundhed, men snarere som et konceptuelt værktøj, der giver en indikation af jordens tilstand baseret på de inkluderede parametre. Dermed er analysen udformet som et koncept, der er med til at illustrere hvordan disse sundhedsparametre kan bruges til vurderinger og kortlægning på større skala, og hvordan en teoretisk model kan anvendes til at vurdere jordens tilstand.

I det følgende vil der være en kort forklaring af jordparametrene hver især og hvorfor de er rangeret som de er 0-4. Den resulterende tabel med deres endegyldige rangeringer er opstillet i tabel 8.4.

Dexter Ratio

Dexter Ratio $[\frac{g\text{ }Ler}{g\text{ }OM}]$ angiver forholdet mellem lerindhold og organisk kulstof, hvilket er en indikator for jordens struktur og dens evne til at danne stabile aggregater. Lav Dexter Ratio (< 10) er fordelagtig for god struktur og reducerer risikoen for erosion. Derimod vil en højere ratio (> 10) indikere at der er en forringet jordstruktur og kompaktering, som hæmmer vandbevægelsen. Rang 0-4: En værdi på ≤ 5 indikerer optimal struktur (rang 0), mens en værdi over 30 (rang 4) indikerer kritisk tilstand med dårlig jordstruktur.

Plantetilgængeligt vand (PAW)

Plantetilgængeligt vand (PAW) $[\frac{cm^3\text{ vand}}{cm^3\text{ jord}}]$ er mængden af vand, der er tilgængelig for planter. Hvis PAW er for lav ($\leq 0,15$), kan planterne opleve vandstress, mens høj PAW ($> 0,15$) er med til at sikre en god plads til opmagasinerings og dermed rigelig vandforsyning til vækst. Rang 0-4: Høj PAW ($> 0,3$) får rang 0, mens lav PAW ($\leq 0,125$) får rang 4, da det kan føre til utilstrækkelig vandforsyning.

Gasdiffusivitet

Gasdiffusivitet $[(\frac{D}{D_o})]$ beskriver jordens evne til at tillade gasudveksling (ilt og CO₂) gennem porehulrummene i jorden, hvilket er essentielt for røddernes respiration og mikrobiologisk aktivitet. En lav gasdiffusivitet ($\leq 0,01$) tyder på dårlig luftudveksling, hvilket kan medføre iltfattige forhold. Rang 0-4: Høj gasdiffusivitet ($> 0,05$) får rang 0, mens lav diffusivitet ($\leq 0,01$) får rang 4, som indikerer iltmangel.

Vandmættet hydraulisk ledningsevne (K_{sat})

Den vandmættede hydrauliske ledningsevne (K_{sat}) måler jordens evne til at transmittere vand, når den er fuldt vandmættet. Optimal dræningsevne (15-85 mm/h) får rang 0, mens meget lav (< 1 mm/h) eller meget høj (> 100 mm/h) K_{sat} kan føre til hhv. vandmætning eller udvaskning af næringsstoffer. Rang 0-4: Værdier mellem 15-85 mm/h rangeres som 0, mens ekstremt lave/høje værdier får rang 4.

Surhedsgrad ($pH\ CaCl_2$)

Jordens pH er kritisk for næringsstofftilgængelighed og mikroorganismers aktivitet. En pH-værdi mellem 6-7 er ideel (rang 0), mens værdier under 4,5 pH eller over 8,5 pH kan føre til næringsstoffmangel eller toksicitet (rang 4).

Elektrisk ledningsevne (EC)

EC måles i [mS/cm] og er et mål for saltholdigheden i jorden. Moderat saltniveau (1,39-2,05 mS/cm) er optimalt (rang 0), mens meget lave eller høje niveauer kan forårsage hhv. næringsstoffmangel eller saltstress. Rang 0-4: Lav EC ($\leq 0,15$ mS/cm) eller høj EC (> 5 mS/cm) får rang 4, hvilket indikerer problematiske forhold for plantevækst.

8.3.2 Vurdering af jordens sundhed ud fra jordparametrene

I følgende tabel 8.4 ses en oversigt over de udvalgte intervaller og hvilke rangeringer de giver, for hver anvendte jordparameter. Yderligere ses også den udvalgte vægtning for hver af dem. Her er den eneste parameter, som har en anden vægtning end 1 $pH\ CaCl_2$, der får en vægtning på 2. Dette er fordi surheden af jorden spiller en særligt signifikant rolle i forhold til netop jordens habitabilitet for planter. Eksempelvis, og som nævnt i afsnit 4.6, har pH som nævnt indflydelse på tilgængeligheden af både nødvendige næringsstoffer for plantevækst og -overlevelse, men også indflydelse på tilgængeligheden af særligt giftige stoffer i jorden, som kan opnå fatale koncentrationer ved eksempelvis for syrlige jordforhold - dette er blandt andet tilfældet ved aluminium, som bliver mere tilgængelig i syrlig jord. Omvendt medfører basiske jorde ofte til at nødvendige næringsstoffer bliver utilgængelige, da de binder sig til andre mineraler og ioner i jorden.

Tabel 8.4. Rangeringer af de udvalgte jordparametre, hvor $\wedge$ er *og*, og $\vee$ er *eller*.

Soil Parameters	Ranks					Weight
	0	1	2	3	4	
Dexter Ratio $\frac{g\ Ler}{g\ OM}$	≤ 5	≤ 10	≤ 20	≤ 30	> 30	1
PAW $\frac{cm^3\ vand}{cm^3\ jord}$	$> 0,3$	$> 0,25$	$> 0,15$	$> 0,125$	$\leq 0,125$	1
Gas Diffusivity $\frac{D(p,100)}{D_o}$	$> 0,05$	$> 0,02$	$> 0,015$	$> 0,01$	$\leq 0,01$	1
$K_{sat}\ mm/h$	$< 85 \wedge \geq 15$	$< 100 \wedge \geq 10$	$\geq 100 \vee < 10$	< 1	$< 0,1$	1
$pH\ CaCl_2$	$> 6 \wedge < 7$	$\leq 6 \vee \geq 7$	$\leq 5,5 \vee \geq 7,5$	$\leq 5 \vee \geq 8$	$\leq 4,5 \vee \geq 8,5$	2
EC mS/cm	$> 1,39 \wedge \leq 2,05$	$> 0,8 \wedge \leq 2,2$	$> 0,3 \wedge \leq 3,5$	$\leq 0,3 \vee > 3,5$	$\leq 0,15 \vee > 5$	1

Parametrene kan derfor med deres vægte indsættes i en endelig ligning til udregning af multikriterieanalysen. Denne ligning vil lyde som 8.2 herefter, hvor rangeringerne og vægtene for hver parameter er defineret i tabel 8.4, hvor $A = Dexterratio$, $B = PAW$, $C = Gasdiffusivitet$, $D = K_{sat}$, $E = pH\ CaCl_2$ og $F = Elektriskkonduktivitet$ i samme rækkefølge som i tabel 8.4.

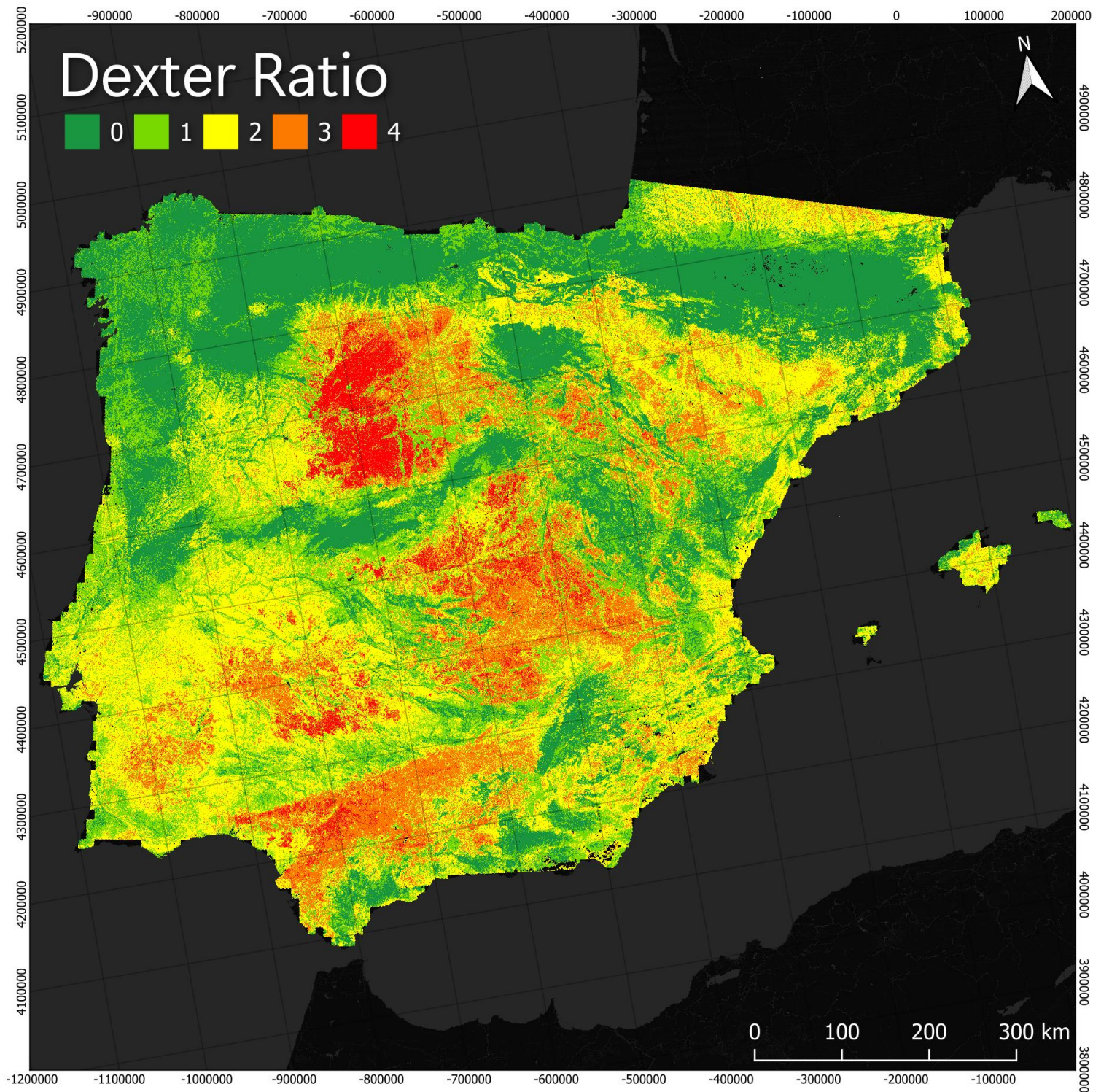
$$MKA = \sum A_R \cdot A_W + B_R \cdot B_W + C_R \cdot C_W + D_R \cdot D_W + E_R \cdot E_W + F_R \cdot F_W \quad (8.2)$$

Før den definitive og samlede multikriterieanalyse kan samles og laves, skal der beregnes rangerings-scores for de seks udvalgte jordbundssundhedsparametre ud fra intervallerne og kravene som fastsat i tabel 8.4 herover. Dette gøres ved en simpel klassificering af cellerne, dvs. pixel, for hver af kortene. Her tildeles cellerne den respektive score, afhængigt af den prædikterede værdi for parameteren. Eksempelvis for klassificering af pH, hvis der er prædikeret til at være en pH på 7, klassificeres denne til en score på 1, da den er ≥ 7 . Når hver jordparameter er blevet klassificeret ud fra hvad der står i tabel 8.4, anvendes ligning 8.2 til at samle dem til ét kort, altså den endelige multikriterieanalyse, eller med andre ord; det endelige kort, der giver et konkret og kvantitativt bud på jordbundssundheden for caseområdet.

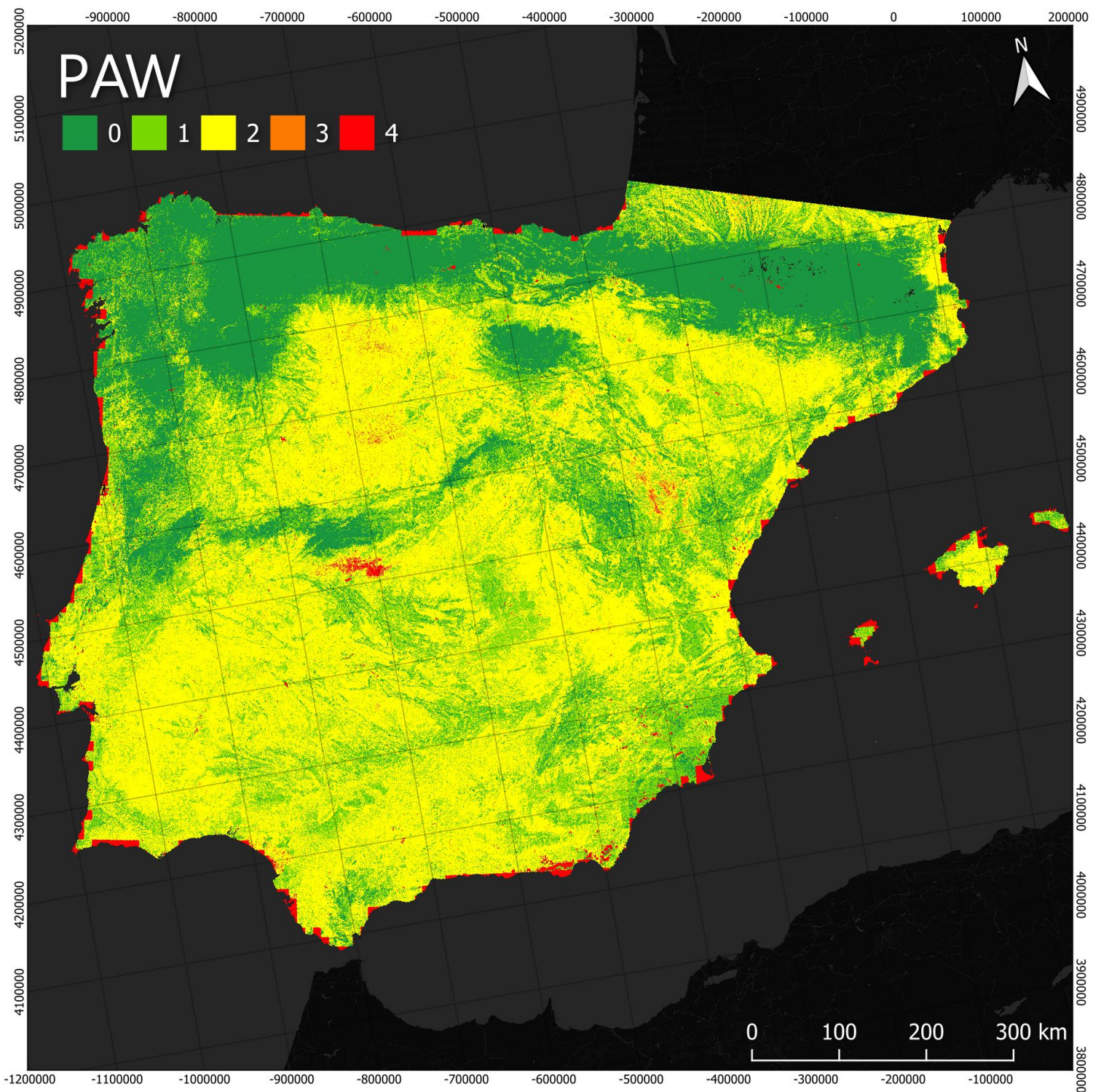
I de følgende figurer 8.16, 8.17, 8.18, 8.19, 8.20 og 8.21 præsenterer de kortlagte klassificerede scorer for jordparametrene, i samme rækkefølge som i tabel 8.4.

Et element der kan udledes fra sundhedsklassificeringen af Dexter Ratio for området er, at den optimale Dexter Ratio i større omfang følger de steder, hvor PAW opnår gode resultater. Kortet for Dexter Ratio rangeringer kan ses i figur 8.16, og for PAW kan de ses i figur 8.17. Ses der tilbage på eksempelvis kortet for köppen klimaklassifikationer som vist i figur 3.3, tyder det ikke på, at dexter ratioen har signifikant korrelation med klimaet, men det kan ikke udelukkes.

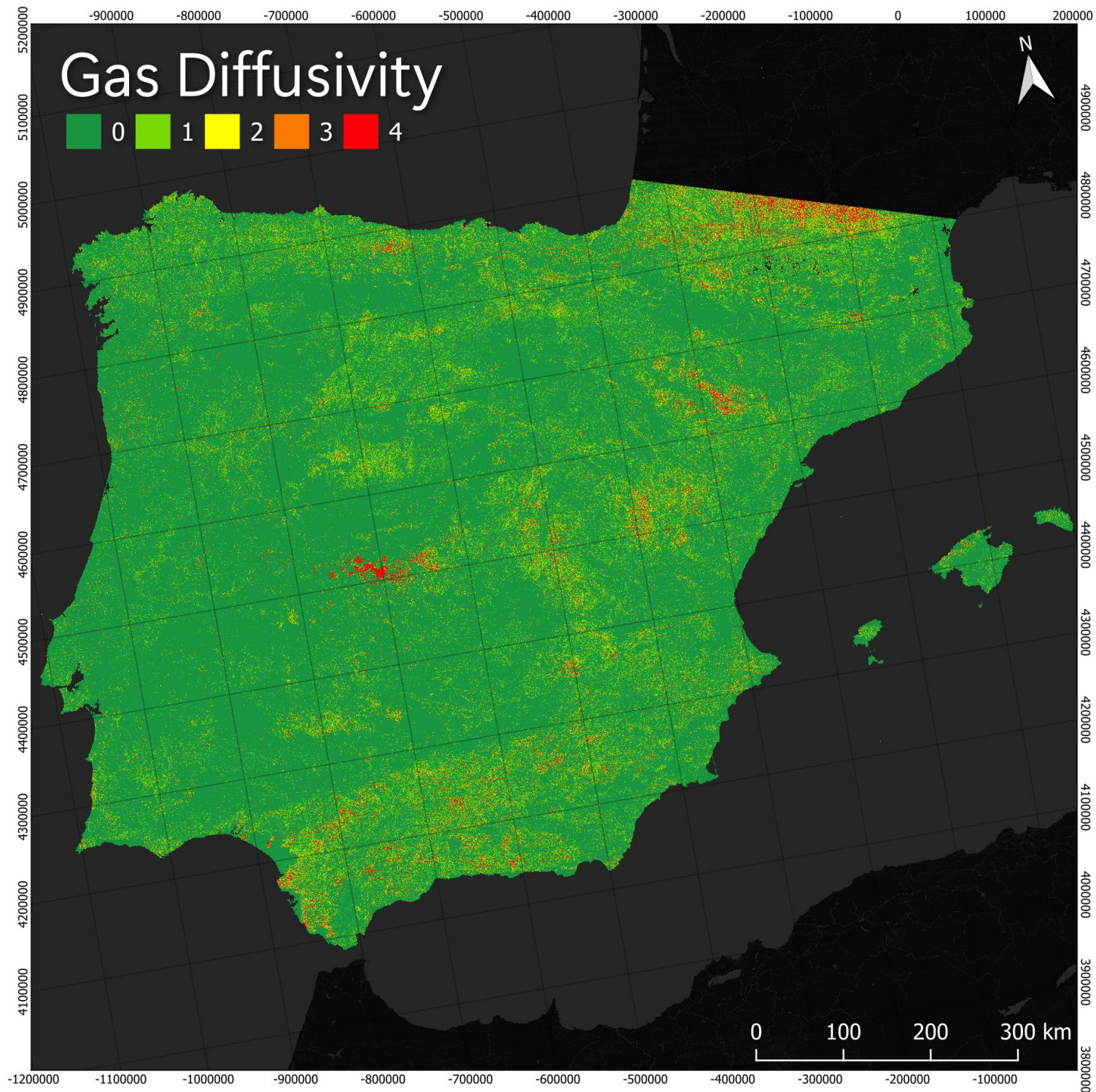
Ved nærmere undersøgelse kan der eksempelvis henvises tilbage til blandt andet de beregnede korrelationsværdier mellem de prædikterede jordparametre med klima- og CORINE-inputparametre i tabel 7.10. Her ses eksempelvis en moderat negativ sammenhæng på $r = -0,490$ mellem ler CORINE-Code, altså mellem det prædikterede lerindhold og klassifikationen af arealanvendelsen for det givne område, og tildels også med planteproduktionen, DMP med ligeledes moderat negativ sammenhæng på $r = -0,424$. OC, som indgår i estimering af Dexter Ratioen, har derimod større positiv korrelation med klimamæssige faktorer, herunder særligt med DMP ($r = 0,535$) og CORINE-Code ($r = 0,573$), men også mængden af nedbør ($r = 0,572$), hvor højere mængder nedbør dermed korrelerer med højnet OC [Copernicus Climate Change Service (C3S), u.d.], og deraf, som set i figur 8.16 bedre dexter ratio rangering. De fleste af de steder, hvor dexter ratioen derfor opnår gode rangeringer, dvs. under 2, vægter tildels til skråninger, højlande, bjerge, skovområder og lignende områder.



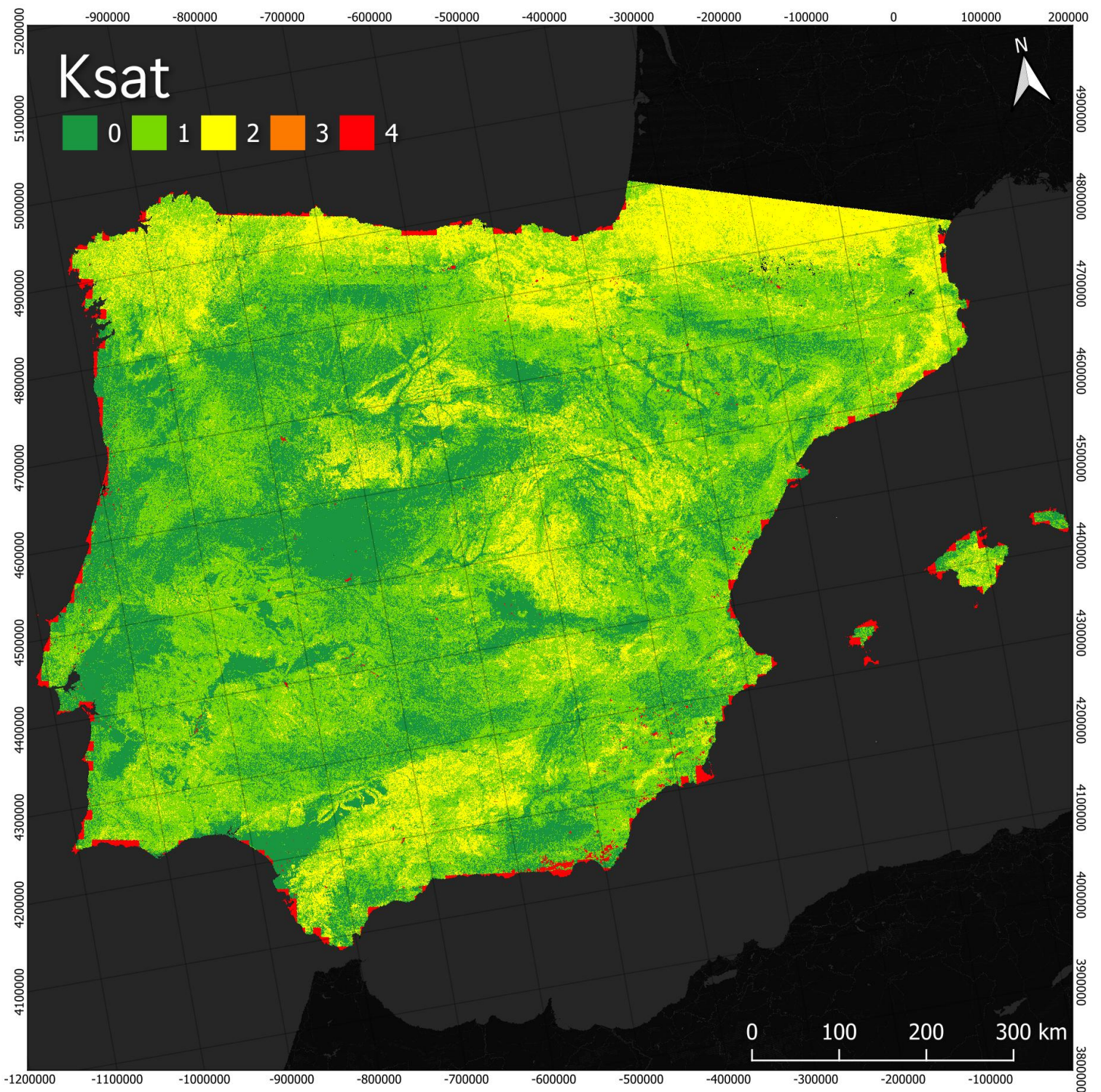
Figur 8.16. Klassificering af Dexter Ratio rangeringerne i caseområdet på baggrund af kravene for Dexter Ratioen som defineret i tabel 8.4. En lavere score indikerer en god sundhedstilstand, mens en højere score indikerer en dårlig sundhedstilstand.



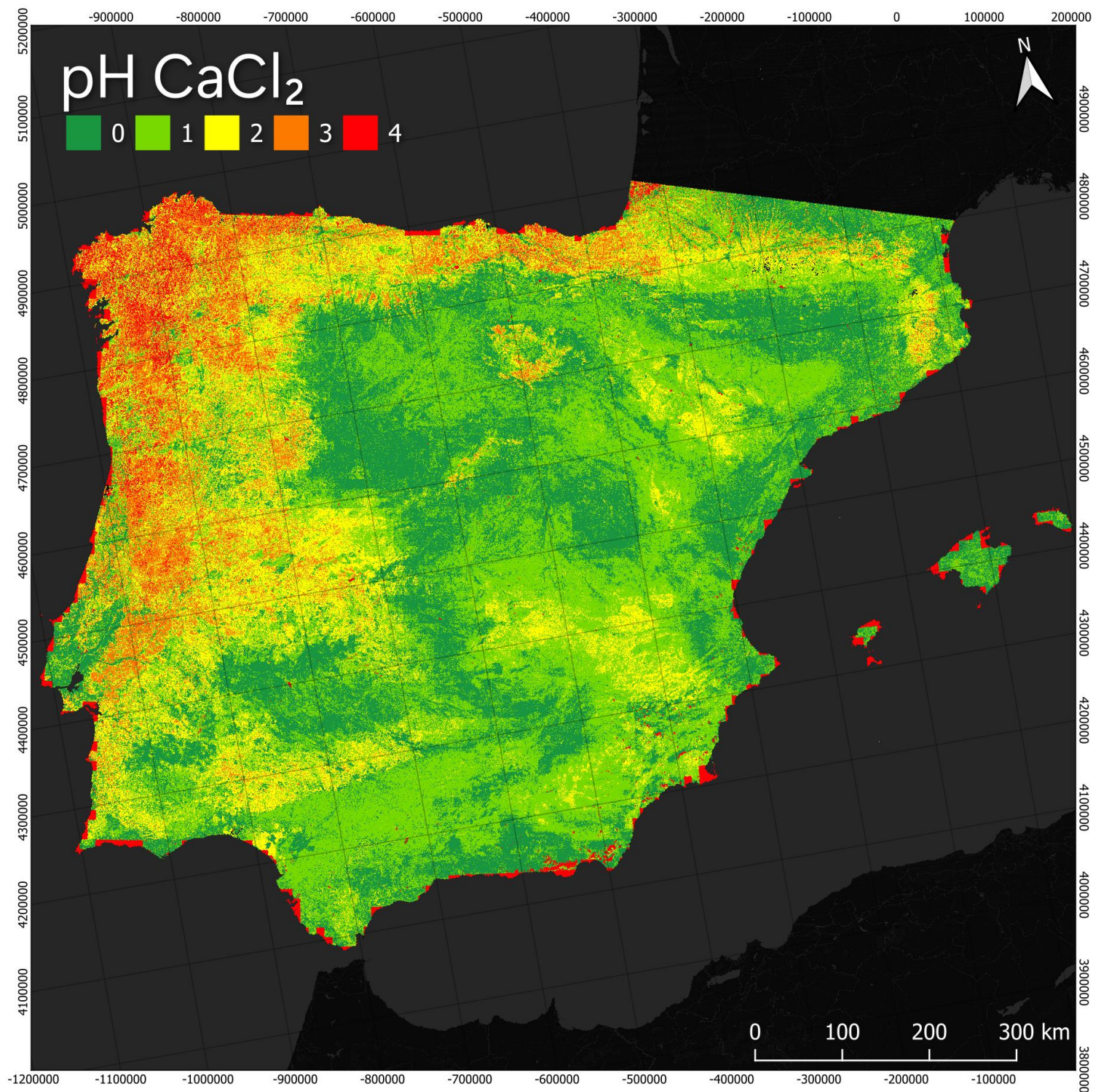
Figur 8.17. Klassificering af PAW-rangeringerne i caseområdet baseret på kriterierne for plantetilgængeligt vand, som defineret i tabel 8.4. En lavere score indikerer en god sundhedstilstand, mens en højere score indikerer en dårlig sundhedstilstand.



Figur 8.18. Klassificering af gasdiffusivitet rangeringerne i caseområdet på baggrund af kravene for gasdiffusivitet som defineret i tabel 8.4. En lavere score indikerer en god sundhedstilstand, mens en højere score indikerer en dårlig sundhedstilstand.



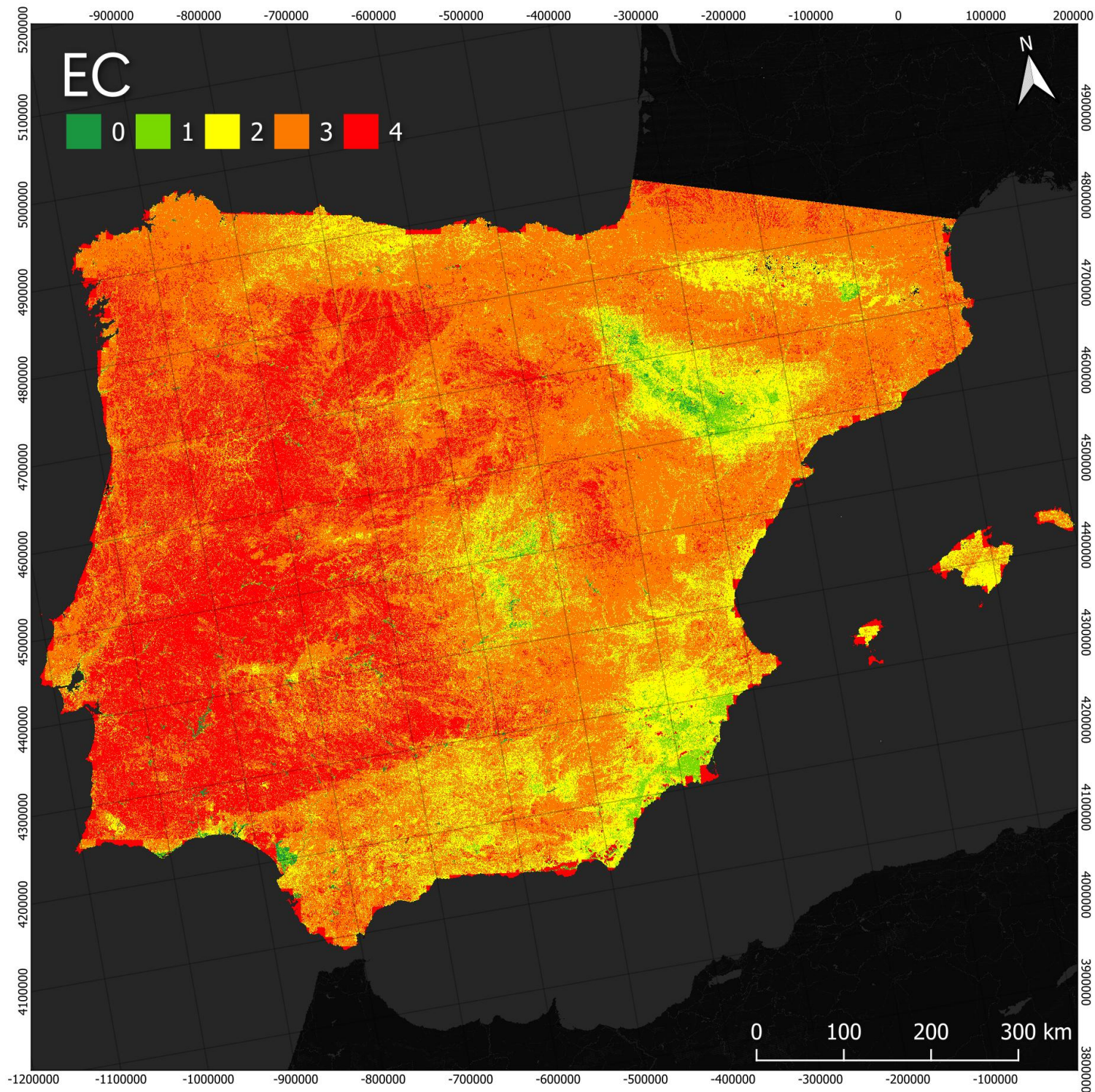
Figur 8.19. Klassificering af K_{sat} rangeringerne i caseområdet på baggrund af kravene for K_{sat} som defineret i tabel 8.4. En lavere score indikerer en god sundhedstilstand, mens en højere score indikerer en dårlig sundhedstilstand.



Figur 8.20. Klassificering af $pH\ CaCl_2$ rangeringerne i caseområdet på baggrund af kravene for pH som defineret i tabel 8.4. En lavere score indikerer en god sundhedstilstand, mens en højere score indikerer en dårlig sundhedstilstand.

Rangeringen af pH i caseområdet beretter om, at med udgangspunkt i de definerede rangeringsintervaller, at det hovedsageligt er vestlige regioner og områder som blandt andet Portugal og Galicien, Asturien, Cantabrien, Baskerlandet samt dele af Extremadura og Castilla y León i Spanien, hvor der er dårlige niveauer af pH . Kigges der på figur 8.5 a

for pH målt via $CaCl_2$, som netop er brugt hertil, kan det ses, at disse nævnte områder estimeres til at have særligt syrlige jorde, med pH alt fra lige under $pH \leq 6$, og i flere andre områder i især Galicien på under $pH \leq 4,5$. Ved at sammenligne med blandt andet indholdet af calciumkarbonat i caseområdet, som kan ses i figur 8.6 b, er netop de syrlige områder kalkfattige, med flere steder mindre kalk end 40 g/kg , som hører til blandt de mest kalkfattige områder per prædiktionerne. Derudover er også dele af den nordvendte bjergside af Pyrenæerne vurderet til en score på 2 og derover. Dette kan måske skyldes udvaskning af basiske mineraler såsom blandt andet kalken, som hjælper med at holde pH i balance. I forlængelse af det, ses det også at både EC-niveauet og kalkindholdet er særligt højt i Ebro-dalen ved Aragon-området syd for Pyrenæerne. Her kan muligvis kalkindholdet påvirke det målte EC, men påvirker også samtidigt pH-balancen, eftersom det her ses, at pH er særligt højt, på $pH \geq 7$ og derover.



Figur 8.21. Klassificering af *EC* rangeringerne i caseområdet på baggrund af kravene for *EC* som defineret i tabel 8.4. *EC* er omregnet fra *mS/m* til *mS/cm* for at passe med teorien om optimale *EC*-niveauer og tilmed rangeringerne i multikriterieanalysen. En lavere score indikerer en god sundhedstilstand, mens en højere score indikerer en dårlig sundhedstilstand.

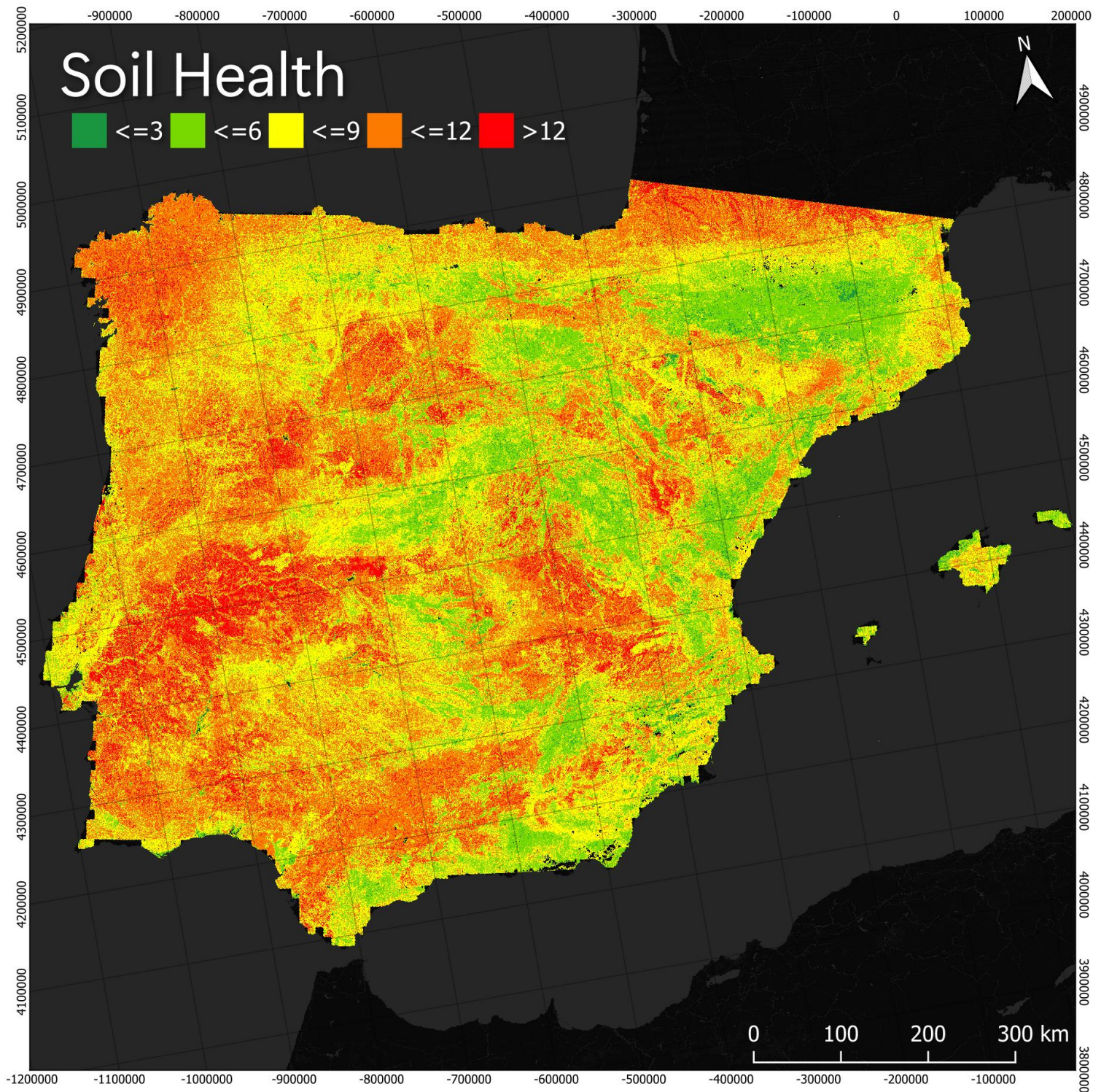
Fra rangeringerne lavet på prædikeret elektrisk konduktivitet indenfor caseområdet, kan det ses, at det forventes, at langt størstedelen af arealet opnår en score på 2 eller derover. Kun få områder, her især i og omkring Ebro floden ved Aragon region og øst-for, samt i

sydlige egne ved blandt andet Murcia og det sydlige område af Valencia-regionen, opnår scorer på 1 eller derunder. Ud fra det, kan det derfor vurderes, at størstedelen af caseområdet, må formodes enten næringsfattig eller potentielt i risiko for høj koncentration af næringsalte med videre, hvilket vil virke giftigt for planter. Netop EC er med til at give et billede på, hvor stor en andel der er af opløste salte i jorden, hvoraf mange er vigtige næringsstoffer, eller kan fungere som pH-regulerende mineraler. Samtidigt kan EC også fortælle om jordens fugtighed, i og med at dette er et essentielt bindeled i forhold til om jorden kan lede en elektrisk ladning. Da majoriteten af aseområdet scorer en EC på mindst 2, kan det derfor formodes, at der, på generel plan, er ugunstige mængder af enten tilgængelige, utilgængelige næringsstoffer eller vand for planter. Der kan dog ikke siges noget om det specifikke næringsstof eller volumetrisk vandvolumen ud fra EC alene. Netop her kan PAW give forklaring for det potentielle vandindhold planter kan tilgå.

For yderligere indsigt i de producerede kort for de anvendte jordparametre, og hvor store arealer de opnår de forskellige rangeringer, kan aflæses i følgende tabel 8.5. Heri kan arealerne for hvad hver rangering udgør læses. Rød farve indikerer høj værdi imens blå det modsatte. Ud fra det kan det aflæses, at for flere af parametrene, udgør rangeringen 0 store arealer af hele caseområdet relativt til rangeringer af 2 og derover. Rangering 1 udgør for de fleste parametre det største areal, hvilket også overføres til det summerede areal for hver rangering, hvoraf rangering 1 udgør det største og næstefter rangering 0.

Tabel 8.5. Arealerne for hver af de klassificerede jordparametre fra figurerne 8.16, 8.17, 8.18, 8.19, 8.20 og 8.21. Rød farve indikerer relativt høj værdi imens blå det modsatte.

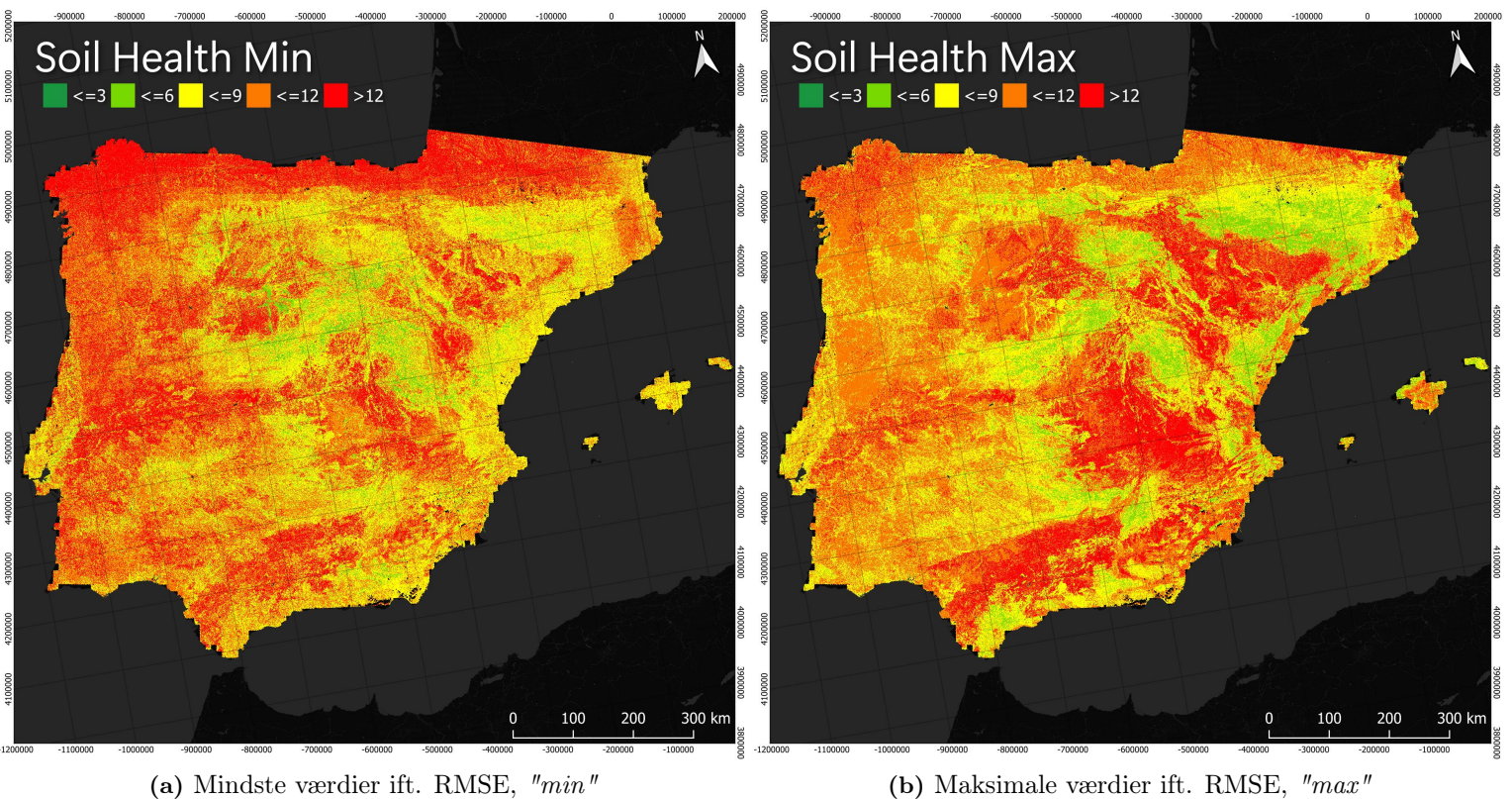
Jordparameter	Rangering (km^2)				
	0	1	2	3	4
Dexter Ratio $\frac{g\ Ler}{g\ OM}$	176359.44	152592.80	190583.58	89732.37	23595.95
PAW $\frac{cm^3\ vand}{cm^3\ jord}$	151289.16	157801.00	319088.67	3506.36	10219.20
Gas Diffusivity $\frac{D_{(p,100)}}{D_o}$	530861.17	86582.45	6370.09	4851.25	13239.43
$K_{sat}\ mm/h$	208266.38	308222.89	116246.42	106.75	9061.94
$pH\ CaCl_2$	199943.64	249954.27	114205.39	56638.41	21162.67
EC mS/cm	4198.75	15756.59	107367.80	353901.34	160679.91
Sum Areal km^2	1270918.54	970910.01	853861.95	508736.48	237959.11



Figur 8.22. Den samlede vurdering af jordens sundhed ud fra parametrene og deres specifikke rangeringer samt dertil bestemte krav, som er defineret i tabel 8.4. Heri indikerer en lavere score bedre sundhedsvurdering, og højere score en værre sundhedsvurdering. Den bedste score er 0, imens den værste beregnede score er 22.

Som udvidelse på MKA resultatet som vist i figur 8.22, er der i figur 8.23 beregnet for scenarierne, hvis alle parametrene kronisk har overvurderet og undervurderet i forhold til normalen. Det vil sige, at der for kortet i figur 8.23 (a) er fratrasket RMSE for

den enkelte parameter, og rangeret derpå, imens der for 8.23 (b) er lagt til, hvoraf de maksimale værdier opnås. Afhængigt af parameteren kan fratrækkelsen eller tillægelsen af den respektive RMSE-værdi medvirke til bedre rangeringer for den enkelte rangering, imens en eller flere af de andre parametre kan opnå værre rangeringer. Den normale værdi for de enkelte parametre som resulterer i kortet i figur 8.22 skal derfor ses som hvad der kan være mest sandsynligt, imens minimum og maksimum som vist i figur 8.23 kan forstås som bud på multikriterieanalyser af jordbundssundheden, ved enten lave eller høje værdier respektivt, og viser derfor ikke nødvendigvis hvordan den bedste eller værste kombination af parametre og respektive RMSE vil lyde.



Figur 8.23. Undersøgelse af hvad det kan betyde for MKA resultaterne, hvis den enkelte jordbundssundhedsparameter justeres for RMSE. Her er der kortlagt for hvis alle parametrene normalværdier er overvurderede ("min"), samt hvis undervurderede ("max").

Betragtes score-resultaterne af den samlede multikriterieanalyse som vist i figur 8.22 sammen med CORINE Land Use/Land Cover som præsenteret i figur 3.5, kan det ses, at mange af de steder, som opnår en score omkring 14 og derover, drives en type af landbrugsdrift på. Disse områder er typisk underlagt landbrug af eksempelvis marker til afgrøder med videre, eller plantager af frugttræer eller vinmarker. Ligeledes er flere af stederne hvor der drives agroskovbrug også vurderet til at være af ringe jordkvalitet. Den ringe-vurderede jordkvalitet kan skyldes potentielt intensiv landbrug der potentielt kan skade jordens balance.

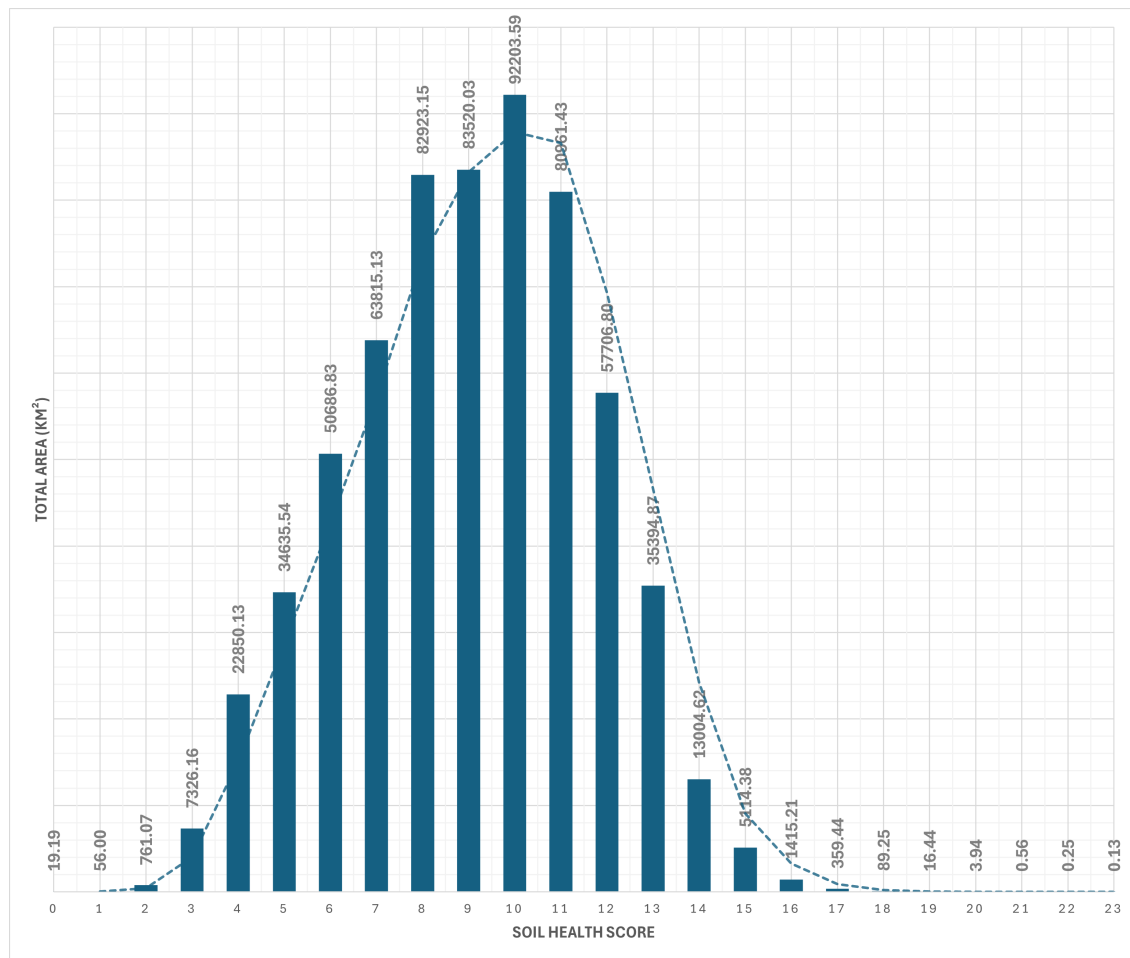
Det samlede caseområde udgør heri omtrent $641.585,23 \text{ km}^2$ jævnfør grænserne som defineret i forbindelse med kortet i figur 2.3. Da der herefter er arbejdet med rasters,

ækvivalent til hvad der er konstrueret af pixels, lyder arealet anderledes heri, eftersom opløsningen er blevet simplificeret. Derfor lyder arealet på $632.864,14 \text{ km}^2$, en forskel på omtrentlige $8.721,09 \text{ km}^2$. Dette er usikkerheder der altid kan opstå, når der arbejdes mellem rasters og vektorer i GIS-miljø. I tabel 8.6 står skrevet hvor store områder de viste rangeringer fra det endelige multikriterie kort, som vist i figur 8.22, udgør. Arealer med en rangering til og med 3 udgør $8.162,43 \text{ km}^2$, svarende til 1,29%. Fra og med scoren 4 til og med 6 udgør $108.172,50 \text{ km}^2$, svarende til 17,09%. 7 til og med 9 udgør et areal på $230.258,31 \text{ km}^2$, svarende til 36,38%. 10 til og med 12 udgør et areal på $230.871,82 \text{ km}^2$, svarende til 36,48%. Fra og med 13 og derover udgør et areal på $55.399,08 \text{ km}^2$, svarende til cirka 8,75% af området. Arealerne hver enkelte score udgør kan aflæses i histogrammet i figur 8.24.

Tabel 8.6. Arealerne for hver af de viste rangeringer på figur 8.22 udgør, samt deres procentmæssige størrelse i forhold til det totale areal.

Samlet Score						
	<=3	<=6	<=9	<=12	>12	Sum Areal
km ²	8162,43	108172,50	230258,31	230871,82	55399,08	632864,14
%	1,29%	17,09%	36,38%	36,48%	8,75%	100,00%

Resultaterne af den samlede multikriterieanalyse peger i retning ad, at flere og store dele af områderne i caseområdet, og deraf Spanien, Portugal og sydlige Frankrig, i større træk lider af dårlige jordforhold, imens Andorra stort set kun opnår gode resultater. De gode resultater for Andorra kan med al sandsynlighed skyldes manglen på data vedrørende CORINE Land Use/Land Cover, som eksempelvis kan ses i kortet for CORINE i figur 3.5. Kun et areal på i alt $108.172,50 \text{ km}^2$ opnår en score på 6 eller derunder. Dette svarer til cirka 17,09% ud af et samlet caseområde på i alt $632.863,82 \text{ km}^2$. Som det yderligere kan aflæses fra histogramanalysen over rangeringerne og deres arealer i figur 8.24, er det kun få fraktioner af caseområdet, som opnår særligt ringe score på eksempelvis 17 og derover. Majoriteten af caseområdet opnår en score mellem 4 og 14.



Figur 8.24. Histogram med arealet for hver score af multikriterieanalysen i caseområdet. Heri kan det eksempelvis aflæses, at omtrent $92.203,59 \text{ km}^2$ af caseområdet opnår en score på 10, og er derfor også den score der dækker største areal, og kun $19,19 \text{ km}^2$ af caseområdet opnår en score på 0.

Hvad resultaterne af både det formidlede kort i figur 8.22 samt histogrammet over arealfordelingen af hver score i figur 8.24, fortæller, er at der er flere og større steder og områder i caseområdet, hvori jordbunden i de øverste 20 cm, er ugunstige. De ugunstige forhold skelner ikke specifikt imellem om den ene eller anden parameter er "god", hvilket grundlæggende er sådan en multikriterieanalyse fungerer. Det vil derfor betyde, at selv hvis *pH* kan vurderes som *perfekt* for små områder eller i et andet scenarie hele det undersøgte område, overføres dette ikke nødvendigvis til den samlede multikriterieanalyse, da der heri kan være andre parametre, som peger på dårlige forhold. Multikriterieanalysen skal derfor anses som værende en holistisk analyse af de multiple inputs der indgår i den, hvori, for eksempelvis samme sted, flere parametre opnår god score, imens andre den værste score.

Diskussion 9

I dette studie blev det undersøgt, om en konceptuel model baseret på forskellige inputparametre i en regressionsmodel kan give prædiktioner på relevante jordparametre, som kan anvendes til at estimere jordens sundhed. Denne metode inkluderede både in-situ jordprøver fra LUCAS-datasættet og forskellige inputparametre i regressionsmodellen, for at lave en kortlægning af jordparametrene på stor skala, og for at anvende disse i en vurdering af jordens sundhedstilstand over større geografiske områder.

Et af de største problemer med modellen er, at distributionen af LUCAS-jordprøver er ujævnt fordelt på tværs af arealanvendelser, klimaklassifikationer og terrænhøjder. Som det blev illustreret i figurerne 3.4, 3.6 og 3.8, er nogle områder underrepræsenteret i forhold til andre i distributionen af LUCAS-jordprøver. Dette vil sige, at den datamæssige opløsning på baggrund af LUCAS-data ved nogle arealer, er ringere, hvorved der derfor er mindre data at træne en maskinlæringsmodel på, versus areal typer hvor der er høj mængde af data. Dette kan potentielt have implikationer for prædiktionspræcisionen, i og med at modellen har færre kombinationer af data at arbejde med, hvorfor den kan overse eventuelle scenarier og derved prædiktere forkerte eller forringede resultater.

De steder hvor datatilgængeligheden er meget høj, relativt til arealstørrelsen af den enkelte eksempelvis areal anvendelsestype eller terrænhøjde, kan måske halvdelen eller mindre endnu være nok til at give alle eller en repræsentativ mængde af forskellige kombinationer, for modellen til at give gode prædiktionsresultater. Med andre ord kan det derfor undres hvad den relevante mængde af data er for hver arealtype, dvs. såsom arealanvendelse, klima, terræn eller andet, før at maskinlæringsmodellen kan give fornuftige og acceptable prædiktioner. De i LUCAS-datasættet observerede jordparametre udgør ikke nødvendigvis alle relevante jordparametre til at give en fuldkommen vurdering af jordbundssundheden, men udgør en stikprøve af nogle af dem, der anses som mest relevant for jorde og deres funktioner, hvorfor det netop er disse, der undersøges i forbindelse med LUCAS-undersøgelsen.

Grundet i at LUCAS ikke har taget direkte målinger eller oplyst direkte værdier for parametre som eksempelvis plantetilgængeligt vand (PAW), kan der desuden potentielt udvikle sig større usikkerheder i resultaterne, idet der bliver nødsaget til at anvende pedotransfer-funktioner til estimering af disse. Dette skaber en vis usikkerhed i resultaterne, da disse estimeringer muligvis ikke nøjagtigt afspejler virkeligheden eller er gældende for hele det kortlagte område, dette projekt beskæftiger sig med. Usikkerheder ved disse parametre kan have større betydning for anvendelsen af det endelige resultat, det vil sige multikriterieanalysen, i og med at disse parametre udgør store dele heri.

Den statistiske analyse af inputparametrene viste, at visse parametre var stærkt

korrelerede, såsom flere af de Sentinel-2 spektrale bånd, indekserne NDVI og SAVI, samt NDMI og Gypsum Index med flere, hvilke kan aflæses ud fra figur 7.4. Korrelerede inputparametre eller multikollinearitet kan skabe udfordringer i modellen, da det kan føre til overfitting, hvilket reducerer modellens evne til at generalisere nye data og kan påvirke modellens prædiktionssevne med andre datasæt eller områder. Ved at inkludere flere parametre, som ikke er stærkt korrelerede, kan det derfor være med til at forbedre modellens robusthed og en højere præcision generelt [Frost, 2023].

En af de største begrænsninger i forhold til metodologien var satellitdataens opløsning og dække af jordprøver fra LUCAS-datasættet. Selvom Sentinel-2 leverer data i høj opløsning (generelt 10-20 meter), kan det stadig være for groft til at fange små, lokale variationer i jorde, især i heterogene landskaber. Desuden var jordprøverne i LUCAS-datasættet ikke jævnt fordelt i hele case-området, hvilket kan have påvirket modellens præcision i områder med færre jordprøver. Det undersøgte caseområde var meget stort, da det dækkede hele den Iberiske Halvø, men der var alligevel en utilstrækkelig prøvedækning i visse områder til at sikre en præcis kalibrering af modellen, hvilket især kan skabe problemer for modellen i områder med høj grad af variation i jordtyperne. Den ujævne fordeling af jordprøver betyder, at modellen forventes at kunne blive forbedret med flere prøver fra disse underrepræsenterede områder.

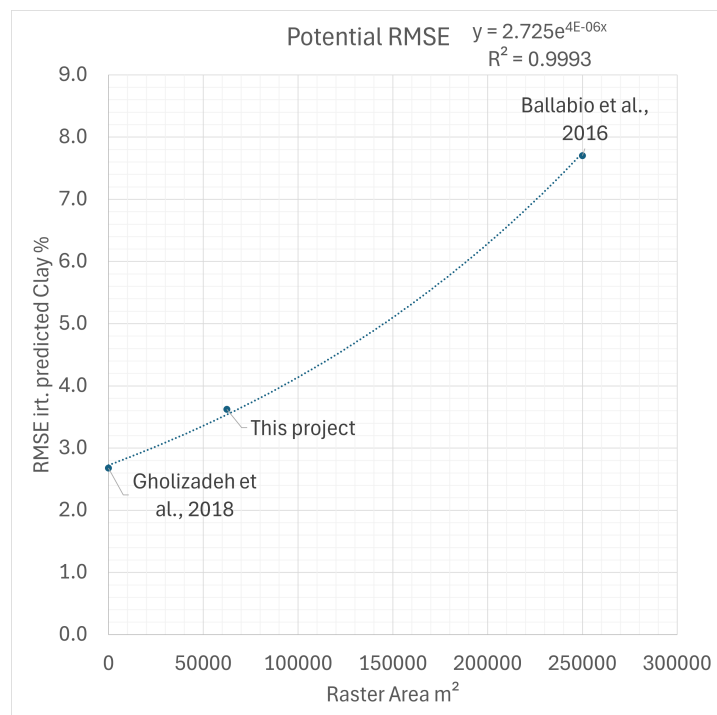
Derudover må det også formodes, at forholdet mellem arealstørrelse per rasterpixels i anvendte data per LUCAS jordprøve, spiller en rolle i forhold til en models evne til at prædiktere nøjagtigt for det punkt der svare til det sted der blev indsamlet jordprøven. Eksempelvis må der under al overbevisning være stor sandsynlighed for, at et område svarende til $250 \times 250 \text{ m}^2$, som brugt her i projektet, dække mere en én komposition af ler-, silt- og sandforhold. Altså at det ikke kan forventes, at en jord på eksempelvis 20% ler og silt og 60% sand ikke er realistisk for et område på $250 \times 250 \text{ m}^2$. Det må derfor forventes at ratioen mellem rasterstørrelse og LUCAS punkter har implikationer for modellens prædiktionssevne, eftersom der er en iboende "støj" i den inputdata den fodres med, eftersom den data, der dækker det store område, bliver fortalt kun repræsenterer én jordkomposition, eller flere, hvis der er flere LUCAS jordprøver indenfor det givne areal. Dette kan hypoteseres at have en vildledende effekt på modellen, hvorfor det kan blive nødvendigt med yderligere mængder af forskelligartet data, der kan bidrage med unikke karakteristika for den enkelte $250 \times 250 \text{ m}^2$ celle i GIS miljøet. Optimalt vil derfor være at den arealmæssige ratio mellem inputdata og LUCAS-data, eller anden data der ønskes trænet på, nærmer sig 1, så de bedst muligt kun repræsenterer den enkelte jordprøve, og ikke andre eller andet.

En hypotese i samspillet mellem den arealdække-mæssige størrelse på anvendte data og præcisionen af data kan lyde på, at der er en sammenhæng. Betragtes eksempelvis $RMSE_{cv}$ for ler som udregnet for jordmodellerne som lavet i forbindelse med studiet af Gholizadeh et al. [2018], som set i tabel 1.1, scorer de alle under 3,5 på gennemsnitligt 2,678. Heri brugte de resamplede satellit rasters til $10 \times 10 \text{ m}^2$ for undersøgelse via satellit [Gholizadeh et al., 2018, s. 92]. I kontrast hertil opnår Ballabio et al. [2016] en $RMSE_{cv}$ på 7,70 for ler, ved brug af en arealopløsning på $500 \times 500 \text{ m}^2$ [Ballabio et al., 2016, s. 112]. Heri projektet blev der anvendt en opløsning på $250 \times 250 \text{ m}^2$, hvor der blev estimeret en $RMSE$ for ler til 3,623. Alle resultater taget i betragtning lyder en eksponentiel trend på

$y = 2,725 \times e^{4 \times 10^{-6} \times x}$ ($R^2 = 0,9993$), se figur 9.1, hvor mindre areal umiddelbart bidrager til potentielt bedre *RMSE*. Disse forskellige resultater *RMSE* sammenlignet med den anvendte opløsning, kan tyde på, at netop opløsning anvendt spiller en rolle i forhold til præcisionen af blandt andet modelleringen og deraf prædiktionssevnen. Der er naturligvis mere end blot størrelse på den enkelte celle der påvirker prædiktionspræcisionen. Ved en eventuel fremtidig undersøgelse af prædiktionsmuligheder på jord ved brug af Remote Sensing, kan det derfor være interessant også at betragte hvordan opløsningen af dataene og modellen kan ændre på resultaterne.

Resultaterne fra analysen kan være et brugbart værktøj i forbindelse med jordforvaltning og for landmænd, især i områder med dårlig jordbundssundhed, eller for at overvåge udviklingen, så der kan igangsættes tiltag og gribes ind i tide, i områder med fare for nedbrydning, inden jorden forværres i en sådan grad, at den bliver usund. Eksempelvis kan det være nødvendigt at optimere irrigationen i områder med lav vandtilgængelighed eller reducere kompakteringsgraden i områder med lav gasdiffusivitet, eller forbedre jordens lagring af kulstof og næringsstoffer, samt pH. Dermed kan modellen være med til at understøtte jordforvaltningen, ved at overvåge jordens sundhed og laves initiativer med en dynamisk og adaptiv forvaltning, hvor man hurtigt kan reagere på disse ændringer og dermed vedligeholde jordens sundhed, eksempelvis ved at forhindre erosion, tab af organiske materiale eller andre jordnedbrydende processer.

I forbindelse med fremtidig forskning kan det være interessant at inddrage flere jordparametre eller finjustere de eksisterende modeller for at øge præcisionen. For eksempel kunne parametre som jordens mikrobiologiske aktivitet, kationbytningssevnen (Cation Exchange Capacity) eller kulstoflagringssevne også inddrages, da disse også er vigtige indikatorer for jordens langsigtede sundhed. Derudover kunne fremtidige studier også inkludere et mere historisk perspektiv med prognoser eller klimamodeller for at undersøge, hvordan ændringer i klimaet påvirker jordens sundhed over tid, hvilket kan være særligt vigtigt i områder, der er udsat for klimaændringer. Muligheden for at kunne prædiktere jordsundhed for fremtiden afhænger endvidere af det tilgængelige data og deres eventuelle



Figur 9.1. Regression på forholdet mellem udregnet *RMSE* på modelleret lerindhold i jorde i forhold til arealmæssige størrelse på de enkelte pixels i det anvendte data.

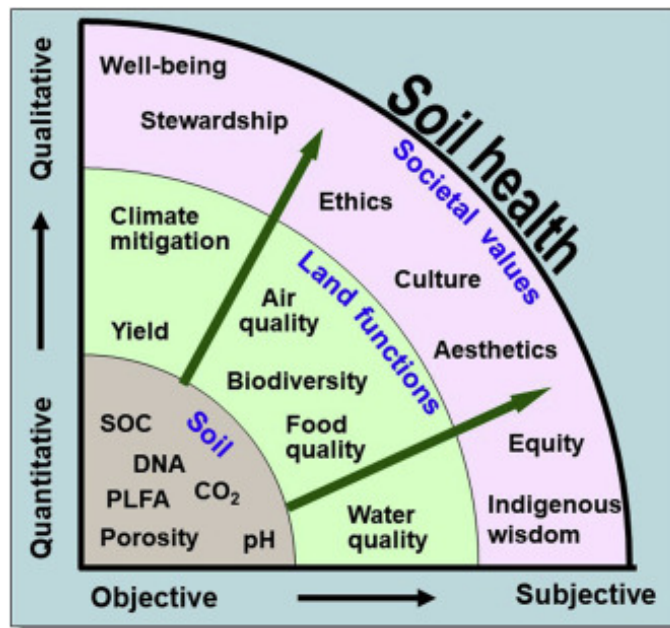
temporale opløsning. Datakilder såsom Sentinel-2 opdateres jævnligt i forbindelse med at de flyver over områder, og andre data såsom klima kan dække over længere såvel som kortere perioder og scenarier. Direkte satellit-foto fortæller dog kun om her og nu, hvorfor der hertil bliver nødvendigt at lave fremtidsprognoser eksempelvis baseret på forrige års data for en given dag, måned, årstid eller sæson. I og med at klimaet, som evident i forbindelse med analysen, spiller en rolle for flere af jordparametrene, kan det forventes, at anvendelse af forskellige fremtidige klimascenarier kan have indflydelse på hvad der prædikteres for fremtiden. Graden terrænhøjden ændrer sig i løbet af fremtiden, eksempelvis igennem erosion og lignende, kan potentielt også ignoreres, hvis den er tilstrækkelig lav. Da alle de andre geomorfologiske data i forbindelse med modelleringen heri baserer sig på en sådan terrænmodel (DEM), er kompleksiteten af en eventuel fremtidsprognostisering af en højdemodel relativt simpel.

9.1 Kritik over projektets tilgang til vurdering af jordens sundhed

Det bør bemærkes, at dette projekt er et koncept, og at der er flere begrænsninger ved brugen af numeriske modeller til at måle jordens sundhed. Selvom numeriske modeller kan give et objektivt grundlag for vurdering af jordens tilstand, er der flere komplekse aspekter af jordens sundhed, der ikke kan indkapsles udelukkende gennem tal. Ifølge studiet af Hughes et al. [2023] varierer sundhedsindikatorer betydeligt afhængigt af den specifikke kontekst. Økosystemer og jordtyper varierer, hvilket gør det problematisk at anvende standardiserede målinger uden at tage højde for de forskellige økosystemer, som jorden understøtter.

I en artikel af Janzen et al. [2021] understreges det, at på trods af det er muligt at måle bestemte jordparametre og bruge disse til at estimere jordens evne til at understøtte plantevækst og jordfunktioner, er jordens sundhed fortsat et konceptuelt begreb. De beskriver jordens sundhed som en metafor, og at den derfor skal betragtes som et udefinerbart begreb. Forskellige planter kræver forskellige jordforhold; hvad der kan betragtes som dårlig jord for én planteart, kan være ideel for en anden. For eksempel trives fyrretræer i sandet jord, mens frugttræer her vil fejle, idet de kræver en anden jordtekstur. Dette understreger begrænsningen ved at vurdere jordens sundhed alene ud fra få parametre og generelle tærskelværdier.

Hertil kritiserer Janzen et al. [2021] tilmed metaforen jordens sundhed for, at en tildelt værdi ikke er nok, hvis man vil vurdere hvor 'god' eller hvor 'egnet' en jord er til bestemte funktioner. Men at man er nødt til at tage i betragtning, hvad der gør en jord 'god' et bestemt sted, og hvorfor, for at få et større indblik i hvordan jorden bedst vedligeholder sine funktioner og understøtter det omkringliggende økosystem.



Figur 9.2. Et diagram der illustrerer hvordan jordmålinger (sundhedsindikatorer) er nødt til at ses i perspektivet af jordens funktioner og samfundsværdi, før man kan sige noget konkluderende om jordens sundhed. Bemærk at diagrammet ikke er en fuldendt liste, men blot med til at vise diversiteten af forskellige faktorer, der bør tages hensyn til, i en vurdering af jordens sundhed [Janzen et al., 2021].

Janzen et al. [2021] mener dermed, at jordens sundhed skal betragtes i dets kontekst, og danne et helhedsbillede (se figur 9.2), der inddrager kvantitative og kvalitative egenskaber, såvel som de objektive og subjektive. At en objektiv, numerisk, kvantitativ model, kan danne et grundlag eller fundament til et koncept, der er med til at forbedre forståelsen for jordens sundhed, men at man også skal medregne landets funktioner og de værdier, det omkringliggende samfund pålægger jorden, for at danne et meningsfuldt og fuldendt billede af jordsundheden. Og dermed, at jordens sundhed findes et sted i de relationelle mekanismer mellem jorden, økosystemet, økosystemtjenesterne og de værdier, de tilsammen skaber [Janzen et al., 2021].

Af disse grunde kan nærværende projekts tilgang kritiseres for ikke at inkludere jordens sundhed ud fra dens kontekst i en mere kvalitativ betragtning. Økosystemers og planters forskellige krav bliver ikke taget i betragtning, hvilket gør, at projektet risikerer at generalisere jordens sundhed baseret på universelle parametre uden at anerkende den nødvendige variation i jordforholdene, og ligeledes i jordens sundhed.

På den anden side tilbyder numeriske modeller en vigtig fordel, idet de gør det muligt at vurdere jordsundhed objektivt og systematisk over store geografiske områder. Ved at anvende målbare parametre som gasdiffusivitet, pH og elektrisk ledningsevne kan modellen skabe et sammenligneligt og kvantificerbart grundlag for vurdering af jorden. Objektiviteten kan være særligt nyttig i praktiske anvendelser, som for eksempel landbrugsplanlægning, hvor der er brug for hurtige og præcise vurderinger. Projektets koncept om anvendelse af maskinlæring og fjernmålingsdata muliggør altså kontinuerlig overvågning på stor skala, hvilket er værdifuldt for landbrug og naturforvaltning. Løbende overvågning kan samtidig hjælpe med at sikre bæredygtig udnyttelse af ressourcerne og

tidlig identifikation af problemer i jordens kvalitet.

Til trods for fordelene bør projektet erkende, at numeriske modeller kan overse vigtige kvalitative faktorer, som ikke kan måles, men som er essentielle for en helhedsforståelse af jordens sundhed. Som Janzen et al. [2021] bemærker, skal en vurdering af jordsundheden også tage højde for kvalitative og kontekstspecifikke faktorer. Derfor bør modellen suppleres med mere nuancerede og kvalitative vurderinger, så den kvantitative tilgang sættes i perspektiv.

Konklusion 10

Dette projekt har vist, hvordan en model, der anvender remote sensing, klimadata og terændata kombineret med en hybrid model, baseret på in-situ data og pedotransfer funktioner, for at kortlægge relevante sundhedsindikatorer. Gennem prædiktions af væsentlige jordparametre såsom plantetilgængeligt vand, jordtekstur, jordens gasdiffusivitet, pH osv., er det demonstreret, at metoden danner et grundlag for at evaluere jordens tilstand på tværs af mange parametre og over et område med mange forskellige regioner og forskelligheder. Dette koncept har vist sig at være en effektiv måde at modellere jordens sundhed, og kan danne et grundlag, for at udvikle et effektivt værktøj til kortlægning af jordbundens sundhed.

Den benyttede fremgangsmetode er i form af en fleksibel model, som afhængigt af tilgængelige data, kan tilpasses mange forskellige områder og skala afhængigt af behov. I dette projekt blev der anvendt et caseområde i form af den Iberiske Halvø der også dækker dele af sydfrankrig, og her vises det hvordan modellen nemt kan skaleres op til større geografiske områder for at monitorere jordbundssundheden på nationalt eller regionalt niveau. Alternativt kan der også nedskaleres til specifikke landbrugsområder som i artiklen af Gholizadeh et al. [2018], hvis højere opløsning og detaljeret overvågning af et mindre område er påkrævet. Derfor er anvendelsen af denne metode anvendelig i mange forskellige kontekster, både på makro- og mikroskala.

Særligt remote sensing har klare fordele, især på grund af metodens lave omkostninger sammenlignet med feltarbejde og laboratorieanalyser. Metoden gør det muligt at overvåge store geografiske områder regelmæssigt via satellitter som Sentinel-2, med løbende overvågning, der kan hjælpe med at identificere tidlige tegn på jordnedbrydning og dermed muliggøre hurtige indgreb for at forbedre jordens sundhed. Desuden giver de satellitbaserede observationer en kontinuerlig mulighed for at følge ændringer over tid og sikre, at der kan reageres hurtigt på potentielle problemer, før de udvikler sig og gør yderligere skade på jordens tilstand.

Derfor kan det være nødvendigt og essentielt for modellen, at anskaffe mere in-situ data på tværs af mange forskellige kombinationer af jorde, for at kalibrere og validere modellen, så den er så præcis som mulig.

Derudover bør det erkendes i forhold til projektets koncept, at numeriske modeller kan overse vigtige kvalitative faktorer, som ikke kan måles direkte, men som stadig er afgørende for at forstå jordens sundhed. Særligt ved at det er nødvendigt at inddrage både kvalitative og kontekstspecifikke faktorer for at danne et mere nuanceret billede af jordsundheden, hvorfor modellen forventes at kunne gavne fra at blive suppleret med mere kvalitative vurderinger, for at sikre en mere helhedsorienteret forståelse af jordens tilstand [Janzen

et al., 2021].

For alle de prædikterede jordparametre samt afledte sundhedsmæssige parametre estimeret ud fra pedotransferfunktioner, ses en kronisk over og undervurdering ved ekstreme værdier (outliers). Der er derfor elementer at optimere og forbedre i forbindelse med maskinlæringsmodellen, således disse afvigelser kan mindskes eller helt fjernes. Det vides ikke om afvigelserne kommer af mangel på inputdata til træning af modellerne eller modsat for meget data, eller om en anden regressionsmodel kan mindske eller fjerne afvigelserne på bekostning af blandt andet præcision.

Resultaterne af sammenlægningen og vurderingen af de seks sundhedsmæssige jordparametre med de udvalgte vægtninger, tærskler og beregnede rangeringer, peger i retning af, at majoriteten af den Iberiske Halvø har forringet jordbundssundhed. Områderne, for både sunde og ugunstige jorde, er ikke nødvendigvis sammenhængene, og der er derfor at finde flere enklaver af områder med eksempelvis god jordsundhed, imens det samme er gældende for ugunstige jorde. Generelt ses jorde være sundere langs bjergskrånninger og lignende, hvor der eksempelvis ikke drives samme grad af landbrug, eller fordi der falder mere regn, med videre. Det organiske indhold i jorden er ikke det eneste element som løfter jordens sundhed. Eksempelvis spiller pH en rolle herved taget planter i betragtning, hvorfor også jordens sundhed bliver sunder imod den østlige del af den Iberiske Halvø, i og med der blandt andet er højere mængde af $CaCO_3$ til at balancere jordens pH.

Kun 18,38% af det undersøgte caseområdet opnår acceptable forhold på en score op til og med 6, imens kun 1,29% estimeres til at have optimale forhold for jordbundssundheden med en score på ≤ 3 . Kun et lille område på i alt $19,19 \text{ km}^2$, svarende til 0,003%, opnår hvad der ud fra multikriterieanalysen kan vurderes som perfekt score, hvori de bedste forhold, på baggrund af anvendte og undersøgte parametre, eksisterer.

Konceptet, som er prædiktion og estimering af jordparametre ud fra remote sensing data, er en potentiel tilgang til dette formål. Anvendelse af maskinlæring til prædiktion af jordparametre og afledt deraf, som arbejdet med heri forbindelse med projektet, sundhedsvurdering af jorden, vil kræve mere en finpudset og optimeret tilgang af både det data der trænes på, samt de inputdata der anvendes, hvis det skal anvendes i en mere professionel sammenhæng, hvor præcision kan være et højt krav. Indtil den tid, hvor remote sensing kan tilbyde data af særligt høj kvalitet og opløsning, kan maskinlæring derfor fortsat anvendes som værktøj til at give bud på hvad der kan forventes af jorde og dets kompositioner indenfor større arealer.

Litteraturliste

- Aciego Pietri og Brookes, 2008.** J.C. Aciego Pietri og P.C. Brookes. *Nitrogen mineralisation along a pH gradient of a silty loam UK soil*. Soil Biology and Biochemistry, 40(3), 797–802, 2008. ISSN 0038-0717. doi: <https://doi.org/10.1016/j.soilbio.2007.10.014>. URL <https://www.sciencedirect.com/science/article/pii/S0038071707004245>.
- ArcGIS, u.d.** ArcGIS. *Weighted Sum (Spatial Analyst)*. ESRI, u.d.
- Ashman og Puri, 2002.** Mark R. Ashman og Geeta Puri. *Essential soil science: a clear and concise introduction to soil science*. Wiley-Blackwell Publishing, 2002. ISBN 978-0-632-04885-4.
- Atlas Scientific, 2023.** Atlas Scientific. *A Guide To Soil Electrical Conductivity (EC): Description, Importance & Soil EC Maps*, 2023. URL <https://atlas-scientific.com/blog/soil-electrical-conductivity/>.
- Ballabio et al., 2016.** Cristiano Ballabio, Panos Panagos og Luca Monatanarella. *Mapping topsoil physical properties at European scale using the LUCAS database*. Geoderma, 261, 110–123, 2016. ISSN 0016-7061. doi: <https://doi.org/10.1016/j.geoderma.2015.07.006>. URL <https://www.sciencedirect.com/science/article/pii/S0016706115300173>.
- Beck et al., 7 2018.** Hylke E. Beck, Niklaus E. Zimmermann, Tim R. McVicar, Noemi Vergopolan, Alexis Berg og Eric F. Wood. *Present and future Köppen-Geiger climate classification maps at 1-km resolution*. Scientific Data 5:180214, 2018. doi: [10.6084/m9.figshare.6396959](https://doi.org/10.6084/m9.figshare.6396959). URL https://figshare.com/articles/dataset/Present_and_future_K_ppen-Geiger_climate_classification_maps_at_1-km_resolution/6396959.
- Ben-Dor et al., 1997.** E. Ben-Dor, Y. Inbar og Y. Chen. *The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process*. Remote Sensing of Environment, 61(1), 1–15, 1997. ISSN 0034-4257. doi: [https://doi.org/10.1016/S0034-4257\(96\)00120-4](https://doi.org/10.1016/S0034-4257(96)00120-4). URL <https://www.sciencedirect.com/science/article/pii/S0034425796001204>.
- Ben-Dor, 2002.** Eyal Ben-Dor. *Quantitative remote sensing of soil properties*. Advances in Agronomy, 75, 173–243, 2002. URL [http://dx.doi.org/10.1016/S0065-2113\(02\)75005-0](http://dx.doi.org/10.1016/S0065-2113(02)75005-0).
- Berlin, u.d.** Freie Universität Berlin. *Sentinel 2*, u.d. URL <https://blogs.fu-berlin.de/reseda/sentinel-2/>. Tilgået: 07-08-2024.

- Bradford et al., 09 2008.** Scott Bradford, Eran Segal, Wei Zheng, Qiquan Wang og Stephen Hutchins. *Reuse of Concentrated Animal Feeding Operation Wastewater on Agricultural Lands*. Journal of environmental quality, 37, S97–S115, 2008. doi: 10.2134/jeq2007.0393.
- Buckingham, 1904.** Edgar Buckingham. *Contributions to our knowledge of the aeration of soils*. U.S. Dept. of Agriculture, Bureau of Soils, Washington, D.C., 25, 1904. URL <https://archive.org/details/contributionstoo25buck/>.
- Buckingham, 1907.** Edgar Buckingham. *Studies on the movement of soil moisture*. U.S. Dept. of Agriculture, Bureau of Soils, Washington, D.C., 38, 1907. URL <https://archive.org/details/studiesonmovemen38buck>.
- Castaldi et al., 2023.** Fabio Castaldi, Muhammed Halil Koparan, Johanna Wetterlind, Renaldas Žydelis, Ialina Vinci, Ayşe Özge Savaş, Cantekin Kıvrak, Tülay Tunçay, Jonas Volungevičius, Silvia Obber, Francesca Ragazzi, Douglas Malo og Emmanuelle Vaudour. *Assessing the capability of Sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands*. ISPRS Journal of Photogrammetry and Remote Sensing, 199, 40–60, 2023. ISSN 0924-2716. doi: <https://doi.org/10.1016/j.isprsjprs.2023.03.016>. URL <https://www.sciencedirect.com/science/article/pii/S0924271623000771>.
- Conrad et al., 2015.** O. Conrad, B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann og J. Boehner. *System for Automated Geoscientific Analyses (SAGA) v. 2.1.4*. Geoscientific Model Development, 8(7), 1991–2007, 2015. doi: 10.5194/gmd-8-1991-2015. URL <https://gmd.copernicus.org/articles/8/1991/2015/>.
- Conrad et al., 2024.** Olaf Conrad, Benjamin Bechtel, Juergen Boehner, Helmut Dietrich, Erik Fischer, Lars Gerlitz, Julian Wehberg, Volker Wichmann og Maik Buerker. *System for Automated Geoscientific Analyses (SAGA) - Version 9.5.1*, 2024. URL <http://www.saga-gis.org/>. Tilgæt: 12-08-2024.
- Copernicus, u.d.** Copernicus. *Copernicus DEM - Global and European Digital Elevation Model (COP-DEM)*, u.d. URL <https://spacedata.copernicus.eu/en/web/guest/collections/copernicus-digital-elevation-model>. Tilgæt: 19-08-2024.
- Copernicus, 2020.** Copernicus. *CORINE Land Cover 2018 (vector), Europe, 6-yearly - version 2020_20u1, May 2020*. Corine Land Cover (CLC), 2020. doi: 10.2909/71c95a07-e296-44fc-b22b-415f42acdfd0. URL <https://doi.org/10.2909/71c95a07-e296-44fc-b22b-415f42acdfd0>.
- Copernicus og ESA, u.d.land12010032.** Copernicus og ESA. *Overview of Sentinel-2 Mission*, u.d.land12010032. URL <https://sentiwiki.copernicus.eu/web/s2-mission>. Tilgæt: 29-08-2024.
- Copernicus Climate Change Service (C3S), u.d.** Copernicus Climate Change Service (C3S). *Copernicus Interactive Climate Atlas (C3S Atlas)*, u.d. URL <https://atlas.climate.copernicus.eu/atlas>. Tilgæt: 20-08-2024.

- Copernicus Land Monitoring Service, u.d.** Copernicus Land Monitoring Service. *Sentinel-2 Mosaic Hub*, u.d. URL <https://s2gm.land.copernicus.eu/mosaic-hub>. Tilgået: 18-06-2024.
- Das et al., 2024.** Soil health assessment and spatial characterization using remote sensing. In *Remote Sensing of Soils*, pages 455–467. Elsevier, 2024. ISBN 978-0-443-18773-5. doi: <https://doi.org/10.1016/B978-0-443-18773-5.00034-X>. URL <https://www.sciencedirect.com/science/article/pii/B978044318773500034X>.
- Datta et al., 2018.** Sumon Datta, Saleh Taghvaeian og Jacob Stivers. *Understanding Soil Water Content and Thresholds for Irrigation Management*, 2018. URL <https://extension.okstate.edu/fact-sheets/understanding-soil-water-content-and-thresholds-for-irrigation-management.html>.
- Dexter et al., 2008.** A.R. Dexter, G. Richard, D. Arrouays, E.A. Czyż, C. Jolivet og O. Duval. *Complexed organic matter controls soil physical properties*. *Geoderma*, 144 (3), 620–627, 2008. ISSN 0016-7061. doi: [10.1016/j.geoderma.2008.01.022](https://doi.org/10.1016/j.geoderma.2008.01.022). URL <https://www.sciencedirect.com/science/article/pii/S0016706108000219>.
- EU-Kommissionen, u.d.** EU-Kommissionen. *Soil strategy for 2030*, Jordbundsstrategi for 2030, u.d. URL https://environment.ec.europa.eu/topics/soil-and-land/soil-strategy_en. Tilgået 22-02-2024.
- EU-Kommissionen, 2021.** EU-Kommissionen. *EU's jordbundsstrategi for 2030: Udnyttelse af fordelene ved en sund jordbund for mennesker, fødevarer, natur og klima*, COM/2021/699 final., 2021. URL <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021DC0699>.
- European Commission et al., 2020.** European Commission, Joint Research Centre, A Jones, O Fernández-Ugalde og S Scarpa. *LUCAS 2015 Topsoil Survey. Presentation of dataset and results*. Publications Office, 2020. doi: doi.org/10.2760/616084.
- European Commission et al., 2022.** European Commission, Joint Research Centre, O Fernández-Ugalde, A Orgiazzi, A Marechal, A Jones, S Scarpa, P Panagos og M Van Liedekerke. *LUCAS 2018 soil module – Presentation of dataset and results*. Publications Office of the European Union, 2022. doi: doi.org/10.2760/215013.
- European Commission Directorate-General Joint Research Centre, 2018.** European Commission Directorate-General Joint Research Centre. *Dry Matter Productivity*, 2018. URL <https://land.copernicus.eu/en/products/vegetation/dry-matter-productivity-v1-0-300m>.
- European Environment Agency, 2022.** European Environment Agency. *Soil carbon*. <https://www.eea.europa.eu/publications/soil-carbon>, 2022. Tilgået: 27-02-2024.
- FAO, 2018.** FAO. *The future of food and agriculture - Alternative pathways to 2050*, page 224. Rom, Italien. Food and Agriculture Organization of the United Nations. Report, 2018. ISBN 978-92-5-133582-6. URL <https://www.fao.org/3/I8429EN/i8429en.pdf>.

- FAO et al., 2020.** FAO, ITPS, GSBI, SCBD og EC. *State of knowledge of soil biodiversity - Status, challenges and potentialities*, page 618. Rom, Italien. Food and Agriculture Organization of the United Nations. Report, 2020. ISBN 978-92-5-130158-6. URL <https://doi.org/10.4060/cb1928en>.
- Ferrer Julià et al., 2004.** M. Ferrer Julià, T. Estrela Monreal, A. Sánchez del Corral Jiménez og E. García Meléndez. *Constructing a saturated hydraulic conductivity map of Spain using pedotransfer functions and spatial prediction*. *Geoderma*, 123(3), 257–277, 2004. ISSN 0016-7061. doi: <https://doi.org/10.1016/j.geoderma.2004.02.011>. URL <https://www.sciencedirect.com/science/article/pii/S001670610400062X>.
- Frost, 2023.** Jim Frost. *Multicollinearity in Regression Analysis: Problems, Detection, and Solutions*, 2023. URL <https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/>. Tilgået: 02-10-2024.
- Gallo et al., 10 2018.** Bruna Gallo, José Alexandre Demattê, Rodnei Rizzo, José Safanelli, Wanderson de Sousa Mendes, Igo Lepsch, Marcus Sato, Danilo Romero og Marilusa Lacerda. *Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with Soil Classes and Geology*. *Remote Sensing*, 10, 1571, 2018. doi: 10.3390/rs10101571. URL <https://doi.org/10.3390/rs10101571>.
- Gatea et al., 08 2018.** Ahmed Gatea, Abbas Kouzani, Akif Kaynak, Sui Yang Khoo, Michael Norton og Will Gates. *Soil Bulk Density Estimation Methods: A Review*. *Pedosphere*, 28, 581–596, 2018. doi: 10.1016/S1002-0160(18)60034-7.
- Geographic Information System of the Commission, 2024.** Geographic Information System of the Commission. *Geodata*, 2024. URL <https://ec.europa.eu/eurostat/web/gisco/geodata>.
- Gholizadeh et al., 2018.** Asa Gholizadeh, Daniel Žižala, Mohammadmehdi Saberioon og Luboš Borůvka. *Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging*. *Remote Sensing of Environment*, 218, 89–103, 2018. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2018.09.015>. URL <https://www.sciencedirect.com/science/article/pii/S0034425718304267>.
- Gong, 2022.** Liyan Gong. *Soil Electrical Conductivity(EC): What's It, Why Important, How to Measure & More*, 2022. URL <https://www.seedstudio.com/blog/2022/07/15/soil-electrical-conductivity>.
- Google, ud.** Google. *What is Supervised Learning?*, ud. URL <https://cloud.google.com/discover/what-is-supervised-learning>.
- Gupta og Larson, 12 1979.** Satish Gupta og W. Larson. *Estimating Soil Water Retention Characteristics From Particle Size Distribution, Organic Matter Percent, and Bulk Density*. *Water Resources Research - WATER RESOUR RES*, 15, 1633–1635, 1979. doi: 10.1029/WR015i006p01633.
- Hughes et al., 2023.** H.M. Hughes, S. Koolen, M. Kuhnert, E.M. Baggs, S. Maund, G.W. Mullier og J. Hillier. *Towards a farmer-feasible soil health assessment that is globally applicable*. *Journal of Environmental Management*, 345, 118582, 2023. ISSN

- 0301-4797. doi: <https://doi.org/10.1016/j.jenvman.2023.118582>. URL <https://www.sciencedirect.com/science/article/pii/S0301479723013701>.
- Janzen et al., 2021.** H. Henry Janzen, David W. Janzen og Edward G. Gregorich. *The 'soil health' metaphor: Illuminating or illusory?* Soil Biology and Biochemistry, 159, 108167, 2021. ISSN 0038-0717. doi: <https://doi.org/10.1016/j.soilbio.2021.108167>. URL <https://www.sciencedirect.com/science/article/pii/S0038071721000390>.
- Juliá et al., 06 2021.** Fernando Juliá, Victor Snyder og Miguel Vazquez. *Validation of Soil Survey Estimates of Saturated Hydraulic Conductivity in Major Soils of Puerto Rico*. Hydrology, 8, 94, 2021. doi: 10.3390/hydrology8030094.
- Loll og Moldrup, 2000.** Per. Loll og Per. Moldrup. *Soil characterization and polluted soil assessment*. Aalborg University, Aalborg, 2000.
- Maaz et al., 2023.** Tai McClellan Maaz, Ronald H. Heck, Christine Tallamy Glazer, Mitchell K. Loo, Johanie Rivera Zayas, Aleric Krenz, Tanner Beckstrom, Susan E. Crow og Jonathan L. Deenik. *Measuring the immeasurable: A structural equation modeling approach to assessing soil health*. Science of The Total Environment, 870, 161900, 2023. ISSN 0048-9697. doi: <https://doi.org/10.1016/j.scitotenv.2023.161900>. URL <https://www.sciencedirect.com/science/article/pii/S0048969723005156>.
- Mahender et al., 2019.** Anumalla Mahender, B. P. Mallikarjuna Swamy, Annamalai Anandan og Jauhar Ali. *Tolerance of Iron-Deficient and -Toxic Soil Conditions in Rice*. Plants, 8(2), 2019. ISSN 2223-7747. doi: 10.3390/plants8020031. URL <https://www.mdpi.com/2223-7747/8/2/31>.
- Meter Group, ud.** Meter Group. *How to measure soil hydraulic conductivity— Which method is right for you?*, ud. URL <https://metergroup.com/measurement-insights/how-to-measure-hydraulic-conductivity-which-method-is-right-for-you/>.
- Minnesota Pollution Control Agency, 2022a.** Minnesota Pollution Control Agency. *Soil chemical properties and processes*, 2022a. URL https://stormwater.pca.state.mn.us/index.php?title=Soil_chemical_properties_and_processes.
- Minnesota Pollution Control Agency, 2023.** Minnesota Pollution Control Agency. *Soil hydrologic properties and processes*, 2023. URL https://stormwater.pca.state.mn.us/index.php?title=Soil_hydrologic_properties_and_processes.
- Minnesota Pollution Control Agency, 2022b.** Minnesota Pollution Control Agency. *Soil physical properties and processes*, 2022b. URL https://stormwater.pca.state.mn.us/index.php?title=Soil_physical_properties_and_processes.
- Mohammed og Kora, 2023.** Ammar Mohammed og Rania Kora. *A comprehensive review on ensemble deep learning: Opportunities and challenges*. Journal of King Saud University - Computer and Information Sciences, 35(2), 757–774, 2023. ISSN 1319-1578. doi: <https://doi.org/10.1016/j.jksuci.2023.01.014>. URL <https://www.sciencedirect.com/science/article/pii/S1319157823000228>.

- Munyaneza et al., 03 2024.** Venuste Munyaneza, Wen Zhang, Sharjeel Haider, Fangsen Xu, Chuang Wang og Guangda Ding. *Strategies for alleviating aluminum toxicity in soils and plants*. Plant and Soil, pages 1–24, 2024. doi: 10.1007/s11104-024-06617-7.
- NSW Department of Primary Industries, 2000.** NSW Department of Primary Industries. *Understanding Soil pH*, 2000. URL https://www.dpi.nsw.gov.au/__data/assets/pdf_file/0003/167187/soil-ph.pdf. Tilgået: 10-09-2024.
- PyCaret, 2024.** PyCaret. *PyCaret Documentation*, 2024. URL <https://pycaret.gitbook.io/docs>.
- Rakshit et al., 2020.** Amitava. Rakshit, Subhadip. Ghosh, Somsubhra. Chakraborty, Varughese. Philip og Avishek. Datta. *Soil Analysis: Recent Trends and Applications*. Springer Singapore, Singapore, 1st ed. 2020. edition, 2020. ISBN 981-15-2039-9.
- Ritchie og Roser, 2024.** Hannah Ritchie og Max Roser. *Land Use*. Our World in Data, 2024. URL <https://ourworldindata.org/land-use>. Først udgivet i september 2019 og sidst revideret i maj 2024.
- Ros et al., 2022a.** Gerard H. Ros, Sven E. Verweij, Sander J. C. Janssen, Janjo De Haan og Yuki Fujita. *An Open Soil Health Assessment Framework Facilitating Sustainable Soil Management*. Environmental Science & Technology, 56(23), 17375–17384, 2022a. doi: 10.1021/acs.est.2c04516. URL <https://doi.org/10.1021/acs.est.2c04516>. PMID: 36399796.
- Ros et al., 2022b.** Gerard H. Ros, Sven E. Verweij, Sander J. C. Janssen, Janjo De Haan og Yuki Fujita. *An Open Soil Health Assessment Framework Facilitating Sustainable Soil Management*. Environmental Science & Technology, 56(23), 17375–17384, 2022b. doi: 10.1021/acs.est.2c04516. URL <https://doi.org/10.1021/acs.est.2c04516>. PMID: 36399796.
- Schjønning et al., 2010.** P. Schjønning, L. W. de Jonge, P. Møldrup, B. T. Christensen og J. E. Olesen. *Searching the critical soil organic carbon threshold for satisfactory tilth conditions – test of the Dexter clay:carbon hypothesis*. Aarhus Universitet, Det Jordbrugsvidenskabelige Fakultet. 1st International Conference and Exploratory Workshop on Soil Architecture and Physico-Chemical Functions "CESAR" - Foulum, Danmark, pages 341–346, 2010. URL <https://orgprints.org/id/eprint/18441/1/18441.pdf>.
- Schjønning et al., 2009.** Per Schjønning, Lis W. de Jonge, Jørgen E. Olesen og Mogens H. Greve. *Organic farming effects on clay dispersion in carbon-exhausted soils*. ICROFS News, p. 8-10, 2009. URL <https://orgprints.org/id/eprint/18442/>.
- Schreiber-Gregory, 2017.** Deanna Naomi Schreiber-Gregory. *Multicollinearity in Regression Models: Problems and Solutions*, National University, SAS Institute, 2017. URL <https://support.sas.com/resources/papers/proceedings17/1404-2017.pdf>. Tilgået: 02-10-2024.
- Scikit-learn developers, 2024.** Scikit-learn developers. *API Reference*, 2024. URL <https://scikit-learn.org/stable/api/index.html>. Tilgået 22-09-2024.

- Simic, 2024.** Milos Simic. *Random Forest vs. Extremely Randomized Trees*, 2024. URL <https://www.baeldung.com/cs/gradient-boosting-trees-vs-random-forests>.
- ST Biologicals, 2024.** ST Biologicals. *How Iron Affects Soil Health and Plant Growth*, 2024. URL <https://stbiologicals.com/how-iron-affects-soil-health-and-plant-growth/>.
- Tempfli et al., 2009.** K. Tempfli, G.C. Huurneman, W.H. Bakker, L.L.F. Janssen, W.F. Feringa, A.S.M. Gieske, K.A. Grabmaier, C.A. Hecker, J.A. Horn, N. Kerle, F.D. van der Meer, G.N. Parodi, C. Pohl, C.V. Reeves, F.J.A. van Ruitenbeek, E.M. Schetselaar, M.J.C. Weir, E. Westinga og T. Woldai. *Principles of remote sensing: an introductory textbook*. ITC Educational Textbook Series. International Institute for Geo-Information Science and Earth Observation, 2009. ISBN 978-90-6164-270-1. URL http://www.itc.nl/library/papers_2009/general/PrinciplesRemoteSensing.pdf.
- United Nations, 2024.** United Nations. *World Population Dashboard*. United Nations, 2024. <https://www.unfpa.org/data/world-population-dashboard>.
- van der Meer et al., 2014.** F. D. van der Meer, H.M.A. van der Werff og F.J.A. van Ruitenbeek. *Potential of ESA's Sentinel-2 for geological applications*. Remote Sensing of Environment, 148, 124–133, 2014. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2014.03.022>. URL <https://www.sciencedirect.com/science/article/pii/S0034425714001084>.
- Xu et al., 2024.** Zhengyuan Xu, Lingzhen Ye, Qiufang Shen og Guoping Zhang. *Advances in the study of waterlogging tolerance in plants*. Journal of Integrative Agriculture, 23(9), 2877–2897, 2024. ISSN 2095-3119. doi: <https://doi.org/10.1016/j.jia.2023.12.028>. URL <https://www.sciencedirect.com/science/article/pii/S2095311923004756>.
- Zolfaghari Nia et al., 2023.** Masoud Zolfaghari Nia, Mostafa Moradi, Gholamhosein Moradi og Ruhollah Taghizadeh-Mehrjardi. *Machine Learning Models for Prediction of Soil Properties in the Riparian Forests*. Land, 12(1), 2023. ISSN 2073-445X. doi: 10.3390/land12010032. URL <https://www.mdpi.com/2073-445X/12/1/32>.

