



KORTLÆGNING AF DRÆNEDE AREALER I DANMARK

ANDERS BJØRN MØLLER, CHRISTEN DUUS BØRGESEN, EVA OVERBY BACH, BO VANGSØ IVERSEN
OG BJARNE MOESLUND

DCA RAPPORT NR. 135 · NOVEMBER 2018



AARHUS
UNIVERSITET

DCA - NATIONALT CENTER FOR FØDEVARER OG JORDBRUG

KORTLÆGNING AF DRÆNEDE AREALER I DANMARK

DCA RAPPORT NR. 135 · NOVEMBER 2018



Anders Bjørn Møller¹⁾, Christen Duus Børgesen¹⁾, Eva Overby Bach¹⁾, Bo Vangsø Iversen¹⁾ og
Bjarne Moeslund²⁾

Aarhus Universitet

¹⁾ Institut for Agroøkologi v. Aarhus Universitet

²⁾ Orbicon

Blichers Allé 20

Postboks 50

8830 Tjele

KORTLÆGNING AF DRÆNEDE AREALER I DANMARK

- Serietitel DCA rapport
- Nr.: 135
- Forfattere: Anders Bjørn Møller, Christen Duus Børgesen, Eva Overby Bach,
Bo Vangso Iversen og Bjarne Moeslund
- Udgiver: DCA - Nationalt Center for Fødevarer og Jordbrug, Blichers Allé 20,
postboks 50, 8830 Tjele. Tlf. 8715 1248, e-mail: dca@au.dk, hjemmeside:
www.dca.au.dk
- Rekvirent: Miljø- og Fødevareministeriet, NaturErhvervstyrelsen
- Fagfælle-
bedømt: Landskonsulent Søren Kolind Hvid (SEGES) har gransket rapporten.
- Fotograf: Forsidefoto: Anders Bjørn Møller. Ortofoto fra Styrelsen for Dataforsyning og
Effektivisering.
- Tryk: www.digisource.dk
- Udgivelsesår: 2018
- Gengivelse er tilladt med kildeangivelse
- ISBN: Trykt version 978-87-93643-87-1, elektronisk version 978-87-93643-88-8
- ISSN: 2245-1684
- Rapporterne kan hentes gratis på www.dca.au.dk

Rapport

Rapporterne indeholder hovedsageligt afrapportering fra forskningsprojekter, oversigtsrapporter over faglige emner, vidensynteser, rapporter og redegørelser til myndigheder, tekniske afprøvninger, vejledninger osv.

Forord

NaturErhvervstyrelsen (NAER Miljø og Fødevarerministeriet) anmodede i henvendelse af 10. januar 2014 DCA – Nationalt Center for Fødevarer og Jordbrug – om en beskrivelse af et projekt til undersøgelse af, hvordan en bedre og mere sikker kortlægning af drænedede arealer kan foretages (Bilag 7). På baggrund af projektbeskrivelsen (Bilag 8) er der efterfølgende gennemført databaserede analyser af sammenhængen mellem dræning og en lang række geografiske datalag. Resultatet er et landsdækkende kort, der viser den sandsynlige udbredelse af drænedede henholdsvis ikke drænedede arealer.

Kortet har en opløsning på 30,4 x 30,4 meter og en nøjagtighed på 79%. Med kortet følger desuden et datalag over sandsynligheden for dræning samt datalag til vurdering af kortets nøjagtighed i specifikke områder.

Denne DCA-rapport præsenterer de udarbejdede kort samt de metoder og data, der er anvendt ved udarbejdelsen af kortene. Projektet og dets resultater indgår som en del af Anders Bjørn Møllers Ph.d.-projekt med titlen "Mapping soil properties and drainage systems with machine learning algorithms".

Indholdsfortegnelse

Forord	3
Resumé	7
1 Introduktion	11
1.1 Baggrund.....	11
1.2 Formål og omfang.....	14
2 Metoder og materialer	15
2.1 Datagrundlag for kortet.....	16
2.1.1 Data fra Olesen (2009).....	16
2.1.2 Data fra Orbicons drænarkiv.....	17
2.1.3 Data fra Landskontoret for planteavl 1973	19
2.1.4 Geografiske datalag.....	20
2.1.5 Endelige datasæt	23
2.2 Kortlægning.....	24
2.3 Kombination af kort	25
2.4 Kortets nøjagtighed	26
2.4.1 Hoveddræn	26
2.4.2 Drænudløb.....	28
2.4.3 Drænprocenter på markniveau.....	28
2.4.4 Drænrør i kabelgrave	29
3 Resultater	32
3.1 Nøjagtighed af de enkelte modeller.....	32
3.2 Udvalgte modeller.....	33
3.3 Kortlægning.....	37
3.3 Kombination af kort	38
3.4 Kortets nøjagtighed	40
3.4.1 Hoveddræn	41
3.4.2 Drænudløb.....	43
3.4.3 Drænprocenter på markniveau.....	43
3.4.4 Drænrør i kabelgrave	44
4 Diskussion	47
4.1 Geografiske datalag.....	47
4.2 Observationer.....	50
4.3 Nøjagtighed.....	51
4.4 Anvendelse og usikkerheder	52
5 Konklusion.....	57

Tak til.....	58
Referencer.....	59
Bilag 1: Geografiske datalag.....	65
Bilag 2: Drænkategorier for afgroeder.....	67
Bilag 3: Begreber	73
Bagging.....	73
Beslutningstræer	73
Boosting.....	73
Bootstrap sampling	73
Diskriminantanalyse.....	74
Ensemble.....	74
k-nærmeste naboer.....	74
Kunstige neurale netværk	74
Lineære modeller.....	74
Rasterfil.....	75
Support Vector Machines	75
Bilag 4: Test af metoden til kombination af modeller	76
Opbygning af modeller.....	76
Kombination af modeller	79
Resultater	81
Bilag 5: Nøjagtighed i udvalgte områder	84
Bilag 6: Overensstemmelse mellem kort og registrerede drænrør	89
Bilag 7: Bestilling	92
Bilag 8: Projektbeskrivelse.....	98
Bilag 9: Møller et al. (2018)	115

Resumé

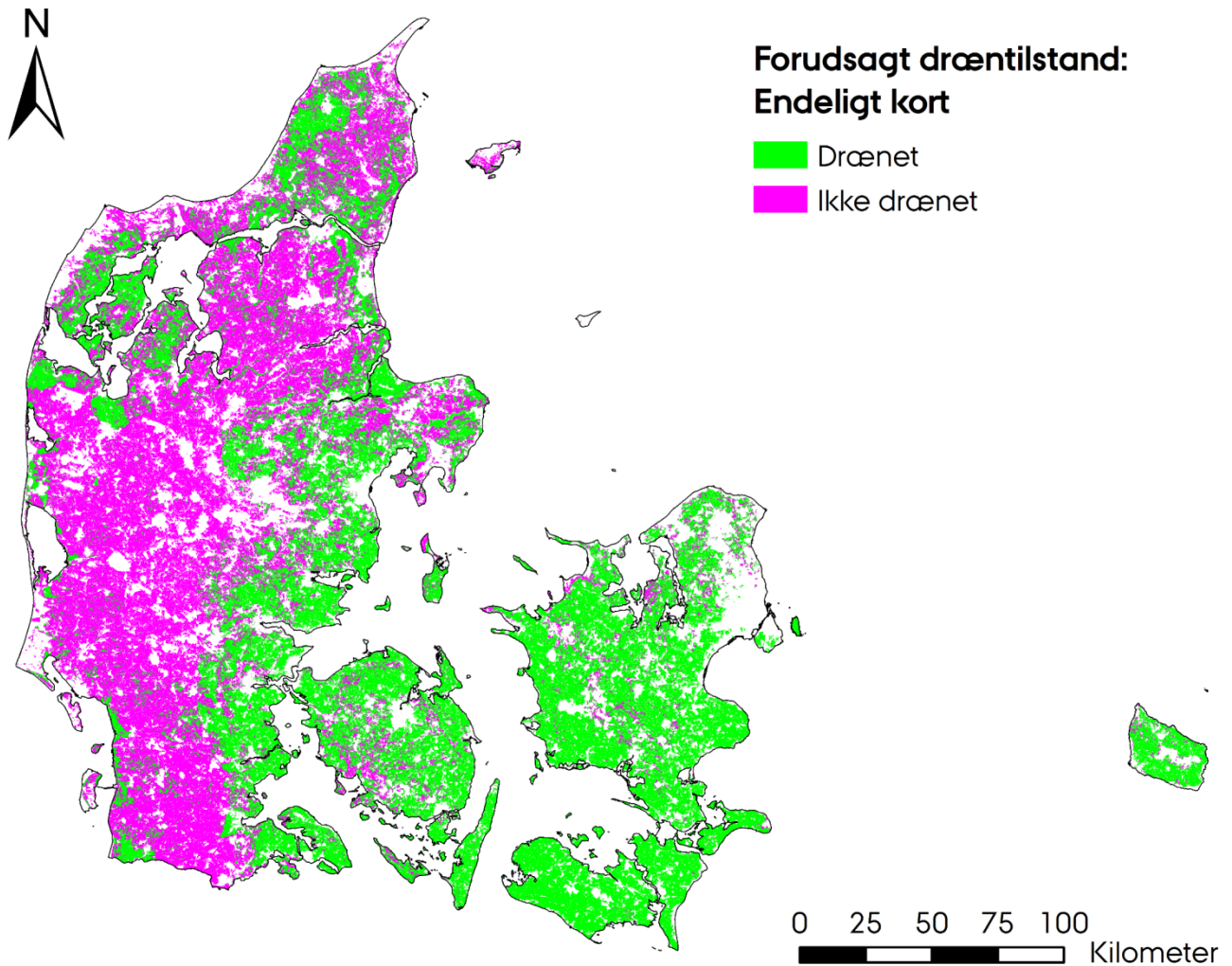
Denne rapport præsenterer et nyt landsdækkende kort over drænede markarealer i Danmark. Kortet viser i en opløsning på 30,4 meter den sandsynlige udbredelse af drænede og ikke drænede arealer. Kortet er baseret på statistiske analyser af sammenhængen mellem dræning og en lang række geografiske datalag, heriblandt topografiske variabler, jordbundsforhold, satellitbilleder m.fl.

Udarbejdelsen af det nye kort er foretaget i to faser. I undersøgelsens første fase blev der udarbejdet et kort ved anvendelse af observationerne fra kortlægningen af drænede arealer udført af Olesen (2009). I undersøgelsens anden fase blev datasættet udvidet med data fra Landskontoret for Planteavl's undersøgelse (Skriver og Hedegård, 1973) og fra Orbicons drænarkiv (Orbicon, 2015).

I begge faser af undersøgelsen blev der anvendt en teknik, hvor der først opbygges et stort antal statistiske modeller. Derefter udvælger de modeller, der tilsammen giver den mest nøjagtige forudsigtelse af drænede arealer for et uafhængigt datasæt. Til sidst kombineredes modellernes forudsigelser til et landsdækkende kort over den mest sandsynlige udbredelse af drænede og ikke drænede arealer.

I undersøgelsens første fase blev der udvalgt 36 relativt enkle modeller, der gav en samlet nøjagtighed på 77%, beregnet på et uafhængigt datasæt (Møller et al., 2018). I undersøgelsens anden fase blev der udvalgt 14 mere komplekse modeller, som gav en samlet nøjagtighed på 78%. I begge tilfælde skal nøjagtigheden forstås som sandsynligheden for, at drænede og ikke drænede arealer gengives korrekt på kortet.

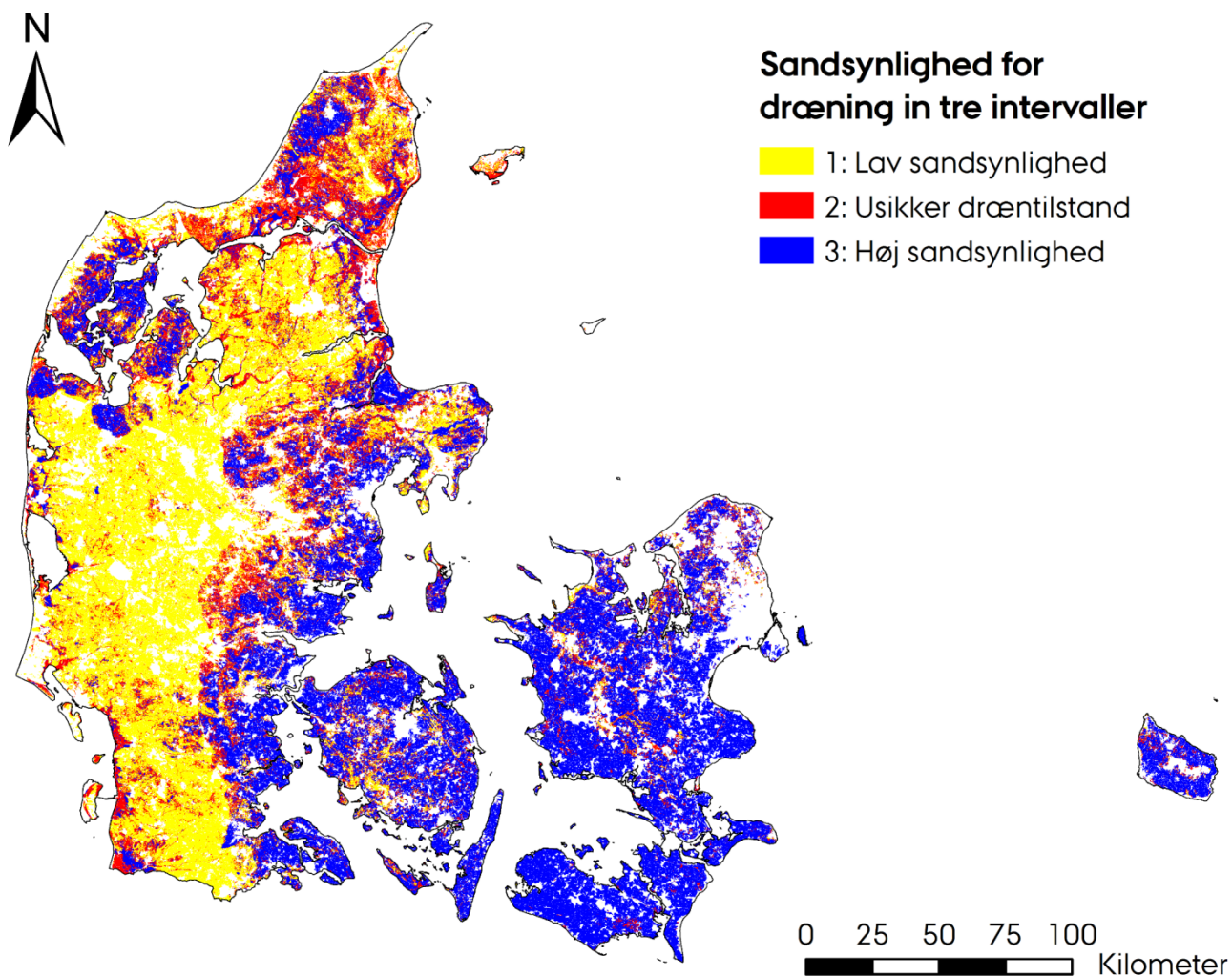
Kortene, der blev udarbejdet i hver af de to faser af undersøgelsen, blev til slut kombineret til ét kort, sammensat af de områder, hvor det ene eller det andet kort gav den største nøjagtighed. Kortet fra undersøgelsens anden fase var mest nøjagtigt på Littorinafladerne (hævet havbund) og i Midtjylland. Kortet fra undersøgelsens anden fase blev derfor anvendt for disse områder, mens kortet fra undersøgelsens første fase blev anvendt for alle øvrige områder. På det endelige kort er 52% af markarealet drænet (Figur 1).



Figur 1: Det endelige kort over drænedede markarealer udarbejdet i denne undersøgelse.

En sammenligning med et uafhængigt datasæt viste, at kortet har en gennemsnitlig nøjagtighed på cirka 79%. Det vil sige, at kortet med en sandsynlighed på cirka 79% forudsiger, om et givent markareal er drænet henholdsvis ikke drænet. Nøjagtigheden var mindst i områder, hvor den forudsagte sandsynlighed for dræning var tæt på 50%. Omvendt var nøjagtigheden størst i områder, hvor den forudsagte sandsynlighed for dræning var markant større eller mindre end 50%.

På denne baggrund blev sandsynligheden for dræning inddelt i tre klasser: Lav sandsynlighed for dræning, usikker forudsigtelse og høj sandsynlighed for dræning (Figur 2).



Figur 2: Sikkerhed forbundet med forudsigelsen af drænedede arealer, beregnet ud fra den forudsagte sandsynlighed for dræning: Lav sandsynlighed for dræning (0 – 40% sandsynlighed, 82% sikkerhed), usikker forudsigelse (40 – 60% sandsynlighed, 70% sikkerhed) og høj sandsynlighed for dræning (60 – 100% sandsynlighed, 80% sikkerhed).

Det endelige korts nøjagtighed blev desuden beregnet på grundlag af en række uafhængige data. De uafhængige data omfattede hoveddræn kortlagt af Orbicon, punkter med målt drænaftømning, drænprocenter kortlagt på markniveau og drænrør kortlagt langs Energinets kabelstrækninger.

Sammenligningen med de uafhængige datasæt viste, at der generelt er god sammenhæng mellem den forudsagte og den observerede forekomst af dræning, og at langt de fleste drænedede arealer er dækket af kortet. Kortet er desuden mere nøjagtigt i sandede end i lerede områder. I lerede områder, og især i dødislandskaber, er den forudsagte drænprocent væsentligt højere end den observerede.

Mange af de vigtigste variabler for forudsigelsen af dræning, eksempelvis jordens lerindhold og dybden til grundvandet har en kendt betydning for behovet for dræning, men også lokale forhold som terrænet kan have stor betydning for drænbehovet og for den faktiske dræning.

Samlet set viser resultaterne af de gennemførte analyser, at det nye kort er mere nøjagtigt end kortet udarbejdet af Olesen (2009) i de fleste områder og især på randmoræner og på sandede ferskvandsaflejringer.

1 Introduktion

1.1 Baggrund

Dræning af jorden har først og fremmest betydning for dens egnethed til planteproduktion, men har også stor betydning for omsætning af næringsstoffer i jorden og for de mekanismer, der styrer transporten af vand og næringsstoffer fra marken og frem til vandløb, søer og kystvande.

Når jorden vandmættes, mindskes luftskiftet mellem jorden og atmosfæren. Ilt bliver mindre tilgængeligt, hvilket hindrer afgrødernes rødder i at vokse og optage næringsstoffer (Brady og Weil, 1996). Vandmætning over længere tid er derfor skadeligt for mange afgrøder. Langvarig vandmætning i rodzonen kan således reducere udbyttet af hvede med 34 – 60% (Collaku og Harrison, 2002), udbyttet af majs med op til 35% (Ren et al., 2014) og udbyttet af byg og havre med henholdsvis 39% og 48% (Watson et al., 1976). Samtidig betyder vandmætning i foråret, at jorden opvarmes langsommere (Brady og Weil, 1996). Vandmættede jorde vil også være mere sårbare over for kompaktering (Hamza og Anderson, 2005).

Denne viden om betydningen af vandmætning i jorden førte allerede på et tidligt tidspunkt til, at man i landbrugsmæssig sammenhæng begyndte at sænke vandspejlet i de øvre jordlag gennem dræning.

Mens man til at begynde med især drænedde jorden med gravede grøfter, har den fremherskende metode i nyere tid været nedgravede drænrør, der kan fremme afvandingen af de øverste 100 – 150 cm af jorden. I Danmark blev de første drænrør nedgravet i 1848 (Madsen, 2010). Dræningen tog dog først for alvor fart, da det med vandløbsloven i 1859 blev lovligt at føre drænrør over anden mands ejendom (Porsmose, 2008). Samtidig blev det tilladt at føre drænvandet ud til søer og åer, og flertallet af vandløbene blev udrettet og uddybet for, at man kunne sænke vandspejlet på markerne til det niveau, der var nødvendigt for dyrkningen af afgrøder.

Dræningen i Danmark fandt primært sted i to perioder (Olesen, 2009). I den første periode, 1860 – 1900, var dræningen fokuseret på de lerede højbundslande i Østdanmark. Her har jorderne en høj bonitet, og de kunne nemt afvandes gennem dræning. I den anden periode, 1930 – 1970, var dræningen derimod fokuseret i det vestlige Danmark og især på lavbundslande.

Fra 1921 til 1988, blev der ydet offentligt tilskud til dræning, blandt andet for at fremme beskæftigelsen i landområderne (Olesen, 2009, Madsen, 2010, Nielsen, 2015).

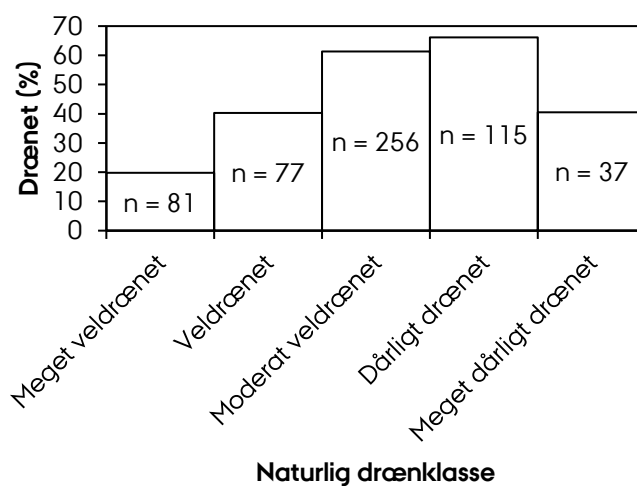
Indtil 1970'erne bestod drænrørene af brændt tegl, der fra begyndelsen af perioden og langt frem blev gravet ned med håndkraft (Porsmose, 2008, Madsen, 2010). Siden da har plastkrør erstattet teglrørene, og i dag bliver de i mange tilfælde ikke gravet ned, men i stedet "pløjet" ned i jorden med moderne maskiner (Nielsen, 2015). I 1800-tallet stod lokale drænmestre for de fleste af drænopgaverne, mens det Danske Hedeselskab i 1900-tallet projekterede og gennemførte de fleste drænprojekter (Madsen, 2010). I 1900-tallet begyndte Hedeselskabet at registrere beliggenheden af projekterne på målebordsblade. Det er

vurderet, at Hedeselskabets drænarkiv i dag rummer oplysninger om cirka 30% af alle drænedede arealer i Danmark (Olesen, 2009).

Flere undersøgelser har vist, at omkring halvdelen af det danske landbrugsareal er drænet med drænrør (Skriver og Hedegård, 1973, Olesen, 2009). Tallet har ikke ændret sig væsentligt siden 1970'erne, da de fleste drænaktiviteter siden da har drejet sig om gendræning af allerede drænedede jorde.

På højbundsjerne kan drænrør fungere i meget lang tid, i nogle tilfælde op til 100 år (Madsen, 2010). På lavbundsjerne fører dræningen derimod typisk til jordsætning, der forkorter drænenes levetid. Det gælder især grad humusjorde, da det sænkede grundvandsspejl øger tilførslen af ilt, hvilket fører til hurtig omsætning af det organiske materiale. Sådanne terrænsætninger betyder, at afvandingsdybden gradvis mindskes, og at man som følge heraf har et grundvandsspejl tæt ved terrænoverfladen i løbet af få år (Olesen, 2009, Madsen, 2010). I nogle tilfælde er den forringede afvandingsdybde som følge af sætning blevet modvirket gennem fornyet hovedafvanding ved omlægning af dræn og uddybning af vandløbene. I mange andre tilfælde er den forringede afvandingsdybde primært søgt imødegået gennem intensiveret vandløbsvedligeholdelse i form af grødeskæring og oprensning.

Samlet set har dræningen i Danmark fundet sted gennem en meget lang periode, og der er derfor mange forhold, der kan have afgjort, om et areal er blevet drænet. Sammenholdes drænprocenten med jordens naturlige drænforhold (dvs. jordens naturgivne evne til at transportere den overskydende nedbør bort), ses der et generelt, men ikke entydigt mønster (Figur 3).



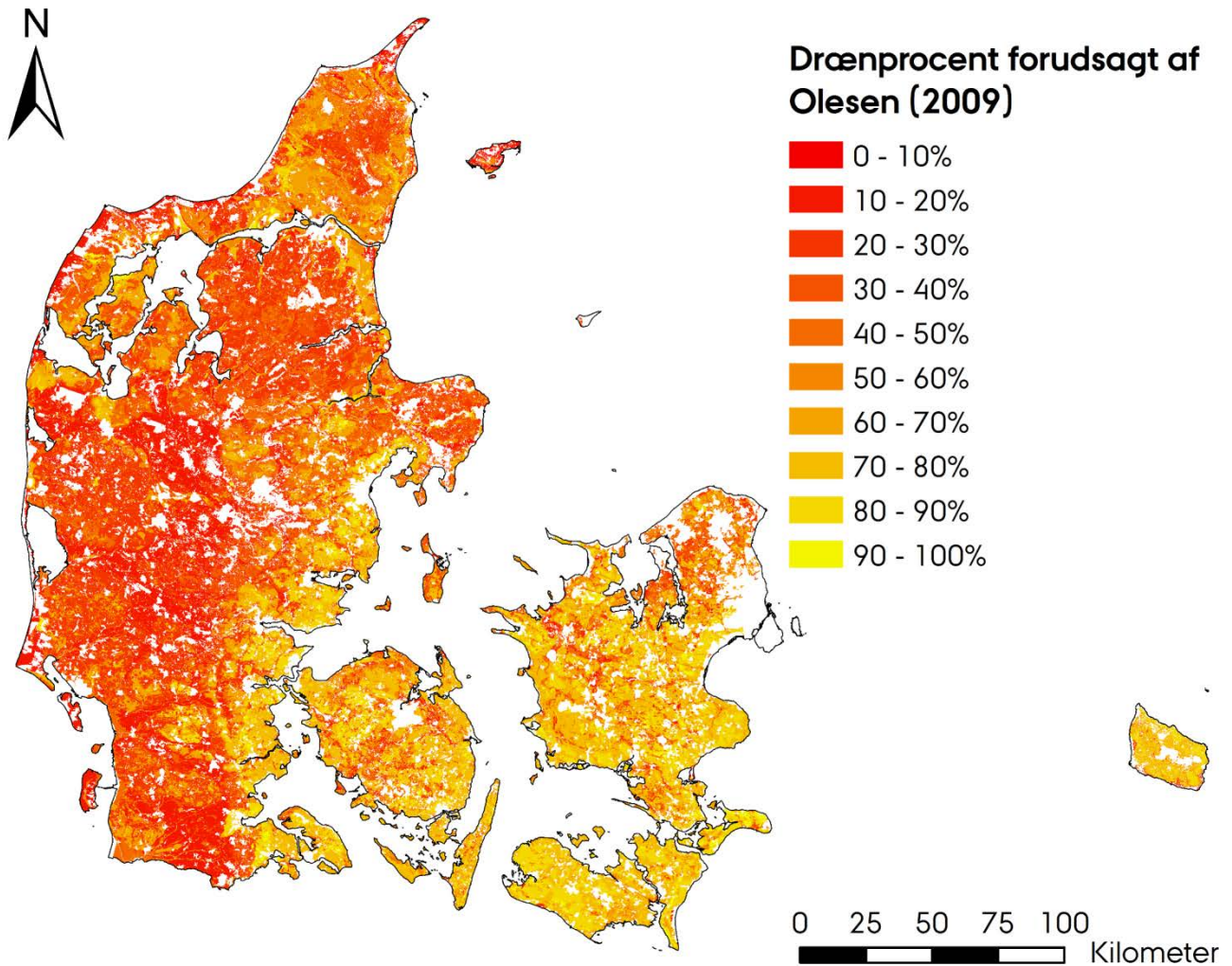
Figur 3: Drænprocenter for landbrugsjorde med hver af de naturlige drænklasser. Baseret på jordbundsprofiler i kvadratnettet, hvor der findes drænoplysninger (Olesen, 2009). Tallene i hver søjle angiver antallet af jordbundsprofiler inden for hver drænklasse. Meget veldrænedede jorde er fortrinsvist sandjorde (JB 1 - 4), mens veldrænedede jorde oftere er enten lerede sandjorde eller sandede lerjorde (JB 3 - 6). De moderat veldrænedede jorde er oftest lerede højbundsjerne (JB 5 - 7), mens dårligt og meget dårligt drænedede jorde primært findes på lavbund (Møller et al., 2017).

Nogle jorde er naturligt veldrænede. Det gælder eksempelvis højtliggende sandjorde. Men analysen af forholdet mellem dræning og de naturlige drænforhold viser, at der også findes drænrør på jorde, der er naturligt veldrænede, eller endda meget veldrænede. Det kan dels skyldes den økonomiske støtte til dræning, at man har drænet en hel mark på én gang for at sikre, at alle de vandlidende områder, der måtte være, blev drænet, eller at man har ført dræn fra vandlidende områder, f.eks. lavninger, gennem naturligt veldrænede områder. Omvendt ses det også, at ikke alle jorde, der fra naturens side er dårligt drænede, er blevet afvandret med drænrør. Det kan skyldes, at de af jordbundsmæssige og/eller terrænmæssige årsager er vanskelige at dræne, eller at arealerne anvendes til landbrugsmæssige formål med et mindre afvandingsbehov end dyrkede afgrøder, eksempelvis kreaturgræsning og høslæt.

Det kan også have betydning, at dræning kan påvirke den naturlige drænklasse. De naturlige drænklasser er defineret ud fra observationer i jordbundsprofilen så som dybden til grundvandspejlet, iltfrie forhold eller tørvelag. Ved dræning kan disse forhold ændres. For eksempel er meget dårligt drænede jorder blandt andet defineret ved, at de har et vandspejl mindre end 50 cm fra jordoverfladen. Dræning vil imidlertid sænke vandspejlet, og jorden vil derfor ikke længere kunne klassificeres som meget dårligt drænet. Dette kan muligvis forklare den relativt lave drænprocent for meget dårligt drænede jorde.

Selv om der er gode landbrugsmæssige grunde til at dræne, er der også negative miljømæssige konsekvenser ved dræning. Dræning kan blandt andet føre til tab af grundvandsafhængige naturtyper og øget udvaskning af okker (jern) til vandløb (Madsen, 2010). Dræning fremmer også transporten af næringsstoffer til vandløb og derfra videre til søer og kystvande, hvor de typisk fører til algeopblomstring og i værste fald iltsvind (Gambrell et al., 1975, Schelde et al., 2006, Ernstsén et al., 2015). Målrettede virkemidler som minivådområder og brug af markvirkemidler såsom efterafgrøder vil kunne nedsætte udvaskningen af næringsstoffer til vandmiljøet (Kovacic et al., 2000, Dinnes et al., 2002, Eriksen, 2014). Brugen af målrettede drænvirkemidler (vådområder, kunstige vådområder, kontrolleret dræning) forudsætter et detaljeret kendskab til drænede arealers forekomst, for eksempel for at sikre, at et vådområde kan dimensioneres i forhold oplandets størrelse.

Den seneste kortlægning af dræningens udbredelse i Danmark blev foretaget af Olesen (2009). Analysen resulterede i et kort, hvor alle arealer blev tildelt en sandsynlighed (%) for, om de var drænede. Kortet var baseret på en GIS-analyse på baggrund af GEUS' geologiske jordartskort (Jakobsen et al., 2015), kort over landskabselementer og vådområder og jordbundsklassificeringen (farvekodekortet) (Madsen et al., 1992). Gennem feltstudier blev det undersøgt, om der var drænet ved 745 punkter placeret i et landsdækkende kvadratnet. På basis af disse punkter blev alle områder (de mulige kombinationer af jordbund, geologi, landskabstype og georegion) tildelt en gennemsnitlig drænprocent (Figur 4).



Figur 4: Drænprocenter forudsagt af (Olesen, 2009).

Ifølge Olesen (2009) bør det udarbejdede kort ikke anvendes på en målestok finere end 1:100.000 på grund af de underliggende datalags grove opløsning. Kortet er derfor ikke egnet til at vurdere sandsynligheden for dræning på lokalt niveau endsige på markniveau.

1.2 Formål og omfang

Behovet for at vide, om givne arealer er drænede eller ej, er blevet aktuelt i forbindelse med anvendelsen målrettede virkemidler mod næringsstofudvaskning til vandmiljøet. Kendskab til drænsystemerne er derudover vigtigt for at kunne modellere næringsstofudvaskningen fra dyrkede arealer, da dræning på flere afgørende punkter påvirker udvaskningen. Kortet udarbejdet af Olesen (2009) har vist sig ikke at være fyldestgørende til disse formål på grund af den begrænsede nøjagtighed og opløsning, og fordi kortet udelukkende angiver drænprocenter og ikke de drænede arealers placering.

Nærværende undersøgelse er derfor gennemført med det formål at skabe et mere nøjagtigt kort jf. bestilling fra NAER 2014 (Bilag 7) og projektoplægget til bestillingen (Bilag 8).

2 Metoder og materialer

I denne rapport præsenteres et nyt kort over sandsynligheden for dræning på danske landbrugsarealer.

Den landsdækkende kortlægning af drænede arealer er i denne undersøgelse baseret på statistiske modeller, der kombinerer punktobservationer med geografiske datalag over variabler, som kan indikere, om arealer er drænede. Analyserne omfatter således data fra 46 geografiske datalag, samt punktdata fra Olesen (2009) og fra Skriver og Hedegård (1973) samt oplysninger fra Orbicons drænarkiv (Orbicon, 2015).

I undersøgelsen blev der først opbygget en lang række modeller, der potentielt kunne anvendes til at forudsige hvor der er drænet. Der er herefter anvendt et uafhængigt datasæt over observationer til at udvælge de modeller, der tilsammen giver den mest nøjagtige forudsigelse af om et landbrugsareal er drænet.

I undersøgelsens første fase blev der udarbejdet et landsdækkende kort over drænede arealer. Denne fase af undersøgelsen anvendte de samme punktobservationer som Olesen (2009) og er detaljeret gennemgået i Møller et al. (2018) (Bilag 9). Det første kort dannede udgangspunkt for den efterfølgende indsamling af yderligere drænobservationer.

Det blev vurderet, at kortet fra undersøgelsens første fase var især unøjagtigt på sandjorde, specielt i Himmerland, Vestjylland og de mest sandede dele af Østjylland. Orbicons kort over hoveddræn viste, at der i disse områder var dræning på lavtliggende arealer og specielt i ådalene. Det udarbejdede kort viste derimod næsten ingen drænede arealer i disse områder.

I undersøgelsens anden fase blev datagrundlaget derfor udvidet. Der blev sat fokus på indsamling af flere data for drænede sandjorde, specielt for områderne i Vestjylland, Himmerland og den sandede del af Østjylland. De nye data blev indsamlet fra Orbicons drænarkiv (Orbicon, 2015) og i datamaterialet fra en tidligere undersøgelse (Skriver og Hedegård, 1973).

Det nye datasæt blev anvendt til at udarbejde et nyt kort over drænede arealer. Det blev herefter beregnet i hvilke områder hvert af de to kort var mest nøjagtige. Det endelige kort blev udarbejdet ved at kombinere de to kort. Det endelige kort er således sammensat af de områder, hvor hvert af de to kort viser udbredelsen af drænede og ikke drænede arealer med størst nøjagtighed

Det endelige kort har en opløsning på 30,4 x 30,4 meter. Det viser både sandsynligheden for dræning, svarende til kortet udarbejdet af Olesen (2009), og udbredelsen af dræning, ved at en sandsynlighed på 50% bliver anvendt som skæringsværdi for, om et areal forventes at være drænet (>50% sandsynlighed) eller ikke drænet (\leq 50% sandsynlighed).

Det endelige korts nøjagtighed blev i øvrigt beregnet ved anvendelse af fire uafhængige datasæt (Validerings-datasæt). De uafhængige data omfattede hoveddræn kortlagt af Orbicon, punkter med målt

drænastrømning, drænpcenter kortlagt på markniveau og drænrør kortlagt i forbindelse med Energinets nedgravning af fire kabelstrækninger.

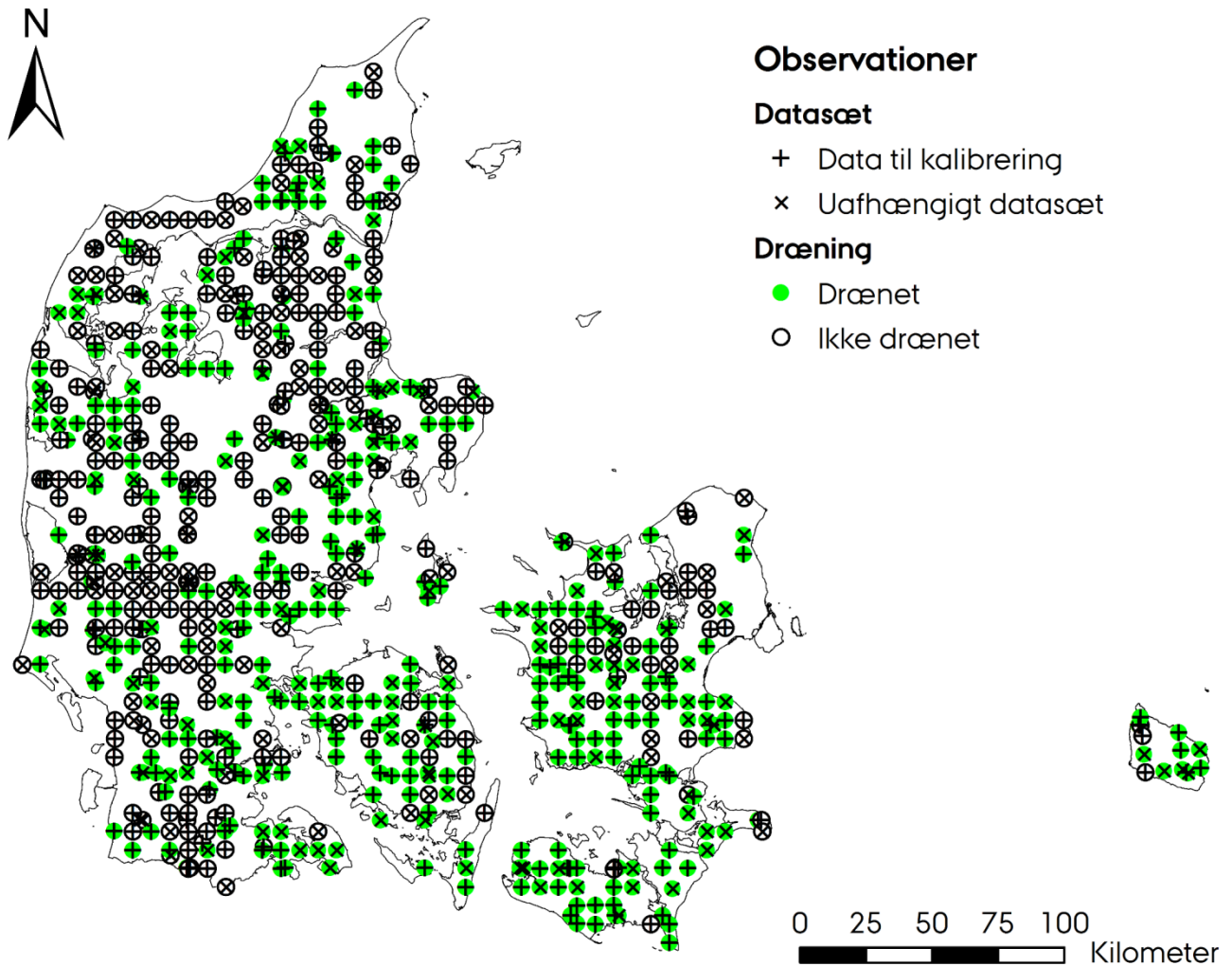
De følgende afsnit gennemgår datagrundlaget for kortet over drænedede arealer, de anvendte metoder og de uafhængige datasæt, som blev anvendt til at beregne kortets nøjagtighed.

2.1 Datagrundlag for kortet

2.1.1 Data fra Olesen (2009)

Undersøgelsens første fase (Møller et al., 2018) anvendte punktobservationerne fra Olesen (2009). Datasættet omfatter 745 punkter, hvor det ved hvert punkt er observeret, om arealet er drænet med drænrør. I 118 tilfælde gjaldt observationen for et areal i stedet for et punkt, men observationerne blev i alle tilfælde lagret som punkter. Arealobservationerne omfattede mellem 0,5 og 54 ha. hver, i gennemsnit 10 ha for hver observation.

I undersøgelsens første fase blev observationerne opdelt i to datasæt. Det første datasæt omfattede to tredjedele af observationerne og blev anvendt til at kalibrere modellerne. Det andet datasæt omfattede en tredjedel af observationerne og blev anvendt til at beregne kortets nøjagtighed. (Figur 5).



Figur 5: Punktobservationer drænedede og ikke drænedede arealer fra Olesen (2009) og opdelingen af dem i datasæt til kalibrering af modeller og et uafhængigt datasæt til vurdering af kortets nøjagtighed.

I undersøgelsens anden fase blev opdelingen af datasættet i to dele bevaret, så den første del (to tredjedele) blev anvendt til kalibrering og den anden del (en tredjedel) blev anvendt til en uafhængig beregning af kortets nøjagtighed. Da den del af datasættet, der blev anvendt til beregning af kortets nøjagtighed var det samme i undersøgelsens første og anden fase, kan den danne grundlag for at sammenligne de to kortes nøjagtighed.

2.1.2 Data fra Orbicons drænarkiv

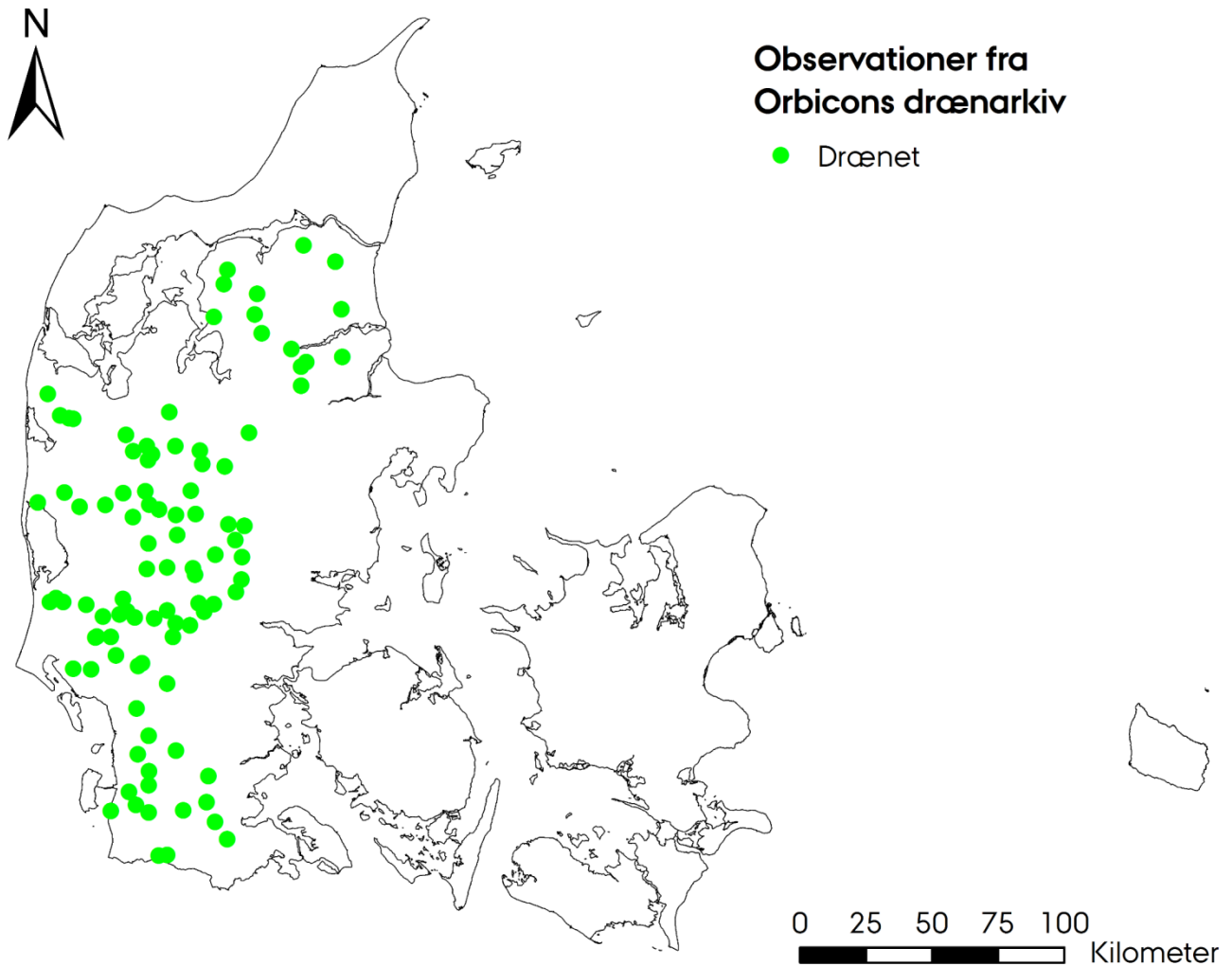
I undersøgelsens anden fase blev datasættet udvidet med data fra Orbicons drænarkiv (Orbicon, 2015) fra de arealer i Vestjylland, Himmerland og Østjylland, der var forudsagt som ikke drænedede på det første kort. Der blev først opstillet et kortlag over arealer i Vestjylland og Himmerland, som var forudsagt som ikke drænedede på kortet fra undersøgelsens første fase. For at kunne indsamle observationer repræsentativt, blev der genereret tilfældigt placerede punkter i disse tre landsdele. Disse punkter dannede herefter udgangspunkt for indsamling af nye oplysninger fra drænarkivet.

De nye oplysninger blev indsamlet fra oversigtskort, der viser drænprojekternes placering på topografiske kort. Fremgangsmåden bestod i at finde det nærmeste drænprojekt ved hvert af de tilfældigt placerede punkter. Følgende kriterier blev anvendt til at afgøre om et drænprojekt eller et punkt skulle udelades fra den videre analyse:

- Drænprojekter markeret som henlagte eller ikke-gennemførte på kortbladene blev valgt fra.
- Drænprojekter, hvor arealet var forudsagt som drænet på det første kort blev valgt fra.
- Hvis der ikke var nogen drænprojekter inden for de omkringliggende kortblade, eller hvis de alle var forudsagt som drænede på det første kort, blev det tilfældigt genererede punkt valgt fra.
- Hvis det nærmeste drænprojekt allerede var tilknyttet et andet punkt, blev det tilfældigt genererede punkt valgt fra.

I alt blev 132 tilfældigt placerede punkter undersøgt. Ved 97 af disse punkter blev et drænprojekt registreret, mens 35 punkter blev valgt fra på grund af de ovenstående kriterier.

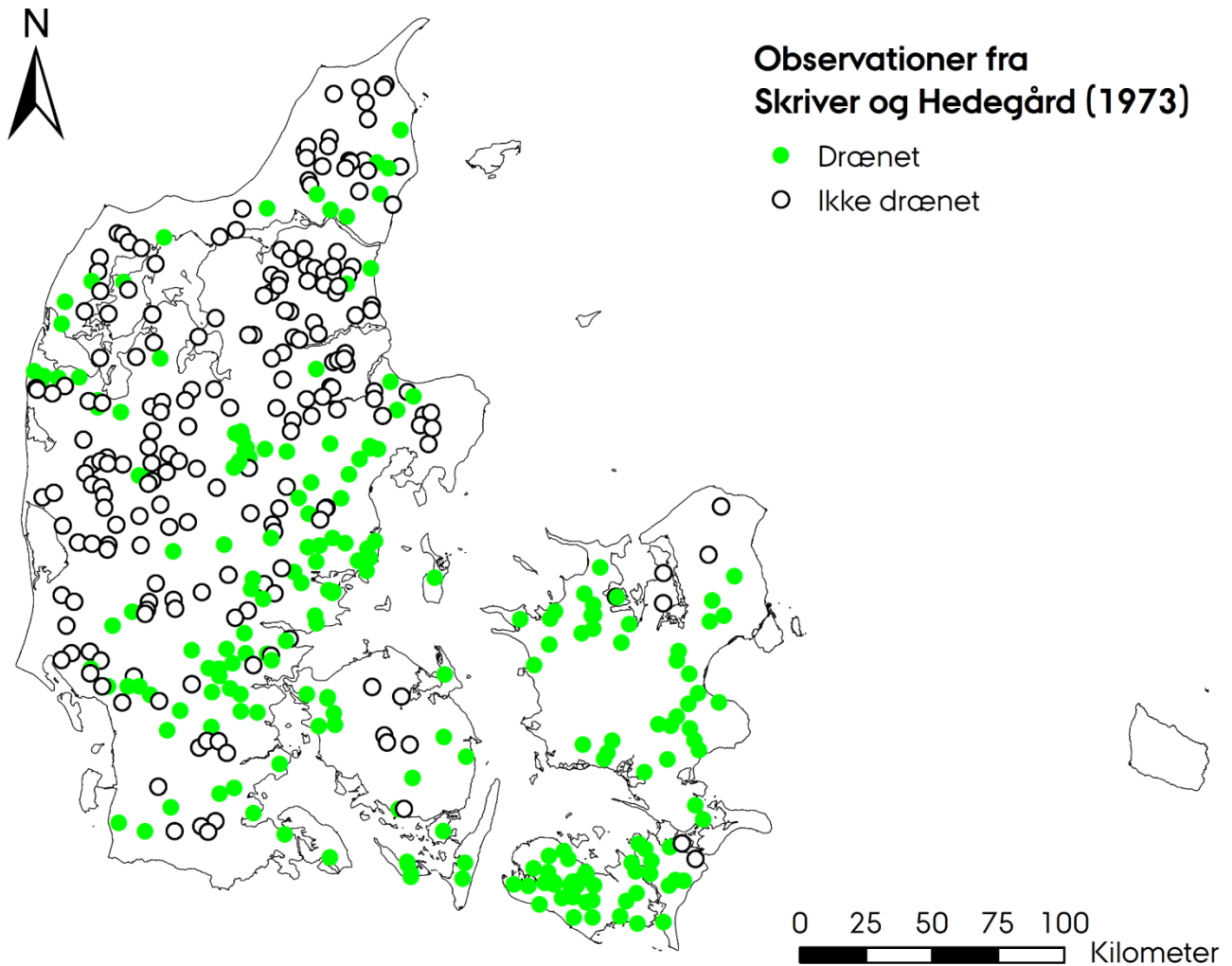
De 97 udvalgte drænprojekters udstrækning på oversigtskortene blev kortlagt som polygoner. Da undersøgelsens formål er at udarbejde et kort over drænede arealer, var det ikke nødvendigt at registrere de enkelte drænrør. For at sikre, at drænprojekterne lå inden for markarealet, blev polygonerne beskåret, så de kun omfattede dyrkede arealer registreret på Internet Markkort (IMK) i hele perioden 2011 – 2014 (Landbrugsstyrelsen, 2014). IMK indeholder landmændenes indmeldinger af dyrkede arealer til ansøgning om arealstøtte, og arealer udenfor IMK er derfor med stor sandsynlighed ikke dyrkede. Hvis dette kriterie fjernede mere end halvdelen af et polygon, blev metoden korrigeret således, at arealet kun behøvede at være med i IMK i ét af periodens fire år. Det korrigerede kriterium blev anvendt på 11 af de 97 drænprojekter. Til slut blev drænprojekterne konverteret til punkter (Figur 6).



Figur 6: Punkter med udvalgte drænprojekter fra Orbicons drænarkiv.

2.1.3 Data fra Landskontoret for planteavl 1973

I tillæg til ovenstående blev der anvendt markoplysninger fra Landskontoret for Planteavls undersøgelse (Skriver og Hedegård, 1973). Undersøgelsen blev foretaget på bedriftsniveau, hvor det blev angivet, hvor stor en del af bedriftens jord, der var drænet. Samtidig blev placeringen af den største mark i hver bedrift markeret som et punkt på et kort. Ved at udvælge bedrifter, hvor enten 0% eller 100% af arealet var drænet, kan det opfattes som sikkert, om der var drænet ved punktet. Af disse bedrifter var 414 placeret inden for det nuværende markareal med lige mange drænede og ikke drænede bedrifter (Figur 7).



Figur 7: Udvalgte drænedede og ikke drænedede punkter fra Landskontoret for Planteavl's undersøgelse (Skriver og Hedegård, 1973).

2.1.4 Geografiske datalag

Punktdata blev kombineret med geografiske datalag, der repræsenterer 46 forklarende variabler, som blev anset for potentielt relevante i relation til dræning (Bilag 1). Lagene indeholdt både variabler, som kan være afgørende for beslutningen om at dræne et areal, eksempelvis jordens indhold af ler, samt variabler, som dræningen kan påvirke. Eksempelvis kan satellitbilleder i nogle tilfælde aflæse dræningens effekt på jordens fugtighed, både fra planternes respons på vandindholdet i rodzonen og fra bar jord (Verma et al., 1996, Northcott et al., 2000, Tetzlaff et al., 2009).

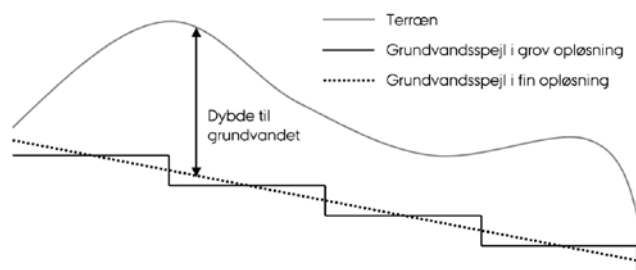
De topografiske variabler blev beregnet på baggrund af en digital højdemodel med en opløsning på 30,4 meter. Modellen blev genereret på baggrund af en højdemodel med en opløsning på 1,6 meter. Hver celle i den grovere højdemodel dækkede således over 19x19 celler, og højden blev beregnet som den gennemsnitlige højde i de disse underliggende celler. De fleste topografiske variabler beskriver forhold, der kan have direkte betydning for beslutningen om at dræne. De var således i de fleste tilfælde enten

forbundet med vandets strømningsveje i terrænet, arealets placering i forhold til overfladevand eller til dybden til grundvandet.

Vandets strømningsveje i terrænet kan have betydning for beslutningen om at dræne, da jorden vil være mere vandlidende, der hvor vandet naturligt samler sig og afstrømmer på eller under terrænet. Den akkumulerede afstrømning og det topografiske vådhedsindeks er eksempler på sådanne topografiske variabler med relevans i forhold til dræning. Den akkumulerede afstrømning angiver, hvor stort oplandet til et givent areal er. Det topografiske vådhedsindeks udregnes ud fra den akkumulerede afstrømning og terrænoverfladens hældning (Bilag 1), da en stor hældning vil betyde, at vandet hurtigt afstrømmer mod et lavereliggende område (Sørensen et al., 2006, Ågren et al., 2014). I begge tilfælde regnes der udelukkende med overfladeafstrømning, og der tages derfor ikke højde for nedsivning. Dybden af afløbsløse lavninger er en anden topografisk variabel, der relaterer til vandets forekomst og afstrømning i landskabet og dermed til behovet for dræning.

Et areals placering i forhold til vandløb, søer og havet har ligeledes potentielt stor betydning for dræningen. Placeringen kan både afgøre, om et areal er vandlidende, og om det er muligt at dræne (Nielsen, 2015). Både den horisontale og vertikale afstand til overfladevand blev derfor inddraget som variabler med relevans i forhold til dræning. Hældningen ned til overfladevand blev også udregnet som et udtryk for faldforholdene mod overfladevand.

Dybden til grundvandet har stor betydning for behovet for at etablere drænanlæg (Madsen et al., 1992, Nielsen, 2015). For at få større sikkerhed i forhold til denne variabel blev dybden til grundvandet inkluderet i form af to forskellige datalag. For det første blev der udregnet et grundvandsspejl ved interpolation af observationer af overfladevand og vandstanden i geologiske borer (Hansen et al., 2004). Herudover blev grundvandsstanden fra DK10-modellen anvendt (Henriksen et al., 2012). Grundvandsstanden blev interpoleret fra modellens opløsning på 500 meter til højdemodellens opløsning på 30,4 meter. Den udregnede grundvandsstand blev i begge tilfælde trukket fra højdemodellen således, at dybden fra jordoverfladen til grundvandsspejlet kunne beregnes (Figur 8).



Figur 8: Illustration af udregningen af dybden til grundvandet. Dybden til grundvandet er udregnet som højden af terrænet fratrukket højden af grundvandsspejlet. Grundvandsspejlet er enten udregnet ved interpolation af punktobservationer eller ved at interpolere grundvandsspejlet fra DK10-modellen til en finere opløsning.

Der blev desuden inddraget en række mere deskriptive topografiske variabler med betydning for udbredelsen af dræning, herunder overfladens hældning, kurvatur og kompasretning. Kompasretningen blev nedbrudt i komponenterne øst-vest og nord-syd, fordi statistiske modeller generelt ikke kan håndtere cirkulære variabler. Solindstrålingen blev også udregnet fra højdemodellen, da arealer med stor solindstråling generelt er varmere og tørrere end andre arealer.

Satellitbillederne bestod af en mosaik af Landsat 8-scener fra marts 2014, der var den eneste måned med skyfri billeder fra hele landet (NASA Landsat Program, 2014). Billederne blev transformeret til den samme opløsning som de topografiske variabler. Billederne gengav overfladereflektiviteten fra syv spektrale bånd og normaliserede indekser for fugt (ndmi), overfladevand (ndwi), vegetation (ndvi) og jordtilpasset vegetation (savi). Bilag 1 viser udregningen for hvert af disse indekser.

Jordens tekstur er en af de væsentligste årsager til, at der drænes i Danmark (Madsen et al., 1992, Nielsen, 2015). For at kunne tage højde for de teksturbetingede drænforhold, blev der indhentet jordbundsdata fra flere kilder. Jordbundskortene omfattede kort over jordens lerindhold (Adhikari et al., 2013), kort over jordens naturlige drænklasser (Møller et al., 2017), GEUS's jordartskort (Jakobsen et al., 2015), kort over landskabselementer (Madsen et al., 1992) og vådområder (Greve et al., 2014).

Adhikari et al. (2013) udarbejdede geografiske datalag over jordens indhold af ler, silt, finsand og grovsand. Jordens drænforhold er ikke kun bestemt af lerindholdet. For eksempel kan indholdet af grovsand også have betydning (Nielsen, 2015). Denne undersøgelse inddrog kun lerindholdet, da dette vurderes at have størst betydning for jordens hydrauliske ledningsevne (Wösten og van Genuchten, 1988) og derfor også for drænforholdene og behovet for dræning. De oprindelige kort angav lerindholdet i seks dybdeintervaller. Til denne undersøgelse blev antallet af dybdeintervaller reduceret fra seks til fire ved at udregne vægtede gennemsnit af de oprindelige seks intervaller. Dette skyldes, at de øverste tre intervaller (0 – 5 cm, 5 – 15 cm, 15 – 30 cm) i Danmark oftest indgår i pløjelaget. Jorden i pløjelaget vendes jævnlige, og jorden i dybden 0 – 30 cm vil derfor i de fleste tilfælde have det samme lerindhold.

Jordens drænforhold har i udstrakt grad betydning for både for arealanvendelsen og for valget af afgrøder. Både arealanvendelsen og dyrkningshistorikken blev derfor inddraget. Arealanvendelsesdata blev inddraget fra CORINE 2012 (European Environment Agency, 2014). Arealanvendelsen blev inddelt i tre kategorier: (1) Naturlig vegetation og græsningsarealer, (2) dyrkede arealer og (3) arealer med vådbundsvegetation. Skove, byer og søer blev skåret fra og indgår ikke i analysen.

Dyrkningshistorikken blev inddraget fra IMK (Landbrugsstyrelsen, 2014). Afgrøderne blev inddelt i tre kategorier baseret på deres afvandingsbehov: (1) drænkkrævende afgrøder, (2) måske drænkkrævende afgrøder og (3) ikke drænkkrævende afgrøder. Klassificering af afgrøder er vist i Bilag 2. Drænkkrævende afgrøder omfattede især vintersæd, mens de måske drænkkrævende afgrøder først og fremmest udgjordes af vårbyg. Der skelnedes mellem drænkkrævende og måske drænkkrævende afgrøder da nedbørsoverskuddet i Danmark ligger om vinteren. Dette betyder, at det øvre grundvand ligger højere om vinteren, og en jord, der

er vandlidende i vinterhalvåret kan derfor have bedre drænforhold i sommerhalvåret. Vinterafgrøder sås om efteråret og overvintrer derfor på marken. Disse afgrøder er derfor mere sårbare overfor dårlige drænforhold om vinteren end vårafgrøder, der sås om foråret (Kronvang et al., 2013). De fleste ikke drænkrevende afgrøder er græsarealer af forskellig art (Bilag 2). Antallet af år med hver kategori på et givent areal blev opgjort for perioden 2011 til 2014. Det blev også opgjort, hvor mange år hvert areal var med i IMK.

Den gennemsnitlige årlige nedbørsmængde i perioden fra 1961 til 1990 blev også inddraget som en relevant variabel. Nedbørsmængden blev interpoleret fra punktobservationer. Nedbørsmængden har ændret sig gennem tiden, men det antages, at de rumlige nedbørsmønstre generelt har været de samme i løbet af perioden, hvor man har drænet. Da undersøgelsen baseres på statistiske analyser, er det den rumlige fordeling af nedbøren snarere end de absolutte værdier, der er afgørende.

2.1.5 Endelige datasæt

De forklarende variabler blev knyttet til punktobservationerne af drænedede og ikke drænedede arealer for at danne datasættet til analysen. Det samlede datasæt til kalibrering indeholdt 1009 punktobservationer, hvoraf der var drænet ved 572 punkter og ikke drænet ved 437 punkter.

Der var således tale om et ubalanceret datasæt, hvilket kan være et problem, fordi mange statistiske metoder forudsiger en større udbredelse af den klasse, der oftest forekommer i datasættet, end man observerer i virkeligheden. Et ubalanceret datasæt med den her opnåede fordeling mellem drænedede og ikke drænedede punkter vil derfor kunne betyde, at flere arealer fremstår som drænedede på kortet end i virkeligheden. Tidligere undersøgelser har vist, at omkring halvdelen af landbrugsarealet er drænet (Skriver og Hedegård, 1973, Olesen, 2009). I det nye datasæt er 57% af arealerne imidlertid drænet, og denne ubalance ville uden korrektion kunne blive afspejlet i det udarbejdede kort.

For at balancere datasættet blev der genereret 135 syntetiske punktobservationer af ikke drænedede arealer med SMOTE-algoritmen [Synthetic Minority Over-sampling Technique] (Chawla et al., 2002). De genererede punkter har ingen rumlig placering, men bygger derimod på statistiske sammenhænge i datasættet. Dette er en rent beregningsmæssig fremgangsmåde, der ikke ændrer på gennemsnitsværdier for drænedede og ikke drænedede arealer. Det ændrer derfor ikke ved datasættets karakter.

Det endelige datasæt til kalibrering indeholder derfor 1144 punkter, med lige mange drænedede og ikke drænedede observationer (Tabel 1). Datasættet til beregning af kortets nøjagtighed indeholdt 133 punkter med dræning og 114 punkter uden dræning.

Tabel 1: Datasæt til at udarbejde og vurdere kort over drænedede arealer med angivelse af datakilderne. For hvert datasæt er det angivet, hvor mange drænedede og ikke drænedede observationer det indeholder, og hvor mange observationer det indeholder i alt.

Datasæt/kilde	Drænet	Ikke drænet	I alt
Opbygning af modeller			
Olesen, 2009 (2/3)	268	230	498
Landskontoret for Planteavl, 1973	207	207	414
Orbicons drænarkiv	97	0	97
Syntetiske punkter (SMOTE)	0	135	135
I alt	572	572	1144
Beregning af nøjagtighed			
Olesen, 2009 (1/3)	133	114	247

Det endelige datasæt omfatter både punktobservationer og arealobservationer konverteret til punkter. I alle tilfælde er det observeret om punktet eller arealet var drænet eller ej, men ikke hvor drænene var placeret. Det er derfor ikke sikkert, at der er drænrør ved de punkter, der indgår i datasættet. Det endelige kort vil afspejle dette forhold og vil derfor vise drænedede arealer i bredere forstand end "arealer med drænrør". I områder hvor der er stor afstand mellem drænrørene, vil det drænedede areal være større end arealet hvor der er drænrør. Det kan for eksempel ske, hvis man kun har lagt drænrør i de lavninger, hvor vandet samles.

2.2 Kortlægning

I denne undersøgelse er udbredelsen af drænedede og ikke drænedede arealer kortlagt med maskinlæring. Maskinlæring er et felt indenfor datalogi, der sigter imod at gøre computere i stand til at afdække og anvende sammenhænge i data uden eksplicit programmering (Samuel, 1959). Feltet er tæt beslægtet med både statistik og forskning i kunstig intelligens og anvender værktøjer fra begge felter. Modeller baseret på maskinlæring kan ofte levere meget nøjagtige forudsigelser under forudsætning af gode læringsdata og har blandt andet været anvendt til tekstgenkendelse (Sebastiani, 2002), kræftdiagnosticering (Wang et al., 2005) og til at opdage mistænkelige finanstransaktioner (Bose og Mahapatra, 2001).

Modeller baseret på maskinlæring er datadrevne, hvilket vil sige, at en algoritme opbygger modellen ud fra de data, der stilles til rådighed. Modellerne indeholder derfor kun sammenhænge, der gennem analyse kan identificeres i de anvendte data. Modeller baseret på maskinlæring omfatter både enkle statistiske metoder, så som lineære modeller og diskriminantanalyse, og mere komplekse tilgange som beslutningstræer og kunstige neurale netværk (Bilag 3).

Inden for jordbundsvidenskaben har maskinlæringsalgoritmer især været anvendt til kortlægning, da de kan anvendes til at forudsige jordbundsforhold for områder, hvor der ikke er observationer eller målinger, men hvor der foreligger kendte variabler, der korrelerer med jordbundsforholdene. De er blandt andet blevet

anvendt til at kortlægge jordens tekstur (Adhikari et al., 2013), humusindhold (Poggio og Gimona, 2014) og de naturlige drænforhold (Zhao et al., 2013, Beucher et al., 2017, Møller et al., 2017). De således kortlagte områder spænder fra enkelte marker (Liu et al., 2008) til global skala (Hengl et al., 2017).

Metoden til udarbejdningen af kortet over drænede arealer bestod dels i at opbygge en række modeller, dels i at kombinere de modeller der tilsammen gav den mest nøjagtige forudsigelse.

Modellerne til forudsigelse af drænede arealer blev i denne undersøgelse opbygget gennem 77 maskinlæringsmetoder. For hver metode blev der opbygget fire modeller med forskellige datasæt, og der blev derfor opbygget i alt 308 modeller.

De fire datasæt udgjordes af ét datasæt med de oprindelige datalag og tre datasæt med Principal Components (PCs). I Principal Component-analyse udregnes der et sæt nye variabler ud fra et datasæt. De nye variabler korrelerer ikke med hinanden, men indeholder så stor en del som muligt af variationen i det oprindelige datasæt. Jo flere Principal Components, der udregnes, desto mere af variationen i det oprindelige datasæt kan de repræsentere. De nye datasæt indeholdt henholdsvis 10, 20 og 44 PCs og forklarede henholdsvis 52%, 71% og 95% af variationen i det oprindelige datasæt. Principal Component-analyse er nærmere forklaret i Bilag 4.

De endelige modeller til forudsigelse af drænede arealer blev herefter udvalgt med en teknik, der kombinerer de modeller, der tilsammen giver den mest nøjagtige forudsigelse. Teknikken består i anvendelse af en trinvis algoritme, der udvælger modeller én ad gangen og til sidst kombinerer forudsigelserne fra de udvalgte modeller. I dette tilfælde blev forudsigelserne kombineret som et gennemsnit af de forudsagte sandsynligheder for dræning fra hver af modellerne. Nøjagtigheden for de samlede forudsigelser blev herefter beregnet for det uafhængige datasæt, som ikke indgik i maskinlæringen.

Teknikken blev testet med en række forskellige indstillinger (Bilag 4). De indstillinger, der gav de mest nøjagtige resultater blev herefter anvendt til at udvælge de modeller, der blev anvendt til at udarbejde det landsdækkende kort over sandsynligheden for dræning. De endelige sandsynligheder for dræning blev udregnet som et gennemsnit af de kort, som de enkelte modeller genererede. Resultat blev beskåret, så det kun dækkede arealer, der var en del af markarealet i mindst ét af årene i perioden fra 2011 til 2014.

For de modeller, der kunne beregne betydningen af de anvendte datalag, blev den relative betydning skaleret til 100 for det vigtigste datalag, og et gennemsnit blev beregnet på tværs af de udvalgte modeller.

Fremgangsmåden for opbygningen og udvælgelsen af modeller er nærmere beskrevet i Bilag 4.

2.3 Kombination af kort

Da indsamlingen af nye observationer var koncentreret i bestemte dele af landet, er det muligt, at kortet fra undersøgelsens anden fase er mere nøjagtigt i nogle områder, mens kortet udarbejdet i undersøgelsens

første fase er mest nøjagtigt i andre områder. For at tage højde for denne mulighed blev der foretaget en analyse af de to korts nøjagtighed opdelt på geologiske klasser, landskabselementer og georegioner.

De to korts nøjagtighed blev beregnet for hver klasse baseret på den tredjedel af observationerne fra Olesen (2009), der ikke indgik i opbygningen og udvælgelsen af modeller. I hvert tilfælde blev nøjagtigheden sammenlignet gennem en binomial test. Hvis kortet fra undersøgelsens anden fase havde en statistisk signifikant bedre nøjagtighed end kortet udarbejdet i undersøgelsens første fase, blev kortet fra undersøgelsens anden fase anvendt for det pågældende område. I alle andre tilfælde blev kortet fra undersøgelsens første fase anvendt.

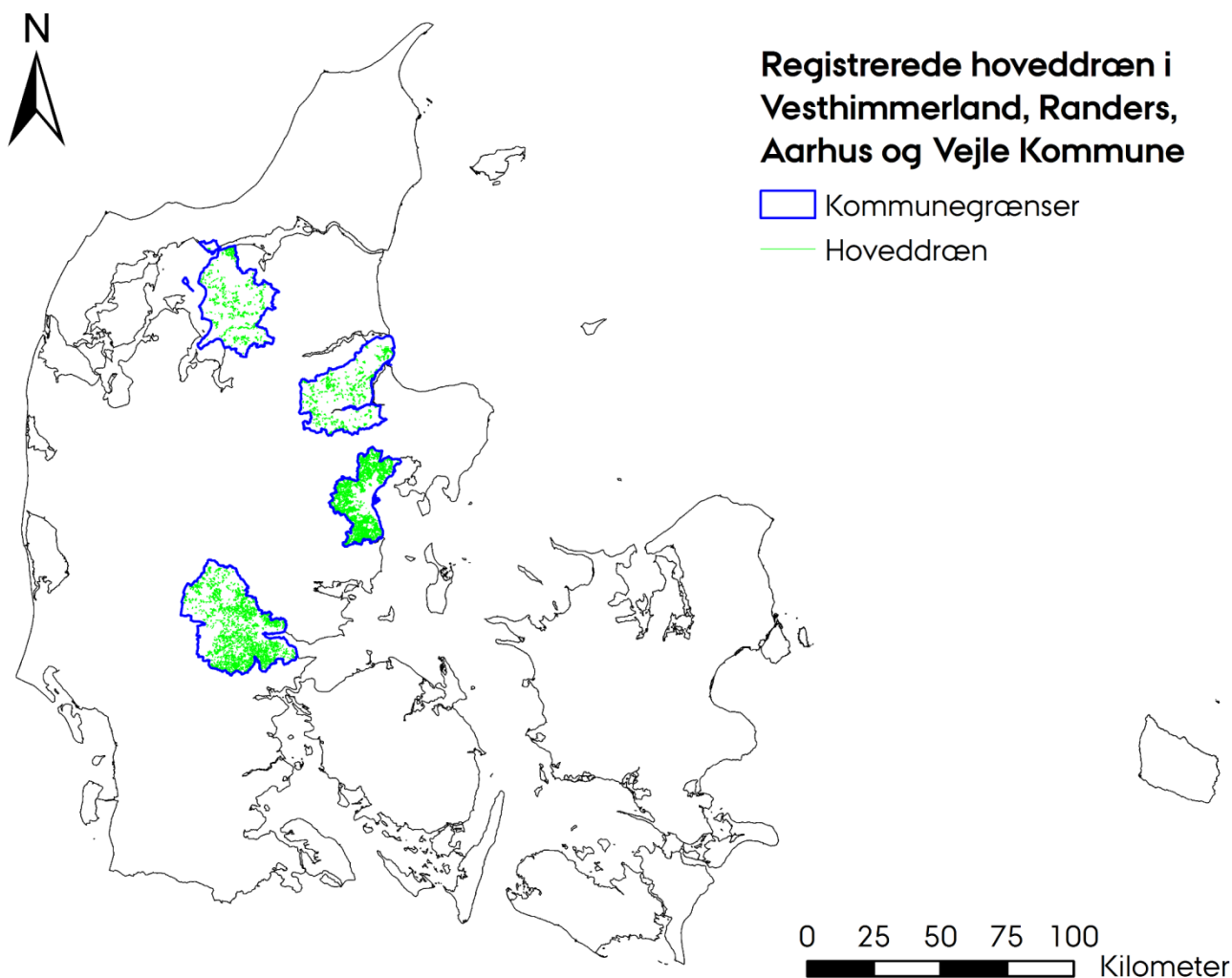
2.4 Kortets nøjagtighed

For at kunne vurdere kortets evne til at forudsige drænedede arealer i forskellige sammenhænge blev der anvendt data fra en række uafhængige kilder. De uafhængige data omfattede hoveddræn fra Orbicons drænarkiv, lokaliteter med målinger af drænastrømning, en række oplande, hvor drænprocenten er kortlagt på markniveau samt drænrør registreret af Energinet i forbindelse med nedgravning af elkabler.

For at sammenligne nøjagtigheden af det endelige kort med kortet udarbejdet af Olesen (2009), blev sidstnævnte kort omdannet til en rasterfil (Bilag 3) med samme opløsning som kortet udarbejdet i denne undersøgelse. Det blev herefter opdelt i drænedede og ikke drænedede arealer, ved at arealer med drænprocenter $> 50\%$ blev betragtet som drænedede, og arealer med mindre drænprocenter blev betragtet som ikke drænedede. Kortet blev yderligere beskåret, så det udelukkende dækkede det nuværende markareal. Nøjagtigheden af dette kort blev derefter beregnet ud fra de uafhængige datasæt på samme måde som kortet udarbejdet i denne undersøgelse.

2.4.1 Hoveddræn

Det første datasæt omfatter hoveddræn kortlagt af Orbicon. Orbicon har registreret og digitaliseret hoveddræn i Vesthimmerland, Randers, Aarhus og Vejle Kommune i forbindelse med opgaver for disse kommuner (Figur 9). Hoveddrænene er registreret både som linjer og polygoner. For at kunne vurdere kortets nøjagtighed med disse data blev både linjer og polygoner konverteret til rasterdatasæt i samme opløsning som kortet og føjet sammen, så de dannede ét lag for hver kommune. For hver kommune blev det beregnet, hvor stor en del af cellerne med hoveddræn, der var forudsagt som drænedede, og det blev beregnet, hvor stor en del af kommunens markareal, der var forudsagt som drænet.



Figur 9: Hoveddræn registreret af Orbicon i Vesthimmerland, Randers, Aarhus og Vejle Kommune.

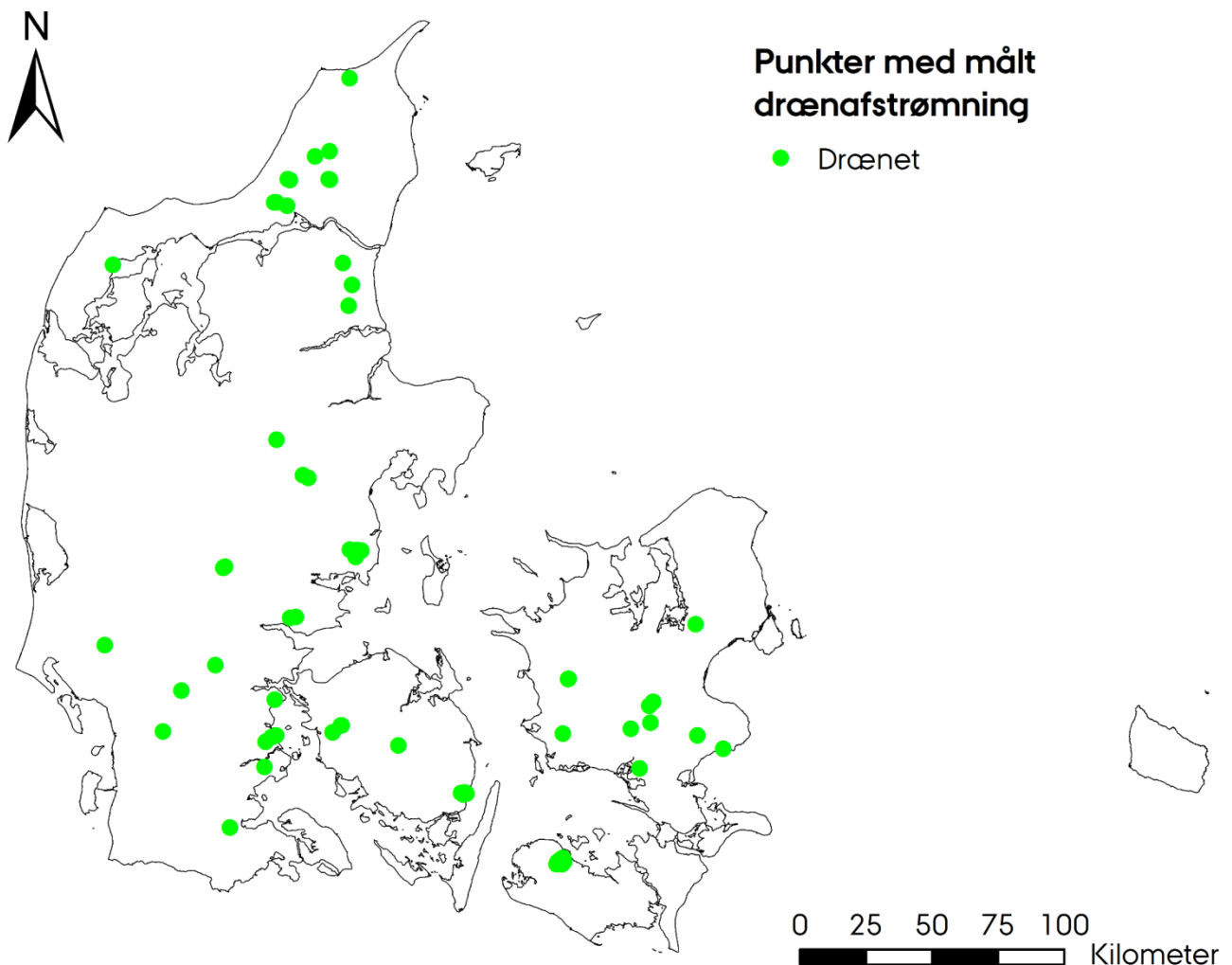
Datasættet har en række begrænsninger. For det første omfatter datasættet ikke sidedræn. For det andet omfatter det udelukkende drænprojekter udført af Hedeselskabet. Der kan derfor være drænprojekter udført af andre entreprenører, som ikke er med i datasættet.

Der er derfor ikke tale om et komplet datasæt over drænedede arealer i de fire kommuner, og det er derfor ikke muligt at vurdere, om kortet forudsiger dræning på arealer, hvor der ikke er drænrør (overprædiktion) ud fra datasættet. På den anden side kan de kortlagte hoveddræn afsløre, om der ligger hoveddræn på arealer, som kortet forudsiger som ikke drænedede (underprædiktion).

En usikkerhed i denne forbindelse er, at hoveddræn i nogle tilfælde føres igennem arealer, som ikke i sig selv er drænkrevende for at føre drænvandet fra et drænet areal ned til en recipient. Det kan eksempelvis ske, hvis et hoveddræn føres igennem en bakke for at afvande en afløbsløs lavning. Det er også muligt, at rørlagte vandløb er blevet registreret som hoveddræn, selv hvis der ikke er sidedræn. Omvendt kan der også være drænet uden et hoveddræn, for eksempel hvis drænrørene er ført direkte ud til en grøft.

2.4.2 Drænudløb

Det andet datasæt omfatter placeringen af drænudløb. I forbindelse med projekter, hvor man har målt afstrømningen fra drænrør, er i alt 66 drænudløb blevet registreret som punkter fordelt over det meste af landet (Figur 10). Der er langt fra tale om et komplet datasæt, men det kan anvendes til at vurdere graden af underprædiktion. Afstrømningen fra hvert udløb stammer fra et drænet areal, men arealets størrelse og afgrænsning er i flere tilfælde ukendt, og derfor anvendes kun punkterne. Det blev beregnet, hvor mange af udløbene, der lå inden for arealer, der blev forudsagt som drænete, og hvor langt hvert udløb lå fra det nærmeste areal, der blev forudsagt som drænet.

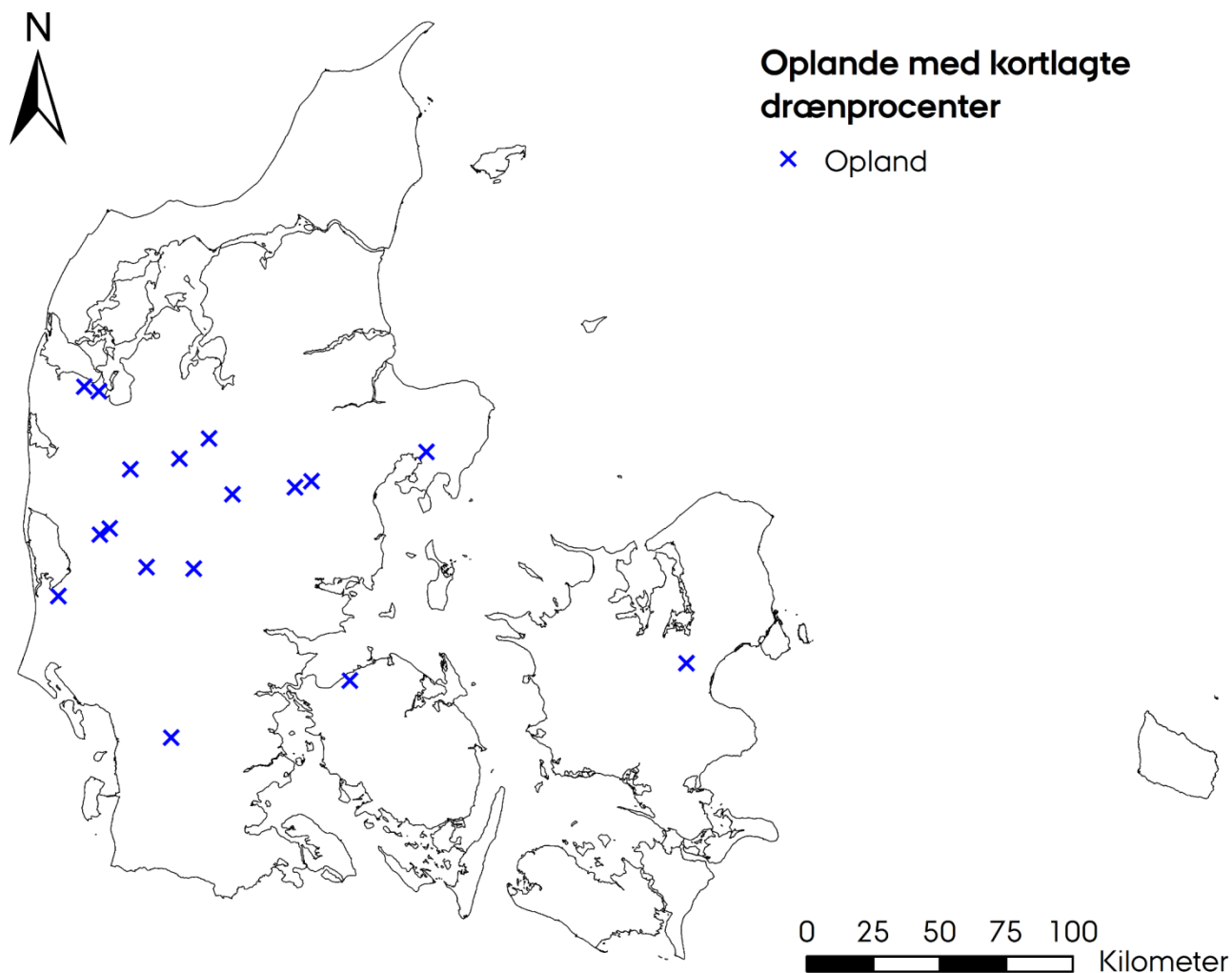


Figur 10: Punkter med målt drænafstrømning.

2.4.3 Drænprocenter på markniveau

Det tredje datasæt omfatter drænprocenter kortlagt på markniveau. I forbindelse med NPo-undersøgelsen (kvælstof (N), fosfor (P) og organisk materiale (o)), blev drænprocenten kortlagt på markniveau for en række mindre oplande i løbet af 1980'erne (Kelstrup og Hansen, 1986, Hansen, 1990). I denne undersøgelse blev der anvendt oplysninger fra 17 oplande (Figur 11). Oplysningerne omfattede i alt 1439 marker med et

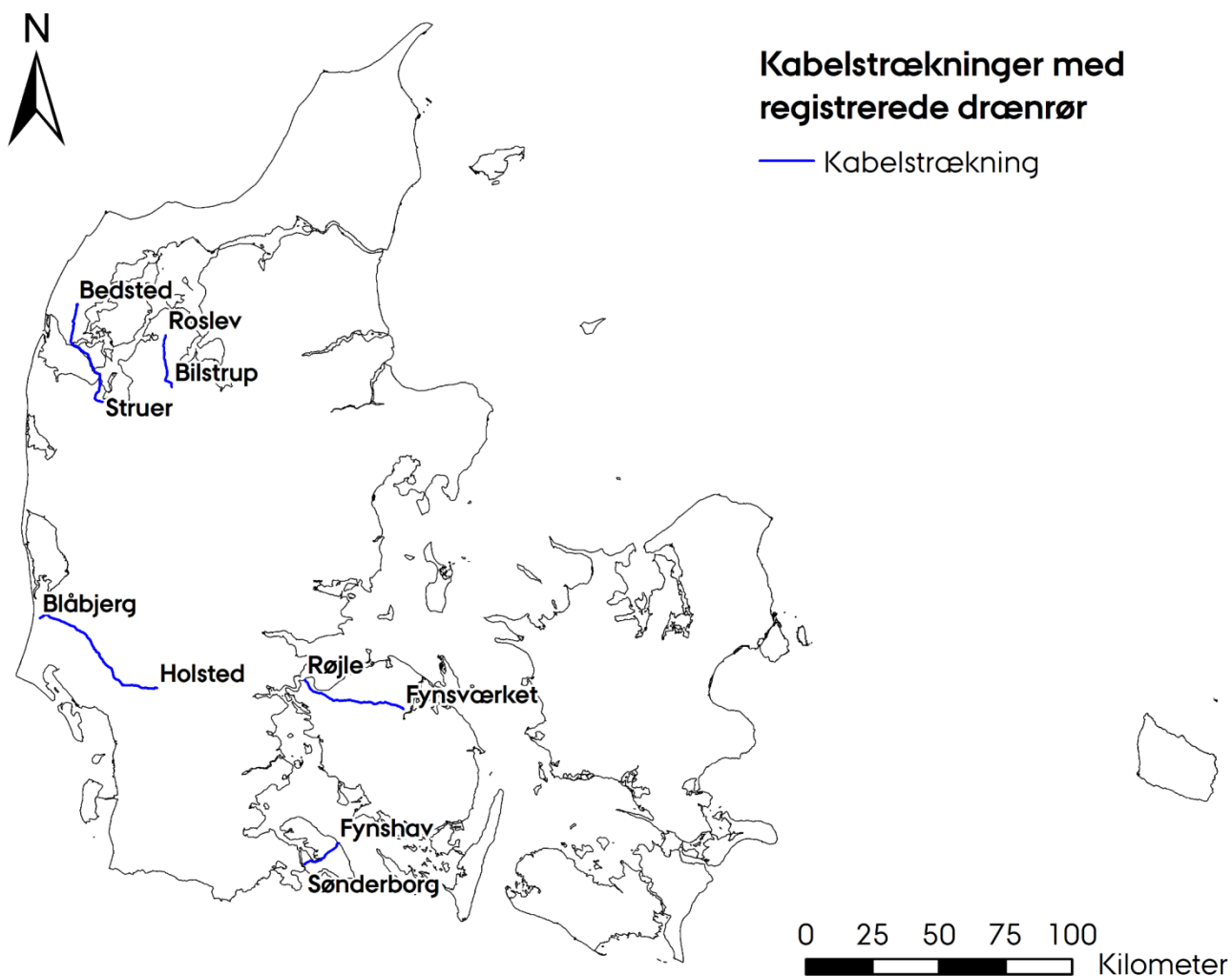
samlet areal på 68 km². Marker, som ikke var tildelt en drænprocent, eller som lå uden for det nuværende markareal, blev udeladt, hvilket reducerede datasættet til 1285 marker med et samlet areal på 59 km². For hver mark blev det beregnet, hvor stor en del af cellerne inden for marken, der var forudsagt som dræned. Resultatet blev sammenlignet med de angivne drænprocenter med en vægtet lineær regression, hvor antallet af celler inden for markerne blev brugt som vægte.



Figur 11: Placering af 17 oplande med drænprocenter kortlagt på markniveau.

2.4.4 Drænrør i kabelgrave

Det fjerde datasæt omfatter drænrør registreret i forbindelse med nedgravning af elkabler. Energinet har i forbindelse med nedgravning af elkabler på et antal strækninger registreret de drænrør, der krydser kabelgraven. Der blev anvendt data fra fem strækninger i Jylland og på Fyn (Figur 12). Kabelgravene blev konverteret fra en linjer til et rasterdatasæt med samme opløsning som kortet over drænedede arealer. For hver celle i det nye lag, blev det angivet, om der var registreret drænrør inden for cellen.



Figur 12: Kabelstrækninger med registrerede drænrør (data anvendt med tilladelse fra Energinet).

Kabelgravene varierede i længde fra 17 km for Sønderborg-Fynshav-strækningen til 59 km for Blåbjerg-Holsted-strækningen, og antallet af drænrør varierede fra 10 per km for Bedsted-Struer-strækningen til 25 per km for Røjle-Fynsværket-strækningen. Andelen af celler med drænrør havde generelt samme mønster som antallet af drænrør per km ($R^2 = 0,94$; $n = 5$; $p < 0,05$).

Tabel 2: Længde af hver kabelstrækning fra Energinet og det samlede antal drænrør, antal drænrør per kilometer af strækningen og andelen af 30,4 x 30,4 meter celler på markarealer, der har registrerede drænrør (data anvendt med tilladelse fra Energinet).

Strækning	Længde (km)	Antal drænrør	Drænrør/km	Celler med drænrør (%)
Bedsted-Struer	47	477	10	22
Blåbjerg-Holsted	59	688	12	24
Roslev-Bilstrup	22	495	23	39
Røjle-Fynsværket	43	1089	25	39
Sønderborg-Fynshav	17	395	23	34

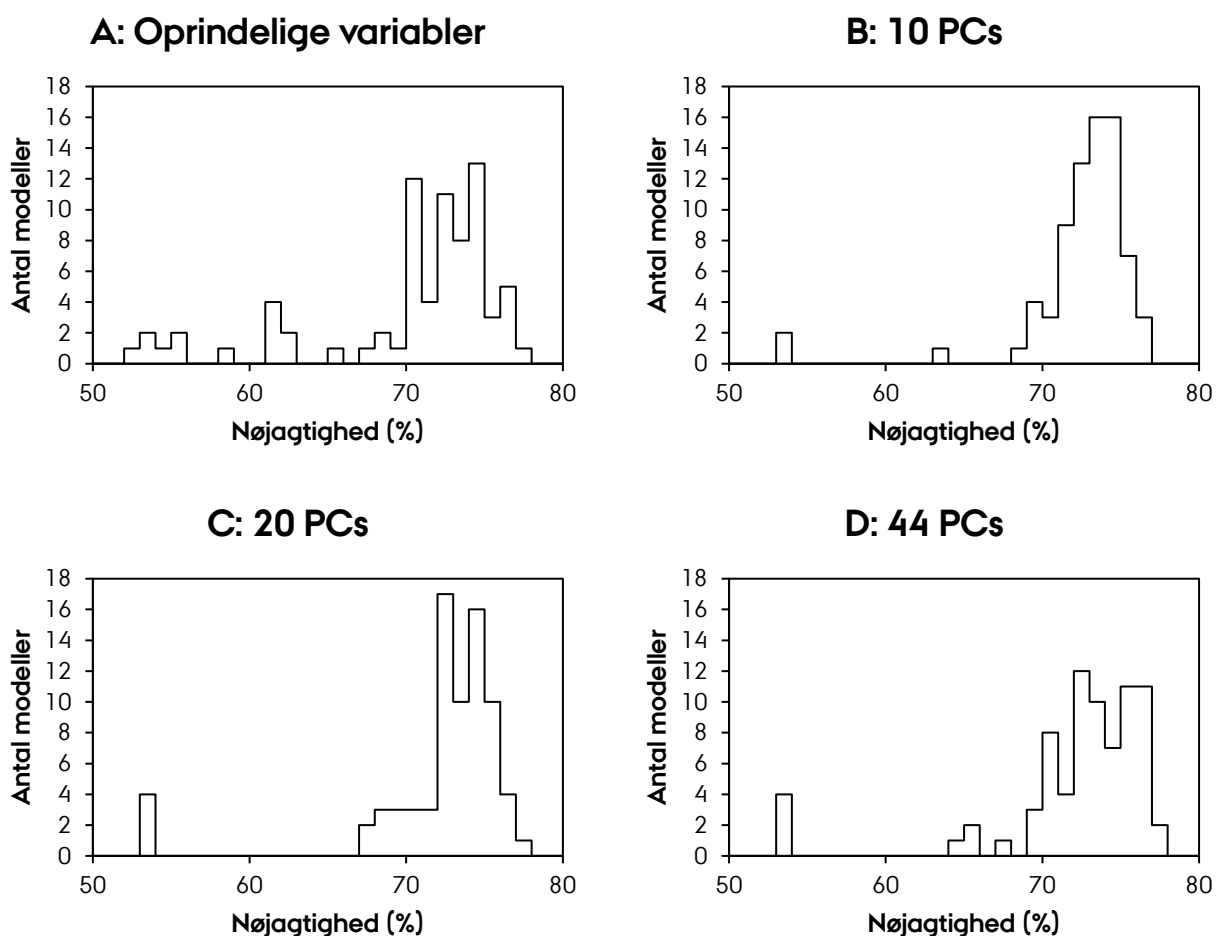
For hver strækning blev overensstemmelsen opgjort mellem celler med registrerede drænrør og celler, som kortet forudsagde som drænede. Det blev også beregnet, hvor stor en del af markarealet på hver strækning, der var forudsagt som drænet. De samme beregninger blev også foretaget på basis af landskabselementer og geologiske klasser for at undersøge, om disse forhold kunne have betydning for overensstemmelsen.

3 Resultater

3.1 Nøjagtighed af de enkelte modeller

De enkelte modellers nøjagtighed i undersøgelsens første fase er vist i Bilag 9.

I undersøgelsens anden fase varierede de enkelte modellers nøjagtighed fra 23% til 77%, i gennemsnit 71%, beregnet for det uafhængige datasæt. De fleste modeller havde en høj nøjagtighed (>65%) (Figur 13), mens relativt få modeller havde en lav nøjagtighed. Modellerne opbygget med de oprindelige variable havde en lavere nøjagtighed end modellerne opbygget med Principal Components (PCs; afsnit 2.2) (Figur 13). De mest nøjagtige modeller var fortrinsvis opbygget med 44 PCs.



Figur 13: Fordeling af modellernes nøjagtighed beregnet for det uafhængige datasæt. Figuren viser fordelingen af nøjagtigheden for modeller opbygget med A: De oprindelige variable, B: 10 PCs, C: 20 PCs og D: 44 PCs. Nøjagtigheden er i alle tilfælde inddelt i intervaller på 1%, så hvert interval angiver, hvor mange modeller, der har en nøjagtighed inden for hvert interval.

3.2 Udvalgte modeller

Algoritmen til udvælgelse af modeller blev anvendt med de indstillinger, der gav den mest nøjagtige forudsigtelse af drænedede arealer (Bilag 4). I undersøgelsens første fase udvalgte algoritmen i alt 36 modeller, fortrinsvist opbygget med 10 PCs (Bilag 9).

I undersøgelsens anden fase udvalgte algoritmen i alt 14 modeller (Tabel 3). Seks af de udvalgte modeller var opbygget med beslutningstræ-algoritmer, mens de restende modeller var opbygget ved enten diskriminantanalyse (2), k-nærmeste naboer (2), kunstige neurale netværk (2), lineær regression (1) eller support vector machines (1). Modeltyperne er beskrevet i Bilag 3.

Otte af modellerne var opbygget med det oprindelige datasæt, mens tre var opbygget med 44 PCs. To var opbygget med 10 PCs, mens en enkelt model var opbygget med 20 PCs. Kun fem af modellerne kunne beregne betydningen af de datalag, der blev anvendt. Ingen af modellerne blev udvalgt mere end én gang.

Tabel 3: Metode og inputdata for modellerne udvalgt af algoritmen ved hvert trin, samt en angivelse af, om metoden indebærer en beregning af betydningen af de enkelte datalag.

Trin	Metode	Forklaring	Inputdata	Betydning af datalag
1	cforest	Conditional inference forest; ensemble af beslutningstræer	Oprindeligt datasæt	Ja
2	Rf	Random forest; ensemble af beslutningstræer	Oprindeligt datasæt	Ja
3	lda2	Lineær diskriminantanalyse	Oprindeligt datasæt	Nej
4	nodeHarvest	Ensemble af beslutningstræer	Oprindeligt datasæt	Nej
5	ranger	Random forest; ensemble af beslutningstræer	Oprindeligt datasæt	Ja
6	stepLDA	Trinvis lineær diskriminantanalyse	10 PCs	Nej
7	Knn	k-nærmeste naboer	44 PCs	Nej
8	xgbLinear	Optimeret lineær model	Oprindeligt datasæt	Ja
9	svmRadialCost	Support vector machine	Oprindeligt datasæt	Nej
10	pcaNNet	Kunstigt neuralt netværk	20 PCs	Nej
11	Nnet	Kunstigt neuralt netværk	Oprindeligt datasæt	Ja
12	adaboost	Ensemble af beslutningstræer	44 PCs	Nej
13	nodeHarvest	Ensemble af beslutningstræer	10 PCs	Nej
14	Kknn	K-nærmeste naboer	44 PCs	Nej

De samlede forudsigtelser fra de udvalgte modeller havde en nøjagtighed på 78%, udregnet med det uafhængige datasæt. Ingen af de 308 modeller, der blev opbygget, havde hver for sig en lige så høj nøjagtighed som den samlede forudsigtelse fra de udvalgte modeller.

I begge undersøgelsens faser var lerindholdet det vigtigste datalag for forudsigtelsen af drænedede arealer (Tabel 4). Dybden til grundvandet og den naturlige drænklasse var også vigtige i begge tilfælde. Flere af de

vigtigste datalag var enten direkte eller indirekte relateret til lerindholdet eller dybden til grundvandet. Eksempelvis indgik udbredelsen af moræneler på GEUS' jordartskort som et vigtigt datalag relateret til lerindholdet, mens den vertikale afstand til det nærmeste overfladevand er kendt som en proxy for dybden til grundvandet. Andre datalag var relateret til vandets bevægelse på i terrænet. For eksempel havde den akkumulerede afstrømning og det topografiske vådhedsindeks forholdsvis stor betydning. Nedbørsmængden og flere af satellitbillederne optrådte i øvrigt blandt de vigtigste 30 datalag.

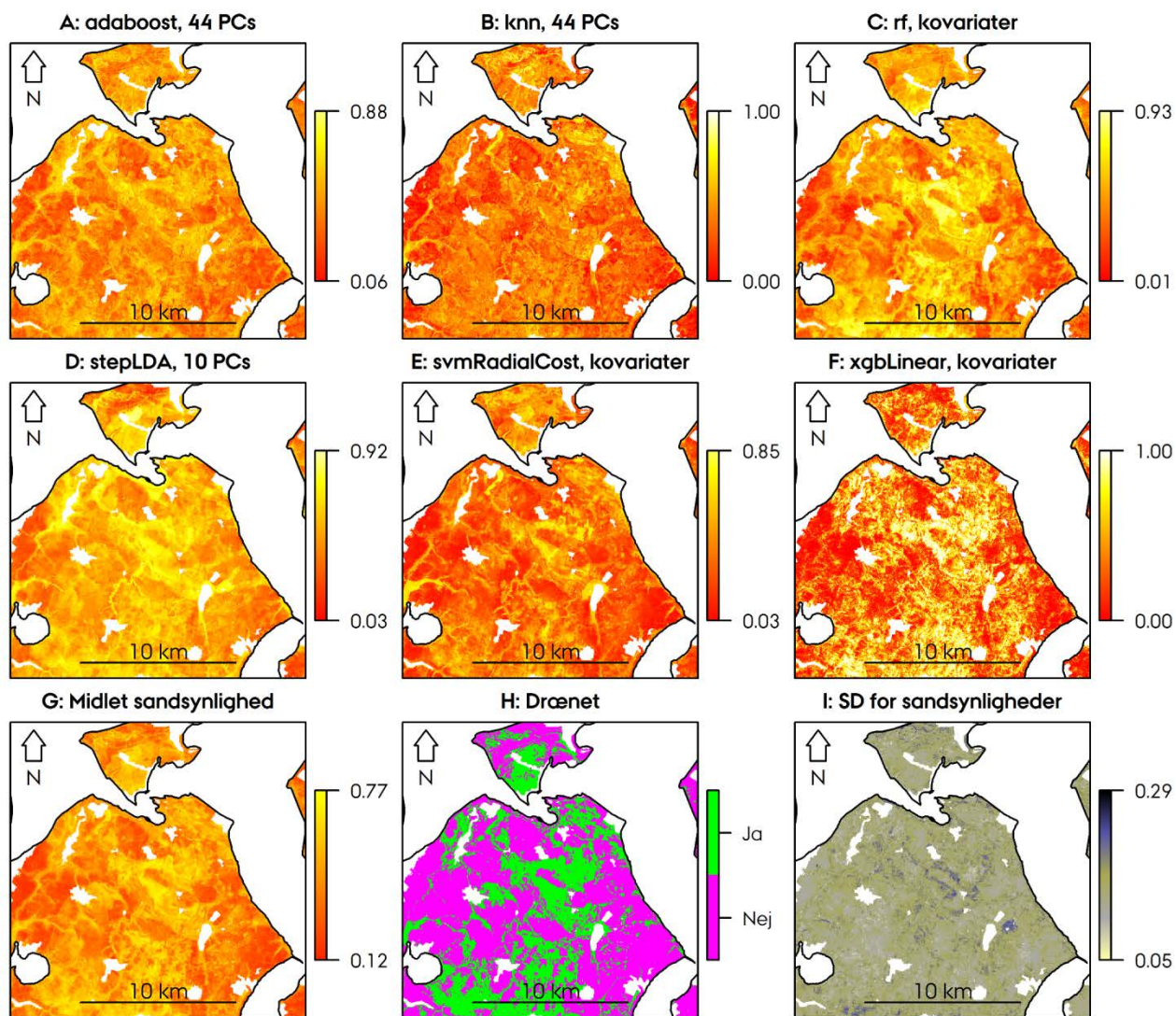
Sammenlignet med undersøgelsens første fase havde datalag relateret til topografi og satellitbilleder større betydning i den anden fase, mens georegioner, geologi og landskabselementer havde mindre betydning.

Tabel 4: De 30 vigtigste datalag for forudsigelsen af drænedede arealer i undersøgelsens to faser. Tabellen angiver desuden hvert datalags gennemsnitlige betydning i de anvendte modeller.

Første fase		Anden fase		
Rang	Datalag	Betydning	Betydning	
1	Ler (%); 60 - 100 cm	99	Ler (%); 0 - 30 cm	87
2	Ler (%); 30 - 60 cm	99	Dybde til grundvand (interpoleret)	52
3	Ler (%); 0 - 30 cm	98	Moræneler (GEUS)	48
4	Ler (%); 100 - 200 cm	95	Ler (%); 30 - 60 cm	44
5	Nedbør (1961 - 1990)	88	Vertikal afstand til overfladevand	43
6	Vestjylland (georegion)	81	Dybde til grundvand (DK10)	32
7	Naturlig drænklasse	74	Hældning til overfladevand	27
8	Moræneler (GEUS)	71	Højde over havet	27
9	Morænesand (GEUS)	70	Horisontal afstand til overfladevand	26
10	Dybde til grundvand (DK10)	69	Naturlig drænklasse	26
11	Smeltevandsslette (landskab)	69	Akkumuleret flow	26
12	Smeltevandssand (GEUS)	68	Nedbør (1961 - 1990)	24
13	Bakkeø (landskab)	67	Ler (%); 60 - 100 cm	23
14	Dybde til grundvand (interpoleret)	67	Ler (%); 100 - 200 cm	22
15	Vertikal afstand til overfladevand	65	Daldybde	22
16	Højde over havet	64	Nærinfrarød reflektivitet (Landsat 8)	20
17	Fugtindeks (Landsat 8)	62	Vådhedsindeks (SAGA GIS)	20
18	Højde fratrukket gennemsnit (4 km radius)	61	Højde fratrukket gennemsnit (4 km radius)	18
19	Vådområder	60	Himmerland (georegion)	18
20	Østdanmark (georegion)	57	Infrarød reflektivitet II (Landsat 8)	18
21	Syddanmark (georegion)	56	Nord-syd-komponent (kompasretning)	17
22	Littorinaflade (landskab)	53	År med drænkrævende afgrøder	16
23	Jordtilpasset vegetationsindeks (Landsat 8)	53	Øst-vest-komponent (kompasretning)	16
24	Horisontal afstand til overfladevand	53	Direkte solindstråling	16
25	Infrarød reflektivitet II (Landsat 8)	53	Grøn reflektivitet (Landsat 8)	15
26	Daldybde	51	Indeks for dalbundes fladhed	14
27	Hældning til overfladevand	51	Østdanmark (georegion)	14
28	Overfladens hældning	51	Midtjylland (georegion)	13
29	Topografisk vådhedsindeks	51	Infrarød reflektivitet I (Landsat 8)	13
30	Indeks for dalbundes fladhed	50	Rød reflektivitet (Landsat 8)	13

De udvalgte modeller gav meget forskellige forudsigelser, selv om en del mønstre gik igen (Figur 14A - F). Den gennemsnitlige sandsynlighed for dræning var generelt tættere på 50% end sandsynlighederne

forudsagt af de enkelte modeller (Figur 14G). Den endelige forudsigelse af drænedede og ikke drænedede arealer var meget detaljeret med specifikke forudsigelser selv for små områder (Figur 14H). På trods af den store variation mellem modellernes forudsigelser af sandsynligheden for dræning var standardafvigelsen (SD; Figur 14I) på tværs af modellernes forudsigelser relativt lille. I de fleste områder var standardafvigelsen cirka ti procentpoint, og store standardafvigelser var sjældne.

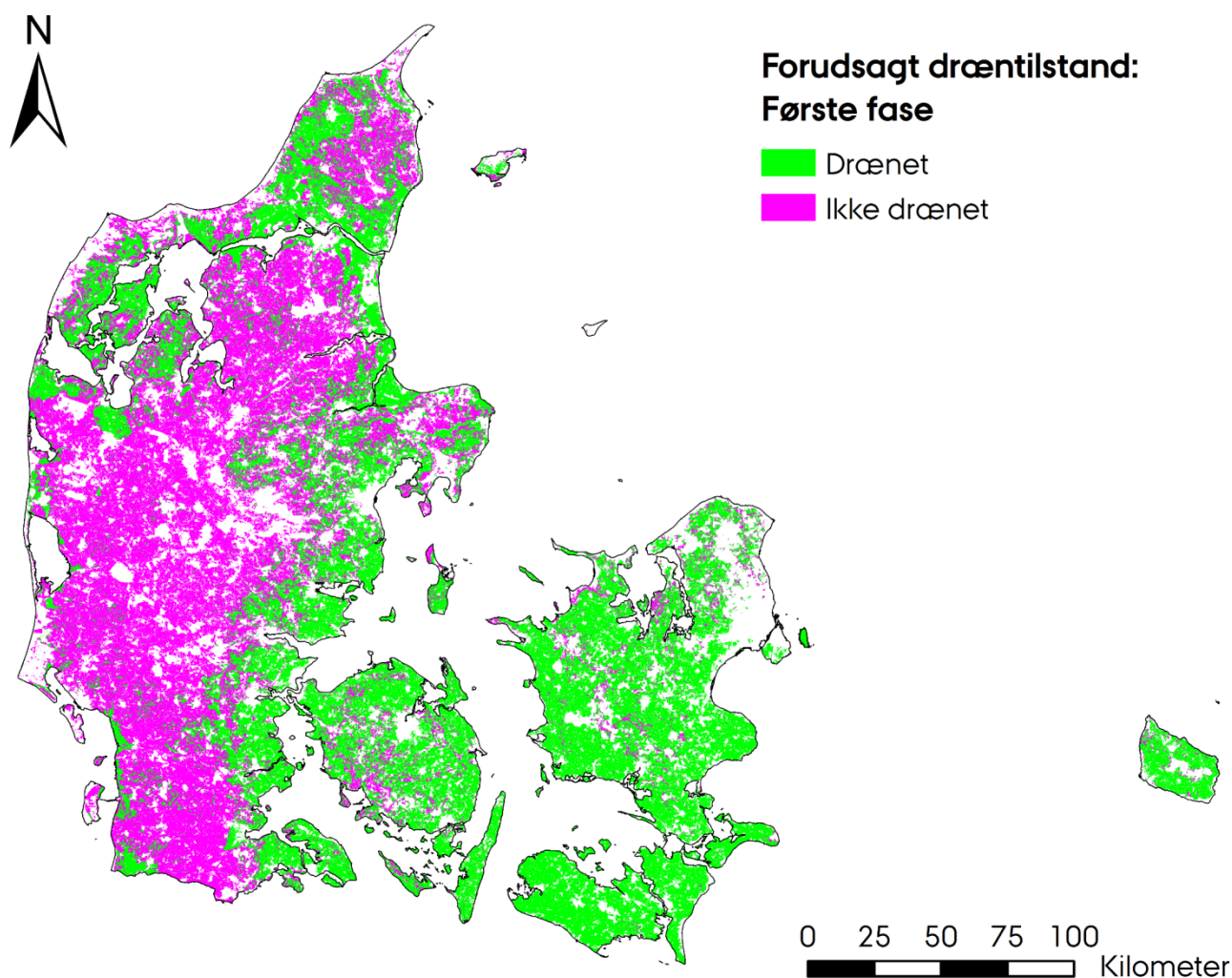


Figur 14: Eksempel på den forudsagte sandsynlighed for dræning i et område på 19x19 km på det nordlige Salling. A – F: sandsynligheden for dræning forudsagt af seks af de udvalgte modeller (1.00 = 100%). A: Ensemble af boostede beslutningstræer, opbygget med 44 PCs. B: k-nærmeste naboer, opbygget med 44 PCs. C: Ensemble af beslutningstræer, baseret på bagging og randomisering, opbygget med det oprindelige datasæt (kovariater). D: Trinvis lineær diskriminantanalyse, opbygget med 10 PCs. E: Support vector machine, opbygget med det oprindelige datasæt (kovariater). F: Boostet lineær model, opbygget med det oprindelige datasæt (kovariater). G: Gennemsnitlige sandsynlighed for dræning forudsagt af alle de udvalgte modeller.

H: Endelig forudsigtelse af udbredelsen af drænede arealer. I: Standardafvigelse (SD) for sandsynlighederne for dræning forudsagt af de udvalgte modeller.

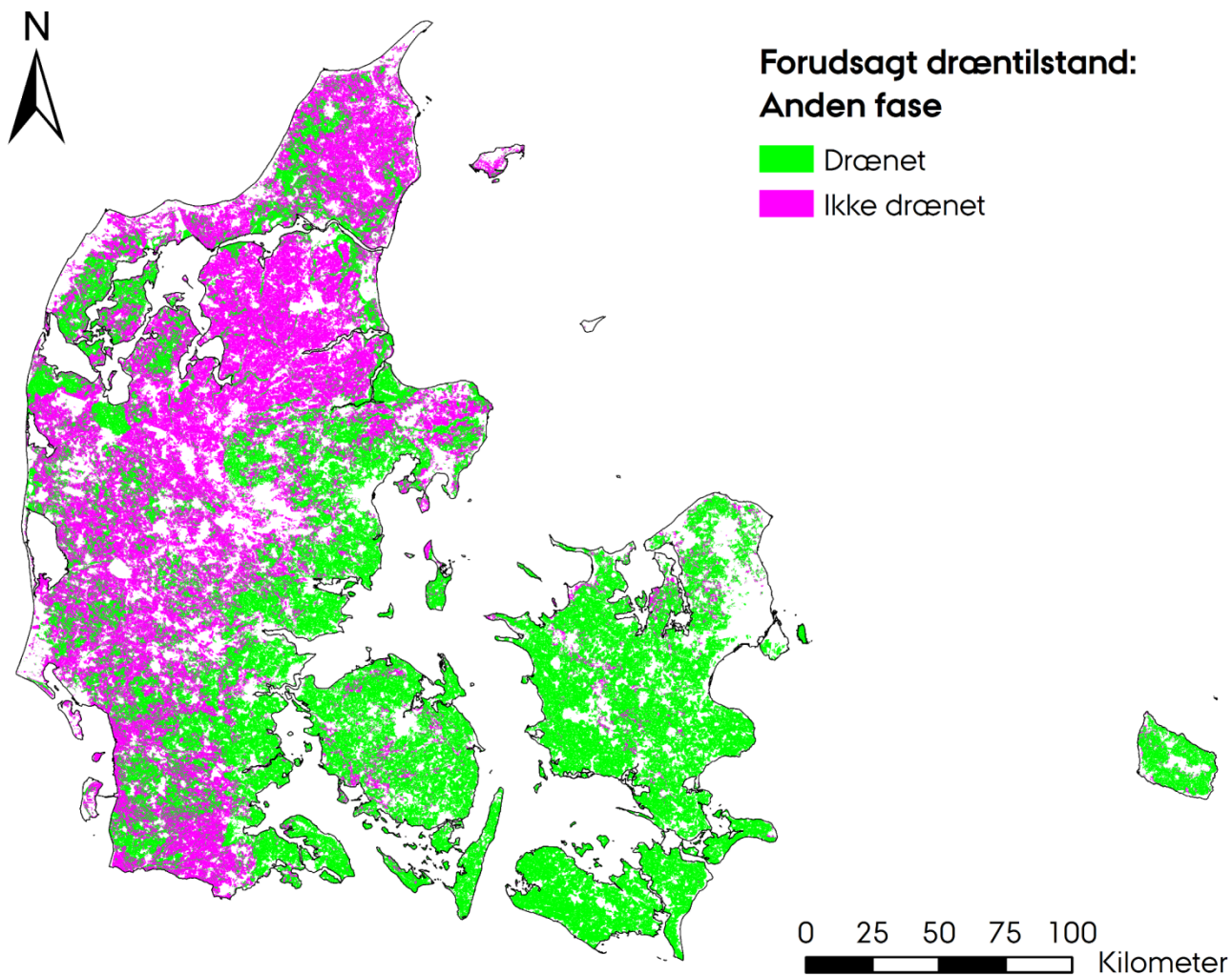
3.3 Kortlægning

På kortet udarbejdet i undersøgelsens første fase var langt størstedelen af markarealerne i Østdanmark forudsagt som drænede (Figur 15). Derudover var der forudsagt større drænede områder i Thy og på Mors, på littorinafladerne i Nordjylland og i marsken i Sydvestjylland. De største arealer uden dræning blev forudsagt i Himmerland, i Vestjylland og på højbundsjorderne i Nordjylland. I alt blev 48% af markarealet forudsagt som drænet.



Figur 15: Kort over drænede arealer udarbejdet i undersøgelsens første fase. Metoden er beskrevet i Møller et al. (2018) (Bilag 9).

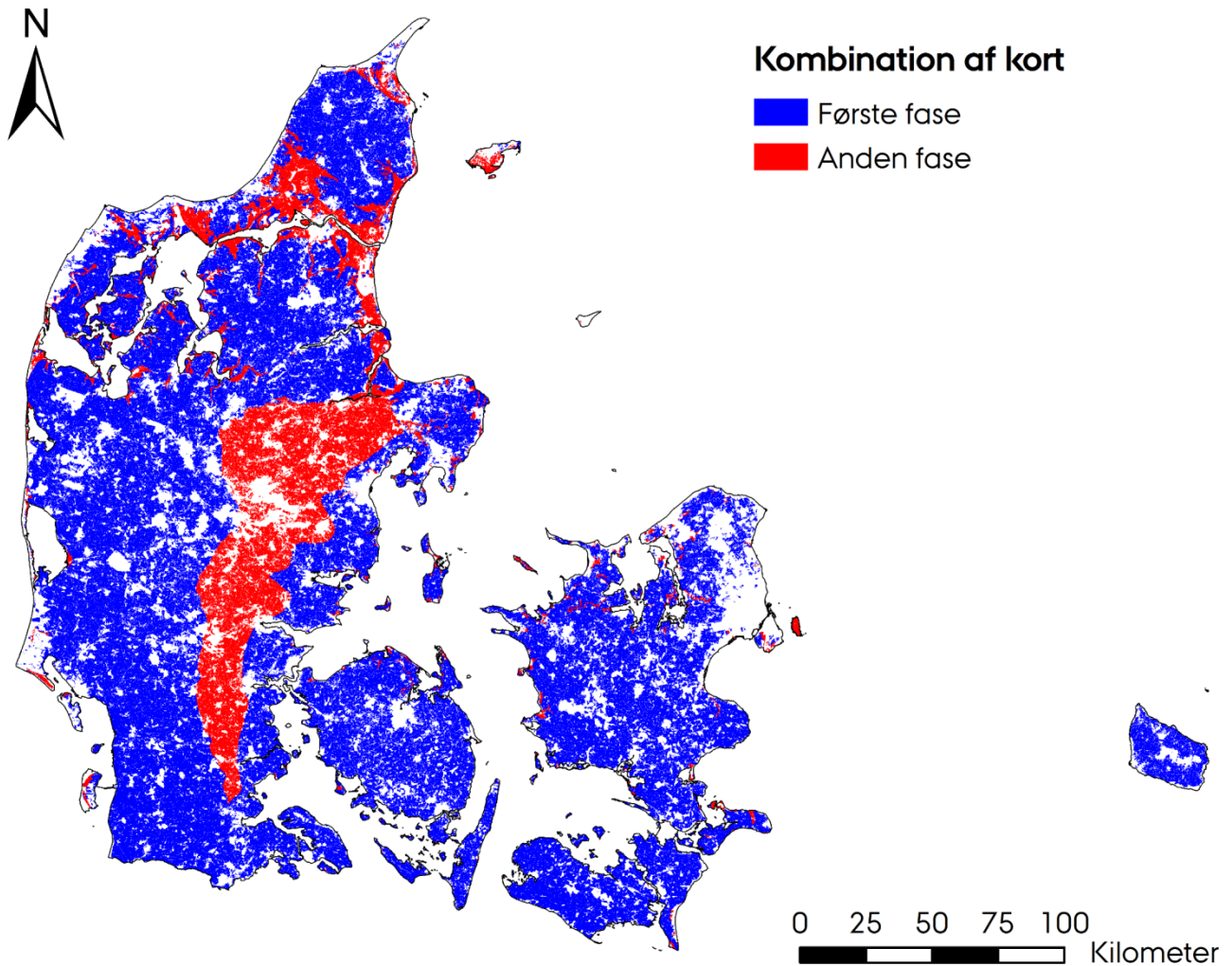
På kortet fra undersøgelsens anden fase var det drænede areal væsentligt større (Figur 16). Det drænede areal var især større i Midtjylland, på den lave del af smeltevandssletterne og på bakkeøerne. Omvendt var det drænede areal i Nordjylland og i marsken mindre. I alt blev 57% af markarealet forudsagt som drænet.



Figur 16: Kort over drænedede og ikke drænedede arealer i Danmark udarbejdet i undersøgelsens anden fase med et større dataset sammensat af observationer fra Olesen (2009), Skriver og Hedegård (1973) og fra Orbicons drænarkiv (Orbicon, 2015).

3.3 Kombination af kort

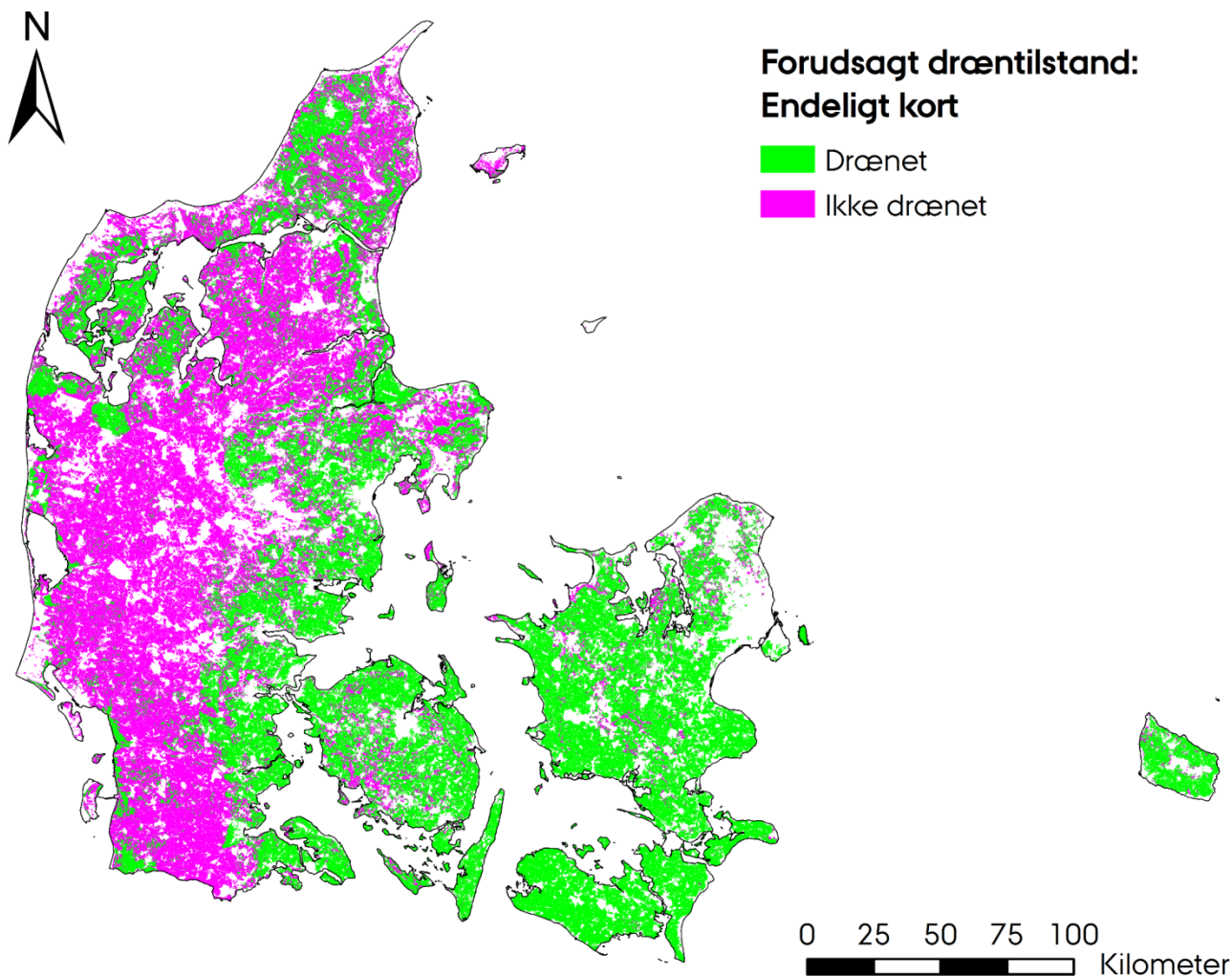
I to områder havde kortet fra undersøgelsens anden fase en statistisk signifikant højere nøjagtighed end kortet fra undersøgelsens første fase. For det første var det mere nøjagtigt på Littorinafladerne. For det andet var det mere nøjagtigt i Midtjylland. For alle andre områder, og for geologiske klasser, var der ingen statistisk signifikante forskelle (Bilag 5). På det endelige kort blev kortet fra undersøgelsens anden fase derfor anvendt for Littorinafladerne og Midtjylland, mens kortet fra undersøgelsens første fase blev anvendt for de øvrige områder (Figur 17).



Figur 17: Kombination af de to kort over drænede arealer. I det blå område blev forudsigelserne fra undersøgelsens første fase anvendt. I det røde område (Georegion 4 og Littorinaflader) blev kortet fra undersøgelsens anden fase anvendt.

På det kombinerede kort over drænede arealer er der meget få drænede arealer i Vestjylland sammenlignet med kortet fra undersøgelsens anden fase (Figur 18). Omvendt er det drænede areal i Østjylland større end på kortet fra undersøgelsens første fase (Figur 15), og udbredelsen af dræning er mere detaljeret kortlagt på Littorinafladerne end på det første kort.

I alt fremstår 52% af markarealet som drænet på det kombinerede kort. Denne andel er højere end drænprocenterne forudsagt af Olesen (2009) (49%) og på kortet fra undersøgelsens første fase (48%). Omvendt var det lavere end på kortet fra undersøgelsens anden fase (57%).



Figur 18: Endeligt kort over drænedede arealer. Kortet er en kombination af kortene fra undersøgelsens to faser.

3.4 Kortets nøjagtighed

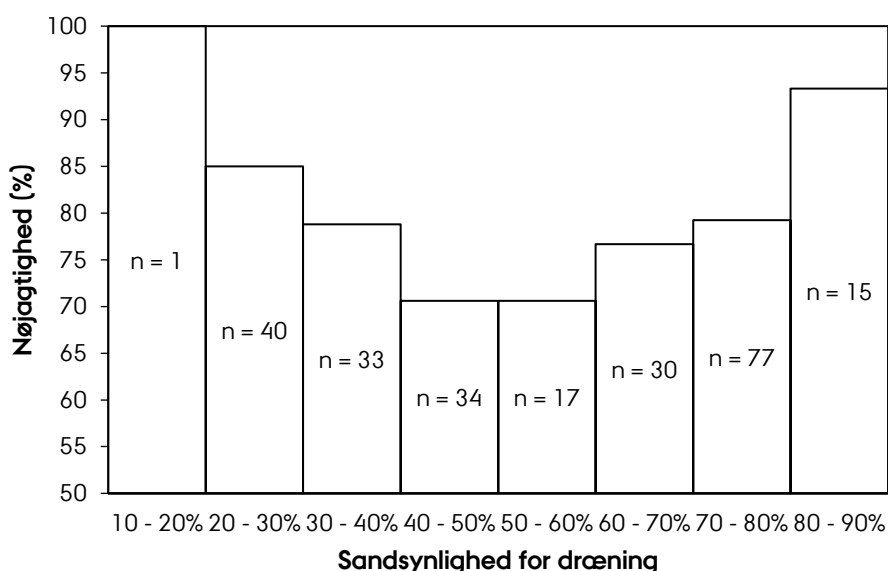
Udregningen på baggrund af det uafhængige datasæt viste, at det kombinerede kort har en nøjagtighed på 79%, hvilket er højere end nøjagtigheden på 77% for kortet udarbejdet i undersøgelsens første fase. Nøjagtigheden blev i begge tilfælde udregnet på baggrund af de samme 247 observationer (valideringsdatasættet), og forbedringen svarer til, at yderligere seks punkter blev forudsagt korrekt på det kombinerede kort. En binomial test viser at forskellen er statistisk signifikant ($n = 247$; $p < 0,05$).

Nøjagtigheden af det kombinerede kort er også højere end nøjagtigheden af kortet udarbejdet af Olesen (2009), beregnet for de samme 247 punkter. For disse punkter havde kortet udarbejdet af Olesen (2009) en nøjagtighed på 72%. Forskellen er bemærkelsesværdig, da Olesen (2009) brugte alle de tilgængelige punkter til at udregne drænprocenter. Punkterne er derfor ikke uafhængige af kortet udarbejdet af Olesen (2009), og man skulle derfor forvente, at der var en høj grad af overensstemmelse mellem punkterne og kortet udarbejdet af Olesen (2009).

For at undersøge, om den forudsagte sandsynlighed for dræning kan angive det kombinerede korts nøjagtighed i specifikke områder, blev observationerne i det uafhængige datasæt inddelt i otte intervaller baseret på den forudsagte sandsynlighed for dræning. Nøjagtigheden blev herefter beregnet separat for observationerne i hvert interval (Figur 19).

Der er generelt lavest nøjagtighed (71%) i sandsynlighedsintervallerne fra 40% til 60%, og nøjagtigheden stiger generelt desto længere sandsynligheden for dræning er fra 50%. En sandsynlighed for dræning tæt på 50% afspejler således, at det er usikkert om arealet er drænet, mens en sandsynlighed for dræning, der enten er markant højere eller lavere, afspejler en større sikkerhed for angivelsen af om arealet er drænet.

Nøjagtigheden er generelt større for intervaller med lav sandsynlighed for dræning end for intervaller med høj sandsynlighed for dræning. Den største nøjagtighed findes i intervallet 10 - 20%, men er kun baseret på en enkelt observation og derfor ikke repræsentativ.



Figur 19: Nøjagtighed, beregnet som procentandelen af korrekt forudsagte punkter på grundlag af det uafhængige datasæt, overfor den forudsagte sandsynlighed for dræning inddelt i otte intervaller.

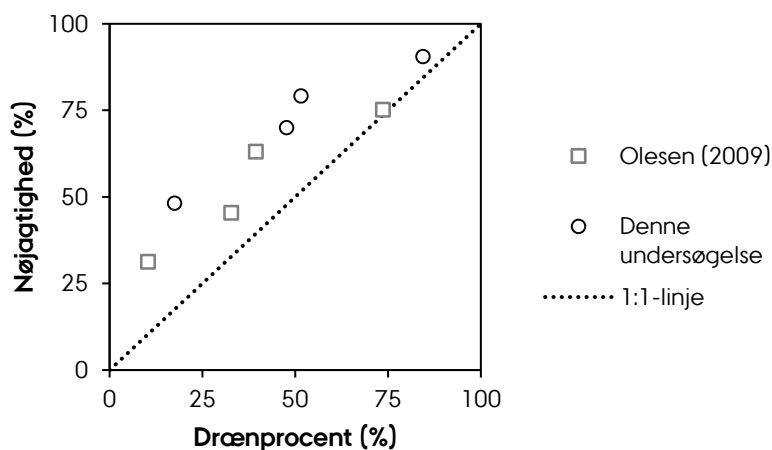
3.4.1 Hoveddræn

På kommunebasis var den forudsagte drænprocent for markarealet størst i Aarhus Kommune (84%) og mindst i Vesthimmerlands Kommune (18%) (Tabel 5). For nøjagtigheden var mønstret det samme. Her havde Aarhus Kommune den største nøjagtighed (91%) og Vesthimmerlands Kommune den mindste (48%). Overordnet fremstår langt de fleste arealer med hoveddræn som drænedede på det endelige kort. Dækningen af hoveddræn er generelt dårligst i kommuner, hvor den forudsagte drænprocent er lav. Af samme grund er det i absolutte tal et begrænset areal, der ikke er forudsagt som drænet, selv om der er hoveddræn.

Tabel 5: Sammenligning med hoveddræn registreret af i Orbicons drænarkiv. Tabellen angiver den forudsagte drænprocent for hver kommunes markareal på kortet udarbejdet af Olesen (2009) og i denne undersøgelse. Tabellen angiver desuden nøjagtigheden af de to kort, udregnet som andelen af celler med registrerede hoveddræn inden for kommunen.

Kommune	Drænprocent (%)		Nøjagtighed (%)	
	Olesen (2009)	Denne undersøgelse	Olesen (2009)	Denne undersøgelse
Randers	33	48	45	70
Vejle	39	52	63	79
Vesthimmerland	10	18	31	48
Aarhus	74	84	75	91

Både den forudsagte drænprocent og nøjagtigheden er større for kortet udarbejdet i denne undersøgelse end for kortet udarbejdet af Olesen (2009). Nøjagtigheden var i begge tilfælde størst for kommuner med en høj drænprocent (Figur 20). Dette var forventet, da nøjagtigheden udelukkende blev udregnet som andelen af celler med hoveddræn, der blev forudsagt som drænede på kortet. En forudsagt drænprocent på 0% ville derfor give en nøjagtighed på 0%, da ingen områder med hoveddræn ville blive forudsagt som drænede. Omvendt ville en forudsagt drænprocent på 100% ville give en nøjagtighed på 100%, da alle arealer med hoveddræn med garanti ville blive forudsagt som drænede.



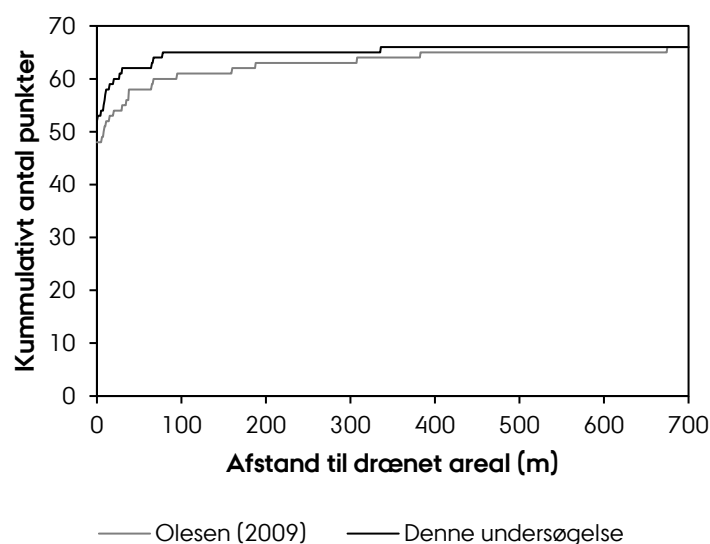
Figur 20: Sammenligning mellem den forudsagte drænprocent og nøjagtigheden af den forudsagte udbredelse af dræning for kortene udarbejdet i denne undersøgelse og af Olesen (2009) for fire kommuner (Tabel 5). Nøjagtigheden er udregnet som andelen af celler med registrerede hoveddræn, der blev forudsagt som drænede på de to kort.

Nøjagtigheden er dog i alle tilfælde højere end den forudsagte drænprocent for kommunen, hvilket viser, at der er et betydeligt sammenfald mellem hoveddræn og forudsigelsen af drænede arealer. Forskellen mellem nøjagtigheden og den forudsagte drænprocent var større for kortet udarbejdet i denne undersøgelse

end for kortet udarbejdet af Olesen (2009) (beregnet som en parvis t-test; $n = 4$; $p < 0,05$). Kortet udarbejdet i denne undersøgelse har således en større nøjagtighed i forhold til den forudsagte drænprocent.

3.4.2 Drænuvløb

Af de 66 registrerede drænuvløb, lå otte uden for markarealet. Af de resterende 58 drænuvløb blev 52 forudsagt som drænede på kortet udarbejdet i denne undersøgelse, hvorimod 48 drænuvløb blev forudsagt som drænede på kortet udarbejdet af Olesen (2009) (Figur 21). Størstedelen af de drænuvløb, som lå uden for arealer, der blev forudsagt som drænede, lå i begge tilfælde inden for en afstand af 100 meter fra et areal, der blev forudsagt som drænnet. På kortet udarbejdet af Olesen (2009) lå fem drænuvløb mere end 100 meter fra et areal, der var forudsagt som drænnet, mens det samme kun gjaldt et enkelt punkt på kortet udarbejdet i denne undersøgelse.



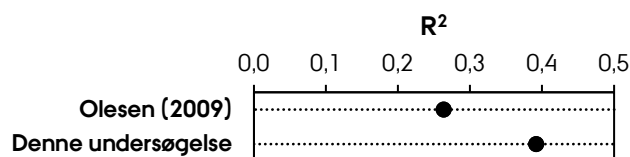
Figur 21: Kumulativ fordeling over punkternes afstand til det nærmeste areal, der blev forudsagt som drænnet på kortene udarbejdet af Olesen (2009) og i denne undersøgelse. Baseret på 66 punkter med målt drænastrømning (Figur 10).

Der er derfor stor overensstemmelse mellem placeringen af drænuvløb og arealer, som er forudsagt som drænede. Overensstemmelsen er desuden større for kortet udarbejdet i denne undersøgelse end for kortet udarbejdet af Olesen (2009).

3.4.3 Drænprocenter på markniveau

Den vægtede regression mellem de forudsagte drænprocenter og drænprocenterne kortlagt på markniveau i 17 oplande viste en statistisk signifikant korrelation ($n = 1285$; $p < 0,05$). Dette var tilfældet både for kortet udarbejdet af Olesen (2009) og for kortet udarbejdet i denne undersøgelse. Korrelationskoefficienten var dog højere for kortet udarbejdet i denne undersøgelse ($R^2 = 0,39$) end for kortet udarbejdet af Olesen (2009) ($R^2 = 0,26$) (Figur 22). Der er derfor en rimelig sammenhæng mellem de forudsagte drænprocenter og

drænprocenterne kortlagt på markniveau, og sammenhængen er stærkere for kortet udarbejdet i denne undersøgelse end for kortet udarbejdet af Olesen (2009).



Figur 22: Korrelationskoefficient (R^2) for den vægtede regression mellem drænprocenter kortlagt på markniveau i 17 oplande (Figur 11) og drænprocenter forudsagt af Olesen (2009) og i denne undersøgelse.

3.4.4 Drænrør i kabelgrave

Langs kabelstrækningerne varierede den forudsagte drænprocent fra 7% langs Blåbjerg-Holsted-strækningen til 92% langs Røjle-Fynsværket-strækningen (Tabel 6). Den forudsagte drænprocent fulgte generelt det samme mønster som andelen af celler med drænrør, men den forudsagte drænprocent var væsentligt større i alle andre tilfælde end på Blåbjerg-Holsted-strækningen. Overensstemmelsen mellem celler med registrerede drænrør og forudsagte drænedede arealer varierede fra 44% på Sønderborg-Fynshav-strækningen til 77% på Blåbjerg-Holsted-strækningen. Overensstemmelsen var således generelt størst på strækninger med en lav drænprocent og mindst på strækninger med en høj drænprocent. Uoverensstemmelserne skyldtes især celler der blev forudsagt som drænedede, uden at der var registreret drænrør (Bilag 6).

På Bedsted-Struer-strækningen havde kortet udarbejdet i denne undersøgelse en lavere nøjagtighed end kortet udarbejdet af Olesen (2009), men for de øvrige strækninger var kortet udarbejdet i denne undersøgelse mest nøjagtigt.

Tabel 6: Sammenligning mellem den forudsagte udbredelse af dræning og drænrør registeret af Energinet i forbindelse med nedgravning af elkabler. For hver strækning er det angivet, hvor mange af cellerne, der indeholdt registrerede drænrør, og hvor mange af cellerne, der var forudsagt som drænedede på kortene udarbejdet af Olesen (2009) og i denne undersøgelse. Samtidig er overensstemmelsen mellem udbredelsen af celler med registrerede drænrør og forudsigelsen af drænedede arealer beregnet for hver strækning (Bilag 6).

Strækning	Celler med drænrør (%)	Drænprocent (%)		Overensstemmelse (%)	
		Olesen (2009)	Denne undersøgelse	Olesen (2009)	Denne undersøgelse
Bedsted-Struer	22	63	58	67	57
Blåbjerg-Holsted	24	22	7	74	77
Roslev-Bilstrup	39	55	67	49	53
Røjle-Fynsværket	39	94	92	41	45
Sønderborg-Fynshav	34	97	90	38	44

Opdelt på geologiske klasser, var der størst overensstemmelse mellem den forudsagte drænprocent og celler med registrerede drænrør på smeltevandssand og morænesand (Tabel 7). Den laveste overensstemmelse forekom derimod på moræneler. Der var generelt mindre overensstemmelse i områder med leret undergrund (47 – 52%) end i områder med sandet undergrund (63 – 81%).

For de fleste geologiske klasser var der ikke nogen stor forskel på nøjagtigheden af kortet udarbejdet i denne undersøgelse og kortet udarbejdet af Olesen (2009). Den største forskel forekom i områder med ferskvandssand, hvor kortet udarbejdet i denne undersøgelse havde en væsentligt højere nøjagtighed.

Tabel 7: Sammenligning mellem den forudsagte udbredelse af dræning og drænrør registeret af Energinet i forbindelse med nedgravning af elkabler. For hver geologisk klasse er det angivet, hvor mange af cellerne, der indeholdt registrerede drænrør (Observeret) og hvor mange af cellerne, der var forudsagt som drænedede på kortene udarbejdet af Olesen (2009) og i denne undersøgelse. Samtidig er overensstemmelsen mellem udbredelsen af celler med registrerede drænrør og forudsigelsen af drænedede arealer beregnet for hver geologisk klasse (Bilag 6).

Geologisk klasse	Celler med drænrør (%)	Drænprocent (%)		Overensstemmelse (%)	
		Olesen (2009)	Denne undersøgelse	Olesen (2009)	Denne undersøgelse
Ferskvandsler	44	82	73	52	52
Ferskvandssand	34	39	19	47	69
Ferskvandstørv	48	70	47	51	50
Havsand	16	46	49	62	63
Moræneler	36	79	82	47	47
Morænesand	17	7	6	81	80
Smeltevandssand	20	0	8	80	81

Opdelt på landskabselementer var der størst overensstemmelse mellem forudsigelsen af drænedede arealer og udbredelsen af registrerede drænrør på smeltevandssletter og bakkeøer og mindst overensstemmelse i dødislandskaber (Tabel 8). Dødislandskaber havde også det største misforhold mellem den forudsagte drænprocent og andelen af celler med registrerede drænrør, da den forudsagte drænprocent var mere end tre gange så stor som andelen af celler med registrerede drænrør. På smeltevandssletter og bakkeøer var den forudsagte drænprocent mindre end andelen af celler med registrerede drænrør.

For de fleste landskabselementer var der ikke nogen stor forskel på nøjagtigheden af kortene udarbejdet i denne undersøgelse og af Olesen (2009). Randmoræner og dødislandskaber udgjorde en undtagelse, da nøjagtigheden var højere for randmoræner og lavere for dødislandskaber på kortet udarbejdet i denne undersøgelse.

Tabel 8: Sammenligning mellem den forudsagte udbredelse af dræning og drænrør registeret af Energinet i forbindelse med nedgravning af elkabler. For hvert landskabselement er det angivet, hvor mange af cellerne, der indeholdt registrerede drænrør (Observeret) og hvor mange af cellerne, der var forudsagt som drænedede på kortene udarbejdet af Olesen (2009) og i denne undersøgelse. Samtidig er overensstemmelsen mellem udbredelsen af celler med registrerede drænrør og forudsigelsen af drænedede arealer beregnet for hvert landskabselement (Bilag 6).

Landskabselement	Celler med drænrør (%)	Drænprocent (%)		Overensstemmelse (%)	
		Olesen (2009)	Denne undersøgelse	Olesen (2009)	Denne undersøgelse
Littorinaflade	24	60	61	59	61
Randmoræne	53	99	88	53	64
Yngre moræneflade	32	65	75	52	51
Dødislandskab	28	74	86	45	39
Smeltevandsslette	25	5	8	75	75
Bakkeø	25	24	8	72	76

Tabel 9: Sammenligning mellem den forudsagte udbredelse af dræning og drænrør registeret af Energinet i dødislandskaber langs de tre kabelstrækninger, der krydser dette landskabselement. For hver strækning er det angivet, hvor mange af cellerne, der indeholdt registrerede drænrør (Observeret) og hvor mange af cellerne, der var forudsagt som drænedede på kortene udarbejdet af Olesen (2009) og i denne undersøgelse. Samtidig er overensstemmelsen mellem udbredelsen af celler med registrerede drænrør og forudsigelsen af drænedede arealer beregnet for hver strækning (Bilag 6).

Strækning	Celler med drænrør (%)	Drænprocent (%)		Overensstemmelse (%)	
		Olesen (2009)	Denne undersøgelse	Olesen (2009)	Denne undersøgelse
Bedsted-Struer	18	6	78	79	33
Røjle-Fynsværket	34	97	97	34	40
Sønderborg-Fynshav	27	94	98	33	43

4 Diskussion

Kortet over drænedede arealer, som er udarbejdet i denne undersøgelse adskiller sig fra kortet af Olesen (2009) ved at bygge på flere geografiske datalag, flere punktobservationer og mere komplekse statistiske modeller. De nye kort har en større nøjagtighed for de fleste områder.

4.1 Geografiske datalag

Olesen (2009) opbyggede det tidligere kort ud fra jordbundstype, geologi, georegion, landskabselementer og vådområder. Det blev på den baggrund vurderet, at et bedre kort ville kunne udarbejdes ved at inddrage topografien, da den blandt andet ville kunne anvendes til at kortlægge drænforholdene på smeltevands-sletterne. Det er derfor bemærkelsesværdigt, at seks af de vigtigste ti datalag i undersøgelsens anden fase relaterer til topografien (Tabel 4).

Næsten alle datalag havde i større eller mindre grad betydning for kortlægningen af drænedede arealer. Dette viser, at placeringen af drænedede og ikke drænedede arealer bedst forstås som et samspil mellem adskillige variabler, der ikke i sig selv, men tilsammen afgør, om et areal bliver drænet. Et fladt areal kan for eksempel være naturligt veldrænet, hvis jorden er sandet, mens det med større sandsynlighed vil være vandlidende, hvis lerindholdet er højt.

Modellerne i denne undersøgelse benytter desuden kun statistiske sammenhænge mellem datalagene og dræning. Der kan derfor være tale om indirekte årsager til dræning, eller om forhold, der påvirkes af dræning, og derfor kan indikere, om der er drænet.

Det ses også, at ikke alle datalag har lige stor betydning for udbredelsen af dræning. Det er derfor ikke alle de datalag, der er blevet inddraget, der bør indgå i en operationel forståelse af, hvor der er og bliver drænet. Der bør derimod fokuseres på de vigtigste datalag.

Kun to af de datalag som Olesen (2009) anvendte (geologi og tekstur) er blandt de vigtigste 10 datalag. Samtidig indgår kortene over landskabselementer og vådområder end ikke blandt de vigtigste 30 datalag (Tabel 4).

I det første tilfælde ser det ud til, at landskabselementer har relativt lille betydning for udbredelsen af dræning. Littorinafladerne indgik dog i kombinationen af kort fra undersøgelsens to faser, og bakkeøer og smeltevandssletter indgik som betydelige datalag i undersøgelsens første fase (Møller et al., 2018) (Bilag 9).

I det andet tilfælde kan den lave betydning af vådområder skyldes, at der er inddraget kort over dybden til grundvandet, der i store træk viser det samme mønster. På kortet fra undersøgelsens første fase indgik vådområder dog blandt de vigtigste 30 datalag.

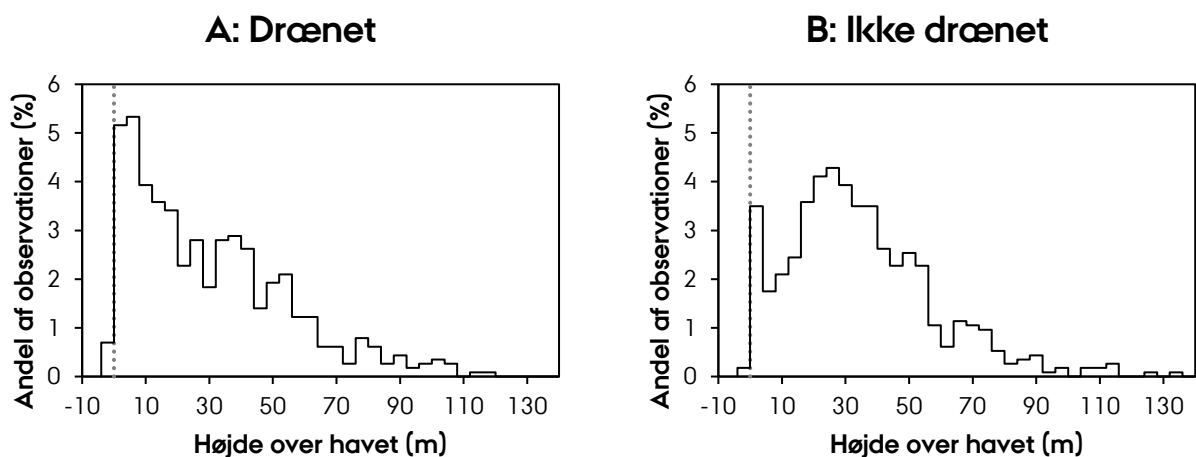
For begge kort var lerindholdet det vigtigste datalag og fremstår derfor som den vigtigste variabel i henseende til dræning. Den store betydning af lerindholdet stemmer overens med de kendte årsager til

dræning. Det kan i øvrigt forventes, at lerindholdet i de dybere intervaller var vigtigst, da drænrør typisk ligger i dybder på 70 – 120 cm under terræn (Nielsen, 2015).

I undersøgelsens første fase var lerindholdet i dybden 60 – 100 cm det vigtigste datalag, hvilket stemmer overens med denne forventning. I undersøgelsens anden fase var lerindholdet i topjorden derimod det vigtigste datalag. Forskellen kan skyldes, at de fleste af observationerne i undersøgelsens første fase var placeret ved jordbundsprofiler med målinger af lerindholdet i dybden. For det udvidede datasæt var lerindholdet derimod ukendt for mange af observationerne, og det var derfor nødvendigt at anvende lerindholdet fra det geografiske datalag. Kortet over lerindholdet i topjorden bygger på flere observationer og derfor er mere nøjagtigt end kortet over lerindholdet i underjorden. Lerindholdet i topjorden er kortlagt på baggrund af mere end 40.000 jordprøver fordelt over hele landet, mens lerindholdet i dybden kun er kortlagt på baggrund af et par tusind jordbundsprofiler (Adhikari et al., 2013). Det forklarer sandsynligvis også, hvorfor udbredelsen af moræneler på GEUS' jordartskort (Jakobsen et al., 2015) var vigtigere end lerindholdet i dybden, da dette kort også bygger på flere observationer. Det har sandsynligvis også betydning, at lerindholdet i topjorden og i underjorden er korreleret, og at lerindholdet i topjorden derfor kan give et estimat af lerindholdet i underjorden. Den relative betydning af lerindholdet i de enkelte dybdeintervaller er derfor primært et udtryk for sikkerheden inden for hvert interval.

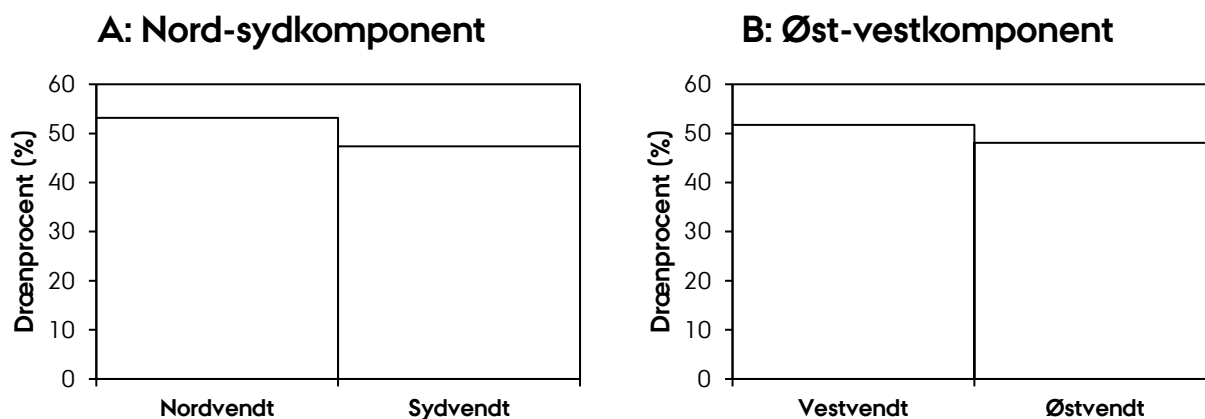
De to kort over dybden til grundvandet var vigtige i begge faser af undersøgelsen. Dybden til grundvandet havde især stor betydning i undersøgelsens anden fase. Grundvandets store betydning stemmer overens med årsagerne til dræning (Nielsen, 2015).

Efter dybden til grundvandet var afstanden og hældningen til overfladevand blandt de vigtigste datalag. Dette er ikke overraskende, da et fladt areal med ringe vertikal afstand til et vandløb potentielt vil være mere vandlidende, men også vanskeligere at dræne (Nielsen, 2015). Højden over havet havde dog også stor betydning for kortlægningen af drænede arealer (Tabel 4). Dette skyldes, at drænede arealer ofte er lavtliggende (Figur 23). Af samme årsag vil det mange steder være vanskeligt at uddybe vandløbene yderligere uden at skulle pumpe vandet bort, hvor der er lav højde over havet eller ringe faldforhold.



Figur 23: Fordeling af henholdsvis drænedede og ikke drænedede observationer overfor deres højde over havet. Baseret på datasættet til opbygning af modeller. Højden over havet er inddelt i intervaller på fire meter. De optrukne linjer angiver, hvor stor en del af de drænedede henholdsvis ikke drænedede observationer, der ligger inden for hvert interval. Den stiplede lodrette linje angiver havniveauet.

Resultaterne viser også, at lokale terrænforhold kan have stor betydning for udbredelsen af dræning. Den relativt store betydning af den akkumulerede afstrømning og det topografiske vådhedsindeks viser, at vandets strømningsveje i landskabet kan have stor betydning. Erfaring viser også, at både hoveddræn og detaildræn er ofte ligger der, hvor overskudsnedbøren naturligt vil samles og strømmer af. Solindstrålingen har også betydning for udbredelsen af dræning, sandsynligvis på grund af relationen til fordampning. Når nord-syd-komponenten af overfladens kompasretning er blandt de vigtigste 30 datalag, er det sandsynligvis fordi den relaterer til solindstrålingen. Omvendt kan betydningen af øst-vest-komponenten skyldes, at skråninger i den hyppigste vindretning generelt modtager større mængder nedbør (Beullens et al., 2014). Datasættet til opbygning af modeller viser en større drænprocent for nordvendte skråninger overfor sydvendte, og vestvendte overfor østvendte (Figur 24). Forskellene er ikke markante (henholdsvis 53% mod 47% og 52% mod 48%), men de er i begge tilfælde statistisk signifikante (beregnet som en binomial test; $n = 1144$; $p < 0.05$).



Figur 24: Drænprocent afhængigt af overfladens kompasretning, opdelt på (A) nord-sydkomponenten og (B) øst-vestkomponenten. Både nord-sydkomponenten og øst-vestkomponenten til overfladens kompasretning er oprindeligt beregnet som kontinuerte variabler (Bilag 1), men er her hver opdelt i to generelle klasser. Baseret på datasættet til opbygning af modeller.

Flere af satellitbilledernes informationer havde også betydning for kortlægningen af drænedede arealer. Det kan for det første skyldes, at fugtig jord har en lavere reflektivitet og derfor er mørkere end tør jord. Jordens reflektivitet kan derfor indikere, om jorden er vandlidende og derfor også, om den er drænet. Det infrarøde spektrum har særligt stor udsagnsværdi, da vand absorberer infrarød stråling (Verma et al., 1996, Northcott et al., 2000, Naz og Bowling, 2008, Naz et al., 2009). For det andet kan betydningen af satellitbilledernes informationer skyldes, at vegetationen reagerer på jordens drænforhold, eller at drænforholdene har betydning for afgrødevalget. Det sidste stemmer overens med, at antallet af år med drænkrevende afgrøder havde betydning for kortlægningen af drænedede arealer (Tabel 4).

Overordnet er årsagerne til, at der drænes generelt meget enkle, og de stemmer overens med eksisterende viden. Dræning er i de fleste tilfælde et virkemiddel mod et for højtliggende grundvandsspejl og jordens tekstur og deraf forringede evne til at bortlede overskudsnedbøren. Denne undersøgelse viser dog, at der også er mange andre lokale forhold, der kan have betydning for at forudsige, hvor der er drænet, herunder topografi og afgrødevalg.

4.2 Observationer

Kortet fra undersøgelsens første fase var primært baseret på modeller, der kun indeholdt 10 PCs (Bilag 9). Det vil sige, at de datalag, der blev anvendt, kun omfattede lidt over halvdelen af den variation, der fandtes i det samlede datasæt (Bilag 4). Der var således tale om forholdsvis simple modeller. I undersøgelsens anden fase var de fleste af de anvendte modeller derimod opbygget med det oprindelige datasæt, der indeholdt flere variabler. Der var derfor tale om mere komplekse modeller.

Den væsentligste forskel på kortene udarbejdet i undersøgelsens to faser er, at undersøgelsens anden fase inddrog observationer fra flere forskellige datakilder. Denne kendsgerning tyder på, at det kræver komplekse

modeller og benyttelse af mange drænrelaterede variabler at opnå nøjagtige forudsigelser af, om givne arealer er drænede.

4.3 Nøjagtighed

Det kombinerede kort havde en nøjagtighed på 79% for det uafhængige datasæt, hvilket er statistisk signifikant højere end nøjagtigheden af kortet fra undersøgelsens første fase. Det kombinerede kort indeholdt kun forudsigelser fra undersøgelsens anden fase i et mindre område, mens kortet fra undersøgelsens første fase blev anvendt i de fleste områder. Indsamlingen af yderligere observationer havde derfor kun betydning for kortets nøjagtighed i et mindre område. Kombinationen af de to kort gav dog en større nøjagtighed end kortene havde hver for sig. Det kombinerede kort havde en væsentlig større nøjagtighed end kortet udarbejdet af Olesen (2009) (72%).

Sammenligningen med Orbicons kort over hoveddræn og placeringen af drænudløb viste, at forudsigelsen af drænede arealer dækkede langt størstedelen af de drænede arealer. Sammenligningen med drænprocenter kortlagt på markniveau viste også, at der var god sammenhæng mellem de kortlagte mønstre. Kortet udarbejdet i denne undersøgelse viste i alle tilfælde en større grad af overensstemmelse med de faktiske forhold end kortet udarbejdet af Olesen (2009).

Overordnet vurderes nøjagtigheden af det endelige kort at være god, men det bemærkes, at kortet ikke viser drænede arealer med 100% nøjagtighed. Det kan blandt andet skyldes, at beslutningen om at dræne ikke kun afhænger af de naturgivne forhold, men også af landbrugspraksis og økonomi. Således kan arealer med lille drænbehov på den ene side være drænede, mens arealer med stort drænbehov i nogle tilfælde ikke er drænet. Både lovgivningen for dræning, støtteordninger og landbrugets generelle økonomi har ligesom arealanvendelsen og afgrødevalget ændret sig betydeligt siden dræning med nedgravede rør blev indført for mere end 150 år siden. I sidste ende er det den enkelte landmands beslutning, om en mark skal drænes. Forudsigelsen af drænede arealer på grundlag af drænrelaterede variabler stemmer derfor ikke altid overens med de faktiske forhold. Hver generation kan have haft sin egen vurdering af, hvad der kunne betale sig. Det forklarer også, at nøjagtigheden er lavest, hvor sandsynligheden for dræning ligger tæt på 50%, idet landmanden i så fald har stået over for spørgsmålet: skal eller skal ikke? For disse jorder har der måske – måske ikke været økonomisk gevinst ved at dræne, og denne usikkerhed påvirker kortets nøjagtighed.

Sammenligningen med drænrør registreret af Energinet viste, at der ikke var drænrør tilstede på en del af de arealer, der var forudsagt som drænede. Dette var især tilfældet i dødislandskaber, mens der var god overensstemmelse i områder med sandet undergrund. Det er muligt, at nogle af uoverensstemmelserne skyldes, at der er drænrør til stede, som ikke er blevet registreret. Inden for en enkelt celle kan det for eksempel tænkes, at drænrørene ikke har krydset kabelgraven, hvis de har ligget parallelt med den. På nogle jorder er det i øvrigt tilstrækkeligt med en drænaftand på op til 40 meter (Nielsen, 2015), og det er derfor muligt, at en celle er drænet af drænrør, der ligger udenfor cellen.

En mere subtil årsag til afvigelser mellem forudsagte og faktisk drænedede arealer kan også være, at et "drænet areal" ikke i alle tilfælde er ensbetydende med et "areal med drænrør". I undersøgelsen af Olesen (2009) blev det blot vurderet, om arealet var drænet eller ej, men ikke hvor drænrørene lå. Kortet udarbejdet af Olesen (2009) har en målestok på 1:100.000, var det ikke noget væsentligt problem, men da kortet i denne undersøgelse har en finere opløsning, kan det være med til at skabe uoverensstemmelser. På samme måde blev drænrørenes præcise placering heller ikke taget i betragtning, da der i denne undersøgelse blev indsamlet informationer om drænprojekter fra oversigtskortene til Orbicons drænarkiv. I stedet blev drænprojekterne indtegnet som polygoner, som derefter blev konverteret til punkter.

Når misforholdet imellem den forudsagte udbredelse af dræning og de registrerede drænrør er størst i dødislandskaber, kan det netop skyldes denne uoverensstemmelse. Dødislandskaber har et stærkt kuperet terræn med mange bakketoppe og lavninger. Det vil derfor ofte være tilstrækkeligt at punktdræne lavningerne, men drænrørene i lavningerne vil modtage vand fra størstedelen af marken. En mark i et dødislandskab kan derfor indeholde relativt få drænrør, selvom marken som helhed er "drænet". At kun en mindre del af marken indeholder drænrør vil kun fremgå, hvis der foreligger drænkort, eller hvis drænrørene er blevet kortlagt på anden vis.

En fremtidig kortlægning af drænedede arealer i Danmark bør derfor sikre, at der er drænrør under de punkter, der betragtes som drænedede. Dette gælder især, hvis der anvendes en finere opløsning end i denne undersøgelse. Drænrørenes placering ville kunne fastlægges ved at indhente drænoplysninger fra Orbicons drænarkiv (Orbicon, 2015), ved feltundersøgelser med sensorer (Allred et al., 2004, Allred og Redman, 2010) eller ved at identificere dem på flyfotos (Verma et al., 1996, Northcott et al., 2000, Naz og Bowling, 2008, Naz et al., 2009, Tetzlaff et al., 2009).

4.4 Anvendelse og usikkerheder

Det kombinerede kort har en høj nøjagtighed og dækker langt de fleste drænedede arealer. Det er derfor velegnet til at vurdere hvor der er drænet inden for større områder. Opløsningen på 30,4 meter betyder desuden, at det er muligt at anvende kortet til at vurdere placeringen af drænedede arealer på lokalt niveau eller endda på for enkelte marker.

For mindre områder kan lokale variationer og forhold dog betyde, at kortets forudsigelser afviger fra virkeligheden. I sådanne tilfælde er det vigtigt at tage den forudsagte sandsynlighed for dræning i betragtning, da kortets nøjagtighed var afhængig af den forudsagte sandsynlighed for dræning. Forudsigelsen var således mest usikker, når sandsynligheden for dræning var tæt på 50%, og mest sikker, når sandsynligheden for dræning var betydeligt mindre eller større end 50%.

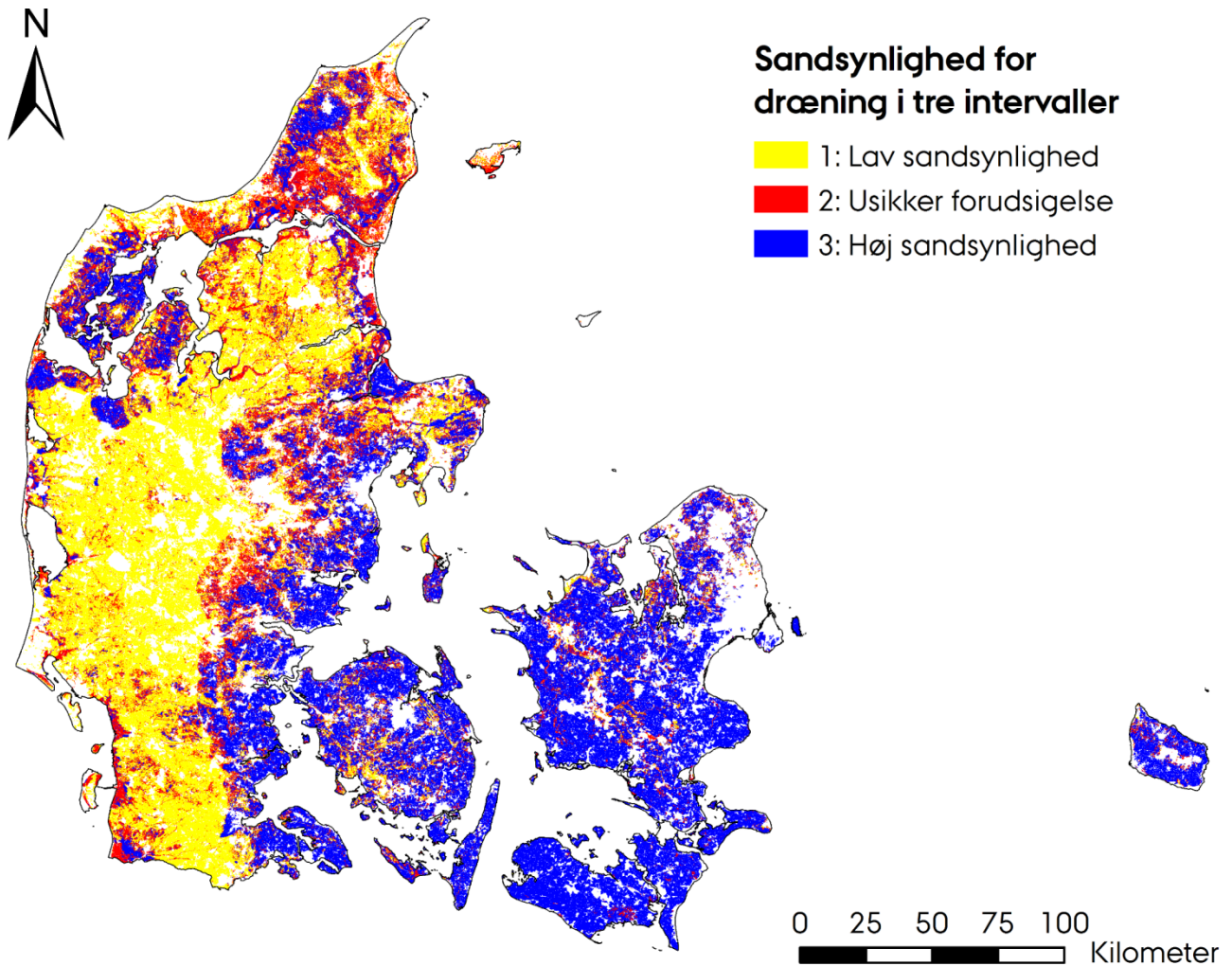
For at indarbejde denne sammenhæng, blev den forudsagte sandsynlighed for dræning opdelt i tre klasser (Tabel 10): Områder med lav sandsynlighed for dræning (0 - 40% sandsynlighed), områder, hvor forudsigelsen er usikker (40 - 60% sandsynlighed), og områder med høj sandsynlighed for dræning (60 -

100% sandsynlighed). Nøjagtigheden af forudsigelsen i de enkelte klasser var henholdsvis 82%, 71% og 80%. Nøjagtigheden i områder med usikre forudsigelser var således markant lavere end i de to øvrige klasser.

Tabel 10: Inddeling af den forudsagte sandsynlighed for dræning i tre klasser: Lav sandsynlighed for dræning, usikker forudsigelse og høj sandsynlighed for dræning. Tabellen angiver for hver klasse sandsynlighedsintervallet, nøjagtigheden af forudsigelsen, antal punkter anvendt til beregning af nøjagtigheden og klassens andel af det samlede markareal.

Klasse	Sandsynlighed (%)	Nøjagtighed (%)	Punkter	Areal (%)
Lav sandsynlighed	0 - 40	82	74	36
Usikker forudsigelse	40 - 60	71	51	21
Høj sandsynlighed	60 - 100	80	122	43

Områder med usikre forudsigelser udgør 21% af markarealet (Tabel 10). De forekommer især i Midtjylland, Nordjylland, i marsken og i ådale (Figur 25). Der bør derfor tages højde for den lavere sikkerhed i forudsigelsen, når kortet anvendes i disse områder. Kortet med de tre klasser indgår sammen med kortet over drænede arealer i det endelige produkt.

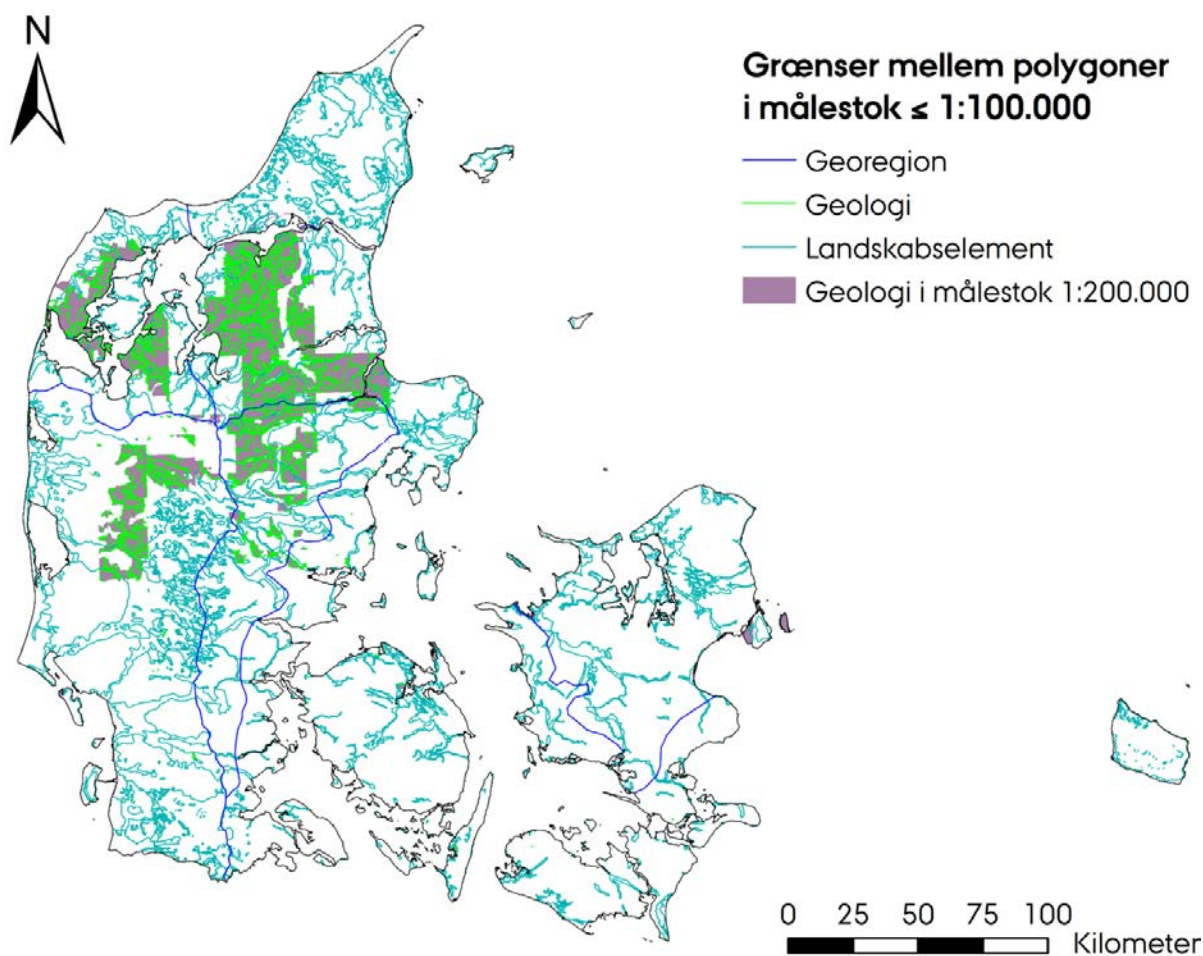


Figur 25: Den forudsagte sandsynlighed for dræning inddelt i tre intervaller: 1: Lav sandsynlighed for dræning (0 – 40%), 2: Usikker forudsigtelse (40 – 60 %) og 3: Høj sandsynlighed for dræning (60 – 100%). Nøjagtigheden af forudsigtelsen i de tre områder var 1: 82%, baseret på 74 punkter, 2: 71% baseret på 51 punkter og 3: 80% baseret på 122 punkter.

Et andet forhold, der skal tages i betragtning, når kortet anvendes lokalt er, at flere af de underliggende geografiske datalag har en grovere opløsning end den digitale højdemodel og satellitbillederne. For eksempel har kortene over landskabselementer og georegioner en målestok på 1:100.000. Hvis man antager, at de trykte kort har en usikkerhed på 1 millimeter, svarer det derfor til en geografisk usikkerhed på 100 meter. Hvis et areal ligger indenfor 100 meter af en grænsedragning imellem to polygoner, er det derfor muligt, at usikkerheden har påvirket forudsigtelsen af, om der er drænet. I disse tilfælde bør man derfor vurdere, om grænsedragningen er pålidelig, og om det kan have påvirket forudsigtelsen. Konkret ville man kunne undersøge, om grænsedragningen stemmer overens med topografien, da mange af grænsedragningerne er fastlagt ud fra topografien (Madsen et al., 1992, Gravesen et al., 2006).

GEUS' jordartskort med geologiske klasser dækker 88% af Danmarks areal i målestok 1:25.000 (Jakobsen et al., 2015). Den geografiske usikkerhed er derfor 25 meter, hvilket er mere nøjagtigt end højdemodellens opløsning på 30,4 meter. I det resterende område er geologien derimod kortlagt i målestok 1:200.000, og den geografiske usikkerhed er derfor langt større i disse områder.

De fleste grænsedragninger i målestok 1:100.000 eller mindre findes i Midtjylland og Himmerland (Figur 26). Der er mulighed for, at en usikker grænsedragning har påvirket forudsigelsen for et areal, hvis det ligger indenfor 100 meter af en grænsedragning tegnet i målestok 1:100.000 eller indenfor 200 meter af en grænsedragning tegnet i målestok 1:200.000. Udregnet på denne måde ligger i alt 3.232 km² af det kortlagte areal inden for områder, der kan påvirkes af usikre grænsedragninger. Dette svarer til 12% af det kortlagte areal.



Figur 26: Grænsedragninger mellem polygoner i lagene over georegioner, geologi og landskabselementer. Grænsedragningerne mellem de geologiske klasse omfatter kun de dele af kortet, der er tegnet i målestok 1:200.000. Geologiske grænsedragninger, der ligger mere end 200 meter fra det nærmeste markareal er udeladt, og grænsedragninger for landskabselementer og georegioner, der ligger mere end 100 meter fra det nærmeste markareal er også udeladt.

35 af observationerne i det uafhængige datasæt ligger på arealer, der kan være påvirket af usikre grænsedragninger, hvilket svarer til 14% af datasættet. Af disse punkter er 28, svarende til 80%, korrekt forudsagt på kortet. Dette ligger meget tæt på kortets overordnede nøjagtighed på 79%, og der er derfor ikke tegn på, at kortet generelt er mere unøjagtigt i områder med usikre grænsedragninger. Det udelukker ikke, at der kan være tilfælde, hvor en grænsedragning påvirker forudsigelsen, men det tyder på, at det kun gælder relativt få arealer. De usikre grænsedragninger og områder, der kan påvirkes af dem, indgår som datalag i det endelige produkt.

5 Konklusion

Kortlægningen af drænede arealer bygger i denne undersøgelse på et stort antal datalag. Betydningen af de enkelte datalag viste, at de vigtigste datalag generelt stemmer overens med de kendte årsager til dræning, såsom jordens tekstur og dybden til grundvandet. Det sås dog også, at mange lokale forhold, herunder topografi og afgrødevalg, kan have betydning for forudsigelsen af, om et areal er drænet eller ej. Samlet set kan årsagerne til dræning opfattes som resultatet af interaktioner mellem mange forskellige variabler. De anvendte modeller gjorde det muligt at medregne en stor del af denne kompleksitet.

Resultatet er et kort over drænede arealer, der er væsentligt mere nøjagtigt end kortet udarbejdet af Olesen (2009). Både sammenligning med observationerne anvendt af Olesen (2009), hoveddræn registreret af Orbicon, placeringen af drænudløb og drænprocenter kortlagt på markniveau viste, at kortet udarbejdet i denne undersøgelse havde en større nøjagtighed. Sammenligning med drænrør registreret af Energinet viste, at kortet udarbejdet i denne undersøgelse især var mere nøjagtigt end kortet udarbejdet af Olesen (2009) på sandede ferskvandsaflejringer og på randmoræner. Sammenligningen viste også, at kortet generelt er mere nøjagtigt i sandede områder end i lerede områder, og at kortets nøjagtighed var dårligst i dødislandskaber.

Mange af uoverensstemmelserne mellem kortet og registrerede drænrør skyldes tilsyneladende, at der ikke er drænrør på flere af de arealer, hvor kortet forudsiger, at der er drænet. Den underliggende årsag er, at kortet dækker drænede arealer i bred forstand, snarere end de enkelte drænrør, og at det ikke skelner mellem systemdrænede og punktdrænede arealer. Dette er et væsentligt forhold at tage højde for, når kortet anvendes i praktiske sammenhænge.

Der bør også tages højde for, at kortets nøjagtighed kan variere lokalt. Flere af de underliggende datalag har usikre grænsedragninger, og den forudsagte sandsynlighed for dræning har også betydning for nøjagtigheden. Grænsedragninger i grov målestok og den forudsagte sandsynlighed for dræning indgår derfor i det endelige produkt.

Overordnet har det udarbejdede kort en høj nøjagtighed og en stor detaljeringsgrad. Dette betyder, at det i langt højere grad end kortet udarbejdet af Olesen (2009) er egnet til analyser på lokalt niveau og i visse tilfælde endda på markniveau. Det endelige kort vurderes dermed at have forbedret kortlægningen af drænede arealer, når der tages højde for de nævnte begrænsninger.

Tak til

Afdelingsleder Henrik Vest Sørensen (Orbicon) tog initiativ til undersøgelsen og startede samarbejdet mellem AU og Orbicon.

Drænteknikker Karl Henrik Laursen (Orbicon) bidrog med praktisk erfaring om markdræning, kortlægning af dræn og forhold, der har betydning for dræning. Teknisk assistent Astrid Merete Borch (Orbicon) hjalp med at indsamle oplysninger fra Orbicons drænarkiv.

Tak til Energinet for tilladelse til brug af oplysninger om drænrør registreret i forbindelse med nedlægning af elkabler.

Landskonsulent Søren Kolind Hvid (SEGES) har gransket rapporten.

Referencer

- Adhikari, K., Kheir, R.B., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B. og Greve, M.H., 2013. High-resolution 3-D mapping of soil texture in Denmark. *Soil Sci. Soc. Am. J.* 77(3), 860-876. <http://dx.doi.org/10.2136/sssaj2012.0275>.
- Allred, B.J., Fausey, N.R., Peters, L., Chen, C., Daniels, J.J. og Youn, H., 2004. Detection of buried agricultural drainage pipe with geophysical methods. *Appl. Eng. Agric.* 20(3), 307-318. <http://dx.doi.org/10.13031/2013.16067>.
- Allred, B.J. og Redman, J.D., 2010. Location of agricultural drainage pipes and assessment of agricultural drainage pipe conditions using ground penetrating radar. *J. Environ. Eng. Geophys.* 15(3), 119-134. <http://dx.doi.org/10.2113/JEEG15.3.119>.
- Banfield, R.E., Hall, L.O., Bowyer, K.W. og Kegelmeyer, W.P., 2007. A comparison of decision tree ensemble creation techniques. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 29(1), 173-180. [http://dx.doi.org/Doi 10.1109/Tpami.2007.250609](http://dx.doi.org/Doi%2010.1109/Tpami.2007.250609).
- Bauer, E. og Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Mach. Learn.* 36(1-2), 105-139. <http://dx.doi.org/10.1023/A:1007515423169>.
- Beucher, A., Møller, A.B. og Greve, M.H., 2017. Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark. *Geoderma*. <http://dx.doi.org/10.1016/j.geoderma.2017.11.004>.
- Beullens, J., Van de Velde, D. og Nyssen, J., 2014. Impact of slope aspect on hydrological rainfall and on the magnitude of rill erosion in Belgium and northern France. *Catena* 114, 129-139. <http://dx.doi.org/10.1016/j.catena.2013.10.016>.
- Bose, I. og Mahapatra, R.K., 2001. Business data mining — a machine learning perspective. *Information & Management* 39(3), 211-225. [http://dx.doi.org/10.1016/s0378-7206\(01\)00091-x](http://dx.doi.org/10.1016/s0378-7206(01)00091-x).
- Brady, N.C. og Weil, R.R., 1996. *The nature and properties of soils*. Prentice-Hall Inc. ISBN 0132431890.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24(2), 123-140. <http://dx.doi.org/10.1023/A:1018054314350>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45(1), 5-32. <http://dx.doi.org/10.1023/A:1010933404324>.
- Breiman, L., Cutler, A., Liaw, A. og Wiener, M. 2015. randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.6-12. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf> (tilgået 18-08-2017).

- Caruana, R., Niculescu-Mizil, A., Crew, G. og Ksikes, A., 2004. Ensemble selection from libraries of models. Proceedings of the twenty-first international conference on Machine learning, ACM, 18. <http://dx.doi.org/10.1145/1015330.1015432>.
- Chawla, N.V., Bowyer, K.W., Hall, L.O. og Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research 16, 321-357. <http://dx.doi.org/10.1613/jair.953>.
- Collaku, A. og Harrison, S.A., 2002. Losses in wheat due to waterlogging. Crop Sci. 42(2), 444-450. <http://dx.doi.org/10.2135/cropsci2002.4440>.
- Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Mach. Learn. 40(2), 139-157. <http://dx.doi.org/10.1023/A:1007607513941>.
- Dinnes, D.L., Karlen, D.L., Jaynes, D.B., Kaspar, T.C., Hatfield, J.L., Colvin, T.S. og Cambardella, C.A., 2002. Nitrogen Management Strategies to Reduce Nitrate Leaching in Tile-Drained Midwestern Soils. Agronomy Journal 94(1), 153. <http://dx.doi.org/10.2134/agronj2002.1530>.
- Eriksen, J., Jensen, Poul Nordemann, Jacobsen, Brian H., 2014. Virkemidler til realisering af 2. generations vandplaner og målrettet arealregulering. DCA rapport. Aarhus University.
- Ernstsen, V., Olsen, P. og Rosenbom, A.E., 2015. Long-term monitoring of nitrate transport to drainage from three agricultural clayey till fields. Hydrol. Earth Syst. Sci. 19(8), 3475-3488. <http://dx.doi.org/10.5194/hess-19-3475-2015>.
- European Environment Agency, 2014. Corine Land Cover (CLC) 2012 - Denmark, Version 1, Oct. 2014 [dataset]. <http://download.kortforsyningen.dk/content/corine-land-cover>.
- Freund, Y. og Schapire, R.E., 1996. Experiments with a new boosting algorithm. ICML, 96, 148-156.
- Gambrell, R.P., Gilliam, J.W. og Weed, S.B., 1975. Denitrification in subsoils of the North Carolina coastal plain as affected by soil drainage. J. Environ. Qual. 4(3), 311-316. <http://dx.doi.org/10.2134/jeq1975.00472425000400030005x>.
- Gravesen, P., Pedersen, S.A.S., Klint, K.E.S. og Jakobsen, P.R., 2006. Geologiske kort i Danmark - hvad viser de kvartærgeologiske kort. GeologiskNyt 16(2). <http://dx.doi.org/10.7146/gn.v0i2.3552>.
- Greve, M.H., Christensen, O.F., Greve, M.B. og Kheir, R.B., 2014. Change in peat coverage in Danish cultivated soils during the past 35 years. Soil Sci. 179(5), 250-257. <http://dx.doi.org/10.1097/ss.0000000000000066>.
- Hamza, M.A. og Anderson, W.K., 2005. Soil compaction in cropping systems. Soil and Tillage Research 82(2), 121-145. <http://dx.doi.org/10.1016/j.still.2004.08.009>.

- Hansen, B., 1990. Landbrugets gødnings- og arealanvendelse i 1983 og 1989 - interviewundersøgelser i syv mindre landbrugsområder. NPo-forskning fra Miljøstyrelsen nr. A21. Miljøstyrelsen.
- Hansen, B., Olesen, S.E. og Ernstsén, V., 2004. Naturlig dræning og grundvandsdannelse, I: Ernstsén, V. (Ed.) Afprøvning af undersøgelsesmetoder med henblik på etablering af et zoneringskoncept for danske lerjorde : Statusrapport. Geological Survey of Denmark and Greenland, Appendix 10.
- Hengl, T., de Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S. og Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLOS ONE 12(2), e0169748. <http://dx.doi.org/10.1371/journal.pone.0169748>.
- Henriksen, H.J., Højberg, A.L., Olsen, M., Seaby, L.P., van der Keur, P., Stisen, S., Trolborg, L., Sonnenborg, T.O. og Refsgaard, J.C., 2012. Klimaeffekter på hydrologi og grundvand - klimagrundvandskort. Aarhus University.
- Jakobsen, P.R., Hermansen, B. og Tougaard, L., 2015. Danmarks digitale jordartskort 1:25000 version 4.0. GEUS.
- Kelstrup, N. og Hansen, B., 1986. Arealanvendelse og geologi - nitrat i grundvand. Miljøprojekt nr. 73. Miljøstyrelsen.
- Kovacic, D.A., David, M.B., Gentry, L.E., Starks, K.M. og Cooke, R.A., 2000. Effectiveness of Constructed Wetlands in Reducing Nitrogen and Phosphorus Export from Agricultural Tile Drainage. J. Environ. Qual. 29(4), 1262. <http://dx.doi.org/10.2134/jeq2000.00472425002900040033x>.
- Kronvang, B., Kristiansen, S.M., Schelde, K. og Børgesen, C.D., 2013. Udredningsprojekt vedrørende dræns betydning for afvanding - og de naturlige og menneskeskabte faktorer som influerer på dræns virke som vand-afleder fra marker. Aarhus University. DCE.
- Landbrugstyrelsen. 2014. Markkort. <http://www.geodatainfo.dk/Portal/ShowMetadata.aspx?id=6e3bc77f-c193-4508-80d4-836e1668db91> (tilgået 29-05-2017).
- Liu, J., Pattey, E., Nolin, M.C., Miller, J.R. og Ka, O., 2008. Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. Geoderma 143(3-4), 261-272. <http://dx.doi.org/10.1016/j.geoderma.2007.11.011>.
- Madsen, H.B., 2010. Drænrørets indførelse og betydning i et landbrugs- og miljømæssigt perspektiv, I: Nielsen, M.A. (Ed.) Det fremmede som historisk drivkraft. Danmark efter 1742. Det Kongelige Danske Videnskabernes Selskab, 158-165.
- Madsen, H.B., Nørr, A.H. og Holst, K.A., 1992. The Danish soil classification. The Royal Danish Geographical Society, Copenhagen, Denmark.

Møller, A.B., Beucher, A., Iversen, B.V. og Greve, M.H., 2017. Prediction of soil drainage classes in Denmark by means of decision tree classification. *Geoderma*. <http://dx.doi.org/10.1016/j.geoderma.2017.10.015>.

Møller, A.B., Beucher, A., Iversen, B.V. og Greve, M.H., 2018. Predicting artificially drained areas by means of a selective model ensemble. *Geoderma* 320, 30-42. <http://dx.doi.org/10.1016/j.geoderma.2018.01.018>.

NASA Landsat Program, 2014. Landsat OLI/TIRS scenes LC81920222014090LGN00, LC81940212014072LGN00, LC81940222014072LGN00, LC81950212014079LGN00, LC81950222014079LGN00, LC81960202014070LGN00, LC81960212014070LGN00, LC81960222014070LGN00, LC81980202014068LGN00, LC81980212014068LGN00, L1T. USGS, Sioux Falls.

Naz, B.S., Ale, S. og Bowling, L.C., 2009. Detecting subsurface drainage systems and estimating drain spacing in intensively managed agricultural landscapes. *Agric. Water Manag.* 96(4), 627-637. <http://dx.doi.org/10.1016/j.agwat.2008.10.002>.

Naz, B.S. og Bowling, L.C., 2008. Automated identification of tile lines from remotely sensed data. *Trans. ASABE* 51(6), 1937-1950. <http://dx.doi.org/10.13031/2013.25399>.

Nielsen, J.A., 2015. Dansk Markdræningsguide. SEGES.

Northcott, W.J., Verma, A.K. og Cooke, R.A., 2000. Mapping subsurface drainage systems using remote sensing and GIS. ASABE Paper No. 002113. ASABE, St. Joseph, Michigan.

Olesen, S.E., 2009. Kortlægning af potentielt dræningsbehov på landbrugsarealer opdelt efter landskabelement, geologi, jordklasse, geologisk region samt høj/lavbund. Aarhus University.

Orbicon. 2015. Hele Danmarks drænarkiv. <http://www.orbicon.dk/Dr%C3%A6narkiv.3302.aspx> (tilgået September 28-2015).

Poggio, L. og Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation - an example from Scotland. *Geoderma* 232-234, 284-299. <http://dx.doi.org/10.1016/j.geoderma.2014.05.004>.

Porsmose, E., 2008. Danske landsbyer. Gyldendal. ISBN 8702016907 9788702016901.

R Development Core Team, 2011. R: a language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.

Ren, B., Zhang, J., Li, X., Fan, X., Dong, S., Liu, P. og Zhao, B., 2014. Effects of waterlogging on the yield and growth of summer maize under field conditions. *Can. J. Plant Sci.* 94(1), 23-31. <http://dx.doi.org/10.4141/cjps2013-175>.

Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development* 3(3), 210-229. <http://dx.doi.org/10.1147/rd.33.0210>.

- Schelde, K., de Jonge, L.W., Kjærgaard, C., Lægdsmand, M. og Rubæk, G.H., 2006. Effects of manure application and plowing on transport of colloids and phosphorus to tile drains. *Vadose Zone J.* 5(1), 445. <http://dx.doi.org/10.2136/vzj2005.0051>.
- Sebastiani, F., 2002. Machine learning in automated text categorization. *ACM Comput. Surv.* 34(1), 1-47. <http://dx.doi.org/10.1145/505282.505283>.
- Seligman, M. 2017. Rborist: extensible, parallelizable implementation of the random forest algorithm. R package version 0.1-8. <https://cran.r-project.org/web/packages/Rborist/index.html>).
- Skriver, K. og Hedegård, J., 1973. Undersøgelser over danske jorders dræningstilstand, I: Olesen, J. (Ed.) *Oversigt over forsøg og undersøgelser i landbo- og husmandsforeningerne*. Andelsbogtrykkeriet i Odense, Odense, 2055-2059.
- Sørensen, R., Zinko, U. og Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrol. Earth Syst. Sci.* 10(1), 101-112. <http://dx.doi.org/10.5194/hess-10-101-2006>.
- Tetzlaff, B., Kuhr, P. og Wendland, F., 2009. A new method for creating maps of artificially drained areas in large river basins based on aerial photographs and geodata. *Irrig. Drain.* 58(5), 569-585. <http://dx.doi.org/10.1002/ird.426>.
- Verma, A.K., Cooke, R.A. og Wendte, L., 1996. Mapping subsurface drainage systems with color infrared aerial photographs. Urbana-Champaign: Department of Agricultural Engineering, University of Illinois.
- Wang, Y., Tetko, I.V., Hall, M.A., Frank, E., Facius, A., Mayer, K.F. og Mewes, H.W., 2005. Gene selection from microarray data for cancer classification--a machine learning approach. *Comput Biol Chem* 29(1), 37-46. <http://dx.doi.org/10.1016/j.compbiolchem.2004.11.001>.
- Watson, E.R., Lapins, P. og Barron, R.J.W., 1976. Effect of waterlogging on the growth, grain and straw yield of wheat, barley and oats. *Animal Prod. Sci.* 16(78), 114-122. <http://dx.doi.org/10.1071/EA9760114>.
- Wright, M.N. og Ziegler, A., 2015. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* 77(1). <http://dx.doi.org/10.18637/jss.v077.i01>.
- Wösten, J.H.M. og van Genuchten, M.T., 1988. Using Texture and Other Soil Properties to Predict the Unsaturated Soil Hydraulic Functions. *Soil Sci. Soc. Am. J.* 52(6), 1762. <http://dx.doi.org/10.2136/sssaj1988.03615995005200060045x>.
- Zhao, Z.Y., Ashraf, M.I. og Meng, F.-R., 2013. Model prediction of soil drainage classes over a large area using a limited number of field samples: a case study in the province of Nova Scotia, Canada. *Can. J. Soil Sci.* 93(1), 73-83. <http://dx.doi.org/10.4141/cjss2011-095>.

Ågren, A., Lidberg, W., Strömgren, M., Ogilvie, J. og Arp, P., 2014. Evaluating digital terrain indices for soil wetness mapping – a Swedish case study. *Hydrol. Earth Syst. Sci.* 18(9), 3623-3634.

Bilag 1: Geografiske datalag

Tabellen angiver geografiske datalag til kortlægning af drænede arealer. Søjlen til højre angiver middelværdien, minimum- og maksimumværdien for hvert datalag, eller antallet af klasser for kategoriske datalag.

Geografisk datalag	Gennemsnit (min - max) /Antal klasser
Topografiske variabler	
Nord-syd-komponent (cosinus til overfladens kompasretning)	0,01 (-1,00 - 1,00)
Øst-vest-komponent (sinus til overfladens kompasretning)	-0,03 (-1,00 - 1,00)
Dybde af afløbsløse lavninger (m)	0,1 (0,0 - 30,8)
Overfladens plankurvatur	0,0 (-5,1 - 5,3)
Overfladens profilkurvatur	0,0 (-7,3 - 6,1)
Højde minus gennemsnitshøjden i 4 km radius (m)	1,0 (-57,9 - 98,8)
Direkte solindstråling (kWh/år), beregnet ud fra breddegraden samt overfladens hældning og kompasretning	1269 (331 - 1691)
Højde over havet (m)	31,0 (-11,9 - 170,3)
Akkumuleret afstrømning (antal celler der fører ned til den pågældende celle)	63 (1 - 110908)
Dybde til grundvandet interpoleret fra boringer og overfladevand (m)	6,6 (0,0 - 111,6)
Dybde til grundvandet fra DK10-modellen.	5,5 (0,0 - 108,2)
Horisontal afstand til det nærmeste overfladevand (m)	235,6 (0,0 - 3238,4)
Mid-slope position; punktets position på en skråning	0,3 (0,0 - 1,0)
Multi-resolution index of valley bottom flatness; indeks for dalbundes fladhed	4,3 (0,0 - 10,8)
Vådhedsindeks fra SAGA GIS; modificeret version af TWI	14,5 (6,9 - 19,1)
Overfladens hældning (grader)	1,5 (0,0 - 60,4)
Hældning ned til det nærmeste overfladevand (grader)	1,1 (0,0 - 52,6)
Topografisk vådhedsindeks (TWI); Beregnet som $TWI = \ln(a/\tan b)$: hvor a er den akkumulerede afstrømning, og b er overfladens hældning	5,9 (0,0 - 54,6)
Dalddybde (m)	7,4 (0,0 - 89,1)
Vertikal afstand til det nærmeste overfladevand (m)	4,1 (0,0 - 96,6)
Satellitbilleder	
Ultrablå refleksivitet (Landsat 8 Bånd 1; Marts 2014)	369 (-399 - 15471)
Blå refleksivitet (Landsat 8 Bånd 2; Marts 2014)	438 (-407 - 15769)
Grøn refleksivitet (Landsat 8 Bånd 3; Marts 2014)	655 (-261 - 15843)
Rød refleksivitet (Landsat 8 Bånd 4; Marts 2014)	697 (-479 - 16000)
Nærinfrarød refleksivitet (Landsat 8; Marts 2014)	2295 (-77 - 15955)
Kortbølget infrarød refleksivitet I (Landsat 8 Bånd 6; Marts 2014)	1993 (-26 - 16000)
Kortbølget infrarød refleksivitet II (Landsat 8 Bånd 6; Marts 2014)	1334 (-5 - 17082)
Normaliseret fugtindeks; (Bånd 5 - Bånd 6)/(Bånd 5 + Bånd 6)	0,06 (-0,94 - 0,97)
Normaliseret vegetationsindeks; (Bånd 5 - Bånd 4)/(Bånd 5 + Bånd 4)	0,52 (-1,00 - 0,98)
Normaliseret vandindeks (Bånd 5 - Bånd 3)/(Bånd 5 + Bånd 3)	-0,55 (-0,97 - 1,00)
Jordtilpasset vegetationsindeks; (Bånd 5 - Bånd 4)*(1 + 0.5)/(Bånd 5 + Bånd 4 + 0.5)	0,3 (-0,2 - 0,7)

Jord og udgangsmateriale

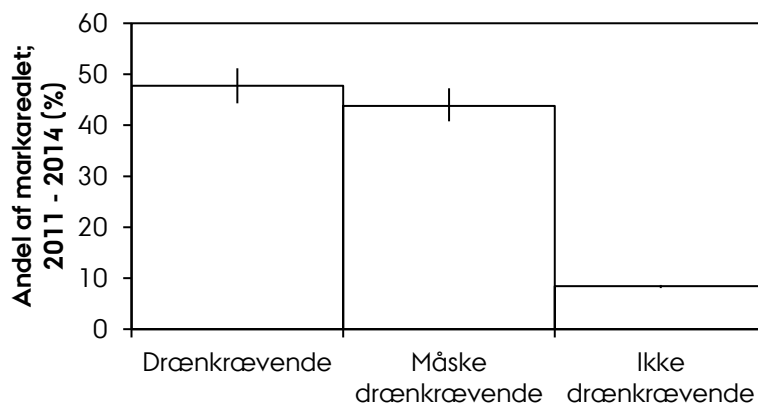
Lerindhold (%), 0 - 30 cm	8,3 (0,0 - 51,2)
Lerindhold (%), 30 - 60 cm	10,0 (0,0 - 62,7)
Lerindhold (%), 60 - 100 cm	11,3 (0,0 - 59,1)
Lerindhold (%), 100 - 200 cm	11,2 (0,0 - 57,1)
Jordens naturlige drænklasse	2,9 (1 - 5)
Jordartskort fra GEUS (Målestok 1:25.000)	10 klasser
Geografiske regioner (Målestok 1:100.000)	7 klasser
Landskabselementer (Målestok 1:100.000)	11 klasser
Viser udbredelsen af højbund (0), vådområder (1), centrale vådområder (2) og tørv (3) (Målestok 1:20.000)	0,3 (0 - 3)

Arealanvendelse, dyrkningshistorik og klima

År med afgrødedata i perioden 2011 - 2014	3,1 (0 - 4)
År med drænkrævende afgrøder i perioden 2011 - 2014	1,5 (0 - 4)
År med måske drænkrævende afgrøder i perioden 2011 - 2014	1,3 (0 - 4)
År med drænuafhængige afgrøder i perioden 2011 - 2014	0,2 (0 - 4)
CORINE arealanvendelse (Målestok 1:100,000)	3 klasser
Gennemsnitlig årlig nedbør i perioden 1961 - 1990 interpoleret fra punktdata (mm)	718 (473 - 963)

Bilag 2: Drænkategorier for afgrøder

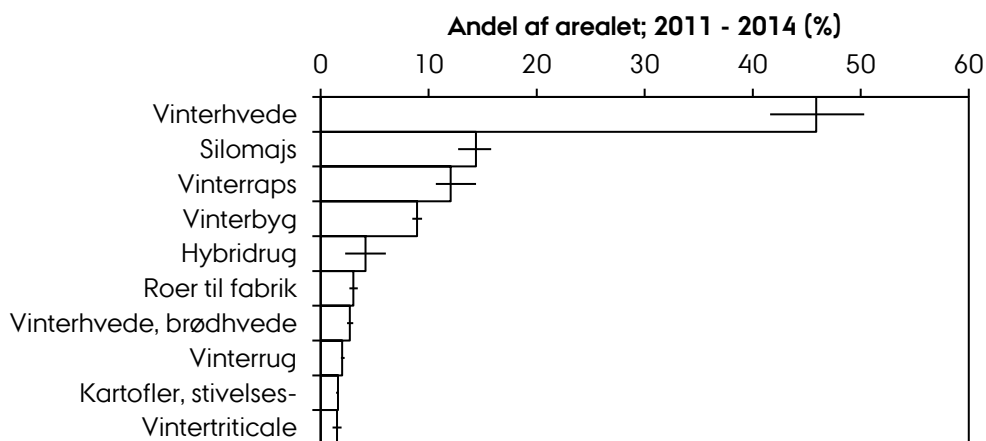
Drænkævende og måske drænkævende afgrøder udgjorde hver især over 40% af markarealet i perioden 2011 - 2014, mens ikke drænkævende afgrøder udgjorde mindre end 10% af arealet (Figur 27).



Figur 27: Andel af markarealet udgjort af henholdsvis drænkævende afgrøder, måske drænkævende afgrøder og ikke drænkævende afgrøder i perioden 2011 - 2014. Søjlerne angiver middelværdierne, mens barerne angiver minimum og maksimum i løbet af perioden.

De drænkævende afgrøder omfattede især vintersæd, majs og roer (Figur 28). De dyrkes fortrinsvist i Østdanmark og især i Østjylland (Figur 31).

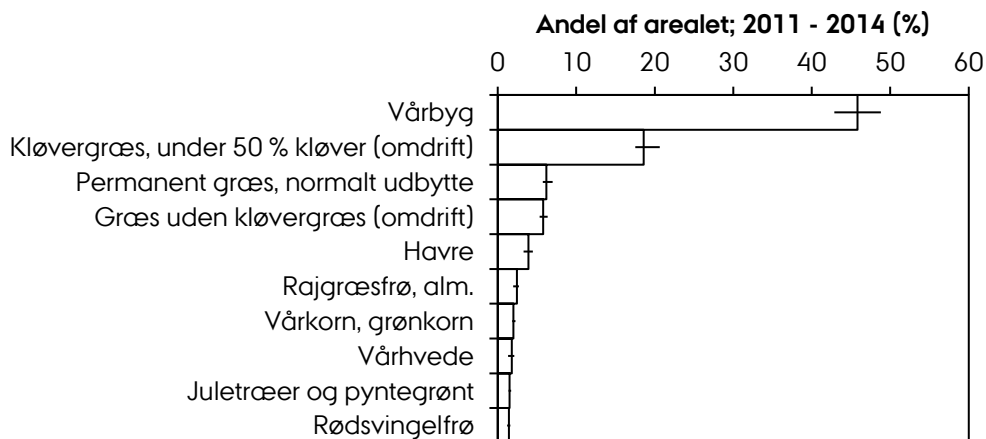
Drænkævende afgrøder



Figur 28: De 10 mest udbredte drænkævende afgrøder. X-aksen angiver deres andel af det samlede areal med drænkævende afgrøder. Søjlerne viser middelværdien for årene 2011 - 2014, mens barerne viser minimum- og maksimumværdierne i løbet af perioden.

De måske drænkrevende afgrøder udgjordes især af vårbyg og kløvergræs, samt en blanding af græsarealer af forskellig art, havre og andre afgrøder (Figur 29). De dyrkes fortrinsvist i Vestjylland, Himmerland og Nordjylland (Figur 32).

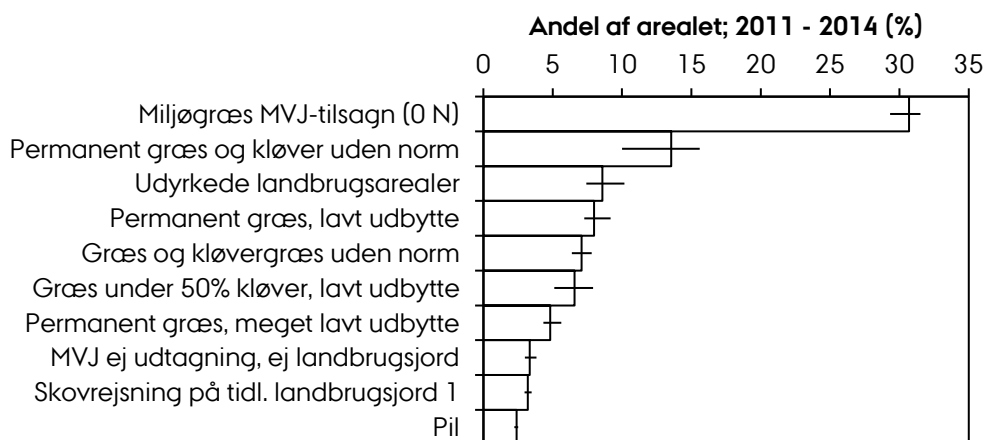
Måske drænkrevende afgrøder



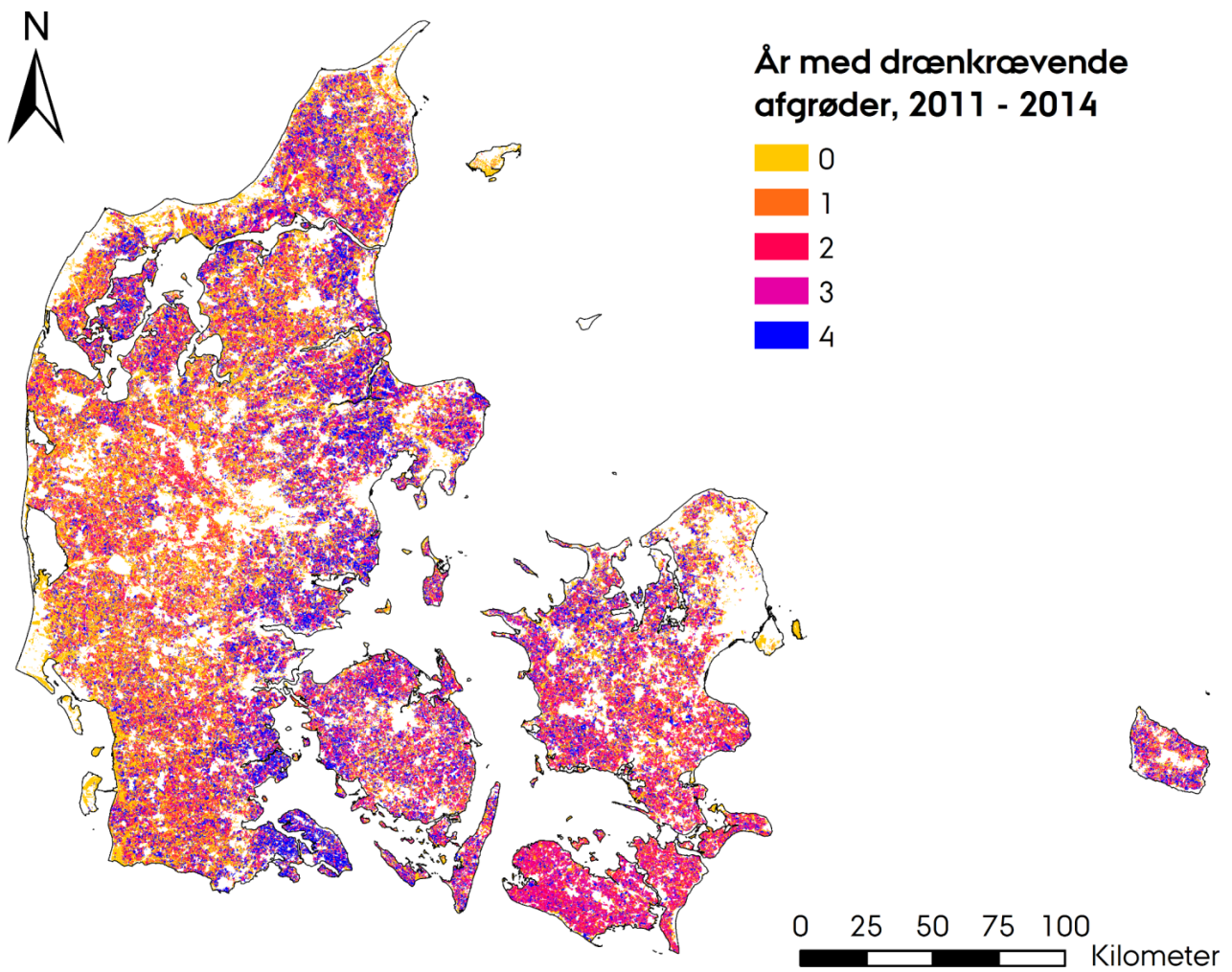
Figur 29: De 10 mest udbredte måske drænkrevende afgrøder. X-aksen angiver deres andel af det samlede areal med måske drænkrevende afgrøder. Søjlerne viser middelværdien for årene 2011 - 2014, mens bærerne viser minimum- og maksimumværdierne i løbet af perioden.

De ikke drænkrevende afgrøder udgjordes især af græsarealer af forskellig art, samt udyrkede arealer, skovrejsning og pil (Figur 30). De dyrkes fortrinsvist på lavbundslande, for eksempel i Ådale (Figur 33).

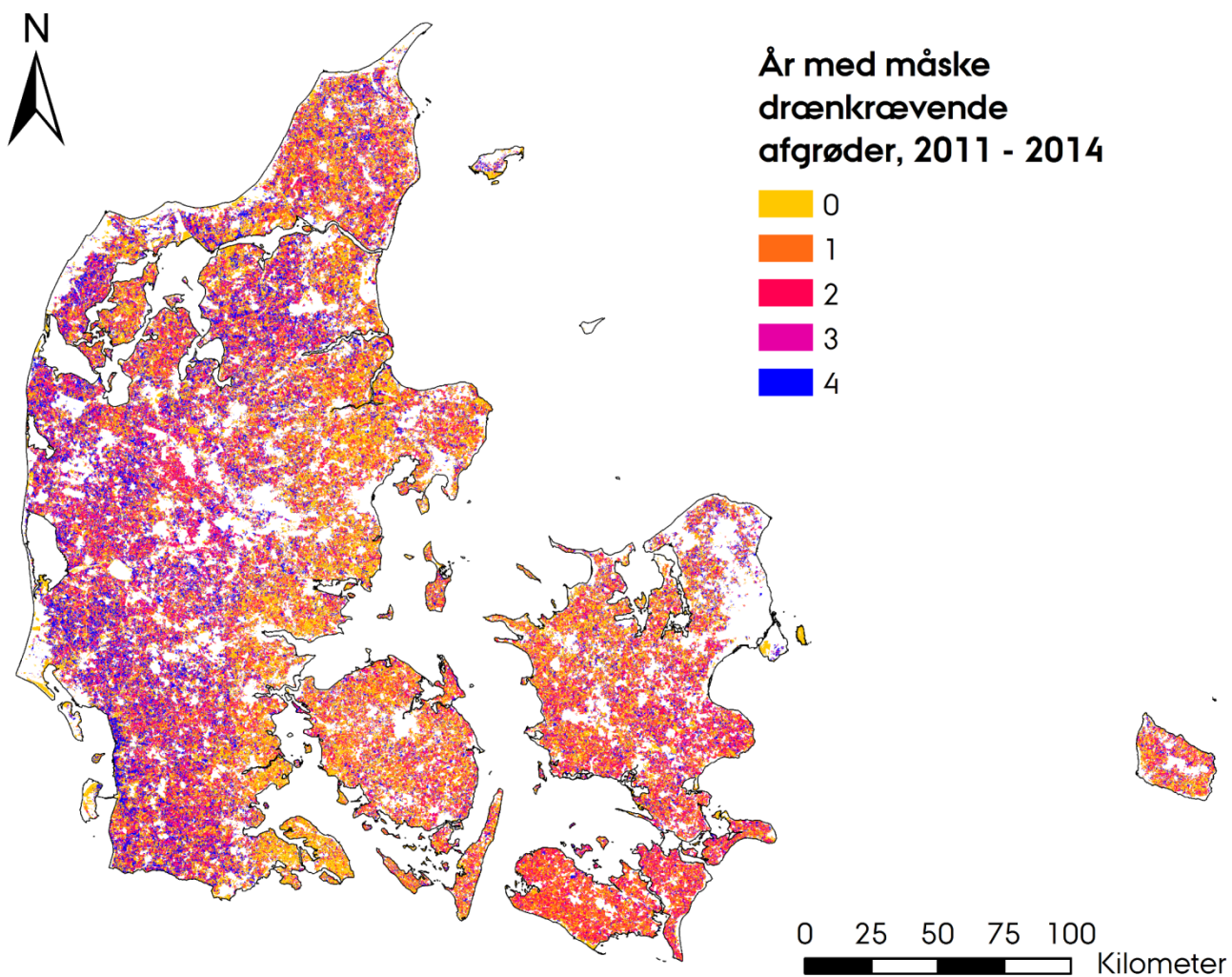
Ikke drænkkrævende afgrøder



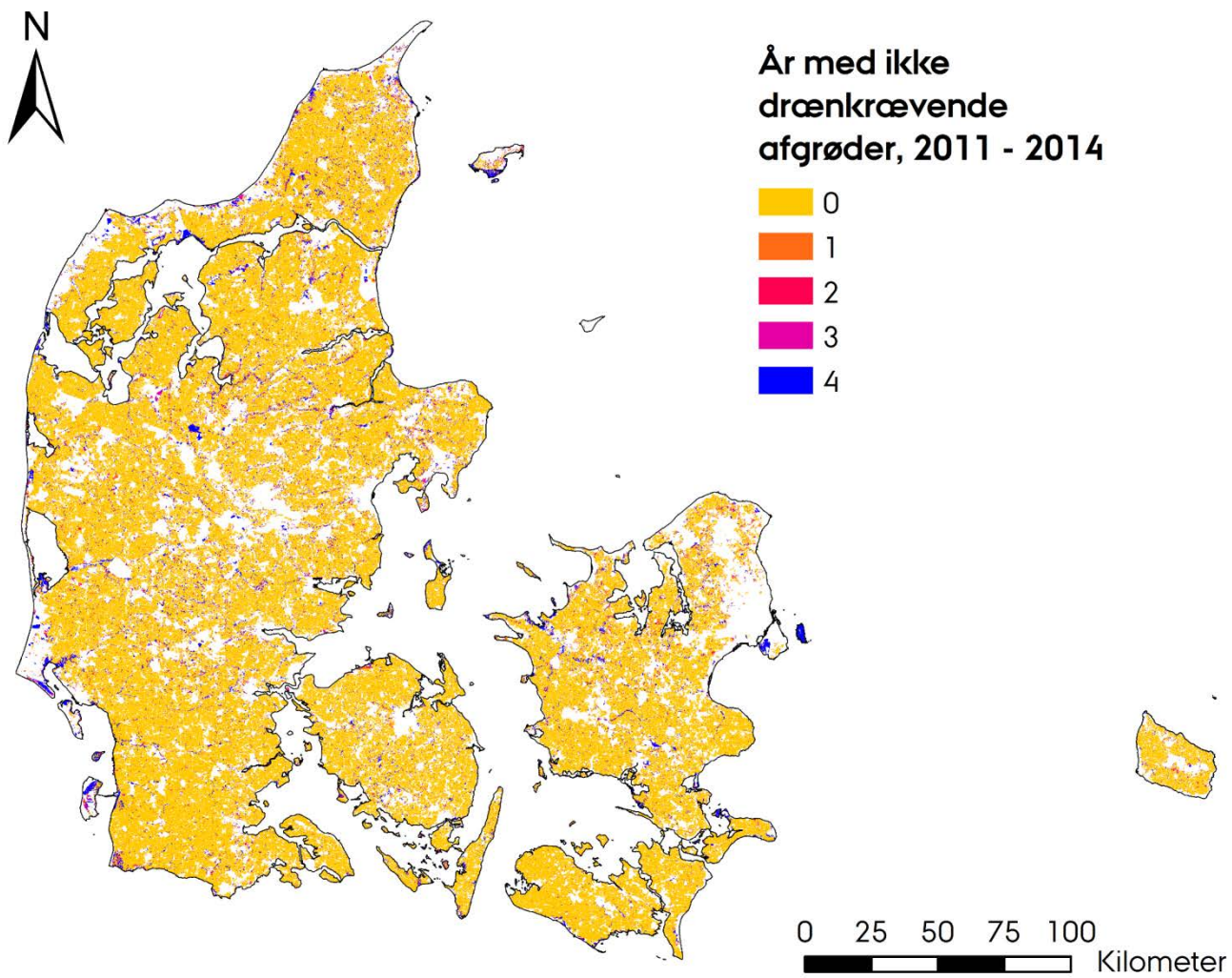
Figur 30: De 10 mest udbredte ikke drænkkrævende afgrøder. X-aksen angiver deres andel af det samlede areal med ikke drænkkrævende afgrøder. Søjlerne viser middelværdien for årene 2011 - 2014, mens barrierne viser minimum- og maksimumværdierne i løbet af perioden.



Figur 31: År med drænkkrævende afgrøder i årene fra 2011 til 2014. Baseret på IMK (Landbrugsstyrelsen, 2014).



Figur 32: År med måske drænkrevende afgrøder i årene fra 2011 til 2014. Baseret på IMK (Landbrugstyrelsen, 2014).



Figur 33: År med ikke drænkkrævende afgrøder i årene fra 2011 til 2014. Baseret på IMK (Landbrugsstyrelsen, 2014).

Bilag 3: Begreber

Bagging

Bagging er en sammentrækning af ordene *bootstrap aggregation*. Metoden anvendes indenfor maskinlæring til at generere *ensembles* af modeller ved at udtage en række *bootstrap samples* fra et datasæt og opbygge en model ud fra hvert sample. Hvis den variabel, der skal forudsiges er kategorisk, kombineres forudsigelserne ved simpel afstemning. Alternativt kan de enkelte modeller forudsige sandsynligheden for hver kategori, hvorefter gennemsnittet af sandsynlighederne beregnes på tværs af modellerne. Ensembles opbygget ved bagging har i mange tilfælde vist sig at kunne forudsige udfald med langt større nøjagtighed end enkeltstående modeller (Breiman, 1996, Bauer og Kohavi, 1999, Banfield et al., 2007).

Beslutningstræer

Beslutningstræer er modeller opbygget ved opdelinger af et datasæt. Ved hver opdeling vil algoritmen søge at finde den bedst mulige opdeling af den variabel, der skal forudsiges, ud fra de forklarende variabler, der er til rådighed. Det kunne eksempelvis tænkes, at lokaliteter under en bestemt kote fortrinsvis er drænede, mens lokaliteter over denne kote sjældnere er drænede. Algoritmen vil herefter søge at opdele de resulterende deldatasæt yderligere og så fremdeles indtil hver forgrening fører til en ensartet klassifikation.

Beslutningstræer bliver ofte beskåret for at fjerne overflødige forgreninger, der ikke bidrager til modellens evne til at forudsige et udfald. Beslutningstræer bliver også ofte kombineret i *ensembles* ved brug af *bagging* eller *boosting* (Dietterich, 2000, Banfield et al., 2007).

Boosting

Boosting er en metode til opbygning af *ensembles*. Den fungerer iterativt, ved at der først opbygges en model og de observationer, som modellen forudsiger forkert, tildeles en højere vægt. Der opbygges herefter en ny model, hvor vægtene indgår, og de observationer, der bliver forkert forudsagt bliver igen tildelt en højere vægt. Proceduren fortsætter på denne måde over et specificeret antal gentagelser, hvorefter modellernes forudsigelser kombineres ved en vægtet afstemning, alt efter modellernes nøjagtighed. På denne måde fokuserer boosting på observationer, som er svære at klassificere. *Ensembles* genereret ved boosting har ofte en langt højere nøjagtighed end enkeltstående modeller (Freund og Schapire, 1996, Bauer og Kohavi, 1999, Dietterich, 2000).

Bootstrap sampling

Et bootstrap sample er en stikprøve af et datasæt udtaget med afløsning. Det vil sige, at en given observation i det oprindelige datasæt stadig kan udtages til stikprøven, selv om den allerede er udtaget. På den måde kan de enkelte observationer i det oprindelige datasæt forekomme i stikprøven én gang, flere gange eller slet ikke. I de fleste tilfælde udtages der det samme antal observationer som der findes i det oprindelige

datasæt, men fordi den samme observation kan udtages flere gange vil et bootstrap sample i gennemsnit indeholde $1 - 1/e = 63,2\%$ af observationerne i det oprindelige datasæt.

Bootstrap sampling kan i mange tilfælde bruges til at generere kunstige stikprøver ud fra et enkelt datasæt. Teknikken kan blandt andet anvendes til at estimere konfidensintervaller i en population ved at generere flere bootstrap samples. Bootstrap sampling anvendes også i *bagging*.

Diskriminantanalyse

Diskriminantanalyse er en metode, der finder en lineær kombination af et antal forklarende variabler, der adskiller klasserne i en afhængig variabel. I diskriminantanalyse behandles de forklarende variabler, ligesom i Principal Component Analysis, som et multidimensionalt rum. Diskriminanten er en lineær vektor, der bedst muligt adskiller klasserne i den variabel, som skal forudsiges.

Ensemble

Et ensemble er en kombination af to eller flere modeller, hvis forudsigelser kombineres. Boosting og bagging er populære metoder til at generere ensembles, men kan kun generere ensembles bestående af én type modeller. Metoden i den nærværende undersøgelse genererer et ensemble bestående af modeller af forskellige typer.

k-nærmeste naboer

I k-nærmeste naboer-analyse (*knn*) behandles de forklarende variabler som et multidimensionalt rum. En model opbygget ved *knn* forudsiger en kategorisk variabel for et punkt i dette rum ved at finde de nærmeste observationer og udregne dets tilhørsforhold ud fra disse observationer. Parameteren *k* er et heltal der angiver, hvor mange observationer, der anvendes til at udregne tilhørsforholdet.

Kunstige neurale netværk

Kunstige neurale netværk (*ANNs*) er modeller bestående af forbundne knudepunkter arrangeret i lag. Kunstige neuralt netværk indeholder i de fleste tilfælde mindst tre lag: Et inputlag, ét eller flere skjulte lag og et outputlag. I inputlaget udgør hver variabel et knudepunkt, der sender et signaler videre til knudepunkterne i det skjulte lag. Knudepunkterne i det skjulte lag modificerer de inputs, som de modtager og sender signaler videre til det næste skjulte lag, hvis der er flere, eller til outputlaget. I outputlaget udgøres knudepunkterne af sandsynligheden for hver af de klasser, som man ønsker at forudsige.

Lineære modeller

Lineære modeller beskriver det lineære forhold imellem en afhængig variabel og én eller flere forklarende variabler. Brugen af lineære modeller kan udvides til kategoriske afhængige variabler ved, at sandsynligheden for et udfald, *p*, omskrives til en lineær variabel med funktionen *logit*.

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad (2)$$

Rasterfil

En rasterfil er en geografisk fladedækkende fil, der består af celler arrangeret i et rektangulært net på samme måde som et digitalt billede består af pixels. Hver celle har en værdi, der repræsenterer den gennemsnitlige værdi inden for det rektangulære område, som cellen dækker. Et eksempel er en digital højdemodel. Hvis højdemodellen har en opløsning på 10 meter, vil hver celle angive den gennemsnitlige værdi indenfor et område på 10 x 10 meter.

Support Vector Machines

Support Vectors Machines (*SVMs*) behandler de uafhængige som et multidimensionalt rum. *SVMs* forsøger at placere en flade i rummet, der adskiller klasserne med så stort et mellemrum mellem observationerne fra hver klasse som muligt. Hvis de ikke kan placere en flade uden, at der er observationer, der overlapper, genereres der nye forklarende variabler ved at kombinere de eksisterende forklarende variabler eller opløfte dem i potenser. Nye forklarende variabler kunne eksempelvis være *lerindhold*kote* eller *kote²*.

Bilag 4: Test af metoden til kombination af modeller

Opbygning af modeller

Modellerne til kortlægningen af drænedede arealer blev opbygget med 77 maskinlæringsmetoder implementeret i programmeringssproget R (R Development Core Team, 2011) (Tabel 11). Metoderne repræsenterede hver især enten forskellige algoritmer eller forskellige implementeringer af den samme algoritme. For eksempel er metoderne *ranger* (Wright og Ziegler, 2015), *Rborist* (Seligman, 2017) og *rf* (Breiman et al., 2015) forskellige implementeringer af beslutningstræ-algoritmen *Random Forest* (Breiman, 2001).

De fleste af metoderne var implementeringer af beslutningstræer (n = 29), diskriminantanalyse (n = 16), regressionsanalyse (n = 15), kunstige neurale netværk (n = 7), eller Support Vector Machines (n = 6). De sidste metoder var Nearest Neighbor-analyser (n = 2), Naïve Bayes (n = 1) og Nearest Centroid (n = 1).

Tabel 11: Tabellen angiver hver metodes navn og generelle type. Den indeholder desuden en kort beskrivelse af hver metode. Modeltypene er yderligere forklaret i Bilag 3.

Metode	Type	Beskrivelse
ada	Beslutningstræ	Boostede klassifikationstræer
adaboost	Beslutningstræ	AdaBoost-klassifikationstræer
AdaBoost.M1	Beslutningstræ	AdaBoost.M1-klassifikationstræer
avNNet	Kunstigt neuralt netværk	Modelmidlet neuralt netværk
bagEarthGCV	Regressionsanalyse	Baggede multivariable tilpassede regression splines med generaliseret krydsvalideringspruning
bagFDA	Diskriminantanalyse	Bagget fleksibel diskriminantanalyse
bagFDAGCV	Diskriminantanalyse	Bagget fleksibel diskriminantanalyse med generaliseret krydsvalideringspruning
bayesglm	Regressionsanalyse	Bayesisk generaliseret lineær model
bdk	Kunstigt neuralt netværk	Selvorganiserende kort
blackboost	Beslutningstræ	Boostede beslutningstræer
C5.0	Beslutningstræ	C5.0-beslutningstræer og -regelsæt
C5.0Rules	Beslutningstræ	Enkelt C5.0-regelsæt
C5.0Tree	Beslutningstræ	Enkelt C5.0-beslutningstræ
cforest	Beslutningstræ	Random Forest (beslutningstræer) med betinget inferens
ctree	Beslutningstræ	Beslutningstræer med betinget inferens
ctree2	Beslutningstræ	Beslutningstræer med betinget inferens
dnn	Kunstigt neuralt netværk	Dybt neuralt netværk med Stacked AutoEncoder
dwdLinear	Diskriminantanalyse	Afstandsvægtet lineær diskriminantanalyse
dwdRadial	Diskriminantanalyse	Afstandsvægtet diskriminantanalyse med radial basisfunktionskerne
earth	Regressionsanalyse	Multivariable tilpassede regression splines
evtree	Beslutningstræ	Beslutningstræer med genetiske algoritmer
fda	Diskriminantanalyse	Fleksibel diskriminantanalyse
gbm	Regressionsanalyse	Stokastisk gradient boosting

gcvEarth	Regressionsanalyse	Multivariable tilpassede regression splines
glm	Regressionsanalyse	Generaliseret lineær model
glmboost	Regressionsanalyse	Boostet generaliseret lineær model
glmStepAIC	Regressionsanalyse	Generaliseret lineær model med trinvis komponentudvælgelse
hdda	Diskriminantanalyse	Højdimensionel diskriminantanalyse
hdrda	Diskriminantanalyse	Højdimensionel regelmæssig diskriminantanalyse
kernelpls	Regressionsanalyse	Delvis mindste kvadraters regression
kknn	Nearest Neighbors	k-nearest neighbors
knn	Nearest Neighbors	k-nearest neighbors
lda	Diskriminantanalyse	Lineær diskriminantanalyse
lda2	Diskriminantanalyse	Lineær diskriminantanalyse
LogitBoost	Regressionsanalyse	Boostet logistisk regression
mlpSGD	Kunstigt neuralt netværk	Regressionsmodel baseret på neuralt netværk med stokastisk gradient decent
nb	Naive Bayes	Naive Bayes analyse
nnet	Kunstigt neuralt netværk	Kunstigt neuralt netværk
nodeHarvest	Beslutningstræ	Ensemble af beslutningstræer
ORFlog	Beslutningstræ	Skrå Random Forest (beslutningstræer) med logistisk regression
ORFpls	Beslutningstræ	Skrå Random Forest (beslutningstræer) med delvis mindste kvadraters metode
ORFridge	Beslutningstræ	Skrå Random Forest (beslutningstræer) med hurtig ridge regression med single value decomposition
ORFsvm	Beslutningstræ	Skrå Random Forest (beslutningstræer) med lineære Support Vector Machines
pam	Nearest Centroid	Nearest shrunken centroids
parRF	Beslutningstræ	Parallel Random Forest (beslutningstræer)
pcaNNet	Kunstigt neuralt netværk	Neuralt netværk med komponentudtræk
pda2	Diskriminantanalyse	Diskriminantanalyse med indbyggede omkostninger
pls	Regressionsanalyse	Delvis mindste kvadraters regression
ranger	Beslutningstræ	Random Forest (beslutningstræer)
Rborist	Beslutningstræ	Effektiv implementering af Random Forest (beslutningstræer)
rf	Beslutningstræ	Random Forest (beslutningstræer)
rlda	Diskriminantanalyse	Regelmæssig lineær diskriminantanalyse
rotationForest	Beslutningstræ	Rotation Forest (beslutningstræer)
rotationForestCp	Beslutningstræ	Rotation Forest (beslutningstræer)
rpart	Beslutningstræ	Beslutningstræer til klassifikation og regression
rpart1SE	Beslutningstræ	Beslutningstræer til klassifikation og regression
rpart2	Beslutningstræ	Beslutningstræer til klassifikation og regression
RRF	Beslutningstræ	Regelmæssig Random Forest (beslutningstræer)
RRFglobal	Beslutningstræ	Regelmæssig Random Forest (beslutningstræer)
sda	Diskriminantanalyse	Diskriminantanalyse med svind
simpls	Regressionsanalyse	Delvis mindste kvadraters regression
slda	Diskriminantanalyse	Stabil lineær diskriminantanalyse
sparseLDA	Diskriminantanalyse	Sparsom lineær diskriminantanalyse

spls	Regressionsanalyse	Sparsom mindste kvadraters regression
stepLDA	Diskriminantanalyse	Lineær diskriminantanalyse med trinvis komponentudvælgelse
stepQDA	Diskriminantanalyse	Kvadratisk diskriminantanalyse med trinvis komponentudvælgelse
svmLinear	Support Vector Machine	Support Vector Machines med lineær kerne
svmLinear2	Support Vector Machine	Support Vector Machines med lineær kerne
svmPoly	Support Vector Machine	Support Vector Machines med polynomisk kerne
svmRadial	Support Vector Machine	Support Vector Machines med radial basisfunktionskerne
svmRadialCost	Support Vector Machine	Support Vector Machines med radial basisfunktionskerne
svmRadialSigma	Support Vector Machine	Support Vector Machines med radial basisfunktionskerne
treebag	Beslutningstræ	Baggede beslutningstræer til klassifikation og regression
widekernelpls	Regressionsanalyse	Delvis mindste kvadraters regression
xgbLinear	Regressionsanalyse	Lineær model med eXtreme Gradient Boosting
xgbTree	Beslutningstræ	Beslutningstræmodel med eXtreme Gradient Boosting
xyf	Kunstigt neuralt netværk	Selvorganiserende kort

Da mange af de algoritmer, der blev anvendt, ikke kan håndtere kategoriske forklarende variabler, blev de kategoriske datalag konverteret til et antal binære numeriske datalag (bestående af tallene 1 og 0) lig med antallet af klasser. Dette øgede antallet af datalag til 73.

Derudover er nogle af algoritmerne sårbare overfor korrelerede eller overflødige forklarende variabler, mens andre bliver unøjagtige, hvis antallet af forklarende variabler er for stort. Der blev derfor, ud over de rå datalag, udregnet yderligere datalag gennem Principal Component Analysis (PCA).

I PCA transformeres variabler, der muligvis er korrelerede til et antal lineært ukorrelerede variabler kaldet principal components (PCs). De variabler som man vil transformere behandles som et multidimensionalt rum, og det første PC (PC1) er den vektor der forklarer den størst mulige andel af variationen i inputvariablerne. Den næste PC (PC2) er derefter den vektor, der forklarer den størst mulige grad af variationen i residualerne fra PC1, mens PC3 forklarer variationen i residualerne fra PC2 og så fremdeles. Antallet af PCs kan fastsættes efter hvor stor en del af variationen i de oprindelige variabler der skal forklares og er oftest væsentligt mindre end antallet af inputvariabler.

PCA blev udført på 100.000 tilfældigt udvalgte punkter fra de geografiske datalag. Analysen blev anvendt til at udarbejde nye kortlag med PCs samt nye datasæt til opbygning af modeller. Da det bedste antal forklarende variabler for hver metode ikke kendes på forhånd, blev der genereret tre nye datasæt med forskellige antal PCs. De nye datasæt indeholdt henholdsvis 10, 20 og 44 PCs, og forklarede henholdsvis 51,7%, 70,5% og 95,2% af variationen i de oprindelige datalag.

De 77 maskinlæringsmetoder blev herefter anvendt til at opbygge modeller ud fra datasættene med de rå geografiske datalag og 10, 20 og 44 PCs. Kombinationen af 77 maskinlæringsmetoder og fire datasæt betød, at der blev opbygget 308 modeller i alt.

Flere af de anvendte metoder havde parametre, der skulle tilpasses for at opnå den bedst mulige nøjagtighed. Parametre blev tilpasset for hver model ved at variere dem i tre trin enkeltvis og kombineret. De værdier, der gav den bedste nøjagtighed blev herefter anvendt i den endelige model. For at kunne vurdere nøjagtigheden uafhængigt af de data, der blev anvendt i opbygningen af modellerne, blev der anvendt krydsberegning. Datasættet til opbygning af modellerne blev inddelt i 10 dele, og hver af de 10 dele blev på skift taget fra mens en testmodel blev opbygget på baggrund af de resterende 9 dele. Testmodellen blev herefter anvendt til at forudsige sandsynlighederne for dræning for den del af datasættet, der var blevet taget fra. Den samlede nøjagtighed blev herefter beregnet på baggrund af de krydsberegnete forudsigelser for alle 10 dele.

For hver af de endelige modeller blev det i øvrigt målt, hvor længe det tog at generere et kort over et areal på 19 x 19 km.

Kombination af modeller

Der blev anvendt en teknik udviklet af Caruana et al. (2004) til at udvælge de modeller, der tilsammen gav den mest nøjagtige forudsigelse. Teknikken består i en trinvis algoritme, der udvælger modeller én ad gangen. Ved hvert trin prøver algoritmen et antal modeller igennem, for at teste, hvilken af dem, der giver den mest nøjagtige forudsigelse i kombination med de modeller, som den allerede har udvalgt. Den udvælger herefter den model, der resulterer i den mest nøjagtige forudsigelse.

Caruana et al. (2004) nåede frem til, at den bedste kombinerede forudsigelse blev opnået ved at tillade algoritmen at vælge den samme model flere gange samtidig med, at den blev forhindret i at vælge blandt alle modellerne i hvert trin. I stedet kunne algoritmen kun vælge i blandt et antal tilfældige modeller for hvert trin. Caruana et al. (2004) nåede frem til, at det var forskelligt, hvor mange modeller algoritmen skulle have at vælge imellem for at opnå den bedste forudsigelse, afhængigt af hvilket problem, der blev behandlet.

Caruana et al. (2004) anvendte tre datasæt: Ét til opbygning af modeller, et andet til at udvælge modeller og et tredje til en uafhængig vurdering af nøjagtigheden. Opdelingen skyldtes, at mange modeller baseret på maskinlæringsalgoritmer opnår en meget højere nøjagtighed på det datasæt som anvendes i opbygningen, end de vil kunne opnå på et uafhængigt datasæt. Hvis det samme datasæt blev anvendt til at opbygge og udvælge modellerne, ville det betyde, at algoritmen favoriserede modeller, der havde en høj nøjagtighed for det pågældende datasæt, men ikke nødvendigvis på nye data. Caruana et al. (2004) mente dog, at man ville kunne anvende det samme datasæt til opbygning og udvælgelse af modeller, hvis man krydsberegnete modellernes forudsigelser, som beskrevet i det forrige afsnit.

Denne undersøgelse anvendte derfor de krydsberegnete forudsigelser til at udvælge modeller. Algoritmen valgte først den af de tilgængelige modeller, der havde den bedste nøjagtighed, beregnet for de krydsberegnete forudsigelser. I næste trin kunne algoritmen vælge imellem et nyt tilfældigt udvalg af modeller. Hver af modellerne blev forsøgsvis udvalgt, og sandsynligheden for dræning blev midlet over den første model og den forsøgsvis udvalgte model for hver observation. Nøjagtigheden af de gennemsnitlige

sandsynligheder blev beregnet, og algoritmen udvalgte den model, der gav den bedste nøjagtighed. I næste trin blev processen gentaget, og så fremdeles. Nøjagtigheden blev beregnet som Kappa (K), der tager højde for tilfældige sammenfald:

$$K = 1 - \frac{1-p_o}{1-p_e} \quad (1)$$

p_o : observeret sammenfald mellem forudsigelser og observationer. p_e : hypotetisk sandsynlighed for sammenfald mellem forudsigelser og observationer.

I denne undersøgelse blev der i øvrigt afprøvet to mulige forbedringer af teknikken. For det første bliver det afprøvet, om nøjagtigheden bliver forbedret ved at nøjagtigheden i hvert udvælgelsestrin bliver vurderet på baggrund af et *bootstrap sample* (Bilag 3) af observationerne med krydsberegne forudsigelser, i stedet for at anvende alle observationerne hver gang. Der indføres derved en grad af tilfældighed i udvælgelsesprocessen, hvilket burde føre til, at de udvalgte modeller kan give nøjagtige forudsigelser i flere forskellige situationer.

For det andet blev det afprøvet, om den tid, som det tog at generere et kort med de udvalgte modeller, også kaldet *prædiktionstiden*, kunne forkortes ved at inddrage prædiktionstiden beregnet for hver model på et mindre område. Dette blev afprøvet ved, at nøjagtigheden i udvælgelsesprocessen blev divideret med den samlede prædiktionstid for de udvalgte modeller, som var blevet målt i området på 19x19 km. Dette burde favorisere modeller, der opnår en høj nøjagtighed men en kort prædiktionstid.

Der blev udført fire eksperimenter for at undersøge effekten af at anvende *bootstrap sampling* og af at inddrage prædiktionstiderne i udvælgelsesprocessen:

- Eksperiment 1: Ingen bootstrap sampling. Ingen inddragelse af prædiktionstider.
- Eksperiment 2: Ingen bootstrap sampling. Inddragelse af prædiktionstider.
- Eksperiment 3: Bootstrap sampling af hillclimb datasættet. Ingen inddragelse af prædiktionstider.
- Eksperiment 4: Bootstrap sampling af hillclimb datasættet. Inddragelse af prædiktionstider.

I hvert eksperiment blev det også varieret, hvor mange modeller algoritmen kunne vælge imellem i hvert udvælgelsestrin. Fraktionen blev varieret fra 5% til 100% af alle 308 modeller i intervaller af 5%. For hver værdi blev algoritmen afprøvet med 100 udvælgelsestrin 100 gange.

Hver gang algoritmen blev afprøvet, blev nøjagtigheden beregnet for to tidspunkter. For det første blev nøjagtigheden beregnet efter alle 100 udvælgelsestrin. For det andet blev nøjagtigheden beregnet ved det sidste udvælgelsestrin før den samlede prædiktionstid oversteg 300 sekunder. Herudover blev den samlede prædiktionstid efter 100 udvælgelsestrin beregnet, og det blev angivet hvor mange forskellige modeller, der blev udvalgt. Alle værdierne, både nøjagtighed, prædiktionstid og antal unikke modeller, blev midlet over de 100 afprøvninger.

Algoritmen blev derefter sat til at udvælge de endelige modeller med den teknik, der gav de bedste resultater. Disse modeller blev anvendt til at udarbejde et kort over drænede arealer for hele Danmark. Det endelige resultat blev beskåret, så det kun dækkede arealer, der var en del af markarealet i mindst ét af årene i perioden 2011 – 2014.

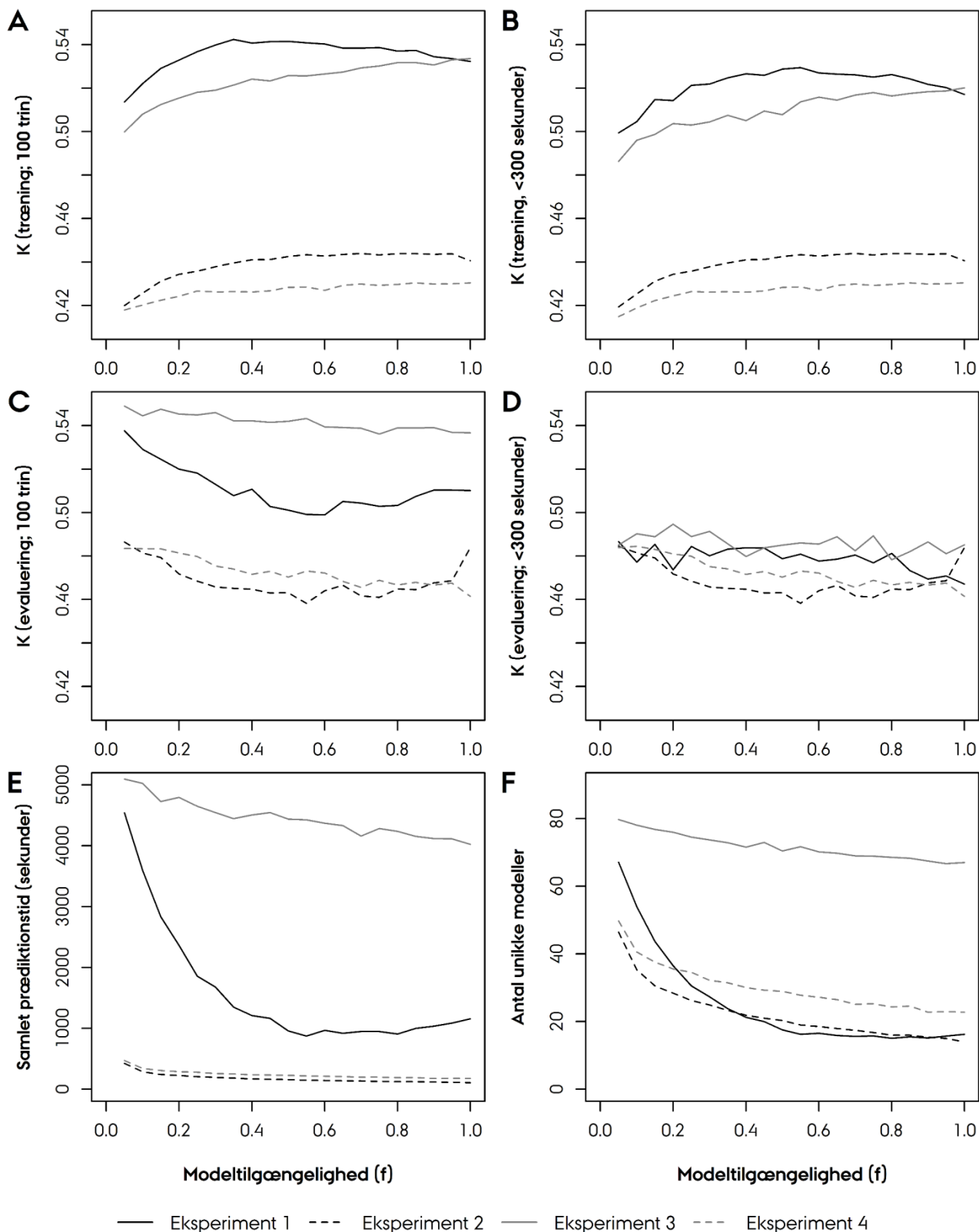
For de modeller der kunne beregne betydningen af de enkelte datalag, blev den relative betydning skaleret til 100 for det vigtigste datalag og et gennemsnit blev beregnet på tværs af de udvalgte modeller.

Resultater

Nøjagtigheden var generelt mindre, når der blev taget højde for prædiktionsiden, end når der ikke blev taget højde for prædiktionsiden (Figur 34). Dette gjaldt både når nøjagtigheden blev beregnet for de krydsberegne forudsigelser (Figur 34A og Figur 34B) og på det uafhængige datasæt (Figur 34C og Figur 34D). Beregnet for de krydsberegne forudsigelser var nøjagtigheden mindre, når der blev anvendt bootstrap sampling end når der ikke blev anvendt bootstrap sampling (Figur 34A og Figur 34B). Beregnet på det uafhængige datasæt var nøjagtigheden derimod større, når der blev anvendt bootstrap sampling. Dette gjaldt både for nøjagtigheden efter 100 udvælgelsestrin (Figur 34C) og med en maksimal prædiktionsid på 300 sekunder (Figur 34D).

Beregning af modellernes prædiktionsider gjorde, at den samlede prædiktionsid blev mindre (Figur 34E). Brug af bootstrap sampling gav generelt længere prædiktionsider og førte til, at der blev valgt flere unikke modeller (Figur 34F), hvilket vil sige, at der var mindre tilbøjelighed til, at algoritmen valgte den samme model flere gange. Brug af bootstrap sampling uden iberegning af prædiktionsider resulterede i det største antal unikke modeller.

Beregnet på det uafhængige datasæt blev den bedste nøjagtighed, før den samlede prædiktionsid oversteg 300 sekunder, opnået med bootstrap sampling, men uden iberegning af prædiktionsider (Figur 34D). Denne teknikken blev derfor anvendt til at udvælge de endelige modeller til kortlægningen. Den bedste nøjagtighed blev opnået ved, at algoritmen kunne vælge imellem 20% af de 308 modeller i hvert udvælgelsestrin, og denne værdi blev derfor anvendt ved udvælgelsen af de endelige modeller til kortlægningen.



Figur 34: Resultater fra de fire eksperimenter med kombination af modeller. X-aksen i hvert plot angiver hvor mange af modellerne i biblioteket, som algoritmen kunne vælge imellem ved hvert trin i udvælgelsen af modeller. For eksempel vil $f = 0,2$ sige, at algoritmen kunne vælge imellem et tilfældigt udpluk af modellerne

svarende til 20% af biblioteket. For hvert trin blev der anvendt et nyt udpluk. A: Gennemsnitlig nøjagtighed beregnet for de krydsberegne forudsigelser (træning) efter 100 trin. B: Gennemsnitlige nøjagtighed beregnet for de krydsberegne forudsigelser før prædiktionsiden oversteg 300 sekunder. C: Gennemsnitlig nøjagtighed beregnet på det uafhængige datasæt (evaluering) efter 100 trin. D: Gennemsnitlig nøjagtighed inden prædiktionsiden oversteg 300 sekunder. E: Gennemsnitlig samlet prædiktionsid efter 100 udvælgelsestrin. F: Gennemsnitligt antal unikke modeller efter 100 udvælgelsestrin.

Testen af teknikken til kombination af modeller viste, at det bedste resultater blev opnået med *bootstrap sampling* i udvælgelsesprocessen, men uden inddragelse af den tid, som modellerne brugte på at generere et kort. Resultaterne viste også, at den bedste nøjagtighed blev opnået, når algoritmen kunne vælge imellem 20% af modellerne i hvert udvælgelsestrin. De endelige modeller til kortlægningen blev derfor udvalgt med *bootstrap sampling* og uden inddragelse af tiden, som modellerne brugte på at generere et kort. Samtidig var 20% af modellerne tilgængelige i hvert udvælgelsestrin.

Da udvælgelsesprocessen ikke inddrog, hvor lang tid det tog for modellerne at generere et kort, var flere af modellerne meget længe om at generere kort over drænede arealer. Det blev derfor besluttet at øge den samlede tilladte tid til generering af kort fra 300 sekunder til 1000 sekunder. Dette førte til, at der blev udvalgt et større antal modeller, hvilket burde øge forudsigelsens robusthed.

Bilag 5: Nøjagtighed i udvalgte områder

Dette bilag præsenterer nøjagtigheden af kortene udarbejdet i undersøgelsens første og anden fase fordelt på geologiske klasser (Tabel 12), landskabselementer (Tabel 13) og georegioner (Tabel 14). Nøjagtigheden er baseret på det uafhængige datasæt. Hver tabel angiver antal observationer i hver klasse og p-værdien for en binomial test af nøjagtighederne. Figur 35, Figur 36 og Figur 37 viser kort over hvert datasæt.

Tabel 12: Nøjagtighed af kortene udarbejdet i undersøgelsens første og anden fase fordelt på geologiske klasser (Figur 35). Kolonnerne n og p angiver antallet af punkter til vurdering af nøjagtigheden og p-værdien for en binomial test af nøjagtighederne.

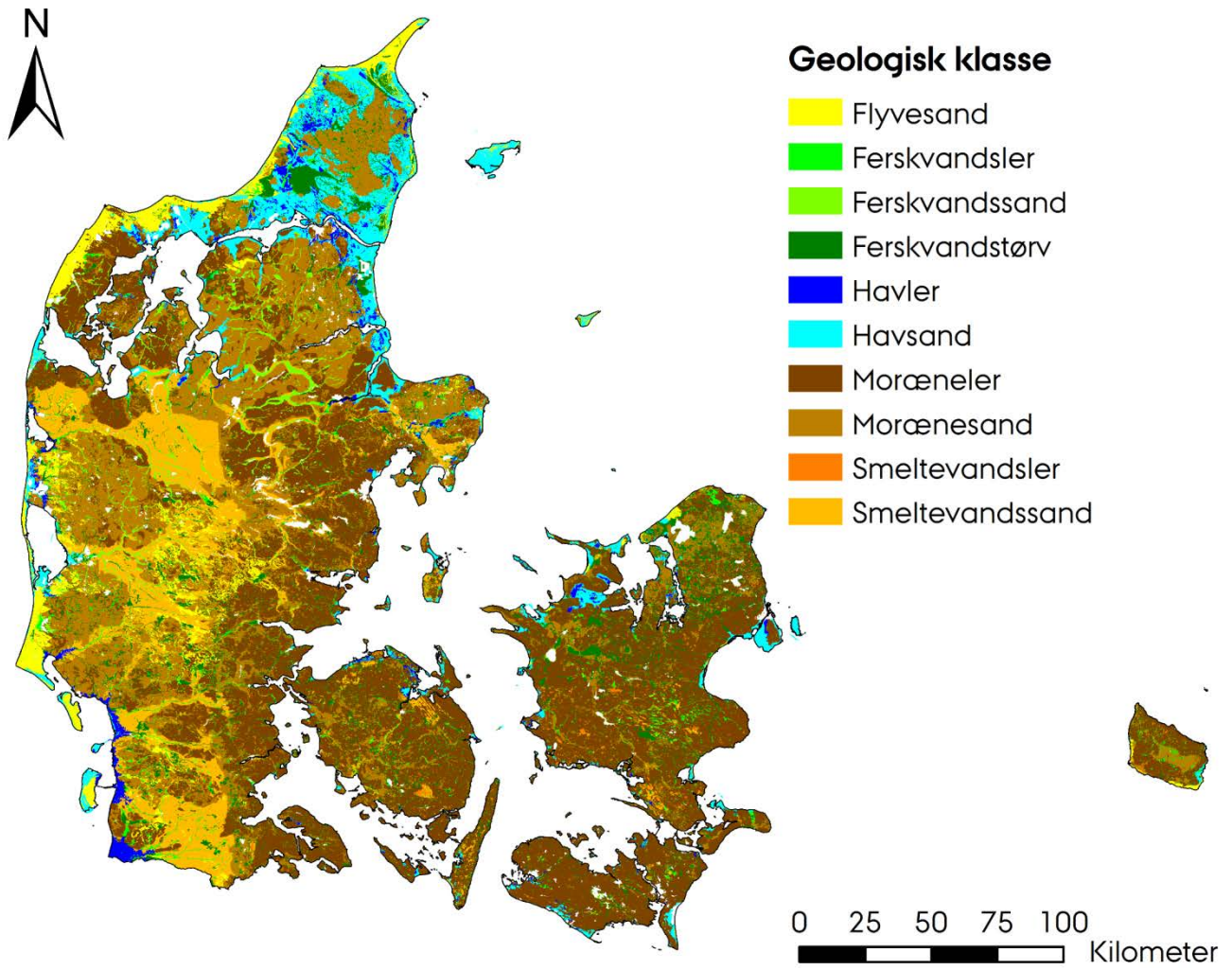
Geologisk klasse	Nøjagtighed (%)		n	p
	Første kort	Andet kort		
Flyvesand	71	71	7	0.319
Ferskvandsler	56	56	9	0.260
Ferskvandssand	88	75	8	0.196
Ferskvandstørv	67	100	3	0.296
Havler	100	100	3	1.000
Havsand	67	83	6	0.263
Moræneler	80	79	131	0.084
Morænesand	76	82	50	0.086
Smeltevandsler	50	50	2	0.500
Smeltevandssand	68	71	28	0.152

Tabel 13: Nøjagtighed af kortene udarbejdet i undersøgelsens første og anden fase fordelt på landskabselementer (Figur 36). Kolonnerne n og p angiver antallet af punkter til vurdering af nøjagtigheden og p-værdien for en binomial test af nøjagtighederne.

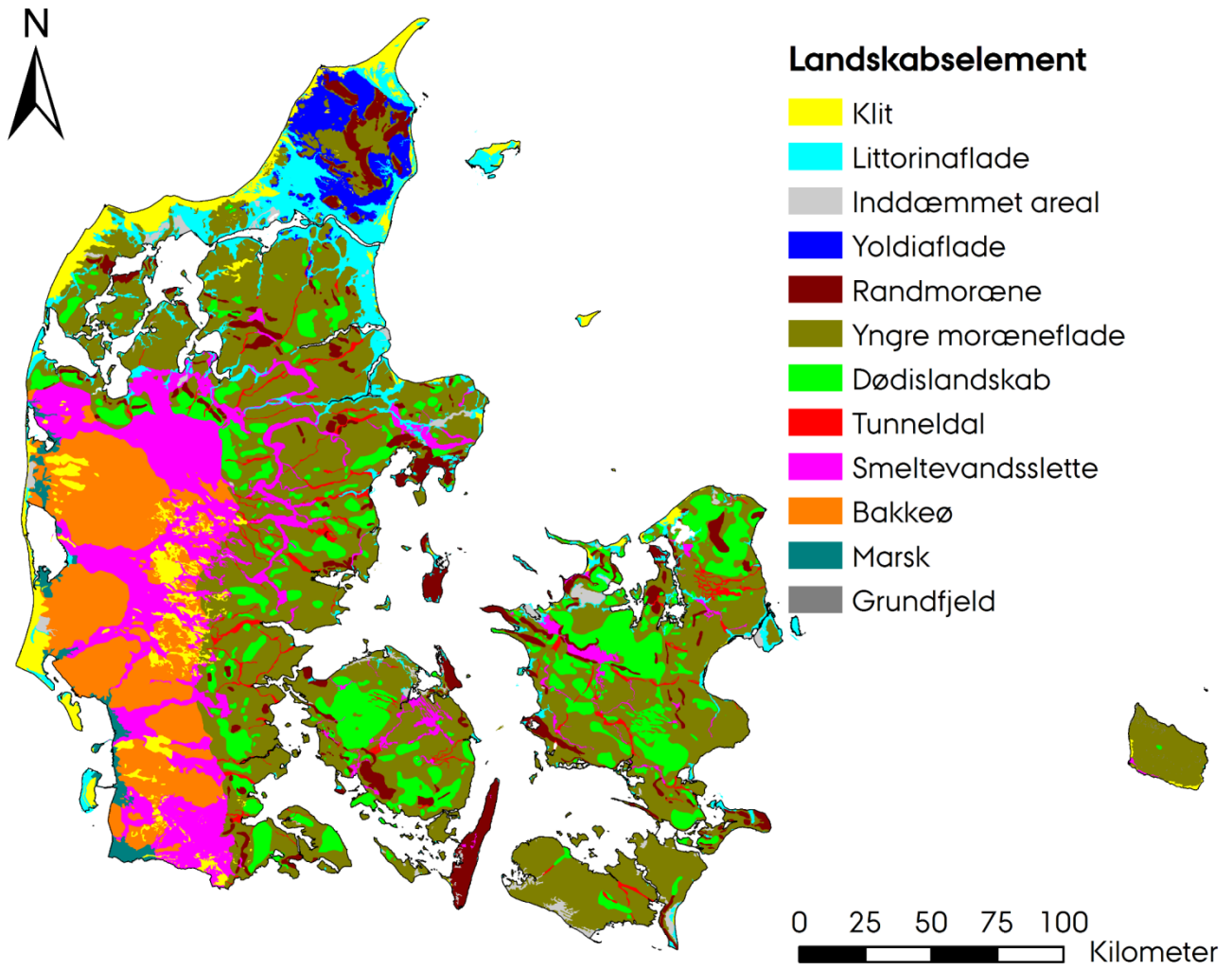
Landskabselement	Nøjagtighed (%)		n	p
	Første kort	Andet kort		
Klit	100	100	7	1.000
Littorinaflade	67	100	9	0.026
Inddæmmed	100	100	2	1.000
Yoldiaflade	100	75	4	0.000
Randmoræne	73	73	15	0.228
Nyere bundmoræne	81	79	119	0.080
Dødislandskab	66	72	32	0.117
Tunneldal	33	67	3	0.444
Smeltevandsslette	77	74	31	0.148
Bakkeø	67	75	24	0.125
Marsk	100	100	1	1.000

Tabel 14: Nøjagtighed af kortene udarbejdet i undersøgelsens første og anden fase fordelt på georegioner (Figur 37). Kolonnerne *n* og *p* angiver antallet af punkter til vurdering af nøjagtigheden og *p*-værdien for en binomial test af nøjagtighederne.

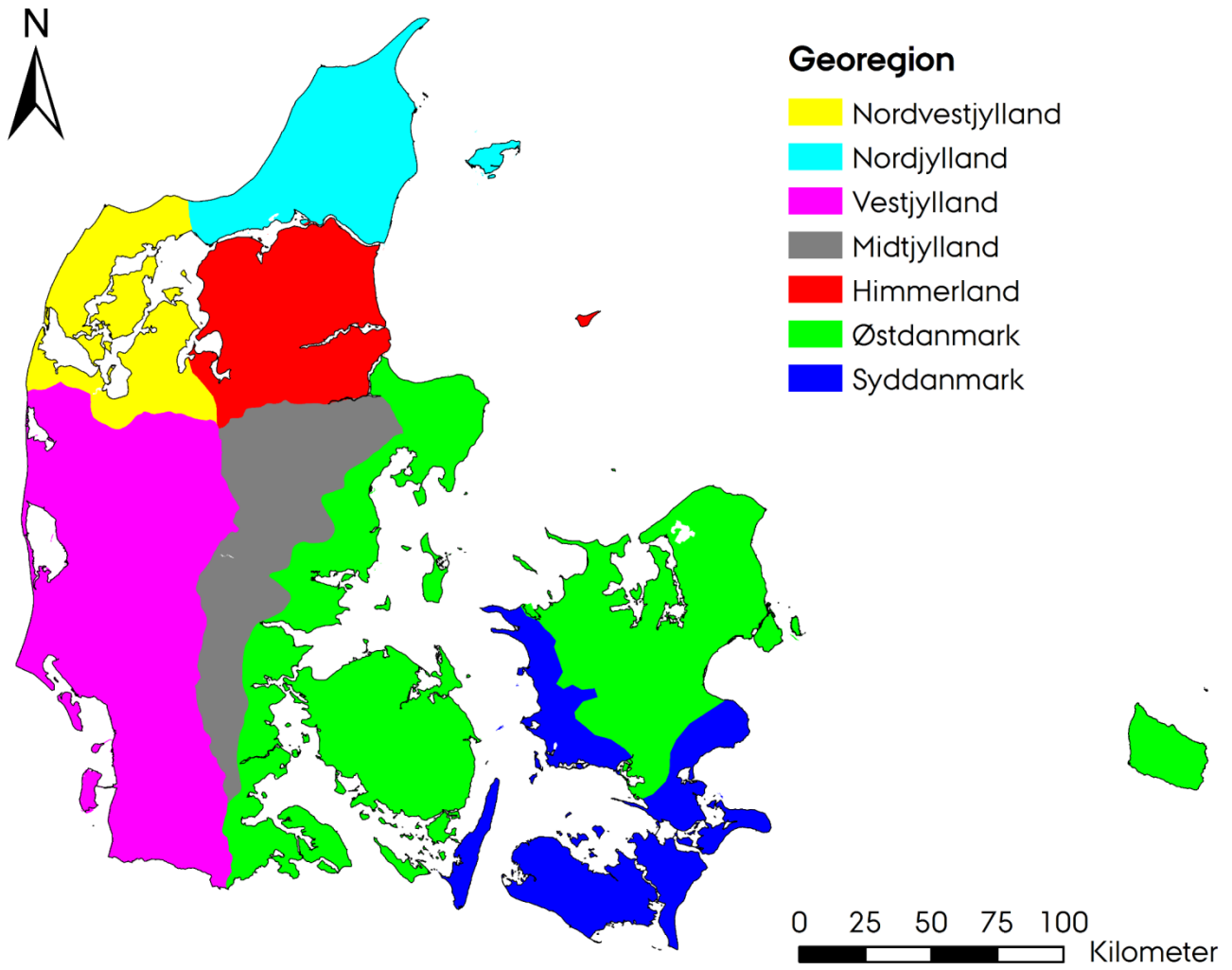
Georegion	Nøjagtighed (%)		n	p
	Første kort	Andet kort		
Nordøstjylland	81	76	21	0.174
Nordjylland	90	90	10	0.387
Vestjylland	75	77	52	0.124
Midtjylland	66	79	29	0.048
Himmerland	86	86	21	0.181
Østdanmark	72	72	86	0.092
Syddanmark	89	89	28	0.237



Figur 35: Kort over geologiske klasser. Forenklet version af GEUS' jordartskort (Jakobsen et al., 2015).



Figur 36: Kort over landskabselementer efter Madsen et al. (1992).



Figur 37: Kort over georegioner (Adhikari et al., 2013).

Bilag 6: Overensstemmelse mellem kort og registrerede drænrør

Tabellerne angiver overensstemmelsen mellem forudsigelsen af drænede arealer (søjler) og drænrør registreret i kabelgraven af Energinet (rækker). Overensstemmelsen er opgjort både for kortet udarbejdet af Olesen (2009) og i denne undersøgelse. Tallene angiver, hvor stor en procentdel af cellerne i rasterlaget langs hver strækning, i hver geologisk klasse eller i hver landskabselement, der har en given kombination af drænforudsigelser og registrerede drænrør. Den overordnede overensstemmelse er udregnet som *% celler uden drænrør forudsagt som ikke drænede + % celler med drænrør forudsagt som drænede*.

Tablet 15: Overensstemmelse mellem forudsigelser og registrerede drænrør. Opgjort for kabelstrækninger.

Strækning	Olesen (2009)		Denne undersøgelse	
	Ikke drænet	Drænet	Ikke drænet	Drænet
Bedsted-Struer				
Ingen drænrør	55	24	39	40
Drænrør	9	13	4	18
Blåbjerg-Holsted				
Ingen drænrør	64	12	73	3
Drænrør	14	10	20	4
Roslev-Bilstrup				
Ingen drænrør	27	34	23	38
Drænrør	17	22	10	30
Røjle-Fynsværket				
Ingen drænrør	4	56	7	55
Drænrør	3	37	1	38
Sønderborg-Fynshav				
Ingen drænrør	3	62	10	56
Drænrør	0	34	0	34

Tabel 16: Overensstemmelse mellem forudsigelser og registrerede drænrør. Opgjort for landskabselementer.

Landskabselement	Olesen (2009)		Denne undersøgelse	
	Ikke drænet	Drænet	Ikke drænet	Drænet
Littorinaflade				
Ingen drænrør	38	38	38	38
Drænrør	3	21	1	23
Randmoræne				
Ingen drænrør	1	47	12	36
Drænrør	0	52	0	52
Yngre moræneflade				
Ingen drænrør	28	41	22	46
Drænrør	7	24	3	29
Dødislandskab				
Ingen drænrør	21	51	13	60
Drænrør	5	23	1	27
Smeltevandsslette				
Ingen drænrør	72	2	71	4
Drænrør	23	2	22	4
Bakkeø				
Ingen drænrør	62	13	72	4
Drænrør	14	11	20	5

Tabel 17: Overensstemmelse mellem forudsigelser og registrerede drænrør. Opgjort for geologiske klasser

Geologisk klasse	Olesen (2009)		Denne undersøgelse	
	Ikke drænet	Drænet	Ikke drænet	Drænet
Ferskvandsler				
Ingen drænrør	13	42	18	38
Drænrør	5	39	10	35
Ferskvandssand				
Ingen drænrør	37	29	58	8
Drænrør	24	10	23	11
Ferskvandstørv				
Ingen drænrør	17	35	27	25
Drænrør	14	34	26	23
Havsand				
Ingen drænrør	51	34	49	35
Drænrør	4	12	2	14
Moræneler				
Ingen drænrør	16	48	14	50
Drænrør	5	31	3,5	33
Morænesand				
Ingen drænrør	78	4	79	4
Drænrør	15	2	15	2
Smeltevandssand				
Ingen drænrør	80	0	77	3
Drænrør	20	0	16	5

Bilag 7: Bestilling

Den gode Bestilling af forskningsbaseret myndighedsrådgivning

Som led i udvikling af det gode samarbejde mellem parterne er det hensigten, at FVM's medarbejdere anvender nedenstående skema og medsender det ved konkrete bestillinger hos DCA – Nationalt Center for Fødevarer og Jordbrug, som varetager myndighedsbetjeningen på jordbrugs- og fødevarerområdet ved Aarhus Universitet.

Bemærk venligst tjeklisten i punkt 13, som er udarbejdet med henblik på at understøtte effektive arbejdsprocesser, give arbejdsglæde og høj kvalitet i resultaterne.

1. Bestillingens titel:

Kortlægning af markdræn

2. Hvem bestiller hvornår, til hvornår og til hvem?

- Departementet
- NaturErhvervstyrelsen
- Fødevarestyrelsen

Dato for bestilling: 25.09. 2014

Navn på kontaktperson i FVM: Tilde Hellsten, Jakob Møgelvang og Julie Borch Friderichsen

E-mail: Tihe@naturerhverv.dk, Jakm@naturerhverv.dk og jubf@naturerhverv.dk

Tlf.: 45 26 38 68, 25 23 83 91 og 25 23 84 73

Tidsfrist for levering: Endelig leverance medio 2016

Svar sendes til (husk evt. cc): Tihe@naturerhverv.dk, Jakm@naturerhverv.dk og jubf@naturerhverv.dk

3. Opgavetype

Det er vigtigt for DCA's interne håndtering af opgaven, at dette punkt udfyldes.

- Ny akut opgave, der ikke er beskrevet i Bilag 2 til aftalen for 2014
- "Bilag 2 – opgave" med ID-nummer:

Bestilles i forlængelse af FM-20

4. Temaområde (kun relevant for nye opgaver, der ikke er beskrevet i Aftalens Bilag 2)

FVM skønner af denne nye opgave hører under følgende tema/temaer (de 13 tema-områder findes kort beskrevet i punkt 14):

- 1.1 Bæredygtig husdyrproduktion (BH)
- 1.2 Bæredygtig planteproduktion (BP)
- 1.3 Bæredygtig teknologiudvikling og teknologivurdering (BT)
- 1.4 Bioøkonomi, bioenergi og landdistriktsudvikling (BL)
- 1.5 Ressourceeffektivitet (RE)

- 2.1 Fødevarerproduktion og dyrkningsjorden (FJ)
- 2.2 Fødevarerproduktion og bioressourcer (FB)
- 2.3 Fødevarerproduktion, landskab, natur og biodiversitet (FN)
- 2.4 Fødevarerproduktion, hjælpestoffer og miljø (FM)
- 2.5 Fødevarerproduktion, klima og drivhusgasser (FD)
- 3.1 Fødevarer kvalitet (KV)
- 3.2 Forbrugeradfærd og -præferencer for fødevarer (AD)
- 3.3 Mad- og måltidsvaners betydning for sundheden (MA)

5. Hvad bestilles?

- Forventingsafstemning vedrørende denne nye opgave
- Et notat til brug for drøftelse i embedsværket
- Et fagligt bidrag til beslutningsoplæg, svar til Folketinget m.m.
- Deltagelse i møde
- Deltagelse i arbejdsgruppe, f.eks. ifm. lovforberedende arbejde
- Andet: Kortlægning af dræn jf. DCA's projektbeskrivelse "Notat om status og muligheder for o

6. Hvad er baggrunden for bestillingen – hvorfor og til brug for hvad?

Beskrives kortfattet – 10–15 linjer – idet det er vigtigt at kende baggrunden/konteksten for bestillingen for at AU/DCA's svar bliver anvendeligt for FVM

DCA har udarbejdet en projektbeskrivelse for, hvorledes en kortlægning af dræn på markniveau kan gribes an jvf opgave med ID-nummer FM-20 på bilag 2 i 2014. Projektbeskrivelsen indeholder afsnit om metoder til kortlægning, beskrivelse af et forskningsprojekt til udvikling af GIS metoden, understøttet med brug af historiske kort og målinger i felten og til sidst et afsnit om samspillet mellem den fysiske drænkortlægning og N-retentionskortet.

NaturErhvervstyrelsen har forstået projektbeskrivelsen således, at kortlægningsprocessen indeholder følgende 3 leveringer:

- Et pilotprojekt til udarbejdelse af en manual til kortlægning af drænedede arealer
- Udarbejdelse af et landsdækkende kort over drænedede arealer (på basis af metode udviklet i nr. 1)
- Metode til lokal kortlægning af drænedede arealer ved brug af flyfotos, georadar og droner

Med nærværende bestilling ønsker NaturErhvervstyrelsen at igangsætte leverance 1 og 2. Endvidere vil NaturErhvervstyrelsen gerne drøfte igangsættelse af leverance nr. 3 (jvf. afkrydsning i punkt nr. 5.) samt datagrundlaget beskrevet i projektbeskrivelsen, særligt ifht. den digitale højdemodel.

NaturErhvervstyrelsen ønsker endvidere, at der afholdes statusmøder undervejs i kortlægningsprocessen således, at Styrelsen kan følge udviklingen og koordinere ifht. det administrative set-up i styrelsen. DCA forventes at indkalde til statusmøderne, når der er materiale at drøfte.

7. Detaljeret beskrivelse af problemstillingen:

Det er meget vigtigt, at bestillingen er præcis og entydig for at de bidragende forskere kan levere et svar, der kan anvendes; evt. kan man bede om en forventningsafstemning forud for bestilling m.h.p. at klarlægge hvad AU/DCA vil kunne levere svar på, se punkt 5.

Der henvises til projektbeskrivelsen. Er vedhæftet som bilag. Desuden:

- NaturErhvervstyrelsen ønsker endvidere, at der afholdes evalueringsmøder undervejs i kortlægningsprocessen således, at Styrelsen kan følge udviklingen og koordinere ifht det administrative set-up i styrelsen.
- NaturErhvervstyrelsen ønsker en vurdering af usikkerheden af kortlægningen i forhold til brug af forskellige datakilder og de forskellige dræntyper.
- Kortlægning af dræn er vigtig for NaturErhvervstyrelsen i forhold til bl.a.:
- Kortlægningen skal indgå som et administrationsgrundlag
- Identificering af risikoområder, hvor der sker udvaskning af kvælstof og fosfor (dræn kan virke som motorveje for stoftransport ud i vandløbene)
- Målrettet placering af virkemidler
- Modellering af retentionsforhold i jord

8. Dimensionering og form af besvarelse

- Kort faktuel svar
- Kort notat kun med allervigtigste litteraturhenvisninger - ca. sidetal:
- Grundig faglig redegørelse, gerne med fyldige litteraturhenvisninger fra ind- og udland
 - Redegørelsen ønskes på dansk
 - Redegørelsen ønskes på engelsk
 - AU/DCA kan vælge, om man ønsker at skrive på dansk eller engelsk

9. Evt. andre relevante oplysninger, som kan have relevans for AU/DCA's opgaveløsning:

Kortgrundlaget skal indgå som en del af NaturErhvervstyrelsens administrationsgrundlag

10. Hvis der anmodes om bemærkninger til et større dokument, skal det her anføres, hvilke afsnit eller sider, der er særligt relevante for modtageren eller som der særligt ønskes bemærkninger til:

Klik her for at angive tekst.

11. Proces

- Eventuelle milepæle med angivelse af tidsfrist
 - Oplysning om eventuelle/forventelige mødedatoer
 - Er andre institutioner inddraget i opgaven
- hvis ja – hvilke:
- hvis ja – om hvad:

- Det er vigtigt ved denne bestilling, at besvarelsen behandles fortroligt, til den er blevet behandlet i den ministerielle beslutningsproces. AU/DCA bedes derfor ikke udlevere den ikke-offentliggjorte besvarelse til personer udenfor FVM og AU/DCA eller - såfremt andre er involveret jf. ovenstående - da kun til de, der specifikt er involveret, og kun i det omfang disse har behov herfor. I sidstnævnte tilfælde skal DCA kræve samme fortrolighed af modtageren, som DCA selv er underlagt.
-

12. Offentliggørelse af rådgivningssvar på AU/DCA's hjemmeside

Bestilte rådgivningsopgaver skal som hovedregel offentliggøres så hurtigt som muligt, og det tilstræbes, at der ved opgavens igangsættelse laves en præcis aftale om offentliggørelse.

I henhold til de retningslinjer, der er vedtaget i aftalen med FVM (Afsnit 4.4), offentliggør AU/DCA sine rådgivningssvar på hjemmesiden: <http://dca.au.dk/myndigheder/>

Nedenfor anføres evt. særlige forhold omkring offentliggørelse af svar på denne bestilling:

- Kan offentliggøres på hjemmeside samtidig med, at svar afsendes til FVM
- Kan offentliggøres 10 arbejdsdage efter afgivelse af svar
- Kan offentliggøres en måned efter afgivelse af svar
- Kan offentliggøres den
- En eksakt dato for offentliggørelse kan ikke gives
- Svaret kan ikke offentliggøres af ganske særlige hensyn, som er anført herunder

Bemærk at i de tilfælde, hvor DCA ikke kan offentliggøre svaret, vil der inden igangsættelse af opgaven blive foretaget en vurdering af, om DCA kan påtage sig opgaven

Begrundelse:

Klik her for at angive tekst.

13. Vigtigt: Tjekliste til brug ved bestilling

- ✓ Husk at bestillingen skal være præcis og entydig for at de bidragende forskere kan levere et svar, der kan anvendes efterfølgende – det kan være en god ide at lade kolleger læse bestillingen, inden den sendes for at være sikker på, at den forstås af andre.
- ✓ Hvis det er vanskeligt at udforme en præcis bestilling, kan bestiller bede om, at der foretages en forventningsafstemning inden fremsendelse m.h.p. at klarlægge, hvad AU/DCA vil være i stand til at levere svar på.
- ✓ Bestillingen sendes altid til AU/DCA – Nationalt Center for Fødevarer og Jordbrug, Postboks 50, 8830 Tjele, e-mail: dca@au.dk.
- ✓ Det anføres i bestillingen, hvem svaret skal sendes til, og om der skal sendes cc til personer eller kontorpostkasser.
- ✓ Alle svar sendes i kopi til Center for Innovation: innovation@naturerhverv.dk.
- ✓ Tidsfrister, omtrentligt omfang og særlige ønsker til formidling aftales mellem bestiller og koordinator.
- ✓ *Bemærk: Henvendelser til enkeltmedarbejdere er i følge Aftalen med AU/DCA ikke sikret besvarelse.*
- ✓ Hastebestillinger følges altid umiddelbart op pr. telefon (hvis koordinator ikke er på arbejde, vil der være forslag i auto-reply til hvem, der kan kontaktes), så FVM medarbejderen sikrer sig, at den bliver håndteret så hurtigt som muligt. Hvis det ikke allerede er klart, hvad grunden er til en ekstraordinær kort frist, må det oplyses. Det fremmer motivationen i alle led.
- ✓ Den opfølgende dialog og uddybning af opgaven kan ske mellem bestiller og den eller de forskere ved AU/DCA, der forestår besvarelse.
- ✓ Husk at skrive afsenders telefonnummer og evt. en opfordring til at tage kontakt med henblik på afstemning af forventninger og uddybende spørgsmål.
- ✓ Husk god e-mailpraksis. Sæt kun de modtagere og cc'er på, der virkelig har brug for at se bestillingen.
- ✓ Meld meget gerne tilbage, når notatet/svaret er modtaget/brugt. Dette tjener flere formål: Dels er det vigtigt for AU/DCA at erfare, om form, omfang mv. har været tilfredsstillende, så der kan ske en stadig udvikling i kvalitet af den forskningsbaserede rådgivning. Dels er det vigtigt for evt. opfølgende rådgivning på området, at DCA får at vide, hvordan svaret har indgået i det videre arbejde i FVM.
- ✓ Hvis svaret giver anledning til uddybende spørgsmål eller evt. yderligere analyser, bedes bestiller henvende sig til en af DCA's koordinatore for myndighedsrådgivning for at aftale nærmere procedure. Navne, e-mails og telefonnumre på DCA's koordinatore:
 - ❖ Susanne Elmholt (Susanne.Elmholt@agrsci.dk , tlf. 871 57685),
 - ❖ Karl Tolstrup (Karl.Tolstrup@agrsci.dk , 871 51265),
 - ❖ og Klaus Horsted (Klaus.Horsted@agrsci.dk , 871 57975).

14. Myndighedsrådgivning på tretten temaområder

DCA – Nationalt Center for Fødevarer og Jordbrug leverer forskningsbaseret myndighedsbetjening på tretten temaområder, der er fordelt på de tre elementer af Fødevareministeriets mission. Ved bestilling af rådgivningsopgaver beder AU/DCA om, at kontaktpersonen i FVM skønner, hvilket eller hvilke temaer, man mener opgaven kan henføres til. Til hjælp herfor er der nedenfor anført de korte beskrivelser af forskningen inden for de 13 områder. Beskrivelserne stammer fra Aftalens Bilag 1b.

Element 1: Udviklings- og vækstorienteret fødevareerhverv

1.1 Bæredygtig husdyrproduktion (BH)

Forskningen skal modsvare fødevareerhvervenes behov for at kunne gennemføre en ansvarlig og bæredygtig husdyrproduktion, der tilgodeser landmandens, erhvervenes og samfundets ønske om god dyrevelfærd og –sundhed, forebyggelse af zoonoser, minimering af restkoncentrationer af medicin og forekomst af resistens over for antibiotika under hensyntagen til produktionsøkonomi, ressourceeffektivitet og produktivitet samt minimering af næringsstofoverskud, uønskede stoffer, emissioner og lugt.

1.2 Bæredygtig planteproduktion (BP)

Forskningen skal modsvare fødevareerhvervenes behov for at kunne gennemføre en effektiv og ansvarlig planteproduktion, under hensyntagen til produktionsøkonomi, ressourceeffektivitet, produktkvalitet og produktivitet samt understøtte samfundets ønsker om et rigt og varieret kulturlandskab, øget biodiversitet samt minimering af påvirkning af miljøet med næringsstoffer, emissioner og pesticider.

1.3 Bæredygtig teknologiudvikling og teknologivurdering (BT)

Forskningen skal modsvare fødevareerhvervenes behov for automatisering og udvikling af nye miljø- og ressourceoptimerende teknologier, med tilhørende managementsystemer, herunder bio- og nanoteknologier, samt myndighedernes behov for en systematisk vurdering af disse i forhold til deres potentielle bidrag til erhvervenes samfundsmæssige udfordringer og effekt. Udvikling af ny teknologi bygger på en procesforståelse i hele værdikæden med ansvarlighed og respekt for samspillet mellem teknik, biologi og etik i relation til f.eks. ressourceforbrug, klima og miljø, fødevaresikkerhed og -kvalitet, husdyrsundhed samt -velfærd.

1.4 Bioøkonomi, bioenergi og landdistriktsudvikling (BL)

Forskningen skal modsvare fødevareerhvervenes og samfundets behov for udvikling og afprøvning af nye bæredygtige forretningsområder og –modeller, forsyningskæder, værdikæder, produktionsprocesser og teknologier, der fremmer en bæredygtig og ressourceeffektiv erhvervsmæssig anvendelse af biologisk materiale og fremmer erhvervsudvikling, beskæftigelse og udvikling i landdistrikter.

1.5 Ressourceeffektivitet (RE)

Forskningen skal modsvare fødevareerhvervenes behov for at kunne fremme ressourceeffektiviteten i hele produktionsprocessen og forsyningskæden, herunder dokumentere kritiske punkter i den enkelte proces eller led i forsyningskæden med optimeringspotentiale.

Element 2: Ansvarlig forvaltning af naturressourcer

2.1 Fødevarerproduktion og dyrkningsjorden (FJ)

Forskningen skal modsvare fødevareerhvervenes og samfundets behov for at dyrkningsjordens frugtbarhed, kvalitet og funktion til stadighed opretholdes eller forbedres til opfyldelse af nuværende og fremtidige anvendelser og funktioner.

2.2 Fødevarerproduktion og bioressourcer (FB)

Forskningen skal modsvare fødevareerhvervenes og samfundets behov for at de biologiske ressourcer, der ligger til grund for fødevarerproduktionen, til stadighed opretholdes og udvikles til gavn for den nuværende og fremtidige anvendelse.

2.3 Fødevarerproduktion, landskab, natur og biodiversitet (FN)

Forskningen skal modsvare fødevareerhvervenes og samfundets behov for opretholdelse og beskyttelse af landskabelige værdier, naturværdier samt biodiversitet på især arealer med landbrugsmæssig anvendelse.

2.4 Fødevarerproduktion, hjælpestoffer og miljø (FM)

Forskningen skal modsvare fødevareerhvervenes og samfundets behov for viden om og dokumentation af fødevarerproduktionens påvirkning af miljøet, risiko, spredning og effekt af hjælpestoffer samt dannelse, risiko, spredning og effekt af naturligt forekommende stoffer, herunder toksiner og hormonforstyrrende stoffer.

2.5 Fødevarerproduktion, klima og drivhusgasser (FD)

Forskningen skal modsvare fødevareerhvervenes og myndighedernes behov for viden om og dokumentation af fødevarerproduktionens emission af drivhusgasser, øvrige klimapåvirkninger samt betydningen af klimaændringerne for den fremtidige teknologianvendelse og produktion.

Element 3: Fødevaresikkerhed, forbrugernes valgmuligheder og sunde kostvaner

3.1 Fødevarekvalitet (KV)

Forskningen skal modsvare fødevareerhvervenes og myndighedernes behov for dokumentation og vurdering af fødevarers egnethed til videre forarbejdning, sensorisk og sundhedsmæssig kvalitet, herunder indflydelse af de primære produktionsforhold, processing, emballering og lagring.

3.2 Forbrugeradfærd og -præferencer for fødevarer (AD)

Forskningen skal modsvare fødevareerhvervenes og myndighedernes behov for viden om udviklingen i forbrugertrends, -adfærd, og -præferencer m.v. som grundlag for bl.a. grøn omstilling af produktionen, produktdifferentiering og produktudvikling, samt myndighedernes behov for viden om forbrugernes adfærd og præferencer for fødevarer og forbrugernes opfattelse af diverse former for sundhedsformidling (f.eks. mærkning) og interventioner.

3.3 Mad- og måltidsvaners betydning for sundheden (MA)

Forskningen skal modsvare fødevareerhvervenes behov for viden om efterspørgsel og afsætningsmuligheder for sunde, ernæringsrigtige fødevarer med som samtidig opfylder forbrugernes krav til sensorisk og kulinarisk kvalitet samt myndighedernes behov for viden om fødevarer, herunder fødevarernes næringsstæthed i relation til sundhed samt mad- og måltidsvaners betydning for befolkningens ernærings- og sundhedstilstand.

Bilag 8: Projektbeskrivelse



DCA - NATIONALT CENTER FOR FØDEVARER OG JORDBRUG
AARHUS UNIVERSITET

NaturErhvervstyrelsen

Vedrørende notat om status og muligheder for og ressourcebehov ved prioriteret kortlægning af dræn på markniveau

Susanne Elmholt

Koordinator for
myndighedsrådgivning

Dato: 14. april 2014

Direkte tlf.: 8715 7685
E-mail:
Susanne.Elmholt@agrsci.dk

Afs. CVR-nr.: 57607556

Side 1/1

NaturErhvervstyrelsen (NAER) har den 10. januar 2014 bedt DCA – Nationalt Center for Fødevarer og Jordbrug udarbejde et notat om kortlægning af dræn på markniveau inklusiv forslag til hvorledes kortlægningsopgaven kan gribes an.

Herunder følger notatet, der indeholder afsnit om metoder til kortlægning, beskrivelse af et forskningsprojekt til udvikling af GIS metoden, understøttet med brug af historiske kort og målinger i felten og til sidst et afsnit om samspillet mellem den fysiske drænkortlægning og N-retentionskortet.

Notatet er udarbejdet af seniorforsker Kirsten Schelde, sektionsleder Mogens H. Greve, seniorforsker Charlotte Kjærgaard, lektor Bo Vangsø Iversen og seniorforsker Christen D. Børgesen, alle Institut for Agroøkologi.

Med venlig hilsen

Susanne Elmholt
Seniorforsker

Status og muligheder for og ressourcebehov ved prioriteret kortlægning af dræn på markniveau

Kirsten Schelde, Mogens H. Greve, Charlotte Kjærgaard, Bo Vangsø Iversen, Christen D. Børgesen, Institut for Agroøkologi

Sammenfatning

NaturErhvervstyrelsen (NAER) har den 10. januar 2014 anmodet DCA – Nationalt Center for Fødevarer og Jordbrug om i et notat at udforme et oplæg til beskrivelse for et projekt om forbedret kortlægning af drænsystemer på landbrugsjord.

Notatet beskriver indledningsvist (afsnit 2) metoder til kortlægning (GIS-baseret kortlægning med tilgængelige kort samt fysisk lokalisering af dræn vha. georadar m.fl.). Derefter skitseres (afsnit 3) et forskningsprojekt til udvikling af GIS metoden, understøttet med brug af historiske kort og målinger i felten. Endelig redegør afsnit 4 for samspillet mellem den fysiske drænkortlægning og retentionskortet, og notatet afsluttes med en anbefaling.

Indhold

Sammenfatning	1
1. Indledning	2
2. Metoder til forbedret kortlægning af markdræn.....	2
2.1 Kortlægning ved hjælp af tilgængelige GIS temalag	2
2.2 Fysisk lokalisering af dræn ved hjælp af flyfotos, georadar og andre metoder.....	5
3. Projektoplæg.....	7
3.1 Kort projektbeskrivelse.....	7
3.2 Skøn over ressourcebehov	8
4. Samspil med N-retentionskortet	9
5. Anbefalinger.....	10
6. Referencer	11
Bilag 1: Brug af historiske kort	12

1. Indledning

NaturErhvervstyrelsen (NAER) har 10. januar 2014 anmodet DCA – Nationalt Center for Fødevarer og Jordbrug – om en skitse/projektbeskrivelse for hvordan et projekt om forbedret kortlægning af markdræn på markniveau kan udformes.

Foruden metode(r) til kortlægning skal der redegøres for forventet ressourcebehov ved kortlægningsmetoderne. AU/DCA bedes endvidere beskrive hvor den pt. manglende viden om dræning på markniveau er problematisk i forhold til differentieret kvælstofregulering, herunder komme med forslag til afgrænsning af opgaven (f.eks. om nogle arealer er mere interessante at kortlægge end andre arealer), dataindsamling og samspil med N-retentionskortet. Endelig bedes AU/DCA vurdere, hvorvidt Orbicons drænkortarkiv kan anvendes i den fremtidige kvælstofadministration samt foreslå alternative datakilder.

2. Metoder til forbedret kortlægning af markdræn

I det efterfølgende beskrives to metoder til at kortlægge markdræn på mark- eller markblokniveau i Danmark. Metodernes anvendelighed og gennemførlighed er skalaafhængige:

- a. Forbedret kortlægning vha. GIS og tilgængelige digitaliserede kort (afsnit 2.1) anses for at være ressourcemæssigt mulig for hele landet.
- b. Feltbaserede metoder (afsnit 2.2) kan anvendes til at opnå mere detaljeret information om den fysiske beliggenhed af markdræn. Det drejer sig om kortlægning vha. flyfotos, georadar, scannede og digitaliserede drænkort m.fl., som alle er forholdsvis ressourcekrævende metoder, der kan gennemføres lokalt.

2.1 Kortlægning ved hjælp af tilgængelige GIS temalag

Status

Den seneste kortlægning af drænedede arealer i Danmark blev foretaget af Olesen (2009). Resultatet af analysen var et kort over potentielt drænedede områder, hvor alle arealer blev tildelt en sandsynlighed (%) for at de er kunstigt drænedede. Kortet var baseret på en GIS-analyse ved kobling af GEUS' geologiske jordartskort, kort over landskabselementer, samt jordbunds-klassificeringen (farvekodekort). Alle områder blev desuden klassificeret som lavbund eller højbund. På basis af et geografisk netværk af observerede drænprocenter blev alle områder (de mulige kombinationer af jordbund, geologi, landskabstype og georegion) tildelt en generaliseret drænprocent.

Analyserne i Olesen (2009) inddrog ikke informationer om terræn.

Metode

Til en forbedret kortlægning af markdræn i Danmark tages der tages udgangspunkt i kortlægningen fra Olesen (2009). Kortet med drænprocenter bruges ikke direkte, men de bagvedliggende kort indgår i GIS analysen. Hvor det er muligt, anvendes der opdaterede eller forbedrede kort.

Basiskortene

- Lavbundstema
- GEUS geologiske jordartskort
- Farvekodekort (overjordstype)
- Opdateret kort over jordtyper (Mark (2014); Adhikari et al. (2013))
- Landskabslementer

kombineres med bl.a.

- Afgrødekort på markniveau vedr. 2010-2013
- Arealanvendelseskort (afgrøder) på markblokniveau for 2005-2009
- Hydrologisk tilrettet digital højdemodel, DEM (DHyD/Rain)
- Afledte af DEM, beregnet af Adhikari et al. (2013) og tilgængelig i AU/DCA
- Vandløbstema
- Lavninger (fra Naturstyrelsen; baseret på hydrologisk DEM)
- Strømningsvejes oplandsstørrelse (fra Naturstyrelsen; baseret på hydrologisk DEM)

Hvis det er muligt at skaffe, kombineres også med

- Kort over øvre grundvandspejl (baseret på målinger, ikke modelleret)
- **Historiske kort** (Bilag 1) kan muligvis anvendes som supplement til ovenstående GIS analyse i en afgrænsning af naturligt dårligt drænedede arealer. Metoden er ressourcekrævende og vil kun være mulig til regional/lokal kortlægning eller hvor der er behov for øget sikkerhed i bestemmelsen vha. GIS.

GIS analysemetoden vil bestå af en række delelementer og er ikke færdigudviklet endnu.

Der arbejdes rasterbaseret, muligvis i et 8-16 m grid eller et andet multiplum af 1,6 m, som er opløsningen i DEM og andre afledte kort.

Som udgangspunkt laves der drænklassificeringer i jordarts- og jordtypekortene, idet de enkelte jordklasser karakteriseres som hhv. 'kræver formentlig dræning' (f.eks. moræneler) eller 'kræver næppe dræning' (f.eks. flyvesand). For afgrødekortene klassificeres de enkelte afgrødekoder som dræningskrævende (f.eks. vinterkorn) eller ej (f.eks. permanent græs med lavt udbytte). En række jordtyper og afgrøder vil falde i kategorier 'måske'.

For alle celler beregnes, i hvilken andel af årrækken 2005-2013 der var dræningskrævende afgrøder på arealet (Figur 1). Idéen er, at hvis geologi og terræn tilsiger, at arealet kræver dræning, og det samtidig har været muligt at dyrke dræningskrævende afgrøder som f.eks. vinterhvede i en årrække, vil der være høj sandsynlighed for at arealet er kunstigt drænet. Modsat – hvis der kun har været dyrket ikke-dræningskrævende afgrøder – er arealet formentlig ikke drænet.

Kortlægningen vil bestå af en lang række algoritmer af følgende type, hvor DrKr står for DræningsKrævende:

Hvis underjord = DrKr og overjord = ej_DrKr og DrKr-afgrødeår > 50% => cellen er DrKr

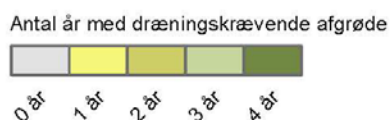
Udover kriterierne i eksemplet herover vil indgå kriterier som 'vandret afstand til nærmeste vandløb/grøft', 'lodret afstand til nærmeste vandløb/grøft', 'cellen er en lavning', 'hældningen > xx grader', osv.

Algoritmerne udgør en beslutnings-træstruktur og bygger på overvejelser om, hvilke områder der er (og ikke er) drænet og hvorfor. Arbejdet vil videreudvikle udenlandske erfaringer med lignende kortlægning (Tetzlaff et al., 2009, Naz et al., 2009).

I denne del af analysen involveres en erfaren medarbejder inden for dræning fra Orbicon (det tidligere Hedeselskabet).



Figur 1. Indledende analyse baseret på afgrødekort i årene 2010-2013 for værkstedsområdet, der også er afbildet i Figur B1_1 i Bilag 1. Kortet viser hvor der har været dyrket dræningskrævende afgrøder som f.eks. vinterkorn.



Begrænsninger og usikkerheder

- Jordartskortene, der tænkes inddraget, er kun repræsentative til en begrænset dybde (max 1,0 m). Hvis der findes dybere vandstandsende lag, som ikke fremgår af kortene, vil deres effekt – som ofte udmønter sig i et dræningsbehov – ikke nødvendigvis afsløres af analysen.
- Forud for 2010 foreligger arealanvendelsen (afgrøder) kun kortlagt på markblokniveau. Analysen vil kræve at hver markblok årligt tildeles netop én dræningsklasse. Tildelingen vil ske efter en algoritme, som kan inddrage årets afgrøder og andre variable i blokken, bl.a. jordtyper og terræn. Metoden medfører at enkeltmarker fejlagtigt kan blive påtvunget et dræningsbehov pga. deres tilhørsforhold i markblokken.
- Metoden vil for nogle kombinationer af landskabstype og jordarter kunne forudsige med ret høj sikkerhed hvilke arealer der er kunstigt drænet. Der vil være andre kombinationer af landskabstyper og jordarter, hvor kortlægningen vil være mere usikker.
- Nøjagtigheden og sikkerheden af kortlægningen vil blive forsøgt kvantificeret.

Muligheder for verifikation

- Check mod punktobservationer fra KVADRATNET som i Olesen (2009), samt information fra LOOP oplande og VAP områder
- Check mod georefererede oversigtskort over projekterede drænanlæg fra Orbicons arkiv, f.eks. i et transekt
- Check mod historiske kort (Bilag 1)
- Check mod feltbaserede metoder (flyfotos, georadar)

Ressourcebehov, generelt

Den mest ressourcekrævende del af den GIS-baserede metode er den iterative metodeudvikling og verifikation. Når det for en række landskabselementer og geologier er sandsynliggjort, at metoden giver et retvisende billede af det faktisk drænedede areal, er kortlægning på landsplan mulig.

2.2 Fysisk lokalisering af dræn ved hjælp af flyfotos, georadar og andre metoder

De fysiske metoder kan give detaljeret information om individuelle drænanlægs og -rørs beliggenhed. Metoderne kan f.eks. anvendes, hvor der planlægges miljømæssige virkemidler på markniveau, der forudsætter at landmanden kan dokumentere, hvordan drænsystemet er anlagt. Metoderne kan desuden anvendes til lokal verifikation af den GIS baserede kortlægning.

Metoder

Luftfotografier

Satellit- og luftfotos kan afsløre struktur- og teksturforskelle i den øvre del af jorden først og fremmest afstedkommet af forskelle i vegetationen. Oftest vil luftfotos være at foretrække, da disse har en høj opløsning og samtidig er relativt billige i anskaffelse. Afhængigt af metoden, hvormed drænrørene er installeret, vil forskelle i tekstur og struktur have indvirkning på jordens vandindhold og dermed afgrødevæksten. Forskellene vil være mest tydelige i starten af vækstsæsonen, og fotografier fra de tidlige forår vil derfor være mest anvendelige. Afgrødetypen vil ligeledes have en stor betydning. Lavt stående kornafgrøder eller græsmarker vil oftest give et tydeligere aftryk end rækkeafgrøder såsom majs og kartofler (Tetzlaff et al., 2009). Tetzlaff et al. (2009) kombinerede luftfotometoden med andre relaterede GIS-temaer såsom jordbundskort, arealanvendelse, og dybde til grundvandspejlet og producerede derved et GIS-tema, der angav områder, man vurderede var potentielt drænedede. Herved kunne de finde et sammentræf på 78 % mellem de områder, man vurderede var drænedede, og de områder, der faktisk var drænedede.

Georadar

Georadar (på engelsk ground penetrating radar (GPR)) er en metode, der benytter sig af en elektromagnetisk puls, der bliver sendt ned i jorden. En modtager på georadaren måler den tid det tager den udsendte puls at blive reflekteret fra et medie i jorden. Refleksioner fra forskellige dybder giver en signalaftegning, som er en funktion af radarbølgeamplituden og tiden. Markante ændringer i lagrækkefølgen i jorden eller nedgravede genstande vil påvirke energiudbredelsen af den elektromagnetiske puls. Anvender man georadar til at lokalisere drænrør, vil den bratte signal-overgang mellem drænrøret og den omgivende jord, mellem drænrøret og vandet/luften i røret eller mellem påfyldningsjorden over drænet og den underliggende uforstyrrede jord kunne detekteres af georadaren. I et amerikansk studie (Allred et al., 2004) blev en række geofysiske metoder testet med

det formål at undersøge om det var muligt at lokalisere drænrør i landbrugsområder. Her viste det sig, at en ground penetration radar (GPR) var den eneste anvendelige geofysiske metode. På 11 forsøgsmarker, hvor tekturen varierede fra 'clay' (omtrent svarende til JB 8 og 9) til 'sandy loam' (omtrent svarende til JB 5 og 6) blev drænene lokaliseret med en gennemsnitlig succesrate på 81 %. Laveste succesrater (0 %) blev opnået på marker med meget høje lerindhold, og hvor drænene var gravet ned i dybder på mere end 1 meter. For marker hvor drænene lå i 1 meters dybde og derover, og hvor tekturen varierede fra 'sandy loam' til 'silty clay' (JB 10) var succesraten i langt de fleste tilfælde på 100 %.

AU har kontakt til en amerikansk forskergruppe med ekspertise på feltet, samt til et dansk firma, der har det rigtige GPR udstyr til formålet.

Indsamling af gamle drænkort, herunder brug af drænkort fra Orbicons arkiv

I Letland foregår i disse år en betydelig indsats på drænkortlægning og opbygning af en national drændatabase. Den store digitaliseringsopgave udføres og vedligeholdes af en enhed under Landbrugsministeriet (Lagzdins og Jansons 2013; www.melioracija.lv). Digitaliseringen lettes af, at de fleste drænanlæg i Letland blev anlagt i Sovjettiden, og at kortene har været opbevaret centralt.

I Danmark er billedet langt mere fragmenteret med flere historiske bølger af dræningsindsatser, mange private aktører, entreprenører mv. Orbicon råder over et betydeligt arkiv med drænanlæg, som Hedeselskabet har projekteret, men ikke nødvendigvis udført. Det vil være muligt at erhverve digitaliserede og georefererede drænkort fra Orbicon.

Det vil ligeledes være muligt for de lokale landboforeninger eller landbrugsrådgivninger at indsamle og digitalisere drænkort fra kredsens landmænd.

Anvendelse af droner og termisk fotografering

Der er mulighed for, at højopløselig termisk fotografering kan anvendes til lokalisering af drænudløb i lavbundsområder eller vandløb på tidpunkter af året hvor der er stor temperaturkontrast mellem overfladen/overfladevand og udstrømmende dræn/grundvand.

Rekognoscering efter drænrørsudløb til vandløb

For marker, der grænser ned til vandløb, vil det ofte være muligt at lokalisere drænrørene ved en visuel inspektion foretaget i selve vandløbet. Her vil det ofte være muligt at lokalisere de drænrør, der afvander markerne direkte ud i vandløbet. Drænrør, der befinder sig under vandspejlet, vil være svære at lokalisere. Inspektionen bør derfor foretages, når vandføringen er lav, og vegetationen langs åbredden er begrænset. Det tidlige forår vil oftest være det bedste tidspunkt.

Begrænsninger

- Flyfotos: Som udgangspunkt er metoden med identifikation via fotos begrænset af, at den skal foretages manuelt. Det vil sandsynligvis være vanskeligt at opstille billedbaserede rutiner, der automatisk kan detektere drænrørsforløb i landskabet. Afhængigt af vegetation, jordbundsforhold, vandmangel og tidspunkt for fotograferingen vil det være usikkert om (alle) drænrør bliver detekteret med metoden. Usikkerheden medfører, at fravær af synlige drænrør på fotos ikke med sikkerhed kan tolkes som fravær af drænsystemer.
- Georadar: Hvor lerindholdet ligger under 40%, og hvor drænrørene ligger i en meters dybde og vandindholdet befinder sig på markkapacitet og derunder, må det forventes, at det vil være muligt at lokalisere drænrør i Danmark med GPR. Lerindholdet for danske jorde ligger sjældent over 40%, og drænrør, specielt i systemdrænedede områder, ligger oftest i omkring 1

meters dybde. For dybereliggende drænrør vil succesraten sandsynligvis være begrænset. Man skal undgå at foretage målingen, når jordens vandindhold er højt.

- Brug af digitaliserede drænkort fra private eller fra Orbicon: Orbicons arkiv indeholder kort over områder, hvor et dræningsbehov har været erkendt, og hvor det i store træk er kendt (ved anmærkninger i de enkelte sagsnumre), i hvilket omfang dræningen er foretaget som projekteret.
Det vil være vanskeligt og logistikmæssigt tungt at indsamle og digitalisere alle drænkort i Danmark. Det vurderes, at gamle detailkort for drænanlæg hovedsageligt skal bruges – og vil være værdifulde - til at verificere en landsdækkende GIS-baseret metode.
- Rekognoscering langs vandløb kan udelukkende bruges til at konstatere, at den tilstødende mark er drænet, men selve dræforløbet i marken vil forblive ukendt.

Muligheder for verifikation

Verifikation af metoderne vil kunne foretages i områder, hvor drænrørsplaceringen er kendt. Det er den f.eks. på arealer, der er tilknyttet hhv. landovervågningen (LOOP) og Varslingssystem for udvaskning af pesticider til grundvand (VAP). Verifikationen bør foretages på jorde med forskellige teksturtyper og dræningsmønstre (systemdrænedede arealer, strengelægning, etc.)

Ressourcebehov, generelt

- Luftfotos er for de fleste statslige institutioner frit tilgængelige og forefindes for en længere årrække. Det største ressourceforbrug vil være i forbindelse med den manuelle processering af kortene.
- GPR kan udføres fladedækkende på marken ved at fastspænde udstyret bag på en ATV. Hastigheden afhænger af det anvendte udstyr. Målingerne kræver en efterfølgende processering og tolkning før drænmønstret kan fastlægges.
- Digitalisering og georeferering af drænkort samt indkøb af digitaliserede drænkort fra Orbicon vil være ressourcekrævende. Se priseksempel i afsnit 3.
- Rekognoscering langs vandløb er ligeledes meget ressourcekrævende.

3. Projektoplæg

3.1 Kort projektbeskrivelse

Det anbefales, at der gennemføres et pilotprojekt til udvikling og verifikation af en ny GIS baseret kortlægning af markdrænsystemer i Danmark. Metoden skal dække væsentlige kombinationer af landskabstyper og jordbundsforhold.

Desuden anbefales afprøvning af metoder til fysisk lokalisering af drænsystemer på lokalt plan. Disse metoder er nødvendige at have til rådighed, hvor der er behov for dokumentation af dræns beliggenhed indenfor en mark.

Udvikling af den GIS-baserede metode i udvalgte værkstedsområder

- Værkstedsområder: Til metodeudviklingen er der behov for at arbejde i et antal (6-9) værkstedsområder, der repræsenterer et udvalg af kombinationer af landskabselementer og jordarter. Efter vores vurdering bør som minimum østdansk moræne, midtjysk moræne,

nordjysk moræne, bakkeø, Littorina og Yoldia være repræsenteret i værkstedsområderne. I udvælgelsen af værkstedsområder vil der også blive taget hensyn til, hvor det er muligt at finde drændata til kalibrering og verifikation.

- GIS-metoden: Der arbejdes iterativt med at fastlægge beslutningstræstrukturen, så alle kombinationer af landskabselementer, jordtypeklasser mv. er dækket ind og kategoriseres tilfredsstillende med hensyn til om der er drænet eller ej.
Der indkøbes ekspertise fra Orbicon, og Orbicons dræneekspert (senior-tekniker Karl Henrik Laursen) inddrages i detaljer vedr. udviklingen af beslutningstræet og tolkning af Orbicons drænkort.
- Kalibrering og verifikation: I værkstedsområderne anvendes som 'kalibreringsdata'
 - 1) information fra kvadratnetpunkter og andre tilgængelige lokale oplysninger,
 - 2) drænprojektkort fra Orbicon i det omfang de forefindes,
 - 3) resultater fra lokale feltundersøgelser med georadar og andre metoder, samt
 - 4) historiske kort over udbredelsen af vandlidende jorde (Bilag 1). Det vil blive undersøgt i hvor høj grad brugen af de historiske kort kan forbedre analysen. Der foreligger allerede processerede historiske kort fra de tre værkstedsområder vist i Bilag 1.

Metode til lokal kortlægning af drænedede arealer

Her foreslås en trinvis følge af metoder, startende med den enkleste og billigste metode.

1. Flyfotos: For arealer i enkelte værkstedsområder (2-3; med kendt drænkortlægning) afprøves i hvor høj grad det er muligt at erkende drænsystemer vha. gamle flyfotos.
2. Georadar metode: Anvendeligheden af georadar til kortlægning af drænsystemer undersøges på repræsentative jordtyper og drændybder for mindre arealer i 2-3 værkstedsområder med kendt drænkortlægning. Georadar kortlægningen gennemføres ved varierende vandindhold på højbund og lavbund. På baggrund af denne undersøgelse fastlægges kriterier for brug af georadar til drænkortlægning.
Desuden gennemføres et GIS-baseret feasibility studie af, for hvilke arealer i Danmark det skønnes muligt at kortlægge dræn ved brug af georadar.
Tidsforbrug og omkostninger ved metoden fastlægges.
3. Anvendelse af droner og termisk fotografering af drænudløb: Anvendeligheden af rekognoscering vha. droner og kameraer undersøges i de samme delområder som metode 1 og 2. Der udføres rekognoscering flere gange over en vækstsæson, startende i det tidlige forår. Der anvendes både optisk og termisk kamera.

3.2 Skøn over ressourcebehov

- Udvikling af GIS baseret metode:
8 mandmåneder til metodeudvikling på 6-9 værkstedsområder
2 mandmåneder til syntese
Anslået 4-7000 kr. pr. værkstedsområde til indkøb af drænkort og yderligere detaljerede

oplysninger fra Orbicon

Anslået 1 mandmåned til indkøb af rådgivning ved seniortechniker Karl Henrik Laursen fra Orbicon

- Undersøgelser med flyfotos og georadar:
1 mandmåned pr værkstedsområde
2 mandmåneder til bearbejdning af data
Drift til GPR er ikke endeligt afklaret
- Undersøgelser med drone kan endnu ikke prisfastsættes, men AU/DCA har det nødvendige udstyr (drone og kameraer) til rådighed.

4. Samspil med N-retentionskortet

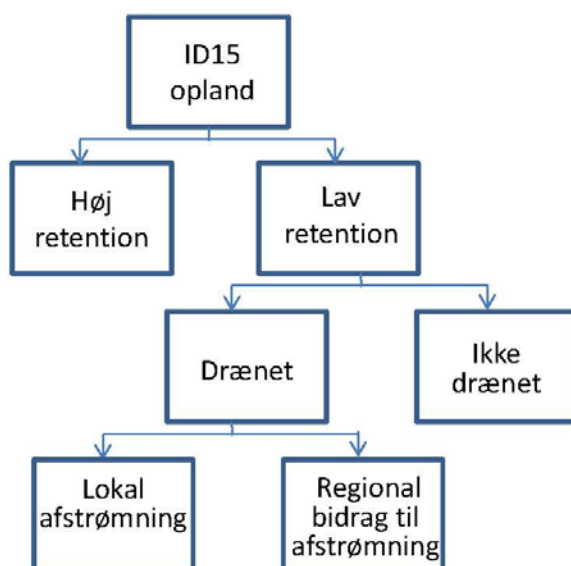
Kortlægning af arealers N-retentionspotentiale forudsætter som udgangspunkt kendskab til strømningssvejen fra mark til recipient, og kendskab til retentionsprocesser under transport. Når perkolat fra rodzonen afstrømmer direkte til vandløb via dræn, hindres den naturlige nedsivning til dybere grundvand, og dermed reduceres potentialet for N-reduktion i forbindelse med vandets transport frem til vandløbet. Omfanget af dræning af danske landbrugsarealer er usikkert, og der foreligger kun begrænsede oplysninger om beliggenheden af dræn, men det er skønnet, at mere end 50 % (1,5 mio. ha) er systematisk drænet (Olesen, 2009).

Det igangværende arbejde med opdatering af N-retentionskortet omfatter en ny metode til kortlægning af N-retentionen, der geografisk opdeler landet i ca. 3200 deloplande med en gennemsnitsstørrelse på ca. 1500 ha (ID15 oplande). N-retention opgøres som forholdet mellem den beregnede N-udvaskning fra rodzonen og kvælstoftransporten fra oplandet til ferske og marine recipienter. N-transporten fra dyrkede og udyrkede arealer antages overordnet at kunne beregnes med DK modellen. Da DK modellen opererer med et 500 m grid, vil viden om markskala transportveje, herunder betydningen af drænsystemer, ikke indgå i den nuværende modelopsætning. Hvert ID15 opland tildeles på baggrund heraf en retentionsklasse, der repræsenterer et vægtet gennemsnit for oplandets arealer (landbrug og ikke landbrug).

En fremtidig videreudvikling og yderligere detaljeret retentionskortlægning forudsætter en bedre beskrivelse af transportvejen mellem mark og recipient, herunder især en mere korrekt opsplitning af rodzoneafstrømning i hhv. drænastrømning og vertikal nedsivning til dybere grundvand. En forbedret beskrivelse af drænastrømningen forudsætter dels en kortlægning af drænedede arealer, dels modeller der kan forudsige drænastrømning på den aktuelle lokale skala. Kortlægningen af drænedede arealer vil i første fase kunne bidrage til en differentiering mellem drænedede og ikke drænedede arealer indenfor ID15 oplande. Der vil i ID15 oplande med et stort reduktionskrav til udledningen være et særligt behov for en mere differentieret opdeling af arealer efter retentionspotentiale. Her vil en differentiering mellem drænedede og ikke drænedede arealer bidrage til at kvalificere en mere detaljeret kortlægning af strømningssveje (Figur 2). Jo mere detaljeret skala, regulering målrettes mod, jo mere øges kravet til sikkerheden i drænkortlægningen.

Det er væsentligt at påpege, at en drænkortlægning alene ikke muliggør N-retentionskortlægning på meget detaljeret skala, fx markblok. Analyser af danske drænastrømningsdata for 19 mindre systemdrænedede drænoplande (jordtype JB5, JB6 og JB7) viser, at andelen af vinternedbør, der afstrømmer via dræn gennemsnitligt varierer mellem 10 og 90% (Iversen og Kjærgaard, 2013). Afstrømningen via dræn afhænger, foruden klimatiske forhold, af drænoplandets hydro-topografiske

placering indenfor oplandet samt jordtype og underliggende geologi. På basis af erfaringer med modellering af drænaflow på lokal skala er det konstateret, at den begrænsende faktor i forhold til prædiktionsforhold på lokal skala er den rumlige opløsning af geologiske data (Refsgaard, 2011). En bedre beskrivelse af transporten mellem mark og recipient forudsætter foruden en kortlægning af drænedede arealer også en forbedret kortlægning af både over- og underjorden.



Figur 2. En videreudvikling af N-retentionskortet kan baseres på en sekventiel detaljeret kortlægning, hvor ID15 oplande med lav retentionsklasse differentieres i drænedede og ikke drænedede arealer. En yderligere differentiering af drænedede arealer i forhold til type og omfang af drænaflow kræver bl.a. implementering af ny viden omkring drænaflow.

Opsummering:

- Forbedret lokal beskrivelse af vandets transportveje (dræn/ikke dræn) kan bruges i en mere kvalificeret underopdeling af ID15 oplande.
- Opbygning af videngrundlag til forbedret prædiktionsforhold på lokal skala vil være afgørende for en fremtidig mere detaljeret N-retentionskortlægning og muligheden for at optimere lokalt målrettede virkemidler.

5. Anbefalinger

- Det anbefales at arbejdet med GIS-baseret metode og verifikation, jævnfør projektforslaget i afsnit 3, iværksættes snart.
- I takt med at retentionskortet opdateres fra 2014 og frem, giver det mulighed for at prioritere arbejdet i forhold til den landsdækkende kortlægning af markdræn. Ud fra ID15 oplandene og deres retentionsklasser kan indsatsen fokuseres på oplande med betydelige reduktionskrav.

6. Referencer

- Adhikari, K., R.B. Kheir, P.K. Böcher, B.P. Malone, B. Minasny, A.B. McBratney og M.H. Greve (2013). High-resolution 3-D mapping of soil texture in Denmark. *Soil Sci. Soc. Am. J.* 77:860-876.
- Allred, B.J., N.R. Fausey, L. Peters, C. Chen, J.J. Daniels og H. Youn (2004). Detection of buried agricultural drainage pipe with geophysical methods. *Appl. Eng. Agric.*, 20:307-318.
- Iversen, B.V. og Kjærgaard, C. (2013). Analyse af danske drænaftstrømningsdata. Igangværende analysearbejde i forbindelse med iDRÆN-projektet, GUDP
- Lagzdins, A. og V. Jansons (2013). Nutrient leaching losses from subsurface drainage systems and digital drainage database in Latvia. NJF Seminar 462, Does climate change demand a new approach to drainage design?, Sarpsborg, Norge, 23-25 sept. 2013. NJF rapport Vol. 9 no. 5, [http://www.njf.nu/filebank/files/20131215\\$194520\\$fil\\$18tY1nnvqUH65d6F4Pe5.pdf](http://www.njf.nu/filebank/files/20131215$194520$fil$18tY1nnvqUH65d6F4Pe5.pdf)
Endvidere <http://www.melioracija.lv/>; zoom ind 6 gange på kortet og navigér i drænkortet.
- Mark (2014). Nye 3D jordbundskort går to meter i dybden. *Magasinet Mark*, no. 4, april 2014, s. 20-23.
- Naz, B.S., S. Ale og L.C. Bowling (2009). Detecting subsurface drainage systems and estimating drain spacing in intensively managed agricultural landscapes. *Agric. Water Management*, 96:627-637.
- Olesen, S.E. (2009). Kortlægning af potentielt drænbehov på landbrugsarealer opdelt efter landskabelement, geologi, jordklasse, geologisk region samt høj/lavbund. DJF Markbrug, rapport nr. 21. 31 sider, Aarhus Universitet.
- Refsgaard, J.C. (2011). Hvordan kan modeller hjælpe os med at vurdere nitratreduktion i undergrunden. NiCA note, November 2011
- Tetzlaff, B., P. Kuhr, H. Vereecken og F. Wendland (2009). Aerial photograph-based delineation of artificially drained areas as a basis for water balance and phosphorus modelling in large river basins. *Physics Chrm. Earth*, 34:552-564.

Bilag 1: Brug af historiske kort

Metode

Historiske kort kan anvendes som kilde til kortlægning af de dårligst (naturligt) drænedes jorder. Analysen udføres med udgangspunkt i de tre landsdækkende historiske kortværker; Original 1 (matrikelkort fremstillet i starten af 1800-tallet), de høje målbordsblade (topografiske kort fremstillet i slutningen af 1800-tallet) og de lave målbordsblade (topografiske kort fremstillet i starten af 1900-tallet). Kortene er nærmere beskrevet i Bilag 1.

De historiske kort indeholder oplysninger om udviklingen i dræningsaktiviteterne på en indirekte måde, idet landskabet på kortværkerne er opdelt i forskellige arealanvendelser. Dyrket mark har normalt ingen signatur og vil derfor kunne pløjes og dyrkes ved alm. landbrugspraksis. På de gamle kort vil en stor del af landskabet have den såkaldte sidbunds-signatur (våde arealer, eng, mose), hvilket indikerer at området har en naturlig dårlig dræning. Efterhånden som områderne blev kunstigt afvandet vha. rørdræn, kunne disse sidbundsområder inddrages i det almindelige landbrugsareal og sidbunds-signaturer forsvandt på næste generation af de historiske kort.

De historiske kort har følgende information om landskabets dræning:

Original 1 (ca. 1770-1867) kortene repræsenterer det udrænedes landskab mange steder i Danmark fra begyndelsen af 1800-tallet. Med det "udrænedes landskab" menes et landskab, der er repræsenteret ved en hydrologisk situation fra før man etablerede rørdræn, men hvor der i stort omfang var gravet grøfter på lavbundsjordene. Bemærk at Original 1 kort findes for hele Danmark (fra Kongeåen) i scannet version men ikke digitaliseret, og der er tale om offentligt tilgængelige data fra Geodatastyrelsen.

Høje målbordsblade (1842-99) repræsenterer det drænedes landskab for hele øst DK og dele af vest DK i sidste halvdel af 1800 tallet. Disse kort findes scannet og georefereret, men ikke digitaliseret.

Lave målbordsblade (1928-45) repræsenterer det drænedes landskab i hele Danmark inklusiv påvirkning af lavbundsjordene ved begyndende grundvandssænkning i forbindelse med vandindvinding til de større byer. Disse kort findes scannet og delvist digitaliseret forstået sådan at våde naturtyper er grupperet i én klasse (DJF's lavbundstema).

Brugen af de historiske kort til kortlægning af successiv dræning er illustreret for tre værkstedsområder i slutningen af dette bilag.

De tre landsdækkende historiske kortværker beskrives nærmere herunder.

Original 1 kortværket består af matrikelkort fremstillet i målestok 1:4.000 (land) og 1:800 (by), der dækker opmåling af ejendomme på landet fra 1806-1822 og fortsatte for mindre områder indtil 1867. Ældre købstæder blev matrikuleret i 1863-ca. 1880 (Korsgaard, 2006). For nærværende projekt er det imidlertid rimeligt at opfatte perioden fra 1806-1822 som oprindelsen af original 1 kortene, da original 1 kortene fra området Bjerringbro/Hvorslev omkring 1815, fra Nordfyn omkring 1805 og fra Tåstrup omkring 1816. Kortværket omfatter hele Danmark minus Slesvig, København og Frederiksberg (www.gst.dk). Original 1 kortværket vurderes at være det ældste og mest homogene kortmateriale der eksisterer til at vise udbredelse af det "naturligt" udrænedes landskab i Danmark.

De høje målbordsblade er topografiske kort i målestok 1:20.000, der blev opmålt i perioden 1842-1899, suppleret med kort i Sønderjylland fra 1921-30. Sjælland og Lolland Falster blev opmålt fra syd mod nord i 1842-62 og igen i 1887-1899, Fyn i 1862-67, Jylland fra syd mod nord i 1867-87, samt Bornholm i 1887. Kortene er udarbejdet i en periode af afvandingshistorien hvor væsentlige dele af de lerede jorder på Sjælland og øerne samt Østjylland blev rørlagt med dræn. Da opmålingen af de høje

målbordsblade startede på Sjælland og Lolland-Falster, er det ikke særlig sandsynligt, at dræningen af vådområderne i Østdanmark kan registreres i de høje målbordsblade. Da store dele af Jylland blev kortlagt sidst i optegningsperioden forventes afvandingen ved rødræning at have sat sine spor i arealfordelingen af eng og mose arealer i Vestdanmark i de høje målbordsblade. Yderligere lægger vandforsyningen til de største danske byer fra ca. 1850 om, så man i stedet for at bruge overfladevand bruger oppumpet grundvand til drikkevandsformål (Odense, 1851; Ålborg, 1854; Åbenrå, 1858; København, 1859; Århus, 1871). I 1894 var der 27 byer med eget vandværk (Barfod, 1977). Det er nok stadig begrænset, hvor meget disse nye vandværker har kunnet påvirke (afvande) vådområder, ud over de helt bynært beliggende vådområder.

De lave målbordsblade er topografiske kort, optegnet i målestok 1:20.000 og opmålt i perioden 1910-45. De lave målbordsblade viser med stor sandsynlighed udbredte effekter på det dræned landskab efter rør-dræning på Sjælland, Øerne og Østjylland. Til gengæld er der ingen forventning om at dræningen af lavbundsjordene i Jylland i perioden 1930-60 i form af de store landvindingsprojekter ved inddæmning af fjorde og tømning af søer fremgår.

Begrænsninger og usikkerheder

Ved opretning af de gamle O1 kort kan det være vanskeligt at opnå en ordentlig geografisk opretning, bl.a. fordi enkelte kort kun har få punkter som genfindes på moderne kort. Der vil ligeledes være en geografisk og personafhængig variation i brugen af vådbundssignatur fra kort til kort.

Muligheder for verifikation

Der vil være mulighed for at verificere jordernes relative vådhed ved en morfologisk beskrivelse af en række jordbundsprofiler på tværs af værkstedsområder. Denne beskrivelse bør omfatte dybde til samt type af de redoximorphe karaktertræk i jordprofilerne.

- Check mod eksisterende drænkort og arealanvendelseskort
- Check mod undersøgelser med georadar eller rekognoscering i vandløb (afsnit 2.3)

Ressourcebehov, generelt

- O1 kort: Digitalisering og georeferering af O1 kortene anslås tage ca. 1-1,5 time pr. kort (19000 kortblade)
- Høje målebordsblade: Disse kort er georefereret, og det anslås at digitaliseringen vil tage i alt ca. 3000 timer.

Referencer

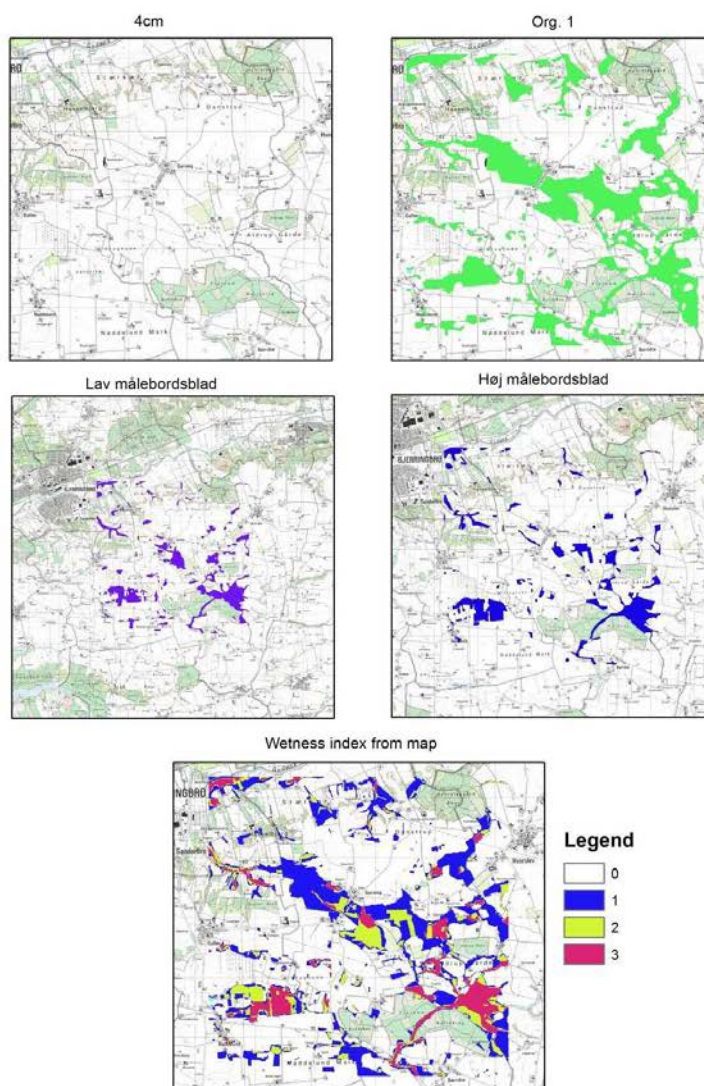
Barfod, J (1977). Vand, i W.E. von Eyben (red.). Dansk Miljøret, 257-288. Akademisk Forlag.

Korsgaard, P. (2006). Kort som kilde – en håndbog om historiske kort og deres anvendelse. Dansk Historisk Fællesråd, København.

Værkstedssområder

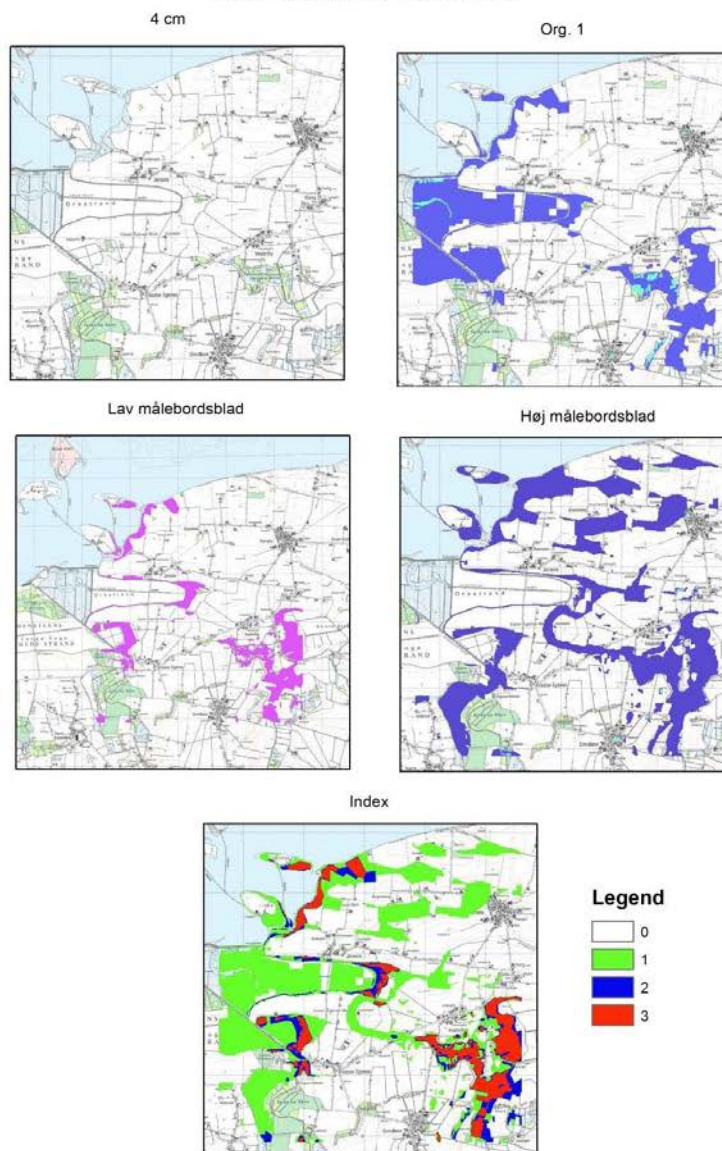
For at illustrere anvendelsen af de historiske kort som kilde til afgrænsning af de dårligst drænedede områder i landskabet, vises eksempler på kortværkerne fra tre områder i Danmark.

Kort baseret wetness

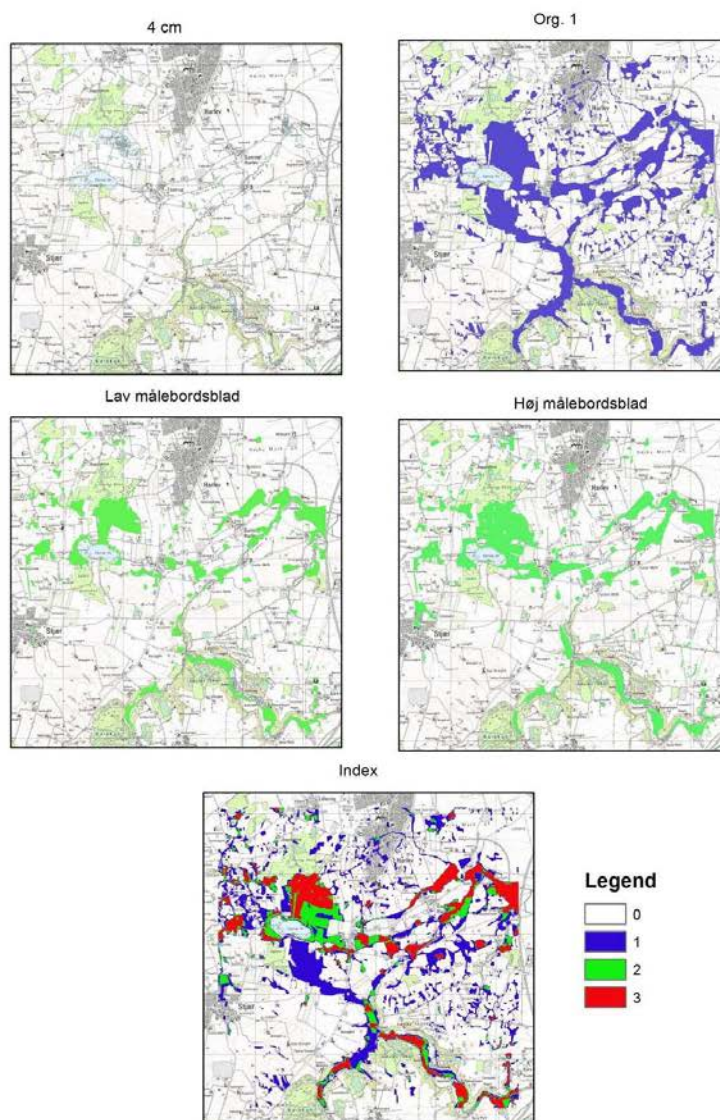


Figur B1_1. Værkstedssområde 1, Bjerringbro-Hvorslev

Kort baseret wetness



Figur B1_2. Værkstedssområde 2, Nordvestfyn.



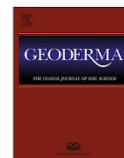
Figur B1_3. Værkstedssområde 3, Tåstrup i Østjylland.



Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma



Predicting artificially drained areas by means of a selective model ensemble



Anders Bjørn Møller*, Amélie Beucher, Bo V. Iversen, Mogens H. Greve

Department of Agroecology, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:

Denmark
Agriculture
Soil management
Water management
Geostatistics
Models

ABSTRACT

Farmers often install subsurface drainage systems to improve yields on wet soils, which has large impacts on the hydrological system. The present study uses an ensemble of machine learning models to map the extent of artificially drained areas in Denmark. The prediction is based on 745 field observations, of which one third is held out for evaluation, and 46 covariate layers. A library of 308 models is trained using 77 machine learning methods and four datasets containing either a combination of topographic variables, satellite imagery, soil properties and land use information or principal components based on these variables.

A stepwise algorithm then selects models from the library, based on their predictions on a hillclimb dataset. The results show that models trained using principal components generally yielded a better performance than the models trained with the raw covariates. Furthermore, the best results were obtained when only a random fraction of the models was available for selection at each step. The covariates that were most important for the prediction of artificially drained areas mostly related to soil properties and topography. Overall, the ensemble predicted artificially drained areas with an accuracy of 76.5%. The study demonstrates machine learning as an accurate method for mapping artificially drained areas, which is likely to benefit both farmers and decision makers.

1. Introduction

Soil drainage is a major agricultural concern, as insufficient drainage can greatly reduce crop yields (Collaku and Harrison, 2002; Ren et al., 2014; Watson et al., 1976). Farmers often respond to poor drainage conditions by installing subsurface drainage pipes in the soil. Information on the location of the pipes is often important to both farmers and environmentalists as it is necessary when new pipes are installed (Allred et al., 2004; Allred and Redman, 2010), and because the systems influence the hydrological cycle (Boland-Brien et al., 2014) and the leaching of nutrients (Ernstsen et al., 2015). The individual contractors that conduct the drainage work rarely keep shared records of the drainage systems, leading to a loss of information. Consequently, there is a strong interest in methods, which can map artificially drained areas.

Studies have shown that ground-penetrating radar can reliably locate drainage pipes at the field level, unless the soil is water saturated or has a high clay content (Allred et al., 2004; Allred and Redman, 2010).

In larger areas, studies have mapped drainage systems based on aerial photography (Northcott et al., 2000; Verma et al., 1996), an approach which later studies automated by means of image processing techniques (Naz et al., 2009; Naz and Bowling, 2008). The two later studies masked out areas without potential for artificial drainage using

a simple decision tree model based on land cover, slope and soil drainage class. Thayn et al. (2011) further developed the use of aerial photography by using photographs taken before and after a rainfall event.

Other studies have favored statistical approaches, usually covering large areas at a coarse resolution. These studies have usually mapped the probability of artificial drainage, rather than the individual drainage pipes. Behrendt et al. (2003) combined information from statistical yearbooks with a survey of expert authorities, while Hirt et al. (2005) combined maps of artificially drained areas with a map of soil classes in order to extrapolate the results.

Feick et al. (2005) mapped the percentages of artificially drained areas of the World on a 5 × 5 minute grid based mostly on international datasets. The final map contained 167 million hectares of artificially drained land globally. Sugg (2007) estimated the percentages of artificially drained areas at county level for 18 states in the USA, using maps of soil drainage classes and the extent of row crops.

Tetzlaff et al. (2009) combined the identification of artificially drained areas from aerial photography with a statistical approach. The authors first identified 2734 artificially drained fields in 231 aerial photographs from northern Germany. The authors then split the study area into 51 parts based on a number of geographic variables and calculated the percentage of artificially drained areas for each of the

* Corresponding author.

E-mail address: anbm@agro.au.dk (A.B. Møller).

<https://doi.org/10.1016/j.geoderma.2018.01.018>

Received 27 October 2017; Received in revised form 12 January 2018; Accepted 16 January 2018

0016-7061/ © 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

parts.

Machine learning present a possible further development of the statistical approaches previously applied. Numerous studies have shown that machine learning models, combining soil observations with maps of known variables, can successfully predict soil properties (McBratney et al., 2003; Scull et al., 2003). The decision to install subsurface drainage systems depends largely on the natural conditions of the soil, and in turn, they affect the surrounding soil and vegetation. It is therefore likely that an approach based on machine learning can predict the extent of artificially drained areas.

Researchers have used a large number of approaches in order to map soil properties. These include discriminant analysis (Bell et al., 1992, 1994; Kravchenko et al., 2002), artificial neural networks (Zhao et al., 2013; Zhao et al., 2008), logistic modelling (Campling et al., 2002) and decision tree analysis (Adhikari et al., 2014; Giasson et al., 2011; Hendersson et al., 2005) amongst others.

Some studies have compared methods or variations on methods in order to optimize the predictions (Giasson et al., 2011; Knotters et al., 1995; Zhao et al., 2013). However, ensembles of diverse models can often achieve a better performance than individual, optimal models (Breiman, 1996; Dietterich, 2000). Despite this finding, only few studies have combined predictions from several methods for mapping soil properties. Malone et al. (2014) tested several methods of model averaging for combining the predictions from a disaggregated conventional soil map and a regression tree model. Later, while mapping soil properties globally, Hengl et al. (2017) averaged the predictions of two machine learning models for each soil property in order to avoid overshooting effects of the individual models.

To our knowledge, no studies have combined predictions of soil properties from more than two models. There are several algorithms for creating ensembles with one type of model. Boosting (Freund and Schapire, 1996) successively adds weight to instances with incorrect predictions and builds new models using the weights, while bagging (Breiman, 1996) trains a number of independent models by drawing bootstrap samples from the training data. However, a different approach is necessary in order to combine models of different types.

Caruana et al. (2004) presented a solution in the form of the selective ensemble technique. The technique first trains a 'library' containing a large number of models based on various machine learning algorithms. An algorithm then builds an ensemble by forward stepwise selection of models from the library.

The present study uses the selective ensemble technique for the prediction of artificially drained areas in Denmark. Firstly, we test two ways of avoiding overfitting while selecting models for the ensemble. Secondly, we test an approach for reducing the prediction time of the ensemble by taking the prediction times of the models into account in the selection process.

We base our study on 745 field observations on the presence of artificial drainage and 46 environmental covariates, including soil properties, topographic variables, satellite imagery, land use information and climatic data.

2. Materials and methods

2.1. Study area

Denmark is a country in northern Europe at 54.56–57.75°N and 8.08–15.20°E (Fig. 1) with a total area of 43,000 km². The terrain is mostly weakly undulating and flat with a mean elevation of 31 m and a maximum elevation of 171 m. The dominant parent material in the eastern part of the country is loamy Weichselian moraine, while sandy deposits dominate in the western part of the country in the form of Weichselian outwash plains and Saalian moraine. The climate is temperate coastal with temperatures ranging from 1.5 °C in January to 16.3 °C in July in the period 2001–2010. The mean annual precipitation ranges from about 650 mm in the eastern part of the country to about

875 mm in the western part with a mean value of 770 mm (Wang, 2013). The main land use is agriculture, accounting for 66% of the area, while natural vegetation and urban areas make up 16% and 10% of the area, respectively (Statistics Denmark, n.d.).

Olesen (2009) estimated that approximately 50% of the agricultural area of Denmark is artificially drained. Drainage work started in 1848 and occurred mainly during two periods when the government subsidized the work. In the second half of the 19th century, drainage work focused on the loamy moraines of eastern Denmark. However, from the 1930s to the 1970s, most drainage work took place in wetland areas in the western part of the country (Hansen et al., 2004; Madsen, 2010).

2.2. Input data

In this study, the full dataset consisted of 745 point observations of the presence or absence of artificial drainage systems (Fig. 1). The observations were collected for a previous study mapping the probability of artificial drainage (Olesen, 2009). 571 observations were collected from the locations of soil profiles situated in a 7 km grid, while 174 observations were collected from additional sites without soil observations. The dataset contained 401 artificially drained locations and 344 locations without artificial drainage. We used two thirds of the observations for the training dataset and one third for evaluation. The full dataset comprised 46 covariates extracted from map layers (Table 1).

The covariate layers comprised topographic variables, data on soil texture and parent material, satellite imagery and data on land use, cropping history and climate (Table 1). We calculated the topographic variables from a digital elevation model (DEM) with a 30.4-meter grid size aggregated from a model with a 1.6-meter resolution as the mean value of the underlying 19 × 19 grid cells. The topographic variables included the depth to the groundwater interpolated from point observations and extracted from a hydrological model (Henriksen et al., 2012). We transformed the groundwater levels from the original 500-meter resolution of the hydrological model to the 30.4-meter resolution of the DEM by means of bilinear interpolation with the tool *Resample* in ArcMap.

The satellite imagery was a mosaic of Landsat 8 scenes from March 2014, which was the only month with cloud-free images in the entire study area, resampled to the 30.4-meter resolution of the DEM (NASA Landsat Program, 2014). The imagery comprised the surface reflectance from the raw bands as well as normalized indices for vegetation, soil-adjusted vegetation, moisture and water.

The study used maps of clay contents produced by Adhikari et al. (2013), aggregating the original depth intervals into four new intervals as weighed means, and a map of soil drainage classes produced by Møller et al. (2017). Soil data also included rasterized choropleth maps of geology (Jakobsen et al., 2015), landscape elements (Madsen et al., 1992) and wetland areas (Greve et al., 2014).

The land use map consisted of CORINE 2012 data (European Environment Agency, 2014). We based the cropping history on a digital field map with the farmers' registration for the common agricultural policy (The Danish Agricultural Agency, 2014). We divided the crops into categories depending on their drainage dependence, understood as the potential negative effect of water saturation on crop yields. We then counted the number of years with each category in the period 2011–2014. Drainage-dependent crops comprised mostly winter cereals while possibly drainage-dependent crops were in most cases spring cereals. Most of the drainage-independent crops were grasses for various uses. We distinguish between drainage-dependent and possibly drainage-dependent crops because Denmark has a precipitation surplus during the winter. Poor drainage conditions in the soil are therefore more likely to affect winter cereals than spring cereals.

We interpolated the precipitation data from point data by means of kriging.

For the point observations with soil profiles, we used the clay

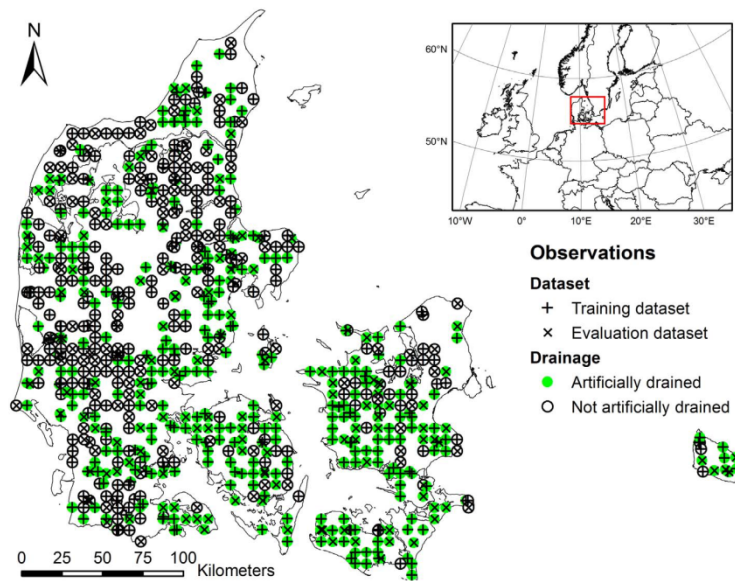


Fig. 1. Study area and input data. Observations on the presence of artificial drainage systems in Denmark from Olesen (2009) are shown along with their division into datasets used for training the statistical models and for evaluating the performance of the models and the model ensembles, respectively. The box in the upper right corner shows the location of Denmark in Europe.

contents, drainage classes, geology, landscape elements and wetland classes observed in the profiles. For the other points and other covariates, we extracted the cell values from the raster layers to the points with the tool *Extract Multi Values to Points* in ArcMap.

As some of the methods used in the study can only handle numeric variables, we converted the categorical covariates to a number of binary numeric variables equal to the number of classes, which increased the number of covariates to 73. As some of the methods are vulnerable to high-dimensional data, redundant covariates, or correlation between the covariates, we produced additional datasets using principal component analysis. We carried out the analysis on 100,000 points randomly sampled from the covariate layers and produced map layers of the principal components (PCs) and three new datasets containing the first 10, 20 and 44 PCs, respectively. The PCs in the new datasets explained 51.7%, 70.5% and 95.2% of the variance in the data, respectively (Fig. 2). Table S2 in the supplementary materials presents the weights of the covariates in the PCs.

2.3. Model library

The library included models trained with 77 different methods representing different machine learning algorithms or different implementations of the same algorithms (Table S1 in the supplementary materials). For example, the methods *rf* (Breiman et al., 2015), *ranger* (Wright and Ziegler, 2015) and *Rborist* (Seligman, 2017) are all implementations of the random forest algorithm (Breiman, 2001). We chose methods by trial and error using all classification methods in the R package *caret* (Kuhn, 2008). We omitted methods if they produced errors, did not provide class probabilities or failed to make predictions within the framework of the *clusterR* function of the R package *raster* (Hijmans et al., 2016).

Most of the methods were implementations of decision tree algorithms ($n = 29$), discriminant analysis ($n = 16$), regression analysis ($n = 15$), artificial neural networks ($n = 7$), or support vector machines ($n = 6$). The remaining methods were nearest neighbor analyses ($n = 2$), Naive Bayes ($n = 1$) and nearest centroid ($n = 1$).

For each method, we trained four models, using the datasets containing the raw covariates and 10, 20 and 44 PCs, respectively. As a

result, the library contained 308 models. Each model was optimized using 10-fold cross calculation in the *caret* R package (Kuhn, 2008) using the same resampling indices for each model. In order to measure the time necessary to produce predictions with the models, we tested the prediction time of each model in ten 19×19 km areas (Fig. 3) and averaged the times over the areas. We selected the center points of the test areas using conditioned Latin Hypercube Sampling (Minasny and McBratney, 2006) on the x - and y -coordinates of the study area in order to ensure a representative spatial distribution.

2.4. Model ensembles

Caruana et al. (2004) split the dataset into three parts: A dataset for training the models, a 'hillclimb' dataset for selecting the models, and an independent evaluation dataset. The hillclimb dataset is a separate dataset held out from the model training process. Each model calculates predictions for the hillclimb dataset, and the ensemble technique uses the predictions on the hillclimb dataset for choosing models. In each step, the technique tests how each of the models in the library affects the accuracy of the ensemble's combined predictions on the hillclimb dataset. The technique then selects the model, which provides the highest accuracy on the hillclimb dataset.

The hillclimb dataset is necessary because many algorithms have a better accuracy on the data used for training the model than on new data. Using the training dataset for model selection would therefore cause the ensemble technique to choose models, which overfit the training data. Caruana et al. (2004) suggested that researchers could reuse the training dataset in place of a separate hillclimb dataset by cross calculating the model predictions.

The study found the best results when the algorithm could only choose from a randomly sampled fraction of the models in the library at each step. If all models in the library were available, the ensemble would overfit as it repeatedly chose the same models. The results also showed that the size of this fraction had different optimal values for various problems.

Caruana et al. (2004) used the term 'validation' for the independent accuracy assessment of the ensemble. However, Oreskes (1998) argues that one cannot validate models of complex natural systems in the true

Table 1
Covariate layers for the prediction of artificially drained areas. The table lists the name of each layer, along with an explanation and its source, its mean value and the range of the values, or the number of classes for categorical variables.

Name	Explanation	Mean (min - max)/number of classes
Topographic variables		
asp_cos	Cosine of the surface aspect	0.01 (-1.00-1.00)
asp_sin	Sine of the surface aspect	-0.03 (-1.00-1.00)
Bluespot	Depth of sinks (m)	0.1 (0.0-30.8)
curv_plan	Plan curvature	0.0 (-5.1-5.3)
curv_prof	Profile curvature	0.0 (-7.3-6.1)
Demdetrend	Elevation minus the mean elevation in a 4 km radius (m)	1.0 (-57.9-98.8)
Dirinsola	Direct insolation (kWh/year)	1269 (331-1691)
Elevation	Elevation above sea level (m)	31.0 (-11.9-170.3)
Flowaccu	Number of upslope cells	63 (1-110,908)
gwd_intp	Depth to groundwater table interpolated from well observations and surface water	6.6 (0.0-111.6)
gwd_model	Depth to groundwater table from hydrological model	5.5 (0.0-108.2)
hdtochn	Horizontal distance to the nearest waterbody	235.6 (0.0-3238.4)
midslpjos	Mid-slope position	0.3 (0.0-1.0)
mvrbf	Multi-resolution index of valley bottom flatness	4.3 (0.0-10.8)
sagawi	SAGA wetness index	14.5 (6.9-19.1)
slpdeg	Surface slope gradient (degrees)	1.5 (0.0-60.4)
slptochn	Downhill gradient to the nearest waterbody (degrees)	1.1 (0.0-52.6)
twi	Topographic wetness index; Calculated as $TWI = \ln(a/\tan b)$; where a is flow accumulation, and b is local slope gradient	5.9 (0.0-54.6)
valldpth	Valley depth (m)	7.4 (0.0-89.1)
vdtochn	Vertical distance to the nearest waterbody (m)	4.1 (0.0-96.6)
Satellite imagery		
LS8_band1	Landsat 8 Band 1 surface reflectance, March 2014 (Ultra blue)	369 (-399-15,471)
LS8_band2	Landsat 8 Band 2 surface reflectance, March 2014 (Blue)	438 (-407-15,769)
LS8_band3	Landsat 8 Band 3 surface reflectance, March 2014 (Green)	655 (-261-15,843)
LS8_band4	Landsat 8 Band 4 surface reflectance, March 2014 (Red)	697 (-479-16,000)
LS8_band5	Landsat 8 Band 5 surface reflectance, March 2014 (Near infrared)	2295 (-77-15,955)
LS8_band6	Landsat 8 Band 6 surface reflectance, March 2014 (Shortwave infrared 1)	1993 (-26-16,000)
LS8_band7	Landsat 8 Band 7 surface reflectance, March 2014 (Shortwave infrared 2)	1334 (-5-17,082)
ndmi	Normalized difference moisture index; (Band 5 - Band 6)/(Band 5 + Band 6)	0.06 (-0.94-0.97)
ndvi	Normalized difference vegetation index; (Band 5 - Band 4)/(Band 5 + Band 4)	0.52 (-1.00-0.98)
ndwi	Normalized difference water index (Band 5 - Band 3)/(Band 5 + Band 3)	-0.55 (-0.97-1.00)
savi	Soil-adjusted vegetation index; (Band 5 - Band 4)/(1 + 0.5)/(Band 5 + Band 4 + 0.5)	0.3 (-0.2-0.7)
Soil and parent material		
Clay_a	Clay content, 0-30 cm (%)	8.3 (0.0-51.2)
Clay_b	Clay content, 30-60 cm (%)	10.0 (0.0-62.7)
Clay_c	Clay content, 60-100 cm (%)	11.3 (0.0-59.1)
Clay_d	Clay content, 100-200 cm (%)	11.2 (0.0-57.1)
dc	Soil drainage class	2.9 (1-5)
Geology	Scanned and registered geological map (Scale 1:25,000)	10 classes
Georeg	Scanned geographical regions map (Scale 1:100,000)	7 classes
Landscape	Landform types (Scale 1:100,000)	11 classes
Wetland	Shows the presence of non-wetlands (0), wetlands (1), central wetlands (2) and peat areas (3) (Scale 1:20,000)	0.3 (0-3)
Land use, cropping history and climate		
Crops_data	Years with crop data in the period 2011-2014	3.1 (0-4)
Crops_drain	Years with drainage-dependent crops in the period 2011-2014	1.5 (0-4)
Crops_maybedrain	Years with possibly drainage-dependent crops in the period 2011-2014	1.3 (0-4)
Crops_nodrain	Years with drainage-independent crops in the period 2011-2014	0.2 (0-4)
Landuse	CORINE land cover data adopted in Denmark (Scale 1:100,000)	3 classes
Precipitation	Mean annual precipitation in the period 1961-1990 interpolated from point data (mm)	718 (473-963)

sense of the word and suggests “evaluation” as a replacement. The presence of artificial drainage systems depends both on the natural conditions in the soil and on human agency over a period of more than a hundred years. We therefore believe that the statement is relevant for this study and adopt the term ‘evaluation’.

The present study used a modified version of the technique described by Caruana et al. (2004) to select models from the library. As the probabilities on the training dataset were predicted using cross evaluation, it was possible to use them for the hillclimb dataset. In each iteration, a randomly sampled fraction *f* of the models in the library were available for selection. The algorithm tested each of these models by adding it to the ensemble and averaging the probabilities of artificial drainage predicted by the model in question and the models already in the ensemble. It then transformed the probabilities into classes, using a probability of 0.50 as the cutoff value, and calculated the performance. The algorithm finally selected the model, which yielded the best

performance for the ensemble. We calculated the performance on the hillclimb and evaluation datasets as kappa (K):

$$K = 1 - \frac{1 - p_o}{1 - p_e} \tag{1}$$

$$p_o = \frac{1}{N} \sum_{i=1}^m E_{ii} \tag{2}$$

$$p_e = \frac{1}{N^2} \sum_{i=1}^m n_{ij} n_{ji} \tag{3}$$

where *p_o* is the observed agreement between observations and predictions, *p_e* is the hypothetical probability of chance agreement, *N* is the total number of observations, *m* is the number of classes, *E_{ii}* is the sum of diagonal elements in the confusion matrix, *n_{ij}* is the number of observations in row *i*, and *n_{ji}* is the number of observations in column *i*.

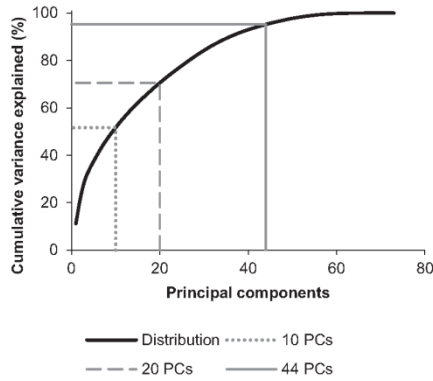


Fig. 2. Cumulative variance explained by the principal components (PCs) calculated from the covariate layers versus the number of PCs. The thick line shows the cumulative distribution. Thin lines mark the cumulative variance explained by the first 10 (dotted line), 20 (dashed line) and 44 PCs (solid line).

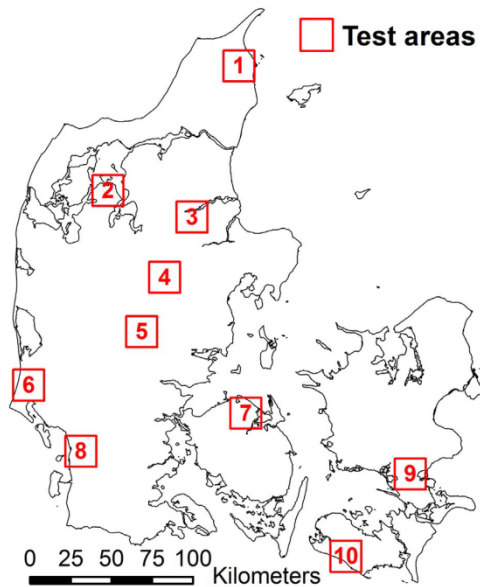


Fig. 3. The location of the ten test areas used to measure the prediction time of each model in the library. The labels show the ID of each test area.

We tested two modifications to the technique. In the first modification, we bagged the hillclimb dataset in each iteration by sampling the original hillclimb dataset with replacement. In this way, the hillclimb dataset becomes a moving target, which might alleviate the tendency to overfitting described by Caruana et al. (2004).

In the second modification, the algorithm divides the performance by the sum of the prediction times of the models in the ensemble as measured in the test areas. The original technique only takes the accuracy of the predictions into account. However, the models that provide the best accuracies may be computationally heavy, leading them to require a long span of time to produce a map. Implementing prediction times in the model selection process should give an advantage towards models, which give a good prediction in a short span of time.

We tested both modifications separately and in combination for a

total of four experiments:

Experiment 1: Unaltered hillclimb dataset; simple performance.

Experiment 2: Bagged hillclimb dataset; simple performance.

Experiment 3: Unaltered hillclimb dataset; performance divided by prediction time.

Experiment 4: Bagged hillclimb dataset; performance divided by prediction time.

In each experiment, the fraction of models available for selection (f) varied from 0.05 to 1.00 in 20 steps. For each value of f , we created 100 ensembles with 100 iterations each and averaged the results.

The evaluation during the experiments comprised the performance of the ensembles after 100 iterations as well as the performance that they reached before the combined prediction times exceeded 300 s as measured in the test areas. We then produced a final ensemble using the approach that provided the best performance within 300 s and used it to predict artificially drained areas for the non-urban, non-forested land area of Denmark.

We chose 300 s as the maximum threshold for the prediction times for the test areas, as we assumed that the prediction times were proportional to the size of the areas. Given this assumption, the maximum prediction time for Denmark, with an area of 43,000 km² should be approximately 10 h, which we considered as the allowable maximum for the final ensemble.

The evaluation of the final ensemble also included simple accuracy (percent correctly predicted) in addition to K, as it is more easily interpreted.

2.5. Covariate importance

We calculated covariate importance measures for the PCs and the raw covariates. For the models in the final ensemble, which were able to provide covariate importance measures, we scaled the importance provided by the models to 100 for the most importance covariate.

In order to measure the importance of the original covariates of the models trained using PCs, we multiplied the importance of each PC with the absolute weight of each of the covariates for the PC in question. We then calculated the sum of these products for each covariate over the PCs and scaled them to 100 for the most important covariate. We repeated the procedure for each of the models.

3. Results

3.1. Model library

Most of the models in the ensemble had prediction times of < 60 s while a few had prediction times of several minutes (Fig. 4A). The models trained using PCs had shorter prediction times than the models trained with the raw covariates. The prediction times generally decreased with the number of PCs. The K achieved on the hillclimb and evaluation datasets was in the range of 0.3–0.5 for most models. A minority of the models had a poorer performance and very few had a better performance (Fig. 4B and Fig. 4C). The models trained on datasets with PCs generally had a higher performance than the models trained on the dataset with raw covariates. Only seven methods achieved their best performance with the raw covariates. On the other hand, 20 methods reached their best performance with 10 PCs, and 35 methods had their best performance with 20 PCs. Only 15 methods achieved their best performance with 44 PCs. Model performance on the evaluation dataset correlated with performance on the hillclimb dataset ($R^2 = 0.66$, $n = 308$, $p < 0.05$).

There was generally little variation in the prediction times between the test areas for each model. The coefficient of variation between the test areas varied from 0.4% to 42.6% with a mean value of 4.4% for the individual models in the library.

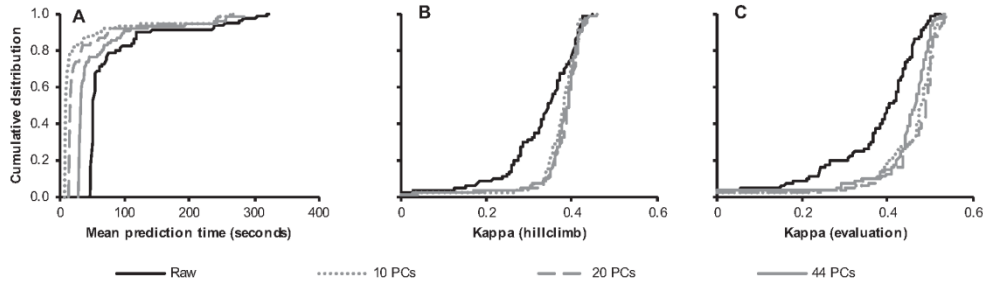


Fig. 4. Cumulative distribution of the models trained on the dataset with raw covariates (Raw) and on the datasets containing 10, 20 and 44 principal components (PCs), respectively, versus their mean prediction time on the test areas (A) and their predictive performance on the hillclimb dataset (B) and the evaluation dataset (C).

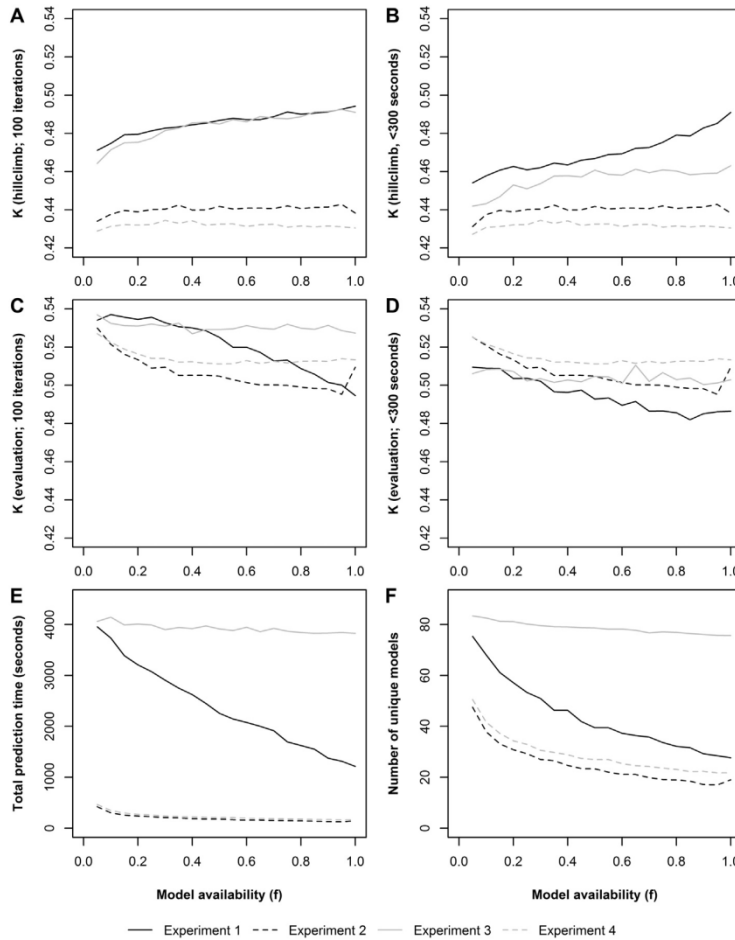


Fig. 5. The predictive performance, total prediction times and number of models in the ensembles created in each experiment versus the model availability parameter f . Lines are averages of 100 repetitions for each experiment. A: Performance achieved on the hillclimb dataset after 100 iterations in each experiment. B: Performance achieved on the hillclimb dataset within 300 s in each experiment. C: Performance achieved on the evaluation dataset after 100 iterations in each experiment. D: Performance achieved on the evaluation dataset within 300 s in each experiment. E: Mean prediction time of the full ensembles created in each experiment. F: Number of unique models in the ensembles after 100 iterations.

3.2. Ensemble testing

In each experiment, the performance on the hillclimb dataset generally improved with higher values of f , while the performance on the evaluation dataset deteriorated (Fig. 5). The ensembles created in Experiment 1 achieved the best performance on the hillclimb dataset

followed by the ensembles created in Experiment 3. This was the case both for the full ensembles and with a prediction time threshold of 300 s. The performance of the ensembles created in Experiment 3 and Experiment 4 depended less on f than the performance of the ensembles created in Experiment 1 and Experiment 2.

With 100 iterations, the ensembles created in Experiment 1 and

Experiment 3 achieved the best performance on the evaluation dataset (Fig. 5C). However, when using a prediction time threshold of 300 s, the best performance was achieved in Experiment 2 and Experiment 4. Within 300 s, the ensembles created in Experiment 4 obtained the best performance on the evaluation dataset across most values of f . At $f = 0.05$, the ensembles created in Experiment 2 and Experiment 4 had nearly the same performance.

The ensembles created in Experiment 3 had the longest total prediction times across all values of f , followed by the ensembles produced in Experiment 1 (Fig. 5E). The total prediction times of the ensembles of Experiment 1 decreased sharply with higher values of f . The ensembles created in Experiment 2 and Experiment 4 had short total prediction times across all values of f . Likewise, the ensembles created in Experiment 3 contained the largest number of unique models across all values of f followed by the ensembles created in Experiment 1 (Fig. 5F). The ensembles created in Experiment 2 and 4 contained a relatively low number of models, but their prediction times were still short compared to the number of models.

3.3. Final ensemble

We used the technique of Experiment 2 and a prediction time threshold of 300 s to create the final ensemble. The performance achieved within 300 s was similar in Experiment 2 and 4, but we chose the technique of Experiment 2 as it was the simplest approach. The resulting ensemble contained 36 models chosen over 82 iterations (Fig. 6). The performance on the hillclimb dataset improved steadily but not uniformly with the number of iterations. On the other hand, the performance on the evaluation dataset improved over the first 20 iterations with some degree of noise followed by stabilization. The final performance on the evaluation dataset was $K = 0.526$ and a simple accuracy of 76.5%. The ensemble chose some models several times, and it therefore contained 36 unique models although the number of iterations was higher.

The final ensemble contained models trained using a wide array of methods (Table 2). On the other hand, the dataset containing 10 PCs was by far the most frequent dataset in the ensemble. The ensemble contained only two models trained on the dataset with 20 PCs. There was only moderate correlation between the prediction times in the test area and the prediction times for the total area ($R^2 = 0.45$, $n = 36$, $p < 0.05$). The total prediction time for the entire study area was 1 h and 35 min.

There was a clear relationship between the standard deviation (SD) of the probabilities of artificial drainage predicted by the models in the

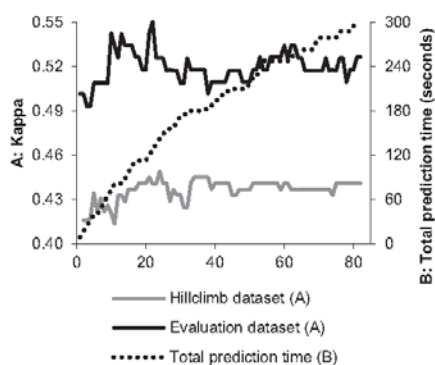


Fig. 6. Performance and prediction time of the final ensemble. A: Predictive performance on the hillclimb dataset, used for selecting the models and evaluation dataset versus the number of iterations in the final ensemble. B: Prediction time of the final ensemble calculated as the sum of the mean prediction times of the constituent models versus the number of iterations.

ensemble and the mean probability of artificial drainage (Fig. 7). At mean probabilities of approximately 0.3 and 0.7, the SD was generally low. The SD was generally larger at an intermediate mean probability of 0.5 and towards the extremes of the probability spectrum.

The relationship between the accuracy of the predictions and the probability of artificial drainage was very clear, as the predictions were generally most uncertain for cases with an intermediate probability (Fig. 8A). In the probability interval from 0.4 to 0.6, the accuracy was approximately 60%, while it was much higher for cases with lower or higher probabilities of artificial drainage. On the other hand, there was no clear relationship between the SD of the probabilities and the accuracy of the predictions (Fig. 8B).

Most of the models in the ensemble had short prediction times. The models in the library had a mean prediction time in the test areas of 49.4 ± 60.9 s (± 1 SD), while the models in the ensemble had a mean prediction time of 8.2 ± 1.7 s (± 1 SD). However, the models in the ensemble had a similar predictive performance to the models in the library. The mean value of K was 0.43 ± 0.11 (± 1 SD) on the evaluation dataset for the models in the library and 0.44 ± 0.10 (± 1 SD) for the models in the ensemble. The final ensemble had a higher predictive performance on the evaluation dataset than the best model in the ensemble ($K = 0.53$ vs. 0.51).

The models in the ensemble displayed a higher degree of internal correlation than the models in the library. Internal R^2 for the probabilities of artificial drainage predicted on the evaluation dataset was 0.71 ± 0.22 (± 1 SD) for the models in the library and 0.77 ± 0.21 (± 1 SD) for the models in the ensemble.

Six of the models in the library achieved a better performance on the evaluation dataset than the final ensemble. However, these models generally had an intermediate performance on the hillclimb dataset, as only two of them were in the top 100 models with the best performance. Furthermore, only one of the models was in the top 50. This model reached the third best performance on the evaluation dataset and the sixth best performance on the hillclimb dataset. Its method was a conditional inference forest algorithm (*cforest*) (Hothorn et al., 2017) and its input data contained 44 PCs. Its mean prediction time in the test areas was 193 s.

3.4. Covariate importance

Only 15 of the models in the final ensemble could provide variable importance measures (Table 2). For these models, PC3 was invariably the most important variable for the prediction followed by PC1 (Table 3). PC3 mostly related to the clay content of the soil, the presence of wetlands and the depth to the groundwater. On the other hand, PC1 mostly related to the satellite-derived covariates (Supplementary Table S2).

The clay contents of the soil at various depths were the most important original covariates for the prediction of artificially drained areas followed by the precipitation. Both layers of the depth to the groundwater proved to be important for the prediction, as was soil drainage class and wetland areas. The categorical covariates mostly related to soil texture. For example, the geological classes *clayey till* and *glaciofluvial sand* relate to clayey and sandy soils, respectively. The Normalized Difference Moisture Index (ndmi) was the only Landsat 8-derived covariate in the top 20.

3.5. Mapping

The probability maps of the individual models in the final ensemble generally showed the same patterns, although there were marked differences (Fig. 9). The SD of the probabilities is generally low, while a few areas have very high SDs amongst the probabilities.

The final ensemble mostly predicted artificially drained areas on the loamy tills of eastern Denmark and in the large wetlands areas of northern Denmark (Fig. 10). In total, the ensemble predicted artificially

Table 2

Models in the final ensemble for predicting artificially drained areas. The method and training data of each model, its weight in the ensemble, the number of the iteration at which it was first chosen, its mean prediction time in the test areas, and its prediction time for the total area is shown. The models are sorted firstly by weight (descending order), secondly by first chosen (ascending order).

Method	Training data	Weight	First chosen (iteration)	Prediction time, test areas (seconds)	Prediction time, total area (seconds)	Covariate importance
LogitBoost	10 PCs	5	2	7.2	131	No
xgbTree	10 PCs	5	14	7.1	119	Yes
gcvEarth	10 PCs	4	3	8.0	163	Yes
fda	10 PCs	4	4	7.6	215	Yes
glmStepAIC	10 PCs	4	5	7.6	158	No
lda2	10 PCs	4	22	7.6	187	No
svmLinear2	10 PCs	4	25	7.3	151	No
rpart2	10 PCs	4	50	7.2	121	Yes
sda	10 PCs	3	17	7.4	152	No
nnet	10 PCs	3	21	7.1	117	Yes
gbm	10 PCs	3	24	7.3	122	Yes
pda2	10 PCs	3	30	7.7	220	No
evtree	10 PCs	3	32	7.2	116	No
rf	10 PCs	3	51	8.5	191	Yes
dwdLinear	10 PCs	2	7	7.7	151	No
bdk	10 PCs	2	9	8.0	177	No
ctree2	10 PCs	2	10	7.9	123	No
rpart	10 PCs	2	27	7.1	116	Yes
dnn	10 PCs	2	29	8.1	186	No
rpart1SE	10 PCs	2	38	7.1	117	Yes
pam	10 PCs	2	40	7.3	175	Yes
pls	10 PCs	2	54	7.7	174	Yes
C5.0	10 PCs	1	1	9.1	374	Yes
C5.0Tree	10 PCs	1	8	9.1	374	Yes
kknn	10 PCs	1	11	9.2	230	No
ctree	10 PCs	1	15	8.0	122	No
bayesglm	10 PCs	1	16	8.5	180	No
xyf	10 PCs	1	23	8.0	180	No
sparseLDA	10 PCs	1	42	7.2	201	No
knn	10 PCs	1	44	8.3	172	No
rf	20 PCs	1	53	14.8	351	Yes
ctree	20 PCs	1	63	14.8	289	No
pcaNNet	10 PCs	1	69	8.6	167	No
simpls	10 PCs	1	70	7.7	169	Yes
blackboost	10 PCs	1	76	8.5	138	No
lda	10 PCs	1	80	7.7	185	No

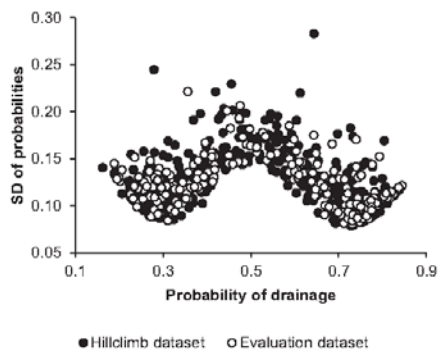


Fig. 7. Relationship between the mean probabilities of artificial drainage, predicted by the models in the final ensemble, and the standard deviations (SD) of the probabilities between the models for the hillclimb dataset (closed symbols) and the evaluation dataset (open symbols).

drained areas in 49.4% of the study area. The SD of the probabilities predicted by the models was smallest on the Weichselian moraine in eastern Denmark and in the outwash plains and Saalian moraine of western Jutland. The SD between the probabilities was largest in small areas such as wetland areas, especially the marsh areas of southwestern Denmark.

4. Discussion

4.1. Model library

The models trained with PCs had shorter prediction times and a better predictive performance than the models trained with the raw covariates. The shorter prediction times are most likely due to the lower dimensionality of the datasets with PCs, which makes the models less computationally expensive. This also explains why the dataset containing only 10 PCs resulted in the lowest prediction times. On the other hand, the higher predictive performance is most likely due to the vulnerability of many methods to redundant or correlated covariates.

We converted all categorical variables to numeric variables to ensure that all methods could use the data. However, some machine learning methods use categorical covariates efficiently. For example, the C5.0 decision tree algorithm can make multiway splits based on categorical covariates (Kim and Loh, 2001).

In this study, we chose methods for the model library by trial and error, as it was a simple and efficient approach. It is possible that we could have built a better model library if we had chosen methods in a more objective manner. This question is a potential object for future studies. Furthermore, it is possible that some of the methods that we discarded would have been useful to the ensemble if we had been able to implement them.

Caruana et al. (2004) used a library with thousands of models, while in this study it contained only 308 models. In addition, in this study we used only the optimized models. The optimization process discards several models for each method, as they do not have the best

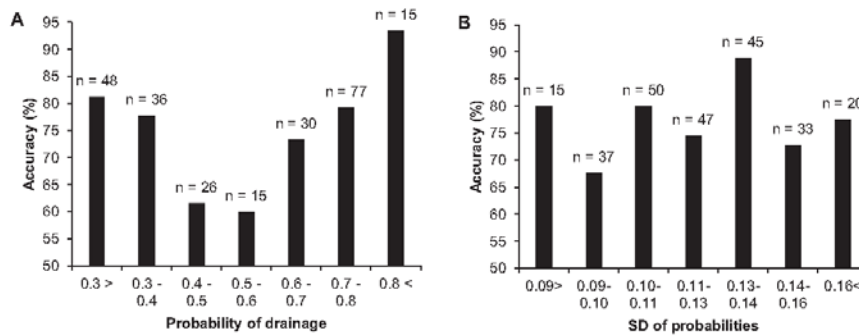


Fig. 8. Accuracy of the predictions for subsets of the evaluation dataset. A) Accuracy for subsets based on the mean probability of artificial drainage predicted by the final ensemble. B) Accuracy for subsets based on the standard deviation (SD) between the probabilities predicted by the models in the final ensemble. The divisions in B are based on the sequence $1 \cdot 10^i$ (0.05^i), where i is an integer, as the SDs of the probabilities are skewed. Labels above the columns show the size of the subsets.

Table 3

Importance of principal components (PCs) and covariates. The table gives the mean importance of the PCs in the models in the final ensemble, which provided variable importance measures as well as the range of the importance measures for each PC. The table also gives the weighed importance of the 20 most important covariates. The weighed importance is the sum of the absolute weights of each covariate over the PCs weighed by the mean importance of each PC.

Rank	PC	Importance; Mean (min - max)	Covariate	Importance; Mean (min - max)
1	PC3	100.0 (100.0–100.0)	Clay_c	98.8 (88.9–100.0)
2	PC1	37.1 (0.0–73.0)	Clay_b	98.8 (90.5–100.0)
3	PC4	14.8 (0.0–36.7)	Clay_a	98.0 (91.6–100.0)
4	PC8	12.9 (0.0–35.2)	Clay_d	95.2 (85.4–98.2)
5	PC10	11.8 (0.0–44.0)	Precipitation	88.4 (74.2–97.6)
6	PC7	10.3 (0.0–47.7)	Georegl3 (Western Jutland)	80.6 (64.0–100.0)
7	PC6	9.8 (0.0–33.2)	dc	74.4 (69.1–78.6)
8	PC2	4.9 (0.0–19.7)	GeologyL7 (Clayey till)	71.0 (57.7–75.5)
9	PC9	3.1 (0.0–24.7)	GeologyL8 (Sandy till)	70.5 (52.4–87.3)
10	PC5	0.8 (0.0–6.8)	gwd_model	69.1 (54.9–77.9)
11	PC11	0.5 (0.0–19.0)	LandscapeL9 (Glacial outwash plain)	69.1 (42.6–100.0)
12	PC20	0.2 (0.0–8.7)	GeologyL10 (Glaciofluvial sand)	68.5 (44.7–96.7)
13	PC13	0.2 (0.0–7.6)	LandscapeL10 (Saalian moraine)	67.4 (42.4–88.7)
14	PC12	0.2 (0.0–6.7)	gwd_intp	67.2 (49.1–78.2)
15	PC15	0.1 (0.0–4.2)	vtochn	64.9 (51.2–70.4)
16	PC18	0.1 (0.0–3.3)	Elevation	63.9 (50.1–81.0)
17	PC19	0.1 (0.0–3.2)	ndmi	61.6 (48.9–71.2)
18	PC14	0.0 (0.0–1.7)	Demdetrend	61.2 (36.9–83.4)
19	PC16	0.0 (0.0–0.7)	Wetland	60.4 (41.5–73.4)
20	PC17	0.0 (0.0–0.0)	Georegl8 (Eastern Denmark)	57.1 (49.2–67.0)

performance for the method in question. It is possible that some sub-optimal models would be useful in an ensemble, as ensembles thrive on dissimilar models (Breiman, 1996; Dietterich, 2000).

Caruana et al. (2004) also split the bagged decision tree models into their constituent decision trees and allowed the algorithm to choose the individual decision trees for the ensemble, which increased the number of models available for selection.

4.2. Ensemble testing

The performance on the hillclimb and evaluation datasets showed clearly distinct patterns, as the performance on the hillclimb dataset

generally increased with f while the performance on the evaluation dataset deteriorated. Caruana et al. (2004) found that this dissimilarity was due to overfitting with high values of f . The results in this study confirm the unsuitability of the hillclimb dataset for assessing the performance of the ensemble.

Another difference in the performance on the hillclimb and evaluation dataset appeared in the performance achieved within a total prediction time of 300 s. In this case, the performance on the hillclimb dataset was highest for the ensembles without prediction times implemented. On the other hand, the performance on the evaluation dataset was highest with prediction times implemented. This shows clearly that, by focusing both on performance and prediction times, the implementation of prediction times allowed the ensembles to achieve a good performance within a short span of time. However, the chosen threshold of 300 s is arbitrary and the results might have been different, if the threshold had been higher or lower. In fact, the total prediction time of the final ensemble for the entire study area was much lower than the maximum threshold of 10 h. This shows that we could have chosen a higher threshold for the test areas without exceeding the threshold for the entire study area.

Bagging the hillclimb dataset improved the performance across most values of f both with and without prediction times implemented. This finding shows that bagging the hillclimb dataset worked as intended, reducing the tendency to overfitting and lowering the importance of adjusting f . However, it ultimately proved unnecessary, as the best performance within 300 s was the same with and without bagging the hillclimb dataset. We cannot rule out the possibility that bagging the hillclimb dataset would be useful in other circumstances, but in this study, adjusting f was sufficient.

4.3. Final ensemble

The models in the final ensemble represented a large variety of methods, but nearly all the models used the dataset with only 10 PCs as input data. The mean performance of the models in the ensemble was on par with the mean performance of the models in the library as a whole, but their prediction times were substantially lower. The prevalence of models trained with only 10 PCs is probably due to their short prediction times relative to their performance. The diversity of the chosen models shows that the algorithm avoided overfitting by choosing dissimilar models. Nevertheless, the probabilities predicted by the models in the ensemble displayed a higher degree of internal correlation than the models in the library as a whole. The use of the same dataset in the training of nearly all the models in the ensemble probably explains this. To some extent, the results are a compromise between model diversity and short prediction times.

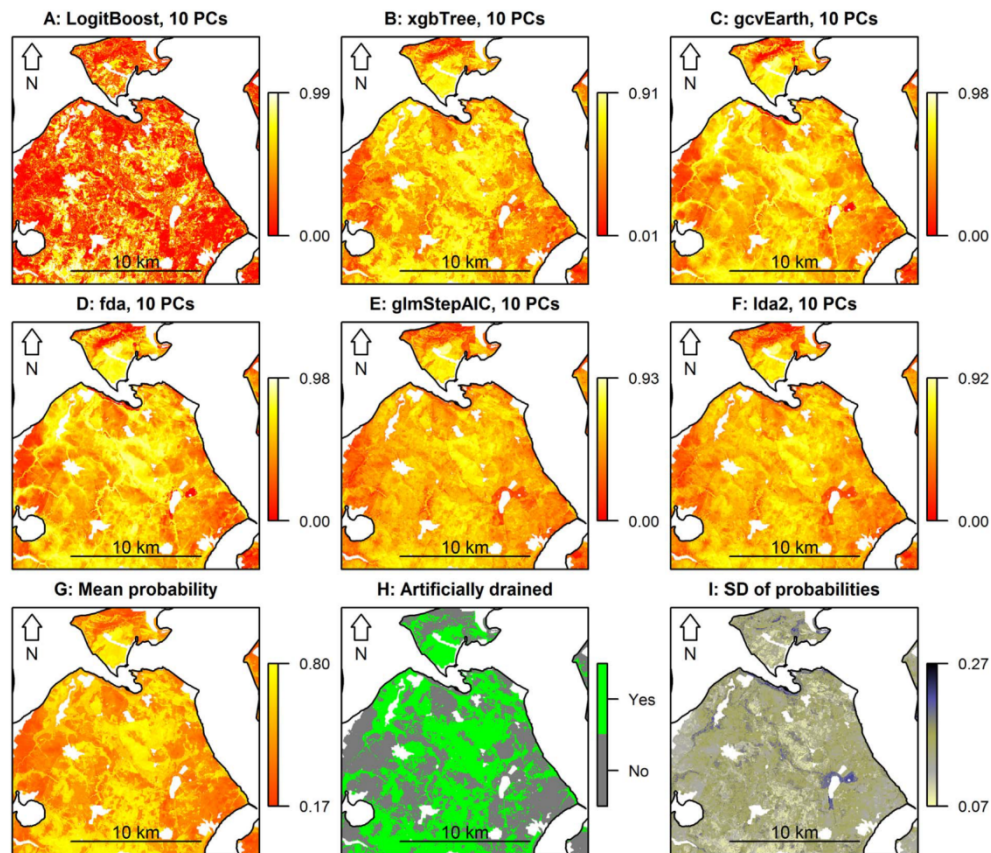


Fig. 9. Example of the prediction of artificially drained areas from Test Area 2. A–F show the probability of artificial drainage predicted by six of the models in the ensemble all trained using the dataset containing the first 10 principal components (PCs). A: Boosted logistic model (LogitBoost). B: Decision tree model with extreme gradient boosting (xgbTree). C: Multivariate adaptive regression splines with generalized cross validation (gcvEarth). D: Flexible discriminant analysis (fda). E: Generalized linear model with stepwise feature selection (glmStepAIC). F: Linear discriminant analysis (lda2). G: Mean probability of artificial drainage predicted by the final ensemble. H: Predicted extent of artificially drained areas. I: Standard deviation (SD) of the probability of artificial drainage predicted by models in the final ensemble.

There was only moderate correlation between the prediction times for the test areas and the total area for the models in the final ensemble. The ensemble is not a representative sample of the model library, and it may explain the absence of a strong correlation. However, at the same time there was little variation in the prediction times between the individual test areas for each model. It is therefore unlikely that a higher number of test areas would improve the prediction times from the test areas as a proxy for the prediction times for the total area. In fact, it is likely that only a single test area would be necessary for the purpose.

Future studies may want to adjust the importance of prediction times in the selection of models. A simple solution would be an exponential transformation for a higher importance or a logarithmic transformation for lower importance.

Overall, the final ensemble had a better performance than nearly all models in the library. Furthermore, the only model that had a consistently better performance than the ensemble also had a very long prediction time of 193 s. This is a shorter prediction time than the total prediction time of the ensemble. However, the fact that only a highly complex model with a long prediction time can achieve a better predictive performance than the ensemble affirms the viability of the ensemble technique.

The complexity of the models or the ensemble necessary to achieve

a good prediction of artificially drained areas casts doubt on the efficacy of very simple decision tree models in the prediction of artificially drained areas. Numerous studies have used these simple models to map artificially drained areas or mask out areas without need of artificial drainage. In fact, Naz et al. (2009) found flaws in the results produced by the decision tree model of the same study. The authors found that artificial drainage systems were present in areas, which should not require them and absent in areas where they should have been necessary according to the model.

There was a clear relationship between the mean probability of artificial drainage predicted by the models in the ensemble and the accuracy of the prediction. The map of the mean probability can therefore serve as an indicator of the reliability of the prediction. On the other hand, the SD between the probabilities predicted by the models in the ensemble was largely unrelated to the accuracy of the prediction. It is therefore not suitable for this purpose.

On the other hand, some of the larger deviations between the probabilities predicted by the models may be results of differences in the ways that the methods calculate class probabilities. For example, decision tree methods usually calculate class probabilities based on the end nodes in the tree, while logistic regression calculates the class probabilities as a continuous function.

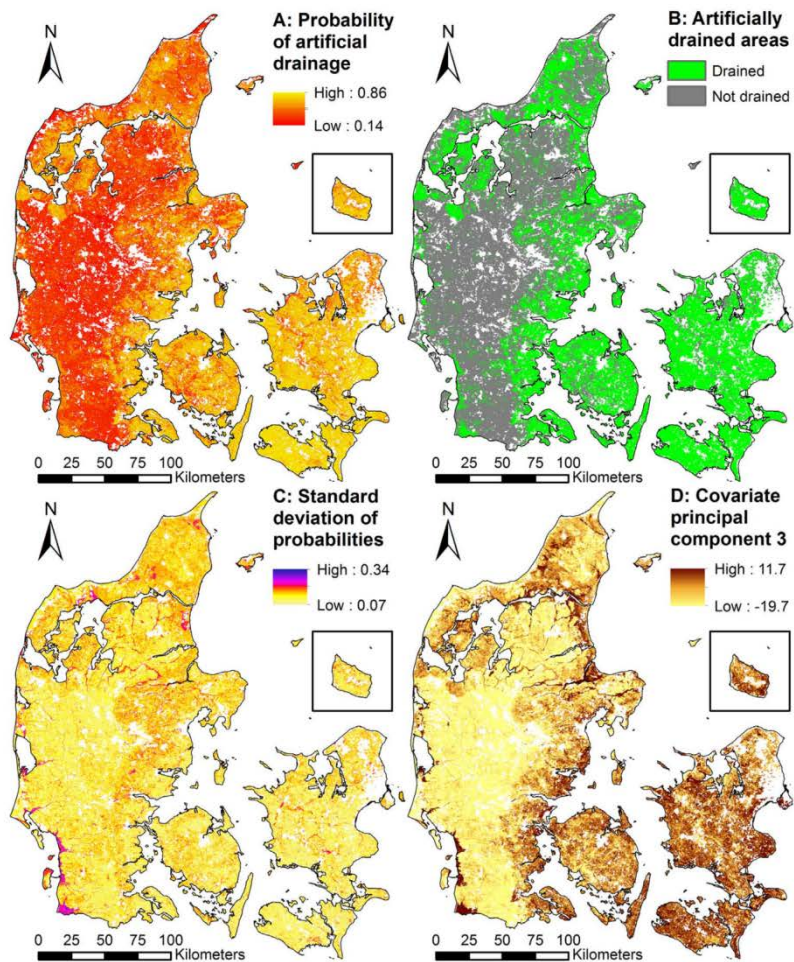


Fig. 10. Ensemble predictions for the study area. A: Mean probability of artificial drainage, predicted by the models in the final ensemble. B: Prediction of artificially drained areas from the final ensemble. C: Standard deviation (SD) of the probability of artificial drainage predicted by the models in the final ensemble. D: Principal component (PC) 3 of the covariate layers; the most important variable for the prediction of artificially drained areas.

The largest SDs between the probabilities predicted by the models in the final ensemble appear in small areas, while the smallest SDs appear in large, relatively uniform areas. This pattern suggests that the largest SDs appear in areas with sparse representation in the training data. However, this circumstance does apparently not affect the reliability of the predictions in these areas.

4.4. Covariate importance

Although only a fraction of the models in the ensemble could provide covariate importance measures, it appears that many of the most important covariates in this study were also important in other studies. Many of the studies mapping artificially drained areas used the soil drainage class to some extent (Gökkaya et al., 2017; Hirt et al., 2005; Naz et al., 2009; Naz and Bowling, 2008; Northcott et al., 2000; Sugg, 2007; Tetzlaff et al., 2009; Verma et al., 1996). In this study, it was the seventh most important covariate.

Studies have also used other soil properties in the mapping, such as soil type, soil texture and geology (Behrendt et al., 2003; Hirt et al., 2005; Tetzlaff et al., 2009; Verma et al., 1996). In the present study, we did not use the soil type directly. However, we used a number of parameters relating to the soil, such as clay content, geology, landscape

elements and georegions, which all proved to be important for the prediction.

Despite the frequent use of aerial and satellite imagery in other studies (Gökkaya et al., 2017; Naz et al., 2009; Naz and Bowling, 2008; Northcott et al., 2000; Tetzlaff et al., 2009; Thayn et al., 2011; Verma et al., 1996), only one of the satellite-derived products were amongst the 20 most important covariates in this study. This covariate, ndmi, represents the relationship between near infrared and shortwave infrared reflectance, which corresponds well with the importance of color and shortwave infrared images in other studies (Gökkaya et al., 2017; Northcott et al., 2000; Thayn et al., 2011; Verma et al., 1996).

Although precipitation and depth to the groundwater were important in this study, only Tetzlaff et al. (2009) used these two covariates for mapping artificially drained areas. On the other hand, the surface slope had only intermediate importance in this study, ranking 28 amongst the covariates, while several studies used it for mapping artificially drained areas (Gökkaya et al., 2017; Naz et al., 2009; Naz and Bowling, 2008; Tetzlaff et al., 2009; Thayn et al., 2011). Indeed, not counting the depth to the groundwater, only three topographic covariates were important in this study: the elevation, the detrended elevation and vertical depth to the channel network. Of these, only the elevation played a role in another study (Northcott et al., 2000).

5. Conclusion

The application of the selective ensemble technique was overall quite successful, as it provided a highly accurate prediction of artificially drained areas within a feasible span of time. In terms of accuracy, the results were on par with or better than the best individual machine learning models in the study.

Bagging the hillclimb dataset showed promising results, as it improved the predictive performance of the ensembles in nearly all cases. However, adjusting the size of the randomly sampled fraction of models available for selection ultimately proved a sufficient procedure for avoiding overfitting. Nevertheless, bagging the hillclimb dataset may be useful in other circumstances.

On the other hand, implementing prediction times into the selection of models for the ensemble greatly improved the performance relative to the total prediction time of the ensemble. Models trained with principal components had a better performance and shorter prediction times than models trained with the raw covariates. Consequently, they dominated the final ensemble.

Many of the covariates, which were important to the prediction of artificially drained areas, have also been used to map drainage pipes or artificially drained areas in other studies. However, the complexity of the best models and the optimal ensembles in this study demonstrate the limitations of models constructed from a priori knowledge. In contrast, the study shows that machine learning is a viable strategy for mapping artificially drained areas, if observations are at hand.

We carried out our study at a national level, and it is likely that machine learning approaches are best suited for large areas. For smaller areas, such as farm and field level studies, the best approaches probably include proximal sensors such as ground-penetrating radar and electromagnetic induction.

Soil drainage has a large impact on agriculture and the hydrological system. Therefore, the prediction of artificially drained areas is integral to protecting the environment and improving agricultural practices. We hope that this study will play a role in the development of better maps of artificially drained areas, both on the field level and in larger areas.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.geoderma.2018.01.018>.

References

- Adhikari, K., Kheir, R.B., Greve, M.B., Bocher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., Greve, M.H., 2013. High-resolution 3-D mapping of soil texture in Denmark. *Soil Sci. Soc. Am. J.* 77 (3), 860–876. <http://dx.doi.org/10.2136/sssaj2012.0275>.
- Adhikari, K., Minasny, B., Greve, M.B., Greve, M.H., 2014. Constructing a soil class map of Denmark based on the FAO legend using digital techniques. *Geoderma* 214–215, 101–113. <http://dx.doi.org/10.1016/j.geoderma.2013.09.023>.
- Allred, B.J., Redman, J.D., 2010. Location of agricultural drainage pipes and assessment of agricultural drainage pipe conditions using ground penetrating radar. *J. Environ. Eng. Geophys.* 15 (3), 119–134. <http://dx.doi.org/10.2113/JEEG15.3.119>.
- Allred, B.J., Fausey, N.R., Peters, L., Chen, C., Daniels, J.J., Youn, H., 2004. Detection of buried agricultural drainage pipe with geophysical methods. *Appl. Eng. Agric.* 20 (3), 307–318. <http://dx.doi.org/10.13031/2013.16067>.
- Behrendt, H., Bach, M., Kunkel, R., Opitz, D., Pagenkopf, W., Scholz, G., Wendland, F., 2003. Nutrient emissions into river basins of Germany on the basis of a harmonized procedure. *UBA-Texte* 82 (03), 201.
- Bell, J.C., Cunningham, R.L., Havens, M.W., 1992. Calibration and validation of a soil-landscape model for predicting soil drainage class. *Soil Sci. Soc. Am. J.* 56, 1860–1866. <http://dx.doi.org/10.2136/sssaj1992.0361599500560060035x>.
- Bell, J.C., Cunningham, R.L., Havens, M.W., 1994. Soil drainage class probability mapping using a soil-landscape model. *Soil Sci. Soc. Am. J.* 58, 464–470. <http://dx.doi.org/10.2136/sssaj1994.03615995005800020031x>.
- Boland-Brien, S.J., Basu, N.B., Schilling, K.E., 2014. Homogenization of spatial patterns of hydrologic response in artificially drained agricultural catchments. *Hydrol. Process.* 28 (19), 5010–5020. <http://dx.doi.org/10.1002/hyp.9967>.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24 (2), 123–140. <http://dx.doi.org/10.1023/A:1018054314350>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <http://dx.doi.org/10.1023/A:1010933404324>.
- Breiman, L., Cutler, A., Liaw, A., Wiener, M., 2015. randomForest: Breiman and Cutler's Random Forests for Classification and Regression. R Package Version 4. pp. 6–12. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>, Accessed date: 18 August 2017.
- Campling, P., Gobin, A., Feyen, J., 2002. Logistic modeling to spatially predict the probability of soil drainage classes. *Soil Sci. Soc. Am. J.* 66 (4), 1390–1401. <http://dx.doi.org/10.2136/sssaj2002.1390>.
- Caruana, R., Niculescu-Mizil, A., Crew, G., Ksikes, A., 2004. Ensemble selection from libraries of models. Proceedings of the twenty first international conference on Machine learning. ACM 18. <http://dx.doi.org/10.1145/1015330.1015432>.
- Collaku, A., Harrison, S.A., 2002. Losses in wheat due to waterlogging. *Crop Sci.* 42 (2), 444–450. <http://dx.doi.org/10.2135/cropsci2002.0444>.
- Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach. Learn.* 40 (2), 139–157. <http://dx.doi.org/10.1023/A:1007607513941>.
- Ernstsen, V., Olsen, P., Rosenbom, A.E., 2015. Long-term monitoring of nitrate transport to drainage from three agricultural clayey till fields. *Hydrol. Earth Syst. Sci.* 19 (8), 3475–3488. <http://dx.doi.org/10.5194/hess-19-3475-2015>.
- European Environment Agency, 2014. Corine Land Cover (CLC) 2012 - Denmark, Version 1, Oct. 2014 [Dataset]. <http://download.kortforsyningen.dk/content/corine-land-cover>.
- Feick, S., Siebert, S., Döll, P., 2005. A Digital Global Map of Artificially Drained Agricultural Areas. (Frankfurt).
- Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. *ICML* 96, 148–156.
- Giasson, E., Sarmento, E.C., Weber, E., Flores, C.A., Hasenack, H., 2011. Decision trees for digital soil mapping on subtropical basaltic steepplands. *Sci. Agric.* 68 (2), 167–174. <http://dx.doi.org/10.1590/S0103-90162011000200006>.
- Gökkaya, K., Budhathoki, M., Christopher, S.F., Hanrahan, B.R., Tank, J.L., 2017. Subsurface tile drained area detection using GIS and remote sensing in an agricultural watershed. *Ecol. Eng.* <http://dx.doi.org/10.1016/j.ecoeng.2017.06.048>.
- Greve, M.H., Christensen, O.F., Greve, M.B., Kheir, R.B., 2014. Change in peat coverage in Danish cultivated soils during the past 35 years. *Soil Sci.* 179 (5), 250–257. <http://dx.doi.org/10.1097/ss.0000000000000066>.
- Hansen, B., Olesen, S.E., Ernstsen, V., 2004. Naturlig dræning og grundvandsdannelse. In: Ernstsen, V. (Ed.), *Afprøvning af undersøgelsesmetoder med henblik på etablering af et zoneringskoncept for danske lerjorde: Statusrapport*. Geological Survey of Denmark and Greenland, Appendix 10.
- Henderson, B.L., Bul, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. *Geoderma* 124 (3–4), 383–398. <http://dx.doi.org/10.1016/j.geoderma.2004.06.007>.
- Hengl, T., de Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilbarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Lennox, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12 (2), e0169748. <http://dx.doi.org/10.1371/journal.pone.0169748>.
- Henriksen, H.J., Højberg, A.L., Olsen, M., Seaby, L.P., van der Keur, P., Sørensen, S., Trolldenier, L., Sonnenborg, T.O., Refsgaard, J.C., 2012. Klimaeffekter på hydrologi og grundvand - Klimagrundvandskort. Aarhus University.
- Hijmans, R.J., van Eten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., Shorridge, A., 2016. Raster: Geographic Data Analysis and Modeling. R Package Version 2. pp. 5–8. <https://cran.r-project.org/web/packages/raster/raster.pdf>, Accessed date: 10 February 2017.
- Hirt, U., Meyer, B.C., Hammann, T., 2005. Proportions of subsurface drainages in large areas—methodological study in the Middle Mulde catchment (Germany). *J. Plant Nutr. Soil Sci.* 168 (3), 375–385. <http://dx.doi.org/10.1002/jpln.200421621>.
- Hothorn, T., Hornik, K., Strobl, C., Zeileis, A., 2017. Party: A Laboratory for Recursive Partitioning. R Package Version 1. pp. 2–3. <https://cran.r-project.org/web/packages/party/party.pdf>, Accessed date: 9 January 2017.
- Jakobsen, P.R., Hermansen, B., Tougaard, L., 2015. Danmarks digitale jordartskort 1:25000 version 4.0. GEUS.
- Kim, H., Loh, W.-Y., 2001. Classification trees with unbiased multiway splits. *J. Am. Stat. Assoc.* 96 (454), 589–604. <http://dx.doi.org/10.1198/016214501753168271>.
- Knotters, M., Brus, D.J., Oude Voshaar, J.H., 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma* 67 (3–4), 227–246. [http://dx.doi.org/10.1016/0016-7061\(95\)00011-c](http://dx.doi.org/10.1016/0016-7061(95)00011-c).
- Kravchenko, A.N., Bollero, G.A., Omonode, R.A., Bullock, D.G., 2002. Quantitative mapping of soil drainage classes using topographical data and soil electrical conductivity. *Soil Sci. Soc. Am. J.* 66 (1), 235–243. <http://dx.doi.org/10.2136/sssaj2002.0235>.
- Kuhn, M., 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* 28 (5), 1–26. <http://dx.doi.org/10.18637/jss.v028.i05>.
- Madsen, H.B., 2010. Drænings indførelse og betydning i et landbrugs- og miljømassigt perspektiv. In: Nielsen, M.A. (Ed.), *Det fremmede som historisk drivkraft. Danmark efter 1742. Det Kongelige Danske Videnskabskabernes Selskab*, pp. 158–165.
- Madsen, H.B., Nørr, A.H., Holst, K.A., 1992. The Danish Soil Classification. The Royal Danish Geographical Society, Copenhagen, Denmark.
- Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B., 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 232–234, 34–44. <http://dx.doi.org/10.1016/j.geoderma.2014.04.033>.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117 (1–2), 3–52. [http://dx.doi.org/10.1016/s0167-7061\(03\)00223-4](http://dx.doi.org/10.1016/s0167-7061(03)00223-4).
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* 32 (9), 1378–1388. <http://>

- [dx.doi.org/10.1016/j.geoderma.2017.10.015](https://doi.org/10.1016/j.geoderma.2017.10.015).
- Møller, A.B., Beucher, A., Iversen, B.V., Greve, M.H., 2017. Prediction of soil drainage classes in Denmark by means of decision tree classification. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2017.10.015>.
- NASA Landsat Program, 2014. Landsat OLI/TIRS scenes LC81920222014090LGN00, LC81940212014072LGN00, LC81940222014072LGN00, LC81950212014079LGN00, LC81950222014079LGN00, LC81960202014070LGN00, LC81960212014070LGN00, LC81960222014070LGN00, LC81980202014068LGN00, LC81980212014068LGN00, LIT. USGS, Sioux Falls.
- Naz, B.S., Bowling, L.C., 2008. Automated identification of tile lines from remotely sensed data. *Trans. ASABE* 51 (6), 1937–1950. <https://doi.org/10.13031/2013.25399>.
- Naz, B.S., Ale, S., Bowling, L.C., 2009. Detecting subsurface drainage systems and estimating drain spacing in intensively managed agricultural landscapes. *Agric. Water Manag.* 96 (4), 627–637. <https://doi.org/10.1016/j.agwat.2008.10.002>.
- Northcott, W.J., Verma, A.K., Cooke, R.A., 2000. Mapping Subsurface Drainage Systems Using Remote Sensing and GIS. ASABE, St. Joseph, Michigan.
- Olesen, S.E., 2009. Kortlægning af potentielt dræningsbehov på landbrugsarealer opdelt efter landskabsselement, geologi, jordklasse, geologisk region samt høj/lavbund. Aarhus University.
- Oreskes, N., 1998. Evaluation (not validation) of quantitative models. *Environ. Health Perspect.* 106 (Suppl. 6), 1453.
- Ren, B., Zhang, J., Li, X., Fan, X., Dong, S., Liu, P., Zhao, B., 2014. Effects of waterlogging on the yield and growth of summer maize under field conditions. *Can. J. Plant Sci.* 94 (1), 23–31. <https://doi.org/10.4141/cjps2013-175>.
- Scull, P., Franklin, J., Chadwick, O., McArthur, D., 2003. Predictive soil mapping: a review. *Prog. Phys. Geogr.* 27 (2), 171–197. <https://doi.org/10.1191/0309133303pp366ra>.
- Seligman, M., 2017. Rborist: Extensible, Parallelizable Implementation of the Random Forest Algorithm. R Package Version 0. pp. 1–8. <https://cran.r-project.org/web/packages/Rborist/index.html>.
- Statistics Denmark, 2006. Arealanvendelse. <https://www.dst.dk/da/Statistik/emner/areal/arealanvendelse>, Accessed date: 9 December 2016.
- Sugg, Z., 2007. Assessing US Farm Drainage: Can GIS Lead to Better Estimates of Subsurface Drainage Extent? World Resources Institute, Washington, DC 20002.
- Tetzlaff, B., Kuhr, P., Wendland, F., 2009. A new method for creating maps of artificially drained areas in large river basins based on aerial photographs and geodata. *Irrig. Drain.* 58 (5), 569–585. <https://doi.org/10.1002/ird.426>.
- Thayn, J.B., Campbell, M., Deloria, T., 2011. Mapping Tile-Drained Agricultural Lands. Institute for Geospatial Analysis and Mapping (GEOMAP). Illinois State University.
- The Danish Agricultural Agency, 2014. Markkort. <http://www.geodatainfo.dk/Portal/ShowMetadata.aspx?id=6e3bc77f-c193-4508-80d4-836e1668db91>, Accessed date: 29 May 2017.
- Verma, A.K., Cooke, R.A., Wendte, L., 1996. Mapping Subsurface Drainage Systems with Color Infrared Aerial Photographs. Department of Agricultural Engineering, University of Illinois, Urbana-Champaign.
- Wang, P.R., 2013. Referenceværdier: døgn-, måneds- og årsverdier for regioner og hele landet 2001–2010, Danmark for temperatur, relativ luftfugtighed, vindhastighed, globalstråling og nedbør. In: Teknisk Rapport 12–24. Danish Meteorological Institute.
- Watson, E.R., Lapins, P., Barron, R.J.W., 1976. Effect of waterlogging on the growth, grain and straw yield of wheat, barley and oats. *Anim. Prod. Sci.* 16 (78), 114–122. <https://doi.org/10.1071/EA9760114>.
- Wright, M.N., Ziegler, A., 2015. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Soft.* 77 (1). <https://doi.org/10.18637/jss.v077.i01>.
- Zhao, Z.Y., Chow, T.L., Yang, Q., Rees, H.W., Benoy, G., Xing, Z.S., Meng, F.R., 2008. Model prediction of soil drainage classes based on digital elevation model parameters and soil attributes from coarse resolution soil maps. *Can. J. Soil Sci.* 88 (5), 787–799. <https://doi.org/10.4141/CJSS08012>.
- Zhao, Z.Y., Ashraf, M.I., Meng, F.R., 2013. Model prediction of soil drainage classes over a large area using a limited number of field samples: a case study in the province of Nova Scotia, Canada. *Can. J. Soil Sci.* 93 (1), 73–83. <https://doi.org/10.4141/cjss2011-095>.

DCA - Nationalt Center for Fødevarer og Jordbrug er den faglige indgang til jordbrugs- og fødevareforskningen ved Aarhus Universitet (AU). Centrets hovedopgaver er videnudvæksling, rådgivning og interaktion med myndigheder, organisationer og erhvervsvirksomheder.

Centret koordinerer videnudvæksling og rådgivning ved de institutter, som har fødevarer og jordbrug, som hovedområde eller et meget betydende delområde:

Institut for Husdyrvidenskab
Institut for Fødevarer
Institut for Agroøkologi
Institut for Ingeniørvidenskab
Institut for Molekylærbiologi og Genetik

Herudover har DCA mulighed for at inddrage andre enheder ved AU, som har forskning af relevans for fagområdet.

An aerial photograph of agricultural fields. The fields are divided into various plots, some of which are darker in color, possibly indicating different crops or soil conditions. A prominent square plot is visible in the center, surrounded by a lighter-colored border. The overall scene shows a typical agricultural landscape with distinct field boundaries and patterns.

RESUME

Denne rapport præsenterer et nyt landsdækkende kort over drænede markarealer i Danmark. Kortet viser i en opløsning på 30,4 meter den sandsynlige udbredelse af drænede og ikke drænede arealer. Kortet er baseret på statistiske analyser af sammenhængen mellem dræning og en lang række geografiske datalag, heriblandt topografiske variabler, jordbundsforhold, satellitbilleder, arealanvendelse m.fl. Kortet har en nøjagtighed på 79%. I alt 52% af markarealet fremgår som drænet på kortet. Med kortet følger desuden et datalag over sandsynligheden for dræning samt datalag til vurdering af kortets nøjagtighed i specifikke områder.