

Den Europæiske Landbrugsfond for Udvikling af Landdistrikterne: Danmark og Europa investerer i landdistrikterne



Miljø- og Fødevareministeriet
Landbrugsstyrelsen



Den Europæiske Landbrugsfond
for Udvikling af Landdistrikterne

LDP 2020



TEKNOLOGISK
INSTITUT

Agro Food Park 15,
Skejby
8200 Aarhus N
Tlf. +45 7220 2000
info@teknologisk.dk

Se EU-Kommissionen, Den Europæiske Landbrugsfond for Udvikling af Landdistrikterne

Notat: Satellitdata-baseret bestemmelse af kløverandelen i slætgræsmarker.

31-01-2019

Philipp Trénel, ph.d., seniorkonsulent, specialist
phtr@teknologisk.dk, +45 72 20 33 92

Formål

- At undersøge, om sentinel-2-satellitdata effektivt kan bruges til at bestemme en marks gennemsnitlige kløverandel (kløver%) på baggrund af et af SEGES indhentet datasæt over observerede kløver%-data bestemt vha. AgroIntellis kamerabaserede CloverCam teknologi.
- Der undersøges også, om fordelingen af kløver% indenfor marken (punktniveau) kan bestemmes vha. satellitdata.

Overordnede konklusioner

- Den her anvendte model bestemmer en marks gns. kløverandel med en præcision på **$\pm 11.2\%$ -point** (testsæt RMSEP, $1/\log(\text{kløverandel})$ -vægtet model).
- Præcisionen på punktniveau er lavere ($\pm 16.6\%$ -point).
- Modellen er generelt bedre til at skelne mellem marker med en kløverandel over og under 30% (balanced accuracy = 0.88) end mellem marker med en kløverandel over og under 10% (balanced accuracy = 0.67) ($1/\sqrt{\text{kløverandel}}$ -vægtet model).
- Funktional PCA kan anbefales som en god strategi, såfremt det ønskes at implementere satellitdata i modeller, da denne metode kan håndtere og delvist kompensere for at data kan være meget sparsomme pga. skydække.
- Det understreges, at de foreliggende dataanalyser bygger på en række eksklusionskriterier, dataudvælgelseskriterier og andre modelrelaterede valg, og at det på nuværende tidspunkt ikke er undersøgt, om modellerne kan forbedres ved ændringer i nogle af de anvendte kriterier og trufne valg.
- Det understreges også, at det på baggrund af det foreliggende begrænsede datamateriale ikke var muligt at fitte modeller, der tager højde for rødkløver vs. hvidkløver-marker, til trods for at der er en indikation af, at denne skelnen ville kunne forbedre prædiktionen. Det anbefales, at dette undersøges nærmere i fremtiden på et større datamateriale. Derudover kunne det undersøges om prædiktionen kan forbedres ved 1) en udvidelse af datasættet med flere marker med en høj kløverandel, 2) data fra et nyt år, 3) ved at bruge f.eks. en to-trins-model eller en særskilt model for markens gennemsnit og 4) ved udelukkelse af datapunkter (punkter i marken) med f.eks. færre end 6 brugbare satellitbilleder indenfor 30 dages vinduet.

Metoder

Datasæt og datahåndtering

- Data og datahåndtering følger beskrivelsen i notatet "Notat - Satellitbaseret prædiktions af TSudbytte og TSprocent i majs og slætgræs" fra den 30-01-2019 (bilag 1).
- AgroIntelli-data blev leveret af SEGES.
- Der brugt funktionel PCA hen over en 30-dages-vindue før kløverbedømmelsesdato for alle satellitbånd (B02, B03, B04, B05, B06, B07, B08, B8A og skydække CLD), samt NDVI, NDRE, NDRE2 (se bilag 1) og NDRE/NDVI. Derudover er for alle variabler fundet værdien fra sidste brugbare satellitdato umiddelbart før kløverbedømmelsesdatoen.
- Der er beregnet ugenummer for kløverbedømmelse og dage mellem sidste brugbare satellitdato og kløverbedømmelsesdato, idet indledende modelkørsler viste, at ugenummer for kløverbedømmelse og dage mellem sidste brugbare satellitdato og kløverbedømmelsesdato er signifikante.
- Der er anvendt følgende eksklusions-/inklusions-kriterier i alle modeller:
 - o $\text{lwrBound_NDVI} = 0.2$ (punkter i marken med $\text{NDVI} < 0.2$ er ekskluderet)
 - o $\text{uprBound_CLD} = 5$ (punkter i marken med CLD (skydække) $> 5\%$ er ekskluderet)
 - o $\text{window_days} = 30$ (til enhver mark/bedømmelsesdato-kombination blev der anvendt satellitdata indenfor et 30 dages vindue før bedømmelsesdato)
 - o $\text{minNobs} = 10$ (mark/bedømmelsesdato-kombinationer med færre end 10 punkter blev ekskluderet)
- Alle analyser og modeller blev udført på punktniveau (indenfor og mellem-mark).
- Prædiktionen for gns. kløverandel er beregnet som den gennemsnitlige prædiktions på punktniveau for en given mark.
- Alle analyserede modeller er GAM modeller med alle inputvariabler fitted som thin-plate spline funktioner med udglatningsgraden bestemt vha. generaliseret krydsvalidering.
- Modellernes performance blev evalueret vha. en 5-fold 2-times repeated cross validation. Testsituationen er udført som en 5-fold 2-times repeated Cross Validation, hvor det altid er hele marken (lokalitet, dvs. også hen over forskellige datoer fra samme mark) der er tilbageholdt i testsættet og udeladt fra træningssættet.
- Da der foreligger langt flere marker med en lille gns. kløverandel og kun få med en stor kløverandel, er alle modellen fittet med dataobservationerne vægtet efter deres kløverandel med følgende vægtninger: 1) uvægtet, 2) proportionalt med den inverse af kløverandelen, 3) proportionalt med den inverse af kvadratroden af kløverandelen, 4) proportionalt med den inverse af log af kløverandelen.

Resultater

- Resultaterne fremgår af Fig. 1 til 3 og tabel 1.



Fig. 1: Den AgroIntelli-bestemte (A, venstre) og satellitdata-baserede modelprædiktation (P, højre) af kløverandelen for alle mark-bedømmelsestidspunkt-kombinationer vist både indenfor marken (punkterne) og for markens gennemsnit (tal i fed). Der er highlightet tre marker: 1) Den ene viser en rigtig god prædiktation af både markens gennemsnitlige kløverandel og indenfor-mark variationen (K.S.13-0, 10-12). 2) Den anden viser en ok god prædiktation af markens gennemsnitlige kløverandel, men en afvigende prædiktation af indenfor-mark variationen (Martin V.V. 17, 09-20) 3) Den tredje viser en dårlig prædiktation af markens gennemsnitlige kløverandel, men en god prædiktation af indenfor-mark variationen (Ekstra A.I. 2, 10-09).

Tabel 1: Modellens testset-performance for modellerne fittet med de fire vægtninger: 1) uvægtet, 2) proportionalt med den inverse af kløverandelen, 3) proportionalt med den inverse af kvadratroden af kløverandelen, 4) proportionalt med den inverse af log af kløverandelen.

1/log(kløverandel)-vægtet			1/sqrt(kløverandel)-vægtet			1/kløverandel-vægtet			uvægtet		
	Mean	s.e.		Mean	s.e.		Mean	s.e.		Mean	s.e.
RMSEP_punkt	16.60	5.42	RMSEP_punkt	19.83	7.38	RMSEP_punkt	21.28	9.10	RMSEP_punkt	17.22	7.41
Se10_punkt	0.51	0.23	Se10_punkt	0.62	0.18	Se10_punkt	0.64	0.18	Se10_punkt	0.47	0.24
Sp10_punkt	0.58	0.24	Sp10_punkt	0.52	0.21	Sp10_punkt	0.49	0.26	Sp10_punkt	0.58	0.22
bAcc10_punkt	0.55	0.13	bAcc10_punkt	0.57	0.11	bAcc10_punkt	0.57	0.10	bAcc10_punkt	0.52	0.11
Se30_punkt	0.14	0.25	Se30_punkt	0.21	0.28	Se30_punkt	0.22	0.23	Se30_punkt	0.10	0.18
Sp30_punkt	0.92	0.08	Sp30_punkt	0.86	0.16	Sp30_punkt	0.81	0.19	Sp30_punkt	0.92	0.11
bAcc30_punkt	0.53	0.11	bAcc30_punkt	0.53	0.12	bAcc30_punkt	0.51	0.16	bAcc30_punkt	0.51	0.10
RMSEP_mark	11.22	3.43	RMSEP_mark	13.57	5.52	RMSEP_mark	14.38	7.07	RMSEP_mark	11.52	5.88
Se10_mark	0.62	0.37	Se10_mark	0.83	0.24	Se10_mark	0.77	0.30	Se10_mark	0.71	0.34
Sp10_mark	0.53	0.29	Sp10_mark	0.50	0.26	Sp10_mark	0.31	0.34	Sp10_mark	0.52	0.39
bAcc10_mark	0.57	0.28	bAcc10_mark	0.67	0.20	bAcc10_mark	0.54	0.27	bAcc10_mark	0.62	0.28
Se30_mark	0.75	0.35	Se30_mark	0.75	0.29	Se30_mark	0.58	0.49	Se30_mark	0.28	0.25
Sp30_mark	0.88	0.18	Sp30_mark	1.00	0.00	Sp30_mark	0.92	0.14	Sp30_mark	0.93	0.12
bAcc30_mark	0.81	0.09	bAcc30_mark	0.88	0.14	bAcc30_mark	0.75	0.31	bAcc30_mark	0.61	0.18

RMSEP_punkt og RMSEP_mark: prædiktionsnøjagtighed på punkt- hhv. markniveau. Se10, Sp10, Se20, Sp30: Sensitivitet hhv. specificitet for skelnen mellem marker over/under hhv. 10% og 30% kløver. bAcc = balanced accuracy.

- Den bedste prædiktionsnøjagtighed målt som **RMSEP** er fundet for den 1/log(kløverandel)-vægtede model med en prædiktionsnøjagtighed af kløverandelen for en punkt (plet) i marken på $16.6 \pm 5.4\%$ -point og på **$11.2 \pm 3.4\%$ -point** for markens gennemsnitlige kløverandel. Til gengæld har denne model en relativ lav evne til at skelne mellem marker med en gns. kløverandel over/under 10% (bAcc = balanced accuracy = 0.57 ± 0.28). Det ser bedre ud med at skelne mellem over/under 30%. Her er balanced accuracy = 0.81 ± 0.09 (dvs. der er 19% fejl).
- Den bedste prædiktionsnøjagtighed målt som **accuracy** for at skelne mellem marker i kategorierne <10%, 10-30% og >30% er den 1/sqrt(kløverandel)-vægtede model. Her er balanced accuracy for at skelne mellem over/under 10% kløverandel på 0.67 ± 0.20 og over/under 30% på 0.88 ± 0.14 . Udtrykker vi modellens performance som sensitivitet, så fanger modellen 83% af alle marker, der har en kløverandel > 10% (Se10_mark), og 75% af alle marker, der har en kløverandel > 30% (Se30_mark).

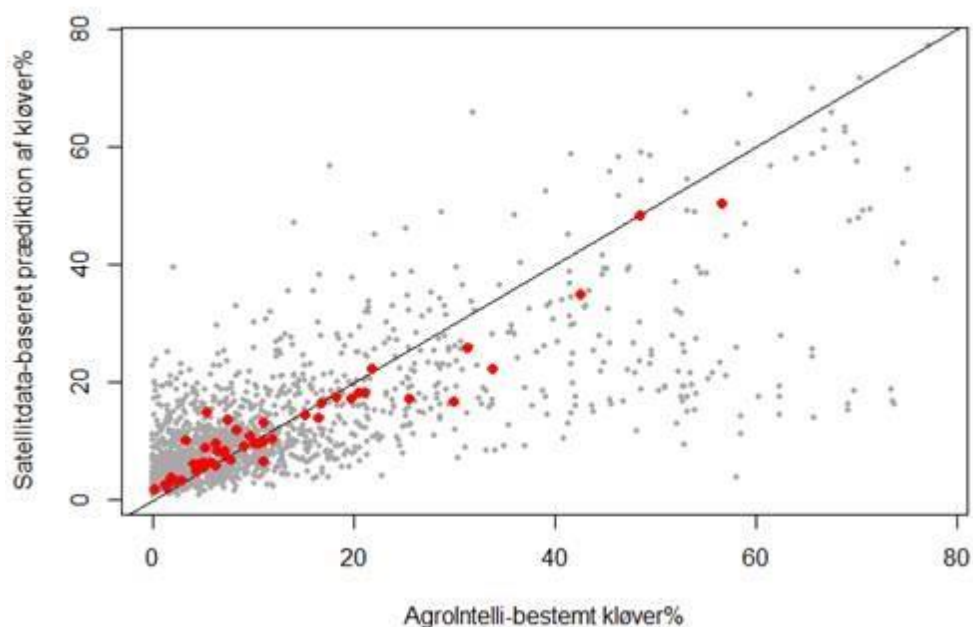


Fig. 2: Prædikeret vs. observeret. Grå punkter: punktniveau, røde punkter: markniveau.

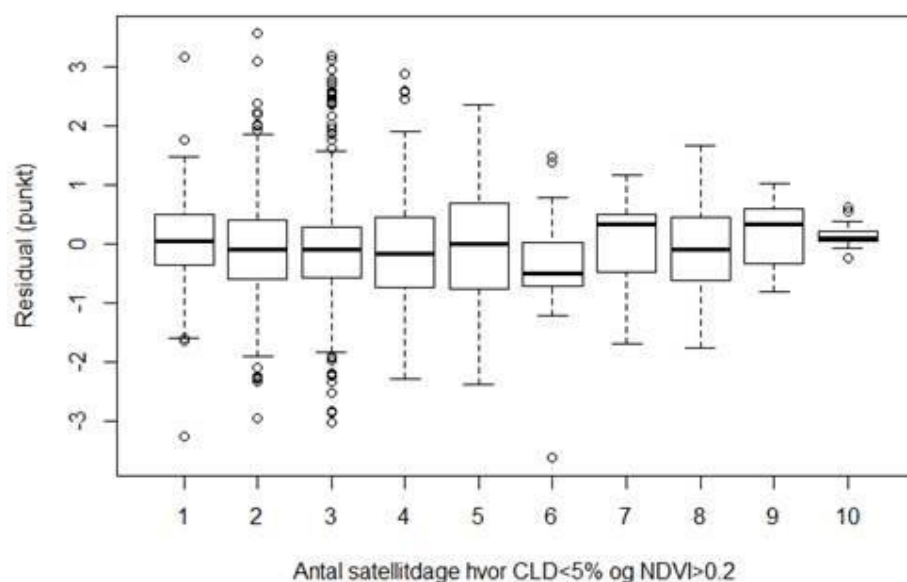


Fig. 3: Prædiktions-ujagtigheden flader med antal satellitdage, der opfylder kvalitetskriterierne. Der er en indikation af, at der skal være 6 eller flere brugbare satellitbilleder før prædiktions-ujagtigheden falder.