

Promilleafgiftsfonden for landbrug

Notat: Prædiktion af tørstofudbytte, tørstofprocent og proteinprocent i majs på baggrund af drone-indhentede spektraldata i forsøg 030191919 (høsttidsforsøg) og 031191919 (sortsforsøg).

10-07-2020

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

Formål

Opgaven er en del af SEGES' projekt "Optimal og bæredygtig foderforsyning på kvægbedrifter". Formålet af denne analyse med indenfor projektet er:

- At udvikle en nemt implementerebar model til prædiktion af tørstofudbytte (TS-udbytte), tørstofprocent (TS%) og proteinprocent (Protein%) i helsædsmajs på baggrund af drone-indhentede spektralbånd (R, G, B, NIR og RedEdge) og heraf afledte vegetationsindekser (VI) indsamlet i forsøgsserierne 030191919 (høsttidsforsøg) og 031191919 (sortsforsøg).
- At undersøge, om en dynamisk kombination og/eller sammenvægtning af VI'er kan forbedre prædiktionsevnen.
- At vurdere modellens prædiktionssikkerhed målt som Root Mean Squared Error of Prediction (RMSEP) langs tidsaksen igennem majsens vækstsæson (tid for dronedeflyvning) og afhængig af høsttidspunkt.

Overordnede konklusioner

- Der er fundet at prædiktionsevnen for TS-udbytte, TS% og Protein% ud fra drone-indhentede spektralbånd og heraf afledte vegetationsindekser (VI) betydeligt kan forbedres ved at inddrage information om sorten.
- Der anbefales forskellige indekser til prædiktion af hhv. TS-udbytte (NDVI), TS% (NDRE, MCARI) og Protein% (flere VI (lasso), særligt i det blå spektrum). Der er brug for flere undersøgelser, særligt for at undersøge om de fundne resultater er lokalitetsspecifikke.
- Der er fundet at prædiktionsevne og optimal VI varierer langs tidsaksen og i forhold til størrelsen af forecast-vinduet.
- Detaljerede resultater fremgår af bilag 1 (resultater_bilag1.xlsx)

Metoder

Forsøgsdesign

- Forsøgsbehandlinger, måleparametre og forsøgsdesign for forsøgsserierne 030191919 og 031191919 er beskrevet i Nordic Field Trial System her:
 - 030191919 (1 forsøg), se her.
 - 031191919 (2 forsøg), se her.
- Yderligere informationer fremgår af resultatsiden i Nordic Field Trial System under de enkelte forsøg.
- Data anvendt i nuværende kontekst består af:
 - forsøgs-id (*trial*) bestående af PLANR og LBNR hentet fra NFTS.
 - Dronedata til tiden *t* bestående af spektralbåndene *R*, *G*, *B*, *NIR* og *RedEdge*. For hvert spektralbånd og parcel er der beregnet parcelvis gennemsnit (*mean*) og standardafvigelse (*sd*) af pixelværdier indenfor parcellen. Ud fra de parcelvise spektralbåndsgennemsnit er der beregnet publicerede vegetationsindekser (*VI*), se tabel 1. Spektralbåndene (*mean* og *sd*) og *VI* kaldes her også tilsammen for *VI*.
 - Dronedata er interpoleret til alle dage mellem første og sidste observationsdag ved hjælp af Stineman-interpolering (monoton polynomial interpolering). Samme interpolering er anvendt for alle responsvariabler mellem høstdage i høsttidsforsøget (030191919-001), hvorved interpoleringen er foretaget på gentagelsesblok-niveau, dvs. parceller høstet på forskellige tidspunkter men af samme sort indenfor gentagelsesblok under antagelse om homogenitet indenfor blok.

Tabel 1: De anvendte vegetationsindekser (*VI*) inkl. spektralbånd. Der er også anvendt rent RGB-baserede *VI*.

Bands: <i>R</i> _{mean} (<i>R</i>), <i>G</i> _{mean} (<i>G</i>), <i>B</i> _{mean} (<i>B</i>), <i>NIR</i> _{mean} (<i>NIR</i>), <i>redEdge</i> _{mean} (<i>redEdge</i>), <i>R</i> _{sd} , <i>G</i> _{sd} , <i>B</i> _{sd} , <i>NIR</i> _{sd} , <i>redEdge</i> _{sd}	
Spectral-band-based indices:	
• Normalised Difference Vegetation Index, Rouse 1974	$NDVI = (NIR - R) / (NIR + R)$
• Normalized Difference Red Edge Index	$NDRE = (NIR - redEdge) / (NIR + redEdge)$
• Corrected Transformed Vegetation Index, Perry 1984	$CTVI = (NDVI + 0.5) / \sqrt{abs(NDVI + 0.5)}$
• Difference Vegetation Index, Richardson1977	$DVI = s * NIR - R$, hvor $s = 1$
• Enhanced Vegetation Index, Huete 1999	$EVI = g * ((NIR - R) / (NIR + C_1 * R - C_2 * B + L_{evi}))$, hvor $g = 2.5$, $C_1 = 6$, $C_2 = 7.5$ og $L_{evi} = 1$
• Two-band Enhanced Vegetation Index,	$EVI2 = g * (NIR - R) / (NIR + 2.4 * R + 1)$
• Green Normalised Difference Vegetation Index,	$GNDVI = (NIR - G) / (NIR + G)$
• Modified Chlorophyll Absorption Ratio Index,	$MCARI = ((redEdge - R) - (redEdge - G)) * (redEdge / R)$
• Modified Soil Adjusted Vegetation Index, Qi 1994	$MSAVI = NIR + 0.5 - (0.5 * \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - (2 * R))})$
• Modified Soil Adjusted Vegetation Index 2, Qi1994	$MSAVI2 = (2 * (NIR + 1) - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - (NIR - R))}) / 2$
• Normalised Difference Water Index, McFeeters 1996	$NDWI = (G - NIR) / (G + NIR)$
• Normalized difference spectral index, Zhao et al. 2018	$NDSI1 = (redEdge - G) / (redEdge1 + G)$ $NDSI2 = (redEdge - B) / (redEdge1 + B)$ $NDSI3 = (redEdge - \text{avg}(G, B)) / (redEdge + \text{avg}(G, B))$, hvor $\text{avg}(\cdot)$ står for gennemsnittet

• NG	$NG = NIR / G$
• Normalised Ratio Vegetation Index, Baret 1991	$NRVI = (R / NIR - 1) / (R / NIR + 1)$
• Optimized soil-adjusted vegetation index,	$OSAVI = (1 + 0.16 * ((NIR - R) / (NIR + R + 0.16)))$
• RDVI	$RDVI = (NIR - R) / (NIR + R)^2$
• Ratio spectral index (RSI), Zhao et al. 2018	$RSI1 = redEdge / B; RSI2 = redEdge / G; RSI3 = redEdge / avg(B, G)$
• Ratio Vegetation Index	$RVI = R / NIR$
• Soil Adjusted Vegetation Index, Huete 1988	$SAVI = (NIR - R) * (1 + L) / (NIR + R + L), \text{ hvor } L = 0.5$
• Simple Ratio Vegetation Index, Birth 1968	$SR = NIR / R$
• Thiam's Transformed Vegetation Index Thiam 1997	$TTVI = \sqrt{\text{abs}((NIR - R) / (NIR + R) + 0.5)}$
• Transformed Chlorophyll Absorption in Reflectance Index, TCARI = 3 * ((redEdge - R) - 0.2 * (redEdge - G) * (redEdge / R))	
• TCARI / OSAVI ratio, Haboudane et al. 2002	$TCARI\text{-}OSAVI\text{-}ratio = TCARI / OSAVI$
• Transformed Vegetation Index, Deering 1975	$VI = \sqrt{((NIR - R) / (NIR + R) + 0.5)}$
RGB-based indices:	
• Brightness Index - Average soil reflectance magnitude	$BI = \sqrt{(R^2 + G^2 + B^2) / 3}$
• Soil Colour Index	$SCI = (R - G) / (R + G)$
• Green leaf index, Louhaichi et al. (2001)	$GLI = (2 * G - R - B) / (2 * G + R + B)$
• Primary colours Hue Index	$HI = (2 * R - G - B) / (G - B)$
• Normalized difference turbidity index Water	$NDTI = (R - G) / (R + G)$
• Normalized green red difference index (sometimes GRVI) NGRDI = (G - R) / (G + R)	
• Spectral Slope Saturation Index	$SI = (R - B) / (R + B)$
• A Visible Atmospherically Resistant Index (VARI)	$VARI = (G - R) / (G + R - B)$
• VVI	$VVI = (1 - (R - 30) / (R + 30)) * (1 - (G - 50) / (G + 50)) * (1 - (B - 1) / (B + 1))$
• GLAI	$GLAI = (25 * (G - R) / (G + R - B) + 1.25)$
• Green-Red Vegetation Index	$GRVI = (G - R) / (G + R)$
• Coloration Index	$CI = (R - B) / R$
• Overall Hue Index	$HUE = \text{atan}(2 * (R - G - B) / 30.5 * (G - B))$
• Overall Saturation Index	$SAT = (\max(R, G, B) - \min(R, G, B)) / \max(R, G, B)$

Model

- Der er anvendt følgende lineære og lineære mixed effects modeller indenfor hvert af de tre forsøg (030191919-001, 031191919-001, 031191919-002), samt kombinationen af forsøg 030191919-001 og 031191919-001:

- (1) $Y_{T,i} \sim \beta_0 + \beta_1 x_{t,i} + g_{T,t,i} + s_{T,t,i} + e_{T,t,i},$ (model I, x)
- (2) $Y_{T,i} \sim \beta_0 + \beta_1 dx_{t,i} + g_{T,t,i} + s_{T,t,i} + e_{T,t,i},$ (model II, dx)
- (3) $Y_{T,i} \sim \beta_0 + \beta_1 x_{t,i} + \beta_2 x_{t,i}^2 + g_{T,t,i} + s_{T,t,i} + e_{T,t,i},$ (model III, poly)
- (4) $Y_{T,i} \sim \beta_0 + \beta_1 x_{t,i} + \beta_2 dx_{t,i} + g_{T,t,i} + s_{T,t,i} + e_{T,t,i},$ (model IV, x*dx)

hvor $g_{T,t,i} \sim N(0, \sigma_e^2)$, $s_{T,t,i} \sim N(0, \sigma_e^2)$ og $e_{T,t,i} \sim N(0, \sigma_e^2)$, hvor $g_{T,t,i}$ modellerer den tilfældige effekt for gentagelse og $s_{T,t,i}$ modellerer den tilfældige effekt af sorten i form af sortens forventede TS% (LSmeans TS% fra Landsforsøgene, hentet fra sortsvalg majs). $Y_{T,i}$ står for én af de tre responsvariabler TS-udbytte i hkg/ha, TS% og Protein% observeret for den i 'te parcel ved høsttid T og $x_{t,i}$ står for én af de i tabel 1 listede VI'er observeret ved drone-overflyvningstid t , hvorved $dx_{t,i}$ indikerer 7-dags-ændringen i samme variabel, og $T \geq t$.

- Alle modeller er også analyseret uden den tilfældige sortseffekt, dvs. uden information om sortens forventede TS%.
- Derudover er der anvendt en model, der kombinerer alle VI. Dertil er der anvendt LASSO regression med alle VI og deres 2-vejs vekselvirkning inkluderet i modellen, med penalty parameteren λ estimeret vha. krydsvalidering.
- Model 1 modellerer sammenhængen mellem responsvariablen og det absolutte **niveau** af VI-variablen.
- Model 2 modellerer sammenhængen mellem responsvariablen og den relative **ændring** i VI-variablen.
- Model 3 og 4 modellerer mere komplekse sammenhænge.
- Da Protein% kun foreligger på led-niveau, udgår $g_{T,t,i}$ og $s_{T,t,i}$ fra modellen, som dermed reduceres til en lineær model.
- Modelperformance er vurderet for alle modeller vha. krydsvalidering (CV). CV-testsæt-modelperformance er vurderet vha. RMSEP på baggrund af en Leave-one-replicate-block-out krydsvalidering. Træningssæt-modelperformance er også vurderet. Hertil er der brugt forklaringsgraden R^2 og Pearsons korrelationskoefficient r .
- Alle analyser er gennemført i R.

Resultater

- Resultater fremgår af vedhæftede excelark (bilag 1) og er visualiseret for forsøg 030191919-001 i Fig. 1.
- Bilagets excel-tabels kolonner er følgende:
 - trial: enkeltforsøget
 - model: angiver, om der er anvendt model I (niveau, x), II (hældning, dx), III (polynomium, poly) eller IV ($x \cdot dx$).
 - sortTS: inkluderer modellen den tilfældige sortseffekt $s_{T,t,i}$ i form af sortens forventede TS% (LSmeans TS% fra Landsforsøgene, hentet fra sortsvalg majs).
 - yvar: responsvariablen
 - dato: droneflyvningsdato (inkl. interpolerede dronedata)
 - hostdato: høstdato (inkl. interpolerede responsværdier)
 - xvar: den forklarende VI variabel. Der er vist resultater for NDVI, NDRE, den VI, der udviste den mindste RMSEP og endeligt VI'en "all" i tilfælde af kombinationen af alle VI'er i lasso modellen. Faktisk foreligger resultater for alle VI'er, men disse er ikke vist i excelfilen og skal rekvireres efter behov.

- RMSEP: værdi for test-set RMSEP. Målt i samme enhed som responsvariablen.
- se: usikkerheden af test-set RMSEP i form af en standard error.
- R2: R^2 for modellen.
- pval: p-værdi for testen, at modellens er bedre end null-modellen (model med kun en intercept).
- r: korrelationskoefficienten for sammenhængen mellem VI og responsvariablen
- SD_sort: den estimerede standardafvigelse for varianskomponenten tilfældige sorteffekt $s_{T,t,i}$ i form af sortens forventede TS%. Målt i samme enhed som RMSEP.
- formula: den anvendte models faste effekter i form af en ligning med de estimerede coefficienter.

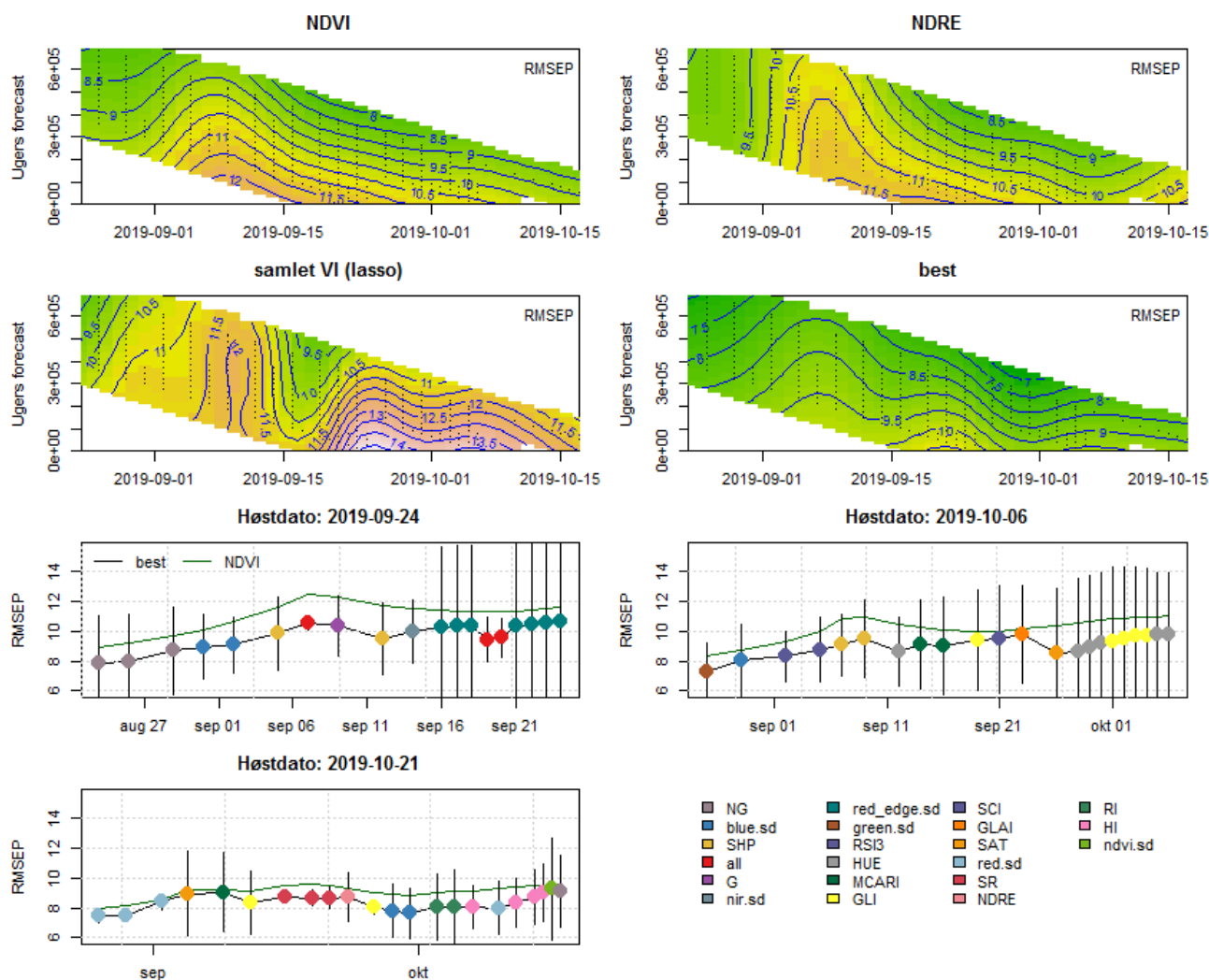


Fig. 1: TS-udbytte prædiktionssevne målt som RMSEP ud fra model I (x) inkl. en tilfældig effekt af sort i form af sortens TS% hentet fra sortsvalg majs for forsøg 030191919-001. RMSEP er vist som isokliner i rummet udspændt af tidsaksen og aksens for forecast-vinduet. Nederste grafer viser, hvorledes forskellige VI'er udviser den bedste prædiktionssevne som funktion af dato og høstdato. Standard error er vist som bjælker. Graferne viser, at prædiktionssevnen generelt er bedre for NDVI sammenlignet med både NDRE og lasso modellen. Sidstnævnte udviser tendens til træningsdata-overfit og en relativ svag test-set prædiktionssevne, hvilket kan ses i excelarket som relativ høje R^2 -værdier, men suboptimale RMSEP-værdier. Tages usikkerheden i RMSEP i betragtning er NDVI ikke signifikant dårligere i sin prædiktionssevne end den bedste prædiktionssevne. Der er fundet en tendens til svagere prædiktionssevne omkring starten af september.

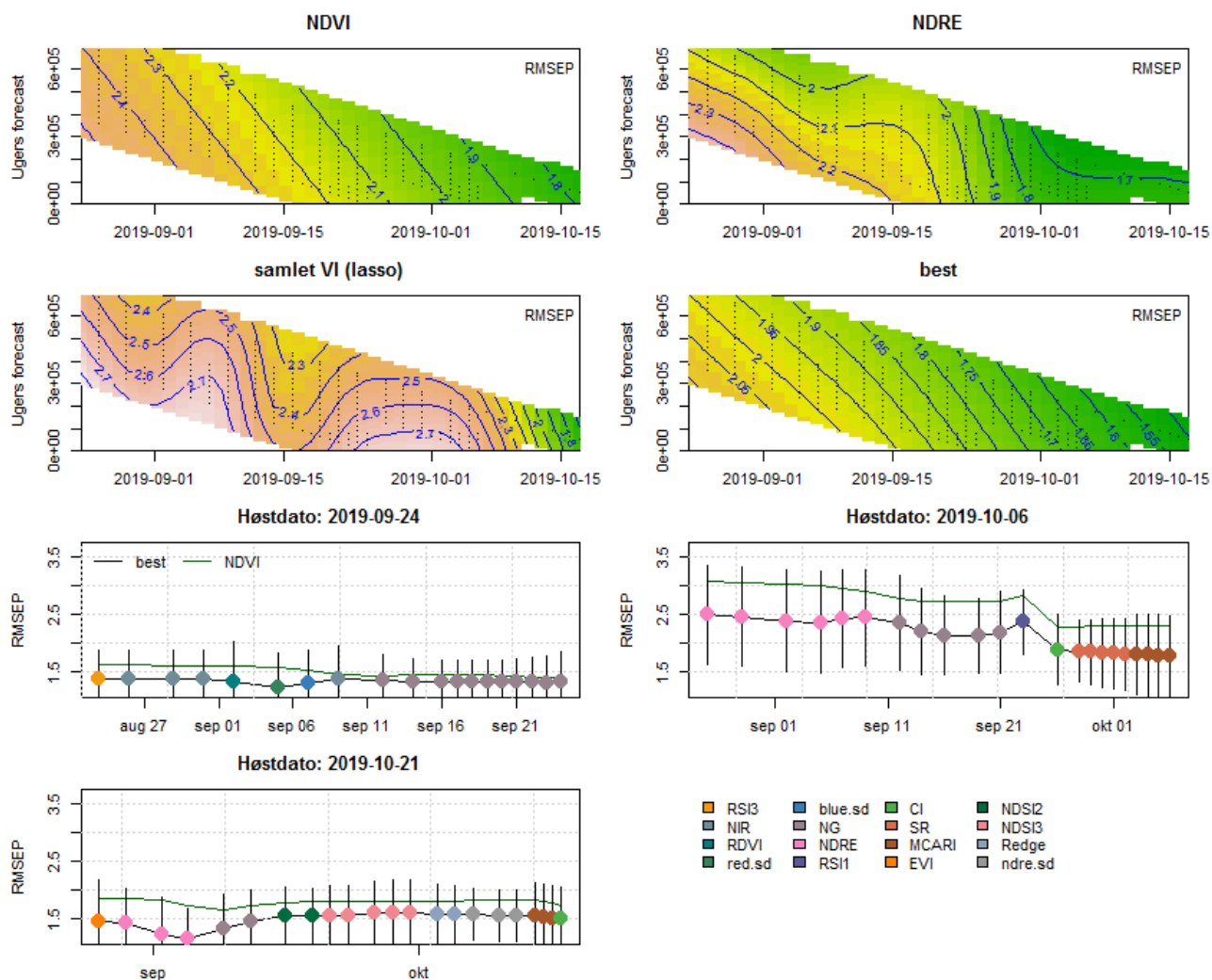


Fig. 2: Som Fig. 1, men for TS% prædiktionsniveau. NDRE udviser en bedre prædiktionsniveau end NDVI. Sidst i sæsonen er MCARI et godt alternativ. Se excelark for flere detaljer.

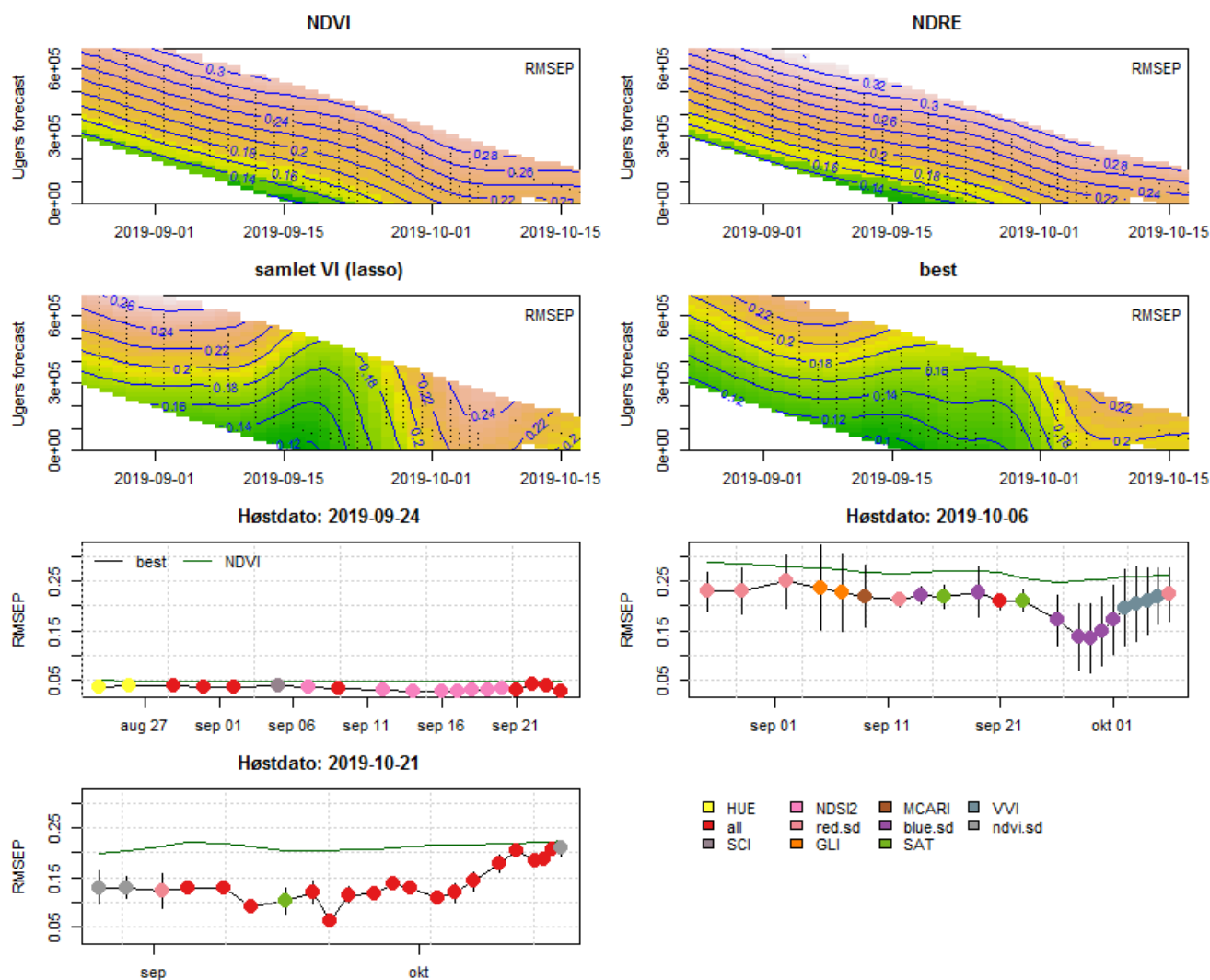


Fig. 3: Som Fig. 1, men for Protein% prædiktionssevne (og uden tilfældig effekt af sort i form af sortens TS%). Den bedste VI udviser signifikant bedre prædiktionssevne sammenlignet med NDVI. Se excelark for flere detaljer.