



Notat: Specialopgørelse af forsøgsserie 031191919 – validering af Solvis algoritme til plantetælling i majs

11-12-2019

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

Formål

1. At validere Solvis algoritme til bestemmelse af plantetal ud fra dronebilleder i majs.
2. Derudover at undersøge, om algoritmen potentielt kan forbedres ved at udnytte yderligere informationer fra dronebilledet.
3. Endeligt er der også udført en validering af bestemmelse af plantehøjde i majs ud fra Teknologisk Instituts algoritme baseret på Solvis højdekort ud fra dronebilleder.

Overordnede konklusioner

- 1) Solvis plantetalsbestemmelse udviser en systematisk underestimering (bias) af plantetal i marken, som bliver mere og mere drastisk, jo senere i sæsonen, droneoverflyvningen er foretaget, se Fig. 5 og tabel 2. Tidligt i sæsonen (14.06.) er biasen fundet til at være mellem -1.3 og -2.7 planter/m² svarende til en 13.6 til 27.7% af markens gennemsnitlige observerede plantetal observeret lige før høst. Senere i sæsonen (25.07.) er biasen fundet til at være på hele 63.5 til 73.3%. Den tilfældige prædiktionsfejl (RMSE) af solvis algoritme varierer mellem 5 og 11% af markens gennemsnitlige observerede plantetal observeret lige før høst. Endeligt er det tydeligt, at variansen i solvibestemt plantetal falder drastisk, jo senere i sæsonen, droneoverflyvningen er foretaget, hvilket betyder, at solvibestemt plantetal senere i sæsonen ikke kan anvendes til en arealspecifik plantetalsbestemmelse, f.eks. til en udpegning af områder med for lavt plantetal.
- 2) Den her udviklede algoritme til en modelkorrigeret solvibestemt plantetalsbestemmelse har effektivt kunne nedbringe særligt den systematiske fejl, se Fig. 6 og 6 og tabel 3. Tidligt i sæsonen er bias, RMSE og varians-ratioen fundet til at være optimal ($|bias_{14.06.}| < 7\%$, $RMSE_{14.06.} < 11\%$ og $variansratio_{14.06.} = 0.38$ til 0.48). Foreliggende analyse indikerer dermed, at der er et potentiale til at forbedre Solvis dronebillede-baserede plantetalsbestemmelse i majs. En implementering af en modelkorrigeret plantetalsbestemmelse ville dog kræve en yderligere undersøgelse af betydningen af lokalitet (kun to marker her), år, tid på sæsonen og en generalisering af de her anvendte parametre, herunder de anvendte tærskelværdier og noise parametre såsom cirkelens radius og dronebilledet opløsning.
- 3) Plantehøjde har kunnet blive bestemt ud fra dronens højdekort og vha. Teknologisk Instituts algoritme akkurat (bias = 5%) og med høj præcision

(RMSE = 8%) i ét af de to forsøg (001), mens der er fundet en betragtelig negativ bias (55%) i det andet forsøg (002). Det anbefales derfor, at der fremover foretages stikprøver af manuelle plantehøjdemålinger i forsøget i f.eks. 4 parceller jævnt fordelt over forsøgsarealet til en kalibrering af plantehøjde ud over at der høstes en 3 m bramme fri omkring forsøget inden droneoverflyvning.

Metoder

Forsøgsdesign

- Forsøget er beskrevet i Nordic Field Trial System [her](#).
- Registreringer i forsøget, herunder link til dronebilleder og solvibestemt plantetal tilgængeligt fra Solvi (<https://solvi.nu/>), fremgår af enkeltforsøgssiderne [her](#) og [her](#).

Baggrund

- Se Fig. 1 for et eksempel af Solvis plantetalsbestemmelse i forsøget.
- Det forventes, at Solvis algoritme underestimerer plantetal i marken, se Fig. 2 for eksempler.
- Der er derfor forsøgt at udvikle en algoritme (jf. formal 2), der med afsat i Solvis plantetalsbestemmelse kan forbedre prædiktionen af plantetal ud fra yderligere informationer tilgængeligt fra dronebilledet, nemlig hvorledes grønhed og NDVI fordeler sig indenfor et cirkulært område lige omkring de solvibestemte planter, se Fig. 3.
- Ground truth plantetal er bestemt lige før høst. Det antages derfor, at plantetal i marken har været konstant igennem sæsonen.

1) Validering

- Validering er foretaget ved at beregne bias, som er en systematisk afvigelse til de observerede værdier, og en tilfældig afvigelse beregnet som Root Mean Squared Error (RMSE) efter at der er taget højde for biasen. Den totale eller samlede afvigelse mellem observeret og prædikeret (i.e., prædiktionsfejl) inkluderer både bias og den tilfældige afvigelse. Der er testet om biasen er signifikant vha. t-tests og et 5%-signifikansniveau.
- Derudover er der beregnet forholdet mellem variansen i det observerede og prædikterede plantetal og der er bestemt, om de observerede og prædikterede værdier har samme varians vha. en F-test. Forskel i varians udtrykt som variansratioen kan derefter bruges til en vurdering om de prædikterede plantetal kan anvendes til en arealspecifik plantetalsbestemmelse, f.eks. til en udpegning af områder med for lavt plantetal. Der gælder, at ønskes der en arealspecifik prædiktion (indenfor mark prædiktion), er det afgørende, at både middelfejlen og variansfejlen er tilstrækkelig lille, mens ønskes der derimod en prædiktion kun på markniveau, er det tilstrækkeligt med en tilstrækkelig lille middelfejl (bias og RMSE) i prædiktionen.

2) Modelkorrigeret solvi-bestemt plantetal

- Til formål 2 er der anvendt en model, der sigter på at korrigere for Solvi-algorithmens eventuelle mangler. Der er kun anvendt informationer

tilgængelige fra Solvi eller dronebilledet til at forbedre Solvis prædiktions af plantetal. Der er anvendt følgende generaliserede additive model (GAM) på alle data på tværs af de to forsøg:

$$y_i \sim y_{solvi,ij} + dy_{ij}$$

$$dy_{ij} \sim \beta_0 + y_{solvi,ij} + s(\text{dato}_j, \text{Mean}(\text{SD}_{ijk})_{ij}, \Sigma(\text{Mean}_{ijk})_{ij}, \Sigma(\text{Jævnhed}_{ijk})_{ij}) + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2)$$

hvor y_i indikerer det observerede plantetal i den i 'te forsøgsparcel blandt de alle forsøgsparceller i de to forsøg, $y_{solvi,ij}$ er det solvi-bestemte plantetal i samme parcel til den j 'te dronedeflyvningsdato, $s(\cdot)$ er en thin-plate spline funktion med udglatningsgraden bestemt vha. generaliseret krydsvalidering, dato er den j 'te dronedeflyvningsdato målt i dage siden første dronedeflyvningsdato, $\text{Mean}(\text{SD}_{ijk})_{ij}$ står for den indenfor parcellen gennemsnitlige værdi af standardafvigelsen af plantepixels indenfor en cirkel med radius 10 cm centreret om hver solvibestemt plante indenfor parcellen, $\Sigma(\text{Mean}_{ijk})_{ij}$ står tilsvarende for summen indenfor parcellen af gennemsnittet af plantepixels indenfor cirklerne, og $\Sigma(\text{Jævnhed}_{ijk})_{ij}$ står for den indenfor parcellen beregnede sum af cirklernes beregnede jævnhedsparameter, se Fig. 3. Jævnhedsparameter er beregnet på følgende måde:

$$\text{Jævnhed} = (1 - \text{SD}_{\text{lav opløsning}} / \text{SD}_{\text{høj opløsning}}) * 100$$

hvor høj opløsning er Solvis højeste opløsning, afhængig af flyvehøjde og kamera (ca. 1-2 cm i tilfældet her), og lav opløsning svarer til $\frac{1}{2}$ radius af cirkler om solvibestemte planter, dvs. 5 cm. SD er standardafvigelsen af plantepixels indenfor en cirkel centreret omkring den solvibestemte koordinat for planten, hvor cirkelns radius = 10 cm. Plantepixels er bestemt ud fra en tærskelværdi i standardiseret grøn (G) og NDVI, se tabel 1, og kan antage værdierne 0 (ingen plante) eller 1 (plante).

- Residualerne e_{ij} er modelleret som værende uafhængige og normalfordelte.
- Alle analyser er udført i R (R Core Team 2019).

3) Dronebestemt plantehøjde

- Til formål 3 er dronebestemt plantehøjde beregnet på følgende måde:

$$\text{Plantehøjde (cm)} = (\text{Dronehøjdekort i m.o.h} - \text{Krigede højdekort i m.o.h}) * 100,$$

hvorved kriging-interpolering over forsøgsarealet er udført på baggrund af dronehøjde fra den omkringliggende mark, se Fig. 4, efterfulgt af parceludklip og en gennemsnitsberegning af plantehøjde per parcel.

Tabel 1: Anvendte tærskelværdier til bestemmelse af en plantepixel i dronebilledet. Tærskelværdierne blev her sat manuelt ud fra en visuel bedømmelse.

Måletid / Dato	Anvendte tærskelværdier					
	Dato	-001		-002		
		NDVI	G	Dato	NDVI	G
P02 / 29.05	28.05	0.33	0.38	29.05.	0.30	0.40
P03 / 14.06.	14.06	0.35	0.40	13.06.	0.45	0.40
P05 / 10.07.	10.07	0.85	0.46	11.07.	0.85	0.50
P06 / 25.07.	25.07	0.85	0.45	26.07.	0.80	0.465
P11 (lige før høst)	25.09.	0.90	0.45	11.09.	0.93	0.42

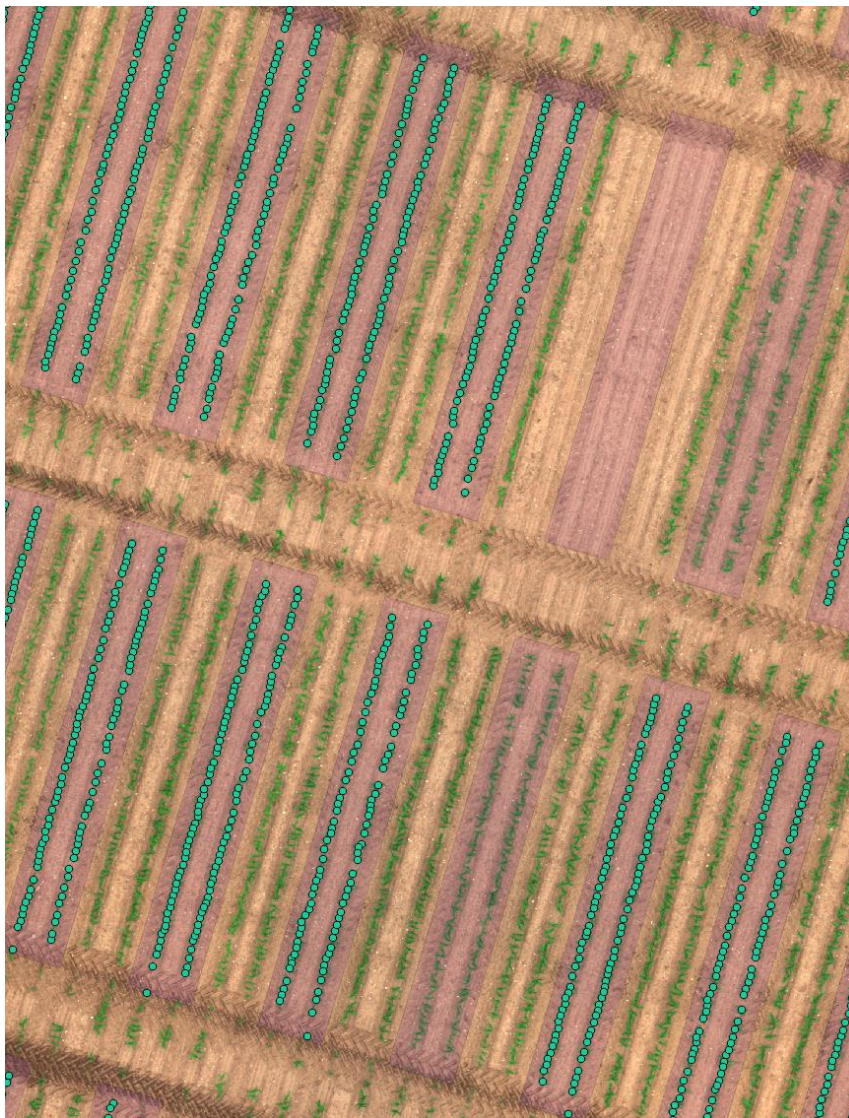


Fig. 1: Eksempel på parcellernes og solvi-bestemte planters placering i dronebilledet (udsnit).

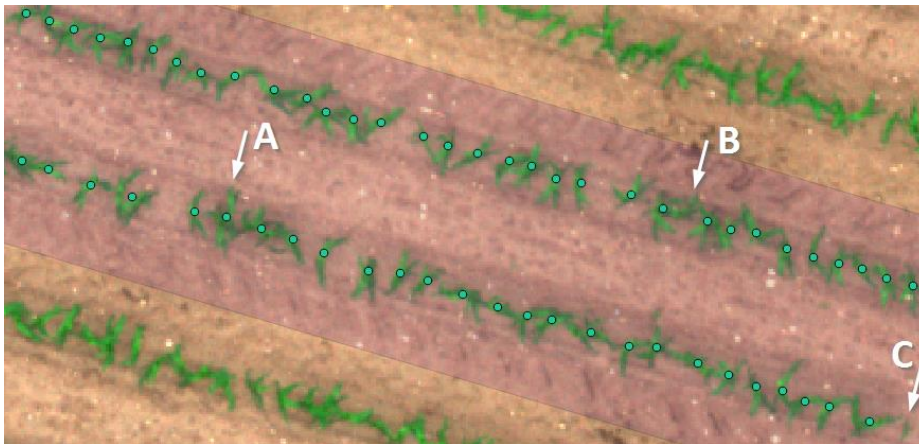


Fig. 2: Zoom in på Fig. 1, der viser udfordringer i Solvis plantegenkendelse. A: Solvi tæller én plante, hvor der formentlig er to samlet i klump. B: Solvi tæller ingen plante, hvor der formentlig er én, formentlig grundet for kort planteafstand indenfor rækken. C: Solvi tæller ingen plante, hvor der formentlig er én, formentlig grundet at planten er for lille.

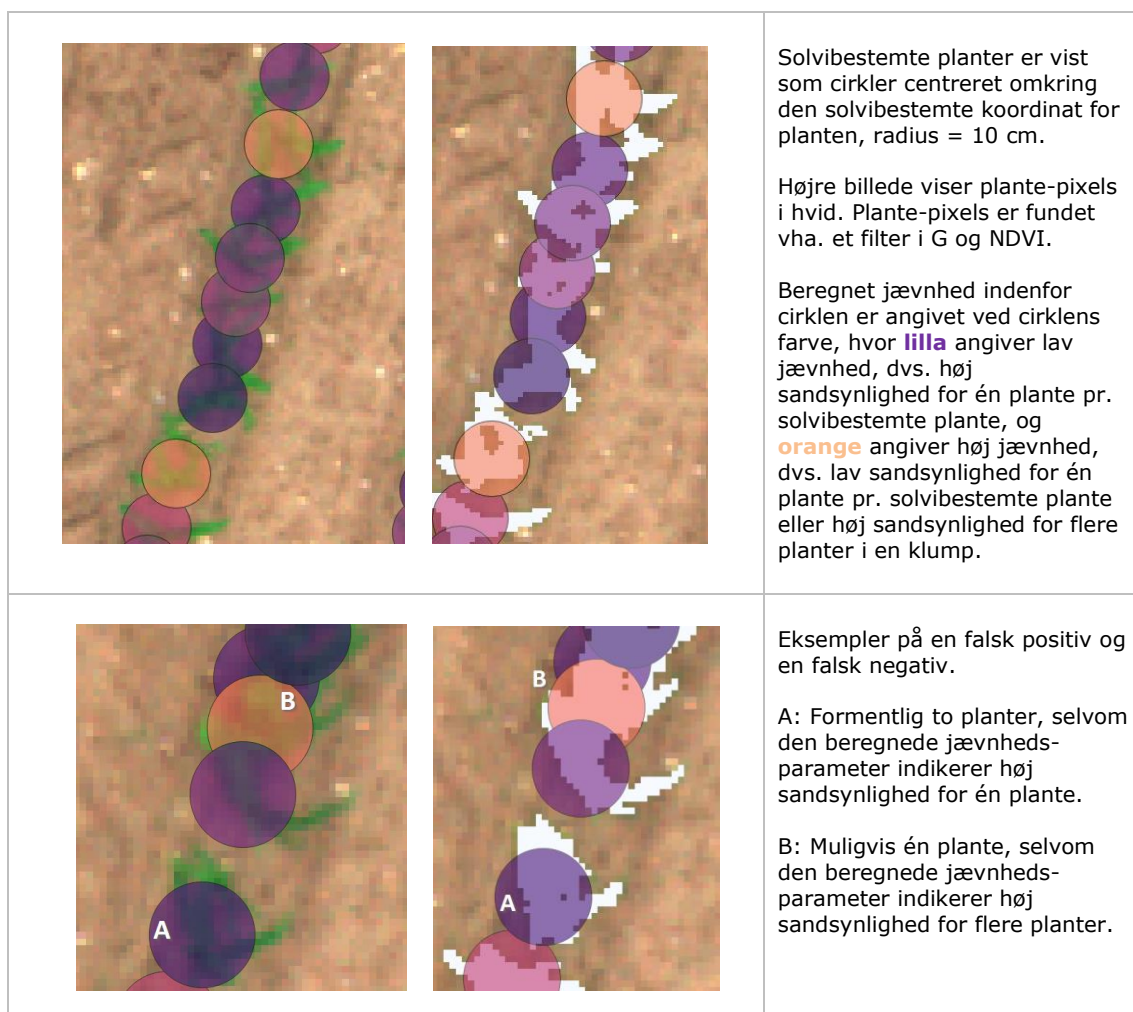


Fig. 3: Eksempler på den beregnede jævnhedsparameter per solvibestemt plante.

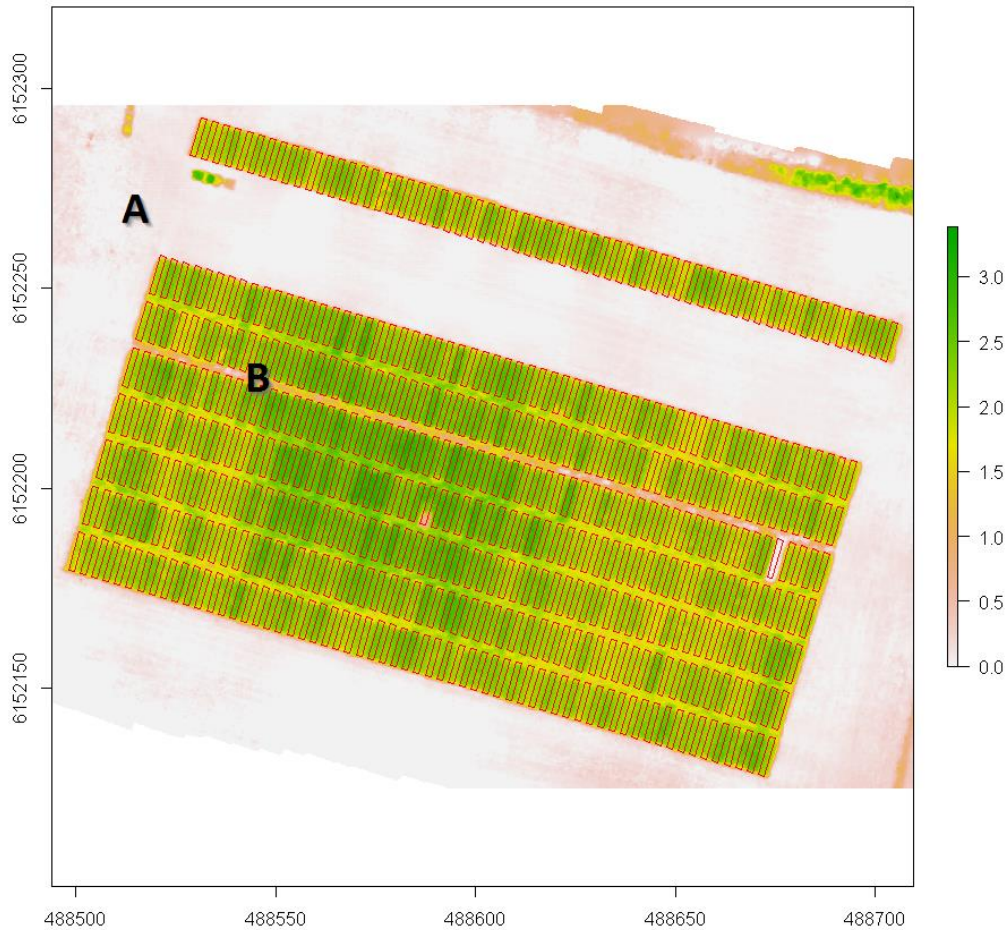


Fig. 4: Eksempel (forsøg 002) på dronebestemt plantehøjdekort, hvor plantehøjden er bestemt som forskellen mellem målt højde i forsøgsmarken og det estimerede højdelandskabet beregnet vha. kriging af den omkringliggende mark, der var høstet fri inden droneflyvningen (A). En stribe igennem forsøget (B) var formentlig også høstet fri, men er for smal og dermed påvirket af parcellerne, der grænser op til, til at blive brugt til kriging. Skala angiver plantehøjde i meter.

Resultater

- Resultater fremgår af tabel 2 til 4, samt figurerne 5 til 8.

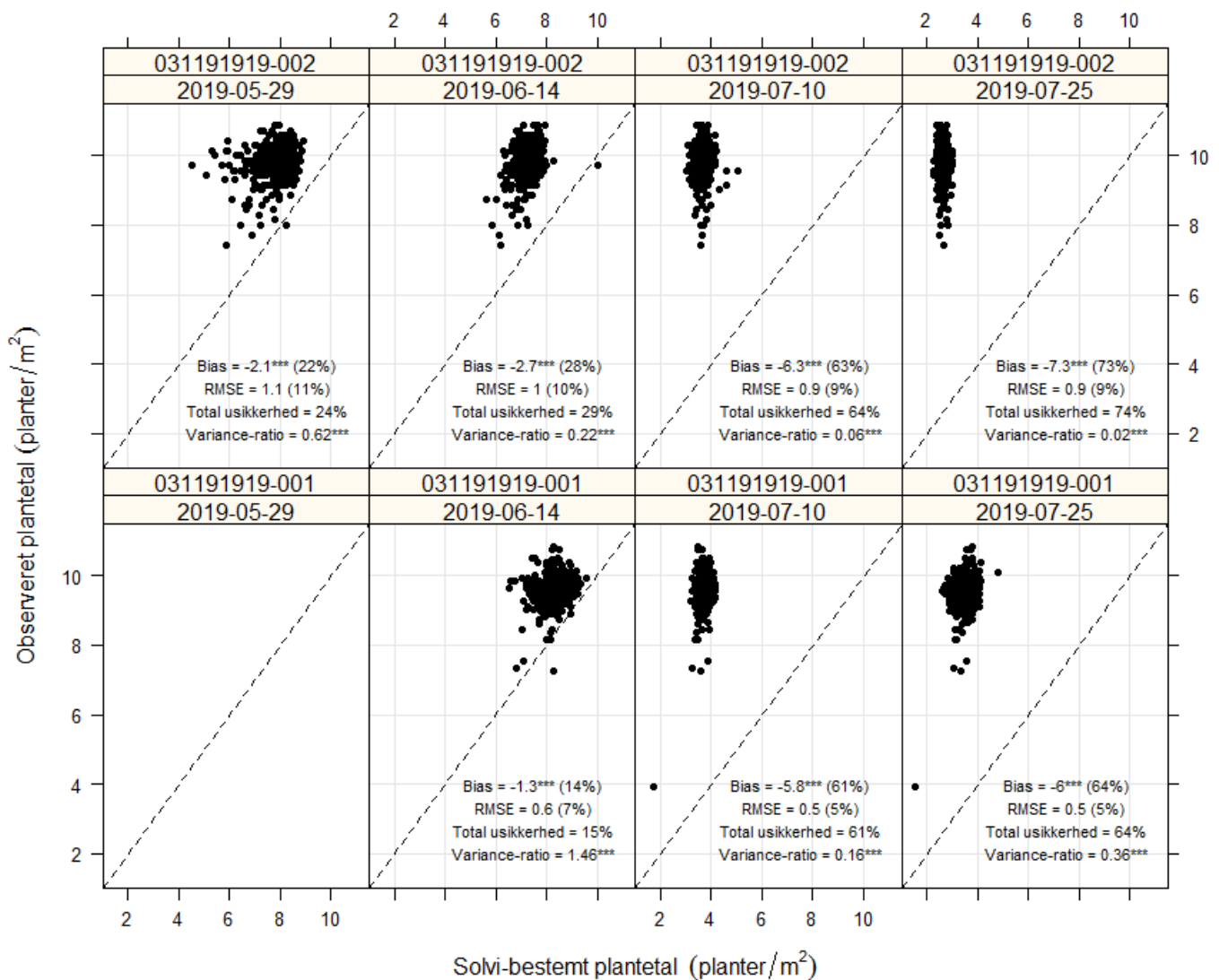


Fig. 5: Observeret plantetal (lige før høst, 12.09.) over **solvi-bestemt plantetal** for droneoverflyvninger på fire forskellige datoer i de to forsøg. Varians-ratioen angiver ratioen af variansen af drone-bestemt plantetal over variansen af det observerede plantetal. Stjerner angiver testsandsynligheder for at ratioen er 1 hhv. at bias er lig 0, hvor *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

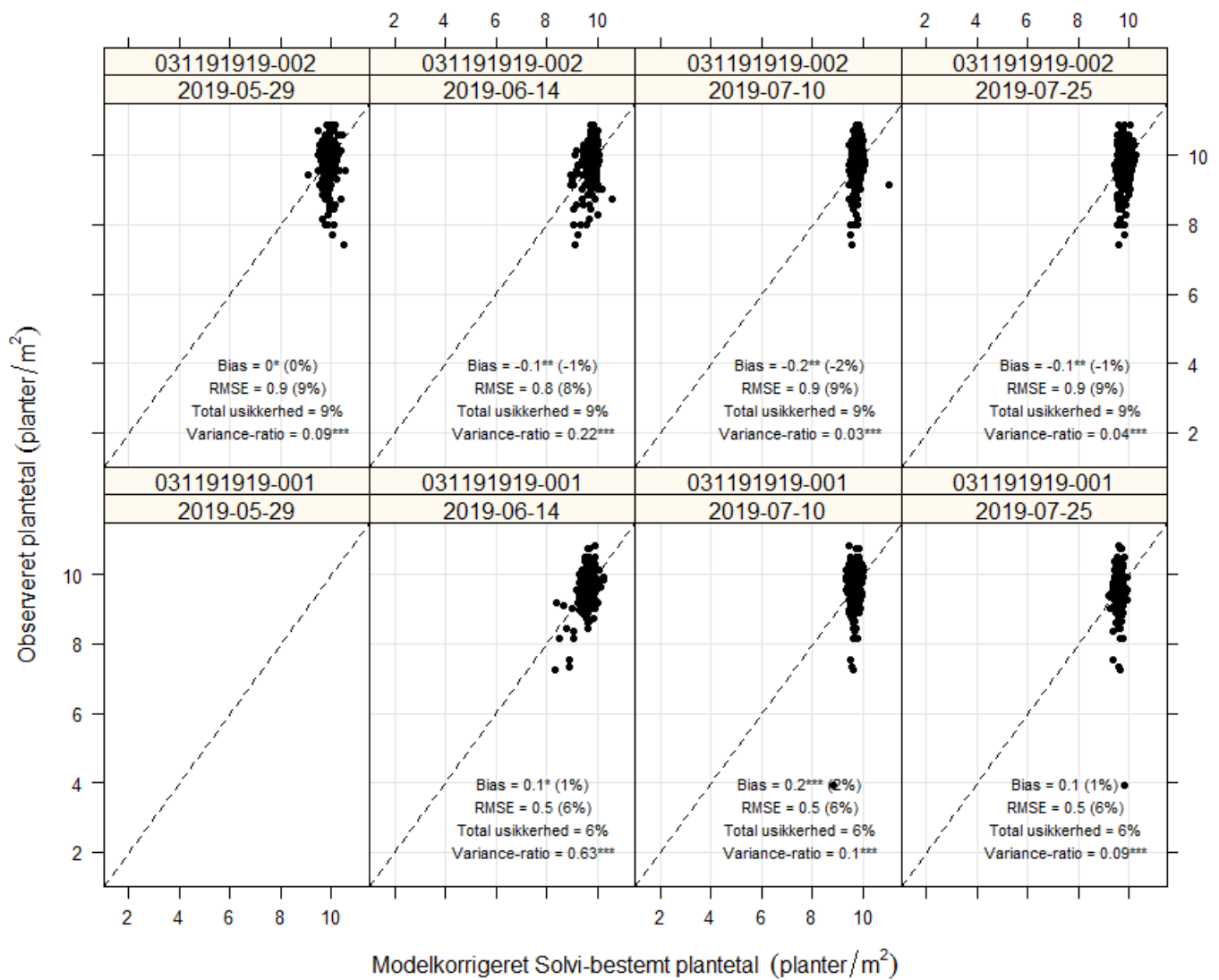


Fig. 6: Observeret plantetal (lige før høst, 12.09.) over **solvi-bestemt og modelkorrigeret plantetal** for droneoverflyvninger på fire forskellige datoer i de to forsøg. Modellens performance er evalueret vha. krydsvalidering, hvor gentagelsesblok er anvendt som hold-out set (8-fold CV).

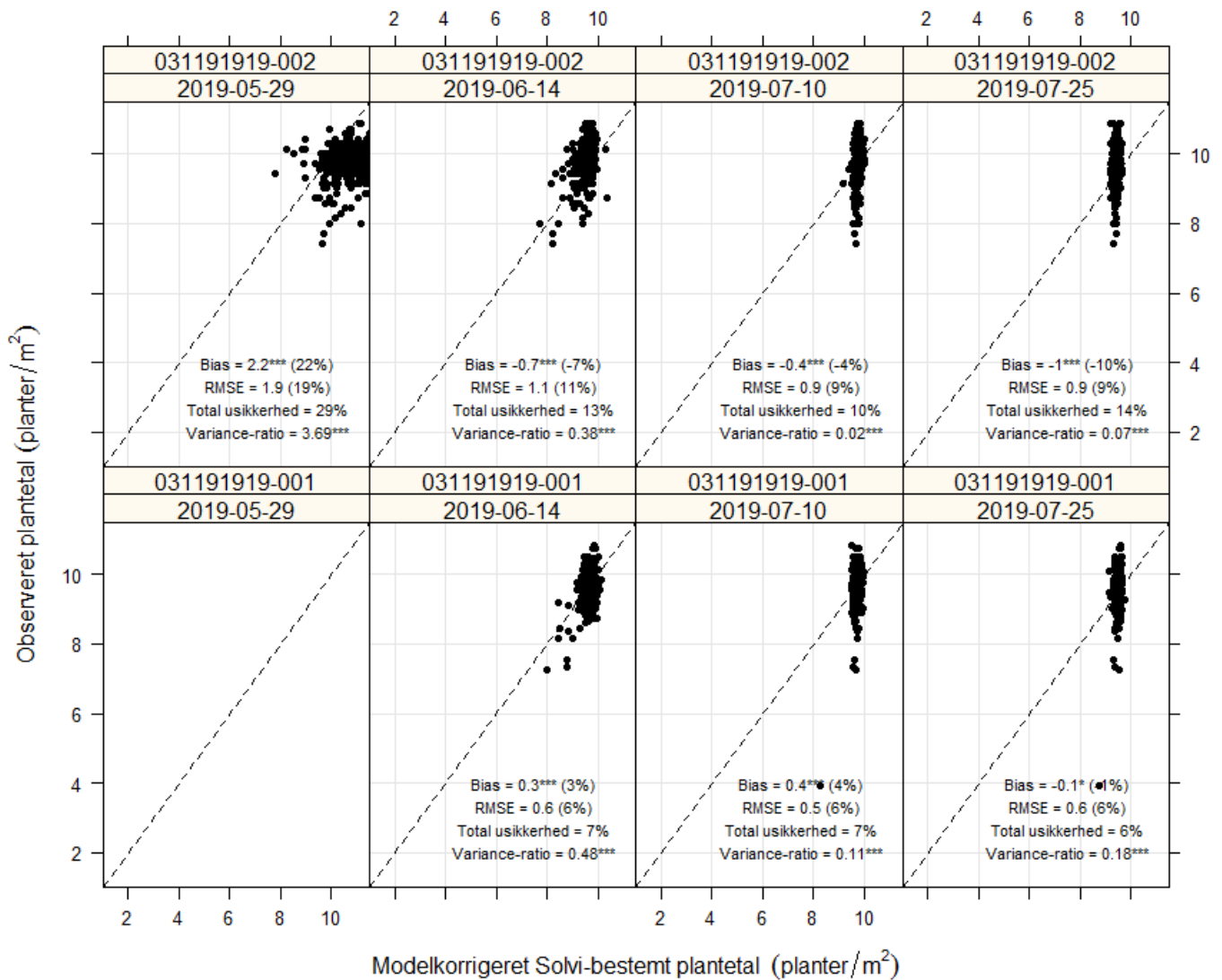


Fig. 7: Observeret plantetal (lige før høst, 12.09.) over **solvi-bestemt og modelkorrigeret plantetal** for droneoverflyvninger på fire forskellige datoer i de to forsøg. Modellens performance er evalueret vha. krydsvalidering, hvor enkeltforsøg er anvendt som hold-out set (2-fold CV).

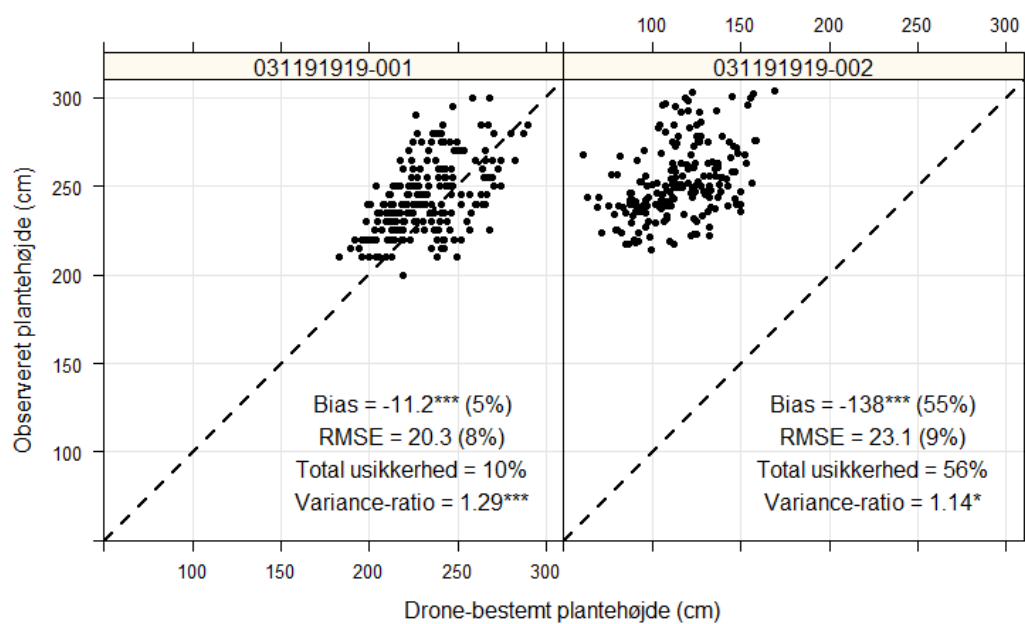


Fig. 8: Observeret over drone-bestemt **plantehøjde** i de to forsøg. Varians-ratio angiver ratioen af variansen af drone-bestemt plantehøjde over variansen af den observerede plantehøjde. Stjerner angiver testsandsynligheder for at ratioen er 1 eller at bias er lig 0, hvor *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.



Tabel 2: Valideringen af **solvibestemt plantetalsbestemmelse**. Bias angiver en systematisk afvigelse til observeret plantetal, mens RMSE (root mean squared error) angiver en tilfældig afvigelse. Den totale prædiktionsusikkerhed inkluderer både bias og RMSE. Variansratioen angiver ratioen af variansen af solvi-bestemt plantetal over variansen af det observerede plantetal. P-værdier angiver testsandsynligheder for at ratioen er 1 hhv. at bias er lig 0.

Forsøg	Plantetal (planter/m ²)				Bias			RMSE		Total usikkerhed		Varians-ratio	p-værdi
	Observeret Gns.±SD	Måletid	Dato	Solvibestemt Gns.±SD	planter/m ²	%	p-værdi	planter/m ²	%	planter/m ²	%		
031191919-001	9.5±0.5	P03	14-6-2019	8.2±0.6	-1.3	13.6	<0.001 ***	0.6	6.6	1.4	15.2	1.5	<0.001 ***
		P05	10-7-2019	3.7±0.2	-5.8	61.3	<0.001 ***	0.5	5.0	5.9	61.5	0.2	<0.001 ***
		P06	25-7-2019	3.5±0.3	-6.0	63.5	<0.001 ***	0.5	5.4	6.1	63.7	0.4	<0.001 ***
031191919-002	9.9±0.9	P02	29-5-2019	7.8±0.7	-2.1	21.6	<0.001 ***	1.1	11.5	2.4p	24.5	0.6	<0.001 ***
		P03	14-6-2019	7.2±0.4	-2.7	27.7	<0.001 ***	1.0	10.1	2.9	29.5	0.2	<0.001 ***
		P05	10-7-2019	3.6±0.2	-6.3	63.3	<0.001 ***	0.9	9.3	6.3	64.0	0.1	<0.001 ***
		P06	25-7-2019	2.6±0.1	-7.3	73.3	<0.001 ***	0.9	9.1	7.3	73.9	0.0	<0.001 ***

Tabel 3: Valideringen af **solvibestemt og modelkorrigeret plantetals-bestemmelse**. Bias angiver en systematisk afvigelse til observeret plantetal, mens RMSE (root mean squared error) angiver en tilfældig afvigelse. Den totale prædiktionsusikkerhed inkluderer både bias og RMSE. Varians-ratioen angiver ratioen af variansen af solvi-bestemt plantetal over variansen af det observerede plantetal. P-værdier angiver testsandsynligheder for at ratioen er 1 hhv. at bias er lig 0. Modelusikkerheden er evalueret på to måder: krydsvalidering, hvor gentagelsesblok er anvendt som hold-out set (8-fold CV) og krydsvalidering, hvor enkeltforsøg er anvendt som hold-out set (2-fold CV).

Forsøg	Plantetal (planter/m ²)				Bias ± s.e.			RMSE ± s.e.		Total usikkerhed ± s.e.		Varians-ratio ± s.e.	p-værdi			
	Observeret Gns.±SD	Måletid	Dato	Solvibestemt Gns.±SD	planter/m ²	%	p-værdi	planter/m ²	%	planter/m ²	%					
gentagelsesblok - 8-fold CV	031191919-001	9.5±0.5	P03	14-6-2019	9.6±0.3	0.1±0.1	1.3	0.020	*	0.5±0.2	5.7	0.5±0.2	5.7	0.63±0.89	<0.001	***
			P05	10-7-2019	9.7±0.1	0.2±0.1	2.0	0.000	***	0.5±0.2	5.6	0.5±0.2	5.8	0.10±0.10	<0.001	***
			P06	25-7-2019	9.6±0.1	0.1±0.1	0.9	0.135		0.5±0.2	5.7	0.5±0.2	5.5	0.09±0.05	<0.001	***
	031191919-002	9.9±0.9	P02	29-5-2019	9.9±0.3	0.0±0.3	0.1	0.012	*	0.9±0.2	9.0	0.9±0.2	9.2	0.09±0.11	<0.001	***
			P03	14-6-2019	9.8±0.4	-0.1±0.2	1.2	0.008	**	0.8±0.2	8.4	0.9±0.2	8.6	0.22±0.22	<0.001	***
			P05	10-7-2019	9.7±0.1	-0.2±0.3	1.8	0.001	**	0.9±0.2	8.7	0.9±0.2	9.0	0.03±0.01	<0.001	***
P06			25-7-2019	9.8±0.2	-0.1±0.3	1.2	0.003	**	0.9±0.2	8.8	0.9±0.2	9.0	0.04±0.02	<0.001	***	
enkeltforsøg - 2-fold CV	031191919-001	9.5±0.5	P03	14-6-2019	9.8±0.4	0.3	3.4	<0.001	***	0.6	6.4	0.7	7.2	0.48	<0.001	***
			P05	10-7-2019	9.9±0.2	0.4	3.8	<0.001	***	0.5	5.7	0.7	6.9	0.11	<0.001	***
			P06	25-7-2019	9.5±0.2	-0.1	0.7	0,022	*	0.6	6.1	0.6	6.2	0.18	<0.001	***
	031191919-002	9.9±0.9	P02	29-5-2019	12.1±1.7	2.2	22.1	<0.001	***	1.9	19.1	2.9	29.2	3.69	<0.001	***
			P03	14-6-2019	9.2±0.5	-0.7	7.1	<0.001	***	1.1	11.0	1.3	13.1	0.38	<0.001	***
			P05	10-7-2019	9.5±0.1	-0.4	3.7	<0.001	***	0.9	8.9	1.0	9.7	0.02	<0.001	***
P06			25-7-2019	8.9±0.2	-1.0	9.8	<0.001	***	0.9	9.3	1.3	13.5	0.07	<0.001	***	



Tabel 4: Valideringen af **dronebestemt plantehøjde**. Bias angiver en systematisk afvigelse til observeret plantehøjde, mens RMSE (root mean squared error) angiver en tilfældig afvigelse. Den totale prædiktionsusikkerhed inkluderer både bias og RMSE. Varians-ratioen angiver ratioen af variansen af solvi-bestemt plantehøjde over variansen af den observerede plantehøjde. P-værdier angiver testsandsynligheder for at ratioen er 1 hhv. at bias er lig 0.

Forsøg	Plantehøjde (cm)		Bias			RMSE		Total usikkerhed		Varians-ratio	p
	Observeret Gns.±SD	Solvibestemt Gns.±SD	cm	%	p	cm	%	cm	%		
001	243.0 ± 20.2	230.0 ± 23.0	-11.2	4.6	<0.001 ***	20.3	8.3	23.2	9.5	1.29	<0.001 ***
002	251.6 ± 20.8	111.2 ± 22.2	-138.0	54.9	<0.001 ***	23.1	9.2	140.0	55.6	1.14	0.032 *

Referencer

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.