



Notat: Spørgeskemaundersøgelse om Lupinmarker

29-12-2023

Svend Vendelbo Nielsen, Konsulent
sven@teknologisk.dk, +45 72 20 17 41

Formål

- Baseret på en serie af spørgeskemaer om dyrkningen af lupin, der løb fra 2021 til 2023, finder vi de faktorer der har betydning for udbyttet af lupin.

Overordnede konklusioner

- Analysen viser at følgende faktorer er relateret til et højere udbytte:
 - En medium sådybde (3,5 cm - 4,0 cm)
 - En tidlig såning (før 1.april)
 - Et lavere ukrudtstal (specielt for forgrenede lupintyper)
- Analysen antyder at følgende faktorer kunne være relateret til et højere udbytte
 - Anvendelse af dybstrøelse og i mindre grad anvendelse af gylle
 - Et lavt reaktionstal
 - Større udsædsmængde for uforgrenede lupintyper
- Analysen finder det ikke sandsynligt at følgende er relateret til betydeligt ændret udbytte
 - Blindharvning
 - Forfrugt
 - Forforfrugt
 - Vanding
 - År

Bemærkninger

- Analysen er baseret på et observationsstudie, som er mere sårbart over for uobserverede, confounding variable. Man skal derfor overveje grundigt om andre ting kan forklare de korrelationer, som analysen viser, før man slutter at der er en kausal sammenhæng.

Data

- Data består af svar fra en spørgeskemaundersøgelse. Hver række i datasættet repræsenterer et felt. Der er 36 rækker med data fra året 2021 og 30 rækker med data fra 2022 og 28 fra 2023. For både 2021 og 2022 findes der et felt uden information om lupinudbyttet, hvilket er udelukket fra analysen. Af de mange faktorer i spørgeskemaerne inkluderer vi vores analyse følgende
 - udbytte ("Udbytte (hkg pr. ha)")
 - plantetal ("Plantebestand (lupin) planter pr. m²")
 - jordtype ("Jordtype på arealerne med lupin")
 - rt ("RT, reaktionstallet")
 - forfrugt ("Forfrugt til lupin")



- forforfrugt ("Angiv forforfrugten til lupin")
 - husdyrgødning ("Er der anvendt husdyrgødning?", "Hvilken type husdyrgødning er anvendt?")
 - renbestand ("Har du lupin i renbestand eller blandsæd?")
 - sort ("Sortsnavn - Lupin")
 - podning ("Er lupin udsæden podet?")
 - usdsædsmængde ("Udsædsmængde pr. ha")
 - blandet_art ("Hvilken anden planteart er lupinen blandet med?")
 - blandet_mængde ("Hvilken udsædsmængde er anvendt af denne anden planteart?")
 - sådato ("Dato for såning (dd.mm.yyyy)")
 - sådybde ("Sådybde (cm)")
 - jordtemperatur ("Jordtemperatur ved såning")
 - blindharvning ("Er marken blindharvet?")
 - vanding ("Er markerne med lupin blevet vandet?")
 - ukrudt ("Ukrudt (% dækning af jorden)")
- For at fokusere analysen på de vigtigste forskelle, har vi grupperet svar for følgende variable:
- forfrugt er grupperet i "Vintersæd", "Vårsæd" og "Andet"
 - forforfrugt er grupperet i "Bælgfrugt" og "Andet"
 - sådato er grupperet i "tidlig" (før slutningen af marts) og "sen" (efter begyndelsen af april)
 - sådybde er grupperet i "lav" (< 3,5 cm), "mellem" (3,5 cm - 4,0 cm) og "høj" (> 4,0 cm)
 - husdyrgødning er grupperet i "dybstrøelse", "gylle" og "ingen".
 - Jordtype er grupperet i "sandet" (<5) og "leret" (>=5).

Metode

For at identificere de betydende faktorer, laver vi først en eksplorativ undersøgelse for at afdække hvilke faktorer der kan inkluderes i modellen og hvordan de kontinuerte og diskrete variable eventuelt skal grupperes.

Vi har derefter identificeret hvilke faktorer og sammenhænge der er mest sandsynlige a priori og fokuserer derefter resten af analysen på disse.

For at håndtere det store antal faktorer i datasættet, bruger vi til at modellere udbyttet en hierarkisk bayesiansk model, hvis matematiske definition kan ses i Appendix. Denne model kan beskrives med en graf (se Figur 1). Plantetal og ukrudt bliver behandlet som medierende faktorer, hvilket vil sige at flere faktorer får både direkte effekter på udbyttet samt indirekte effekter igennem plantetal og ukrudt.

Valget af hvilke forklarende faktorer, der bruges til at prædiktere henholdsvis udbytte, ukrudtstal og plantetal, bygger på ekspertviden. For at begrænse modellen til de vigtigste sammenhænge, har vi reduceret modellen guidet af WAIC og approksimative "Leave-One-Out" crossvalidation. Konkret undersøgte vi iterativt om fjernelse af den numerisk mindste ekspertprædikterede sammenhæng ville føre til en bedre model ifølge kriterierne.

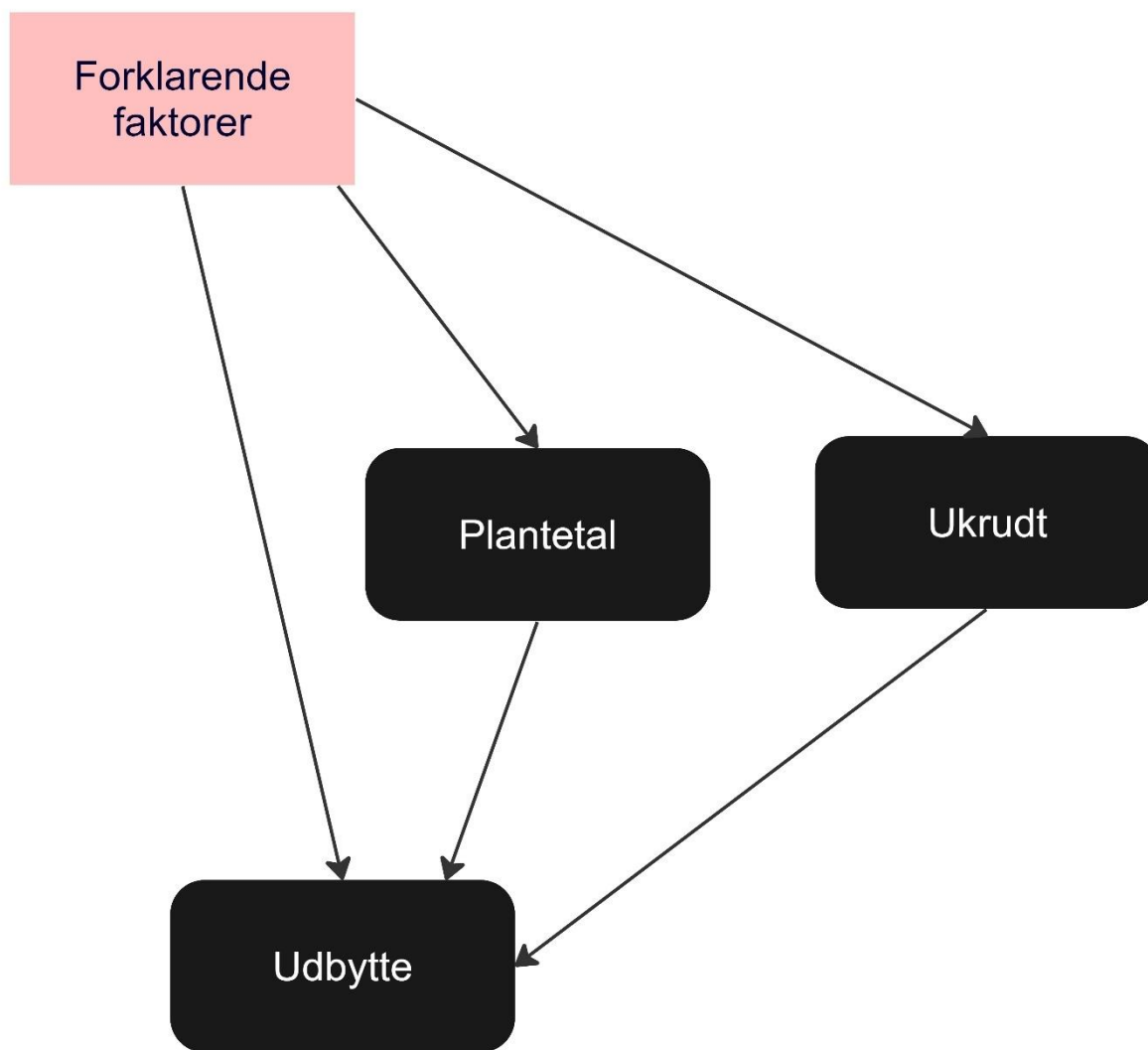
Selvom modellen repræsenteres med en graf vil det være farligt ukritisk at fortolke alle sammenhænge som kausale. Forsøget er ikke designet og randomiseret, så den forestående analyse er en undersøgelse af korrelationerne i datasættet. Et problem kan for eksempel opstå, når vi betragter effekten af vanding på udbyttet. Antag for eksemplets skyld, at det kun er de mest tørkestressede marker som bliver vandet. Det ville betyde at de vandede marker har et dårligere udgangspunkt end de ikke-vandede marker og at vandingen kun



ville få dem til at "indhente" de ikke-vandede marker. Derfor ville vi ikke forvente at modellen ville opfange en positiv effekt af vanding – også selvom alle marker havde godt af at blive vandet.

Modellen tjekkes gennem posterior predictive checks (Se appendix).

Alle faktorer testes også individuelt i en marginal model. Dette kan belyse kompleksiteten af relationen mellem en faktor og et outcome, da forskellige konklusioner i den marginale og i den fulde model tyder på, at der kan være uopdagede årsagssammenhænge.



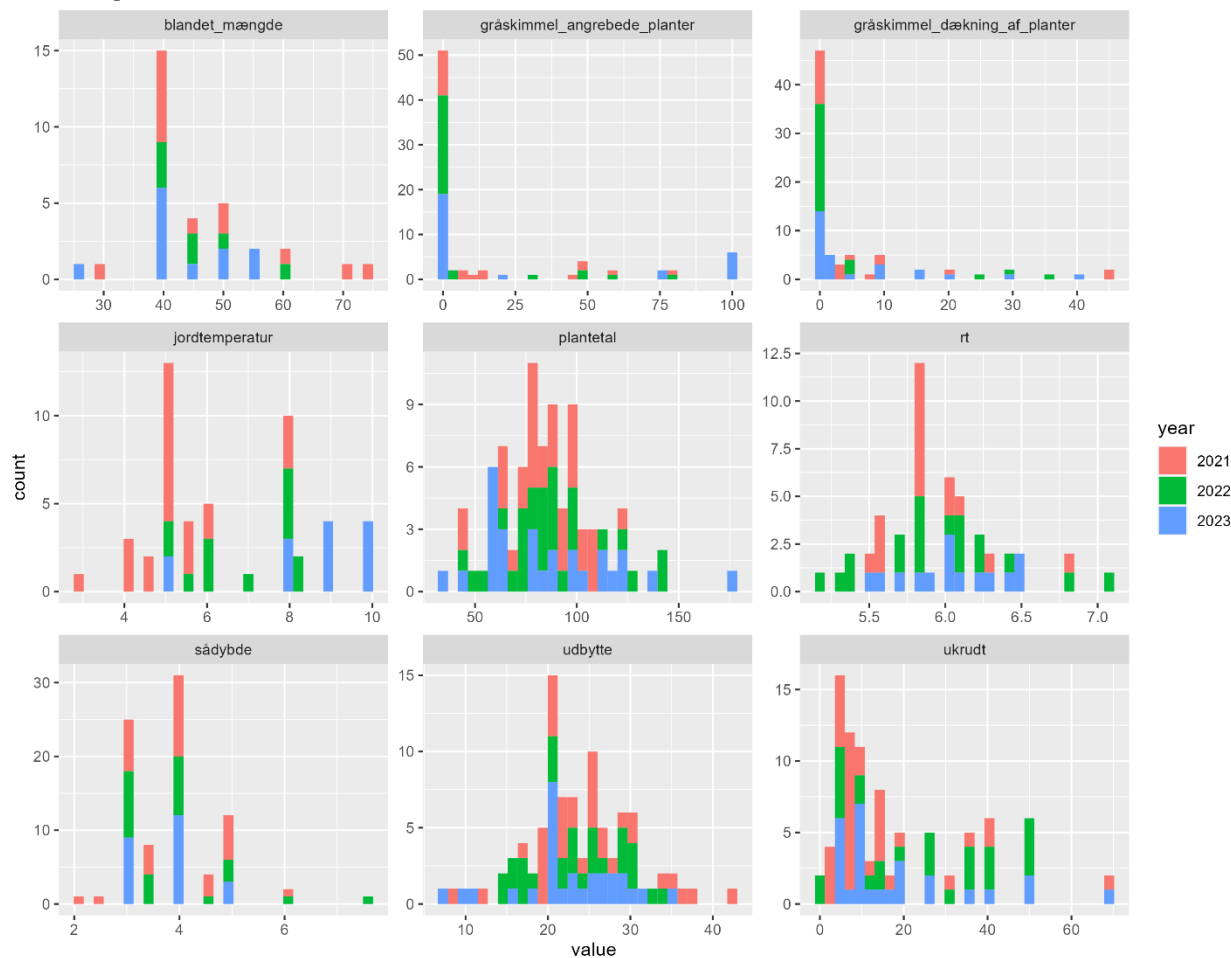
Figur 1: Modellens opbygning der illustrerer hvordan nogle af de forklarende faktorer kan påvirke udbyttet indirekte gennem Plantetal og Ukrudt, men at de forklarende faktorer også kan påvirke udbyttet direkte.

Resultater

Resultaterne fra den eksplorative analyse ses i Figur 2-5. Vi har fjernet podning som faktor, da der var et meget lille antal ikke-podede lupinmarker. Derudover er blandet_mængde, blandet_art, gråskimmel_angrebede_planter og gråskimmel_dækning_af_planter vist, men disse er ikke inkluderet i den videre analyse.



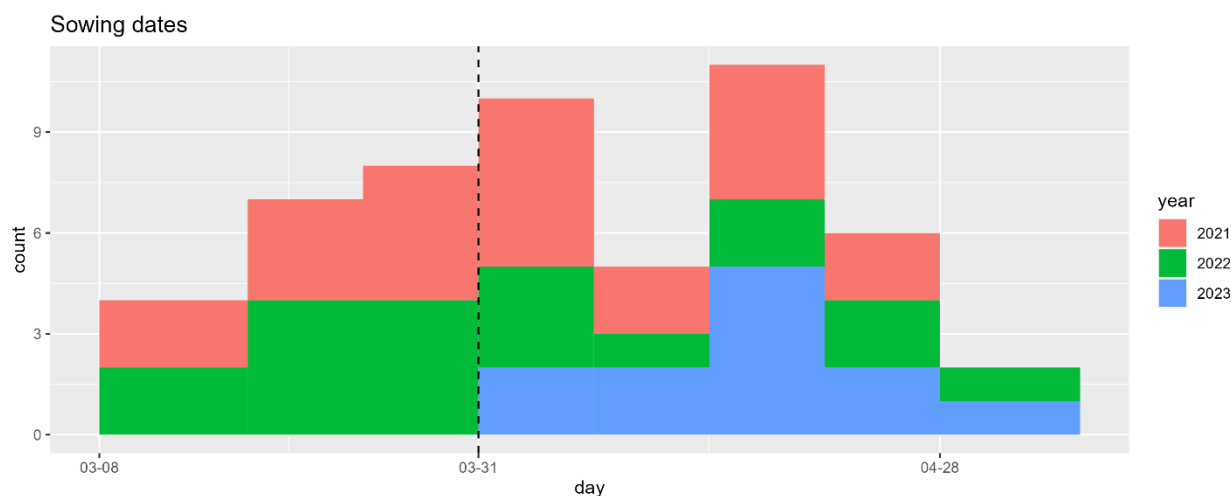
Histograms of the numeric variables



Figur 2: Histogram over de numeriske variable, fordelt på år.



Figur 3: Histogram af observationerne fra de diskrete og diskretiserede faktorer.

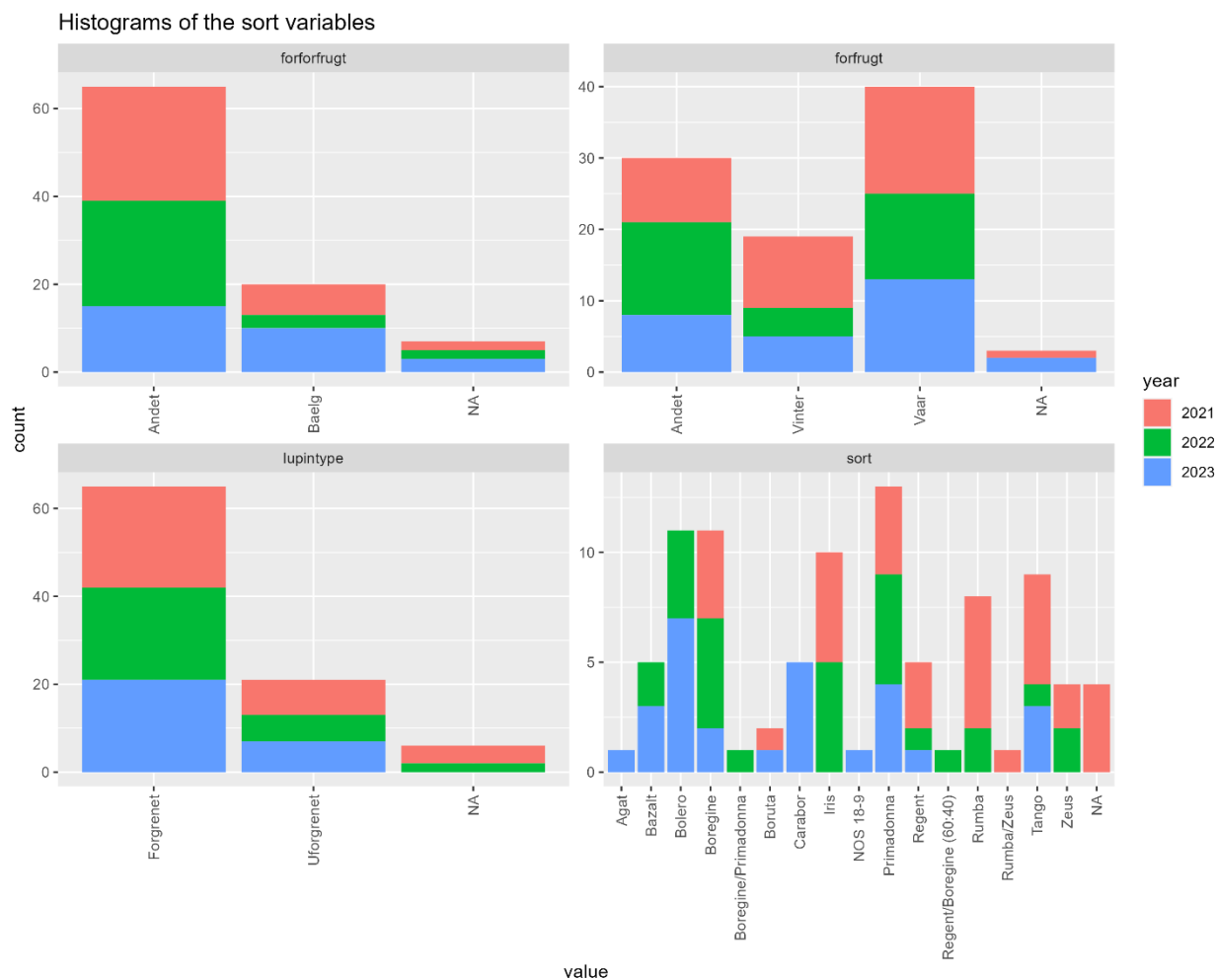


Figur 4: Sådato stratificeret på de tre år. Den stiplede linje indikerer skillelinjen mellem "tidlig" og "sen" såning.

Den bayesianske hierarkiske model blev afprøvet med forskellige sæt af forklarende faktorer (Tabel 1). En endelig model blev udvalgt, og dens estimerede parametre er vist i Tabel 2a, 2b og 3 med 95% HDI-interval. Flere af de inkluderede forklarende faktorer har både en direkte effekt på udbyttet samt en indirekte effekt på udbyttet gennem ukrudts- og/eller plantetallet. Summen af de direkte og indirekte effekter giver totaleffekten. Den bayesianske model fungerer ved at den har en standardantagelse om alle parametre - også kaldet prior - som bliver opdateret baseret på det observerede data, hvilket giver den posterior fordeling (Figur 7 og 8).

Den endelige bayesianske model har en R^2 på 0.43, hvilket vidner om at der stadig er meget usikkerhed, som ikke bliver forklaret af modellen. Det kan ses på sammenligningen mellem de prædikterede og faktiske værdier i Figur 6. Selvsamme figur viser også at der ikke er alvorlige outliers i datasættet. Datasættet indeholder dog flere udbytteværdier omkring 20 end modellen forventer (Appendix, Figur A1 og A2). Dette kan til dels skyldes respondenternes afrunding af deres svar.

På trods af den marginale signifikans af reaktionstallet (RT) er dens ikke effekt ikke signifikant i den initiale model. Når RT stiger med 0.38 (svarende til en standardafvigelse) påvirker det udbyttet på sandede jorde med -2.4 (Highest density interval (HDI) [-6.3, 1.5]). For lerede jorde bliver udbyttet påvirket med 0.50 (HDI = [-3.1, 4.3]). Dog har den initiale model problemer at estimere WAIC og LOO-kriteriet, fordi der er så mange parametre og missing værdier i modellen. Da RT indeholder hele 43 missing værdier i datasættet, forbedrer det modellens evne til at estimere kriterierne, når den bliver kørt uden RT. Vi valgte derfor at fortsætte analysen uden denne faktor. Der kan stadig godt være en reel effekt af RT på udbyttet - især for sandede jorde - men det vil formentligt kræve flere RT-observationer at estimere dette i en samlet model.



Figur 5: Sortsfaktorerne vist i histogrammer stratificeret på de tre år. Det store antal lupinsorter, forsimples til to lupintyper: Forgrenet og uforgrenet.



Tabel 1: Der blev overvejet flere mulige årsagssammenhænge for den hierarkiske, bayesianske model. Denne tabel giver et overblik over alle dem, der blev testet

FORKLARENDE FAKTOR	OUTCOME	KOMMENTAR	INKLUDERET I FULDE MODEL
JORDTYPE	Udbytte		Ja
JORDTYPE * VANDING	Udbytte	Ekskluderet fra modellen på grund WAIC/LOO-kriterierne	Nej
SÅDATO * ÅR	Udbytte		Ja
SÅDYBDE * BLINDHARVNING	Udbytte		Ja
HUSDYRGØDNING	Udbytte		Ja
JORDTYPE*RT	Udbytte	Ekskluderet fra modellen fordi robustheden blev øget ved at fjerne den. Dette kan skyldes at der var mange missing værdier for RT	Nej
FORFRUGT	Udbytte	Ekskluderet fra modellen på grund af WAIC- kriteriet	Nej
PLANTETAL * LUPINTYPE	Udbytte		Ja
UKRUDT * LUPINTYPE	Udbytte		Ja
LUPINTYPE	Ukrudt		Ja
BLINDHARVNING	Ukrudt		Ja
RENBESTAND	Ukrudt		Ja
UDSÆDSMÆNGDE	Plantetal		Ja
RENBESTAND	Plantetal		Ja
SÅDYBDE * BLINDHARVNING	Plantetal		Ja
HUSDYRGØDNING * RT	Udbytte	Var ikke en del af den initiale model. Den blev afprøvet men forbedrede ikke WAIC -kriteriet.	Nej
SÅDATO * ÅR- EFFEKT	Plantetal	Var ikke en del af den initiale model. Det blev testet om den effekt, som sådato*år havde på udbyttet kunne skaleres om til en effekt på plantetallet. Det forværrede WAIC-kriteriet og havde ikke nogen nævneværdig effekt.	Nej
SÅDATO * ÅR- EFFEKT	Ukrudt	Var ikke en del af den initiale model. Det blev testet om den effekt, som sådato*år havde på udbyttet kunne skaleres om til en effekt på plantetallet. Det forværrede WAIC-kriteriet og havde ikke nogen nævneværdig effekt.	Nej
SÅDATO	Plantetal	Var ikke en del af den initiale model. Den blev afprøvet men øgede ikke WAIC-kriteriet.	Nej
SÅDATO	Ukrudt	Var ikke en del af den initiale model. Den blev afprøvet men øgede ikke WAIC-kriteriet.	Nej



Tabel 2a: Denne tabel viser estimaterne i den bayesianske model. Alle parametre måler en effekt udtrykt i forventet udbytte når man går fra "PARAMETER SAMMENLIGNING" til "PARAMETER LEVEL". Den direkte effekt angiver effekten når plantetal og ukrudtstal er kendt, mens den totale effekt angiver effekten for et ukendt plantetal og ukrudtstal. Tallene i firkantede parenteser angiver et 95% Highest Density Interval(HDI) og eventuelle stjerner derefter angiver størrelsesordenen af 2 gange den posterior sandsynlighed for at parameteren er større end eller mindre end 0 – dvs. en slags bayesiansk p-værdi. Niveauerne er ***: >0.99, **:>0.95, *: >0.8.

PARAMETER LEVEL	PARAMETER SAMMENLIGNING	DIREKTE EFFEKT	INDIREKTE EFFEKT (FORGRENET /UFORGRENET)	TOTAL EFFEKT (FORGRENET / UFORGRENET)
År=2021, sådato=sen	År=2023, sådato=sen	1.1 [-2.0,4.2]		1.1 [-2.0,4.2]
År=2022, sådato=sen	År=2023, sådato=sen	-0.2 [-3.6,3.5]		-0.2 [-3.6,3.5]
År=2021, sådato=tidlig	År=2021, sådato=sen	4.0 [-0.0,7.9]*		4.0 [-0.0,7.9]*
År=2022, sådato=tidlig	År=2022, sådato=sen	6.7 [2.7,10,9]***		6.7 [2.7,10,9]***
Sådybde=dyb, blindharvning=nej	Sådybde=lav, blindharvning=nej	0.3 [-3.5,4.3]	Gennem plantetal: -0.1 [-0.9,0.7] / -0.7 [-3.2,1.1]	0.2 [-3.6,4.0] / -0.4 [-4.9, 3.7]
Sådybde=medium, blindharvning=nej	Sådybde=lav, blindharvning=nej	4.0 [0.7, 7.2]**	Gennem plantetal: -0.0 [-0.6,0.4] / -0.4 [-2.3, 0.9]	4.0 [0.7,7.1]** / 3.6 [-0.1, 7.2]*
Sådybde=dyb, blindharvning=ja	Sådybde=lav, blindharvning=nej	0.2 [-4.4, 4.7]	Gennem plantetal: -0.3 [-1.4, 0.8] / 0.1 [-2.1,2.7] Gennem ukrudt: -0.3 [-1.2, 0.5] / -0.2 [-1.0, 0.5]	-0.4 [-5.2, 4.6] / 0.1 [-5.0, 5.5]
Sådybde=medium, blindharvning=ja	Sådybde=lav, blindharvning=nej	2.6 [-1.3, 6.6]*	Gennem plantetal: -0.1 [-0.7,0.5] / -0.5 [-2.4, 1.3] Gennem ukrudt: -0.3 [-1.2, 0.5] / -0.2 [-1.0, 0.5]	2.3 [-1.8, 6.3] / 2.0 [-2.5, 6.3]
Sådybde= lav, Blindharvning=ja	Sådybde=lav, blindharvning=nej	0.5 [-3.0, 4.1]	Gennem plantetal: -0.0 [-0.5,0.4] / .0.3 [-1.9, 0.9] Gennem ukrudt: -0.3 [-1.2, 0.5] / -0.2 [-1.0, 0.5]	0.2 [-3.3, 3.9] / 0.1 [-3.8, 3.7]



Tabel 2b: Fortsættelse af Tabel 2a. I denne del af tabellen indgår blanding og udsædsmængde som ikke har en direkte effekt på udbyttet.

PARAMETER LEVEL	PARAMETER SAMMENLIGNING	DIREKTE EFFEKT (FORGRENET / UFORGRENET)	INDIREKTE EFFEKT (FORGRENET /UFORGRENET)	TOTAL EFFEKT (FORGRENET / UFORGRENET)
Jordtype=sand	Jordtype=ler	1.3 [-2.4,5.4]		1.3 [-2.4,5.4]
Husdyrgødning= dybstrøelse	Husdyrgødning= ingen	3.5 [0.4, 6.8]**		3.5 [0.4, 6.8]**
Husdyrgødning= gylle	Husdyrgødning= ingen	2.7 [-1.5,7.2]		2.7 [-1.5,7.2]
Lupintype = forgrenet, plantetal= gennemsnitlig	Lupintype = uforgrenet, plantetal= gennemsnitlig	0.7 [-2.2, 3.7]	Gennem ukrudt: 0.1 [-0.5, 1.1]	0.9 [-2.2, 4.0]
Blanding= renbestand	Blanding = blandet		Gennem plantetal: 0.0 [-0.4, 0.4] / -0.0 [-1.3,1.3] Gennem ukrudt: -0.2[-1.1, 0.6] / -0.1[-0.8, 0.5]	-0.2 [-1.1,0.6] / -0.1 [-1.5, 1.3]
Udsædsmængde stiger med 24.49 (en std. afvigelse)	Udsædsmængde konstant		Gennem plantetal: 0.1 [-0.4,0.6] / 0.7 [-0.4,2.0]*	0.1 [-0.4,0.6] / 0.7 [-0.4,2.0]*
Plantetal stiger med 24.7 planter pr. m ² (en std. afvigelse)	Plantetal konstant	0.2 [-1.1, 1.6] / 2.2 [-1.0,5.4]*		0.2 [-1.1, 1.6] / 2.2 [-1.0,5.4]*
Ukrudtstal stiger med 16 procentpoint (en std. afvigelse)	Ukrudtstal konstant	-1.8 [-3.5, -0.0]** / -1.0 [-3.2, 1.1]		-1.8 [-3.5, -0.0]** / -1.0 [-3.2, 1.1]

Vi testede også en mulig interaktion mellem RT og husdyrgødning. En forøgelse af RT med 0.38 giver en forventet ændring på -3.23 (HDI = [-6.9, 1.6]) for dybstrøelse, mens den for gylle giver en ændring på 0.01 (HDI = [-10.0,9.3]). Dog giver dette også problemer med at estimere alle parametrene i modellen, så interaktionen er udeladt fra de videre modeller.



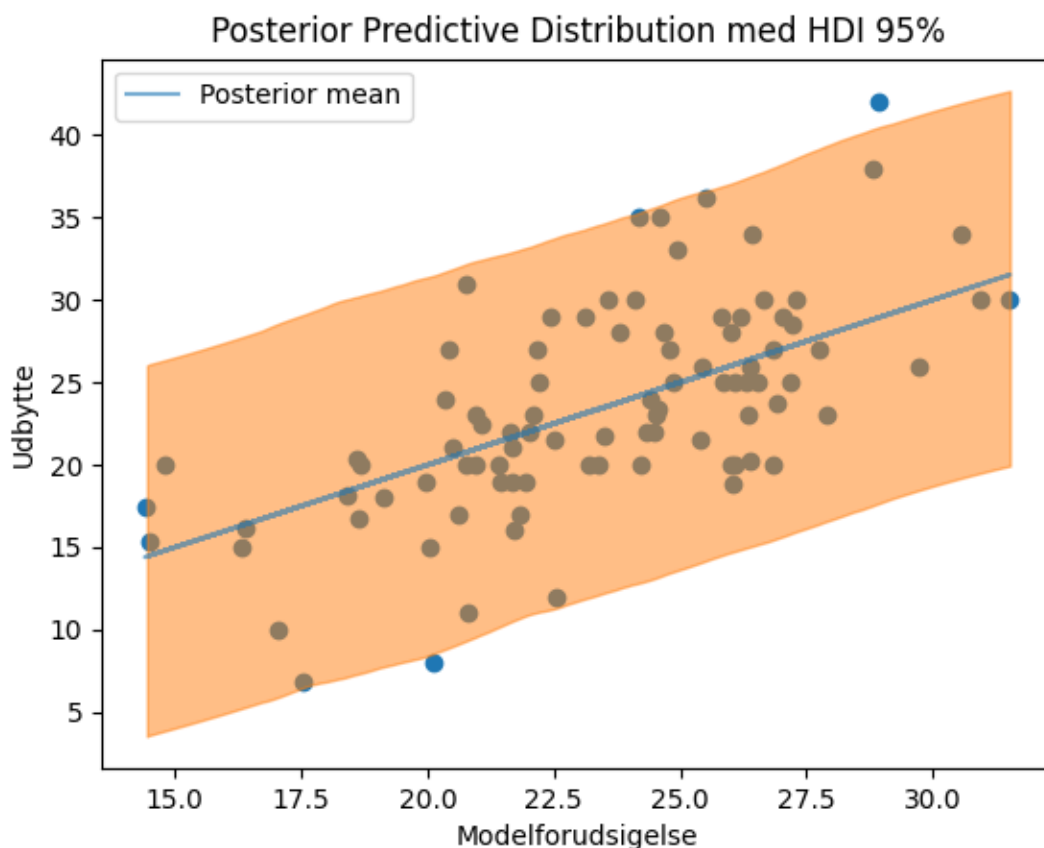
Tabel 3: Denne table viser parameterestimerne for de medierende outcomes, ukrudtstal og plantetal.

PARAMETER LEVEL	PARAMETER SAMMENLIGNING	OUTCOME VARIABEL	EFFEKT
Lupintype= forgrenet	Lupintype = uforgrenet	Ukrudt	-3.2 [-8.7, 1.9]
Blindharvning= ja	Blindharvning= nej	Ukrudt	2.4 [-3.9, 9.1]
Blanding= renbestand	Blanding = blandet	Ukrudt	1.6 [-4.8, 7.9]
Udsædsmængde stiger med en standardafvigelse	Udsædsmængde konstant	Plantetal	7.9 [1.5,14.0]**
Blanding= renbestand	Blanding = blandet	Plantetal	-0.00 [-11.2, 10.4]
Sådybde=dyb, blindharvning=nej	Sådybde=lav, blindharvning=nej	Plantetal	-8.2 [-25.2, 9.4]
Sådybde=medium, blindharvning=nej	Sådybde=lav, blindharvning=nej	Plantetal	-4.5 [-17.0, 8.4]
Sådybde=dyb, blindharvning=ja	Sådybde=lav, blindharvning=nej	Plantetal	3.5 [-15.8, 24.0]
Sådybde=medium, blindharvning=ja	Sådybde=lav, blindharvning=nej	Plantetal	-4.2 [19.0, 11.6]
Sådybde=lav, blindharvning=ja	Sådybde=lav, blindharvning=nej	Plantetal	-3.5 [-15.1, 6.4]

Sådatoen kan tænkes at have en effekt på ukrudts- og plantetallet, så vi undersøgte om disse forbindelser kunne forbedre modellen. En tidlig såning resulterer i en ændring af ukrudtstallet med +5.5 procentpoint (HDI = [-1.2, 12.3]), mens effekten på plantetallet er beskedne -0.1 planter per m² (HDI = [-10.3, 10.1]). Inddragelsen af disse parametre forværrer dog LOO og WAIC-kriterierne, så de er ikke med i den fulde model. Vi afprøvede også om vekselvirkningen mellem år og sådato på udbyttet kan omskaleres til en effekt på ukrudt og plantetal, men også dette forværrer WAIC-værdien.

Af de inkluderede forklarende faktorer er især en tidlig sådato relateret til et højt udbytte. Sammenlignet med sen såning, stiger udbyttet ved tidlig såning med 6.7 (HDI = [2.7,10.9]) i 2022 og 4.0 (HDI = [-0.0,7.9]) for i 2021. Der var ingen tidlige såninger i 2023 og der er ikke nogen tydelig forskel ved sen såning mellem de tre år. Det samme mønster ses også i den rå marginale fordeling, men med større forskel på tidlig sånings-effekten mellem 2021 og 2022 (Tabel 4).

Modellen viser også at en medium sådybde er relateret til et højere udbytte. Uden blindharvning, sammenlignet med lav sådybde, har en medium sådybde et højere udbytte med 4.0(HDI = [0.7, 7.1]) for forgrenede lupintyper og 3.6(HDI = [-0.1, 7.2]) for uforgrenede lupintyper. Forskellen på effekten mellem de to lupintyper skyldes, at sådybden også har en indirekte effekt på udbyttet via plantetallet, som medieres af lupintypen. Hvis der også blindharves mindskes effekten for forgrenede lupintyper til 2.3 (HDI = [-1.8, 6.3]) og for uforgrenede til 2.0 (HDI = [-2.5, 6.3]). For andre kombinationer af sådybde og blindharvning er der ikke nævnefærdig forskel til lav sådybde, ingen blindharvning.



Figur 6: Den bayesianske models forudsigelser sammenlignet med de faktiske værdier, der viser en moderat sammenhæng, samt pålidelige prediktioner. Den bayesianske R^2 -værdi er 0.43.

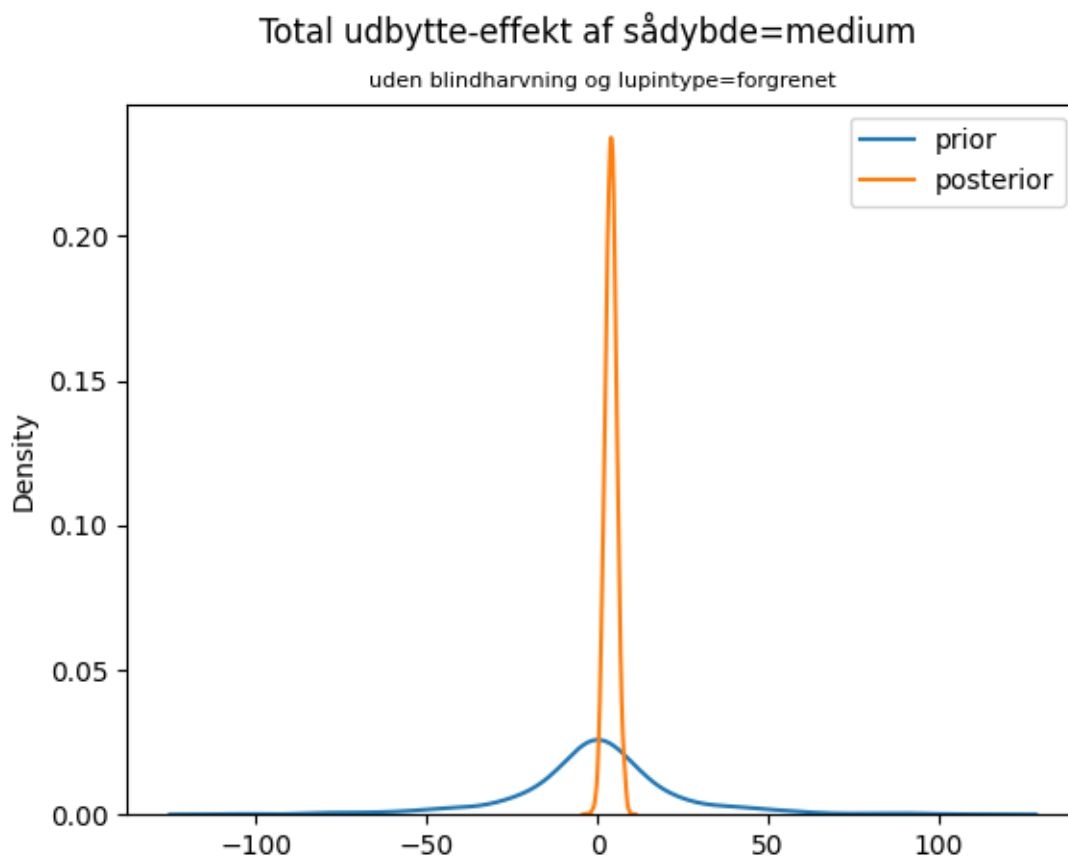
Den marginale, rå effekt af sådybde og blindharvning på udbyttet kan ses i Tabel 5, hvori der ikke findes signifikante forskelle mellem de forskellige kombinationer.

Tabel 4: De marginale rå gennemsnit for interaktionen mellem sådato og år. Compact letter displays (angivet med bogstaver) er udregnet med parvise Mann-Whitney U tests.

	2021	2022	2023
SÅNING = TIDLIG	25,8 ^{ab}	27,9 ^a	
SÅNING = SEN	23,1 ^{abc}	18,9 ^c	23,0 ^{bc}

Tabel 5: De marginale gennemsnit for interaktionen mellem sådybde og blindharvning.

	SÅDYBDE = LAV	SÅDYBDE = MEDIUM	SÅDYBDE = HØJ
BLINDHARVN. = JA	21,6 ^a	24,6 ^a	24,1 ^a
BLINDHARVN. = NEJ	20,5 ^a	26,4 ^a	22,9 ^a



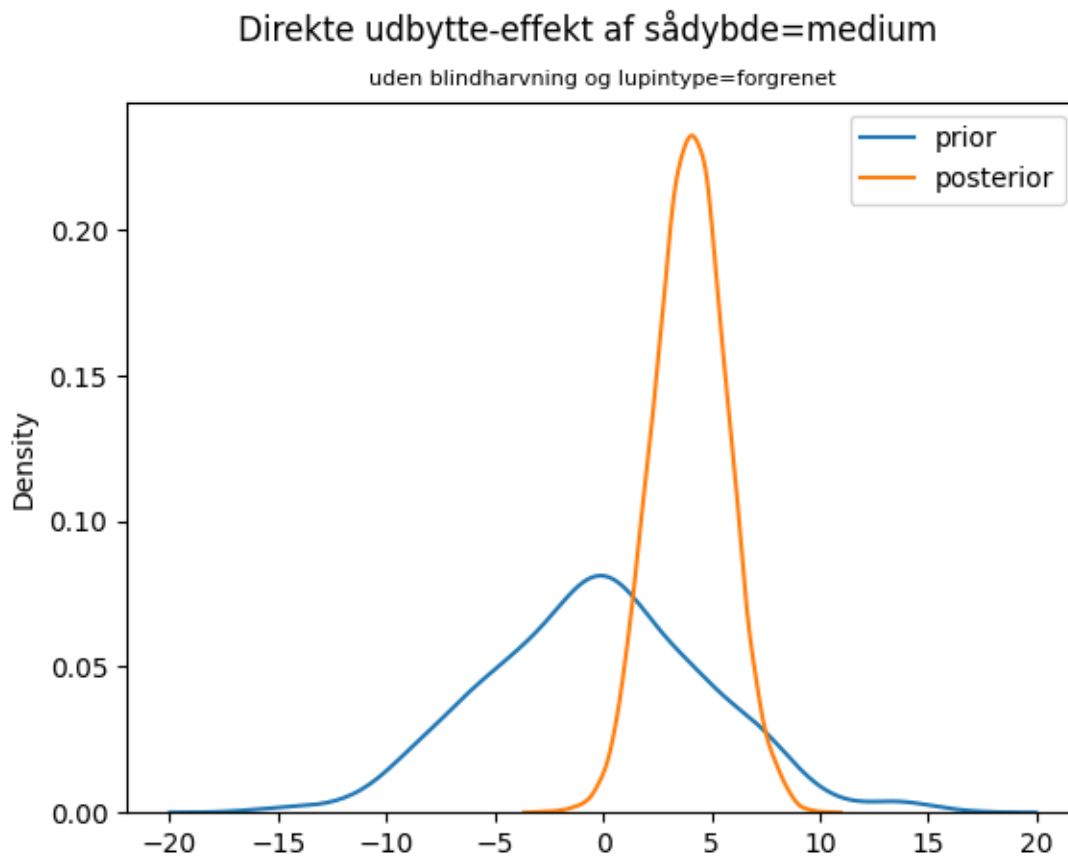
Figur 7: Her ses modellens prior fordeling af den totale udbytte-effekt før modellen havde "set" data sammenlignet med den posterior, opdaterede fordeling af den totale udbytte effekt.

Sammenlignet med ingen husdyrgødning øger anvendelse af dybstrøelse udbyttet med 3.5 (HDI = [0.4, 6.8]), hvorimod gylle øger udbyttet med 2.7 (HDI = [-1.5,7.2]). Denne effekt er lidt større og mere signifikant end effekterne i den rå, marginale tabel (Tabel 6).

Jordtypen har ikke en tydelig effekt på udbyttet i modellen. En sandet jord øger her udbyttet med 1.3 (HDI = [-2.4,5.4]) sammenlignet med en leret jord. I den marginale er den rå effekt dog signifikant med en udbytte-fordel på 5.96 for de sandede jorde.

Ukrudts effekt på udbyttet er stratificeret på forgrenede og uforgrenede jordtyper. Når ukrudtstallet stiger med 16 procentpoint (svarende til en standardafvigelse i populationen) ændres udbyttet med -1.8 (HDI = [-3.5, -0.0]) for forgrenede lupintyper og -1.0 (HDI = [-3.2, 1.1]) for uforgrenede. En lignende effekt ses i den rå, marginale tabel, hvor der er signifikant negativ korrelation mellem ukrudt og udbytte (Tabel 7).

Plantetallets effekt er også separat for de to lupintyper. Når plantetallet stiger med 24.7 planter pr. m², stiger udbyttet med 2.2 (HDI = [-1.0,5.4]) for uforgrenede lupiner, mens de forgrenede stiger med blot 0.2 (HDI = [-1.1, 1.6]). Den lille effekt understøttes af den lave korrelation i den rå, marginale tabel (Tabel 7)



Figur 8: Ligesom Figur 7, vises den prior og posterior fordeling af sådybde-effekten – her blot for den direkte effekt (svarende til konstant plante- og ukrudtstal).

Tabel 6: Gennemsnit inden for hver faktor. Compact letter displays (CLD) er udregnet med Mann-Whitney's U-test uden korrektion for multiple testing.

FAKTOR	NIVEAU	GENNEMSNI TLIGT UDBYTTE	ANTAL	CLD
HUSDYRGØDNING	Dybstrøelse	24,71	8	a
	Gylle	25,06	7	a
	Ingen	22,79	70	a
	Missing	26,86	7	a
JORDTYPE	1	24,35	45	b
	2	25,75	4	ab
	3	22,82	16	ab
	4	24,18	17	b



	6	19,25	9	a
	7	8	1	ab
JORDTYPE, GRUPPERET				
	Leret (JB>5)	18,13	10	a
	Sandet (JB<5)	24,09	82	b
LUPINTYPE				
	Forgrenet	23,63	65	a
	Uforgrenet	23,19	21	a
	<i>Missing</i>	22,17	6	a
SÅDYBDE				
	High/Deep	23,53	19	ab
	Medium	25,84	31	b
	Low	21,3	35	a
	<i>Missing</i>	23,29	7	ab
VANDING				
	Ja	25,42	19	a
	Nej	22,71	62	a
	<i>Missing</i>	24,14	11	a
SÅDATO				
	Tidlig	26,88	27	b
	Sen	22,07	62	a
	<i>Missing</i>	20,67	3	ab
BLINDHARVNING				
	Ja	22,98	38	a
	Nej	23,84	45	a
	<i>Missing</i>	23,33	9	a
ÅR				
	2021	24,41	35	a
	2022	23,18	29	a
	2023	22,49	28	a
FORFRUGT				
	Andet	24,39	30	a
	Vår	22,64	40	a
	Vinter	23,51	19	a
	<i>Missing</i>	24	3	a
FORFORFRUGT				
	Andet	23,23	65	a
	Bælg	23,89	20	a
	<i>Missing</i>	24,14	7	a
ANDEN GØDNING				
	Ja	25,95	9	a
	Nej	22,81	74	a
	<i>Missing</i>	26,11	9	a



PODNING				
	Ja	23,4	82	a
	Nej	21,33	3	a
	Missing	24,79	7	a
BLANDING				
	Blandsæd	24,99	33	a
	Renbestand	22,66	58	a
	Missing	17	1	a

Tabel 7: De kontinuerte forklarende faktorer sammenlignes her med udbyttet. Kategorierne er diskretiserede for at kunne opdage de mest non-lineariteter. I yderste kolonne er den numeriske pearson's korrelation angivet med p-værdien for at korrelation er 0.

FAKTOR	NIVEAU	GENNEMSITLIGT UDBYTTE	ANTAL	CLD	KORRELATION (P-VÆRDI)
REAKTIONSTAL (RT)					
	[5.2,5.8]	25,22	25	b	
	(5.8,6.1]	23,68	12	ab	
	(6.1,7.1]	21,9	12	ab	
	Missing	22,77	43	a	
			49		-0.40 (0.004)
UDSÆDSMÆNGDE					
	[120,180]	24,07	33	a	
	(180,200]	22,74	37	a	
	(200,270]	25,32	11	a	
	Missing	22	11	a	
			81		-0.04 (0.74)
BLANDET MÆNGDE					
	[25,40]	23,71	17	ab	
	(40,48.3]	28,4	4	ab	
	(48.3,75]	26,29	11	b	
	Missing	22,51	60	a	
			32		0.16 (0.38)
JORDTEMPERATUR					
	[3,5]	23,81	19	a	
	(5,8]	24,23	20	a	
	(8,10]	19,74	10	a	
	Missing	23,76	43	a	
			49		-0.15 (0.31)



PLANTETAL					
	[33,76]	22,37	31	a	
	(76,95.1]	23,5	27	a	
	(95.1,175]	25,29	29	a	
	<i>Missing</i>	19	5	a	
			87		0.11 (0.32)
UKRUDT					
	[1,7.1]	23,08	30	a	
	(7.1,19.3]	25,94	29	b	
	(19.3,70]	21,51	30	a	
	<i>Missing</i>	22,17	3	ab	
			89		-0.22 (0.04)
GRÅSKIMMEL, ANTAL ANGREBEDE PLANTER					
	[0,2]	23,05	51	ab	
	(2,100]	25,27	26	b	
	<i>Missing</i>	21,6	15	a	
			77		0.21 (0.07)
GRÅSKIMMEL, PROCENT DÆKNING AF PLANTER					
	[0,1]	23,18	52	ab	
	(1,45]	25,08	25	b	
	<i>Missing</i>	21,6	15	a	
			77		0.13 (0.25)

Diskussion

Vi har fittet en bayesiansk, hierarkisk, lineær model med en R^2 på 0.43, hvilket betyder at det stadig er svært at forudsige udbyttet med de inkluderede forklarende faktorer. Usikkerheden ved en enkelt udbytte-observation er for eksempel +/-10. Vi kan dog stadig identificere overordnede trends.

Den bayesianske model kan udregne relationen mellem hver forklarende faktor og udbyttet, hvor der er taget højde for alle de andre forklarende faktorer. Derfor ser vi forskel mellem den bayesianske model og de marginale modeller.

En medium sådybde, lav ukrudtsmængde, og en tidlig såning er tydeligt forbundet med et højere udbytte i både de marginale tabeller og i den bayesianske model. Det er dermed dette studies bedste bud på hvilke sammenhænge, man bedst kan regne med. Da studiet er observationelt er det dog meget muligt, at der findes andre uobserverede confounding variable, der er de egentlige kausale grunde til det højere udbytte.

Jordtype er den eneste forklarende faktor, der er marginalt signifikant, men ikke i den bayesianske model. Det indikerer at en stor del af udbytteforskellen mellem de lerede og sandede jorde kunne skyldes de andre forklarende faktorer.

Dybstrøelse, derimod, er den eneste forklarende faktor, der er signifikant for udbyttet i den bayesianske model, men ikke marginalt. Det skyldes at anvendelse af dybstrøelse-faktoren kan forbedre forudsigelserne produceret



af de andre forklarende faktorer. Det er altså en indikation på at der faktisk er en kausal, positiv effekt af dybstrøelse. Denne sammenhæng er dog ikke lige så sikker som sådybde, sådato og ukrudts-effekterne, da den nemmere kan forsvinde, hvis modellens antagelser ændres, eller der bliver inkluderet andre faktorer i modellen.

Derudover er der små indikationer, som ikke er signifikante, hverken marginalt eller i modellen. Det gælder

- Anvendelse af gylle øger måske udbyttet
- Et højere plantetal for uforgrenede lupintyper via en større udsædsmængde øger måske udbyttet

Nogle faktorer blev udeladt i modellen. De relativt få blandsædsmarker indikerer for eksempel at der kunne være en gevinst ved en stor mængde af den anden afgrøde (Tabel 7). Reaktionstallet har en tydelig effekt marginalt, som ikke blev genfundet at have samme størrelse i den fulde bayesianske model. Dette skyldes formentligt, at modellen ikke var stabil med RT inkluderet.

Appendix

Den bayesianske model er af formen

$$udbytte_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = \sum_{l \in \text{lupintyper}} (\beta_{\text{plantetal},l} \text{plantetal}_i + \beta_{\text{ukrudt},l} \text{ukrudt}_i) + \sum \beta_j x_{j,i}$$

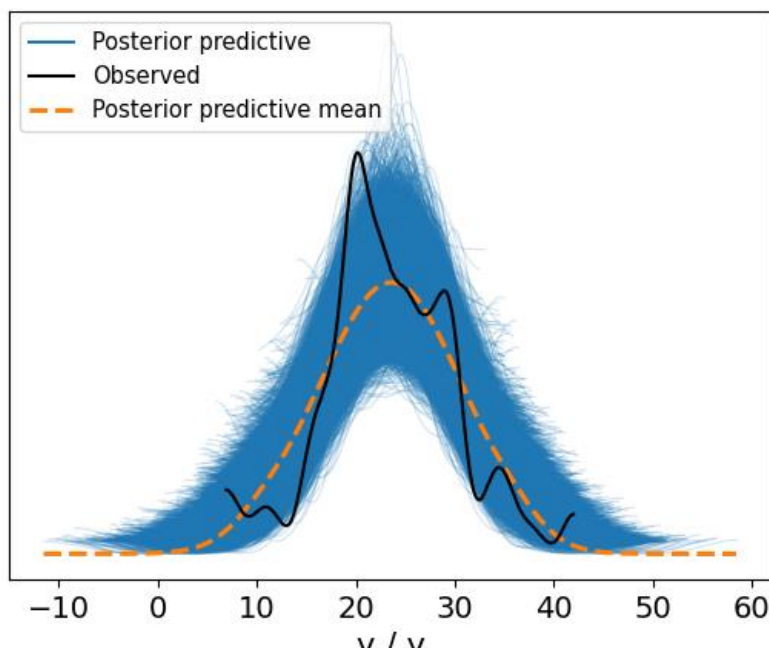
$$\text{plantetal}_i \sim N\left(\sum \gamma_j x_{j,i}, \sigma_\gamma^2\right)$$

$$\text{ukrudt}_i \sim N\left(\sum \delta_j x_{j,i}, \sigma_\delta^2\right)$$

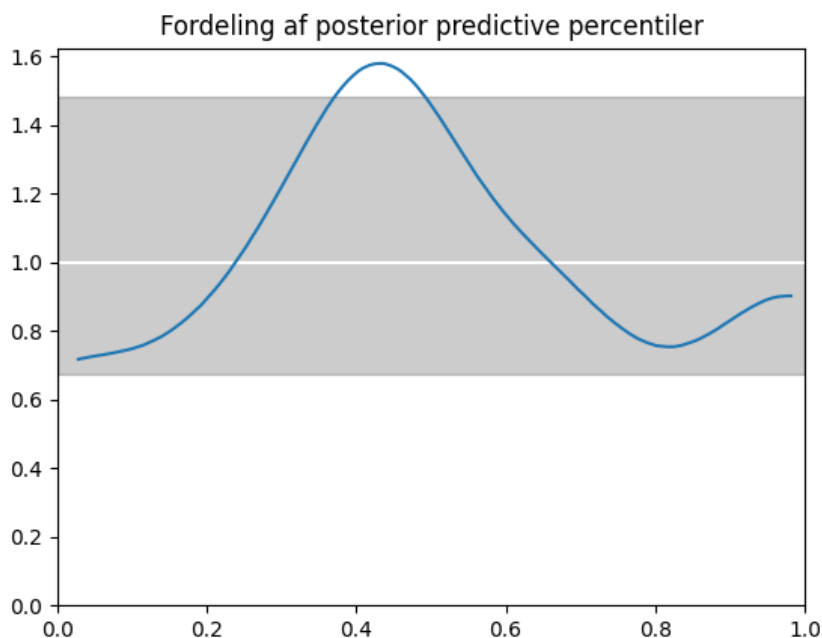
$$\beta_j \sim N(0, 5^2), \delta_j \sim N(0, 1), \gamma_j \sim N(0, 1), \sigma \sim \text{HalfCauchy}(10), \sigma_\gamma \sim \text{HalfCauchy}(1), \sigma_\delta \sim \text{HalfCauchy}(1)$$

Hvor $x_{j,i}$ er den normaliserede faktor for den i 'te observation og plantetal_i og ukrudt_i er det normaliserede plantetal og ukrudtsprocent for den i 'te observation, hhv. Plantetal og ukrudt bliver altså behandlet som mediating faktorer for udbyttet.

Modelkontrollen udføres til dels med Figur 6, der viste at der ikke var outliers i modellen og at modellen beskriver afvigelserne godt. På Figur A1 og A2 vises yderligere check. Der er for lidt mange observationer omkring 20 i forhold til den fittede model. Det ses både på Figur A1, der viser at antallet er ekstremt i forhold til hvad modellen ville forudsige. Dette kan ses tydeligere på den mere abstrakte A2, der viser u -værdierne.



Figur A1: Med de observerede forklarende faktorer ville den bayesianske models forudsigelser fordele sig inden for det blå bånd. De faktiske observationer ligger dog lige uden for dette område omkring 12 og 20.



Figur A2: Forskellen mellem modellens forudsigelser (normaliseret til det grå bånd og de observerede udbytter (blå) viser at modellen ikke kan forudsige det store antal observationer omkring 40%-fraktilen.