

Notat

SEGES Digital

Projekt: 5517, Datadrevet management i mælkeproduktionen Arbejdsmappe 1	Ansvarlig	KAND
	Oprettet	02-12-2022
	Side	1 af 9

Dynamisk Lineær Model til prædiktion af sygdomme og lidelser hos malkekøer på baggrund af ændringer i kropsvægt og mælkeydelse.

Indhold

Baggrund	1
Datagrundlag	2
Grundlæggende registreringer	2
Inddeling efter race og laktationsnummer	2
Inddeling efter behandling for sygdom eller lidelse	2
Afgrensning på baggrund af dage med registreringer	3
Dynamisk Lineær Model	3
Træning af modellen	4
Test af modellen	4
Alarm for risikokøer	4
Performance	4
Perspektivering – prototype.....	5
Appendix.....	7

Baggrund

I danske malkekvægsbesætninger med malkerobotter, eller automatiske malkesystemer (AMS), går køerne sammen i større hold. Landmanden tilsår køerne dagligt, og observerer, om alle dyr er sunde og raske. Det kan dog være vanskeligt for landmanden at se de tidligste tegn på sygdomme og lidelser hos den enkelte ko under de daglige rutiner.

Hvis kropsvægten hos en ko ændres markant, eller hvis koens daglige mælkeydelse afviger systematisk fra det forventede, kan det betyde, at koen er syg, eller har en lidelse, der endnu ikke kan konstateres uden en grundig undersøgelse. Ved at overvåge data om køers vægt og ydelse automatisk, er det muligt at opdage uventede ændringer i vægt og ydelse, og give landmanden besked om, at en ko er i risiko for at fejle noget, og derfor har brug for et grundigere tilsyn. Det giver landmanden mulighed for at iværksætte forebyggende tiltag, eller ændre management i besætningen, så velfærd og produktivitet optimeres.

Dette notat er skrevet med formål at beskrive databehandling og resultater i forhold til anvendelsen af en dynamisk lineær model (DLM) til prædiktion af sygdomme og lidelser hos malkekøer på baggrund af ændringer i kropsvægt og mælkeydelse. Notatet beskriver ikke modellen i

detaljer, da dette er beskrevet i publikationen 'Early detection of abnormal weight patterns in dairy cows based on AMS weighing data'.

Opgaven er en del af Projekt 5517, MAF – Datadrevet Management i Mælkeproduktionen. AP1 Vejedata.

Datagrundlag

Grundlæggende registreringer

Datagrundlaget dannes af data fra 125 danske malkekvægsbesætninger med automatisk malkesystem (AMS) samt registreringer om samme besætninger fra Kvægdatabasen. Ved hver malkning vejes den aktuelle ko, og vægten registreres. Samtidig måler malkerobotten koens ydelse ved den givne malkning. Fra Kvægdatabasen indhentes for hver ko registreringer omkring reproduktion, behandlinger, race, paritet, dage efter kælvning, status (gold eller lakterende), dyre-ID samt besætnings-ID.

Ud fra vægtregistreringerne beregnes et dagligt gennemsnit af koens vægt, så der er én samlet vægtregistrering pr. døgn. Hvis koen har mulighed for at blive malket og vejlet i mere end én malkerobot i den indeværende laktation, korrigeres vægten for eventuelle forskelle mellem robotterne som beskrevet i publikationen 'Oprensning af vægtdata fra automatiske malkesystemer'. Vægten af foster, moderkage med mere, beregnes i forhold til den aktuelle drægtighedsdag og koens race. Denne vægt trækkes fra den daglige vægtregistrering fra malkerobotten, så det er koens reelle vægt, der anvendes.

Ydelsesmålingerne summeres over døgnet fra midnat til midnat, så det er én samlet ydelse pr. døgn, der anvendes i modellen.

Inddeling efter race og laktationsnummer

I alt indgår 19845 individuelle køer fra 125 besætninger i datasættet. Køerne inddeles i tre grupper efter laktationsnummer (1. laktation, 2. laktation og 3. laktation samt ældre) og yderligere i to grupper efter race (Jersey og Ikke-jersey). Den samme ko indgår altså i datasættet flere gange, hvis den har flere laktationer i besætningen. Der indgår i alt 75028 individuelle ko-laktationsnummer-race kombinationer i datasættet.

Tabel 1 viser fordelingen af de 75028 ko-laktationsnummer-race-kombinationer.

Tabel 1

1. Jersey	2. Jersey	3+ Jersey	1. Ikke-jersey	2. Ikke-jersey	3+ Ikke-jersey
1584	1341	2293	21423	18727	28660

Inddeling efter behandling for sygdom eller lidelse

Formålet med denne arbejdsmappe (AP1 – Vejedata) er at udvikle et automatiseret overvågningsværktøj, der kan give en alarm for sygdom eller lidelse hos en ko så landmanden kan undersøge koen grundigt på et tidligere tidspunkt end, hvornår koen ville være blevet behandlet, hvis ikke sådan et værktøj var til rådighed.

Køerne i hver laktationsgruppe inddeles derfor i tre grupper på baggrund af registrerede diagnoser og behandlinger i kvægdatabase. De tre grupper er:

- Raske dyr: Køer, der ikke har nogle registrerede behandlinger i indeværende laktation
- Syge dyr: Køer, hvor der er registreret en behandling i indeværende laktation
- Kronisk syge dyr: Køer, der er behandlet for en kronisk lidelse i en tidligere laktation

De raske dyr udgør det data, overvågningsmodellen trænes på, kaldet træningsdata. Formålet er, at modellen lærer, hvordan vægt og ydelse udvikler sig over en laktation hos en ko, der ikke er ramt af en behandlingskrævende sygdom eller lidelse. Dette giver modellen viden om en normal/forventet vægt- og ydelseskurve for en rask ko, og det danner grundlag for senere at kunne skelne afvigende vægt- og ydelseskurver.

De syge dyr udgør det data, overvågningsmodellen testes på, kaldet testdata. Det er udelukkende data fra den laktation, hvor behandlingen finder sted, der indgår i testdata. Data fra efterfølgende laktationer uden behandlingsregistreringer indgår i træningsdata.

Formålet er dels at undersøge, hvor god modellen er til at genkende afvigende vægt- og ydelseskurver i tidsrummet omkring en behandling af en ko, og dels at undersøge, om det kun er i forbindelse med sygdom, eller lidelse, der sker systematiske ændringer i vægt- eller ydelseskurven hos en ko.

De kronisk syge dyr indgår i testdata fra og med den første laktation med registrering af en behandling for en kronisk lidelse ud fra antagelsen om, at en kronisk lidelse påvirker vægt- og/eller ydelseskurven på alle tidspunkter efter lidelsen er konstateret. Laktationer forud for en laktation med behandling for en kronisk lidelse indgår enten i træningsdata (hvis der ikke er nogen behandling i laktationen), eller i testdata (hvis der er en behandling for en ikke-kronisk lidelse).

Tabel 2 viser fordelingen af raske, syge og kroniske dyr i forhold til race og laktationsnummer.

Tabel 2

	1. Jersey	2. Jersey	3+ Jersey	1. Ikke-jersey	2. Ikke-jersey	3+ Ikke-jersey	I alt
Rask	843	686	697	9558	7502	8432	27718
Syg	617	462	996	10313	9037	14559	35984
Kronisk	124	193	600	1552	2188	5669	10326

Afgrænsning på baggrund af dage med registreringer

I træningsdata, hvor modellen skal lære normale vægt- og ydelseskurver, anvendes kun data til og med dag 305 efter kælvning.

I testdata, hvor modellen skal kunne identificere afvigende kurver, skal en ko have minimum 250 dage i en laktation med data fra malkerobotten for at indgå.

Dynamisk Lineær Model

Det overordnede mål for den dynamisk lineær model (DLM) i dette projekt er at forudsige næste dags vægt og ydelse for den enkelte ko på baggrund af de vægt- og ydelsesdata, der er registreret i den aktuelle laktation til og med indeværende dag.

Når næste dags data registreres, sammenlignes forudsigelsen med den nye registrering, og en eventuel forskel mellem forudsigelse og registrering gemmes som *forecast errors*.

Hvis modellen har estimeret for højt i forhold til næste dags observation, bliver *forecast error*'en negativ, og hvis modellen har estimeret for lavt, bliver *forecast error*'en positiv. Disse *forecast errors* bruges til at generere alarmer om risikokøer, som beskrevet nedenfor.

Træning af modellen

Som beskrevet ovenfor, trænes modellen på vægt- og ydelseskurver fra laktationer, hvor koen ikke er behandlet for sygdom eller lidelse. Træningsdata inddeles efter race (Jersey/Ikke-jersey) og laktationsnummer (1., 2. og 3+ laktation), så modellen lærer vægt- og ydelseskurver for hver race-laktationsnummerkombination. Alle raske køer på tværs af alle besætninger indgår i træningsdata. Modellen lærer således en 'global' normalkurve for vægt og for ydelse indenfor race og laktationsnummer, men på tværs af alle 125 besætninger.

Test af modellen

Testdata inddeles, ligesom træningsdata, i grupper baseret på race og laktationsnummer. Herudover inddeles testdata efter besætning. Der kan være forskellige grænser for, hvornår der skal gives alarmer i forskellige besætninger, og derfor skal tærskelværdien for, hvornår en alarm skal gives, tilpasses den enkelte besætning.

Alarm for risikokøer

De *forecast errors*, der kommer af forskellen på modellens forudsigelse af næste observation og den faktiske observation, summeres kumulativt for hver dag i laktationen. Herved fås en kumulativ sum, eller en CUSUM, der kan afbildes som en kurve (se Figur 1, 2 og 3). Det er forventeligt, at der nogle dage er positive *forecast errors* og nogle dage er negative. De udligner hinanden, og CUSUM vil derfor fluktuere omkring værdien nul.

I det tilfælde, at en ko uventet taber sig, eller tager på, er ændringen systematisk, og det vil vise sig som gentagne positive eller negative *forecast errors*, der ikke udligner hinanden. Når dette sker, vil CUSUM-kurven stige eller falde – og i begge tilfælde bevæge sig længere og længere væk fra nul. Og det skal udløse en alarm til landmanden, hvis afvigelsen overskrider en tærskelværdi.

En tærskel for, hvornår en alarm skal udløses, defineres for hver race-laktationsnummer-besætning kombination. Når CUSUM-kurven rammer tærsklen, får landmanden således besked om, at en ko er i risiko for at fejle noget, og kan iværksætte en mere grundig undersøgelse af koen.

Figurene 1, 2 og 3 (i appendix) viser eksempler på CUSUM-kurver for tre køer fra tre forskellige besætninger.

Yderligere uddybning af modellen og alarmer kan læses (på engelsk) i publikationen '*Early detection of abnormal weight patterns in dairy cows based on AMS weighing data*'.

Performance

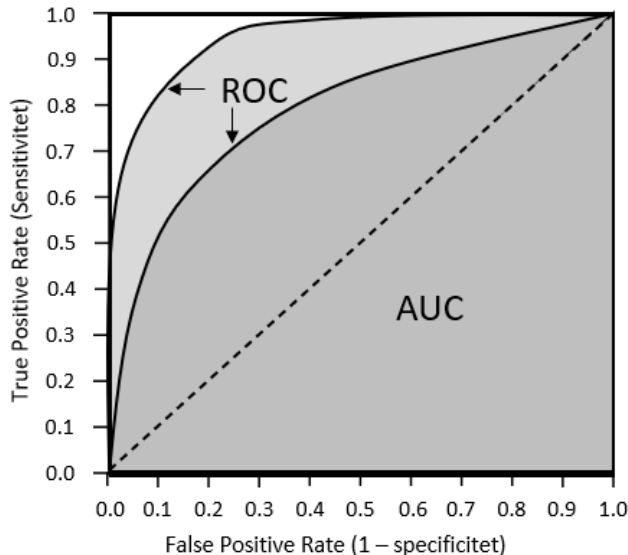
Performance er et udtryk for, hvor god modellen er til at identificere både raske dyr korrekt (sensitivitet) og syge dyr korrekt (specificitet). Antallet af alarmer afhænger af, hvor høj eller lav tærskelværdien i CUSUM sættes. Er den sat højt, er der færre alarmer. Er den sat lavt, er der mange alarmer. En alarm betragtes som sand, hvis den opstår fra 5 dage før en behandling til 2 dage efter en behandling.

Performance kan også udtrykkes så den er uafhængig af placeringen af alarmtærsklen som *Area Under the ROC Curve (AUC)*. Det gøres ved at registrere sensitiviteten og specificiteten

for hver værdi i et interval for alarmtærsklen fra lav til høj, og herefter plote en kurve med *False Positive Rate* ($1 - \text{specificitet}$) ud af x-aksen og *True Positive Rate* (sensitivitet) op ad y-aksen. Arealet under kurven udtrykker modellens performance, og det ønskes så tæt på 1 som muligt.

I Figur 4 ses et eksempel på to ROC kurver og AUC.

Figur 4



Performance beregnes indenfor besætning og indenfor laktationsgruppe. For to tilfældigt udvalgte besætninger med Ikke-jerseykøer er performance for DLM med variablene mælkeydelse og kropsvægt som følger:

Besætning xxx44

1. laktation: AUC = 0.57
2. laktation: AUC = 0.46
3. laktation: AUC = 0.59

Besætning xxx74

1. laktation: AUC = 0.66
2. laktation : AUC = 0.44
3. laktation : AUC = 0.67

Performance er ikke imponerende i de to besætninger, og den ligger markant lavere for køer i 2. laktation end for de øvrige i begge besætninger. Det kan dog vise sig at være anderledes i andre besætninger.

Perspektivering – prototype

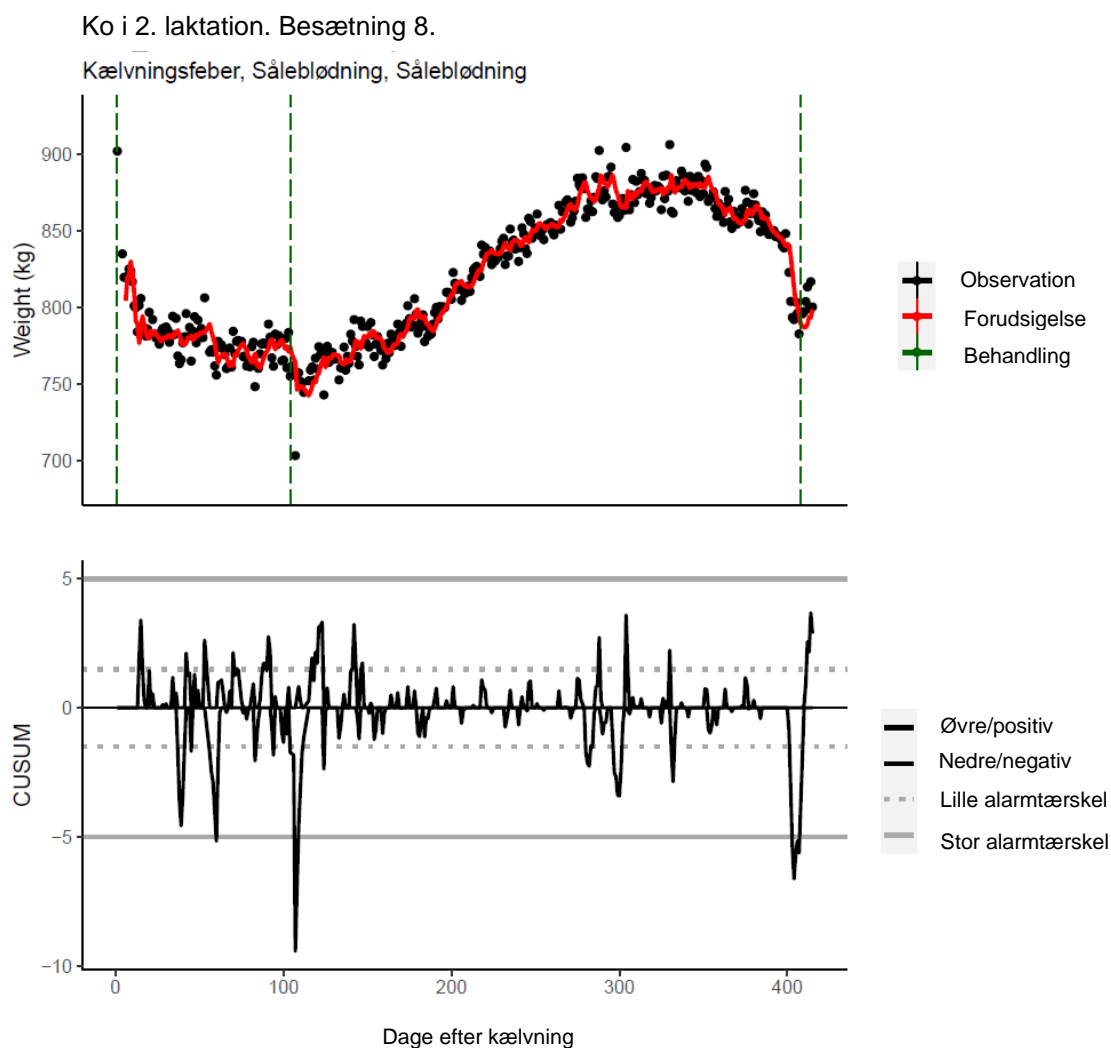
Heldigvis er der gode muligheder for, at modellen vil fungere tilfredsstillende ved implementering i en besætning trods den middelmådige performance, der afrapporteres her.

Dels skal nogle lidelser ikke behandles, men kan give en alarm alligevel, og dels kan nogle lidelser have en subklinisk periode (periode, hvor koen er syg, men ikke viser symptomer), som kan påvirke vægt og ydelse, men som ikke er til at erkende ved daglige tilsyn af dyrene.

Kun ved at afprøve modellen som beslutningsstøtteværktøj i kommercielle besætninger, kan modellens egentlige performance afklares. Herved kan den optimale tærskelværdi for alarmer i den enkelte besætning fastlægges, ligesom feedback fra landmanden er essentiel i arbejdet med at skelne mellem alarmer for lidelser, der ikke skal behandles og lidelser, der kræver behandling.

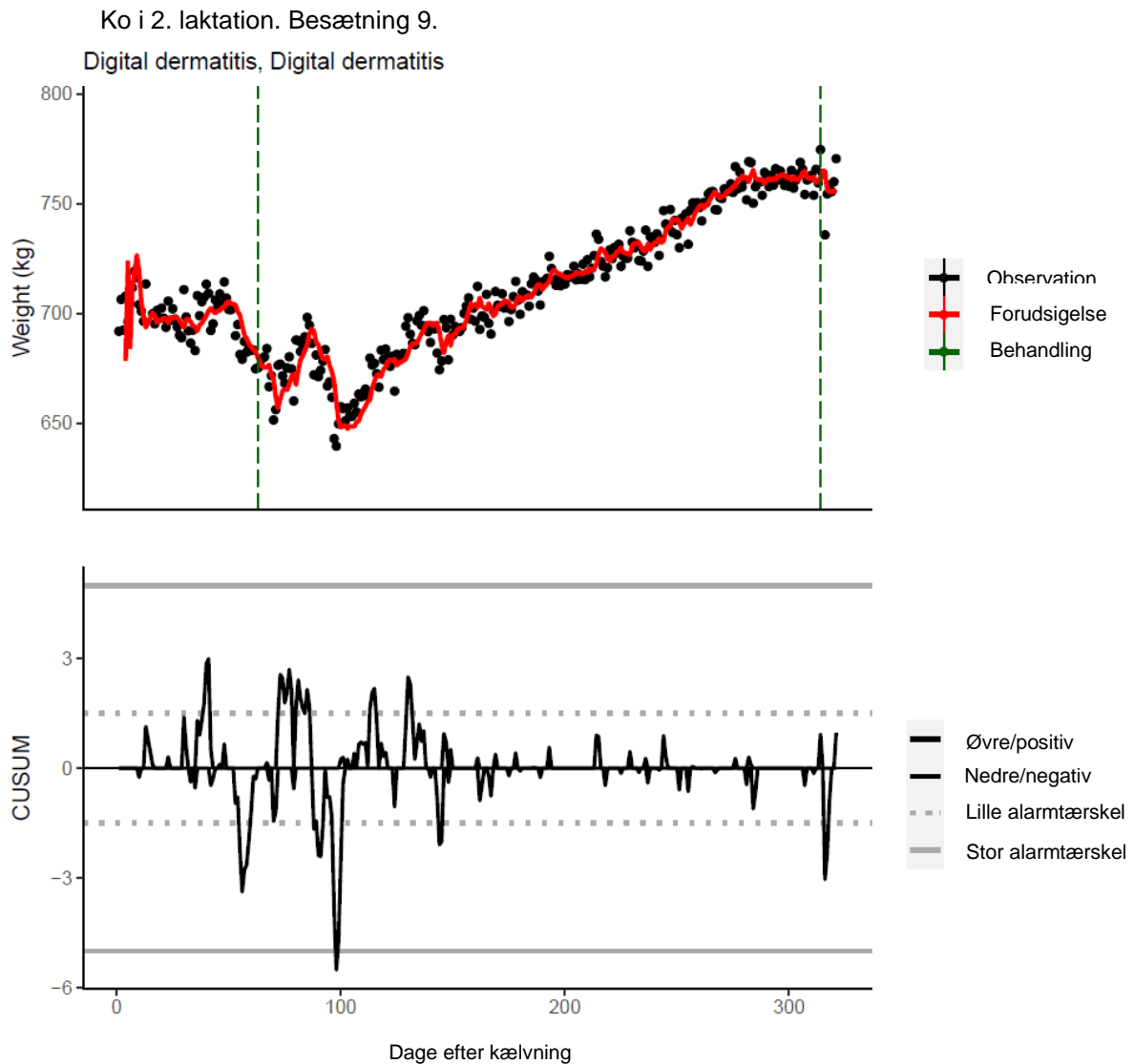
Appendix

Figur 1



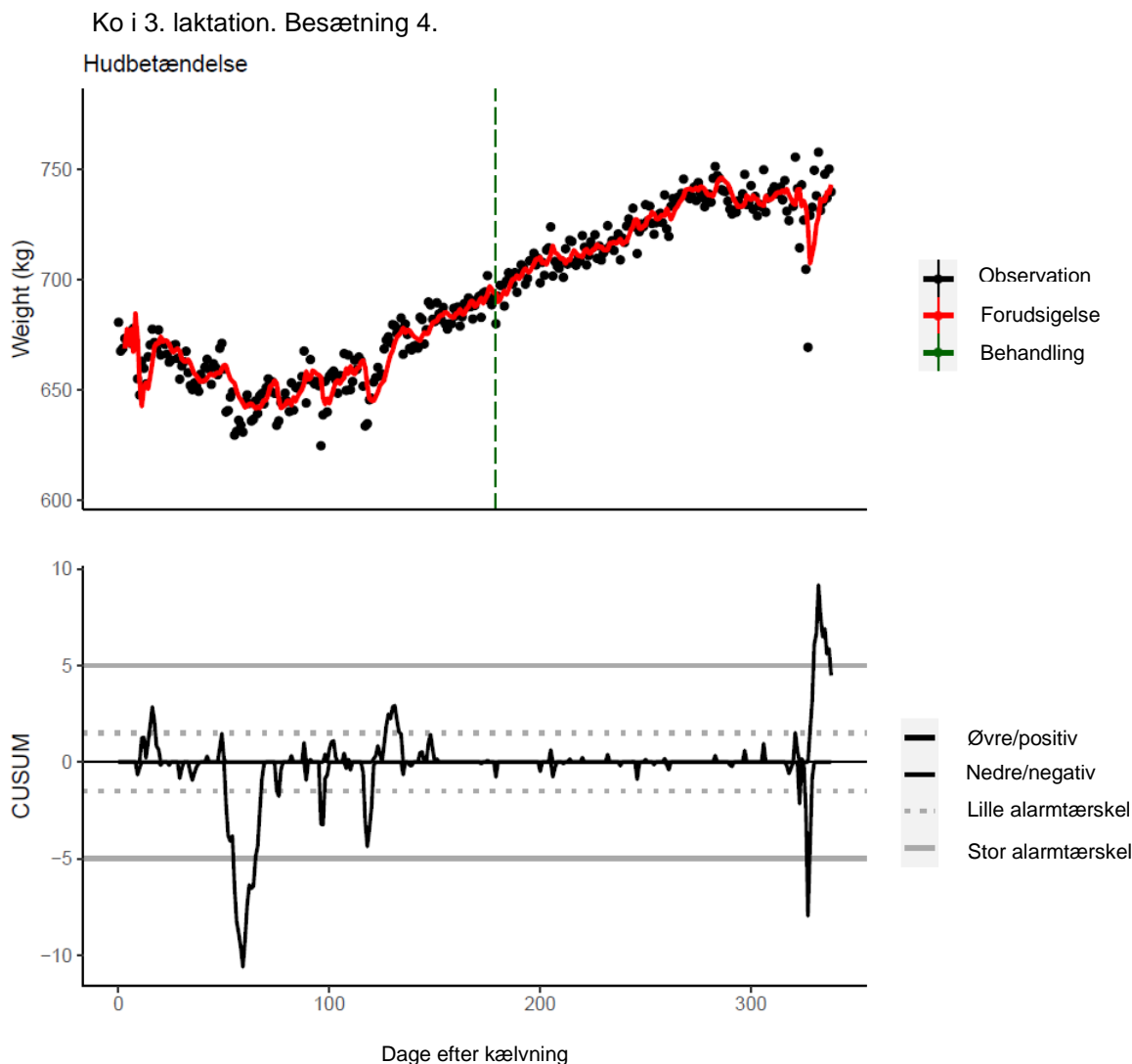
Figur 1: Et eksempel på sammenhængen mellem CUSUM alarmer og uventede ændringer i vægtkurven for en 2. laktations-ko i Besætning 8. Den øverste figur viser den observerede vægt pr. dag efter kælvning (sorte prikker) samt den vægtkurve, der er forudsagt af modellen (rød linje). Den nederste figur viser CUSUM-kurven, hvor kurven afviger med negative værdier, der rammer tærskelværdien for alarm (grå linje) omkring tidspunktet for de to sidste af de tre behandlinger.

Figur 2



Figur 2: Et eksempel på sammenhængen mellem CUSUM alarmer og uventede ændringer i vægtkurven for en 2. laktations-ko i Besætning 8. Den øverste figur viser den observerede vægt pr. dag efter kælvning (sorte prikker) samt den vægtkurve, der er forudsagt af modellen (rød linje). Den nederste figur viser CUSUM-kurven, hvor koen har en ustabil vægt fra kælvning til omkring 150 dage efter kælvning. Dette opfanges af CUSUM-kurven, der overskrider den lille og den store alarmtærskel gentagne gange i den periode.

Figur 3



Figur 3: Et eksempel på sammenhængen mellem CUSUM alarmer og uventede ændringer i vægtkurven for en 2. laktations-ko i Besætning 8. Den øverste figur viser den observerede vægt pr. dag efter kælvning (sorte prikker) samt den vægtkurve, der er forudsagt af modellen (rød linje). Den nederste figur viser CUSUM-kurven, hvor koen har en ustabil vægtkurve fra kælvning til omkring 150 dage efter kælvning, og igen kort før goldning. Dette opfanges af CUSUM-kurven, der overskrider alarmtærsklerne gentagne gange i denne periode. Den lidelse, koen er behandlet for (hudbetændelse) påvirker ikke vægten, men der kan være andre årsager til udsvingene. Det kan ikke afgøres på baggrund af de tilgængelige data.