

# Notat

Alarmer på tankmælk	Ansvarlig	Amk/Sokb
	Oprettet	21-12-2022
Projekt: 5517 Datadrevet management	Side	1 af 9

## Udvikling af alarmer på tankmælk

Anne Mette Kjeldsen, Søren Boldsen og Henrik Martinussen

I dette dokument er beskrevet, hvordan man kan overvåge mælkeproduktionen på det enkelte kvægbrug ud fra leverancerne til mejeriet. Formålet er at give kvægbrugeren alarmer, hvis udviklingen i produktionen afviger fra forventet ud fra den tidligere produktion, og også være med til ud fra data at give Kvægbrugeren de rette handlingsanvisninger, hvis produktionen afviger fra forventet.

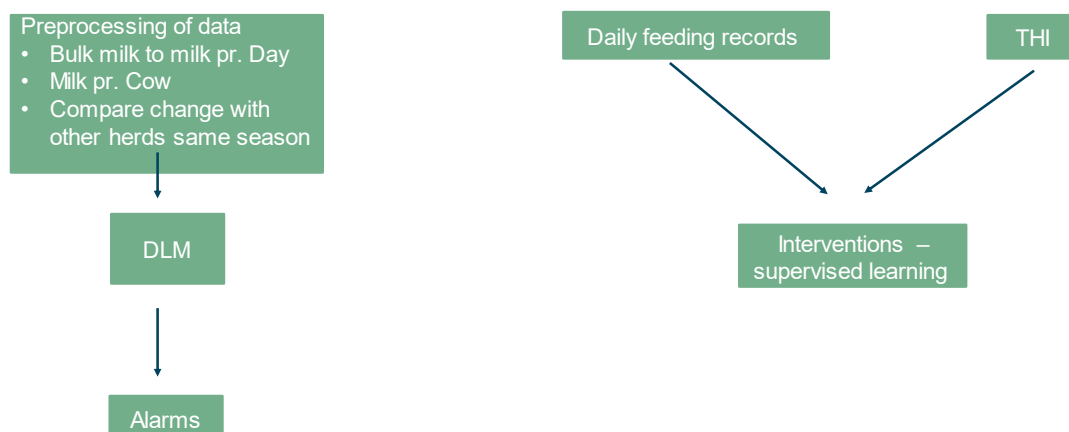
Systemet er lavet, så det både overvåger:

- Den totale mælkeproduktion på gården i kg mælk og kg EKM, samt fedt- og proteinprocent, ureaindhold og celletal.
- Kg mælk og kg EKM pr. ko, primært anvendeligt for besætninger, der indberetter golddato'er til Kvægdatabasen.
- Kg mælk og kg EKM pr. ko i forhold til målydelsen beregnet ud fra kælvningsfordelingen.
- Udviklingen på den enkelte gård i forhold til udviklingen af en tilsvarende gruppe af besætninger på landsplan. Med tilsvarende besætninger menes besætninger med samme racegruppe (stor race/Jersey) og driftsform (økologisk/konventionel).

Udover at give alarmer og der også udviklet et system der kan hjælpe med til at forklarer, om alarmerne kan skyldes varmestress eller ændringen i fodringen.

Figur 1. viser en oversigt over systemet. Da der sker rigtigt mange ting på en Kvægbedrift, som der ikke er data på, er det valgt at lægge alarmer på data ved først at forbehandle data, derefter køre en dynamisk liner model (DLM) og ud fra resultatet af denne at give alarmer på data. Dernæst prøver vi at finde årsagen til alarmerne ved supervised learning eller rent deskriptiv statistisk.

### Outline



Figur 1. Oversigt over fremgangsmåde til at give alarmer på tankmælk og finde årsagen til alarmerne.

## Datamateriale

Til udvikling af modellerne er brugt oplysninger fra Kvægdata-basen kombineret med temperatur oplysninger fra Danmarks Meteorologiske Institut (DMI). Der er brugt tankmælksdata fra d. 1. januar 2020 og frem. I alt indgik der 954.767 observationer fra 2.337 besætninger.

Oplysningerne blev kombineret med oplysninger om malke-system, driftsform og leverance.

Derudover blev der trukket oplysninger, om antal køer og om køerne var golde eller ej fra Kvægdata-basen. Ligeledes blev den beregnede målydelse for de enkelte køer på alle dage trukket fra Kvægdata-basen, så de kunne bruges til at korrigere for forskelle i ko sammensætning.

Som forklaring årsag blev der trukket foderkontroller fra besætninger med automatisk "daglige" foderkontroller (FBO-besætninger).

For alle besætninger blev der trukket varme data, for den 10 km \* 10 km kvadrat som besætningen ligger i. For alle de kvadrater blev der på time basis udregnet THI-indekset, som er en sammenvejning af temperatur og luftfugtighed:

$THI = AirTemp \cdot 0.8 + (AirTemp - 14.4) \cdot \frac{AirRh}{100} + 46.4$  hvor AirTemp er temperaturen og AirRh er den relative luftfugtighed.

På dags basis udregnes følgende tal:

$$THI_{70} = \sum(THI > 70)$$
$$THI_{70Vgt} = \sum THI \cdot (THI > 70)$$

$$THI = \frac{\sum THI}{24}$$
$$THI_{try} = \frac{\sum(THI - VindSpeed + GloRad)}{24}$$
$$THI_{try70Vgt} = \sum(THI - VindSpeed + GloRad) \cdot (THI - VindSpeed + GloRad > 70)$$

hvor vindspeed er vindhastigheden og GloRad er den globale stråling.

Der blev taget gennemsnit af variablene ovenfor løbende over 7 og 14 dage. Alt dette blev der set på men vi endte med kun at bruge  $THI_{70}$  for 14 dage.

## Forbehandling af data

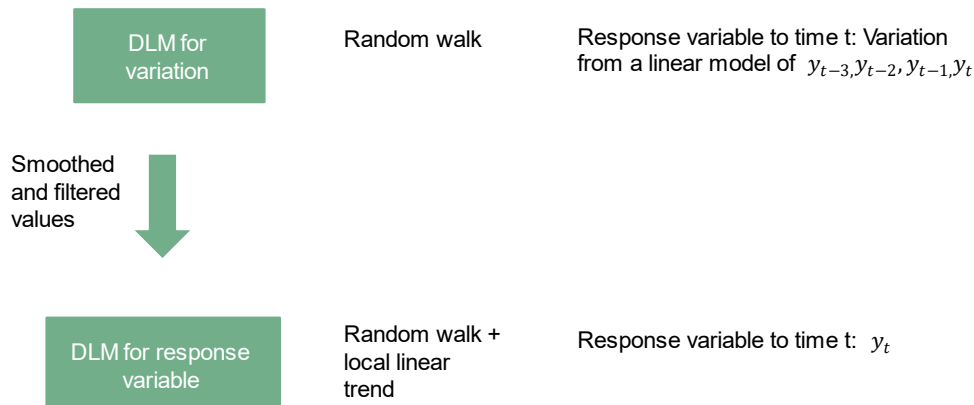
Den forventede ydelse blev den korrigeret for hvor mange malkninger som der blev forventet mellem 2 efterfølgende leveringer. Generelt blev det forsøgt at gøre dette på samme måde, som det gøres på Kvægdata-basen. Dvs. at hvis der blev brugt malkerobot så blev det antaget, at malkningerne var fordelt jævnt over døgnet timer, ellers blev der set på om det kunne forventes, at der var hele malkninger som manglede/var for meget.

Forskellen til landsgennemsnittet for tilsvarende besætninger blev udregnet ved at dele perioden op i 2-dagsintervaller og for hver af disse blev udregnet et gennemsnit for besætninger med samme race-gruppe og samme driftsform. Besætningens værdi på den pågældende dato blev derefter udregnet som den leverede værdi – landsgennemsnittet. Besætningerne blev opdelt i stor race eller Jersey, alt efter om den gennemsnitlige fedtprocent i de sidste år var under eller over 5,22%. Kun observationer hvor mælkeydelsen lå mellem 10 og 50 kg pr. ko og hvor der maksimalt var gået to mælkedage siden sidste malkning blev medtaget i gennemsnittene.

## Dynamisk lineær model - Opdeling af målinger i støj og real udvikling

En DLM-model bruges til at beskrive udviklingen i målingerne opdelt på en real udvikling og tilfældig støj (observations varians). Da den tilfældige støj kan variere mellem besætninger og over tid inden for besætning blev der først opbygget en DLM der beskrev udviklingen observationsvariansen indenfor besætningen. Som observerede værdier blev her brugt variationen omkring en linje mellem de sidste 4 punkter, se figur 2. Den reale udvikling af observationsvariens blev beskrevet ved en såkaldt random walk, dvs. en den udvikler sig tilfældigt. Der blev ikke lagt alarmer på denne model, men de smootede og filtrere værdier indgik i den efterfølgende DLM-model for udviklingen i responsvariablene, som observations varians.

### Two step DLM



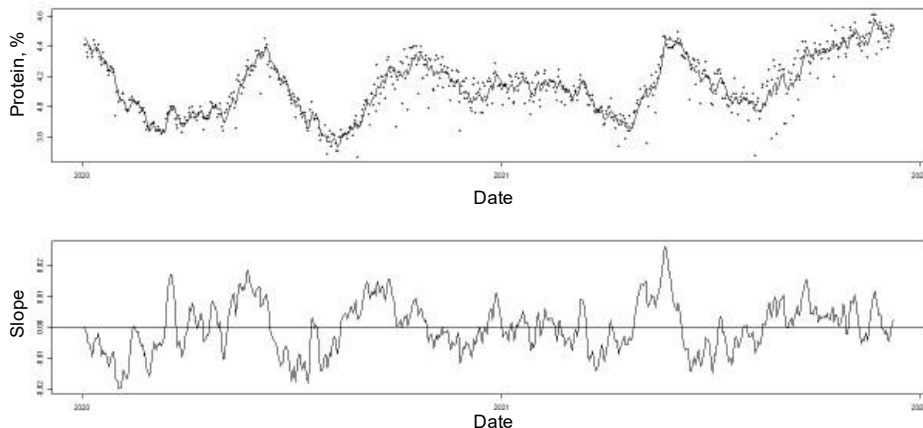
Both models estimated by maximum likelihood estimation. The ratio between different variances is bound between certain limits e.g. the system variance ( $\mathbf{W}$ ) is bound to lie between  $10^{-5}$  and  $10^{-1}$  times the observations variance ( $\mathbf{V}$ ).

SEGES  
INNOVATION

Figur 2. Oversigt over DLM-modeller på data.

I DLM-modellerne på de observerede værdier blev udviklingen i niveauet beskrevet ud fra en random walk, en lokal lineær trend og en evt. lineær effekt af en forklarende variabel. I figur 3 er for en enkelt besætning vist, hvordan niveauet for besætningen i perioder går op og i andre perioder falder, og hvordan det afspejler sig i den lokale lineære trend (hældning/slope).

## DLM - level and slope



SEGES  
INNOVATION

Figur 3. Observerede værdier og reallt niveau for proteinprocenten i besætningen (øverste figur), samt den lokale lineære trend /hældning/slope for samme besætning i samme periode (nederste figur).

Som nævnt inkluderede vi også en lineær effekt af forskellige forklarende variabel i fastlæggelsen af niveauet i nogle af modellerne f.eks. kunne antal køer eller den forventede ydelse indgå i modellerne for kg mælk eller kg EKM. Her blev startværdien sat til korrelationen mellem variabelen og output beregnet ud fra første års data. Den lineære effekt af de forklarende variable kunne udvikle sig med tiden, men kun med meget lav hastighed. Dette blev opnået med at låse forholdet mellem forskellige varianser i modellerne og det blev også på denne måde sikret, at udviklingen i niveauet blev passende fordelt mellem tilfældig udvikling og en lokal trend.

Alle model blev estimeret inden for besætning og er indrettet så de tager hensyn til at tiden mellem afhentninger kan variere. I de viste figurer, er brugt alle data til at estimere modellens parametre, men det blev tjekket om man fik et OK resultat ved at lave en løbende estimering og det fik man.

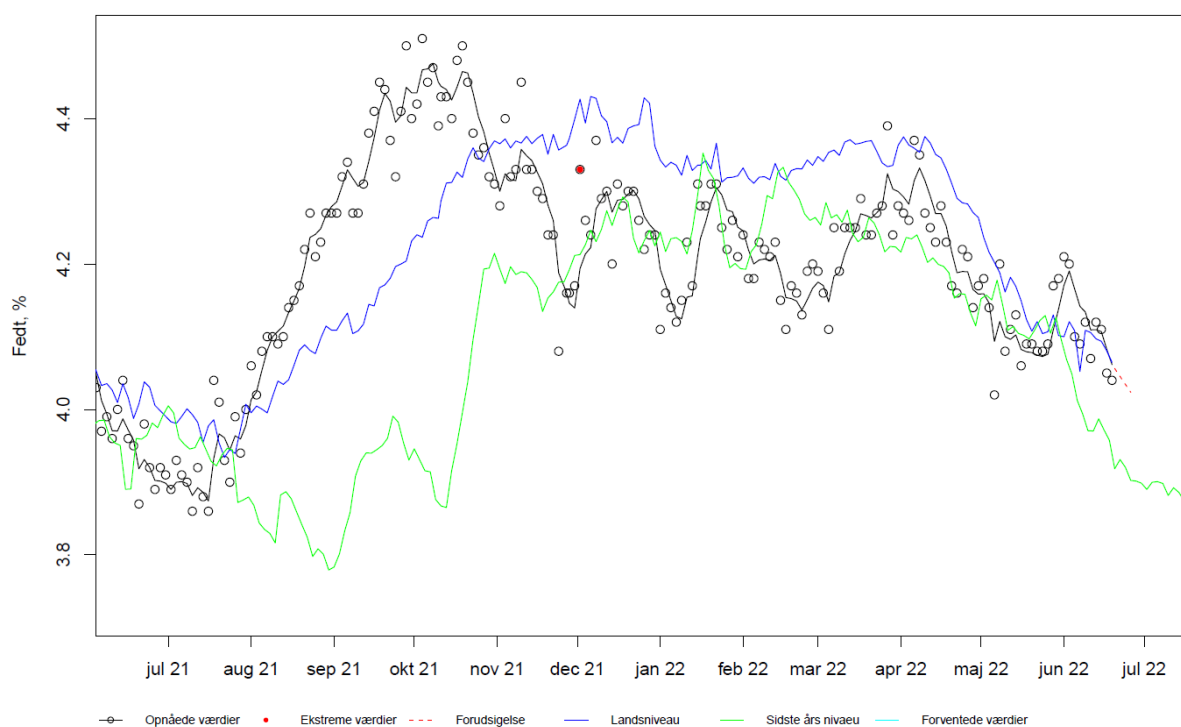
Modeller på celletallet blev både kørt på de rå celletal og for det logaritmetransponerede celletal.

### Alarmer

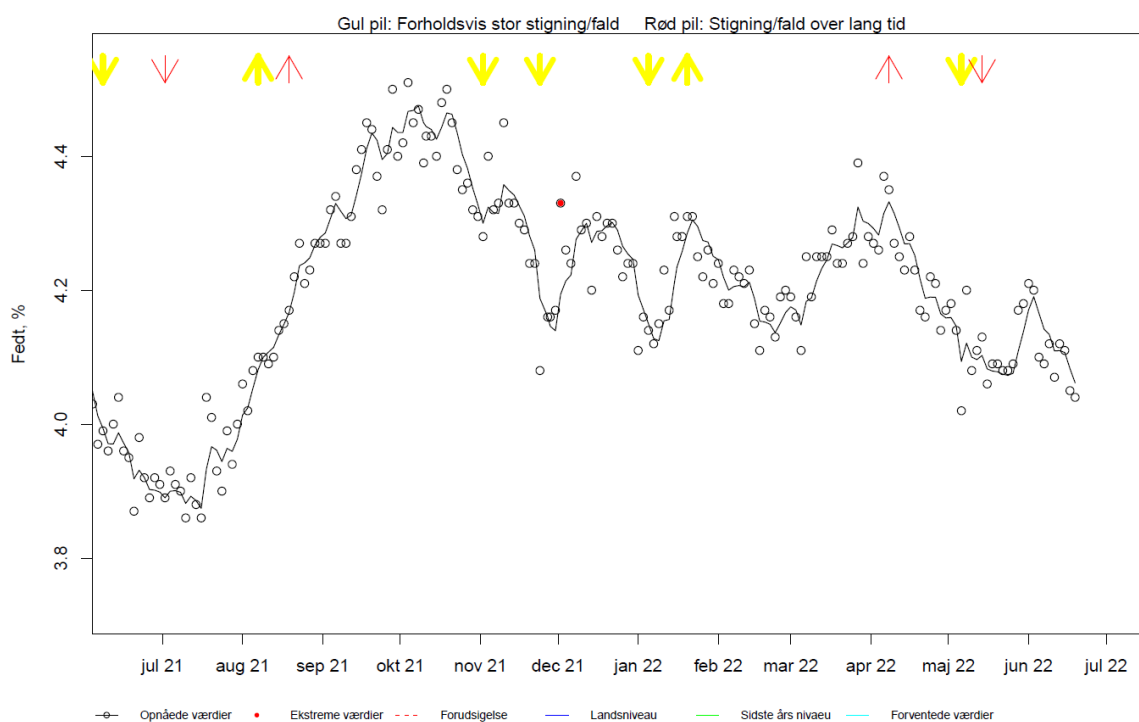
Modellerne blev brugt til at lægge alarmer på data. Alarmer blev givet som følgende:

- ❖ Ekstreme værdier: Givet tidligere målinger er den sidste måling meget usandsynlig ( $P \leq 0,0002$ )
- ❖ Alarm hurtige ændringer:
  - Hældning er signifikant større eller mindre end nul (95%)
  - Der er gået minimum 4 målinger siden sidst alarm
  - Hældningen har passeret nul siden sidste alarm
  - Hvis det ønskes er der mulighed, for at sige at ændringen skal være over en hvis størrelse, for der kommer en alarm ud, man som udgangspunkt er der ikke sat nogen grænse.
- ❖ Alarm, langsom udvikling i samme retning
  - Hældning har i 30 dage har været konstant over eller under nul

Der bliver både givet alarmer, når parameteren går op og ned. Men kunne godt overveje f.eks. for mælkeydelse kun at give alarmer, når mælkeydelsen går ned. I figur 4 a og b er vist hvordan man kan vise data og et eksempel på alarmer på en parameter på en besætning.

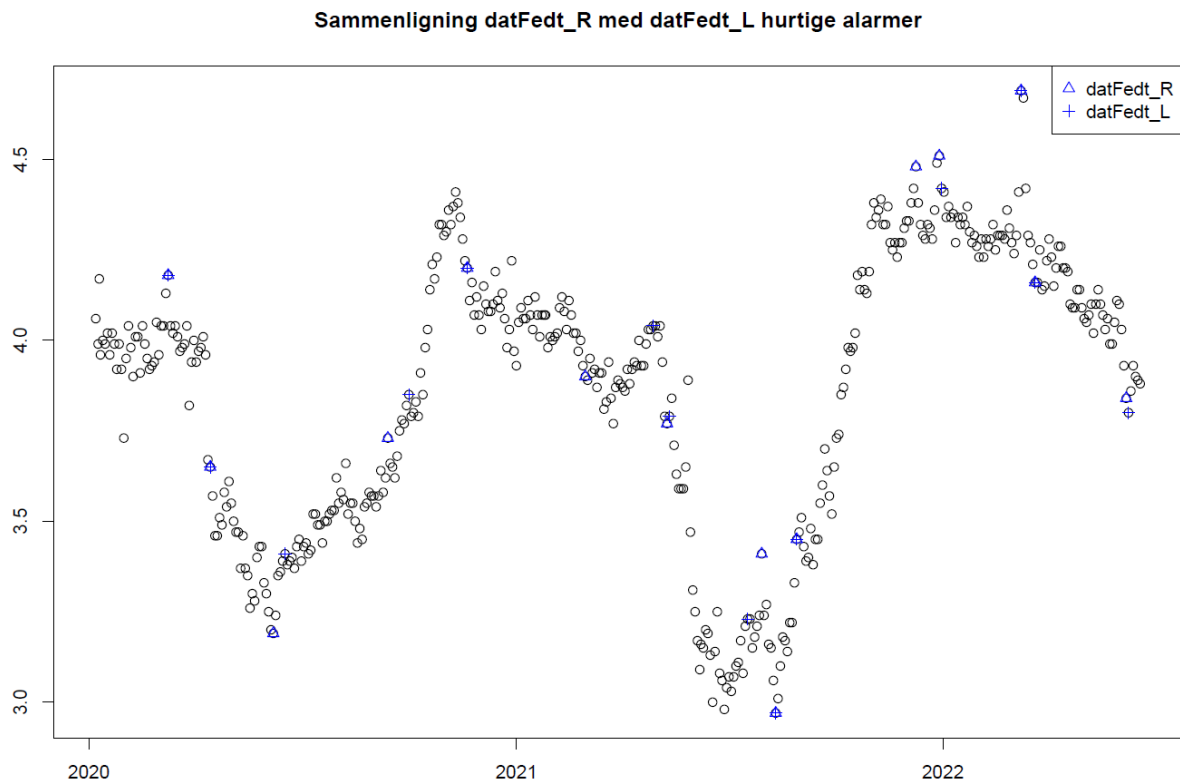


Figur 4.A. Eksempel på udviklingen i en besætning (sort), sammenholdt med sidste års niveau (grøn) og landsniveau (mørke blå)



Figur 4. B. Eksempel på udviklingen i en besætning (sort), sammenholdt med hvornår modellerne vil give alarm for hurtige alarmer (gul) og langsomme alarmer (rød).

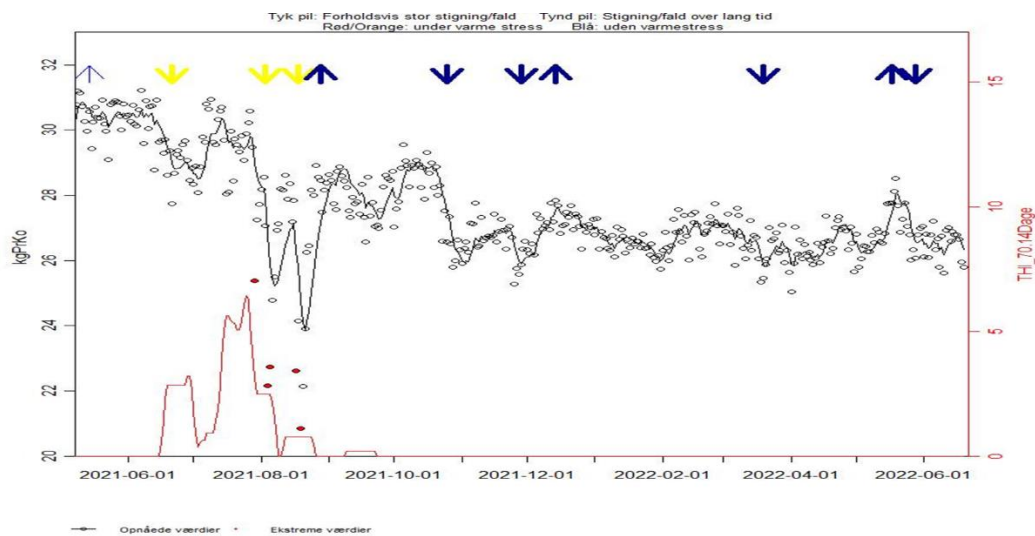
Tilsvarende eksempler som vist i figur 4 er regnet for alle parameter for minimum 500 besætninger. Der er også lavet en grafisk sammenligning for forskellige typer af alarmer, se figur 5. Ofte rykkede det ikke meget på alarmerne, om de stammede fra en model med råværdier eller de stammede fra en model for afvigelsen fra landsgennemsnit. Især for urea, hvor landsniveauet nogle gange rykker sig meget, giver det god mening at modellere på afvigelsen fra landsniveauet, mens det måske ikke giver så god mening for celletallet at modellere på afvigelsen fra landsniveauet, da man her ville komme til at korrigere den ofte forekommende top sidst på sommeren ud.



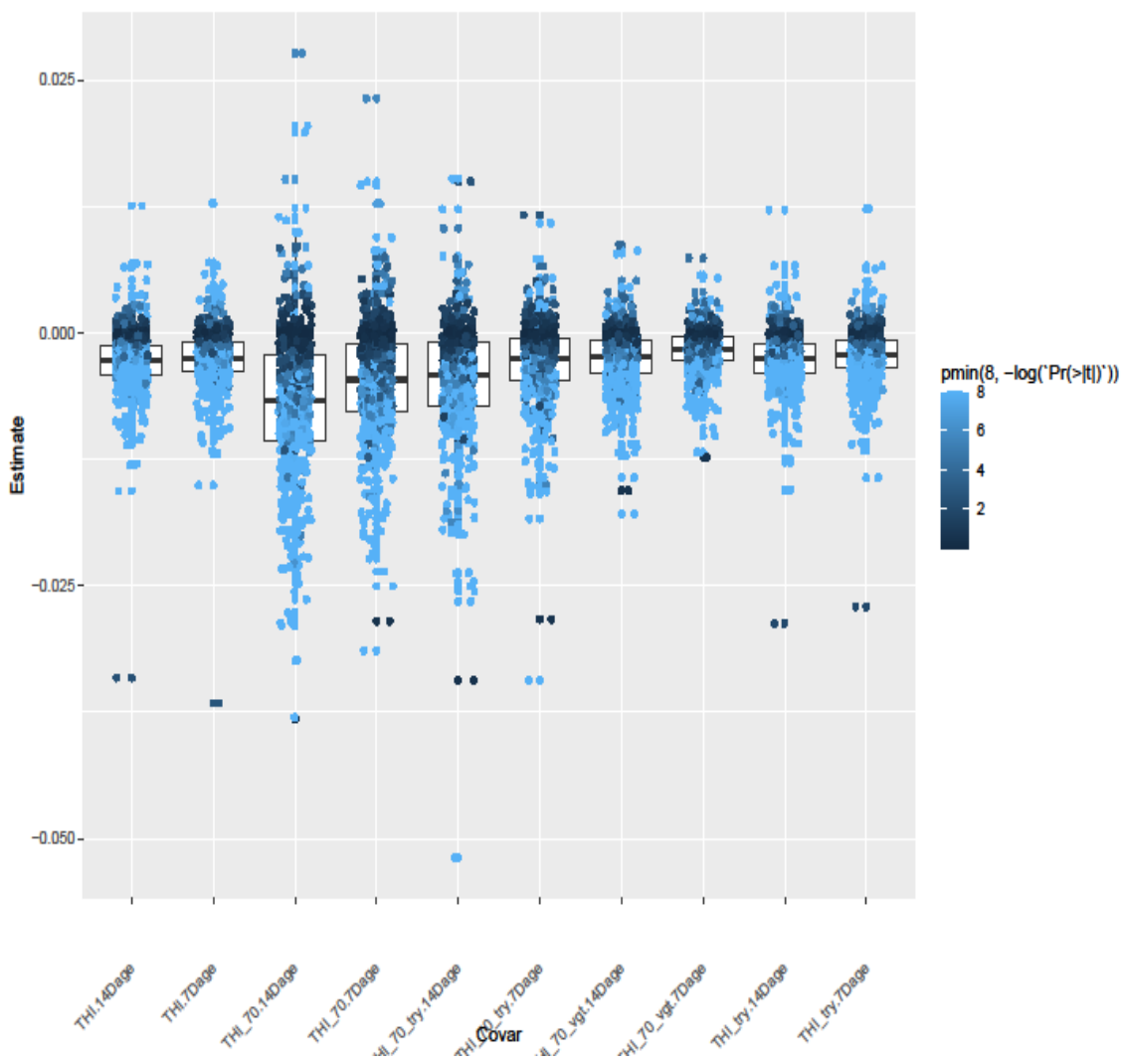
Figur 5. Hurtige alarmer dels ud fra en model for det rå niveau af fedt (datFedt\_) og dels ud fra afvigelsen fra landsgennemsnit (dataFedt\_L).

### Varmestress som årsag til alarmer

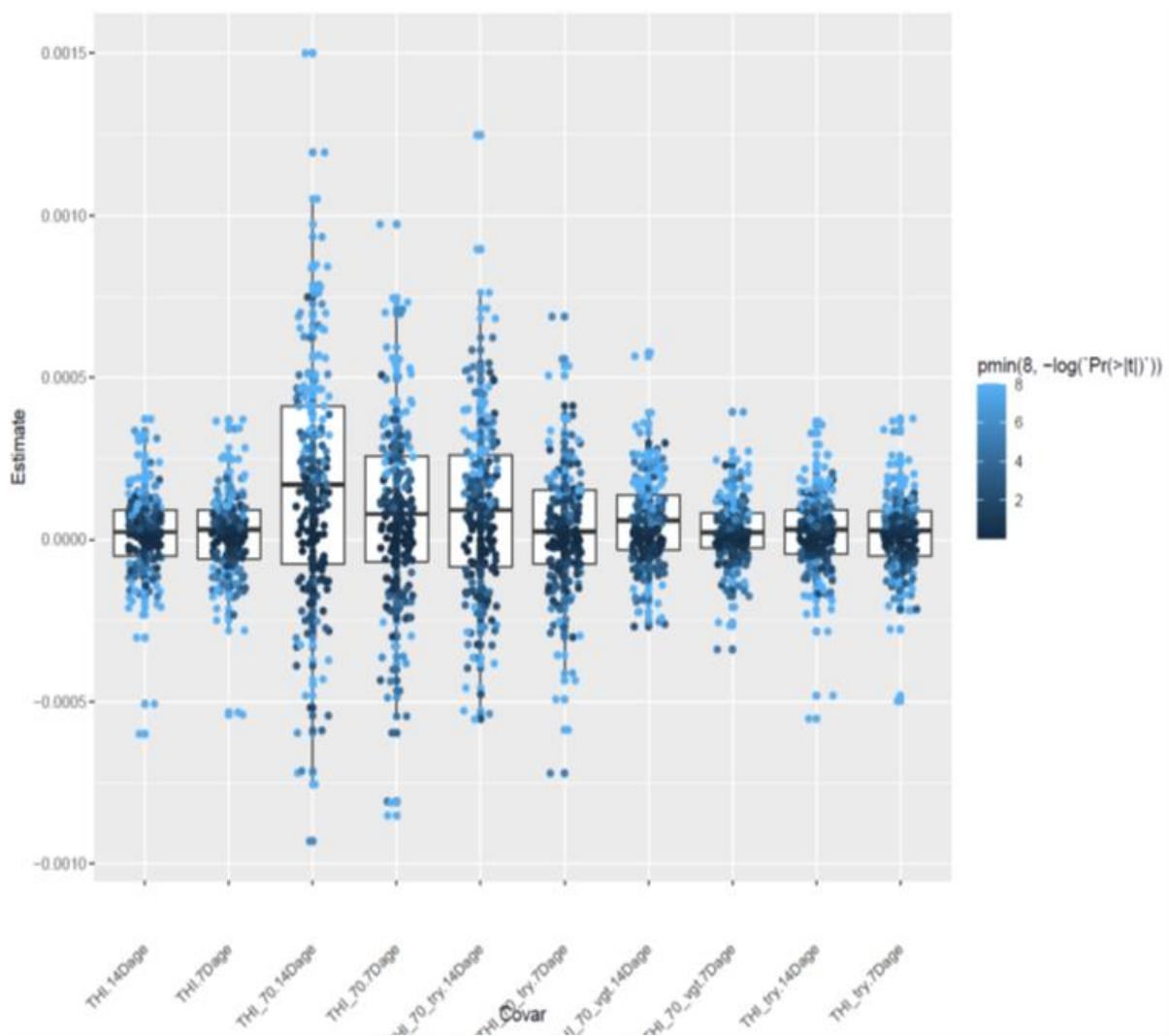
En af årsagerne til at ydelsen falder og celletallet stiger om sommeren kan være varmemstress. I figur 5 er vist et eksempel på, hvilke af alarmerne i en besætning, som kunne forklares med varme stress (gule pile). Ud over den rent deskriptive visning af sammenhængen er der også indenfor hver besætning, lavet en statistik analyse af sammenhængen mellem varmemstress og om mælkemængden/celletallet i besætningen går op eller ned udtrykt som hældningen/slop fra DLM' modellerne. I figur 6.A er vist resultatet af disse for mælkemængden og i figur 6.b er vist resultatet fra disse analyser for celletallet. Det fremgår af figurene, at for en del af besætningerne, så er der en statistisk sikker sammenhæng. Dvs. at man ved en alarm ikke kun ville kunne sige, at der har været varmemstress i perioden op til alarmen, men at det er en meget sandsynlig årsag til alarmen, da varmemstress i besætningen normalt plejer at føre til nedgang i produktionen/højt celletal på denne ejendom.



Figur 5. Eksempel på varmemstress, som forklaring på alarmer.



Figur 6.A. Sammenhængen mellem forskellige mål for THI og ændringen i kg mælk. Hver prik en besætning. Generelt negativ sammenhæng. Specielt stort for  $THI_{70}$  for 14 dage.



Figur 6.B. Sammenhængen mellem forskellige mål for THI og ændringen i Celletallet. Hver prik en besætning. Generelt "positiv" sammenhæng, dvs. når THI går op, så højere celletal. Specielt stort for  $THI_{70}$  for 14 dage.

### Fodring og alarmer

En af årsagerne til alarmerne kan være ændringen i fodringen. For besætninger med daglige foderregistreringer (ca. 350), blev der set på sammenhængen mellem foderregistreringer og alarmer. Dette blev gjort på to måder dels ved at se på ændringer i fodermidler og dels ved at se på indholdet af næringsstoffer og alarmer.

Den bedste sammenhæng mellem ændringen i brugen af fodermidler blev fundet ved at se på besætninger med store ændringer i fodermiddelsammensætningen og hurtige alarmer. I tabel 1 er vist, at hvis man tager de dage med de 20 % største ændringer i fodersammensætningen så var der 22% flere alarmer i forhold til resten af dagene, mens hvis man tager de af 5% dagene med de største ændringer i fodersammensætningen, så var der ca. 50% større risiko for en alarm. Disse tal er til trods for, at man ved en foderændring ikke altid tilstræber en ændring i ydelsen.

Tabel 1. Sammenhæng mellem store ændringen i fodersammensætningen og alarmer.

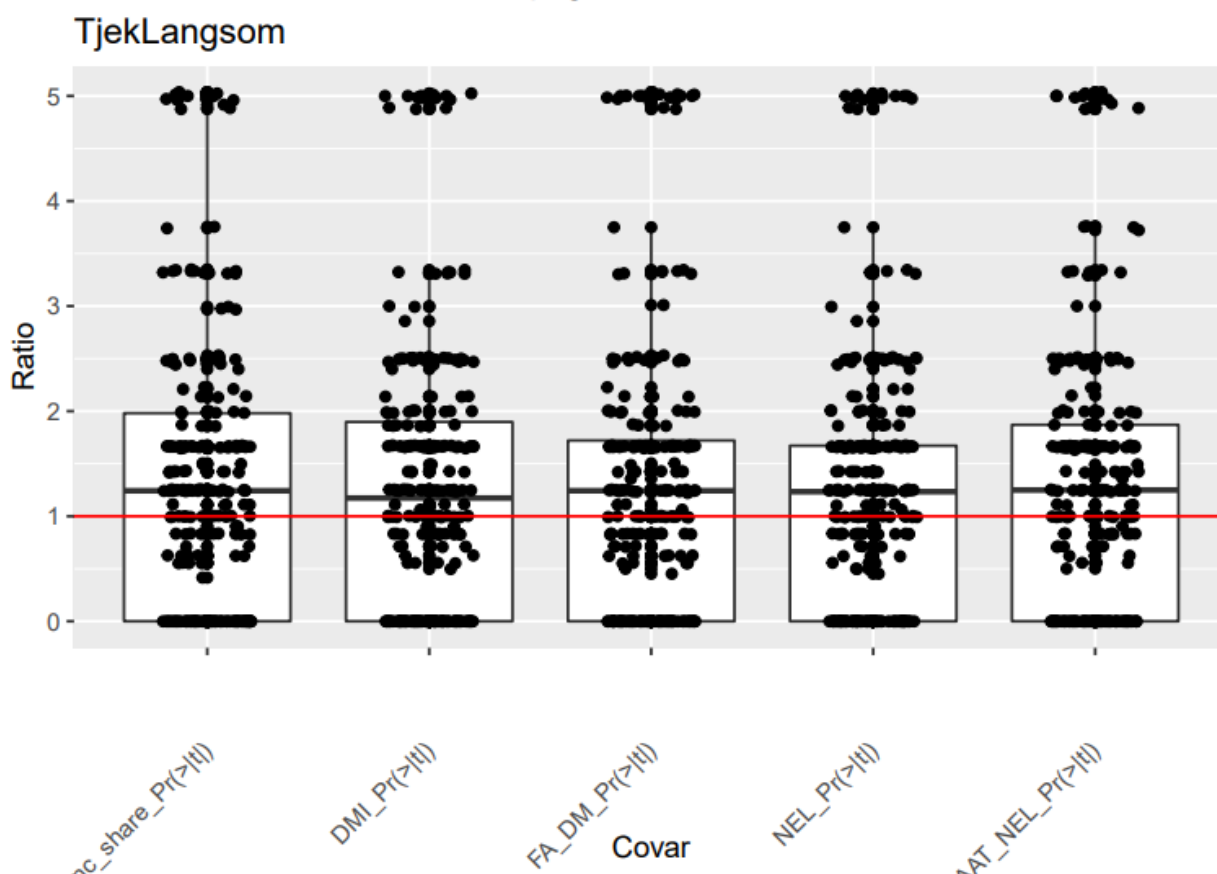
Ratio alarms pr day with change and  
alarms pr day



Days with the largest 20% of alteration in feedstuff within the last 14 days	1.22
Days with the largest 5% of alteration in feedstuff within the last 14 days	1.51

Sammenhængen mellem indholdet af næringsstoffer i foderet og alarmer, blev undersøgt ved at tage de sidst 30 observationer i besætningen og lave model over sammenhængen mellem produktionen og næringsstofindholdet i foderet. Det blev vurderet, at det vigtigste parameter, som kunne forklare ændringer i produktionen, var netto energi (NEL), tørstofindtag (DMI), andel kraftfoder (conc-share), fedtsyre gram pr. kg tørstof (FA\_DM) og AAT-indholdet (AAT\_NEL). Der blev kørt en model for hver parameter hver for sig, da en samlet model gav problemer med tolkningen pga. multicollinearitet. Resultatet af den statistiske analyse viste den bedste sammenhæng til langsomme alarmer, hvor der for de fleste parametre var en øget frekvens af signifikante udslag, når der var en alarm, se figur 7.

page 3 of 20



Figur 7. Ration af alarmer for effekt på mælkeydelse for forskellige parameter.

En måde at bruge denne sammenhæng på, kunne være ved en alarm at udskrive de parametre med den største betydning (højste standardisere koefficient). Eksempelvis:

- Alarm langsom stigning d. 18/11 2020
- Stor betydning for udviklingen:
  - Fedtsyre g pr. kg tørstof +0,46 i relativ betydning – Høj værdi for din besætning
  - NEL +0,35 i relativ betydning – Høj værdi for din besætning