

Notat: Editerings- og valideringsproce- dure for spisekvalitet, foder, metan og vægt registreringer – FutureBeefCross – **Foreløbig udgave**

Forfattere:

Rasmus Skovgaard Stephansen, Konsulent HusdyrInnovation SEGES

Line Hjortø, Specialkonsulent HusdyrInnovation SEGES

Kresten Johansen, Studentermedhjælper HusdyrInnovation SEGES

Anders Fogh, Projektleder HusdyrInnovation SEGES

Peter Løvendahl, Senior forsker AU-MBG

Morten Kargo, Senior rådgiver AU-MBG

Martin Bjerring, IT- medarbejder AU-ANIS

Bent Dueholm, Teknisk ansvarlig Allflex-Danmark

Programmer:

Indhold

Programmer:.....	1
Formål og opbygning	2
Datalagring	Fejl! Bogmærke er ikke defineret.
Kritiske målepunkter – foder og metan data.....	2
Kritiske målepunkter - foderdata	2
Kritiske målepunkter – metan data	5
Protokoller for kallibrering	7
Foder	7
Metan	7
Handlingsplan ved nedbrud på Allfeed og/eller metan målere.....	7

Data flow for data i arbejdsmappe 2	8
Editoringsprocedure for foderdata:	9
1. Fjerne lange bokstider pga. vinter og sommertid	Fejl! Bogmærke er ikke defineret.
2. Korrektion for outliers.....	9
3. Indlæsning af NorFor data	10
4. Beregn døgn foderoptagelse	Fejl! Bogmærke er ikke defineret.
5. Indsæt pedigree	Fejl! Bogmærke er ikke defineret.
6. Udlæsning af data.....	Fejl! Bogmærke er ikke defineret.
Metan data:	Fejl! Bogmærke er ikke defineret.
Editoringsprocedure for vægt data:	12
Editoring	Fejl! Bogmærke er ikke defineret.
Regression for vægt.....	Fejl! Bogmærke er ikke defineret.
Gentagelighed	Fejl! Bogmærke er ikke defineret.
Bilag 1 – SOB til landmænd ved nedbrud	13

Formål og opbygning

Formålet er at dokumentere editorings- og valideringsproceduren for spisekvalitets-, foder-, metan- og vægtdata, der er registreret i FutureBeefCross projektet.

Den generelle opbygning af dette notat er en beskrivelse af dataovervågning efterfulgt af en editering og validering af data fra projektet.

Kritiske målepunkter – foder og metan data

For at sikre en tilstrækkelig overvågning af data fra Allfeed og gasmåleranlæggene er der opstillet kritiske målepunkter (se nedstående tabel 1 og tabel 2), der giver alarmer for hvornår anlæggene kræver ekstra service og data evt. smides væk. Det er f.eks. ved nedbrud, kalibrering, ombygninger på anlægget, etc. Disse fejlmeldinger registreres i en fælles log på den fælles Dropbox.

Log filen har 2 formål:

1. Registrer tidsrum hvor data evt. skal slettes
 - Rent praktisk slettes data på SQL-serveren/eller ved indlæsning af grunddata
2. Registrer kalibrering af vægte og gasmålere, som anvendes til i kombination med KMPerne til at identificere vægte/metan målere, der laver fejl målinger.

Kritiske målepunkter - foderdata

De første punkter fra tabel 1 omhandler generel overvågning af anlægget, der udføres af Allflex. Allflex overvåger følgende KMPer:

- At data kommer løbende data ind – herved opdages nedbrud på anlæggene
- Teste om anlæggene er "live" – udføres ved en software forespørger om anlæggene kører
- Identificere hvis der er mange missing ID'er i en sti
 - Kalve uden elektroniske ørermærker
 - Antenne nedbrud
- Store afvigelser for enkelt foderkasser mellem slutvægten ved et besøg og start vægten ved næste besøg

De næste punkter i tabel 1 er en overvågning over en længere tidshorisont. SEGES/AU er ansvarlig for at overvåge følgende:

- Drift på vægte overvåges med fokus på
 - Antal daglige besøg ved foderkassen
 - Daglig udvejet foder
 - Opgøres på 14 dages data og sendes til interessenter hver mandag morgen
- Udvejningspræcision
 - Dette er mængden af udvejet foder versus det foder der er droppet
 - Opgøres på 14 dages data og sendes til interessenter hver mandag morgen
- Afvigelser GLM
 - Der udviklet en metodik (se under editering af foderdata) til at beskrive outliers i data, hvor udviklingen i antal og type opgøres
 - Opgøres på 14 dages data og sendes til interessenter hver mandag morgen
- Gentagelighed
 - Beregnes på månedsniveau for de enkelte besætninger
 - Opgøres på hele datasættet og sendes til interessenter første mandag i måneden

I tilfælde af ovenstående fejl skal disse fejlmeldes, udbedres (evt. kalibrering) og noteres i log'en.

Tabel 1. Kritiske målepunkter for foderdata fra Allfeed anlægget.

KMP	Beskrivelse for KMP	Ansvar for udførsel	Frekvens	Ansvar for handling ved fejl	Maks tid fra fejlmelding til handling
Dataoverførsel	Løbende tjekke for overførsel af datafiler med indhold	Allflex	Hver 6. time	Allflex/Bramtech	Til næste hverdag
Tjek for nedbrud	Anlæggene testes i "live" ved at en software spørger om anlægget kører	Allflex	Hver 6. time	Allflex/Bramtech	Til næste hverdag
Missing id	Der overvåges løbende på anlægsniveau antal missing værdier.	Allflex	Dagligt	Allflec/Bramtech	Til næste hverdag
Vægt afvigelser	Løbende overvåges forskellen mellem startvægt og slut vægt ved forrige besøg	Allflex	Dagligt	Allflec/Bramtech	Til næste hverdag
Gns daglig foderoptagelse	Der beregnes gns daglig foderoptagelse samt besøg og sættes nedre alarmgrænser.	SEGES/alarmgrænser	Ugentligt		
Drift på vægte	Der udarbejdes følgende plots til overvågning af drift for fodervægtene de seneste 14 dage: • Summeret udvejet foder dagligt • Antal besøg dagligt	SEGES	Ugentligt	Allflex/Bramtech	Til næste hverdag
Udvejningsandel	Mængden af foder der droppes versus det den udvejede mængde foder. Plot vises fra -10 til +10 kg og outliers sendes i tabel. Viser fra de seneste 14 dage og på kvartalsniveau i kombination med kalibrering.	SEGES	Ugentligt	Allflex/Bramtech	Til næste hverdag
Afvigelser GLM	I prækorrektionen af foderdata defineres outliers i data. Outliers indikere fejlregistreringer - plottes over tid.	SEGES	Ugentligt	Allflex/Bramtech	Til næste hverdag
Gentagelighed	Gentageligheden beregnes på månedsbasis for hver besætning	SEGES	Månedligt	SEGES, Landmanden, Allfelx/Bramtech	Til næste hverdag

Kritiske målepunkter – metan data

Allflexs overvågning af fodringsdata anvendes også for metan data til at sikre at der ikke er nedbrud, dyrene har funktionelle ID øremærker og antennerne virker. I tabel 2 er de kritiske målepunkter for metan data specificeret.

De første KMPer omhandler generel overvågning af data, som udføres på AUs SQL server:

- Under udvikling – forventes færdig sommeren 2020

De næste punkter i tabel 2 er overvågning over en længere tidshorisont. SEGES/AU er ansvarlig for at overvåge følgende:

- Under udvikling – forventes færdig sommeren 2020

I tilfælde af ovenstående fejl skal disse fejlmeldes, udbedres (evt. kalibrering) noteres i log'en.

Tabel 2. Kritiske målepunkter for metan data – under udvikling – forventes færdig sommeren 2020

KMP	Beskrivelse for KMP	Ansvar for udførsel	Frekvens	Ansvar for handling ved fejl	Maks tid fra fejlmelding til handling
Dataoverførsel	Løbende tjekke for overførsel af datafiler med indhold	AU-MBG	Hver 6. time	AU-ANIS	Til næste hverdag
Andel af totale besøg	Andelen af foderoptagelsesbesøg med metan	AU-MBG	dagligt	AU-ANIS	Til næste hverdag
Drift på måler	Løbende tjek for drift, ved stor drift skal måler kalibreres – se på gns metan				
Antal peaks pr besøg					
Forskel mellem bund og peak for-søg					
Andel negative værdier					
Forskel til reference					
Andel besøg med få peaks					
Antal dyr med under 20 metan besøg					

Protokoller for kalibrering

Foder

Beskrivelsen for rutinekalibreringen af Allfeed anlægget er indfattet i følgende protokol:

1. Frekvens for kalibrering af vægt

- Under udvikling – forventes færdig sommeren 2020

2. Beskrivelse af kalibreringen

- Under udvikling – forventes færdig sommeren 2020

Metan

Beskrivelsen for rutinekalibreringen af Allfeed anlægget er indfattet i følgende protokol:

3. Frekvens for kalibrering af metan målere

- Under udvikling – forventes færdig sommeren 2020

4. Beskrivelse af kalibreringen

- Under udvikling – forventes færdig sommeren 2020

Handlingsplan ved nedbrud på Allfeed og/eller gasmålere

Der er flere forskellige kategorier af nedbrud på anlægget, der skal håndteres forskelligt efter hvilken type nedbrud der er tale om – se tabel 3. Det vigtigste ved nedbrud på anlægget er at disse ikke går ud over dyrenes ve og vel. Ved type 4 nedbrud er der udarbejdet en SOB til landmænd i bilag 1, der beskriver hvordan landmanden kan afbryde anlægget og fodre dyrene manuelt, samt sørge for at kontakte de relevante personer.

Tabel 3. Handlingsplan ved nedbrud - under udvikling – forventes færdig sommeren 2020

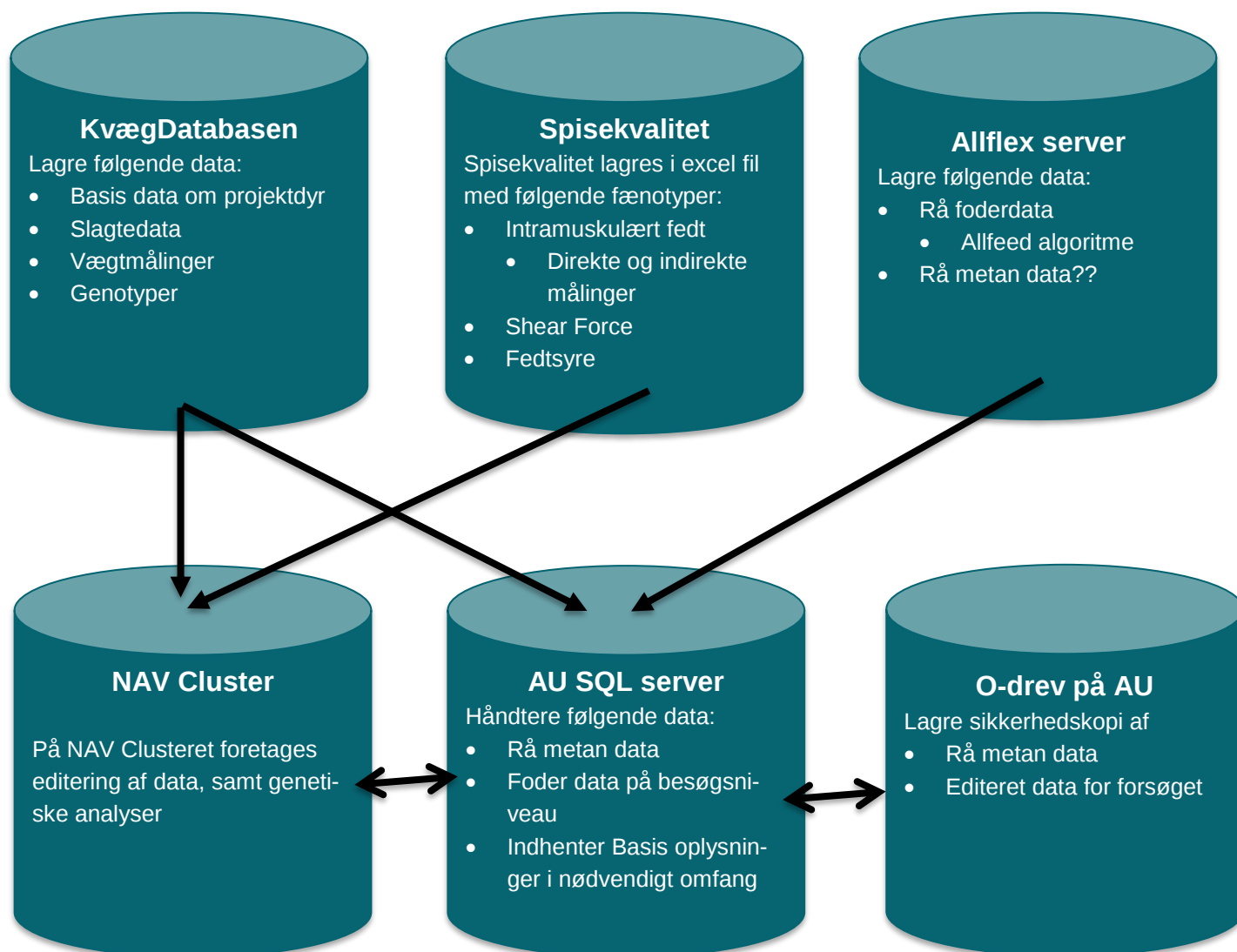
Type nedbrud	Beskrivelse af nedbrud	Handling påkrævet	Ansvarlig og reaktions tid
Type 1 – Ingen mobil kontakt og strøm nedbrud	Forbindelsen til anlæggene kan gå ned. Computerens batterier sikre at data lagres i 6 timer	Computer tjekkes, og nedbruddet registreres i logen. Dette er vigtigt for at kunne korrigere data hvor foderskodder har været åbnet	Bramtech til næste hverdag
Type 2 – Computer nedbrud, kompressor defekt	Forskellige stykker hardware omkring anlægget kan gå i stykker, af forskellige årsager	Hardware udskiftes og dette noteres i logen	Bramtech til næste hverdag
Type 3 – drift på anlæg	Vægte eller metan målerne kan drifte over tid, hvilket overvåges i KMP'erne	Ved kraftig drift er det nødvendigt ud over rutinekalibreringen at kalibrere anlægget	Bramtech til næste hverdag

Type 4 – Brand eller fysisk skade på anlægget	Ved brand eller fysisk skade på anlægget vil anlægget skulle lukkes ned og dyrene fodres manuelt	Landmanden anvender SOB til at lukke anlægget ned og kontakter straks projektet	Landmand og Bramtech kører ud til skaden så snart det er rapporteret fra landmanden
---	--	---	---

Kritiske målepunkter – spisekvalitet

- Under udvikling – forventes færdig sommeren 2020

Data flow – under udvikling



Editerings og valideringsprocedure for foder-, vægt-, og slagtedata:

Editering og validering af foder-, vægt- og slagtedata udføres i PC SAS programmet `foder_edit.sas`. Formålet er at editere de besøg væk, hvor der er sket fejlregistreringer i All-feed algoritmen. Baseret på de valide besøg kan der beregnes døgnfoderoptagelser hvorpå der udføres en udglatningsmodel. Derudover editere programmet også veje- og slagtedata.

1. Fejlregistreringer

Registreringer slettes hvis der er lange eller negative bokstider som konsekvens af vinter/sommertid eller fejlregistreringer. Derudover slettes data fra dage hvor dette er registreret i loggen og dage før opstilling af anlægget var færdigt.

2. Outliers

Proceduren til definition af outliers er udviklet på basis af notatet "Cleaning of TMR feed intake data from KFC". Denne procedure anvendes til dataeditering af rådata fra Insentic boksene på AU-Foulum.

Proceduren udføres i 2 step:

Step 1 – Fjernelse af outliers med følgende regressionsmodel:

$$\text{Foder optag} = \text{ædetid}$$

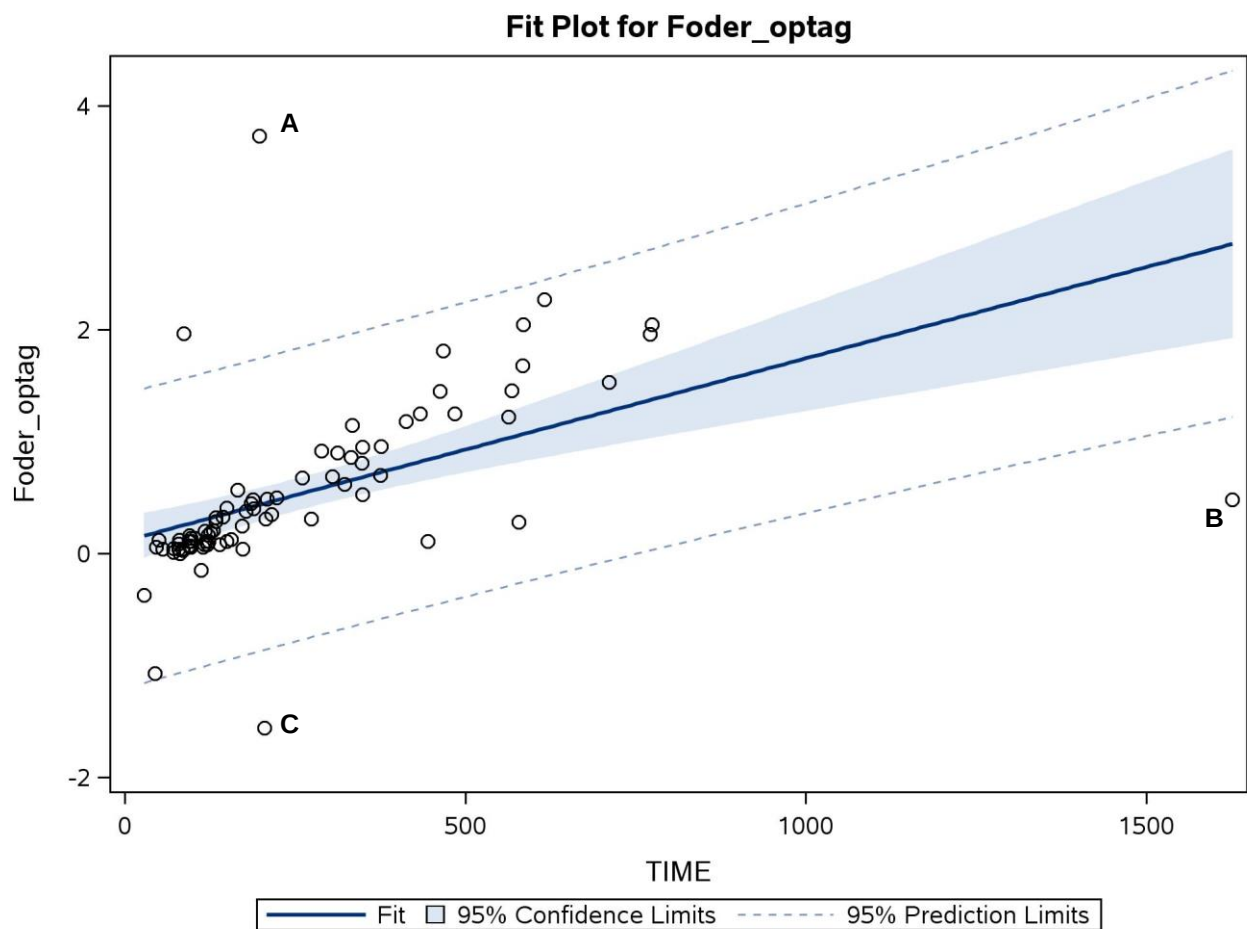
Dette udføres med Proc GLM proceduren i SAS for alle dyr. Fra modellen outputtes parameteren leverage, der er et kendt diagnosticeringsværktøj inden for statistik. Data punkter med ekstreme leverage værdier er data punkter, der påvirker regressionen ekstremt. Data punkter hvor leverage er større eller mindre end $\mu_{\text{alle dyr}} \pm 5 \times \sigma_{\text{alle dyr}}$ slettes.

Step 2 – Korrektion af outliers:

Den samme model som ovenstående udføres med Proc GLM proceduren i SAS. Fra modellen anvendes residualerne til definition af outliers. Data punkter hvor residualerne er større eller mindre end $\mu_{\text{alle dyr}} \pm 5 \times \sigma_{\text{alle dyr}}$ sættes missing. Outliers sættes missing ud fra 3 kriterier (se figur 1):

- A. Foderoptaget er større end den øvre outlier grænse - Denne værdi markeres som "afvigende foderoptag" og foderoptaget sættes missing.
- B. Foderoptaget er positivt men under den nedre outlier grænse - Denne værdi markeres som "afvigende besøgstid" og æde tiden sættes missing.
- C. Foderoptaget er negativt og under den nedre outlier grænse. Denne markeres også som "afvigende foderoptag" og foderoptaget sættes missing.

Når outliers er sat missing udføres den samme procedure igen. De besøg hvor der er missing foderoptag eller missing ædetid slettes, men der er mulighed for at indsætte en prædikeret værdier ud fra regressionslinjen i den sidste Proc GLM kørsel.



Figur 1: Plot af regressionen af foderoptag på besøgstid. Markeret med A, B, og C er outliers som vil blive korri-
geret

3. NorFor data

Værterne i projektet registrer løbende, hvilke foderkomponenter deres foderblanding indeholder. Disse næringsstovværdier indlæses i et Excel ark hvorfra det er muligt at importere relevante næringsstovværdier (tørstof%, energi, etc.) til SAS programmet. Herved er det muligt at beregne den daglige indtag af kg tørstof, MJ, etc.

4. Effekt af editering på besøgsniveau

Største delen af det slettede data fra editeringen (tabel 3) fjernes pga. data var opsamlet før færdig opstilling, dubletter, fejl på anlægget og påfyldningsregistreringer.

Tabel 4. Overblik over effekt af de enkelte editeringskrav på besøgsniveau. Data var trukket den 19. december 2019. Andel slettede data er sat i forhold til indlæst data.

Editeringskrav på besøgsniveau	Antal obs efter editering	N slettede obs	Andel slettede data
Indlæst data	513052	0	0,0%
Dubletter fjernes	479952	33100	6,5%
Data fra før færdig opstilling af udstyr	446152	33800	6,6%
Fjernelse af data via log	445029	1123	0,2%
Fjernelse dage hvor der har været korrektion på hardware	438841	6188	1,2%
Fjernelse af flange bokstider og fejl på tidsregistreringer			
Fjernelse af dyr med under 50 obs	435253	3588	0,7%
Fjernelse af ukorrekt ID og påfyldninger	408853	26400	5,1%
Leverage korrektion med 5xSD	405877	2976	0,6%
Outlier korrektion med 5 x SD	399527	6350	1,2%
Besøg uden NorFor foderkode	399525	2	0,0%
Fjernelse negative foderoptag og ædetider	390081	9444	1,8%
Sum		122971	24%

5. Udglatning af døgn foderoptagelsen

Der udregnes et rå dagligt gennemsnit for alle dyr. Daglige foderoptag slettes hvis:

- Dagligt foderoptag på 1 kg tørstof
- Under 2 besøg på en dag
- Ædetiden overstiger 6 timer for en dag
- Under 7 dage med data for det enkelte dyr

Effekterne af disse editeringskrav på besøgsniveau er vist i tabel 4.

Tabel 5. Effekt af editering på besøgsniveau. Andel slettede data er sat i forhold til antal besøg indlæst.

Editeringskrav på dagsniveau	Antal obs efter editering	N slettede obs	Andel slettede data
Antal dagsbesøg indlæst	32246	0	0%
Minimum 2 daglige besøg, 1 kg tørstof optaget og maks 6 timers ædetid total	30705	1541	4,8%
Færre end 7 dage med data	30455	250	0,8%

- Udglatningsmodellen er under udvikling – forventes færdig sommeren 2020

6. Editering vægt- og slagtedata

På AU-SQL-serveren er der udarbejdet en tabel med slagte- og vægtdata fra værterne i projektet.

Vægtdata fjernes hvis:

- Målinger før maj 2019
- Vejetyperne afviger fra "Vejning" og "Afgang"
- Vejemålinger foretaget 15 dage før/efter første/sidste dag for foderregistrering
- Vejninger hvor der har været en brutto tilvækst siden fødsel der er lavere end 0.5-2.0

Baseret på slagtedata estimeres der en levende vægt - dokumentation for denne kan erhverves ved KvægtIT SEGES.

7. Estimering af bruttotilvækst og "mid-test" kropsvægt

Baseret på vejningerne anvendes en Proc GLM procedure med følgende model:

$$Vejning = \text{alder ved vejning}$$

Følgende parametre fra modellen anvendes:

- Hældningen er fænotypen for bruttotilvækst
- Interceptet + hældningen anvendes til at beregne "mid-test" kropsvægt

Dyr med kun 1 vejning på gården, får beregnet deres bruttotilvækst og "mid-test" vejning baseret på denne 1 vejning samt levendevægt ved slagtning.

8. Udlæsning af foder og vægt data

Fra programmet udlæses der et SAS datasæt der indeholder følgende variabler:

Variabel navn	Beskrivelse
CKRDYRNR	Dyrets danske ID
Slagt_bes	CHR på opfedningsbesætning
DM_kg	Daglig tørstof optagelse i kg
FE_intake	Daglig foderenhed optagelse
NEL_intake	Daglig MJ indtag
Eat_speed	Ædehastigheden
visitdate	Dag for besøg ved Allfeed
N_visit_day	Antal besøg den pågældende dag
Eat_time_day	Total ædetid den pågældende dag i sekunder
Outlier	Outlier fra udglatningsmodellen
Pre_data	Prædikterede dagsobservationer fra udglatningsmodellen
Percent predict	Andel af dagsbesøg der er prædikeret for det enkelte dyr
Slope	Daglig tilvækgt i kg
Inter	Interceptet fra regressionsmodellen på vægt – anvendes til at beregne mid test kropsvægt
Slagt_dato	Dato for slagtning

Slagt_vægt	Slagtevægt registrert på Holsted
E_levvægt	Estimeret levendevægt ved slagtning (se dokumentation i bilag)
LEVENDEVA-EGT	Levendevægt målt på Holsted
Slagt_used	Hvis der er anvendt slagtedata til at beregne Slope og Inter er denne 1, ellers 0

Editerings og valideringsprocedure for metandata:

Dette afsnit vil blive udviklet i 2020, baseret på principperne beskrevet i Madsen et al. (2010) og Lassen et al. (2012).

Editerings og valideringsprocedure for metandata:

Dette afsnit vil blive udviklet i 2020

Bilag 1 – SOB til landmænd ved nedbrud

Dette afsnit vil blive udviklet i 2020

Referencer

Lassen, J., P. Løvendahl, and J. Madsen. 2012. Accuracy of noninvasive breath methane measurements using Fourier transform infrared methods on individual cows. *Journal of Dairy Science* 95(2):890-898.

Madsen, J., B. S. Bjerg, T. Hvelplund, M. R. Weisbjerg, and P. Lund. 2010. Methane and carbon dioxide ratio in excreted air for quantification of the methane production from ruminants. *Livestock Science* 129(1-3):223-227.