



**HAL**  
open science

## Combining datasets in a dynamic residual feed intake model and comparison with linear model results in lactating Holstein cattle

Pauline Martin, Vincent Ducrocq, Amélie Fischer, Nicolas N.C. Friggens

### ► To cite this version:

Pauline Martin, Vincent Ducrocq, Amélie Fischer, Nicolas N.C. Friggens. Combining datasets in a dynamic residual feed intake model and comparison with linear model results in lactating Holstein cattle. *Animal*, 2021, 15 (12), pp.100412. 10.1016/j.animal.2021.100412 . hal-03536564

**HAL Id: hal-03536564**

**<https://hal.inrae.fr/hal-03536564>**

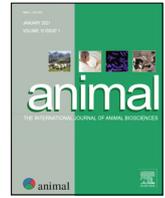
Submitted on 6 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



## Combining datasets in a dynamic residual feed intake model and comparison with linear model results in lactating Holstein cattle



P. Martin <sup>a,\*</sup>, V. Ducrocq <sup>a</sup>, A. Fischer <sup>b,c</sup>, N.C. Friggens <sup>d</sup>

<sup>a</sup>UMR GABI, INRAE, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

<sup>b</sup>PEGASE, INRAE, Institut Agro, 35590 Saint Gilles, France

<sup>c</sup>Institut de l'élevage, 149 rue de Bercy, 75595 Paris, France

<sup>d</sup>UMR 0791 MoSAR, INRAE, AgroParisTech, Université Paris-Saclay, 75005 Paris, France

### ARTICLE INFO

#### Article history:

Received 2 July 2021

Revised 22 October 2021

Accepted 25 October 2021

Available online 26 November 2021

#### Keywords:

Feed efficiency

Lactation

Methods

Modelling

Random regression

### ABSTRACT

A new method to estimate residual feed intake (**RFI**) was recently developed based on a multi-trait random regression model. This approach deals with the dynamic nature of the lactation, which is in contrast with classical linear approaches. However, an issue remains: pooling data across sites and years, which implies dealing with different (and sometimes unknown) diet energy contents. This will be needed for genomic evaluation. In this study, we tested whether merging two individual datasets into a larger one can lead to valuable results in comparison to analysing them on their own with the multi-trait random regression model. Three datasets were defined: the first one with 1 063 lactations, the second one with 205 lactations from a second farm and the third one combining the data of the two first datasets (1 268 lactations). The model was applied to the three datasets to estimate individual RFI as well as variance components and correlations between the four traits included in the model (fat and protein corrected milk production, BW, feed intake and body condition score), and a fixed month-year-farm effect was used to define the contemporary group. The variance components and correlations between animal effects of the four traits were very similar irrespective of the dataset used with correlations higher than 0.94 between the different datasets. The RFI estimates for animals from their single farm only were also very similar ( $r > 0.95$ ) to the ones computed from the merged dataset (Dataset 3). This highlights that the contemporary group correction in the model adequately accounts for differences between the two feeding environments. The dynamic model can thus be used to produce RFI estimates from merged datasets, at least when animals are raised in similar systems. In addition, the 205 lactations from the second farm were also used to estimate the RFI with a linear approach. The RFI estimated by the two approaches were similar when the considered period was rather short ( $r = 0.85$  for RFI for the first 84 days of lactation) but this correlation weakened as the period length grew ( $r = 0.77$  for RFI for the first 168 days of lactation). This weakening in correlations between the two approaches when increasing the used time-period reflects that only the dynamic model permits the regression coefficients to evolve in line with the physiological changes through the lactation. The results of this study enlarge the possibilities of use for the dynamic RFI model.

© 2021 The Authors. Published by Elsevier B.V. on behalf of The Animal Consortium. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Implications

This paper shows that a new multi-trait random regression model, estimating feed efficiency in a continuous manner across the lactation, is suitable for combining datasets from different origins, at least for similar farming systems. It adequately accounts for differences in feeding environments and allows residual feed intake estimates to be produced for datasets too small to be

analysed on their own. This opens up for deriving better residual feed intake estimates from pooled data and implies that there would be considerable value in a centralised database for combining efficiency datasets from different sites.

### Introduction

With feed costs representing above 50% of total production costs in dairy cattle farms (European Commission, 2018), improving feed efficiency has become of great importance in dairy cattle. Although different mathematical ways of estimating feed

\* Corresponding author.

E-mail address: [pauline.martin@inra.fr](mailto:pauline.martin@inra.fr) (P. Martin).

efficiency exist, residual feed intake (RFI) is nowadays the most commonly used one. First proposed in cattle by Koch et al. (1963), RFI is the difference between the actual feed intake of an animal and its predicted feed intake based on its performance, measured over a fixed time-period. The predicted intake is the intake necessary to cover the demands of the different energy sinks (milk, growth, maintenance, etc.), estimated by regression. One of the major advantages of this method is, in contrast with ratio traits such as feed conversion ratio, the built-in absence of correlation (classically at the phenotypic level but also proposed at the genetic level by Kennedy et al., 1993) between the efficiency trait and its predictors such as milk production or BW. This is particularly interesting for selection purposes (Berry and Crowley, 2013). However, some issues remain with this approach. The first issue is related to the use of static linear coefficients in the regression equation. When applied to lactating animals, this does not reflect the biological reality of different processes occurring at different lactation stages, such as body reserve mobilisation in early lactation or reserve accretion associated with pregnancy in late lactation. Therefore, estimating RFI over the whole lactation with a linear regression leads to biased assessments (Li et al., 2017). Various studies have explored further the methodology of RFI or the relationships between its components (e.g. Manzanilla Pech et al., 2014; Strathe et al., 2014; Lu et al., 2015) which led recently to the development of new methods based on multi-trait random regression models that incorporate the dynamic dimension of the data and propose a solution to this first issue (Islam et al., 2020; Martin et al., 2021b). However, as the number of parameters to estimate in the model is considerably greater than for a classical linear regression, the number of records required for its estimation is greater, leading to a second issue. Residual feed intake datasets tend to be relatively small due to the difficulty and cost of measuring intake. Therefore, large-scale phenotyping (in order to provide a reference population for genomic selection for example) requires pooling of data across different farms and years. However, pooling across sites and years implies merging animals with different diets that all have their own energy contents. Then, if DM intake is used as the measure of intake, the efficiency estimation is polluted by the different feed energy densities. Using a conversion to energy intake based on book values could also introduce bias and potentially squeeze out any variance in digestive efficiency (Martin et al., 2021a). The literature offers ideas on how to deal with this issue. For example, Tempelman et al. (2015) and Lu et al. (2017) used random ration-specific regressions of DM intake on milk energy and BW. However, all the needed information about the cows' diet is not always available. Therefore, it is useful to search for simpler methods that could be used at least within broad farm feeding types, especially while using methodologies such as Martin et al. (2021b) that deals with the first issue as well. Accordingly, this study has two objectives. First, we test whether this simplification of using only a contemporary group fixed effect is suitable to deal with diet differences, as well as merging a small dataset with a larger one and compare results obtained when computed on the merged dataset or on the initial datasets analysed separately. Second, we continue testing the dynamic approach by comparing its RFI estimates with estimates produced from a linear regression model computed on the same datasets.

## Material and methods

### Population resources and feeding management

The records used in this study came from two different datasets. Dataset 1 included records collected between 2002 and 2016 at the Danish Cattle Research Centre (Foulum, Denmark) coming from

Holstein cows in first to third lactation. Housing conditions and details on the partial mixed ration were extensively described in Martin et al. (2021b). In short, the diet was based on rapeseed meal, barley, dried sugarbeet pulp, grass silage, maize silage, urea and minerals for a typical net energy of lactation of 6.9 MJ per kg of DM. The nutritional values of the partial mixed ration were all within the range of typically recommended requirements, formulated to support the milk yield level of the herd, and allocated in amounts allowing approximately 10% orts to ensure ad libitum intake. In addition, cows were supplemented daily with a maximum of 3 kg of concentrate in the milking parlour. Dataset 2 consisted of records collected between 2014 and 2017 at the INRAE UMR PEGASE Méjusseume experimental farm (Le Rheu, France) coming from Holstein cows in first to seventh lactation. Animals were housed in a freestall barn and fed with a total mixed ration previously described in Fischer et al. (2018). In short, this diet was based on maize silage, energy concentrate, soybean cake, dehydrated alfalfa and minerals. It had a net energy density of 6.6 MJ per kg of DM. Similarly to dataset 1, cows were fed ad libitum with a daily target of 10% orts per cow.

### Phenotypes and data editing

For dataset 1, a weekly measure of average daily milk yield per cow was obtained by averaging daily milk yield records per cow in each week. Milk samples were taken weekly for fat and protein analyses. Dry matter content of the mixed ration and concentrates were analysed regularly and the compositions were aligned and merged with feed intake records to obtain weekly DM intake (DMI) values for each cow. Animals were also automatically weighed at each milking so that BW records were averaged to obtain a weekly record of BW per cow in each week. Body Condition Score (BCS) was evaluated every two weeks and scored on a scale from 1 to 5 (Kristensen, 1986). For dataset 2, milk quantity, BW and DMI were obtained daily, milk fat and protein were analysed two days a week and BCS was evaluated monthly on a scale from 0 to 5 (Bazin, 1984). All daily measurements were averaged per week.

A corrected milk (cmilk) trait was created for both datasets following the FAO formula which defines as a standard milk with 4.0% fat and 3.3% protein (FAO, 2010):

$$Cmilk (kg) = raw\ milk (kg) * (0.337 + 0.116 * fat\ content (\%) + 0.06 * protein\ content (\%))$$

The same filters as the ones used in (Martin et al., 2021b) were applied to avoid non-sensical performances. Therefore, records implying differences between two consecutive records higher than 12 kg for cmilk, 50 kg for the BW, 7 kg for the daily DM intake and one unit of BCS were discarded. Less than 2% of the data were discarded for cmilk and BCS and about 4% for BW and DMI. Both datasets included at least 200 days of records within each lactation for each animal. The final number of animals and lactations was 564 animals and 1 063 lactations for dataset 1 and 149 animals and 205 lactations for dataset 2. A third dataset, Dataset 3, was created by merging Dataset 1 and Dataset 2 together, leading to 1 268 lactations from 713 animals.

### Statistical analyses

Dynamic model. After some analyses to get descriptive statistics performed using Proc Means under SAS/STAT® software for each farm separately, the methodology presented in (Martin et al., 2021b) was applied to the three datasets using the Wombat software (Meyer, 2007) to estimate a dynamic RFI for each animal at each time point. Briefly, the approach used was as follows:

First, a multi-trait random regression model was applied to the data using each of the four traits (cmilk, BW, DMI and BCS) the following equation:

$$y_{ilmr} = c_i + \sum_{n=0}^3 \beta_{ln} \varphi_{nr}(t) + \sum_{n=0}^2 \alpha_{mn} \varphi_{nr}(t) + e_{ilmr}$$

where  $y_{ilmr}$  is the observation of the  $r$ -th trait,  $c_i$  is the fixed effect of the contemporary group, i.e. the  $i$ -th month-year combination for Datasets 1 and 2 and the  $i$ -th farm-month-year combination for Dataset 3, corresponding to the record date,  $\beta_{ln}$  is the  $n$ -th fixed regression coefficient specific to parity class  $l$ ,  $\alpha_{mn}$  is the  $n$ -th random regression coefficient of the animal  $m$ ,  $\varphi_{nr}(t)$  is the  $n$ -th coefficient of a Legendre polynomial of degree 2 for the animal effect and degree 3 for the parity class, evaluated at day in milk (DIM)  $t$ , and  $e_{ilmr}$  is the random residual effect. Successive lactations from the same animal were considered as coming from different animals. Within trait, residual effects were assumed to have a homogenous residual variance as no differences were observed during tests using heterogeneous residual variance except an increase in calculation time.

For each trait and lactation, an estimated animal effect was generated for each day of lactation. These animal effects will be denoted  $a_{cmilk}(t)$ ,  $a_{weight}(t)$ ,  $a_{DMI}(t)$  and  $a_{BCS}(t)$ . It is worth noting that as no pedigree nor genomic information was included in this study, the animal effect includes both the genetic and the permanent environmental effects.

Then, by multiplying the variance components of the random regression by the corresponding time coefficients of the Legendre polynomials, variances and covariances at each time point were estimated for each trait and used to calculate correlations between traits across the lactation.

Using elements from the resulting variance-covariance matrix and the animal effects for cmilk, weight and BCS, a predicted intake was estimated by a multivariate regression. Finally, the RFI estimate was obtained as the difference between the actual animal effect for DMI and the one predicted from the three other variables (cmilk, BW and BCS). As this RFI is dependent on time, we also defined RFI<sub>tot</sub> as the averaged RFI of each animal over the whole lactation.

Linear model. In order to compare the results obtained from the dynamic approach to a more classical approach, RFI estimates were also calculated on animals from Dataset 2 from a linear regression model as proposed by Fischer et al. (2018). Briefly, after centring the independent variables, the following mixed model was applied as follows:

$$\begin{aligned} DMI_j = & (\mu + \mu_j) + a \times NEMilk_j + b \times (BW^{0.75})_j \\ & + c \times (BWgain * BCS)_j + d \times (BWloss * BCS)_j \\ & + parity + Year + Fortnight * Year + \varepsilon_j \end{aligned}$$

where Fortnight is the fixed effect of lactation fortnight (14-day period),  $DMI_j$  is the mean daily DMI calculated over each fortnight for cow  $j$ ,  $\mu$  is the intercept,  $NEMilk_j$  is the mean daily net energy in milk of cow  $j$  calculated for the considered fortnight according to the equation below,  $BW^{0.75}$  is the mean metabolic BW of cow  $j$  on the considered fortnight,  $(BWgain * BCS)$  is the body reserve gain in the considered fortnight calculated as the fortnight average of the product between BW gain and BCS, and  $(BWloss * BCS)$  is the equivalent for body reserve loss, Year is the fixed effect of year of experimentation and Fortnight\*Year is the interaction between the fixed effect of fortnight within each year of experimentation.  $\mu$  is the fixed part of the intercept and  $\mu_j$  is the individual animal deviation of the intercept,  $a$ ,  $b$ ,  $c$  and  $d$  are the slopes. The residual feed intake is defined here as  $\mu_j$ , the individual animal deviation of the intercept. The model also included the simple and two-way interactions

with parity and Year. This mixed model was carried out using the Proc Mixed of the SAS/STAT® software, using a repeated effect of cow nested within lactation with a first order autoregressive variance-covariance matrix and a random effect of cow nested within lactation, as defined in Fischer et al. (2018).

The net energy in milk was estimated according to the equation of INRA (2010):

$$\begin{aligned} NEMilk = & (MY \times [0.44 + 0.0055 \times (FC - 40) \\ & + 0.0033 \times (PC - 31)]) \times 7.12 \end{aligned}$$

where NEMilk is the net energy in milk in MJ/d, MY is the milk production in kg/d, FC is the milk fat concentration in g/kg and PC is the milk protein concentration in g/kg.

Mean RFI for the two farms of Dataset 3 were compared using Proc Ttest of the SAS/STAT® software (significance threshold of 0.05). Correlations between the results obtained from the different datasets or from the different models were calculated using Proc Corr.

## Results

### Descriptive statistics

Descriptive statistics for both farms, averaged over the lactation, are reported in Table 1 for the raw data of the four considered traits. As already observed by Martin et al. (2021b), some values are especially low (for example, 3.1 kg of cmilk or 5.0 kg of DM) but they are unlikely to be measurement errors because first, data were averaged on a weekly basis and second, they are relatively close to the previous value for a given animal (due to the filtering step). These extreme values are possibly due to health issues. The raw data of the two farms are very similar with the only exception being the body condition score, which shows a difference of one point between the farms. As the standard deviations are nearly the same, it is likely due to a difference in the scale of notation (scale of 1 to 5 for Aarhus and 0 to 5 for Méjusseaume). The similarity between farms lasts during the entire lactation as illustrated by Fig. 1.

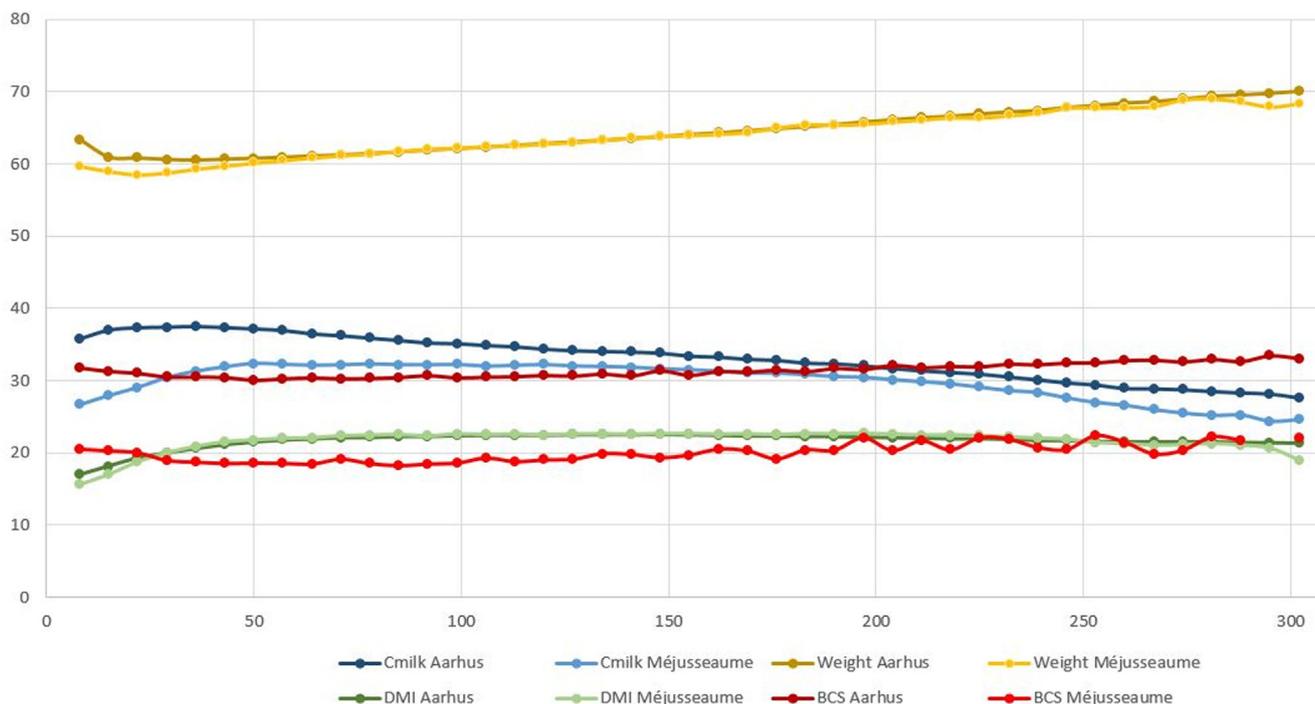
### Comparison of the variance-covariance matrix obtained from the three datasets

Each dataset leads to 430 variance components and 258 correlations (four traits and 43 weeks of lactation). The correlations between the three datasets for these results (the variance components and the correlations between traits) are presented in Table 2. As expected given the large role that Dataset 1 plays in Dataset 3 constitution, the correlations between Dataset 1 and Dataset 3 are almost one. However, the correlations between Dataset 2 and Dataset 3 are extremely high as well (>0.95). This is also the case for the correlations between Dataset 1 and Dataset 2, i.e. between the two farms analysed separately (>0.94). It is interesting to note that the slight difference between dataset 2 and the two other datasets is mainly linked to BCS. In fact, for the 258 correlations among the four traits estimated within each dataset, the average value for a given correlation estimated from dataset 2 is 0.08 higher than the same correlation estimated from one of the two other datasets. But this average difference of 0.08 hides variability: for correlations between cmilk, weight and DMI, the average difference between estimates from dataset 2 and estimates from dataset 1 is 0.03 (range of -0.08 to 0.15) while this difference averages 0.12 (range of -0.07 to 0.34) for correlations between BCS and one of the three other traits. The difference between datasets also increases through days in milk.

The residual variances of the four traits estimated from the three datasets are presented in Table 3. Some differences are

**Table 1**  
Number of observations (n), means, SD, minimum (min) and maximum for the four predictor traits over all lactations in each of the two dairy cattle farms considered.

Trait	Farm	N	Mean	SD	Min	Max
Corrected Milk	Aarhus	40 619	33.4	7.8	3.1	68.9
	Méjusseume	7 734	30.4	5.9	4.2	58.0
BW	Aarhus	40 662	642.5	74.8	398.9	982.9
	Méjusseume	7 131	632.2	69.2	404.0	901.0
DM Intake	Aarhus	42 177	21.7	3.4	6.0	38.7
	Méjusseume	7 266	21.7	3.3	5.0	30.4
Body Condition Score	Aarhus	19 661	3.13	0.34	1.50	4.75
	Méjusseume	1 698	1.98	0.35	1.25	3.67



**Fig. 1.** Evolution of the mean corrected milk (cmilk), BW, DM intake (DMI) and body condition score (BCS) by dairy cattle farm over the lactation (expressed in day in milk). With cmilk expressed in kg, weight expressed in 10 kg, DMI expressed in kg and BCS expressed in 0.1points on the y-axis and the x-axis representing days in milk.

**Table 2**  
Correlations between the equivalent variance components and correlations obtained from the three dairy cattle datasets.

Datasets	Correlation on the variance components	Correlation on the correlations between traits
Dataset 1/Dataset 3	0.999	0.996
Dataset 2/Dataset 3	0.953	0.951
Dataset 1/Dataset 2	0.940	0.964

Dataset 1: Animals from Aarhus farm only; Dataset 2: Animals from Méjusseume farm only; Dataset 3: Animals from both farms.

observed between the two farms when analysed individually. The residual variances from the model combining the two farms are intermediate but closer to Dataset 1, which represents 80% of the data from Dataset 3.

*Comparison of residual feed intake estimates depending on farm origin in dataset 3*

The average RFI of the 1 063 lactations from Aarhus on one hand, and of the 205 lactations from Méjusseume on the other hand, were estimated at each time point and are presented in Fig. 2. Animals from Aarhus seem to be on average slightly more

**Table 3**  
Residual variances of the four traits for the three dairy cattle datasets.

Trait	Dataset 1	Dataset 3	Dataset 2
Corrected milk (kg <sup>2</sup> )	5.96	5.42	2.61
BW (kg <sup>2</sup> )	8.31	10.00	19.62
DM Intake (kg <sup>2</sup> )	1.54	1.54	1.49
Body Condition Score	0.31	0.30	0.12

Dataset 1: Animals from Aarhus farm only; Dataset 2: Animals from Méjusseume farm only; Dataset 3: Animals from both farms.

efficient in early lactation than animals from Méjusseume but after 120 days in milk, this difference switches with animals from Méjusseume appearing on average more efficient. However, this difference never exceeded 0.2 kg of DMI, which is very small relative to daily intake, and given the sample size, this difference is never significant over the lactation ( $P > 0.05$ ).

*Comparison of residual feed intake estimates and efficiency ranks between datasets*

Estimates for RFI and ranking of the animals based on their RFI averaged over the entire lactation were calculated for each dataset.

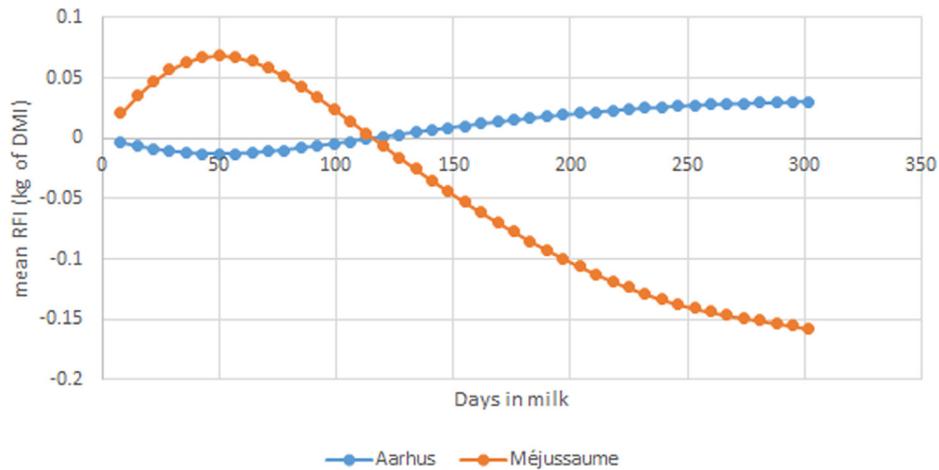


Fig. 2. Evolution of the mean residual feed intake (RFI) by dairy cattle farm over the lactation (from the results of Dataset 3).

For the animals from the Aarhus farm, the correlation between their RFI estimated from Dataset 1 and from Dataset 3 is almost one (>0.99) and the same is observed for the efficiency ranks. For Méjusseume animals, a correlation of 0.95 is obtained between Dataset 2 and Dataset 3 for RFI estimates and 0.94 for efficiency ranks, indicating very limited reranking. These correlations are illustrated in Fig. 3. For both farms, the standard deviations of RFI (mean s.d. of the daily s.d.) are similar for the two dataset estimations, but slightly higher for Aarhus (s.d. = 1.15 kg DM/d in both cases for Aarhus from datasets 1 and 3 vs 0.72 and 0.78 kg DM/d for Méjusseume RFI evaluated from Datasets 2 and 3, respectively).

Comparison of residual feed intake estimates between the dynamic and the linear residual feed intake approach

RFI estimations were available for Dataset 2 both from the random regression model and a linear model. Correlations between the results obtained from these two approaches are presented in Table 4 for two different time intervals (1 to 84 and 1 to 168 DIM). The correlation between the two time intervals is very high when using the dynamic approach (0.9) and quite high as well between the two time intervals when using the linear approach (0.72). The correlations between the two approaches were also

high for each period: 0.77 and 0.85 for 1 to 84 and 1 to 168 DIM, respectively.

When comparing RFI results of a given animal on successive years, the correlations are higher on both periods for the linear model, which included the animal as an effect in the model, than in the dynamic model, which considered different lactations independently as different animals (1 to 84 DIM: 0.55 vs 0.25; 1 to 168 DIM: 0.75 vs 0.55). These correlations are, however, estimated from a relatively small number of animals (50). The same calculation when performed on the Aarhus animals on the overall lactation gives correlation results of 0.43 between lactations 1 and 2 (345 animals), 0.47 between lactations 2 and 3 (154 animals) and 0.32 between lactations 1 and 3 (143 animals).

Discussion

Pooling datasets: A useful opportunity

The first objective of this study was to test the behaviour of the dynamic model of RFI analyses when merging data from different origins and test its suitability to deal with the differences between feeding environments when combining datasets. When combining datasets, a correction of the environmental conditions is necessary

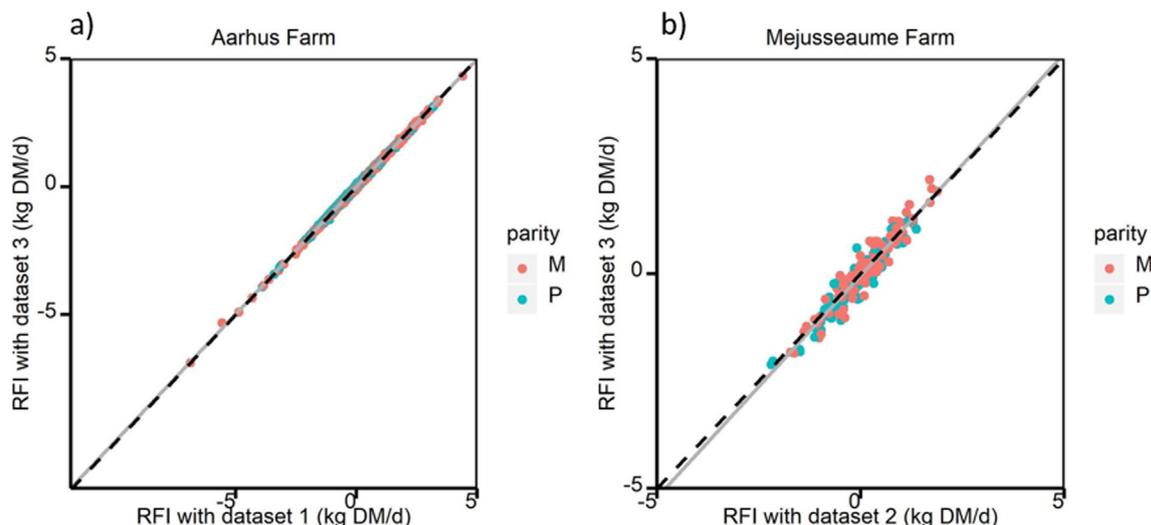


Fig. 3. Comparison of the residual feed intake (RFI) obtained for each dairy cow in the single-farm VS multi-farm analysis. Where a) is the graph for animals from the Aarhus farm and b) the graph for animals from the Méjusseume farm. P means primiparous and M multiparous cows. Each dot corresponds to the average RFI of one animal over all the lactation.

**Table 4**

Correlations between RFI estimates of dairy cows obtained from the dynamic estimation approach and from the linear approach for the periods: days in milk 1 to 84 and days in milk 1 to 168.

Item	Linear RFI DIM 1 to 168	Dynamic RFI DIM 1 to 84	Dynamic RFI DIM 1 to 168
Linear RFI DIM 1 to 84	0.72	0.85	0.85
Linear RFI DIM 1 to 168		0.55	0.77
Dynamic RFI DIM 1 to 84			0.91

Abbreviations: RFI = Residual feed intake, DIM = days in milk.

to account for the different nutrient contents of the feeds used in the different experiments. It also serves to adjust for all other environmental differences and thereby avoid bias such as the Simpson's paradox (i.e. misjudgement of an overall result when combining data of different groups without taking the group information into account). Accordingly, a contemporary group effect (farm-month-year) was used as a corrective effect. This is a usual way to combine data from different farms in a RFI model (Tempelman and Lu, 2020; Taussat et al., 2020). The fact that no significant difference of RFI was found in the present study between the average levels of the two farms strongly suggests that such a correction adequately accounts for differences in the feeding environment. In addition, the high correlations between estimates when considering farms separately or together confirm the suitability of the RFI dynamic approach for pooled datasets, at least when the different original datasets are quite similar.

Pooling data from different farms in the same model presents several benefits. First, increasing the sample size leads to improved accuracy of the estimates of variance-covariance components. Therefore, when estimated from a larger dataset, more representative of the population, the correlations between traits can be considered as more reliable for further use on other animals. Being able to pool phenotypes from different farms is also particularly important for genetic and genomic analyses, not only for an increase of accuracy in estimation of the genetic parameters but also because intake is a phenotype that is difficult to obtain. The implementation of a genomic selection for feed efficiency would be of great interest. Such selection requires a large reference population of several thousands of animals and therefore pooling data from different farms is necessary. Even though we were unable to do it in this article due to the lack of genomic information, the dynamic model is easily adaptable to include such a component (when genotypes are available) and it could be used in the implementation of a genomic evaluation. Finally, pooling datasets may also be useful to obtain estimates on very small datasets. When analysing alone very small datasets with a random regression model, it is possible that too few information are available to allow convergence. Therefore, pooling information from different datasets appears to be a promising solution to get estimates from datasets too small to be considered alone.

#### *An unknown extend of these conclusions due to farm similarity*

If the situation exposed here showed very good results on merging data from different origins, it is worth noting that the two farms used in this study were rather similar in terms of management and performance levels. Apart from the differences in feed composition, the only notable difference in this study was the BCS that was not recorded on the same scale. This difference had an effect on the variance-covariance components and caused changes in correlations among traits up to 0.3–0.4 in late lactation, which is not negligible. Even if in the present case, the model

correctly handles estimations of animals from different farms (and anyway such scale effects could be corrected for), this small difference in BCS and its consequences provide a warning with respect to the limits of using a simple linear adjustment to deal with between farm differences. One could ask if the approach would be able to deal with data from very different origins and diets. It could be hypothesised that the model would produce biased estimates for a small dataset of, for instance, grazing animals if it is combined with a large dataset of indoor highly producing cows. Therefore, the present results need to be confirmed, or infirmed, on farms with markedly different management strategies and performance levels. However, currently, most farms with intake data on a large number of animals use indoor feed recording systems with more or less the same type of management, which limits the possibility of testing on radically different situations.

Another limit of this study holds in using a corrective contemporary group fixed effect when pooling datasets. It is indeed possible that a true difference of efficiency values between the animals of the farms (not due to diet and environmental conditions but real existing difference of level of efficiency) may be hidden by the corrective effect. In the present case, this was unlikely given that the two farms had similar management and performance levels. Nonetheless, in a more general situation where true differences in efficiency level may exist between animals of two farms, it would be necessary to include additional information in the analysis. Genetic information to connect the farms would help distinguish environmental effects from true differences between the animals. Farm specific feed energy values could also be used, either directly or by including random ration-specific coefficients in the regressions of DMI on milk energy and BW, as proposed by Tempelman and Lu (2020).

#### *Comparing residual feed intake estimates from different approaches*

The second objective of this study was to provide the first comparison of RFI results from a linear approach with results from the dynamic method. The correlations between the estimates of the two models were quite high (0.85 and 0.77) and similar or higher to other (all linear) RFI model comparisons. For instance, Knott et al. (2008) compared linear RFI models based on different predictors on growing sheep and found correlations above 0.60 and 0.76 when evaluated at 6 months old and 13 months old, respectively, for the efficiency ranking, and above 0.64 and 0.78 at these two ages for the RFI estimates. This is also similar to the correlation of 0.72 found by Liu et al. (2000) between RFI estimated using residual metabolisable energy consumption calculated from regression analyses and RFI estimated using nutritional prediction equations. It is worth noting that in all cases, even if the correlations are high, they are not equal to one anyway and some reranking is occurring between models which implies that an animal could be considered as more efficient by one model and less efficient by another without any way for the observer to discriminate the two estimations.

It is interesting to observe that the correlation between the linear model in early lactation (1 to 84 DIM) and the dynamic model in early lactation was the same (0.85) as the correlation between the linear model in early lactation and the dynamic model in early + mid-lactation (1 to 168 DIM). In contrast, the correlation of the linear model in early + mid-lactation with the dynamic model in early lactation was considerably lower (0.55). It has been shown that the middle of lactation is the most stable period in terms of RFI (Connor et al., 2019; Martin et al., 2021b). However, the RFI in early lactation is more susceptible to short-term variation, which reflects changes in the physiology of the energy sinks (Prendiville et al., 2011; Martin et al., 2021b). During the first 5–7 weeks of lactation, a substantial increase of milk production

occurs, associated with a pronounced mobilisation of body reserve, and a time-shifted increase of feed intake. This period is associated with an increased risk of health and fertility issues (Esposito et al., 2014). The two RFI approaches (linear and random regression) produce similar estimations when computed on a relatively short period but this similarity decreases when the period is longer and includes multiple lactation stages. The differences in these correlations highlight the key difference between the RFI estimation approaches. One can hypothesise that for the linear approach on the 1 to 168 DIM period, the regression coefficients are largely influenced by the stable mid-lactation period, which masks the changes that occurred in early lactation. In contrast, with the dynamic estimation approach, the regression coefficients are allowed to change at each time point and to follow as much as possible the underlying biological changes. The RFI value reported for the dynamic estimation approach on a given period of time (1 to 84 DIM or 1 to 168 DIM) is in fact the mean of the successive RFIs calculated weekly over the period. Thus, the dynamic estimation process allows a better description of the changing RFI profile through lactation (Martin et al., 2021b).

## Conclusion

We have shown that a dynamic RFI estimation approach, from a multi-trait random regression model, is suitable for use with data combined from different origins, indicating that the contemporary group correction in the model adequately accounts for differences in feeding environments, at least as long as the animals from the different datasets are raised in similar systems with comparable performances levels. This allows RFI estimates to be produced for datasets too small to be analysed on their own. It thus allows a possible use of this model on pooled datasets in a genomic evaluation. In addition, a comparison with a linear RFI approach showed generally high correlations with the dynamic estimation approach. The correlations between linear and dynamic estimation approaches decrease when increasing the time-period used to estimate feed efficiency. This is because only the dynamic model permits the regression coefficients to evolve in line with the physiological changes through the lactation.

## Ethics approval

During the experiment, all animals were kept indoors and handled with care, in line with Aarhus University's ethics policy and INRAE's ethics policy, respectively, in compliance with the European Union legislation for the protection of animals used for scientific purposes. No specific ethical approval was sought for this experiment considering the collected data.

## Data and model availability statement

None of the data were deposited in an official repository. The data that support the study findings are confidential.

## Author ORCIDs

**Pauline Martin:** <https://orcid.org/0000-0003-2950-2728>  
**Vincent Ducrocq:** <https://orcid.org/0000-0002-1503-5199>  
**Amélie Fischer:** <https://orcid.org/0000-0003-4694-822X>

## Author contributions

**P. Martin:** Conceptualization, Formal analysis, Writing – Original Draft.

**V. Ducrocq:** Conceptualization, Methodology, Writing – review & editing.

**Fischer:** Formal analysis, Writing – review & editing.

**N.C. Friggens:** Conceptualization, Writing – review & editing, Project administration, Funding acquisition.

## Declaration of interest

The authors declare no conflicts of interest.

## Acknowledgments

We would like to thank Aarhus University, Center for Quantitative Genetics and Genomics, for the kindly provided data. This study was carried out within GenTORE ([www.gentore.eu](http://www.gentore.eu)), a Horizon 2020 project running from 1 June 2017 to 31 May 2022, as well as from the French national DEFFILAIT project.

## Financial support statement

The research leading to these results has received funding from European Union's Horizon 2020 research and innovation programme - GenTORE - under grant agreement N° 727213, as well as from the French National Agency for Research (ANR-15-CE20-0014) and Apis-Gene for the DEFFILAIT project

## References

- Bazin, S., 1984. Grille de notation de l'état d'engraissement des vaches Pie Noires. Institut technique de l'élevage bovin, Paris, France.
- Berry, D.P., Crowley, J.J., 2013. Cell Biology Symposium: genetics of feed efficiency in dairy and beef cattle. *Journal of Animal Science* 91, 1594–1613.
- Connor, E.E., Hutchison, J.L., Van Tassel, C.P., Cole, J.B., 2019. Defining the optimal period length and stage of growth or lactation to estimate residual feed intake in dairy cows. *Journal of Dairy Science* 102, 6131–6143.
- Esposito, G., Irons, P.C., Webb, E.C., Chapwanya, A., 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. *Animal Reproduction Science* 144, 60–71.
- European Commission 2018. EU Agricultural and Farm Economics Briefs. N°17. EU milk margin index estimate up to 2018. Retrieved on 12 November 2019, from [https://ec.europa.eu/agriculture/sites/agriculture/files/rural-area-economics/briefs/pdf/milk-margin-estimate-2018-no17\\_en.pdf](https://ec.europa.eu/agriculture/sites/agriculture/files/rural-area-economics/briefs/pdf/milk-margin-estimate-2018-no17_en.pdf).
- FAO (Food and Agriculture Organization of the United Nations) 2010. Greenhouse gas emissions from the dairy sector: A life cycle assessment. Retrieved on 30 October 2019, from <http://www.fao.org/3/k7930e/k7930e00.pdf>.
- Fischer, A., Friggens, N.C., Berry, D.P., Faverdin, P., 2018. Isolating the cow-specific part of residual energy intake in lactating dairy cows using random regressions. *Animal* 12, 1396–1404.
- Islam, M.S., Jensen, J., Løvendahl, P., Karlskov-Mortensen, P., Shirali, M., 2020. Bayesian estimation of genetic variance and response to selection on linear or ratio traits of feed efficiency in dairy cattle. *Journal of Dairy Science* 103, 9150–9166.
- Kennedy, B.W., van der Werf, J.H., Meuwissen, T.H., 1993. Genetic and statistical properties of residual feed intake. *Journal of Animal Science* 71, 3239–3250.
- Knott, S.A., Cummins, L.J., Dunshea, F.R., Leury, B.J., 2008. The use of different models for the estimation of residual feed intake (RFI) as a measure of feed efficiency in meat sheep. *Animal Feed Science and Technology* 143, 242–255.
- Koch, R.M., Swiger, L.A., Chambers, D., Gregory, K.E., 1963. Efficiency of Feed Use in Beef Cattle. *Journal of Animal Science* 22, 486–494.
- Kristensen, T., 1986. Method for estimation of body condition of dairy cows. Report No. 615. National Institute of Animal Science, Copenhagen, Denmark.
- Li, B., Berglund, B., Fikse, W.F., Lassen, J., Lidauer, M.H., Mäntysaari, P., Løvendahl, P., 2017. Neglect of lactation stage leads to naive assessment of residual feed intake in dairy cattle. *Journal of Dairy Science* 100, 9076–9084.
- Liu, M.F., Goonewardene, L.A., Bailey, D.R.C., Basarab, J.A., Kemp, R.A., Arthur, P.F., Okine, E.K., Makarechian, M., 2000. A study on the variation of feed efficiency in station tested beef bulls. *Canadian Journal of Animal Science* 80, 435–441.
- Lu, Y., Vandehaar, M.J., Spurlock, D.M., Weigel, K.A., Armentano, L.E., Staples, C.R., Connor, E.E., Wang, Z., Bello, N.M., Tempelman, R.J., 2015. An alternative approach to modeling genetic merit of feed efficiency in dairy cattle. *Journal of Dairy Science* 98, 6535–6551.
- Lu, Y., Vandehaar, M.J., Spurlock, D.M., Weigel, K.A., Armentano, L.E., Staples, C.R., Connor, E.E., Wang, Z., Coffey, M., Veerkamp, R.F., de Haas, Y., Tempelman, R.J., 2017. Modeling genetic and nongenetic variation of feed efficiency and its partial relationships between component traits as a function of management and environmental factors. *Journal of Dairy Science* 100, 412–427.

- Manzanilla Pech, C.I.V., Veerkamp, R.F., Calus, M.P.L., Zom, R., van Knegsel, A., Pryce, J.E., De Haas, Y., 2014. Genetic parameters across lactation for feed intake, fat- and protein-corrected milk, and liveweight in first-parity Holstein cattle. *Journal of Dairy Science* 97, 5851–5862.
- Martin, P., Ducrocq, V., Faverdin, P., Friggens, N.C., 2021a. Invited review: Disentangling residual feed intake—Insights and approaches to make it more fit for purpose in the modern context. *Journal of Dairy Science*.
- Martin, P., Ducrocq, V., Gordo, D.G.M., Friggens, N.C., 2021b. A new method to estimate residual feed intake in dairy cattle using time series data. *Animal* 15, 100101.
- Meyer, K., 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University. Science. B* 8, 815–821.
- Prendiville, R., Pierce, K.M., Delaby, L., Buckley, F., 2011. Animal performance and production efficiencies of Holstein-Friesian, Jersey and Jersey × Holstein-Friesian cows throughout lactation. *Livestock Science* 138, 25–33.
- Strathe, A.B., Mark, T., Nielsen, B., Do, D., Kadarmideen, H.N., Jensen, J., 2014. Deriving genomic breeding values for residual feed intake from covariance functions of random regression models. In: *Proceedings of the 10th World Congress on Genetic Applied to Livestock Production*, 18–22 August 2014, Vancouver, Canada. Abstract 116.
- Taussa, S., Fossaert, C., Cantalapiedra-Hijar, G., Griffon, L., Renand, G., 2020. Genetic parameters of different feed efficiency traits of Charolais young bulls. In: *Book of abstracts of the 71st Annual Meeting of the European Federation of Animal Science (EAAP 2020)*, 1–4 December 2020, Virtual meeting, Wageningen, Netherlands, p.114.
- Tempelman, R.J., Lu, Y., 2020. Symposium review: Genetic relationships between different measures of feed efficiency and the implications for dairy cattle selection indexes. *Journal of Dairy Science* 103, 5327–5345.
- Tempelman, R.J., Spurlock, D.M., Coffey, M., Veerkamp, R.F., Armentano, L.E., Weigel, K.A., de Haas, Y., Staples, C.R., Connor, E.E., Lu, Y., VandeHaar, M.J., 2015. Heterogeneity in genetic and nongenetic variation and energy sink relationships for residual feed intake across research stations and countries. *Journal of Dairy Science* 98, 2013–2026.