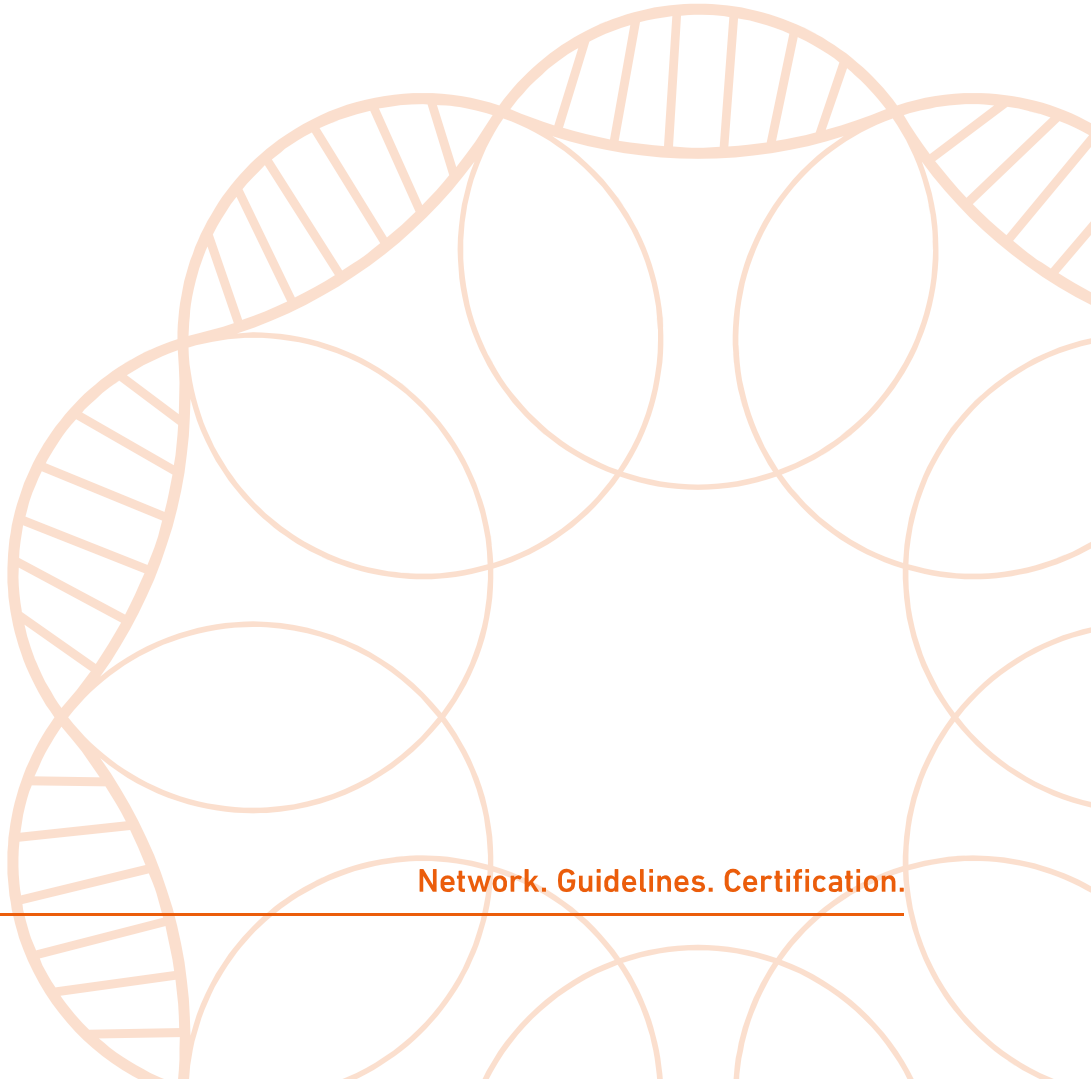
THE GLOBAL STANDARD
FOR LIVESTOCK DATA

Appendix 2 of Section 2 of the ICAR Guidelines – Renewed estimation method for 24-hour fat percentage in AM/PM milk recording scheme

Version June, 2023



Network. Guidelines. Certification.

Table of Contents

Tables

Figures

Change Summary

Date of Change	Nature of Change
June 2023	Creation of document. Update of requirements for milk recording using Automatic Milking Systems (AMS), e.g. definitions, data formats, calculations, plausibility checks and data processing.

1 Renewed estimation method for 24-hour fat percentage in AM/PM milk recording scheme

R.M.G. Roelofs*, G. de Jong and A.P.W. de Roos
NRS, P.O. Box 454, 6800 AL Arnhem, The Netherlands

Paper presented at the ICAR meeting in Kuopio, Finland, 2006.

1.1 Abstract

Based on comments on imprecision of the estimation method for 24-hour fat % in AM/PM milk recording schemes the regression formula was extended and re-estimated. Non-linearity for the existing effects of protein % of the milk sample, interval before sampling, milk amount of sample, milk amount of previous milking and interval before the previous milking was incorporated by using polynomials. Extensions were made by adding the effects of time of sampling, parity and month of sampling as class variables and lactation stage as polynomial. In total a reduction of the standard deviation of the difference between true and estimated 24-hour fat % of 2.4% was reached (0.2856 to 0.2788). The correlation between the two fat %s increased from 0.898 to 0.903, the b-factor of the linear regression between the two fat %s increased from 0.807 to 0.817.

Keywords: estimation, fat %, AM/PM.

1.2 Introduction

The AM/PM milk recording routine is based on only one morning (a.m.) or evening (p.m.) milk sample which are collected in an alternating way. A condition to take part in this AM/PM milk recording in The Netherlands is that on farm electronic milk measurements (EMM) are available. EMM-data consists of time of milking and milk quantity of every milking. Based on one milk sample and the EMM-data the 24-hour fat % is estimated (Peeters and Galesloot, 2002). Also for farms with an automatic milking system (AMS) this estimation is used when only one milk sample is available for analysis on milk composition.

Based on comments from farmers on fluctuations in 24-hour fat % preliminary research was conducted. This showed that the current estimation caused an underestimation of 24-hour fat % based on an a.m.-sample of 0.09% while the estimate based on a p.m.-sample was overestimated by 0.05%. Possible causes for this fluctuation are differences in milk-fat synthesis between day- and night-time as was shown by Gilbert et al. (1972) and Lee and Wardorp (1984). Other factors of imprecision in the current estimation can be caused by lactation stage and parity, two factors that are accounted for in the method of Liu et al. (2000).

The objective of this research is to re-estimate the regression formula which is used to estimate the 24-hour fat %s in AM/PM milk recording and AMS recordings with only one sample. By testing for non-linearity of current effects and introducing new explanatory variables the aim is to increase the accuracy of the estimated 24-hour fat %.

1.3 Material and Methods

The data needed for the objective had to meet a number of criteria. The most important criteria were that the data comprised:

- differences in interval between milking times;
- different milking times;
- multiple samples per cow per herd test date;
- milking time and quantity of all milkings;

Only data of farms that use an AMS met all of these criteria. Therefore the research was conducted on data of all farms that used an AMS from January 20th 2001 until July 1st 2004. Records with only one sample per herd test date were excluded from the analysis.

In order to estimate as well as validate the new regression formula the each herd test date was assigned at random into two separate datasets. Dataset 1 was used for estimation and contained 371.528 samplings on 50.591 cows on 537 farms. Dataset 2 was used for validation and contained 371.885 milkings on 50.643 cows on 538 farms. Some characteristics of variables of both datasets are presented in Table 1.

Table 1. Characteristics of variables in dataset 1 (estimation) and dataset 2 (validation).

Variable	Dataset 1 (estimation)		Dataset 2 (validation)	
	Mean	Std	Mean	Std
Sample milk amount (kg)	10.1	3.1	10.1	3.1
Sample fat (%)	4.40	0.76	4.41	0.76
Sample protein (%)	3.49	0.35	3.49	0.35
Time at sampling	12.29	7.24	12.31	7.24
Interval before sample (min)	520	154	521	155
Interval before prev. milking (min)	526	158	527	159

1.4 Methods

The analysis started with the currently used regression formula which uses the effects: fat %, protein %, milk amount of sampling, interval before sampling, milk amount of the previous milking and interval before the previous milking (Peeters and Galesloot, 2002). All these effects are considered to be linear. As an extra check of the data this regression formula was re-estimated and compared to the currently used regression formula. In order to estimate the regression formula first of all the 24-hour fat % was determined by using a weighted average of all milk samples for that cow on that herd test date.

Subsequently, a number of changes to the regression formula were tested for their effect on the accuracy of the 24-hour fat %. The changes that are tested are:

1. non-linearity of the current effects;
2. effect of time at sampling;
3. effect of lactation stage;
4. effect of parity;
5. month of milk recording;

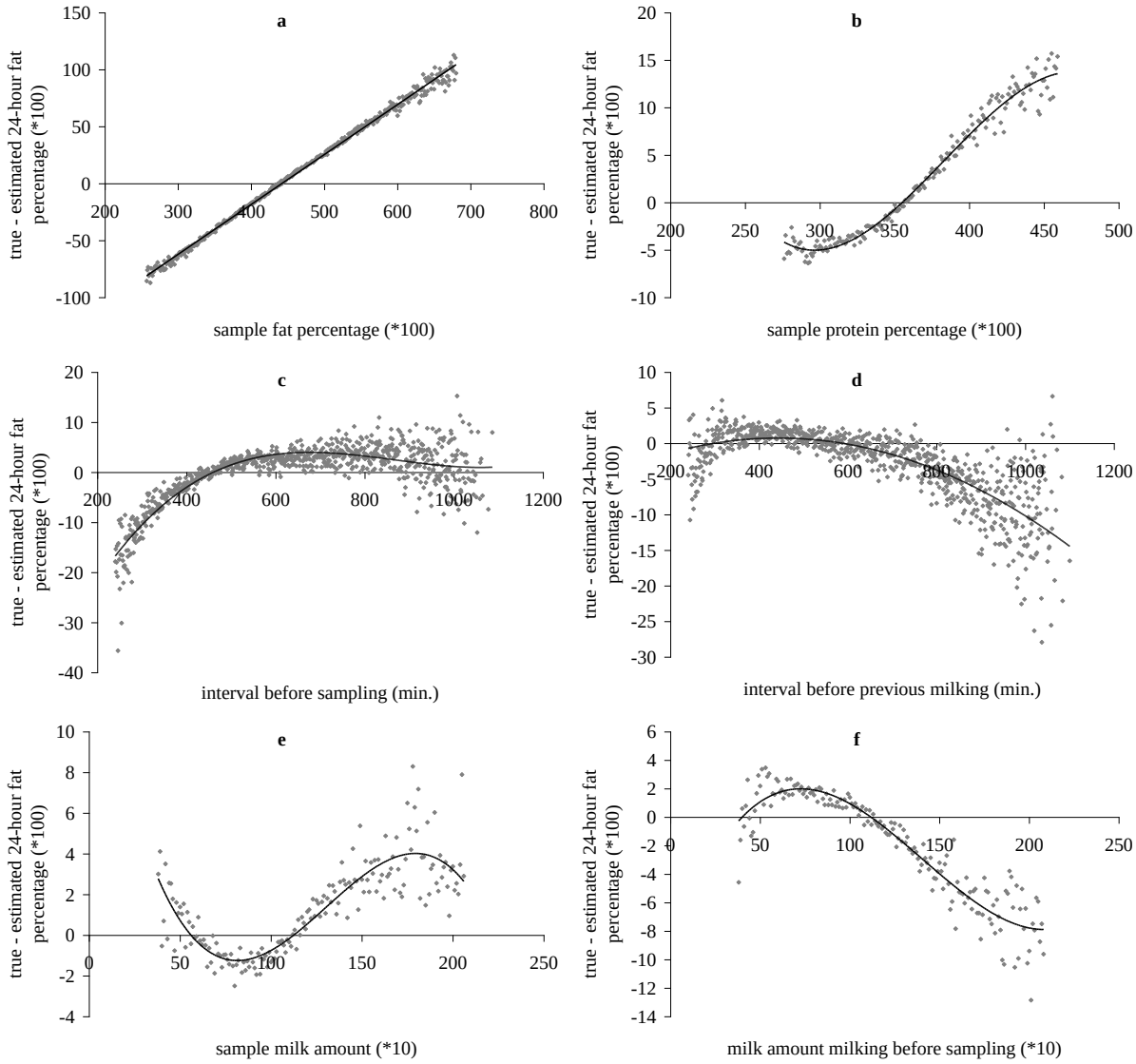
The effects were all tested in a similar way by plotting the residuals of the regression formula without the effect that is tested to the tested effect. Based on this plot a possible relation

between residual and effect becomes clear and the best way of incorporating the effect is shown. The conclusion if an effect had a positive effect on the accuracy of the regression formula was based on the standard deviation of the difference between estimated and true 24-hour fat %. Also the correlation between the two fat %s and the b-factor (regression coefficient) of the linear regression between the two fat %s were considered.

1.5 Results

The regression coefficients of the re-estimated regression formula differed slightly from the estimates by Peeters and Galesloot (2002), probably due to the different dataset.

Figure 1a to 1f show the effect of the variables in the regression formula on the difference



between the true and estimated 24-hour fat %.

Figure 1. Average residual per class for the variables sample fat % (a), sample protein % (b), interval before sampling (c), interval before previous milking (d), sample milk amount (e) and milk amount before sampling (f).

Figure 1a to 1f show the effect of the variables in the regression formula on the difference between the true and estimated 24-hour fat %. Of all variables, only fat % of the milk sample (Figure 1a) seemed to be linear. A 2nd order polynomial fitted the interval before the previous milking. The other variables, i.e. protein % of the milk sample, interval before sampling, milk amount of sample and milk amount of the previous milking were described by a 3rd order polynomial. For all variables except fat % of the sample higher order polynomials were found significant. This however was caused by the large amount of data and no longer a possible biological effect since it also had no effect on the accuracy of the estimation.

The effect of time of sampling showed a large amount of variability over time. Using a polynomial to fit the data was therefore difficult. Estimation of the effect by hourly intervals was a good alternative as is shown in Figure 2. Lactation stage had mainly an effect in the first 50 days of lactation as is shown by Figure 3. A 3rd order polynomial fitted the data properly.

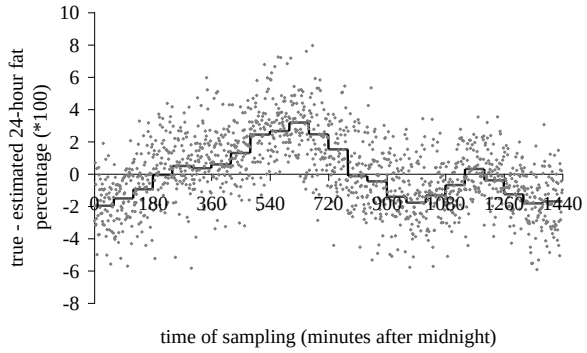


Figure 2. Average residual per class for time of sampling (minutes after midnight).

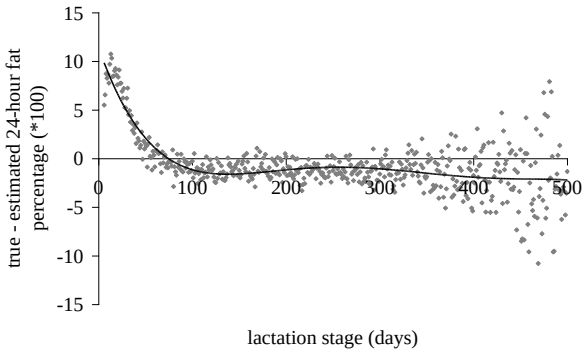


Figure 3. Average residual per class for lactation stage (days).

The effects of parity and month of milk sampling were both considered as class variables. For parity the effects of parity 1 to 6 and 7 or higher were considered. Table 2 shows that mainly for the lower parities the estimated 24-hour fat % was overestimated. Also the months May to October, usually the pasture period, showed an overestimation of 24-hour fat %.

*Table 2. Effect of parity and month of sampling on estimated 24-hour fat % (*100).*

Parity	Estimate	Month of sampling	Estimate
1	-6.58	January	-0.24
2	-3.56	February	-0.28
3	-1.42	June	-0.54
4	-0.48	April	-0.27
5	-0.35	May	-2.07
6	-0.35	June	-3.36
7+	-0.00	July	-4.32
		August	-5.52
		September	-4.74
		October	-2.24
		November	-0.97
		December	-0.00

Table 3. Statistics of the difference between true and estimated 24-hour fat % for six regression formulas (current, re-estimated + five steps), each also including preceding steps.

Regression	Std.	Min	Max	Cor	b-factor
Current, re-estimated	0.2856	-1.840	2.224	0.898	0.807
Non-linearity	0.2820	-1.890	2.198	0.901	0.812
Time of sampling	0.2817	-1.877	2.211	0.901	0.813
Lactation stage	0.2803	-1.883	2.196	0.902	0.814
Parity	0.2794	-1.887	2.179	0.903	0.816
Month of sampling	0.2788	-1.868	2.175	0.903	0.817

Table 3 shows some statistics of the difference between the true and estimated 24-hour fat % based on dataset 2 (validation) of the different regression formulas. Each of the five changes to the regression formula had a (minor) positive effect on either the standard deviation of the difference between the true and estimated 24-hour fat % (Std.), the correlation (Cor) between the two fat %s, the b-factor of the linear regression between the two fat %s or a combination of the these. All changes together reduced the standard deviation with 2.4% from 0.2856 to 0.2788, increased the correlation from 0.898 to 0.903 and increased the b-factor from 0.807 to 0.817.

1.6 Conclusions

The regression formula to estimate the 24-hour fat % based on one milk sample was improved. Improvements were first of all considering non-linearity of the variables by using polynomials for protein % of the milk sample (3rd order), interval before sampling (3rd order), milk amount of sample (3rd order), milk amount of previous milking (3rd order) and interval before the previous milking (2nd order). Secondly, adding the effects of time of sampling (class variable), lactation stage (3rd order polynomial), parity (class variable) and month of sampling (class variable) gave a further reduction of the difference between true and estimated 24-hour fat %. The total reduction in standard deviation of the difference between true and estimated 24-hour fat % is 2.4% (0.2856 to 0.2788). The correlation between the two fat %s increased from 0.898 to 0.903, the b-factor of the linear regression between the two fat %s increased from 0.807 to 0.817.

1.7 References

1. Gilbert, G.R., G.L. Hargrove and M. Kroger, 1972. Diurnal variations in milk yield, fat yield, milk fat % and milk protein % by the test interval method. J. Dairy Sci. 56, 409-410.
2. Lee, A.J. and Wardorp, 1984. Predicting daily milk yield, fat percent, and protein percent from morning or afternoon tests. J. Dairy Sci. 67, 351-360.
3. Liu, Z., R. Reents, F. Reinhardt and K Kuwan, 2000. Approaches to estimating daily yield from single milk testing schemes and use of a.m.-p.m. records in test-day model genetic evaluation in dairy cattle. J. Dairy Sci. 83, 2672-2682.
4. Peeters, R. and P. Galesloot, 2002.Estimating daily fat yield from a single milking on test day for herds with a robotic milking system. J. Dairy Sci. 85, 682-688.