From Calibration of a FTIR Model for Milk Fat Composition to the Estimation of Genetic Parameters

M.J.M. Rutten[*], H. Bovenhuis[*] and J.A.M. van Arendonk[*]

Introduction

Fourier transform infrared (IR) spectroscopy is a suitable method to determine bovine milk fat composition (Soyeurt et al., 2006; Rutten et al., 2009). In this way, fat composition data for a large number of animals can be generated. Hence, IR determined fat composition data could be used by dairy breeding organizations to estimate genetic parameters and breeding values. Genetic parameters, i.e. specifically genetic correlations between observations determined by gas chromatography (GC) and predicted by means of IR, are required to reveal the potential genetic gain that can be achieved from selection on IR predicted fat composition instead of observations determined by GC, which is expensive and time consuming.

Rutten et al. (2009) showed that the number of observations used for calibration of an IR prediction model for fat composition is strongly related to accuracy of prediction. The relation between the number of calibration samples, and therewith accuracy of prediction, and the accuracy of estimated genetic parameters, however, needs to be established. A guideline with respect to the number of observations required for calibration of an IR prediction model for fat composition and its effect on estimated genetic parameters is indispensable for animal breeding organizations if estimation of genetic parameters on IR predicted fat composition is targeted.

Material and methods

Data. In total, we had data available on 1917 milk samples. Milk-fat extraction and the GC procedure were described by Rutten et al. (2009). The individual fatty acids that were included in this study were: C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C18:0 and C18:1*cis*9. Groups of fatty acids included: C6-C12 containing C6:0, C8:0, C10:0, and C12:0; C14-16 containing C14:0 and C16:0; C18u containing 5 identified C18:1 isomers, C18:2 *cis*9,12, C18:3 *cis*9,12,15 and conjugated linoleic acid (CLA i.e. C18:2 *cis*9 *trans*11); and, the ratio of saturated to unsaturated fatty acids.

Fourier Transform IR absorption spectra were recorded using MilkoScan FT 6000 equipment (FOSS, Denmark). FTIR spectra consisted of 1060 IR frequencies (wavenumbers) ranging from 925 to 5008 cm^{-1}. For 1815 milk samples in total, both GC and IR information was available.

[*] Animal Breeding and Genomics Centre, WUR, PO Box 338, 6700 AH, Wageningen, The Netherlands

Infrared prediction model and resampling scheme. To calibrate IR fat composition models, the IR wavenumbers identified by Rutten et al. (2009) for their scenario AA were used. These wavenumbers were selected based on the highest average correlation between fat composition determined by GC and absorption in the IR spectra. In the current study, subsets of observations on milk samples were taken at random, and an IR prediction model for fat composition was calibrated on this subset. Next, the IR model was used to predict the group of fatty acids C18u for all 1817 milk samples. The former procedure was repeated 235 times: first, we sampled 150 subsets of 100 observations for calibration; then, we sampled 50 subsets of 250 observations; then, we sampled 25 subsets of 500 observations; and, finally we sampled 10 subsets of 1000 observations. The resampling scheme thus resulted in 235 different predictions for each milk sample. Subsequently we performed 235 bivariate genetic analyses to estimate genetic correlations between C18u determined by GC and predicted based on IR. C18u represents a trait having high accuracy of prediction (Rutten et al., 2009).

Genetic model. The model used in the genetic analyses was adopted from Stoop et al. (2008). The genetic model accommodated random terms for herd, animal and residual effects. A fixed effect for the general mean, regression coefficients for lactation stage modeled by a Wilmink (1987) lactation curve, regression coefficients for linear and quadratic effects of age at first calving, fixed effects of season of calving (3 classes: 06/2004-08/2004, 09/2004-11/2004 and 12/2004-02/2005) and fixed effects of sire code (3 classes: 1) proven bulls aimed to have at least 200 daughters in the data, 2) test bulls aimed to have at least 20 daughters in the data and 3) other bulls who did not qualify for one of the latter groups) were included. Solutions were generated using ASReml release 2.0. Intraherd heritabilities were calculated as: $h^2_{IH}=\sigma^2_a/(\sigma^2_a+\sigma^2_e)$ i.e. ignoring herd variance. In all genetic analyses, GC parameters were fixed to earlier obtained univariate estimates.

Comparison of alternative selection indices. The objective of the use of the IR prediction model is to replace expensive observations determined by GC. The relative effectiveness of selection indices with either IR- or GC fat composition in dairy cattle breeding programs with performance testing of , say, 100 daughters, can be approximated by the magnitude of the genetic correlation. Hence, in the next section we will focus mainly on the genetic correlations between IR and GC obtained observations.

Results and discussion

Relationship of the number of calibration samples and genetic parameters. In Figure 1, the results from the resampling scheme are plotted for the genetic correlations between C18u determined by GC and predicted using the IR prediction model. When n=100 calibration samples were used, the estimated genetic correlations varied roughly from 0 to 0.99. Validation r-square varied roughly from 0.25 to 0.60. These ranges narrowed down in a stepwise manner when n=250, n=500 and finally when n=1000 calibration samples were used. With the use of n=1000 calibration samples, genetic correlations varied within a range of approximately 0.05. A quadratic regression line forced through the point (1,1) shows that there is a strong relationship between the number of calibration samples and validation r-square on the one hand and genetic correlation on the other hand. The use of all observations (n=1815) for calibration resulted in a validation r-square of 0.74 and a corresponding genetic

correlation of 0.96. This did not translate into a higher genetic correlation and therewith higher potential genetic gain suggesting that there is clearly an optimum for the number of calibration samples. Other fatty acids analyzed in this study showed similar patterns of validation r-square versus genetic correlations.

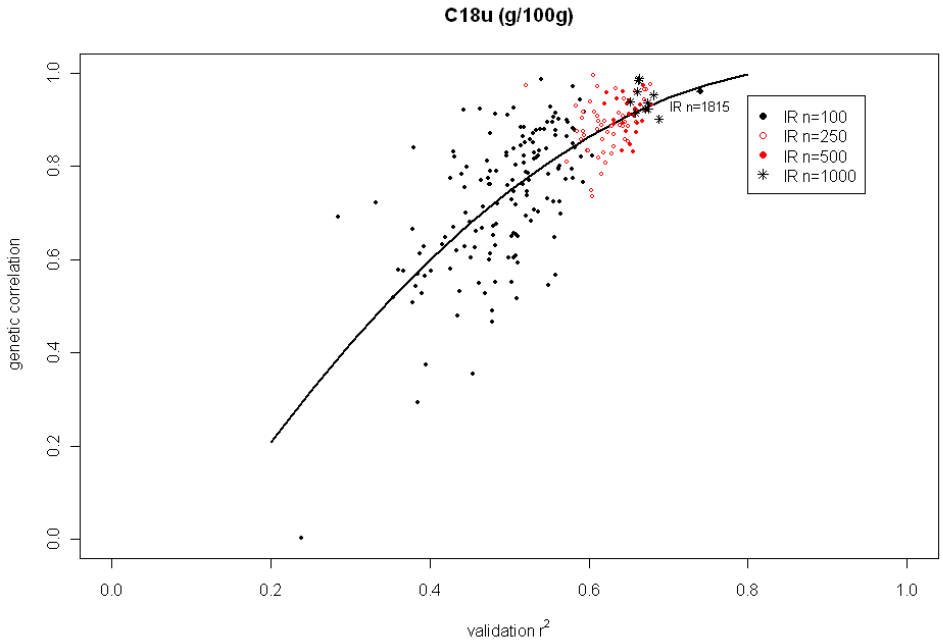

Figure 1: The relationship of validation r-square and the genetic correlation between observations of C18u (g/100g) obtained by GC and by IR prediction models calibrated on different numbers of samples (see legend). The solid line represents a quadratic regression forced trough the point (1,1).

Standard errors. In Figure 1, many estimates of genetic correlations were found at the level of r_A=0.9, regardless of the number of calibration samples that was used. At n=100, the corresponding standard error was approximately s.e.=0.06 (not shown). By looking at Figure 1, it immediately becomes clear that this standard error does not reflect the ranges of genetic correlations that were actually found for the models calibrated on n=100 calibration samples. Thus, there is a discrepancy which is caused by the fact that the standard error for validation r-square is not discounted in the standard error for the genetic correlation. Therefore, particular models resulting in high genetic correlations are not desirable unless the standard error of validation r-square is low. Since standard errors of validation r-square are not commonly provided by statistical software packages, this is identified as a potential pitfall.

Potential genetic gain in fat composition. In Table 1, estimated genetic parameters are presented (here all data was used). Validation r-square varied from 0.43 to 0.77. The heritabilities for predictions of fat composition based on IR spectra were generally either roughly equal or lower than the corresponding GC heritabilities which was according

expectation with the exception of C18:0. Genetic correlations ranged from 0.77 to 0.99 and indicate that the replacement of observations of fat composition by predictions based on IR spectra will yield at least 77% of the potential genetic gain which can be achieved by selection on GC determined data. These results show that the use of IR prediction of milk fat composition does not compromise the achievable selection response from selection on GC values to a large extend. When considering that the routine recording of IR traits is cheap and relatively straightforward, we conclude that IR prediction of milk fat composition provides an excellent means for dairy breeding organizations to cheaply improve the composition of milk fat.

Table 1: Validation r-square (r^2), intra-herd heritabilities for GC (h^2_{GC}) and IR (h^2_{IR}) fatty acids and their genetic correlation (r_a). Fatty acids are expressed as g/100g

Trait	Validation r^2	h^2_{GC}	h^2_{IR}	r_A
Individual fatty acids				
C4:0	0.62	0.42 ±0.09	0.42 ±0.06	0.94 ±0.03
C6:0	0.69	0.46 ±0.10	0.37 ±0.05	0.97 ±0.02
C8:0	0.70	0.61 ±0.11	0.37 ±0.05	0.99 ±0.01
C10:0	0.71	0.71 ±0.12	0.47 ±0.05	0.98 ±0.01
C12:0	0.60	0.63 ±0.11	0.53 ±0.06	0.97 ±0.02
C14:0	0.66	0.59 ±0.11	0.49 ±0.06	0.98 ±0.01
C16:0	0.53	0.43 ±0.11	0.30 ±0.07	0.86 ±0.06
C18:0	0.43	0.23 ±0.07	0.52 ±0.09	0.82 ±0.08
C18:1*cis*9	0.72	0.25 ±0.08	0.26 ±0.05	0.93 ±0.04
Groups of fatty acids				
C6-12	0.77	0.67 ±0.11	0.51 ±0.05	0.99 ±0.01
C14-16	0.57	0.16 ±0.08	0.19 ±0.06	0.77 ±0.12
C18u	0.74	0.26 ±0.08	0.25 ±0.05	0.96 ±0.03
Ratio SFA:UFA	0.77	0.28 ±0.09	0.25 ±0.04	0.99 ±0.02

Conclusion

A strong relationship between the number of calibration samples and validation r-square on the one hand and genetic correlation on the other hand was found. But, only the use of n=1000 calibration samples for the calibration of an IR prediction model for fat composition lead accurate genetic parameters. The replacement of GC fat composition observations by IR predictions potentially yields at least 77% of the genetic gain in the fatty acids analyzed in this study. We conclude that the use of IR prediction models provides an excellent means to the dairy industry to improve milk fat composition.

References

Rutten, M.J.M., Bovenhuis, H., Hettinga, K.A. *et al.* (2009). *J. Dairy Sci.*, 92:6202-6209.

Soyeurt, H., Dardenne, P., Dehareng, F. *et al.* (2006). *J. Dairy Sci.*, 89:3690-3695.