Veerkamp, R., Honerlagen, H., Flossdorf, D., van Breukelen, A., Manzanilla-Pech, C., Bouwman, A. (2026) Fitting many variance components simultaneously using factor analytic models: using the microbiome to predict methane, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23912: https://doi.org/10.31274/wcgalp.23912 *©2026 Veerkamp, et al.* CC BY-NC-ND

Fitting many variance components simultaneously using factor analytic models: using the microbiome to predict methane

R.F. Veerkamp[1], H. Honerlagen[1], D.A.L. Flossdorf[1], A.E. van Breukelen[1], C.I.V. Manzanilla-Pech[1], and A.C. Bouwman[1]

[1]*Wageningen University & Research, Wageningen, The Netherlands; *Corresponding author: roel.veerkamp@wur.nl*

Introduction

Reducing enteric methane emissions from dairy cattle is a key objective in sustainable livestock production. Genetic selection offers a permanent and cumulative solution, but large-scale implementation is constrained by the cost and logistical difficulty of methane recording. Indicator traits that are genetically correlated with methane emission may therefore improve selection accuracy. The rumen microbiome is of particular interest, as microbial composition is biologically linked to methanogenesis.

However, incorporating high-dimensional microbiome data into genetic evaluations remains a challenge. Recent studies have proposed various statistical and machine-learning approaches to exploit microbiome data for methane prediction. Many of these approaches fall outside standard quantitative genetic frameworks used for genetic evaluation. However, standard quantitative methods require estimation of many (co)variance components. The combination of a small dataset with many (microbial) traits is expected to lead to non-positive definite matrices (Hill and Thompson 1978), and therefore poor convergence or software workarounds when estimating variance components. Factor analytical models have been proposed earlier (Meyer 2009), de los Campos and Gianola, (2007) in a Bayesian implementation and by Tyrisevä et al. (2011) for international sire evaluations. In this study, the objective was to estimate variance components treating microbial features as phenotypes in a multitrait model, and factor analytic models were used for dimension reduction, using a particular challenging dataset set with only 645 cows and 31 traits.

Materials & Methods

The dataset consisted of 645 Holstein cows with methane emission records and rumen microbiome samples. For this study, thirty microbial features were selected a priori based on heritability and estimated genetic correlation with methane in a bivariate analysis. All traits were standardized to a phenotypic variance of 1 before the analysis.

Analysis were performed using ASReml (Gilmour 2022) with a multitrait animal model including random effects for herd, the additive genetic effects and the residual. The genetic covariance matrix was modelled using a factor analytic structure $S = LL' + P$, where L contains factor loadings and P is a diagonal matrix of specific variances. This approach substantially reduces the number of parameters to be estimated for the 31 by 31 matrix. To achieve convergence, models of increasing order k (abbreviated as XFA(k)) were fitted sequentially, using !continue to start each model from the previous lower dimension XFA(k-1). ASReml fixed the variance components from the earlier model in the initial iterations. During the iterations ASReml rotated the latent vectors, and fixed redundant variance components to zero. Hence the number of parameters fitted might be lower than the number of parameters modelled. Using the VPREDICT statement the complete matrix

with the 31 by 31 (co)variances between all original traits was obtained. Heritability for each trait was calculated by dividing the genetic variance by the sum of the residual and genetic variance, ignoring the herd variances. The genetic and phenotypic covariance matrices were used to predict methane from the microbiome both phenotypically and genetically, using selection index theory.

Results

Comparing models with reducing parsimoniousness (Table 1) indicated that XFA3 provided the optimal balance between goodness of fit and parsimony (based on Akaike Information Criterion). In total, 634 parameters described the full parameter space of 1023, and the 496 genetic (co)variances were sufficiently described with 107 parameters for the latent factors. Interestingly, when modelling the genetic component as independent variances for each trait (XFA0), already 7 of the variance components were redundant, which indicated that there was some confounding in this small data set.

Table 1 Number of variance components (VC) in the model and finally fitted, and the AIC for models with increasing dimensions for the factor analytical model (XFA). Model 0 includes residual (co)variances only, 0+ adds 31 random herd variances, and XFA0 31 independent genetic variances

XFA	**# VC modelled**	**# VC fitted**	**AIC**
0	496	496	4227.30
0+	527	527	2508.42
XFA0	558	551	2454.67
XFA1	589	579	2382.17
XFA2	620	608	2347.21
XFA3	**651**	**634**	**2341.53**
XFA4	682	662	2358.25
XFA5	713	689	2374.15
XFA6	744	705	2368.44

The heritability estimate for methane was 0.37 (se 0.08) and ranged from 0.04 to 0.77 for the microbial features, with an average of 0.37 (se ranged from 0.03 to 0.08) (Figure 1). Genetic correlations with methane ranged from -0.25 to 0.21, but standard errors were large and around 0.1. Genetic correlations between microbial features were considerably larger and there seemed to be blocks of traits that were associated antagonistically, and also in opposite direction with methane. Lower order XFA gave more extreme estimates of the genetic correlation (Figure 2), whilst XFA3 gave more dimensions for the genetic correlations to vary.

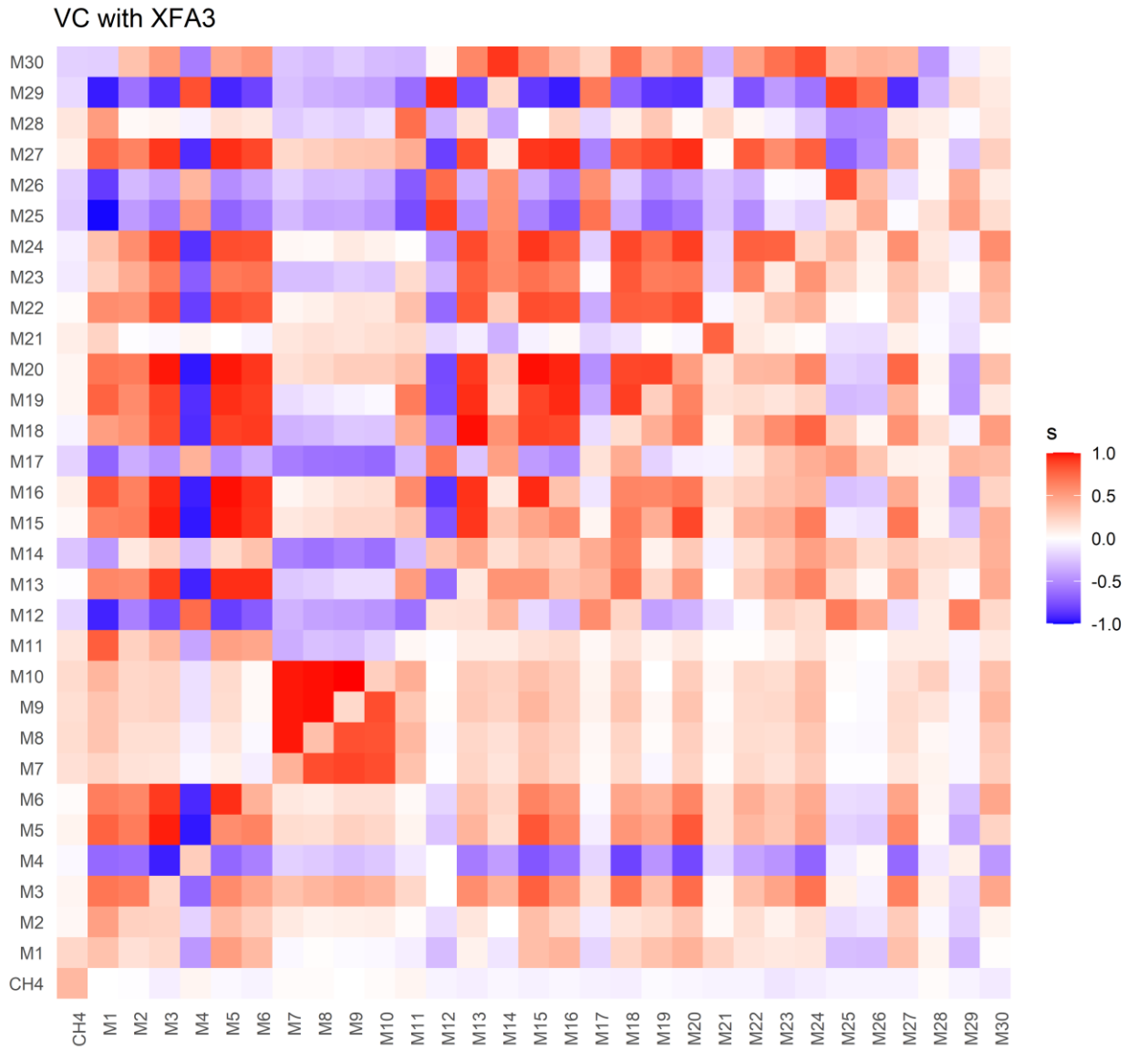

Figure 1 Heatmap of the variance components estimated with the XFA3 model. Heritability on diagonal, genetic correlations above, and phenotypic correlations below diagonal

The lowest eigenvalue of the phenotypic correlation matrix was 0.05, and eleven eigenvalues for the genetic covariance matrix were below 0.01 or even slightly negative, probably due to rounding the values. When phenotypically predicting methane from the microbiome observations (i.e. equivalent to a multiple regression of phenotypic values) the accuracy of prediction was only 0.17, with an R^2 of 0.03. The same values were obtained when predicting the breeding value for an animal, which is lower than the accuracy of 0.61 when having a methane record available on an animal with $h^2 = 0.37$. However, if a very large number of daughters of a bull have the microbiome sampled for these 30 features an accuracy of 0.98 can be achieved.

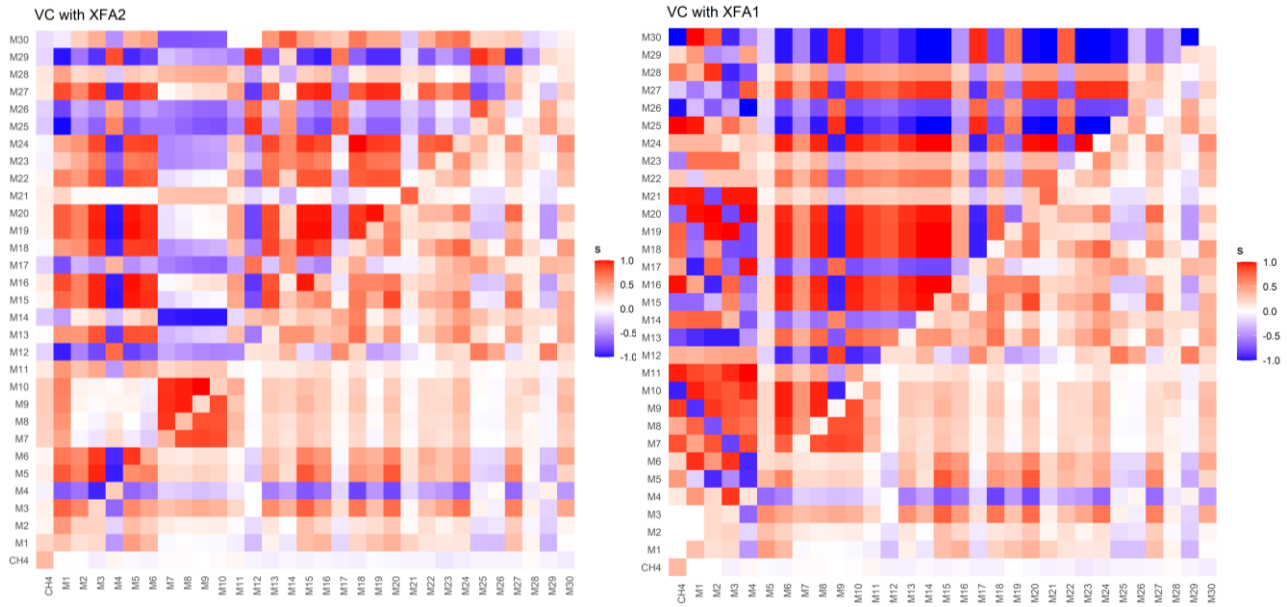

Figure 2 Heatmaps of the variance components estimated with the XFA2 and XFA1models

Discussion

Factor analytic models might be useful and provide a practical solution for incorporating high-dimensional data into routine genetic evaluations. Because XFA allows estimation of the most parsimonious matrix describing a large matrix without prior assumptions. As an example, we used a worst-case scenario data set with 31 traits and only 645 animals with records. Estimating the full matrix without dimension reduction is likely to be impossible, since it is expected to be none positive definite (Hill and Thompson, 1978). With XFA we were able to estimate the full matrix and we expect that with more data we are able to get a more precise estimation of the genetic correlations and probably are capable of fitting a higher order XFA.

Conclusions

Factor analytic models enable estimation of genetic parameters for many traits simultaneously, for example for methane and a large number of microbial traits. The full matrix allows routine quantitative theory to be applied in animal breeding.

References

De Los Campos G., and Gianola D. (2007) Genetics Selection Evolution 39(5)**:** 481–494. https://doi.org/10.1051/gse:20070016

Gilmour A.G., Bj; Cullis, Br; Welham, Sj;Thompson, R. (2022) ASReml User Guide Release 4.2 Functional Specification, pp. VSN International Ltd, Hemel Hempstead, HP2 4TP, UK.

Hill W., and Thompson R. (1978) Biometrics 34(3)**:** 429–439. https://doi.org/10.2307/2530605

Meyer K. (2009) Genetics Selection Evolution 41 https://doi.org/10.1186/1297-9686-41-21

Tyrisevä A., Meyer K., Fikse W., Ducrocq V., Jakobsen J. *et al.* (2011) Genetics Selection Evolution 43 https://doi.org/10.1186/1297-9686-433