Diss. ETH No. 14 223

Genetic evaluation for traits with repeated time-dependent observations (longitudinal data), with an application to feed intake of pigs

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH
for the degree of
Doctor of technical Sciences

presented by
URS SCHNYDER
Dipl. Ing.-Agr. ETH
born 24th September 1971
from Diessbach bei Büren (BE)

accepted on the recommendation of:
Prof. Dr. N. Künzi, supervisor
Dr. K. Meyer, co-referee
Dr. A. Hofer, co-referee

Zürich 2001

Dank

Herrn Prof. Dr. Niklaus Künzi, Institut für Nutztierwissenschaften der ETH Zürich, danke ich für die Überlassung des interessanten Themas, für den mir während der Arbeit gewährten Freiraum sowie für die Übernahme des Referates.

Herrn Dr. Andreas Hofer, Institut für Nutztierwissenschaften der ETH Zürich, danke ich für die gute Betreuung während dieser Arbeit, für die vielen hilfreichen Anregungen und Diskussionen, die sorgfältige Durchsicht der Arbeit und die Übernahme des Korreferates.

I thank Dr. Karin Meyer, Animal Genetics and Breeding Unit, University of New England, Armidale, Australia, for accepting the task of a co-referee, her interest in my work and the many valuable comments that helped to improve this thesis.

Je remercie Dr. Florence Labroue et Laurence Maignel, Institut Technique du Porc, Le Rheu, France, pour leur aide avec les données de comportement alimentaire, les performances et les ascendances des porcs contrôlés en France, qu'elles m'ont mises à disposition pour ce travail.

Ich danke allen Kollegen der Gruppe Tierzucht für das gute Arbeitsklima und die zahlreichen angeregten Diskussionen, insbesondere Regula Schwager, Peter von Rohr, Christian Stricker und Matthias Schelling, mit welchem ich lange Zeit das Büro teilte und der so manches Computerproblem zu lösen wusste.

Ein ganz spezieller Dank gebührt meinen Eltern, die mir stets fördernd und motivierend zur Seite standen und mir das Agronomiestudium ermöglicht haben.

Meiner Frau Katrin danke ich für ihre Liebe und die moralische Unterstützung während der letzten beiden Jahre an dieser Arbeit. Beides war und ist für mich von unschätzbarem Wert.

Den Verantwortlichen des Laur-Fonds der ETH danke ich für den Beitrag an die Druckkosten.

Dieses Projekt wurde durch einen ETH Forschungskredit finanziert.

Table of contents

Table of contents i

Zusammenfassung iv

Summary vii

Chapter 1:

General introduction 1

1. Background 3

2. Purpose and outline of the thesis 5

References 6

Chapter 2:

Random regression and Bayesian inference via Gibbs sampling 9

1. Introduction 11

2. Random regression models and covariance functions 11

2.1. Analysis of longitudinal data in animal breeding 11

2.2. Random regression models 12

2.3. Covariance functions 13

2.4. Recent developments 14

3. Bayesian inference using Markov chain Monte Carlo methods 15

3.1. Bayesian inference 15

3.2. Markov chain Monte Carlo 16

3.3. Bayesian inference and MCMC in animal breeding 18

4. Gibbs sampling for a random regression model 19

4.1. Random regression model 19

4.2. Gibbs sampling 20

4.3. Post-Gibbs analysis 29

References 31

Chapter 3:

Impact of variation of length of individual testing periods on estimation of (co)variance components of a random regression model for feed intake of growing pigs 37

Abstract 38

1. Introduction 39

2. Material and methods 41

2.1. Data simulation 41

2.2. Estimation of (co)variance components 44

3. Results and discussion 47

3.1. Estimates of (co)variance components 47

3.2. Mean squared errors 50

3.3. Effective sample size 52

3.4. General Discussion 54

4. Conclusions 55

References 55

Chapter 4:

Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs 57

Abstract 58

1. Introduction 59

2. Material and methods 60

2.1. Data 60

2.2. Model 62

2.3. Variance components estimation 65

2.4. Post-Gibbs analysis 66

2.5. Eigenfunctions and eigenvalues 68

3. Results and discussion 70

3.1. Behaviour of Gibbs chains 70

3.2. Heritabilities and correlations 74

3.3. Course of variances and heritabilities 78

3.4. Model fit 82

3.5. Genetic eigenfunctions and eigenvalues 84

4. Conclusions **86**

References **87**

Chapter 5:

Multiple trait model combining random regressions for daily feed intake with single measured performance traits of growing pigs **91**

Abstract **92**

1. Introduction **93**

2. Material and methods **94**

2.1. Data 94

2.2. Model 95

2.3. Variance components estimation 100

2.4. Post-Gibbs analysis 101

3. Results and discussion **104**

3.1. Behaviour of the Gibbs sampler 104

3.2. Heritabilities and correlations 109

3.3. Monte Carlo errors 116

4. Conclusions **117**

References **118**

Chapter 6:

General discussion **121**

1. Introduction **123**

2. Choice of data **123**

3. Polynomial random regression model **124**

3.1. Why ordinary quadratic polynomials? 124

3.2. Polynomials in days on test 126

3.3. Problems in feed intake records 127

4. Gibbs sampling for random regression models **128**

5. Feed intake and efficiency of lean growth **130**

6. Final conclusions **133**

References **133**

Zusammenfassung

Genetische Analyse von Merkmalen mit wiederholten, zeitabhängigen Beobachtungen (Longitudinaldaten), angewandt auf den Futterverzehr wachsender Schweine

Das Ziel der vorliegenden Dissertation war die Entwicklung eines Random Regression Modells für die genetische Analyse täglich erhobener Futterverzehrsdaten wachsender Schweine. Es sollte die Frage geklärt werden, ob die Futterverzehrskurve wachsender Schweine durch gezielte Selektion derart verändert werden kann, dass dadurch die Effizienz des Fleischansatzes verbessert wird. Als Regressionsfunktion für die Analyse des täglichen Futterverzehrs wurde ein quadratisches Polynom in Abhängigkeit des Testtages verwendet. Für die Parameterschätzung (Varianzkomponenten) wurde ein Bayes'scher Ansatz unter Verwendung von Gibbs Sampling gewählt.

In einer Simulationsstudie wurde der Einfluss von Unterschieden in der individuellen Mastdauer auf die Schätzung der Varianzkomponenten eines Random Regression Modells untersucht. Da keine Hinweise auf eine Verzerrung der Parameterschätzwerte gefunden wurden, wurden anschliessend reale Futterverzehrsdaten mit ähnlichen Modellen untersucht. Dazu standen uns Daten von Schweinen der Rassen Large White (LW) und Französische Landrasse (FL) zur Verfügung, welche mit Hilfe elektronischer Fütterungsautomaten in drei verschiedenen Mastleistungsprüfstationen in Frankreich erhoben wurden. Die Testperiode begann für alle Tiere bei einem Lebendgewicht von 30 kg und endete für die Eber bei 95 kg Lebendgewicht, während die Kastraten bis 100 kg im Test verblieben und anschliessend geschlachtet wurden. In einer ersten Analyse wurden für zwei Datensätze (je einer pro Rasse: LW und FL) bestehend aus Wochenmittelwerten des täglichen Futterverzehrs (hauptsächlich von Ebern) die Varianzkomponenten für ein Einmerkmals Random Regression Modell geschätzt. Auf Grund der geschätzten Kovarianzmatrizen wurden Heritabilitäten,

genetische und phänotypische Korrelationen von Regressionsparametern, sowie genetische Eigenwerte und Eigenfunktionen berechnet. Anhand genetischer Eigenfunktionen und der dazugehörigen Eigenwerte lässt sich für die untersuchte Population eine Aussage über die genetische Variation des Futterverzehrs im Verlauf der Wachstumsperiode machen. In einer zweiten Analyse wurden sowohl Wochenmittelwerte des täglichen Futterverzehrs, als auch die vier einzeln gemessenen Merkmale durchschnittliche Tageszunahme, Futterverwertung, Magerfleischanteil und Fleischqualitätsindex von Kastraten der Rasse Large White einbezogen. Das Random Regression Modell für den täglichen Futterverzehr wurde mit einem konventionellen Mehrmerkmals Tiermodell kombiniert, um einfach und wiederholt gemessene Merkmale gemeinsam auszuwerten. Mit dieser Methode erhielten wir Schätzwerte für genetische und phänotypische Korrelationen zwischen einfach gemessenen Leistungsmerkmalen und Random Regression Parametern sowie Wochenmittelwerten des täglichen Futterverzehrs.

Die geschätzten Heritabilitäten der Regressionsparameter sowie die genetischen Eigenfunktionen weisen darauf hin, dass es sehr schwierig sein wird, den Verlauf der Futterverzehrskurven durch Selektion zu verändern. Dieser Eindruck wird durch die während der gesamten Wachstumsperiode annähernd konstanten Korrelationen zwischen dem täglichen Futterverzehr und den einzeln gemessenen Leistungsmerkmalen zusätzlich verstärkt. Die beste Möglichkeit zur Steigerung des Futterverzehrs zu Beginn der Wachstumsperiode scheint die Selektion für einen höheren y-Achsenabschnitt (erster Term des quadratischen Polynoms) der Futterverzehrskurve zu bieten. Die Vorteile eines solchen Selektionsschemas gegenüber der Selektion für durchschnittlichen Futterverzehr sind wegen der unvorteilhaften Korrelation zwischen y-Achsenabschnitt und Magerfleischanteil jedoch gering.

Aus der vorliegenden Arbeit können folgende Schlussfolgerungen gezogen werden:

- Random Regression Modelle in Form quadratischer Polynome sind kombiniert mit der Berechnung genetischer Eigenfunktionen und der dazugehörigen Eigenwerte ein geeignetes Werkzeug zur Analyse longitudinaler Daten und zur Abschätzung der in einer Population entlang eines (Zeit-)Verlaufes vorhandenen genetischen Variation.
- Gibbs Sampling für Random Regression Modelle ist wegen der hohen Autokorrelationen zwischen aufeinanderfolgenden Gibbs Samples und der dadurch bedingten hohen Anzahl Iterationsrunden sehr rechenintensiv.
- Es wird sehr schwierig sein, die Futterverzehrskurve wachsender Schweine durch gezielte Selektion derart zu verändern, dass dadurch die Effizienz des Fleischansatzes verbessert wird.

Summary

Genetic evaluation for traits with repeated time-dependent observations (longitudinal data), with an application to feed intake of pigs

The purpose of this thesis was to develop a random regression model for the genetic evaluation of daily feed intake data of growing pigs. The objective was to investigate whether it is possible to change the average feed intake curve by selection to improve efficiency of lean growth. A quadratic polynomial in days on test was chosen as a regression function for the analysis of weekly means of daily feed intake. Bayesian methodology using Gibbs sampling was applied for the estimation of (co)variance components.

In a simulation study, concerns about the validity of a random regression model using quadratic polynomials in days on test in a situation with different length of testing periods were investigated. As no evidence of bias in estimates of (co)variance components was found, similar models were then used to analyse real feed intake data. Daily feed intake data recorded in French central testing stations by means of electronic feed dispensers on group housed French Landrace (FL) and Large White (LW) growing pigs was available for this study. Boars were performance tested between 30 and 95 kg live body weight, while castrated males were on test until they were slaughtered at a live body weight of 100 kg. In a first analysis (co)variance components of a single trait random regression model were estimated for two data sets (one per breed: FL and LW) containing weekly means of daily feed intake (mainly from boars). Heritabilities, genetic and phenotypic correlations of regression parameters were calculated from estimated covariance matrices, as well as genetic eigenvalues and -functions. Genetic eigenfunctions together with their associated eigenvalues give an indication of the amount of genetic variation of feed intake available in a population during the growing period. In a second analysis weekly means of

daily feed intake and the four single measured traits average daily gain on test, feed conversion ratio, carcass lean content and meat quality index of performance tested castrated male Large White pigs were analysed jointly. For this purpose the random regression model for weekly means of daily feed intake was combined with a conventional multiple trait animal model for a joint analysis. This resulted in estimates of genetic and phenotypic correlations between single measured performance traits and random regression parameters as well as weekly means of daily feed intake.

Heritabilities of regression parameters for daily feed intake and genetic eigenfunctions indicate that it will be very difficult to change the shape of average feed intake curves by selection. This impression is supported by correlations between daily feed intake and single measured performance traits, which are almost constant throughout the entire testing period. The best way to improve daily feed intake in the beginning of the testing period might be to select for a higher intercept parameter of feed intake curves. But advantages of such a selection scheme compared to selection for average daily feed intake are limited due to the unfavourable genetic correlation of the intercept parameter with carcass lean content.

The following conclusions can be drawn from this thesis:

- Polynomial random regression models in combination with calculation of eigenfunctions and their associated eigenvalues are a useful tool to analyse longitudinal data and to assess the amount of (genetic) variation available in a population along such a trajectory (e.g. feed intake curve).
- Gibbs sampling for random regression models is very computer intensive because a high number of rounds is needed for reliable estimates, due to high autocorrelations between Gibbs samples.
- Changing the shape of feed intake curves by selection to improve the efficiency of lean growth will be very difficult.

Chapter 1:

General introduction

Seite Leer /
Blank leaf

1. Background

Selection of pigs concentrated for many years on improved leanness and feed efficiency. This was achieved by reducing backfat thickness and by selection for a lower feed conversion ratio, which is the ratio of total feed intake and live weight gain during the fattening period. From 1960 to 1990, this selection regime reduced the percentage of fat in the carcass from 40 % to 20 % and in turn increased the percentage of lean meat in the carcass from about 40 % to 60 % [20]. Rather than just redirecting resources from fat deposition into lean deposition and thus resulting in a higher growth rate, selection for increased leanness and improved feed conversion under ad libitum feeding has lead to a decrease of the feed intake capacity (FIC) [18]. This is confirmed by Cole and Chadd [3], who show that "modern" genotypes of pigs have a lower mean voluntary feed intake and feed intake increases at a lower rate with body weight compared to "older" genotypes. The reason for this reduction is the negative genetic correlation of average daily feed intake with the leanness of the carcass and feed conversion ratio, as shown by Cameron and Curran [2] in a recent selection experiment with British Large White and Landrace pigs under an ad libitum feeding regimen.

In the long run, FIC might become a limiting factor for a further improvement of the efficiency of lean growth. As optimum levels of backfat thickness are or will soon be reached, further improvement of feed conversion by reduction of the rate of fat deposition is limited and thus other routes to improve feed efficiency have to be found [9, 11, 17].

Based on the relationship between FIC and optimum level of feed intake, which is realised when lean deposition rate is at its maximum and fat deposition rate at its minimum for the given lean deposition rate [6], de Vries and Kanis [7] suggested to divide the growing period of pigs into 3 phases:

1) early fattening period where FIC of pigs is determined by mechanical constraints and FIC is less than the optimum level of feed intake (FI(opt)),
2) intermediate fattening period where FIC is still determined by mechanical constraints but FIC > FI(opt),
3) late fattening period where FIC is determined by metabolic constraints with FIC > FI(opt).

Increasing FIC in period 1 to its optimum level should increase growth rate without affecting the leanness of the carcass, while increasing FIC in periods 2 or 3 would lead to fatter carcasses. Increasing FIC in period 1 while keeping FIC in periods 2 and 3 constant should lead to animals growing more efficiently. Webb [18, 19] supports this view and stresses the need of further research on genetic and environmental effects on the shape of feed intake curves.

Electronic feeders installed in central testing stations allow for the measurement of individual daily feed intake of performance tested growing pigs housed in group pens. This form of performance testing presumably reflects the situation in commercial housing systems for growing fattening pigs better than previously used individual penning of pigs in performance testing stations. Data from such electronic feeding stations were first described in a Dutch study [4, 5] and a little later in French [12, 13] and German [15, 16] studies. In these studies, individual daily feed intake records were averaged over parts of the testing period and the resulting traits were evaluated in a multivariate analysis. This approach was also used in a recent study by Hall *et al.* [10]. Alternatively, appropriate functions can be fitted to individual daily feed intake records, and the resulting feed intake curve parameters for each tested animal can be put in a multivariate analysis, as recently shown by Eissen [8]. Instead of this two step approach, inferences on feed intake curve parameters can be drawn directly from daily feed intake records by means of a random regression model [14]. Andersen and Pedersen [1] used a random regression model to show differences in feed intake and growth curves between two sexual types of growing fattening pigs.

2. Purpose and outline of the thesis

The purpose of this study is the development of a random regression model for the genetic evaluation of daily feed intake data of growing pigs (or other traits with repeated time-dependent observations). Genetic variation in feed intake curves and the relationship between feed intake curve parameters or feed intake in different growing periods and carcass traits will be analysed to investigate possible routes for future improvement of the efficiency of high quality pork production.

After the general introduction in this first chapter, chapter 2 gives an introduction to the methodology used in this thesis. The history of the development of random regression models and covariance functions is reviewed. Development and principles of Markov chain Monte Carlo methods (Metropolis-Hastings algorithm, Gibbs sampler) are outlined and the necessary steps to implement the Gibbs sampler for a random regression model are presented. In chapter 3, a simulation study investigates the impact of variation of length of individual testing periods on estimates of (co)variance components of a random regression model for feed intake of growing pigs. Chapter 4 presents an analysis of feed intake data from performance tested French Landrace and Large White growing pigs, focusing on variation in feed intake curves. Chapter 5 gives a multivariate extension of the model used in chapter 4 and evaluates the relationship between feed intake curve parameters and other performance traits. Finally, the findings of the above chapters are summed and discussed in chapter 6.

References

[1] Andersen S., Pedersen B., Growth and food intake curves for group housed gilts and castrated male pigs, Anim. Sci. 63 (1996) 457-464.

[2] Cameron N.D., Curran M.K., Selection for components of efficient lean growth rate in pigs. 4. Genetic and phenotypic parameter estimates and correlated responses in performance test traits with ad-libitum feeding, Anim. Prod. 59 (1994) 281-291.

[3] Cole D.J.A., Chadd S.A., Voluntary food intake of growing pigs, in: Forbes J.M., Varley M.A., Laurence T.L.J. (Eds.), The voluntary food intake of pigs, Occasional Publication of the British Society of Animal Production 13, 1989, pp. 61-70.

[4] De Haer L.C.M., Relevance of eating pattern for selection of growing pigs, doctoral thesis, Research Institute for Animal Production (IVO-DLO) 'Schoonoord', P.O. Box 501, 3700 AM Zeist, The Netherlands, 1992.

[5] De Haer L.C.M., de Vries A.G., Effects of genotype and sex on the feed intake pattern of group housed growing pigs, Livest. Prod. Sci. 36 (1993) 223-232.

[6] De Vries A.G., Kanis E., A growth model to estimate economic values for food intake capacity in pigs, Anim. Prod. 55 (1992) 241-246.

[7] De Vries A.G., Kanis E., Swine breeding goals, short- and long-term considerations, in: Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada, August 7-12, vol. 17, 1994, pp. 390-397.

[8] Eissen J.J., Breeding for feed intake capacity in pigs, doctoral thesis, Animal Breeding and Genetics Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands, 2000.

[9] Fowler V.R., Biological advances towards genetic improvement in pigs, in: Proc. 3rd World Congress on Genetics Applied to Livestock Production,

Lincoln, Nebraska, USA, July 16-22, vol. XI, Genetics of Reproduction, Lactation, Growth, Adaptation, Disease, and Parasite Resistance, 1986, pp. 345-354.

[10] Hall A.D., Hill W.G., Bampton P.R., Webb A.J., Genetic and phenotypic parameter estimates for feeding pattern and performance test traits in pigs, Anim. Sci. 68 (1999) 43-48.

[11] Kanis E., de Vries A.G., Optimization of selection for food intake capacity in pigs, Anim. Prod. 55 (1992) 247-255.

[12] Labroue F., Aspects génétiques du comportement alimentaire chez le porc en croissance, Thèse de Doctorat, Ensar, Rennes, 1996.

[13] Labroue F., Gueblez R., Sellier P., Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace pigs, Genet. Sel. Evol. 29 (1997) 451-468.

[14] Schaeffer L.R., Dekkers J.C.M., Random regression in animal models for test-day production in dairy cattle, in: Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada, 7-12 August, vol. 18, 1994, pp. 443-446.

[15] Von Felde A., Genetische Analyse der Futteraufnahme-Informationen von Jungebern aus Gruppenprüfung mit automatischen Fütterungsanlagen, Schriftenreihe des Institutes für Tierzucht und Tierhaltung der Christian-Albrechts-Universität zu Kiel, Heft 90, Dissertation, Kiel, 1996.

[16] Von Felde A., Roehe R., Looft H., Kalm E., Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars, Livest. Prod. Sci. 47 (1996) 11-22.

[17] Webb A.J., Genetic control of growth, composition, appetite and feed utilisation: non-ruminants, Options for genetic change, in: Proc. 3rd World Congress on Genetics Applied to Livestock Production, Lincoln, Nebraska, USA,

July 16-22, vol. XI, Genetics of Reproduction, Lactation, Growth, Adaptation, Disease, and Parasite Resistance, 1986, pp. 337-344.

[18] Webb A.J., Future challenges in pig genetics, Animal Breeding Abstracts 63 (1995) 731-736.

[19] Webb A.J., Objectives and strategies in pig improvement: an applied perspective, J. Dairy Sci. 81, Suppl. 2 (1998) 36-46.

[20] Whittemore C.T., The Science and practice of pig production, 1st ed., Longman, London, 1993.

Chapter 2:

Random regression and Bayesian inference via Gibbs sampling


Seite Leer /
Blank leaf

1. Introduction

In recent years, random regression models have become very popular among animal breeders and classical quantitative geneticists for the analysis of longitudinal data. Possible traits include test day records for milk yield, as well as growth and feed intake of farm animals. Likelihood based inference using restricted maximum likelihood (REML) has been considered the method of choice for estimating (co)variance components in mixed linear models [21]. Recently, Bayesian methods of inference on model parameters gained popularity among quantitative geneticists. They have become feasible even for complicated models through the availability of numerical integration techniques such as the Metropolis-Hastings algorithm [19, 35] and its special case, the Gibbs sampler [10], two Markov chain Monte Carlo (MCMC) methods.

This chapter reviews the development of random regression models and covariance functions, as well as Bayesian inference using MCMC methods. Bayesian inference for a random regression model using Gibbs sampling is discussed in detail.

2. Random regression models and covariance functions

2.1. Analysis of longitudinal data in animal breeding

Random regression models and covariance functions are two independently developed approaches to analyse longitudinal data, i.e. traits repeatedly measured along some trajectory, e.g. time. Before covariance functions and random regression models were introduced into animal breeding, less adequate methods were used for the analysis of longitudinal data. Repeated measurements of a trait at different points in time were often aggregated (e.g. sums, averages) within time periods (or other categories of a continuous covariable) and then analysed fitting a multivariate model assuming that aggregated measurements of different time periods are different traits. This results in the need for estimating covariance

matrices of dimension equal to the number of time periods considered, which might contain more parameters than necessary to adequately describe the covariance structure. On the other hand, if very long time periods are chosen (e.g. a whole lactation), nothing can be said on the development of the trait in time, i.e. the covariance structure is not reproduced accurately. Simple repeatability models assume that differences between repeated measurements on the same subject (at different points in time) are only due to measurement errors, which are often assumed independent. A test day model using fixed regressions only [39], is basically just an extension of this model, as differences between measurements on a given subject at different points in time are assumed to be levelled out by the (same) fixed regression function for all measured subjects. Other (possibly more appropriate) models have been used for the analysis of repeated measurements in other fields of applied statistics (see e.g. [32]). Littell *et al.* [33] give a tutorial introduction to parametric modelling of covariance structure for repeated measures data in a generalised linear mixed model context.

2.2. Random regression models

In 1962, Elston and Grizzle [5] first distinguished population (fixed) and individual (random) regression coefficients, assuming that each individual regression coefficient is an independently normally distributed random variable. In 1982, Henderson Jr. [20] described the concept of random regression coefficients in a linear mixed model context, mentioning the possibility for non-null covariances between random regression coefficients. Three months later, Laird and Ware [31] independently described a general formulation of the linear model based on Harville [18], which includes growth models as well as repeated measures models as special cases.

In 1994, Schaeffer and Dekkers [42] used the concept of Henderson Jr. [20] to extend the fixed regression test day model (TDM) previously developed by Ptak and Schaeffer [39] into a random regression model for the evaluation of test day production of dairy cows. Andersen and Pedersen [1] used a polynomial random

regression model to describe cumulated feed intake and weight gain of growing pigs. They estimated sex-differences in feed intake and growth curves between gilts and castrated male pigs, as well as phenotypic (co)variances among random regression parameters. In 1997, Jamrozik and Schaeffer [23] and Jamrozik *et al.* [24] presented parameter estimates obtained by Gibbs sampling and a genetic evaluation using a random regression TDM for production traits of dairy cows. They showed how solutions of their random regression model can be used for selection by calculating 305-day equivalents or different measures of persistency from random regression coefficients.

Random regression models are also called random coefficient models and have found applications in other fields of applied statistics, e.g. in econometrics. Longford [34] describes random coefficient models and gives various examples of their application.

2.3. Covariance functions

Covariance functions were introduced in 1989 by Kirkpatrick and Heckman [27] together with the concept of infinite dimensional characters and their eigenfunctions. Infinite dimensional characters are traits where the phenotype and genotype of an individual is described by a function of a continuous (co)variable (e.g. time, age or environmental conditions), rather than by a finite set of measurements. Examples of such infinite dimensional characters are growth, morphological shape and reaction norms [27]. Eigenfunctions of covariance functions are the infinite dimensional equivalent to eigenvectors of covariance matrices. Together with their associated eigenvalues they describe the kind and amount of variation of a trait along some trajectory. In 1990, Kirkpatrick *et al.* [29] applied the methods described by Kirkpatrick and Heckman [27] to growth of mice, using orthogonal Legendre-polynomials to describe the shape of growth curves. They fitted full and reduced order covariance functions to the previously estimated additive genetic covariance matrix between measurements of body weight at different ages. Subsequent

papers by Gomulkiewicz and Kirkpatrick [16] and Kirkpatrick and Lofsvold [28] dealt with reaction norms [16] and constraints imposed by missing additive genetic variation on possible selection responses in the evolution of growth trajectories [28]. In 1994, Kirkpatrick *et al.* [30] presented an alternative method to fit covariance functions to phenotypic and additive genetic covariance matrices using asymmetric coefficients. They applied this method to daily lactation records of dairy cows and showed how to correct for inflation of estimates of phenotypic variances due to measurement errors by extrapolating from covariances to the diagonal. In 1997, Meyer and Hill [37] showed how covariance functions can be fitted to any source of variation, genetic or permanent environmental, directly from the data by REML.

Covariance functions are also used in other fields of applied statistics, e.g. spatial statistics (geostatistics), astrophysics and electrical engineering.

2.4. Recent developments

The equivalence of random regression models and covariance functions was shown by Meyer [36] in 1998, when she estimated full and reduced rank covariance functions directly from the data using REML and a random regression model. Van der Werf *et al.* [46] also showed this equivalence, when they compared the two step approach of Kirkpatrick *et al.* [29] to fitting covariance functions directly from the data by REML using a random regression model transformed to canonical scale. Since then, many other applications of random regression models were presented, many of them for test day records of production traits of dairy cows.

In recent years other approaches were suggested for the analysis of longitudinal data in animal breeding. White *et al.* [49] fitted cubic smoothing splines to lactation curves of dairy cows, which is a semiparametric approach as opposed to the parametric modelling of the trait in a random regression model. Pletcher and Geyer [38] developed the so called character process model for the analysis of

function-valued traits, which uses parametric functions to model the covariance structure, while the approach of Kirkpatrick and Heckman [27] is nonparametric (linear combinations of orthogonal polynomials). By the term "function-valued traits" (or characters) Pletcher and Geyer [38] mean the same as Kirkpatrick and Heckman [27] with "infinite dimensional traits". Jaffrézic and Pletcher [22] included a residual variance in the character process model and compared this approach with orthogonal polynomial covariance functions and random regression models in their ability to fit different covariance structures.

Which method will be used in future practical animal breeding applications for the analysis of longitudinal data, will largely depend on the type of data at hand (covariance structure) and the capability of each method to handle large data sets and multiple traits [22]. Such practical difficulties (e.g. limited computing resources) may also have been one main reason for using simpler but less appropriate methods for the analysis of repeated measures data in the past.

3. Bayesian inference using Markov chain Monte Carlo methods

3.1. Bayesian inference

Consider the observed data y and some parameter θ of the statistical model chosen to describe this data. From a Bayesian perspective, there is no fundamental distinction between observable data and model parameters, both are considered random variables [15]. The process of Bayesian data analysis starts by setting up a full probability model, which means specifying the joint distribution of observed data and unknown model parameters [9]. This joint distribution $p(y, \theta)$ can be written in terms of the prior distribution of the parameter $p(\theta)$ and the likelihood $p(y|\theta)$, or equivalently in terms of the posterior distribution of the parameter given the data $p(\theta|y)$ and the prior distribution of the data $p(y)$:

$$p(y,\theta) = p(\theta)p(y \mid \theta) = p(y)p(\theta \mid y)\,. \tag{1}$$

This is known as Bayes' theorem and yields the formula for the posterior distribution of the parameter given the data, which is the object of all Bayesian inference:

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{p(y)} = \frac{p(\theta)p(y \mid \theta)}{\int p(\theta)p(y \mid \theta)d\theta} \tag{2}$$

Any characteristic of the posterior distribution can be used for Bayesian inference on the model parameters θ. All these characteristics can be expressed in terms of posterior expectations of functions $f(\theta)$ of the parameters [15]:

$$E[f(\theta) \mid y] = \frac{\int f(\theta)p(\theta)p(y \mid \theta)d\theta}{\int p(\theta)p(y \mid \theta)d\theta} \tag{3}$$

For most applications, the integrals in (3) can not be evaluated analytically. One possible alternative is Monte Carlo integration, including Markov chain Monte Carlo.

The integral in the denominator of (2) and (3) needs not to be known, as the factor $p(y)$ in (2), which does not depend on θ, can be considered a constant if the data y is fixed, yielding the unnormalised posterior density, which is the right hand side of (4):

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta) \tag{4}$$

Inferences on the model parameters θ can also be drawn if the posterior density is only known up to the normalising constant.

3.2. Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods combine Monte Carlo integration with cleverly constructed Markov chains to sample from the required distribution $\pi(.)$ [15]. Monte Carlo integration evaluates $E[f(X)]$ by drawing samples $\{X_t,\ t = 1,\ldots,\ n\}$ from $\pi(.)$ and then approximating

$$E[f(X)] \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t). \quad (5)$$

If the samples $\{X_t\}$ are independent, by laws of large numbers, any desired accuracy of this approximation can be reached by increasing the sample size n. As $\pi(.)$ is often non-standard, drawing independent samples may not be possible. In this case, (dependent) samples can be drawn using a Markov chain with $\pi(.)$ as its stationary distribution, which is Markov chain Monte Carlo. A Markov chain is a sequence of random variables $\{X_0, X_1, X_2, ...\}$, where at each time $t \geq 0$, the next state X_{t+1} only depends on the current state X_t and is reached with transition probability $p(X_{t+1}| X_t)$. If the chain is run long enough, it will gradually forget its initial state and eventually converge to a unique stationary or invariant distribution. The time required to converge is called the burn-in period. After the burn-in, the chain will yield dependent samples from its stationary distribution, on which inference will be based. The main issue of MCMC is how to construct a Markov chain (and its transition kernel) such that its stationary distribution is precisely our distribution of interest [15].

MCMC is based on the Metropolis-Hastings algorithm [19, 35] and all other possible ways of constructing these chains, including the Gibbs sampler [10], are special cases of this algorithm [15]. The Metropolis-Hastings algorithm, as its name suggests, has been developed by Metropolis *et al.* [35] in 1953, and was originally applied to equations of (energy) state of interacting individual molecules such as liquids. Subsequently, it has been extensively used in statistical mechanics [19]. In 1970, Hastings [19] presented a generalisation of the original Metropolis algorithm and showed its potential for applications to numerical problems arising in statistics. Despite of this, it lasted another twenty years until it was used in a broader range of statistical applications. Without reference to Hastings' [19] results, Geman and Geman [10] presented the Gibbs sampler, an algorithm based on the method of Metropolis *et al.* [35] for sampling of Gibbs distributions in Bayesian image restoration. In 1990, Gelfand and Smith

[6] showed the relationship between the Gibbs sampler and the method of data augmentation for the calculation of posterior distributions [45] and revealed the potential of the Gibbs sampler for applications in a wide variety of statistical models. In the same year, Gelfand *et al.* [7] illustrated Bayesian inference in normal data models using Gibbs sampling. These papers generated interest in Gibbs sampling among many statisticians and helped to spread MCMC methods. Gelman [8] later showed, that the Gibbs sampler is a special case of the Metropolis-Hastings algorithm.

In the Metropolis-Hastings algorithm, at time t, the next state X_{t+1} is chosen by first sampling a candidate point Y from a proposal distribution $q(.|X_t)$, which is then accepted with probability $\alpha(X_t, Y)$ where

$$\alpha(X,Y) = \min\left(1, \frac{\pi(Y)q(X \mid Y)}{\pi(X)q(Y \mid X)}\right) \tag{6}$$

If the candidate point is accepted, it becomes the next state of the chain ($X_{t+1}=Y$), otherwise the chain does not move ($X_{t+1}=X_t$). The Gibbs sampler is a special case of the Metropolis-Hastings algorithm where samples are drawn from full conditional distributions, which result in an acceptance probability (6) of $\alpha=1$, i.e. candidate points are always accepted.

Casella and George [2] give a simple exposition of the Gibbs sampler. A tutorial introduction to the Metropolis-Hastings algorithm is given by Chib and Greenberg [3].

3.3. Bayesian inference and MCMC in animal breeding

Bayesian inference in animal breeding theory has been discussed in 1986 by Gianola and Fernando [12]. In 1990, Gianola and Foulley [13] presented a method to estimate variance components in a univariate mixed linear model based on exact or approximate posterior distributions. In 1991, Guo and Thompson [17] used the Gibbs sampler jointly with the EM-algorithm for estimating variance component models for large complex pedigrees in human

genetics. Gilks *et al.* [14] describe how to use the Gibbs sampler with a random-effects model for longitudinal data in medicine (long term response to hepatitis B vaccination). But only after the first applications of the Gibbs sampler in an animal breeding context were presented in 1993 and 1994 by Wang *et al.* [47, 48], who estimated variance components for a univariate mixed linear model, Bayesian inference was considered an alternative to REML for practical applications. Subsequently, Bayesian inference on variance components using Gibbs sampling was described by Jensen *et al.* [25], for a model with direct and maternal genetic effects, and Sorensen *et al.* [43], for a threshold model. Jamrozik and Schaeffer [23] adapted the Gibbs sampling algorithms for the maternal and direct genetic effects model of Jensen *et al.* [25] to a random regression test day model for yield traits of dairy cows. Rekaya *et al.* [41] use Bayesian inference and Gibbs sampling to compare different models (multitrait, repeatability and random regression) for the genetic evaluation of test day production of dairy cows.

4. Gibbs sampling for a random regression model

4.1. Random regression model

In Matrix notation, a simple random regression model can be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wp} + \boldsymbol{\varepsilon} \tag{7}$$

where $\mathbf{y}$ is a vector of n observations along a trajectory (e.g. time); $\mathbf{b}$ is a vector of fixed effects for regression parameters; $\mathbf{a}$ and $\mathbf{p}$ are vectors containing random additive genetic and permanent environmental regression coefficients, respectively; $\boldsymbol{\varepsilon}$ is a vector of n residuals and $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$ are incidence matrices containing regression covariables for each observation.

Expected value and variance of $\mathbf{y}$ are given by:

$$\begin{aligned} \mathrm{E}(\mathbf{y}) &= \mathbf{Xb} \\ \mathrm{Var}(\mathbf{y}) &= \mathbf{Z}(\mathbf{A} \otimes \mathbf{G}_0)\mathbf{Z}' + \mathbf{W}(\mathbf{I}_l \otimes \mathbf{P}_0)\mathbf{W}' + \mathbf{I}_n\sigma_\varepsilon^2 \end{aligned} \tag{8}$$

where $\mathbf{G_0}$ and $\mathbf{P_0}$ are covariance matrices for random additive genetic and permanent environmental regression coefficients; $\mathbf{A}$ is the numerator relationship matrix between the m animals in the pedigree; $\mathbf{I_l}$ and $\mathbf{I_n}$ are identity matrices of dimensions equal to the number of permanent environmental effects (l animals with records) and total number of records (n), respectively; σ_ε^2 is the residual variance and $\otimes$ denotes the Kronecker- or direct matrix product (see e.g. Searle [44]).

The mixed model equations (MME) for this model can be written as:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}W} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{A^{-1}} \otimes \mathbf{G_0^{-1}} & \mathbf{Z'R^{-1}W} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}Z} & \mathbf{W'R^{-1}W} + \mathbf{I_l} \otimes \mathbf{P_0^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \\ \mathbf{W'R^{-1}y} \end{bmatrix}$$

where: $$\mathbf{R} = \mathbf{I_n}\sigma_\varepsilon^2 \tag{9}$$

Coefficient matrix and right hand side of the MME can be multiplied by the residual variance and written as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \left(\mathbf{A^{-1}} \otimes \mathbf{G_0^{-1}}\right)\sigma_\varepsilon^2 & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \left(\mathbf{I_l} \otimes \mathbf{P_0^{-1}}\right)\sigma_\varepsilon^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix}$$

or shorter: $$\mathbf{C}\hat{\boldsymbol{\theta}} = \mathbf{r} \tag{10}$$

This works only with a single trait random regression model assuming the same residual variance for all records throughout the trajectory (independent of time).

4.2. Gibbs sampling

For the implementation of the Gibbs sampler we begin with the specifications of the distributional assumptions about the data and the model parameters. For convenience and simplicity of notation we assume in the following derivation that the same k regressions are fitted for all fixed and random effects.

The conditional distribution of the data is assumed to be normal:

$$\mathbf{y} \mid \mathbf{b},\mathbf{a},\mathbf{p},\sigma_\varepsilon^2 \sim N(\mathbf{Xb}+\mathbf{Za}+\mathbf{Wp},\mathbf{I}_n\sigma_\varepsilon^2) \quad (11)$$

As there are no fixed effects in a Bayesian analysis, the prior distribution for classical "fixed" effects is usually assumed to be proportional to a constant, reflecting the lack of prior knowledge about their distribution:

$$\mathbf{b} \sim \text{constant}$$

Conditional distributions of additive genetic and permanent environmental effects are assumed to be multivariate normal:

$$\mathbf{a} \mid \mathbf{A},\mathbf{G}_0 \sim N(\mathbf{0},(\mathbf{A} \otimes \mathbf{G}_0))$$
$$\mathbf{p} \mid \mathbf{P}_0 \sim N(\mathbf{0},(\mathbf{I}_l \otimes \mathbf{P}_0))$$

Variance parameters in univariate normal models follow scaled inverted chi-square ($Inv\text{-}\chi^2$) distributions [9], which is therefore a natural choice of a (proper) prior distribution for the residual variance:

$$\sigma_\varepsilon^2 \mid \nu_\varepsilon, S_\varepsilon^2 \sim Inv-\chi^2(\nu_\varepsilon, S_\varepsilon^2)$$

where hyperparameters ν_ε and S_ε^2 denote the prior degrees of freedom and scale parameter of the scaled inverted chi-square distribution, respectively.

Similarly, k-dimensional inverse Wishart (IW) distributions are assumed for the (k by k) covariance matrices $\mathbf{G_0}$ and $\mathbf{P_0}$:

$$\mathbf{G}_0 \mid \mathbf{V}_\mathbf{G}, \nu_G \sim IW_k(\mathbf{V}_\mathbf{G}, \nu_G)$$
$$\mathbf{P}_0 \mid \mathbf{V}_\mathbf{P}, \nu_P \sim IW_k(\mathbf{V}_\mathbf{P}, \nu_P)$$

where k is the number of regression covariables (inclusive intercept), $\mathbf{V_G}$ and $\mathbf{V_P}$ are prior scale matrices and ν_G and ν_P are prior degrees of freedom of inverse Wishart distributions of covariance matrices $\mathbf{G_0}$ and $\mathbf{P_0}$, respectively.

Degrees of freedom chosen for scaled inverted chi-square and inverse Wishart prior distributions reflect the degree of belief in the prior knowledge about the true values of (co)variance components. If the true value of a (co)variance component is known almost certainly (e.g. based on many previous experiments),

a high number of prior degrees of freedom is chosen. If only little is known about the true value a priori, the prior degrees of freedom should be low. Prior distributions of (co)variance components presented here are proper if the degrees of freedom are chosen appropriately (i.e. such that the expectation of the density function is defined and the density integrates to 1). Alternatively, uninformative (flat) priors could be chosen, which usually are improper. One has to be cautious when using improper prior distributions and check carefully whether the resulting posterior distributions are proper, as inferences drawn from improper posterior distributions are not valid.

Now we can specify the joint posterior distribution of the parameters (omitting the conditioning on hyperparameters of prior distributions of (co)variance components and the known numerator relationship matrix):

$$\begin{aligned} & p(\mathbf{b},\mathbf{a},\mathbf{p},\mathbf{G}_0,\mathbf{P}_0,\sigma_\varepsilon^2 \mid \mathbf{y}) \\ & \propto p(\mathbf{y} \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{G}_0,\mathbf{P}_0,\sigma_\varepsilon^2)\, p(\mathbf{b},\mathbf{a},\mathbf{p},\mathbf{G}_0,\mathbf{P}_0,\sigma_\varepsilon^2) \\ & \propto p(\mathbf{y} \mid \mathbf{b},\mathbf{a},\mathbf{p},\sigma_\varepsilon^2)\, p(\mathbf{a} \mid \mathbf{G}_0)\, p(\mathbf{p} \mid \mathbf{P}_0)\, p(\mathbf{G}_0)\, p(\mathbf{P}_0)\, p(\sigma_\varepsilon^2) \end{aligned} \tag{12}$$

If we plug in the formulae for normal, inverse Wishart and scaled inverted chi-square distributions, respectively (one line per factor in the last line of (12), same order), this yields:

$$\begin{aligned} & = (2\pi)^{-n/2} \left|\mathbf{I}_n \sigma_\varepsilon^2\right|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})' \left(\mathbf{I}_n \sigma_\varepsilon^2\right)^{-1} (\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})\right] \\ & \times (2\pi)^{-m/2} \left|\mathbf{A} \otimes \mathbf{G}_0\right|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{a}'(\mathbf{A} \otimes \mathbf{G}_0)^{-1}\mathbf{a}\right] \\ & \times (2\pi)^{-l/2} \left|\mathbf{I}_l \otimes \mathbf{P}_0\right|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{p}'(\mathbf{I}_l \otimes \mathbf{P}_0)^{-1}\mathbf{p}\right] \\ & \times \left(2^{\nu_G k/2} \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left(\frac{\nu_G + 1 - i}{2}\right)\right)^{-1} \left|\mathbf{V}_\mathbf{G}\right|^{\nu_G/2} \left|\mathbf{G}_0\right|^{-(\nu_G+k+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{V}_\mathbf{G}^{-1}\mathbf{G}_0^{-1}\right)\right] \\ & \times \left(2^{\nu_P k/2} \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left(\frac{\nu_P + 1 - i}{2}\right)\right)^{-1} \left|\mathbf{V}_\mathbf{P}\right|^{\nu_P/2} \left|\mathbf{P}_0\right|^{-(\nu_P+k+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{V}_\mathbf{P}^{-1}\mathbf{P}_0^{-1}\right)\right] \\ & \times \left(\nu_\varepsilon/2\right)^{\nu_\varepsilon/2} \left(\Gamma\left(\nu_\varepsilon/2\right)\right)^{-1} S_\varepsilon^{\nu_\varepsilon} \left(\sigma_\varepsilon^2\right)^{-\left(\nu_\varepsilon/2+1\right)} \exp\left[-\frac{\nu_\varepsilon S_\varepsilon^2}{2\sigma_\varepsilon^2}\right] \end{aligned} \tag{13}$$

where $\Gamma(\alpha)$ is the Gamma function:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \quad \alpha > 0$$

The Gamma function is found in the formulae of scaled inverted chi-square and inverse Wishart densities, because the chi-square distribution is a special case of the Gamma distribution and the Wishart distribution is a multivariate extension of the chi-square distribution [9].

From here, fully conditional probability densities can be derived for each parameter by including all terms from (13) that involve the parameter of interest (all others are assumed known and thus constant with respect to the parameter of interest).

Using the form of the mixed model equations from (10), the posterior distribution of the location parameters $\boldsymbol{\theta}$ ("fixed" and random effects), given (co)variance components and data is given by (e.g. [12, 25, 48]):

$$\boldsymbol{\theta} \mid \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y} \sim N\left(\hat{\boldsymbol{\theta}}, \mathbf{C}^{-1}\sigma_\varepsilon^2\right) \tag{14}$$

where $\hat{\boldsymbol{\theta}} = \mathbf{C}^{-1}\mathbf{r}$ and $\mathbf{C}$ and $\mathbf{r}$ are given in (10).

Select elements of interest from $\boldsymbol{\theta}$, denote them $\boldsymbol{\theta}_1$ and reorder (for convenience):

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} \; ; \; \mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} \; ; \; \mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \; ; \; \mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix}$$

As shown e.g. by Wang *et al.* [48] and Jensen *et al.* [25], standard multivariate normal theory yields for the conditional expectation and variance of $\boldsymbol{\theta}_1$:

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2, \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y}\right] &= \mathrm{E}\left[\boldsymbol{\theta}_1 \mid \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y}\right] + \mathbf{C}^{12}\left(\mathbf{C}^{22}\right)^{-1}\left(\boldsymbol{\theta}_2 - E\left[\boldsymbol{\theta}_2 \mid \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y}\right]\right) \\
&= \mathbf{C}^{11}\mathbf{r}_1 + \mathbf{C}^{12}\mathbf{r}_2 + \mathbf{C}^{12}\left(\mathbf{C}^{22}\right)^{-1}\left(\boldsymbol{\theta}_2 - \mathbf{C}^{21}\mathbf{r}_1 - \mathbf{C}^{22}\mathbf{r}_2\right) \\
&= \mathbf{C}^{11}\mathbf{r}_1 + \mathbf{C}^{12}\mathbf{r}_2 + \mathbf{C}^{12}\left(\mathbf{C}^{22}\right)^{-1}\boldsymbol{\theta}_2 - \mathbf{C}^{12}\left(\mathbf{C}^{22}\right)^{-1}\mathbf{C}^{21}\mathbf{r}_1 - \mathbf{C}^{12}\mathbf{r}_2 \\
&= \left(\mathbf{C}^{11} - \mathbf{C}^{12}\left(\mathbf{C}^{22}\right)^{-1}\mathbf{C}^{21}\right)\mathbf{r}_1 + \mathbf{C}^{12}\left(\mathbf{C}^{22}\right)^{-1}\boldsymbol{\theta}_2 \\
&= \mathbf{C}_{11}^{-1}\mathbf{r}_1 - \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\boldsymbol{\theta}_2 \\
&= \mathbf{C}_{11}^{-1}\left(\mathbf{r}_1 - \mathbf{C}_{12}\boldsymbol{\theta}_2\right)
\end{aligned} \tag{15}$$

and

$$\begin{aligned} \mathrm{Var}\left[\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2, \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y}\right] &= \left(\mathbf{C}^{11} - \mathbf{C}^{12}\left(\mathbf{C}^{22}\right)^{-1}\mathbf{C}^{21}\right)\sigma_\varepsilon^2 \\ &= \mathbf{C}_{11}^{-1}\sigma_\varepsilon^2 \end{aligned} \tag{16}$$

From (15) and (16) we get for the full conditional distributions of "fixed" and random effects:

$$\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2, \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y} \sim N\left(\hat{\boldsymbol{\theta}}_1, \mathbf{C}_{11}^{-1}\sigma_\varepsilon^2\right) \tag{17}$$

The computations involved in (14) correspond to those used for obtaining solutions of linear systems of equations by the Gauss-Seidel algorithm [25]. Based on this result, we can write the full conditional distributions of "fixed" and random regression coefficients. The k regressions on each level of "fixed" and random effects are processed jointly, slightly modifying the calculations of the (scalar) Gauss-Seidel algorithm. Vectors $\mathbf{b}$, $\mathbf{a}$ and $\mathbf{p}$ were split into sub-vectors ($\mathbf{b}_i$, $\mathbf{a}_i$ and $\mathbf{p}_i$) of length k containing the k regression coefficients pertaining to one particular level of the respective effects.

For the "fixed" effects this yields:

$$\begin{aligned} \mathbf{b}_i \mid \mathbf{b}_{-i}, \mathbf{a}, \mathbf{p}, \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y} &\sim N\left(\hat{\mathbf{b}}_i, \left(\mathbf{X}_i'\mathbf{R}^{-1}\mathbf{X}_i\right)^{-1}\right) \\ \hat{\mathbf{b}}_i &= \left(\mathbf{X}_i'\mathbf{R}^{-1}\mathbf{X}_i\right)^{-1}\left(\mathbf{X}_i'\mathbf{R}^{-1}\mathbf{y} - \mathbf{X}_i'\mathbf{R}^{-1}\mathbf{X}_{-i}\mathbf{b}_{-i} - \mathbf{X}_i'\mathbf{R}^{-1}\mathbf{Z}\mathbf{a} - \mathbf{X}_i'\mathbf{R}^{-1}\mathbf{W}\mathbf{p}\right) \\ \hat{\mathbf{b}}_i &= \left(\mathbf{X}_i'\mathbf{R}^{-1}\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{X}_{-i}\mathbf{b}_{-i} - \mathbf{Z}\mathbf{a} - \mathbf{W}\mathbf{p}\right) \end{aligned} \tag{18}$$

where $\mathbf{X}_i$ is a submatrix of $\mathbf{X}$ consisting of the k columns pertaining to the k regressions on level i of the fixed effect in $\mathbf{b}_i$, $\mathbf{X}_{-i}$ is the complement of $\mathbf{X}_i$ to $\mathbf{X}$ and $\mathbf{b}_{-i}$ contains all the elements of $\mathbf{b}$ except the k regression coefficients in $\mathbf{b}_i$.

For random additive genetic regression coefficients, the full conditional distribution becomes:

$$
\begin{aligned}
&\mathbf{a}_i \mid \mathbf{b}, \mathbf{a}_{-i}, \mathbf{p}, \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y} \sim N\left(\hat{\mathbf{a}}_i, \left(\mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{Z}_i + \mathbf{A}_{i,i}^{-1}\mathbf{G}_o^{-1}\right)^{-1}\right) \\
&\hat{\mathbf{a}}_i = \left(\mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{Z}_i + \mathbf{A}_{i,i}^{-1}\mathbf{G}_o^{-1}\right)^{-1} \\
&\quad \times \left(\mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{y} - \mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} - \left(\underbrace{\mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{Z}_{-i}}_{\mathbf{0}} + \mathbf{A}_{i,-i}^{-1} \otimes \mathbf{G}_o^{-1} \right) \mathbf{a}_{-i} - \mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{W}\mathbf{p} \right) \\
&\hat{\mathbf{a}}_i = \left(\mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{Z}_i + \mathbf{A}_{i,i}^{-1}\mathbf{G}_o^{-1}\right)^{-1} \left(\mathbf{Z}_i'\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{W}\mathbf{p}\right) - \left(\mathbf{A}_{i,-i}^{-1} \otimes \mathbf{G}_o^{-1}\right)\mathbf{a}_{-i}\right)
\end{aligned}
\tag{19}
$$

where $\mathbf{Z}_i$ is a submatrix of $\mathbf{Z}$ consisting of the k columns pertaining to the k regressions of animal i (i = 1,..., m) in $\mathbf{a}_i$, $\mathbf{Z}_{-i}$ is the complement of $\mathbf{Z}_i$ to $\mathbf{Z}$ and $\mathbf{a}_{-i}$ contains all the elements of $\mathbf{a}$ except the k regression coefficients in $\mathbf{a}_i$. $\mathbf{A}_{i,i}^{-1}$ is the diagonal element of the inverse of the numerator relationship matrix $\mathbf{A}^{-1}$ corresponding to animal i, and $\mathbf{A}_{i,-i}^{-1}$ is a (row) vector containing all elements of row i of the inverse of the numerator relationship matrix $\mathbf{A}^{-1}$, except the diagonal element.

The last line in (19) holds if the term $\mathbf{Z}_i'\mathbf{R}^{-1}\mathbf{Z}_{-i}$ equals zero, which is correct if $\mathbf{a}$ contains just direct (and no maternal) additive genetic effects.

For permanent environmental regression coefficients, the full conditional distribution becomes:

$$
\begin{aligned}
&\mathbf{p}_i \mid \mathbf{b}, \mathbf{a}, \mathbf{p}_{-i}, \mathbf{G}_0, \mathbf{P}_0, \sigma_\varepsilon^2, \mathbf{y} \sim N\left(\hat{\mathbf{p}}_i, \left(\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{W}_i + \mathbf{P}_o^{-1}\right)^{-1}\right) \\
&\hat{\mathbf{p}}_i = \left(\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{W}_i + \mathbf{P}_o^{-1}\right)^{-1} \\
&\quad \times \left(\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{y} - \mathbf{W}_i'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} - \mathbf{W}_i'\mathbf{R}^{-1}\mathbf{Z}\mathbf{a} - \left(\underbrace{\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{W}_{-i}}_{\mathbf{0}} + \underbrace{\mathbf{I}_{i,-i}}_{\mathbf{0}} \otimes \mathbf{P}_o^{-1} \right) \mathbf{p}_{-i} \right) \\
&\hat{\mathbf{p}}_i = \left(\mathbf{W}_i'\mathbf{R}^{-1}\mathbf{W}_i + \mathbf{P}_o^{-1}\right)^{-1} \left(\mathbf{W}_i'\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a}\right)\right)
\end{aligned}
\tag{20}
$$

where $\mathbf{W}_i$ is a submatrix of $\mathbf{W}$ consisting of the k columns pertaining to the k regressions of permanent environment i (i = 1,..., l) in $\mathbf{p}_i$, $\mathbf{W}_{-i}$ is the complement of $\mathbf{W}_i$ to $\mathbf{W}$ and $\mathbf{p}_{-i}$ contains all the elements of $\mathbf{p}$ except the k regression coefficients in $\mathbf{p}_i$. $\mathbf{I}_{i,-i}$ is a (row) vector containing all the elements of row i of $\mathbf{I}$ except the diagonal element, and thus a vector of zeros.

The full conditional distributions of the (co)variance components are derived by selecting the terms from the joint posterior distribution (13), that involve this (co)variance component.

The full conditional distribution of the residual variance involves the distribution of the data conditional on location parameters and the residual variance, as well as the prior for the residual variance:

$$
\begin{aligned}
& p(\sigma_\varepsilon^2 \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{G}_0,\mathbf{P}_0,\mathbf{y}) \\
& \propto p(\mathbf{y} \mid \mathbf{b},\mathbf{a},\mathbf{p},\sigma_\varepsilon^2)\,p(\sigma_\varepsilon^2) \\
& = (2\pi)^{-n/2}\left|\mathbf{I}_n\sigma_\varepsilon^2\right|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})'(\mathbf{I}_n\sigma_\varepsilon^2)^{-1}(\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})\right] \\
& \times\left(\nu_\varepsilon/2\right)^{\nu_\varepsilon/2}\left(\Gamma\left(\nu_\varepsilon/2\right)\right)^{-1}S_\varepsilon^{\nu_\varepsilon}(\sigma_\varepsilon^2)^{-(\nu_\varepsilon/2+1)}\exp\left[-\frac{\nu_\varepsilon S_\varepsilon^2}{2\sigma_\varepsilon^2}\right]
\end{aligned}
$$

pulling together similar expressions and omitting terms that are constant with respect to the residual variance, this is proportional to:

$$
\begin{aligned}
& \propto (\sigma_\varepsilon^2)^{-((n+\nu_\varepsilon)/2+1)}\exp\left[-\frac{(\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})+\nu_\varepsilon S_\varepsilon^2}{2\sigma_\varepsilon^2}\right] \\
& = (\sigma_\varepsilon^2)^{-(\tilde{\nu}_\varepsilon/2+1)}\exp\left[-\frac{\tilde{\nu}_\varepsilon\tilde{S}_\varepsilon^2}{2\sigma_\varepsilon^2}\right]
\end{aligned}
$$

which has the form of a scaled inverted chi-square distribution:

$$\sigma_\varepsilon^2 \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{G}_0,\mathbf{P}_0,\mathbf{y} \sim Inv-\chi^2(\tilde{\nu}_\varepsilon,\tilde{S}_\varepsilon^2) \qquad (21)$$

with:

$$
\begin{aligned}
\tilde{\nu}_\varepsilon &= n+\nu_\varepsilon \\
\tilde{S}_\varepsilon^2 &= \left((\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Za}-\mathbf{Wp})+\nu_\varepsilon S_\varepsilon^2\right)/(n+\nu_\varepsilon)
\end{aligned}
$$

where n is the total number of records.

The full conditional distribution of the additive genetic covariance matrix $\mathbf{G_0}$ involves the distribution of additive genetic effects conditional on $\mathbf{G_0}$ and the prior for $\mathbf{G_0}$:

$$
\begin{aligned}
& p(\mathbf{G}_0 \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{P}_0,\sigma_\varepsilon^2,\mathbf{y}) \\
& \propto p(\mathbf{a} \mid \mathbf{G}_0)p(\mathbf{G}_0) \\
& = (2\pi)^{-m/2}\left|\mathbf{A}\otimes\mathbf{G}_0\right|^{-1/2}\exp\left[-\frac{1}{2}\mathbf{a}'(\mathbf{A}\otimes\mathbf{G}_0)^{-1}\mathbf{a}\right] \\
& \times\left(2^{\nu_G k/2}\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{\nu_G+1-i}{2}\right)\right)^{-1}\left|\mathbf{V}_\mathbf{G}\right|^{\nu_G/2}\left|\mathbf{G}_0\right|^{-(\nu_G+k+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{V}_\mathbf{G}^{-1}\mathbf{G}_0^{-1}\right)\right]
\end{aligned}
$$

pulling together similar expressions and omitting terms that are constant with respect to the additive genetic covariance matrix, this is proportional to:

$$
\begin{aligned}
& \propto \left|\mathbf{G}_0\right|^{-(m+\nu_G+k+1)/2}\exp\left[-\frac{1}{2}\left(\mathbf{a}'\left(\mathbf{A}^{-1}\otimes\mathbf{G}_0^{-1}\right)\mathbf{a}+\mathrm{tr}\left(\mathbf{V}_\mathbf{G}^{-1}\mathbf{G}_0^{-1}\right)\right)\right] \\
& = \left|\mathbf{G}_0\right|^{-(m+\nu_G+k+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\left(\mathbf{U}+\mathbf{V}_\mathbf{G}^{-1}\right)\mathbf{G}_0^{-1}\right)\right]
\end{aligned}
\qquad (22)
$$

with:

$$
\mathbf{U}=\left\{\mathbf{a}_i'\mathbf{A}^{-1}\mathbf{a}_j\right\}=\begin{bmatrix}\mathbf{a}_1'\mathbf{A}^{-1}\mathbf{a}_1 & \cdots & \mathbf{a}_1'\mathbf{A}^{-1}\mathbf{a}_k \\ \vdots & \ddots & \vdots \\ \mathbf{a}_k'\mathbf{A}^{-1}\mathbf{a}_1 & \cdots & \mathbf{a}_k'\mathbf{A}^{-1}\mathbf{a}_k\end{bmatrix}
$$

for covariables $i, j = 1,...,k$, where $\mathbf{a}_j$ is a sub-vector of $\mathbf{a}$ containing breeding values corresponding to covariable j for the m animals in the pedigree. Note that this definition of a sub-vector is different from the one presented in equation (19).

The second line in equation (22) has the form of a inverse Wishart distribution of dimension k with $m+\nu_G$ degrees of freedom and scale matrix $\left(\mathbf{U}+\mathbf{V}_\mathbf{G}^{-1}\right)^{-1}$. Thus the full conditional distribution of the additive genetic covariance matrix $\mathbf{G_0}$ of regression parameters can be written as:

$$
\mathbf{G}_0 \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{P}_0,\sigma_\varepsilon^2,\mathbf{y} \sim IW_k\left(\left(\mathbf{U}+\mathbf{V}_\mathbf{G}^{-1}\right)^{-1},m+\nu_G\right) \qquad (23)
$$

Analogously, the full conditional distribution of the covariance matrix $\mathbf{P_0}$ of permanent environmental regression coefficients involves the distribution of permanent environmental effects conditional on $\mathbf{P_0}$ and the prior for $\mathbf{P_0}$:

$$
\begin{aligned}
& p(\mathbf{P}_0 \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{G}_0,\sigma_\varepsilon^2,\mathbf{y}) \\
& \propto p(\mathbf{p} \mid \mathbf{P}_0)p(\mathbf{P}_0) \\
& = (2\pi)^{-l/2}\left|\mathbf{I}_l \otimes \mathbf{P}_0\right|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{p}'(\mathbf{I}_l \otimes \mathbf{P}_0)^{-1}\mathbf{p}\right] \\
& \times\left(2^{\nu_P k/2}\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{\nu_P+1-i}{2}\right)\right)^{-1}\left|\mathbf{V}_\mathbf{P}\right|^{\nu_P/2}\left|\mathbf{P}_0\right|^{-(\nu_P+k+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\mathbf{V}_\mathbf{P}^{-1}\mathbf{P}_0^{-1}\right)\right]
\end{aligned}
$$

pulling together similar expressions and omitting terms that are constant with respect to the permanent environmental covariance matrix, this is proportional to:

$$
\begin{aligned}
& \propto \left|\mathbf{P}_0\right|^{-(l+\nu_P+k+1)/2}\exp\left[-\frac{1}{2}\left(\mathbf{p}'(\mathbf{I}_l \otimes \mathbf{P}_0)^{-1}\mathbf{p} + \mathrm{tr}\left(\mathbf{V}_\mathbf{P}^{-1}\mathbf{P}_0^{-1}\right)\right)\right] \\
& = \left|\mathbf{P}_0\right|^{-(l+\nu_P+k+1)/2}\exp\left[-\frac{1}{2}\mathrm{tr}\left(\left(\mathbf{P}+\mathbf{V}_\mathbf{P}^{-1}\right)\mathbf{P}_0^{-1}\right)\right]
\end{aligned}
\qquad (24)
$$

with:

$$
\mathbf{P} = \left\{\mathbf{p}_i'\mathbf{p}_j\right\} = \begin{bmatrix} \mathbf{p}_1'\mathbf{p}_1 & \cdots & \mathbf{p}_1'\mathbf{p}_k \\ \vdots & \ddots & \vdots \\ \mathbf{p}_k'\mathbf{p}_1 & \cdots & \mathbf{p}_k'\mathbf{p}_k \end{bmatrix}
$$

for covariables $i, j = 1,...,k$, where $\mathbf{p}_j$ is a sub-vector of $\mathbf{p}$ containing permanent environmental effects corresponding to covariable j for the l animals with records. Note that this definition of a sub-vector is different from the one presented in equation (20).

The second line in (24) has the form of a inverse Wishart distribution of dimension k with $l+\nu_P$ degrees of freedom and scale matrix $\left(\mathbf{P}+\mathbf{V}_\mathbf{P}^{-1}\right)^{-1}$. Thus the full conditional distribution of the covariance matrix $\mathbf{P_0}$ for permanent environmental regression coefficients can be written as:

$$
\mathbf{P}_0 \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{G}_0,\sigma_\varepsilon^2,\mathbf{y} \sim IW_k\left(\left(\mathbf{P}+\mathbf{V}_\mathbf{P}^{-1}\right)^{-1}, l+\nu_P\right) \qquad (25)
$$

We have now derived all the full conditional distributions needed to implement the Gibbs sampler. The Gibbs sampling algorithm consists of sequentially sampling (many times) from (18), (19), (20), (21), (23) and (25). After a high

enough number of rounds, called the burn-in period, the Gibbs chain will yield dependent samples from its stationary distribution, on which inference will be based.

4.3. Post-Gibbs analysis

An important issue of post-Gibbs analysis is the question whether a Gibbs sampling sequence has converged to its stationary distribution and how many initial rounds (burn-in) should be discarded. According to Cowles and Carlin [4], who compared 13 methods of MCMC convergence diagnostics, it is not possible to say with certainty that a finite sample from an MCMC algorithm is representative of an underlying stationary distribution. All the methods they reviewed can fail to detect convergence failure, thus one needs to be cautious when utilising these methods to determine the length of the burn-in period. Nevertheless, convergence diagnostics should be an integral part of inference based on MCMC methods. In this thesis two different methods were used to determine burn-in: the coupling chain method of Johnson [26] and the method of Raftery and Lewis [40]. For the method of Johnson [26], two or more Gibbs sampling chains are created using the same stream of uniform random deviates, but different starting values for location and variance parameters. He showed, that such coupled or parallel chains must all eventually converge to a single sample path. The length of the burn-in period is determined when the samples of all these chains agree to some small tolerance, as the chains then have "forgotten" their initial values. Typically, the coupling chain method is inapplicable to Metropolis-Hastings samplers, as differences in the number of uniform random deviates used per round of sampling can occur and thus coupling of the chains is not guaranteed. In these situations the method of Raftery and Lewis [40] was used, which is available as a Fortran program (Gibbsit) from Statlib (URL: http://lib.stat.cmu.edu). Their approach is based on two-state Markov chain theory, as well as standard sample size formulae involving binomial variance. Based on the samples of a single Gibbs chain, the

program estimates how many initial rounds should be discarded and for how many rounds the sampler should be run to estimate a quantile of interest to the desired accuracy. Both methods used for burn-in analysis were combined with graphically checking the behaviour of several independent Gibbs chains.

After convergence of the Gibbs chain to its stationary distribution, (dependent) samples are used for inference on model parameters. Based on the ergodic theorem, posterior means of parameters can be estimated from the n samples after burn-in using equation (5) as:

$$\hat{\mu}_n = \frac{1}{n}\sum_{t=1}^{n} X_t \tag{24}$$

Similarly, other functions of the posterior distribution can be estimated from Gibbs samples, e.g. variances or marginal posterior densities. All these estimators are subject to Monte Carlo sampling error, which is reduced by prolongation of the Gibbs chain. Monte Carlo variances can be estimated running several independent chains and calculating the empirical variance of the estimates of each chain. Alternatively, Monte Carlo variances can be estimated from the Gibbs chain itself. Geyer [11] gives a survey of the current methods used for this purpose. In this thesis, the initial monotone sequence estimator [11] was used to estimate the Monte Carlo variance, i.e. the variance of the sample mean. As proposed by Sorensen *et al.* [43], an effective number of independent samples was calculated as the ratio of the sample variance and the variance of the sample mean. This "effective sample size" can help to assess the effect of the correlation of samples on the estimate of the Monte Carlo variance [43] and is thus useful for the interpretation of parameter estimates based on MCMC methods.

References

[1] Andersen S., Pedersen B., Growth and food intake curves for group housed gilts and castrated male pigs, Anim. Sci. 63 (1996) 457-464.

[2] Casella G., George E.I., Explaining the Gibbs sampler, The American Statistician 46/3 (1992) 167-174.

[3] Chib S., Greenberg E., Understanding the Metropolis-Hastings algorithm, The American Statistician 49/4 (1995) 327-335.

[4] Cowles M.K., Carlin B.P., Markov chain Monte Carlo convergence diagnostics: A comparative review, J. Am. Stat. Assoc. 91 (1996) 883-904.

[5] Elston R.C., Grizzle J.E., Estimation of time-response curves and their confidence bands, Biometrics 18 (1962) 148-159.

[6] Gelfand A.E., Smith A.F.M., Sampling-based approaches to calculating marginal densities, J. Am. Stat. Assoc. 85 (1990) 398-409.

[7] Gelfand A.E., Hills S.E., Racine-Poon A., Smith A.F.M., Illustration of Bayesian inference in normal data models using Gibbs sampling, J. Am. Stat. Assoc., 85 (1990) 972-985.

[8] Gelman A., Iterative and non-iterative simulation algorithms, Computing Science and Statistics 24 (1992) 433-438.

[9] Gelman A., Carlin J.B., Stern H.S., Rubin D.B., Bayesian data analysis, Chapman & Hall, London, 1995.

[10] Geman S., Geman D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on pattern analysis and machine intelligence, PAMI 6 (1984) 721-741.

[11] Geyer C.J., Practical Markov chain Monte Carlo, Stat. Sci. 7 (1992) 473-511.

[12] Gianola D., Fernando R.L., Bayesian methods in animal breeding theory, J. Anim. Sci. 63 (1986) 217-244.

[13] Gianola D., Foulley J.L., Variance estimation from integrated likelihoods (VEIL), Genet. Sel. Evol. 22 (1990) 403-417.

[14] Gilks W.R., Wang C.C., Yvonnet B., Coursaget P., Random-effects models for longitudinal data using Gibbs sampling, Biometrics 49 (1993) 441-453.

[15] Gilks W.R., Richardson S., Spiegelhalter D.J., Introducing Markov chain Monte Carlo, in: Gilks W.R., Richardson S., Spiegelhalter D.J. (Eds.), Markov chain Monte Carlo in Practice, Chapman & Hall, London, 1996, pp. 1-19.

[16] Gomulkiewicz R., Kirkpatrick M., Quantitative genetics and the evolution of reaction norms, Evolution 46 (1992) 390-411.

[17] Guo S.W., Thompson E.A., Monte Carlo estimation of variance component models for large complex pedigrees, IMA J. Math. Appl. Med. Biol. 8 (1991) 171-189.

[18] Harville D.A., Maximum likelihood approaches to variance component estimation and to related problems, J. Am. Stat. Assoc. 72 (1977) 320-340.

[19] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97-109.

[20] Henderson C.R. Jr., Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions, Biometrics 38 (1982) 623-640.

[21] Hofer A., Variance component estimation in animal breeding: a review, J. Anim. Breed. Genet. 115 (1998) 247-265.

[22] Jaffrézic F., Pletcher S.D., Statistical models for estimating the genetic basis of repeated measures and other function-valued traits, Genetics 156 (2000) 913-922.

[23] Jamrozik J., Schaeffer L.R., Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins, J. Dairy Sci. 80 (1997) 762-770.

[24] Jamrozik J., Schaeffer L.R., Dekkers J.C.M., Genetic evaluation of dairy cattle using test day yields and random regression model, J. Dairy Sci. 80 (1997) 1217-1226.

[25] Jensen J., Wang C.S., Sorensen D.A., Gianola D., Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler, Acta Agric. Scand., Sect. A, Animal Sci. 44 (1994) 193-201.

[26] Johnson V.E., Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths, J. Am. Stat. Assoc. 91 (1996) 154-166.

[27] Kirkpatrick M., Heckman N., A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters, J. Math. Biol. 27 (1989) 429-450.

[28] Kirkpatrick M., Lofsvold D., Measuring selection and constraint in the evolution of growth, Evolution 46 (1992) 954-971.

[29] Kirkpatrick M., Lofsvold D., Bulmer M., Analysis of the inheritance, selection and evolution of growth trajectories, Genetics124 (1990) 979-993.

[30] Kirkpatrick M., Hill W.G., Thompson R., Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle, Genet. Res. 64 (1994) 57-69.

[31] Laird N.M., Ware J.H., Random-efects models for longitudinal data, Biometrics 38 (1982) 963-974.

[32] Lindsey J.K., Models for repeated measurements, Oxford University Press, Oxford, 1999.

[33] Littell C.R., Pendergast J. and Natarajan R., Modelling covariance structure in the analysis of repeated measures data, Statistics in Medicine 19 (2000) 1793-1819.

[34] Longford N.T., Random coefficient models, Clarendon Press, Oxford, 1993.

[35] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller W., Equation of state calculations by fast computing machines, J. Chem. Phys.21 (1953) 1087-1092.

[36] Meyer K., Estimating covariance functions for longitudinal data using a random regression model, Genet. Sel. Evol. 30 (1998) 221-240.

[37] Meyer K., Hill W.G., Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood, Livest. Prod. Sci. 47 (1997) 185-200.

[38] Pletcher S.D., Geyer C.J., The genetic analysis of age-dependent traits: modeling the character process, Genetics 151 (1999) 825-835.

[39] Ptak E., Schaeffer L.R., Use of test day yields for genetic evaluation of dairy sires and cows, Livest. Prod. Sci. 34 (1993) 23-34.

[40] Raftery A.E., Lewis S., How many iterations in the Gibbs sampler?, in: Bernardo J.M., Berger J.O., Dawid A.P., Smith A.F.M. (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, U.K., 1992, pp. 763-773.

[41] Rekaya R., Carabaño M.J., Toro M.A., Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle, Livest. Prod. Sci. 57 (1998) 203-217.

[42] Schaeffer L.R., Dekkers J.C.M., Random regression in animal models for test-day production in dairy cattle, in: Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada, 7-12 August, vol. 18, 1994, pp. 443-446.

[43] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in Threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229-249.

[44] Searle S.R., Matrix algebra useful for statistics, John Wiley & Sons, New York, 1982.

[45] Tanner M., Wong W.H., The calculation of posterior distributions by data augmentation (with discussion), J. Am. Stat. Assoc. 82 (1987) 528-550.

[46] Van der Werf J.H.J., Goddard M.E., Meyer K., The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records, J. Dairy Sci. 81 (1998) 3300-3308.

[47] Wang C.S., Rutledge J.J., Gianola D., Marginal inferences about variance components in a mixed linear model using Gibbs sampling, Genet. Sel. Evol. 25 (1993) 41-62.

[48] Wang C.S., Rutledge J.J., Gianola D., Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs, Genet. Sel. Evol. 26 (1994) 91-115.

[49] White I.M.S., Thompson R., Brotherstone S., Genetic and environmental smoothing of lactation curves with cubic splines, J. Dairy Sci. 82 (1999) 632-638.

Chapter 3:

Impact of variation of length of individual testing periods on estimation of (co)variance components of a random regression model for feed intake of growing pigs

Schnyder Urs, Hofer Andreas, Künzi Niklaus

Institute of Animal Science

Swiss Federal Institute of Technology (ETH)

CH-8092 Zürich, Switzerland

This chapter has previously been published and is reprinted with permission:

Journal of Animal Breeding and Genetics, Volume 118, Issue 4, August 2001, pp. 235-246.

Abstract

A simulation study was conducted to assess the influence of differences in length of individual testing periods on estimates of (co)variance components of a random regression model for daily feed intake of growing pigs performance tested from 30 to 100 kg live weight. A quadratic polynomial in days on test with fixed regressions for sex, random regressions for additive genetic and permanent environmental effects and a constant residual variance was used for a bivariate simulation of feed intake and daily gain. (Co)variance components were estimated for feed intake only by means of a Bayesian analysis using Gibbs sampling and REML. A single trait random regression model analogous to the one used for data simulation was used to analyse two versions of the data: full data sets with 18 weekly means of feed intake per animal and reduced data sets with individual length of testing periods determined when tested animals reached 100 kg live weight. Only one significant difference between estimates from full and reduced data (REML estimate of genetic covariance between linear and quadratic regression parameters) and two significant differences from expected values (Gibbs estimates of permanent environmental variance of quadratic regression parameters) occurred. These differences are believed to be negligible, as their number lies within the expected range of the type-I-error when testing at the five percent level. The course of test day variances calculated from estimates of additive genetic and permanent environmental covariance matrices also supports the conclusion that no bias in estimates of (co)variance components occurs due to individual length of testing periods of performance tested growing pigs. Lower number of records per tested animal only results in more variation among estimates of (co)variance components from reduced compared to full data sets. Compared to full data, effective sample size of Gibbs samples from reduced data has decreased to 18 % for residual variance and increased up to five times for other (co)variances. Data structure seems to influence mixing of Gibbs chains.

1. Introduction

Typically, random regression models are applied to traits measured over a constant period of time, such as 305 days for test day records of dairy cows [10]. In growing pigs performance tested in central testing stations, the testing periods of individual pigs are defined between 30 and 100 kg live body weight and thus vary between tested animals according to their average live weight gain. If we apply a random regression model to feed intake of pigs performance tested from 30 to 100 kg live body weight, not all animals have the same number of feed intake records available. Consider a reference situation with a constant length of the testing period, such that even the slowest growing animals reach the desired slaughter weight. In the real life situation data is missing for the faster growing pigs towards the end of this constant testing period. These data can not be regarded as missing at random, because feed intake and individual length of testing (age at slaughtering) are correlated.

One implication of this situation is that feed intake curves of faster growing animals are estimated with less precision than feed intake curves of animals that grow more slowly. Consequently, (co)variance components for a random regression model estimated from such data are expected to be less precise than if all the animals had the same number of records available as the slowest growing pigs. If a quadratic polynomial is used as regression function, the quadratic term is expected to be most affected, as it describes the inflexion of feed intake curves. The question now was, if this would be the only implication of this missing data situation, or whether it might result in biased estimates of (co)variance components. Rubin [9] postulated that 1) data have to be missing at random and 2) the process that leads to missing data has to be independent of the trait analysed in order to be able to ignore the selection process for inferences on model parameters. These conditions do not hold for our situation, which implies that the process that leads to missing data should be accounted for in genetic evaluation. The only way the selection process is accounted for in a single trait

random regression model is indirectly over the length of the testing period. It is not possible to use a two trait random regression model for daily feed intake and live weight gain, as live weight is only recorded at the beginning and towards the end of the testing period. A possibility would be to include average daily gain on test as a trait correlated with regression parameters for feed intake, if necessary. However, the missing data problem may be less important for a random regression model, as the influence of missing records on (co)variance components is only indirect. (Co)variance components are estimated for curve parameters, not for single measurements on a given day on test. For these artificial traits no records are missing, as feed intake curves can be estimated for all tested animals, given a minimum number of daily feed intake records.

The objective of this study was to gain some experience with random regression models by simulating daily feed intake and live weight gain. By analysing the simulated feed intake data, we wanted to check our Gibbs sampling programs for variance components estimation, before applying them to real data. Another goal was to examine the influence of differences in individual length of testing periods on the estimation of (co)variance components for a random regression model for daily feed intake of growing pigs, when the process that leads to missing data is ignored.

2. Material and methods

2.1. Data simulation

Daily feed intake (kg) and daily weight gain (kg) were simulated simultaneously from the following multivariate random regression model, which is a second order polynomial in days on test d_m:

$$\begin{aligned} y_{ijkm} = \; & sex_{0ij} + sex_{1ij} * d_m + sex_{2ij} * d_m^2 \\ + \; & a_{0ik} + a_{1ik} * d_m + a_{2ik} * d_m^2 \\ + \; & p_{0ik} + p_{1ik} * d_m + p_{2ik} * d_m^2 \\ + \; & \varepsilon_{ijkm} \end{aligned} \tag{1}$$

where y_{ijkm} is the record of trait i for animal k of gender j on day d_m, sex_{hij} are fixed regressions for the gender of the animals, a_{hik} are random regressions for animal additive genetic effects, p_{hik} are random regressions for animal permanent environmental effects and ε_{ijkm} is a random residual error which accounts for daily deviations of feed intake and daily gain from the expected trajectory (error around curves). Random regression parameters for feed intake and daily gain were simulated multivariate normally distributed with covariance matrices **G** for animal additive genetic regression parameters a_{hik} and **P** for permanent environmental effects p_{hik}.

Fixed effects and (co)variance components of regression parameters for the simulation of feed intake were derived from results of Andersen and Pedersen [1]. Their phenotypic covariance matrix of the regression parameters was divided into additive genetic and permanent environmental variance with a ratio of 3:2. (Co)variance components of regression parameters for daily gain and their covariances with regression parameters for feed intake were chosen arbitrarily. Residuals were assumed bivariate normally distributed with a constant residual variance over the whole testing period and low positive correlation between residuals of feed intake and daily gain. These assumptions resulted in heritabilities and correlations for the traditional traits average daily feed intake,

average daily gain and feed conversion ratio calculated from the simulated data, that were similar to values found in literature (e.g. [5, 14]).

Twenty replicates with 80 sows and eight boars per generation and three offspring generations were simulated. Each boar was randomly mated to ten sows producing one litter each with one male and one female offspring for the performance test and one female as a parent for the next generation. Additionally, each boar had one male offspring chosen randomly from one of the ten matings as a parent for the next generation. This lead to 480 animals with measurements and a total of 832 animals in the pedigree. Due to random mating, the additive genetic relationship matrix varied at random over the 20 replicates.

For each tested animal, additive genetic and permanent environmental random regression parameters for daily feed intake and daily gain were generated, as well as residuals for the two traits for 126 test days. From these parameters phenotypic values for daily feed intake and daily weight gain were calculated for 126 test days according to equation (1). In order to reduce the amount of data and to save computing time for the evaluations, daily feed intake values of seven consecutive days were combined into weekly averages and written to the file as the value for the middle day of the week. This reduces the residual variance to 1/7 of the simulated value, as the variance of an arithmetic mean of n independent values is equal to the original variance of these values divided by n (see e.g. [13]). This rule holds, as residuals of consecutive test days were generated independently of each other from identical normal distributions. Daily gain served to calculate the actual live weight of tested animals on each day on test and thus as a criterion to determine individual length of testing periods. This was done by cumulating daily weight gains over the whole testing period of 126 days, starting on test day one with a live body weight of 30 kg. The end of test was defined at the end of the week when tested animals reached 100 kg live body weight. Daily weight gain was not included in the evaluations otherwise.

Table I: Average distribution of number of records per animal in reduced data sets.

records	8	9	10	11	12	13	14	15	16	17	18
% of animals	< 0.1	3.9	24.6	36.2	23.0	8.5	2.7	0.5	0.2	< 0.1	0.3

Two versions of the data were evaluated (20 replicates each):

1) full data sets with a constant length of the testing period, 18 weekly means of feed intake per tested animal and a total of 8 640 measurements per data set.

2) reduced data sets with individual length of the testing period, slaughtering at the end of the week when 100 kg live weight were reached and an average of 5 380 measurements per data set. All animals had eight or more records. Table I shows the average distribution of the number of records per animal.

The number of replicates was chosen in order to attain maximum power and yet to keep the evaluations computationally feasible within a reasonable period of time. In order to learn about the extent of variation produced by the simulation process, realised values of (co)variance components for feed intake were calculated from simulated effects of all data sets as [8]:

$$\begin{aligned} \mathbf{G} &= \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} \,/\, \#\text{animals} \\ \mathbf{P} &= \mathbf{p}' \mathbf{p} \,/\, \#\text{tested animals} \\ \sigma_\varepsilon^2 &= \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} \,/\, \#\text{records} \end{aligned} \qquad (2)$$

where **a** and **p** are matrices of dimension three by number of animals and number of tested animals respectively, containing additive genetic and permanent environmental effects of random regression coefficients. $\mathbf{A}^{-1}$ is the inverse of the additive genetic relationship matrix and ε is a vector containing residual effects for each record. While **G** and **P** are the same for full and reduced data sets, there may be differences in the realised value calculated for σ_ε^2, as it depends on the number of records used. Mean squared errors of these realised values were calculated as an indicator of variation between replicates. Table II shows the

values of (co)variance components used for data simulation (expected values), mean realised values calculated from simulated random effects, as well as mean squared errors of these realised values.

2.2. Estimation of (co)variance components

The model for the evaluation of simulated feed intake data was identical with the one used for simulation, except that it was a single trait random regression model for daily feed intake only, as daily gain was not evaluated.

For the estimation of (co)variance components two different methods were used:

1) Bayesian methodology using Gibbs sampling

2) REML

Table II: Values of (co)variance components used for data simulation (expected values) together with realised values and their mean squared errors (MSE)

covariance component		expected value	realised value	MSE
G(1,1)		7.380e-02	7.215e-02	9.411e-06
G(2,1)		-1.176e-03	-1.149e-03	2.321e-09
G(2,2)		3.456e-05	3.444e-05	1.110e-12
G(3,1)		0.000e+00	6.159e-08	4.627e-13
G(3,2)		0.000e+00	-3.043e-10	1.831e-16
G(3,3)		6.000e-09	6.072e-09	9.340e-20
P(1,1)		4.920e-02	4.883e-02	1.020e-05
P(2,1)		-7.840e-04	-7.873e-04	3.141e-09
P(2,2)		2.304e-05	2.343e-05	2.325e-12
P(3,1)		0.000e+00	-1.308e-07	3.507e-13
P(3,2)		0.000e+00	5.539e-09	1.991e-16
P(3,3)		4.000e-09	4.061e-09	7.119e-20
σ_ε^2	full	3.571e-02	3.546e-02	2.397e-07
	reduced	3.571e-02	3.541e-02	3.814e-07

For (co)variance components estimation with Gibbs sampling, our own programs were used. Solutions of the mixed model equations with expected values for the (co)variance components obtained by Gauss-Seidel iteration served as starting values for fixed and random effects. MME were processed block wise by means of Cholesky decomposition and backsubstitution during Gauss-Seidel iteration and when generating new solutions in the Gibbs sampler. Informative priors with low numbers of degrees of freedom were used for the (co)variance components, such that their expectations corresponded to the expected values and propriety of priors is guaranteed. Priors for the covariance matrices of regression parameters **G** and **P** were inverse Wishart distributions with five degrees of freedom. For the residual variance σ_ε^2 a scaled inverse Chi-square distribution also with five degrees of freedom was assumed a priori. Fixed regression parameters for each sex were assumed constant and random regression parameters were assumed normally distributed with zero mean and variance $\mathbf{I}\otimes\mathbf{P}$ for permanent environmental and $\mathbf{A}\otimes\mathbf{G}$ for additive genetic regression parameters, respectively.

Each data set was analysed with 100 000 rounds of Gibbs sampling. The first 30 000 samples were discarded from the burn in period, determined by the coupling chain method [4]. Gibbs samples averaged over the remaining 70 000 rounds served as estimates of posterior means of the covariance matrices **G** and **P** and the residual variance σ_ε^2. Effective sample size [12] was calculated using estimates of Monte Carlo variance obtained by the method of initial monotone sequence estimator [2] from the 70 000 samples after burn-in. This estimator was preferred by Geyer [2] over the initial positive sequence estimator, because of making large reductions in the worst overestimates while doing little to underestimates.

After analysing Gibbs sampling results, the influence of the priors used was questioned. Restricted maximum likelihood (REML) was used as a reference method for (co)variance component estimation instead of rerunning the Gibbs sampler with completely flat priors. REML estimates correspond to the mode of

the joint posterior distribution of the desired parameters, marginalized with respect to unknown location parameters in a Bayesian analysis with flat priors for fixed effects and (co)variance components [3]. REML estimates of the (co)variance components were obtained with the program REMLF90, which uses the EM-algorithm with acceleration [7]. Expected values were provided as starting values for the (co)variance components. The influence of starting values on the REML-analysis was tested on three full and the corresponding three reduced data sets. Expected values provided as starting values for (co)variance components reduced the number of iterations required for convergence on average by 19 %, compared to an evaluation with starting values that were far from expectation. Both sets of starting values lead to the same estimates.

Estimated covariance matrices were analysed in two ways. Variances and covariances of random regression coefficients were compared element by element, and test day variances were calculated from covariance matrices for the middle day of all 18 weeks of the testing period. The main focus of the study was on differences between estimates from full and reduced data sets. For both methods, Gibbs sampling and REML, these differences were tested at the five percent level by means of a two-sided paired t-test. The second question was whether estimates of (co)variance components from full and reduced data sets were equal to their expected values. Gibbs sampling and REML results were analysed separately. A comparison of the two methods would not be very meaningful, as they do not use the same prior distributions and because mean (Gibbs) and mode (REML) of respective marginal distributions need not be the same. Each method yielded two sample means, which are not independent of each other. The Bonferroni-method was used to account for multiple testing and to control the overall type-I-error of the experiment at the five percent level. Thus differences between estimates and expected values were tested at the 2.5 % level by means of a one-sample t-test. For all tests function t.test of the statistical software package S-Plus version 3.4 was used [6].

3. Results and discussion

3.1. Estimates of (co)variance components

Table III shows the values of (co)variance components used for data simulation (expected values) and estimates of posterior means of (co)variance components obtained by Gibbs sampling from full and reduced data sets, averaged over 20 replicates. Two significant differences of (co)variance components from expected values were found with Gibbs sampling, one for full and one for reduced data. Both of them concerned the permanent environmental variance of quadratic regression parameters, which describe the inflexion of feed intake curves. None of the differences between estimates from full and reduced data proved to be significant.

Table III: Gibbs sampling estimates of posterior means of (co)variance components (average over 20 replicates)

covariance component	expected value	estimates of posterior mean	
		full data	reduced data
G(1,1)	7.380e-02	7.100e-02	6.968e-02
G(2,1)	-1.176e-03	-1.147e-03	-1.129e-03
G(2,2)	3.456e-05	3.542e-05	3.515e-05
G(3,1)	0.000e+00	-8.006e-07	-1.589e-07
G(3,2)	0.000e+00	3.984e-09	2.005e-08
G(3,3)	6.000e-09	7.111e-09	6.600e-09
P(1,1)	4.920e-02	4.973e-02	5.025e-02
P(2,1)	-7.840e-04	-7.914e-04	-7.842e-04
P(2,2)	2.304e-05	2.273e-05	2.216e-05
P(3,1)	0.000e+00	1.805e-07	-6.494e-07
P(3,2)	0.000e+00	1.640e-08	2.517e-08
P(3,3)	4.000e-09	**3.174e-09**	**3.169e-09**
σ_ε^2	3.571e-02	3.556e-02	3.552e-02

bold: significant differences from expected values

No significant differences from expected values were found with REML, but contrary to the Gibbs sampling estimates, one difference between estimates from full and reduced data appeared significant (Table IV). This was the genetic covariance between the linear and the quadratic regression parameter (G(3,2)), which has been simulated to be zero and thus very small values were estimated.

Table IV: REML estimates of (co)variance components (average over 20 replicates)

covariance component	expected value	REML estimates full data	REML estimates reduced data
G(1,1)	7.380e-02	6.332e-02	6.408e-02
G(2,1)	-1.176e-03	-9.980e-04	-1.073e-03
G(2,2)	3.456e-05	3.108e-05	3.524e-05
G(3,1)	0.000e+00	-3.449e-07	7.777e-07
G(3,2) §	0.000e+00	-5.429e-10	-4.410e-08
G(3,3)	6.000e-09	6.294e-09	6.716e-09
P(1,1)	4.920e-02	5.508e-02	5.518e-02
P(2,1)	-7.840e-04	-9.084e-04	-9.089e-04
P(2,2)	2.304e-05	2.645e-05	2.806e-05
P(3,1)	0.000e+00	4.349e-08	-2.231e-07
P(3,2)	0.000e+00	1.164e-08	-3.938e-09
P(3,3)	4.000e-09	3.737e-09	3.910e-09
σ_ε^2	3.571e-02	3.551e-02	3.528e-02

§ significant difference between estimates from full and reduced data sets

A total of 78 differences were tested, 26 at the five percent level and 52 at the 2.5 %-level, from which 2.6 significant results can be expected to occur by chance. We found a total of three significant results, from which we believe that they only appeared by chance, as their number lies within the expected range of the type-I-error. These tests indicate that there is no bias in estimates of

(co)variance components due to the reduction of available data associated with individual length of testing periods.

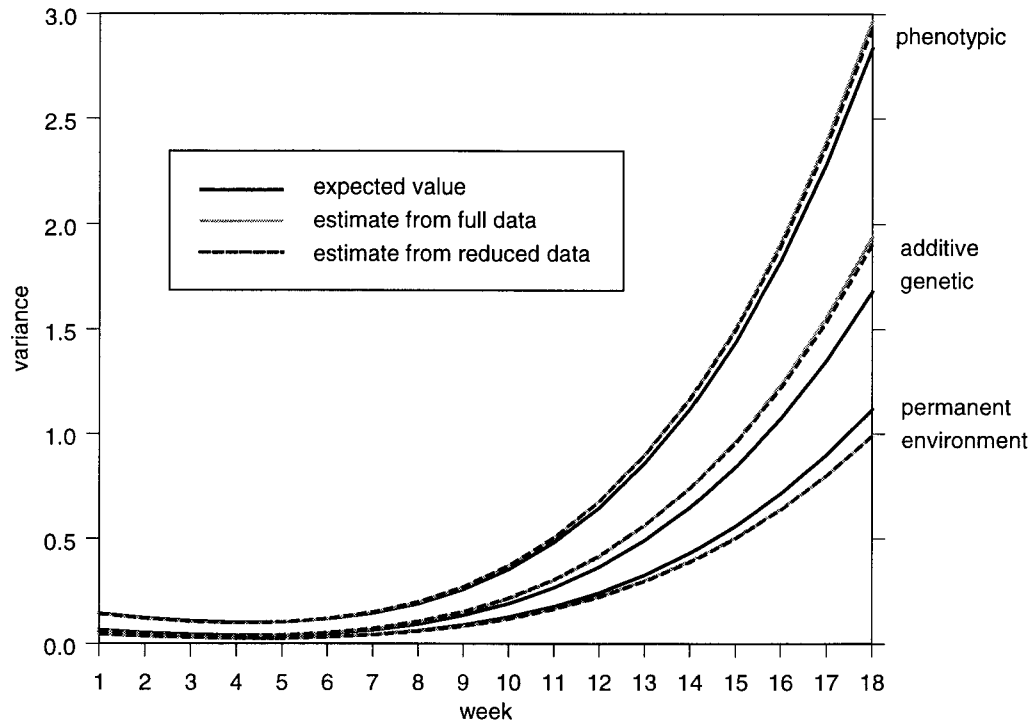


Figure 1: Expected values and Gibbs sampling estimates from full and reduced data sets of genetic, permanent environmental and phenotypic test day variances of weekly means of daily feed intake (kg).

Variances for weekly means of daily feed intake calculated from estimates of (co)variance components (Tables III, IV) as well as expected values are shown in Figures 1 (Gibbs sampling) and 2 (REML). Curves of estimates from full and reduced data sets were almost identical with both REML and Gibbs sampling. This supports the view that there was no bias in (co)variance component estimation due to the reduction in available data. With Gibbs sampling, additive genetic variance has been overestimated in the second half of the testing period, while permanent environmental variance has been underestimated (Figure 1). This is due to overestimation of additive genetic and underestimation of permanent environmental variance of quadratic regression parameters, respectively (Table III). Overestimation of additive genetic variance was

compensated by underestimation of permanent environmental variance, resulting in only a slight overestimation of the phenotypic variance (Figure 1). Test day variances calculated from REML estimates were closer to expected values in this second part of the testing period. With both methods, Gibbs sampling and REML, none of the differences from expected values of additive genetic and permanent environmental test day variances proved to be significant.

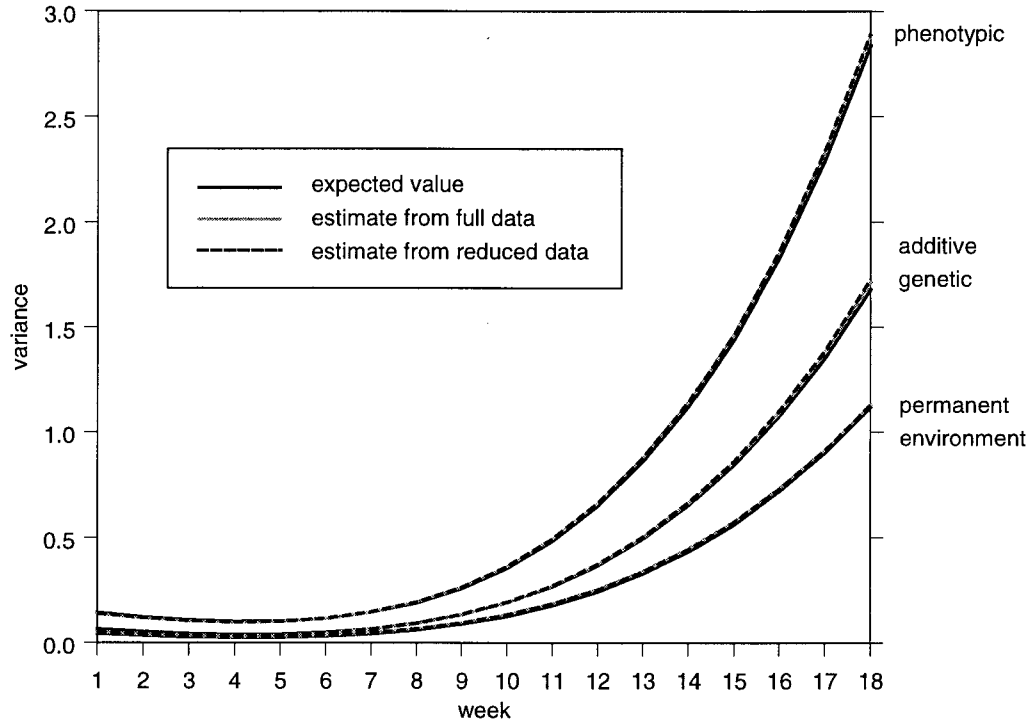


Figure 2: Expected values and REML estimates from full and reduced data sets of genetic, permanent environmental and phenotypic test day variances of weekly means of daily feed intake (kg).

3.2. Mean squared errors

Table V shows mean squared errors (MSE) of estimates of (co)variance components, obtained by Gibbs sampling and REML from full and reduced data sets, relative to MSE of values realised with data simulation. With both methods, MSE of estimates were very large compared to realised values (Table V). The exception was the residual variance, whose estimates showed a much smaller

variation than other parameters. The variation among estimates of (co)variance components resulted in standard deviations of estimates of daily variance that were two to five times bigger than standard deviations of daily variances calculated from realised values of (co)variance components (data not shown). This shows that estimates from single data sets were not very precise. Only averaging estimates over the 20 replicates gave good estimates for (co)variance components. For good estimates from single data sets, much more animals per data set are needed.

Table V: Mean squared errors (MSE) of estimates of (co)variance components divided by MSE of values realised with simulation

covariance component	Gibbs sampling		REML	
	full data	reduced data	full data	reduced data
G(1,1)	47.2	39.4	57.0	53.3
G(2,1)	76.6	70.2	83.8	88.2
G(2,2)	106.7	168.6	116.7	202.2
G(3,1)	37.0	36.7	30.5	40.3
G(3,2)	47.7	71.3	29.3	88.4
G(3,3)	57.5	42.7	34.3	45.1
P(1,1)	24.2	21.2	28.4	28.4
P(2,1)	34.7	40.2	38.6	56.5
P(2,2)	31.5	44.9	39.1	81.7
P(3,1)	17.4	39.2	17.7	45.7
P(3,2)	19.7	41.0	15.6	72.5
P(3,3)	27.5	32.0	14.8	41.3
σ_ε^2	1.2	1.7	1.2	2.0

For most parameters, MSE of estimates from full data sets are smaller than those of reduced data sets (Table V). Exceptions are G(1,1) for both REML and Gibbs sampling, as well as G(2,1), G(3,3) and P(1,1) for Gibbs sampling only. Together

with insignificant differences between estimates from full and reduced data sets, this shows that the influence of the amount of data available is rather on precision (variance) than accuracy (bias) of estimates. While precision of estimated covariance matrices of additive genetic and permanent environmental regression parameters mainly depends on the number of animals with records, if the number of measurements per animal is sufficiently high, estimation of residual variance depends on the total number of measurements containing this residual. Reduced data sets contain 5 380 measurements on average compared to 8 640 in full data sets, which is a reduction of over 37 %. This reduction of information for the estimation of the residual variance in reduced compared to full data sets results in less precise estimates for residual variance. Information for estimation of covariance matrices of regression parameters should not differ too much, as both data sets contained the same number of tested animals with at least eight measurements. Precision of estimates from reduced data sets was reduced because of less precise estimation of regression parameters due to fewer measurements per animal.

3.3. Effective sample size

Effective sample size was estimated for all parameters from all Gibbs chains. Averages over the 20 replicates and standard deviations of estimates of effective sample size are presented in Table VI. Except for the residual variance, estimates of effective sample size of parameters were rather small compared to the 70 000 samples used to estimate posterior means of parameters. A reliable estimation of the marginal posterior density would only have been possible for the residual variance, as numbers of effective samples of other parameters were insufficient. The large standard deviations showed that estimates of effective sample size varied a lot between replicates. For full data sets, estimates of effective sample size for variances and covariance of intercept and linear regression parameters were clearly higher than for variances of quadratic regression parameters and their covariances. This suggests, that variances of intercept and linear regression

Table VI: Effective sample size estimated from Gibbs samples of (co)variance components by the method of initial monotone sequence estimator

	full data		reduced data	
covariance component	mean of estimates	standard deviation	mean of estimates	standard deviation
G(1,1)	91.4	40.6	145.3	52.8
G(2,1)	81.4	43.6	145.3	52.7
G(2,2)	80.6	33.0	150.8	59.0
G(3,1)	32.7	17.6	121.5	41.5
G(3,2)	36.3	21.8	121.0	85.2
G(3,3)	29.4	16.9	151.6	89.8
P(1,1)	93.2	38.8	146.5	55.0
P(2,1)	87.6	51.7	143.8	45.3
P(2,2)	81.6	34.2	147.1	51.5
P(3,1)	33.7	22.0	126.3	34.9
P(3,2)	40.3	34.6	118.3	60.3
P(3,3)	31.3	27.6	128.5	55.1
σ_ε^2	39 574.9	3 567.1	6 587.4	2 947.7

parameters were easier to estimate than quadratic regression parameters. Surprisingly, estimates of effective sample size from Gibbs chains of reduced data sets showed a different pattern. Effective sample size of the residual variance was reduced to about 18 % of its value from full data sets, while estimates of effective sample size of other parameters increased up to five times. Still the same parameters had most effective samples, but differences between parameters became much smaller. The amount of information in the data set and especially the number of records per tested animal seemed to influence the mixing of the Gibbs chain. With longer testing periods and thus more records per animal in full data sets, regression parameters that describe feed intake curves of tested animals are determined more restrictively than in reduced data sets. This restriction might reduce the Gibbs samplers ability to explore the parameter

space and thus result in slower mixing of the Gibbs chains for covariance matrices **G** and **P** for full compared to reduced data sets. Effective sample size of the residual variance σ_ε^2 seems to be influenced by the mixing of Gibbs chains for covariance matrices **G** and **P**. If **G** and **P** move around slowly, σ_ε^2 can explore its parameter space relatively free. If **G** and **P** move around faster, σ_ε^2 has to react more to their movements and to adapt to the new situation. This restricts its freedom to explore the parameter space and results in slower mixing and smaller effective sample size for the residual variance with reduced compared to full data sets.

3.4. General Discussion

The model assumptions used for generating the data may be too simple to reflect the real life situation correctly. Non-zero covariances with quadratic regression parameters and a residual variance that changes with age would have been more realistic. However, this would not change our conclusions when the same model is used for simulation and analysis of the data, as the true values are very likely to lie within the range of values realised by the simulation process (Table II). Zero covariances with quadratic regression parameters were chosen because of lack of knowledge about their true values, and the residual variance was assumed constant for simplicity. A random regression model incorporating a residual variance that changes with age was used for the analysis of real feed intake data in a different study [11].

Regressing feed intake on body weight instead of on days on test would remedy the problem of different lengths of testing periods, as all animals would have the same range of values for regression covariables on the weight scale. However, live weights of performance tested growing pigs are not measured regularly throughout the testing period in most testing schemes, but only at the beginning and towards the end. Therefore, the necessary live weight measures for regressing feed intake on live weight are usually not available in real data sets.

4. Conclusions

Both methods used for estimation of (co)variance components, REML and Gibbs sampling, showed that there was no systematic bias in (co)variance components estimates due to reduction in available data towards the end of a constant testing period. Therefore, it seems not to be necessary to include daily gain into the evaluations, which is the process that causes the missing data in this situation. For evaluations of real data, larger data sets than in this study should be used, in order to obtain good estimates of parameter matrices.

Gibbs chains should be much longer than the 100 000 samples run in this study, if marginal posterior densities of parameters were to be estimated. Effective sample size and mixing of the Gibbs chains can be influenced by the data structure. Optimising the number of records per tested animal might be used to improve mixing of the Gibbs chain in random regression analysis. To verify this hypothesis further research is needed.

References

[1] Andersen S., Pedersen B., Growth and food intake curves for group housed gilts and castrated male pigs, Anim. Sci. 63 (1996) 457-464.

[2] Geyer C.J., Practical Markov chain Monte Carlo, Stat. Sci. 7 (1992) 473-511.

[3] Harville D.A., Maximum likelihood approaches to variance component estimation and to related problems, J. Am. Stat. Assoc. 72 (1977) 320-340.

[4] Johnson V.E., Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths, J. Am. Stat. Assoc. 91 (1996) 154-166.

[5] Labroue F., Gueblez R., Sellier P., Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace pigs, Genet. Sel. Evol. 29 (1997) 451-468.

[6] MathSoft, S-Plus, Version 3.4 for Unix, Supplement, Data Analysis Products Division, MathSoft, Incorporation, Seattle, Washington, USA, 1996.

[7] Misztal I., REMLF90 Manual, Animal and Dairy Science, University of Georgia, Athens, Georgia, USA, 1998.

[8] Rao C.R., Estimation of variance and covariance components in linear models, J. Am. Stat. Assoc. 67 (1972) 112-115.

[9] Rubin D.B., Inference and missing data, Biometrika 63 (1976) 581-592.

[10] Schaeffer L.R., Dekkers J.C.M., Random regression in animal models for test-day production in dairy cattle, in: Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada, 7-12 August, vol. 18, 1994, pp. 443-446.

[11] Schnyder U., Hofer A., Labroue F., Künzi N., Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs, Genet. Sel. Evol. 33 (2001) xxx-xxx.

[12] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in Threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229-249.

[13] Stahel W.A., Statistische Datenanalyse, Eine Einführung für Naturwissenschaftler, Vieweg, Braunschweig/Wiesbaden, 1995.

[14] Von Felde A., Roehe R., Looft H., Kalm E., Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars, Livest. Prod. Sci. 47 (1996) 11-22.

Chapter 4:

Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs

Urs Schnyder[a], Andreas Hofer[a], Florence Labroue[b], Niklaus Künzi[a]

[a] Institute of Animal Science,

Swiss Federal Institute of Technology (ETH),

CH-8092 Zürich, Switzerland

[b] Institut technique du porc,

La Motte au Vicomte, BP 3,

35651 Le Rheu cedex, France

This chapter has been accepted for publication and is reprinted with permission:

Genetics Selection Evolution, Volume 33, Number 6, November/December 2001, pp. xxx-xxx.

Abstract

Daily feed intake data of 1 279 French Landrace (FL, 1 039 boars and 240 castrates) and 2 417 Large White (LW, 2 032 boars and 385 castrates) growing pigs was recorded with electronic feed dispensers in three French central testing stations in years 1992-94. Group housed pigs fed *ad libitum* were performance tested from 35 to 95 (males) or 100 kg (castrates) live body weight. A quadratic polynomial in days on test with fixed regressions for sex and batch, random regressions for additive genetic, pen, litter and individual permanent environmental effects was used, with two different models for the residual variance: constant in model 1 and modelled with a quadratic polynomial depending on the day on test d_m as follows in model 2: $\sigma^2_{\varepsilon_m} = \exp(\gamma_0 + \gamma_1 * d_m + \gamma_2 * d_m^2)$. Variance components were estimated from weekly means of daily feed intake by means of a Bayesian analysis using Gibbs sampling. Posterior means of (co)variances were calculated using 800 000 samples from four chains (200 000 each). Heritability estimates of regression coefficients were 0.30 (FL model 1), 0.21 (FL model 2), 0.14 (LW1) and 0.14 (LW2) for intercept, 0.04 (FL1), 0.04 (FL2), 0.11 (LW1) and 0.06 (LW2) for linear, 0.03 (FL1), 0.04 (FL2) 0.11 (LW1) and 0.06 (LW2) for quadratic term. Heritability estimates for weekly means of daily feed intake were lowest in week 4 (FL1: 0.11, FL2: 0.11) and week 1 (LW1: 0.09, LW2: 0.10), and highest in week 11 (FL1: 0.25, FL2: 0.24) and week 8 (LW1: 0.19, LW2: 0.18), respectively. Genetic eigenfunctions reveal that altering the shape of the feed intake curve by selection is difficult.

1. Introduction

Today, selection of pigs for growth performance considers average daily feed intake and average daily live weight gain over the whole growing period and/or the ratio of the two, i.e. feed conversion. Average daily feed intake is negatively correlated with the leanness of the carcass. Selection for increased leanness and improved feed conversion has lead to a decrease of the feed intake capacity (FIC) [27]. "Modern" genotypes of pigs have a lower mean voluntary feed intake and feed intake increases at a lower rate with body weight compared to "older" genotypes [2]. In the long run, FIC might become a limiting factor for a further improvement of the efficiency of lean growth. In the past, improvement of feed conversion was mainly achieved by a reduction of the rate of fat deposition. But according to several authors, optimum levels of backfat thickness are or will soon be reached and other routes to improve feed efficiency have to be found [7, 14, 26]. De Vries and Kanis [5] suggested to divide the growing period into 3 phases:

1) early fattening period where FIC of pigs is determined by mechanical constraints and FIC is less than the optimum level of feed intake (FI(opt)), where lean deposition rate is at its maximum and fat deposition rate at its minimum for the given lean deposition rate [4],
2) intermediate fattening period where FIC is still determined by mechanical constraints but FIC > FI(opt),
3) late fattening period where FIC is determined by metabolic constraints with FIC > FI(opt).

Increasing FIC in period 1 to its optimum level should increase growth rate without affecting the leanness of the carcass, while increasing FIC in periods 2 or 3 would lead to fatter carcasses. Increasing FIC in period 1 while keeping FIC in periods 2 and 3 constant should lead to animals growing more efficiently. Webb [27] supports this view and stresses the need of further research on genetic and environmental effects on the shape of feed intake curves.

Electronic feeders installed in central testing stations allow for the measurement of individual daily feed intake of performance tested growing pigs. Analyses of feed intake curves might lead to new interesting traits for pig breeders, e.g. curve parameters or feed intake capacity at different ages. One possibility to analyse feed intake curves is by means of polynomials [1] using a random regression model [22].

The objective of this study was to estimate genetic variation in feed intake curves of growing pigs and to assess possibilities to change the feed intake curve by selection.

2. Material and methods

2.1. Data

1 279 French Landrace (FL, 1 039 boars and 240 castrates) pigs from 697 litters and 2 417 Large White (LW, 2 032 boars and 385 castrates) pigs from 1 259 litters were performance tested in three French central testing stations in years 1992-94. For each tested animal, pedigree information of three generations of ancestors was available, which resulted in 3826 (FL) and 7784 (LW) animals in the pedigree, respectively. Growing pigs were housed in groups of 6 to 15 animals in 316 (FL) and 370 (LW) pens, respectively. Pens were equipped with one electronic feed dispenser each (Acema-48, Acemo, Pontivy, Morbihan, France), where *ad libitum* daily feed intake was recorded. Groups that were on test during the same period of time on the same testing station form a batch. There was a total of 35 batches with French Landrace and 36 batches with Large White pigs. After about one week of adaptation to the automatic feed dispensers, boars were tested from 35 to 95 kg and castrated males from 35 to 100 kg live body weight. Raw data contained daily feed intake records for the whole period during which the animals were on the testing station, but records from the adaptation period were discarded. Test day one was defined as the day when animals reached 35 kg live body weight. Starting from there, weekly means of

feed intake per day were calculated and saved as the record for the middle day of the week, in order to reduce the amount of data for the evaluations. This resulted in records for days 4, 11, 18,..., 74, 81, 88 (Table I). The last record of an animal represents feed intake of the last week before leaving the testing station after reaching 95 kg (entire males, candidates to selection) or 100 kg live body weight (castrates, slaughtered contemporaries).

Table I: Number of animals with records of weekly means of feed intake per day (LW = Large White; FL = French Landrace; % = proportion of tested animals).

week	1	2	3	4	5	6	7	8	9	10	11	12	13
day	4	11	18	25	32	39	46	53	60	67	74	81	88
LW	2312	2263	2229	2173	2292	2255	2213	2137	1907	1227	509	131	14
%	95.7	93.6	92.2	89.9	94.8	93.3	91.6	88.4	78.9	50.8	21.1	5.4	0.6
FL	1214	1192	1163	1156	1224	1183	1178	1160	1042	752	323	103	19
%	94.9	93.2	90.9	90.4	95.7	92.5	92.1	90.7	81.5	58.8	25.3	8.1	1.5

The variance of an arithmetic mean of n independent values is equal to the original variance of these values divided by n (see e.g. [24]). Averaging daily records into weekly means therefore results in a reduction of the residual variance proportional to the number of records included in this average. Whenever records of more than one day per week were missing, all the records of this week were discarded and the weekly mean was set to missing, to avoid a major influence of missing records on the estimate of residual variance. Animals with less than five records of weekly means for the estimation of feed intake curves were deleted from the data set. This was also necessary if no records were available in the first three weeks of the testing period, as this might lead to poor estimates for polynomials, especially negative values for the intercept, which is not plausible.

2.2. Model

The following random regression model, which is a quadratic polynomial in days on test d_m was fitted to weekly means of daily feed intake records:

$$\begin{aligned} y_{ghijkm} = {} & sex_{0g} + sex_{1g} * d_m + sex_{2g} * d_m^2 \\ & + batch_{0h} + batch_{1h} * d_m + batch_{2h} * d_m^2 \\ & + a_{0i} + a_{1i} * d_m + a_{2i} * d_m^2 \\ & + p_{0j} + p_{1j} * d_m + p_{2j} * d_m^2 \\ & + l_{0k} + l_{1k} * d_m + l_{2k} * d_m^2 \\ & + e_{0i} + e_{1i} * d_m + e_{2i} * d_m^2 \\ & + \varepsilon_{ghijkm} \end{aligned} \qquad (1)$$

where sex_{ng} and $batch_{nh}$ are fixed regressions for the gender of the animals, and the period and station of their test, respectively; a_{ni} are random regressions for animal additive genetic effects; p_{nj}, l_{nk} and e_{ni} are random regressions for permanent environmental effects of pen, litter and the tested individual, respectively; ε_{ghijkm} is a random residual error which accounts for daily deviations of feed intake from the expected trajectory of animal i on day d_m.

Model (1) can also be presented in hierarchical form, using a quadratic polynomial as a regression function and fitting fixed (*sex, batch*) and random (*a, p, l, e*) effects to regression coefficients, which can be regarded as artificial traits. What is called "permanent environmental effect of the tested individual" above, is nothing else than a residual for regression coefficients. The quadratic polynomial was chosen as a regression function for (weekly means of) daily feed intake based on results of Anderson and Pedersen [1], who showed that a cubic polynomial is sufficient to fit cumulated feed intake of growing pigs. A cubic polynomial for cumulated feed intake corresponds to a quadratic polynomial for daily feed intake, as daily feed intake can be written as the first derivative of cumulated feed intake. A higher order polynomial would fit the data better (reduce the residual variance), but would also substantially increase the number of covariances to be estimated. This additional effort seems not to be justified, as

feed intake is expected to evolve smoothly (almost linear) within the growing period considered.

Fixed and random effects for regression coefficients were chosen based on results of Labroue [16, 17], who analysed daily feed intake averaged within three growing periods (based on the same raw data) using a multivariate model. Instead of fitting a fixed effect for group size (number of pigs in a pen), a random permanent environmental effect for each pen (group of pigs housed together) was included in the model. The same fixed and random effects were applied to all three regression coefficients to guarantee a proper definition of heritability for these artificial traits (see section 2.4).

Normal distribution of feed intake data is assumed:

$$\mathbf{y} \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{l},\mathbf{e},\sigma^2_{\varepsilon_m} \sim N\left\{\mathbf{Xb}+\mathbf{Za}+\mathbf{Up}+\mathbf{Vl}+\mathbf{We},\mathbf{I}\sigma^2_{\varepsilon_m}\right\} \qquad (2)$$

y is a vector containing feed intake data; **b** is a vector containing fixed regressions for sex and batch of dimension three times total number of levels of fixed effects; **a**, **p**, **l** and **e** are vectors containing random regressions for additive genetic and permanent environmental effects of dimension three times number of animals in the pedigree, number of pens, number of litters and number of animals on test, respectively; $\sigma^2_{\varepsilon_m}$ is the residual variance of day on test d_m and **X**, **Z**, **U**, **V** and **W** are incidence matrices containing regression covariables for each record.

The residuals are assumed to be independent. Two different models were applied for the residual variance. In the first model it was assumed constant over the whole testing period for all animals and in the second model all the animals were assumed to have the same residual variance on a given day on test d_m, but the course of the residual variance was modelled as follows:

$$\sigma^2_{\varepsilon_m} = \exp\left(\gamma_0 + \gamma_1 * d_m + \gamma_2 * d_m^2\right) \qquad (3)$$

This second model is expected to fit the data better, because the residual variance is likely to change during the testing period due to scale effects.

The following assumptions were used for the distributions of fixed and random effects (regressions):

$$\begin{aligned} &\mathbf{b} \sim \text{constant} \\ &\mathbf{a} \mid \mathbf{A}, \mathbf{G}_0 \sim N\{\mathbf{0}, (\mathbf{A} \otimes \mathbf{G}_0)\} \\ &\mathbf{p} \mid \mathbf{P}_0 \sim N\{\mathbf{0}, (\mathbf{I} \otimes \mathbf{P}_0)\} \\ &\mathbf{l} \mid \mathbf{L}_0 \sim N\{\mathbf{0}, (\mathbf{I} \otimes \mathbf{L}_0)\} \\ &\mathbf{e} \mid \mathbf{E}_0 \sim N\{\mathbf{0}, (\mathbf{I} \otimes \mathbf{E}_0)\} \end{aligned} \tag{4}$$

A is the numerator relationship matrix, $\mathbf{G_0}$ is the (co)variance matrix of random regressions of additive genetic effects and $\mathbf{P_0}$, $\mathbf{L_0}$ and $\mathbf{E_0}$ are (co)variance matrices for random regressions of permanent environmental effects. All these (co)variance matrices are of dimension 3 x 3.

Informative priors with low numbers of degrees of freedom were used for the variance components. For the 3 x 3 (co)variance matrices of regression coefficients $\mathbf{G_0}$, $\mathbf{P_0}$, $\mathbf{L_0}$ and $\mathbf{E_0}$, inverse Wishart distributions with five degrees of freedom were used. Prior scale matrices were equal for all four covariance matrices. Elements of scale matrices corresponding to intercept and linear regression coefficients were chosen such that their expected value corresponded to one fourth of the phenotypic (co)variances derived from Andersen and Pedersen [1]. Expected values for phenotypic (co)variances of the quadratic regression coefficient were arbitrarily set to 1.0e-8 (variance) and zero (covariances), as Andersen and Pedersen [1] included random effects for intercept, linear and quadratic regression coefficients only, when fitting a cubic polynomial in days on test for cumulated feed intake. The resulting elements of scale matrices for covariance matrices of random regression coefficients are shown in Table II. For the constant residual variance σ_ε^2 a scaled inverse Chi-square distribution with five degrees of freedom and scale parameter $s_\varepsilon^2 = 0.015$ was used. Priors for parameters γ_0, γ_1 and γ_2, that describe the course of the residual variance $\sigma_{\varepsilon_m}^2$ in the second model, were assumed independent of each

other and normally distributed with standard deviations of 1.5 (γ_0), 0.1 (γ_1) and 0.01 (γ_2).

Table II: Lower diagonal elements of symmetric scale matrix **S** for inverse Wishart prior distributions of additive genetic ($\mathbf{G_0}$) and permanent environmental ($\mathbf{P_0}$, $\mathbf{L_0}$, $\mathbf{E_0}$) covariance matrices of random regression coefficients.

Element	S(1,1)	S(2,1)	S(2,2)	S(3,1)	S(3,2)	S(3,3)
Value	3.075e-2	-4.900e-4	1.440e-5	0.0	0.0	2.500e-9

2.3. Variance components estimation

For the estimation of (co)variance components our own programs were used, applying Bayesian methodology using Gibbs sampling [9]. The joint posterior distribution of the parameters given the data is the product of the likelihood and the prior distributions of all parameters [8]. From there marginal distributions are derived easily, as they only have to be known up to proportionality. This results in normal distributions for fixed and random regressions and in inverse Wishart distributions for the (co)variance matrices for additive genetic and permanent environmental effects. For model 1, with a constant residual variance, the marginal distribution of σ_ε^2 is a scaled inverted Chi-square distribution. The parameters γ_0, γ_1 and γ_2, that describe the course of the residual variance $\sigma_{\varepsilon_m}^2$ in the second model, had to be sampled via a Metropolis-Hastings algorithm [12, 19], as their distribution is not a standard one. In each round of Gibbs sampling a new set of parameters γ_i was sampled with a random-walk Metropolis algorithm [21]. Deviations from the current parameter values were generated from independent normal proposal densities with zero mean and fixed standard deviations (0.04, 0.002 and 0.00002 for γ_0, γ_1 and γ_2, respectively). The acceptance probability for this set of candidate points depends only on the ratio of the product of the likelihood and the prior densities of the parameters to be

sampled, evaluated at the candidate points and the current parameter values. In each round of Gibbs sampling, this in-built Metropolis-Hastings algorithm was run until a new set of parameters γ_0, γ_1 and γ_2 was accepted.

Mixed model equations (MME) were processed block wise by means of Cholesky decomposition and backsubstitution when generating new solutions in the Gibbs sampler. For each combination of data sets (French Landrace and Large White) and models (constant and variable residual variance), four Gibbs chains were run, with 250 000 samples each.

2.4. Post-Gibbs analysis

Burn-in for the first chain of model 1 was determined by the coupling chain method [13]. For this, a shorter chain (100 000 samples) was run with different starting values for (co)variance components and fixed and random effects, but identical pseudo random number sequence. Line plots of samples of (co)variance components from every 100^{th} round of Gibbs sampling were used to monitor convergence of the chains to identical sample values. For the other three chains of model 1 and the four chains of model 2 the same burn-in period was adopted and checked graphically on the single chains only. The coupling chain method could not be used for model 2, because in each round of Gibbs sampling the in-built Metropolis-Hastings sampler for parameters γ_0, γ_1 and γ_2 may cause a shift in the pseudo random number sequence relative to coupled chains. For all graphical analysis of Gibbs chains the statistical software package S-Plus [18] was used.

Effective sample size [23] of samples after burn-in was estimated for each chain using estimates of Monte Carlo variance obtained by the method of initial monotone sequence estimator [10]. This estimator was preferred by Geyer [10] over the initial positive sequence estimator, because of making large reductions in the worst overestimates while doing little to underestimates.

Samples from the burn-in period were discarded and posterior means calculated from the remaining samples of each chain served as estimates of (co)variance components. Heritabilities and genetic and phenotypic correlations of regression coefficients were calculated from estimates of posterior means of (co)variance components as well as from samples from every 100^{th} round of Gibbs sampling after burn-in. Density plots of calculated samples of heritabilities and correlations were made in S-Plus [18] to illustrate their distributions.

The concept of heritability for regression coefficients is comparable to the heritability of a trait averaged over the whole testing period (e.g. average daily feed intake), it should clearly be distinguished from the heritability of a single measurement as defined in a simple repeatability model. The phenotypic covariance matrix used for calculating heritabilities and phenotypic correlations of regression coefficients is defined as the sum of additive genetic ($\mathbf{G_0}$) and permanent environmental ($\mathbf{P_0}$, $\mathbf{L_0}$, $\mathbf{E_0}$) covariance matrices. Residuals ε_{ghijkm} (daily deviations from the fitted curve) in model (1) are expected to sum to zero within each animal, as any overall deviation from zero should be incorporated into the intercept of the fitted polynomial. The variance of these residuals depends on the length of the (time) interval which is specified rather arbitrarily (one day, one week, entire growing period) when recording feed intake. Residuals are not part of regression coefficients and therefore the residual variance is excluded from the phenotypic covariance matrix of these artificial traits. It must be included in the definition of the phenotypic variance (and thus influence the heritability) of a single record of the trait evaluated with a random regression model, though.

For the whole testing period, additive genetic and permanent environmental variances of weekly means of daily feed intake were computed from posterior means of (co)variance components as (shown for additive genetic variance):

$$\sigma^2_{G_m} = \boldsymbol{\varphi}'_{\mathbf{m}} \mathbf{G}_0 \boldsymbol{\varphi}_{\mathbf{m}} \qquad (5)$$

where $\sigma^2_{G_m}$ is the additive genetic variance for the day on test d_m; $\mathbf{G_0}$ is the estimate of posterior mean for the additive genetic covariance matrix of regression coefficients and $\boldsymbol{\varphi}'_m = (1 \quad d_m \quad d_m^2)$ is a row vector containing regression covariables for the day on test d_m.

Daily variances calculated based on estimates of (co)variance matrices of additive genetic and the three permanent environmental effects as well as the residual variance were summed to get model estimates of phenotypic daily variances. These estimates of genetic and phenotypic daily variance were used to calculate heritabilities for weekly means of daily feed intake. Estimates of variances and heritability for weekly means of daily feed intake were plotted for the whole testing period.

The fit of the two models with different modelling of the residual variance was judged based on phenotypic daily variances. Model estimates calculated as shown above were compared to phenotypic daily variances calculated from data corrected for fixed effects included in the model. Two different methods were used to correct data for fixed effects. On the one hand estimates of fixed regression curves obtained with the respective models were used, and on the other hand fixed effects were estimated for each test day separately with analysis of variance function "aov" in S-Plus [18] using a fixed effects model.

2.5. Eigenfunctions and eigenvalues

In order to assess the potential for genetic changes of the feed intake curve, genetic eigenfunctions and eigenvalues were calculated from additive genetic (co)variance matrices $\mathbf{G_0}$. In order to allow for meaningful comparisons between the eigenvalues, eigenfunctions have to be adjusted to a norm of unity [15]. Therefore, estimates of genetic (co)variance matrices $\mathbf{G_0}$ of regression coefficients were transformed into (co)variance matrices of regression coefficients based on normalised orthogonal polynomials. For this purpose normalised Legendre polynomials were used [15]:

$$\mathbf{C} = \mathbf{\Phi G_0 \Phi'} = \mathbf{\Phi_1 K \Phi_1'}$$
$$\mathbf{K} = (\mathbf{\Phi_1' \Phi_1})^{-1} \mathbf{\Phi_1' \Phi G_0 \Phi' \Phi_1} (\mathbf{\Phi_1' \Phi_1})^{-1} \quad (6)$$

C is a matrix containing genetic (co)variances between daily measurements of feed intake of dimension *n* x *n*, where *n* is the number of days with measurements; $\mathbf{G_0}$ is the genetic (co)variance matrix between random regression coefficients using quadratic polynomials; **K** is the genetic (co)variance matrix between random regression coefficients using normalised second order Legendre polynomials; Φ is a matrix of *n* rows by three columns containing covariables for quadratic polynomials and Φ_1 is a matrix of *n* rows by three columns containing covariables for normalised second order Legendre polynomials.

After transformation of $\mathbf{G_0}$ into **K**, eigenvalues and eigenvectors were calculated from **K** with S-Plus [18]. The three resulting eigenvectors were multiplied with Φ_1 in order to obtain the three eigenfunctions evaluated for the *n* corresponding days with measurements. The corresponding eigenvalues indicate how much of the genetic variance of a population is explained by a given eigenfunction [15]. Therefore, eigenvalues were transformed to a percent scale, with their sum equal to 100 %.

3. Results and discussion

3.1. Behaviour of Gibbs chains

The coupled chains with identical pseudo random number sequence [13], to determine burn-in with model 1, resulted for both data sets in identical samples within 40 000 rounds of Gibbs sampling. In order to be on the save side for model 2, another 10 000 samples were discarded.

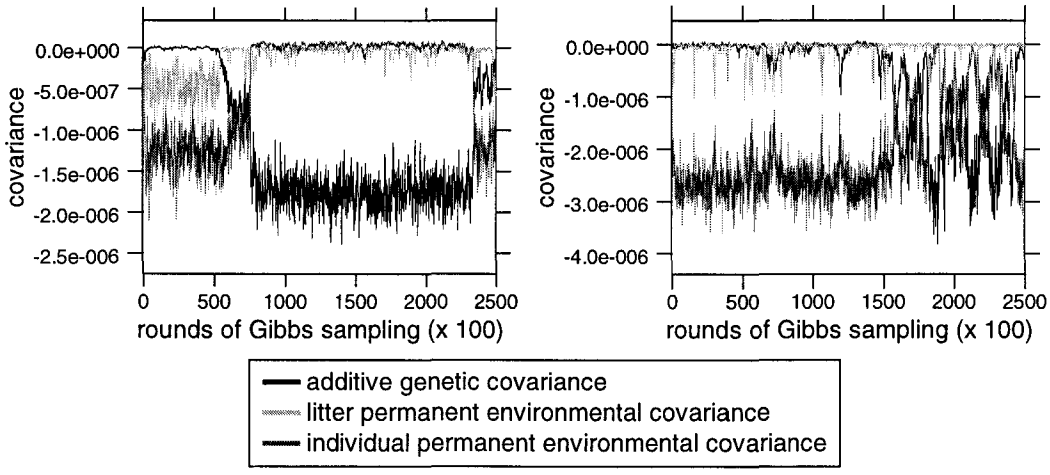


Figure 1: Gibbs samples of additive genetic, litter and individual permanent environmental covariance between linear and quadratic regression coefficients from every 100th round of the Gibbs chain with irregular behaviour under model 1 for Large White (left panel) and French Landrace (right panel) data.

When graphically checking whether Gibbs chains had converged to a stationary distribution within the 50 000 rounds of burn-in chosen, an irregular pattern was discovered for both breeds in one of the four chains run under model 1. Especially (co)variance components of additive genetic, litter and individual permanent environmental effects of linear and quadratic regression coefficients were affected. This is illustrated in Figure 1 with samples of covariances between linear and quadratic regression coefficients. For Large White, the affected chain behaves "normal" for somewhat more than 50 000 rounds, before the additive

genetic effect absorbs most of the covariance of litter and individual permanent environmental effects. Towards the end of the chain, partition of covariance among effects is again about the same as in round 50 000 (Figure 1). Other (co)variances show a similar pattern. Only the variance of the intercept regression coefficient (for all effects) and permanent environmental effects of the pen (for all (co)variances) were not affected. For French Landrace the change in partition of (co)variances occurred after 150 000 rounds, as shown in Figure 1 for the covariance between linear and quadratic regression coefficients. For the remaining rounds fluctuations of samples were rather large compared to earlier rounds and not as stable as in the affected period of the Large White chain. For French Landrace the variance of the intercept regression coefficient was also affected, but no changes in pen and litter (co)variances were found. For both breeds none of the other Gibbs chains showed a similar pattern, neither the three other chains run with model 1, nor the four chains run with model 2. For these chains a burn-in period of 50 000 rounds of Gibbs sampling seems to be sufficient by far. They seem to have reached their stationary distribution already after a few thousand rounds. The reasons for this strange behaviour discovered in two Gibbs chains are not entirely clear. With the proper prior distributions chosen for random effects and (co)variance components, property of the posterior distribution should be guaranteed. Gibbs sampling programs were carefully checked for errors, and were found to work correctly. Pseudo random number sequences used were different for the affected chains of the two breeds, and showed no problems when used for the other model-breed combinations. We therefore believe, that the Gibbs sampler reached this different configuration of (co)variance distribution among additive genetic and permanent environmental effects for regression coefficients in the affected chains just by chance. This configuration may be supported by the data with some low probability, but is not likely to represent the true state of nature. Slow mixing of Gibbs chains may be the reason why the sampler got stuck in this configuration for so many rounds of Gibbs sampling. Because samples of (co)variances left what is believed to be the

true highest density region of the stationary distribution for a substantial number of rounds, we decided not to use the affected chains for inferences on model parameters. Increase in additive genetic and decrease in permanent environmental (co)variance of regression coefficients would have had a major impact on estimates of heritabilities. To guarantee a fair comparison of results between the two models, one additional Gibbs chain was run for both breeds with model 1, which behaved completely normal for both breeds. Thus, inferences on model parameters are based on four chains with a total of 800 000 samples (after burn-in) for all four model-breed combinations.

Table III: Sums of estimates of effective sample size for elements of covariance matrices of intercept, linear and quadratic regression coefficients for daily feed intake (both models), the constant residual variance (model 1) and parameters γ_i describing the course of the residual variance under model 2, based on samples after burn-in of four Gibbs chains (800 000 samples total). Large White data.

Model	effect / element	(1,1)	(2,1)	(2,2)	(3,1)	(3,2)	(3,3)
	additive genetic	250	265	58	241	52	52
	perm. env. pen	20451	18198	20113	18523	21703	23717
Model 1	perm. env. litter	1983	254	90	243	88	87
	ind. perm. env.	800	692	493	594	431	396
	residual variance	199702					
	additive genetic	216	128	51	106	53	61
	perm. env. pen	20102	19180	21121	20277	22475	25498
Model 2	perm. env. litter	1527	282	105	291	109	115
	ind. perm. env.	505	329	484	253	443	414
	γ_0, γ_1, γ_2	38025	33130	34576			

Sums of estimates of effective sample size of the four chains run for each model-breed combination are shown in Table III (Large White) and IV (French Landrace). For all model-breed combinations lowest estimates of effective

sample size were found for estimates of additive genetic (co)variance components. Low estimates of effective sample size indicate slow mixing of Gibbs chains, which is considered the main reason for the long burn-in period that was chosen. Within effects, estimates are lowest for variances of linear and quadratic regression coefficients and their covariance, with the exceptions of permanent environmental effect of pens for Large White (both models) and individual permanent environmental effects under model 2 for both breeds.

Table IV: Sums of estimates of effective sample size for elements of covariance matrices of intercept, linear and quadratic regression coefficients for daily feed intake (both models), the constant residual variance (model 1) and parameters describing the course of the residual variance under model 2, based on samples after burn-in of four Gibbs chains (800 000 samples total). French Landrace data.

Model	effect / element	(1,1)	(2,1)	(2,2)	(3,1)	(3,2)	(3,3)
	additive genetic	635	259	196	222	205	245
	perm. env. pen	18064	9874	2910	8846	2651	2781
Model 1	perm. env. litter	3506	1147	546	931	529	554
	ind. perm. env.	1028	996	1260	1071	1394	1446
	residual variance	189164					
	additive genetic	636	275	140	267	199	207
	perm. env. pen	17138	10380	2656	9568	2414	2523
Model 2	perm. env. litter	3558	1683	358	1507	347	366
	ind. perm. env.	1235	1370	850	1162	764	715
	γ_0 , γ_1 , γ_2	24512	10598	10954			

Highest estimates of effective sample size were found for parameters describing the residual variance and for (co)variances of permanent environmental effects of pens. On average, permanent environmental effects of pens were estimated based on records of 6.5 animals for Large White and 4.1 animals for French Landrace, respectively. For all other random effects of regression coefficients the average

number of animals with records per level of effect is much lower. The number of animals with records was 1.9 per litter for Large White and 1.8 per litter for French Landrace, respectively, one per level of individual permanent environmental effect and considerably less than one per level of additive genetic effect (0.31 for Large White and 0.33 for French Landrace, respectively, including ancestors in the pedigree). Mixing of Gibbs chains for (co)variance components of random regression coefficients thus seems to depend on the amount of information available in the data to estimate each level of the random effect considered. For most parameters estimates of effective sample size are not high enough to allow for accurate density estimates. For this purpose at least a few thousand independent samples from the posterior distribution are required [20]. Therefore only estimates of posterior means will be given and density plots of every 100^{th} sample can only give an indication of distributions.

3.2. Heritabilities and correlations

Estimates of heritabilities and correlations of regression coefficients for daily feed intake are shown in Table V. French Landrace pigs show a quite high heritability for the scalar regression coefficient with model 1, which is reduced substantially under model 2, but still remains higher than for Large White pigs. Heritabilities for linear and quadratic regression coefficients are higher for Large White pigs than for French Landrace, but also reduced under model 2 compared to model 1 (Table V). These heritabilities already show, that it is easier to change the overall level (associated with the intercept regression coefficient) than the shape of feed intake curves (associated with linear and quadratic regression coefficients).

Phenotypic correlations are very similar for both breeds and also between models, whereas genetic correlations differ substantially between breeds (Table V). For French Landrace genetic correlations between the intercept and linear as well as quadratic regression coefficients are more in line with phenotypic correlations than for Large White. Differences between genetic and

phenotypic correlations between linear and quadratic regression coefficients are smaller in Large White than in French Landrace. The reason for these differences may be found in (co)variance components of individual permanent environmental regression coefficients, as pen and litter explain only a small part of permanent environmental variation (data not shown).

Table V: Heritabilities (bold), genetic (above diagonals) and phenotypic (below diagonals) correlations of intercept, linear and quadratic regression coefficients for daily feed intake.

Model \ Breed		Large White			French Landrace		
Model 1	intercept	**0.14**	0.01	0.02	**0.30**	-0.62	0.36
	linear	-0.47	**0.11**	-0.84	-0.48	**0.04**	-0.27
	quadratic	0.29	-0.89	**0.11**	0.31	-0.91	**0.03**
Model 2	intercept	**0.14**	-0.04	0.10	**0.21**	-0.51	0.26
	linear	-0.52	**0.06**	-0.73	-0.53	**0.04**	-0.36
	quadratic	0.33	-0.89	**0.06**	0.36	-0.92	**0.04**

Eissen [6] estimated heritabilities and correlations of feed intake curve parameters in a two step approach. First he fitted linear polynomials depending on days on test to daily feed intake records of growing Duroc pigs. Afterwards he used intercept, linear regression coefficient and residual standard deviation of the fit for individual pigs in a multivariate analysis. For both intercept and linear regression coefficient he found a heritability estimate of 0.32, which is except for intercept of French Landrace much higher than our estimates for the corresponding parameters (Table V). His estimates of genetic and phenotypic correlations between intercept and linear regression coefficient are -0.38 and -0.62, which is in the same range as our estimates, except for genetic correlations of Large Whites (Table V).

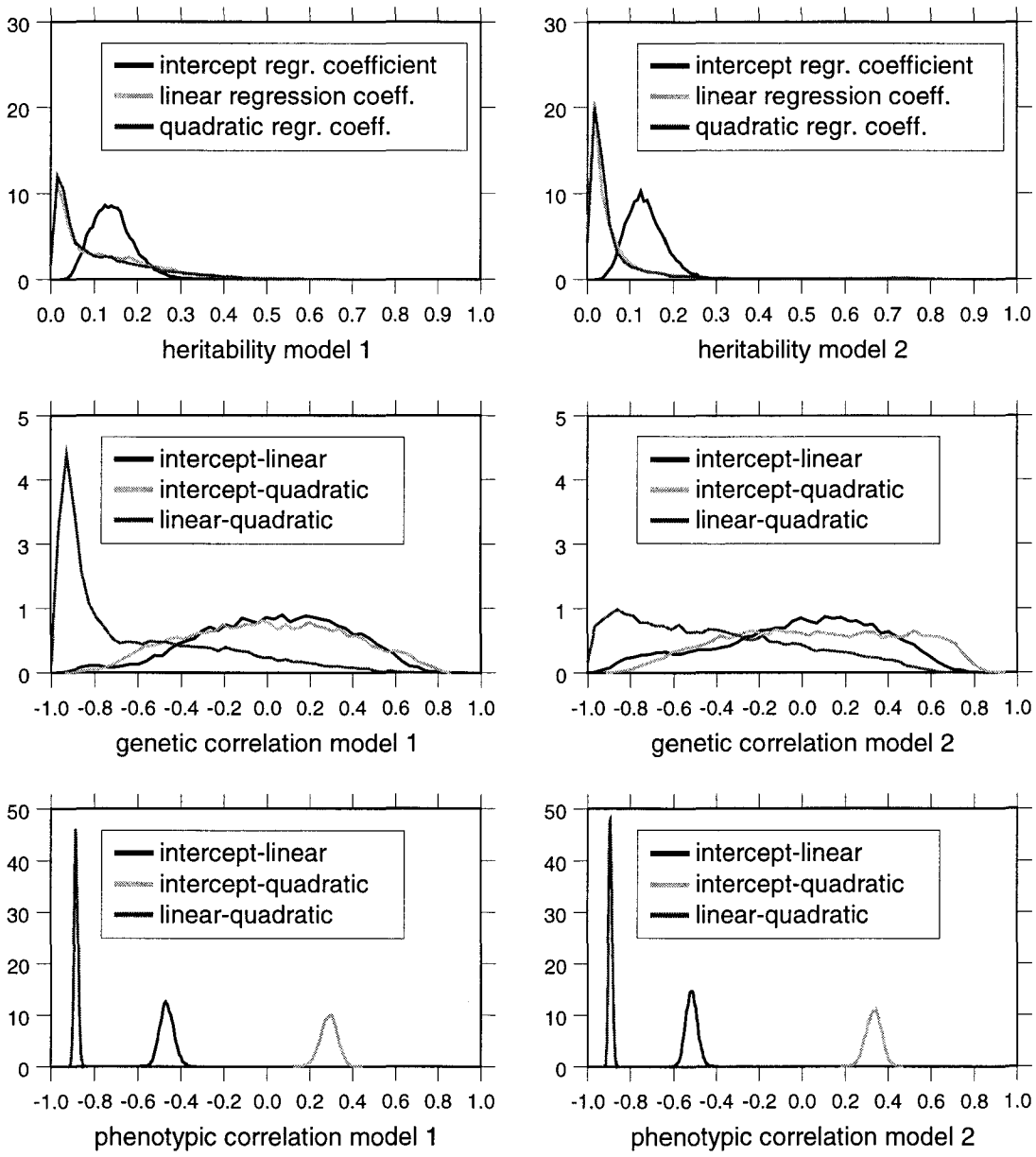


Figure 2: Density plots of heritabilities, genetic and phenotypic correlations of intercept, linear and quadratic regression coefficients for Large White, calculated from every 100th Gibbs sample of covariance matrices used for inferences under model 1 and model 2 (8 000 samples each).

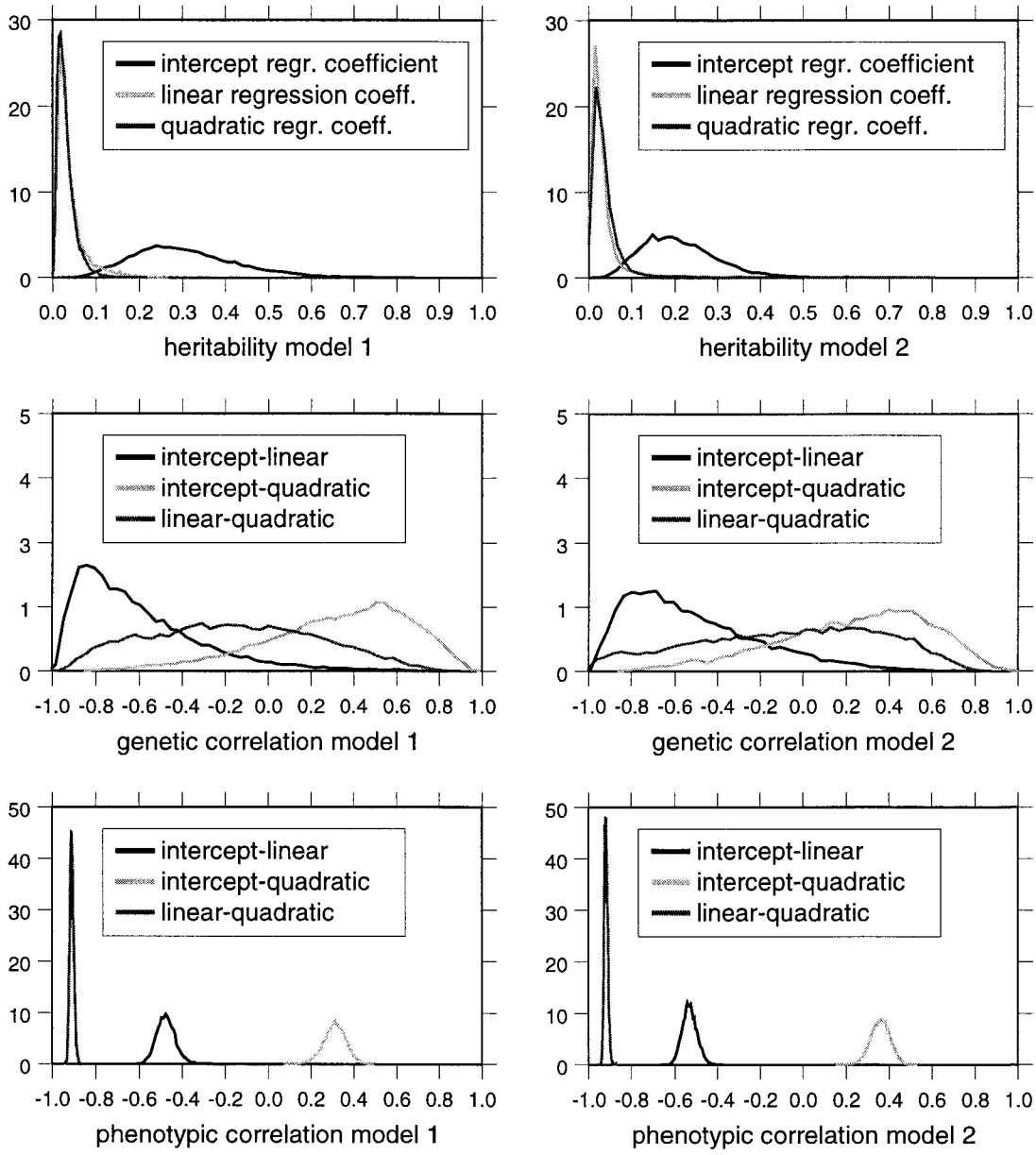


Figure 3: Density plots of heritabilities, genetic and phenotypic correlations of intercept, linear and quadratic regression coefficients for French Landrace, calculated from every 100th Gibbs sample of covariance matrices used for inferences under model 1 and model 2 (8 000 samples each).

Density plots (Figures 2 and 3) of heritabilities and correlations calculated from every 100th sample of (co)variances indicate how accurate these parameters can be estimated from our data. Phenotypic correlations are very well estimated for all model-breed combinations, as can be seen from their high and narrow density plots. Quite contrary, density plots of genetic correlations (note the different scales) are flat over almost the whole parameter space. This indicates, that information on genetic parameters is very limited in both data sets, which may also be the reason for the slow mixing of genetic parameters. Differences between models in the shape of density plots of correlations are small and must be interpreted carefully, as estimates of effective sample size were very low for additive genetic (co)variances (Tables III and IV). A little difference can be found in Large Whites for the genetic correlation between linear and quadratic regression coefficients (Figure 2), which shows a little peak close to the lower end of the parameter space in model 1 and not in model 2. For phenotypic correlations only positions of means differ slightly. Density plots of heritabilities are intermediate in height and width compared to genetic and phenotypic correlations. Heritabilities show more accentuated peaks for linear and quadratic regression coefficients than for the intercept regression coefficient. This may be due to the fact that these low heritabilities are situated close to the lower limit of the parameter space.

3.3. Course of variances and heritabilities

Figures 4 and 5 show the course of the additive genetic variance, the sum of the three permanent environmental variances and the residual variance for weekly means of daily feed intake estimated with models 1 and 2 for Large White and French Landrace growing pigs. As litter and pen explain only a very small part of the total variation, permanent environmental variances were summed to reduce the number of lines in the figures. Variances were plotted for the first eleven weeks on test only, as there are substantially less animals with records in weeks 12 and 13 (Table I). Course of variances is quite similar for both models, except

for the residual variance, which is constant in model 1, while it starts low in model 2 and gets quite high towards the end of the testing period. For both breeds, the sum of permanent environmental variances for model 1 is smaller in the beginning and larger towards the end of the testing period than for model 2.

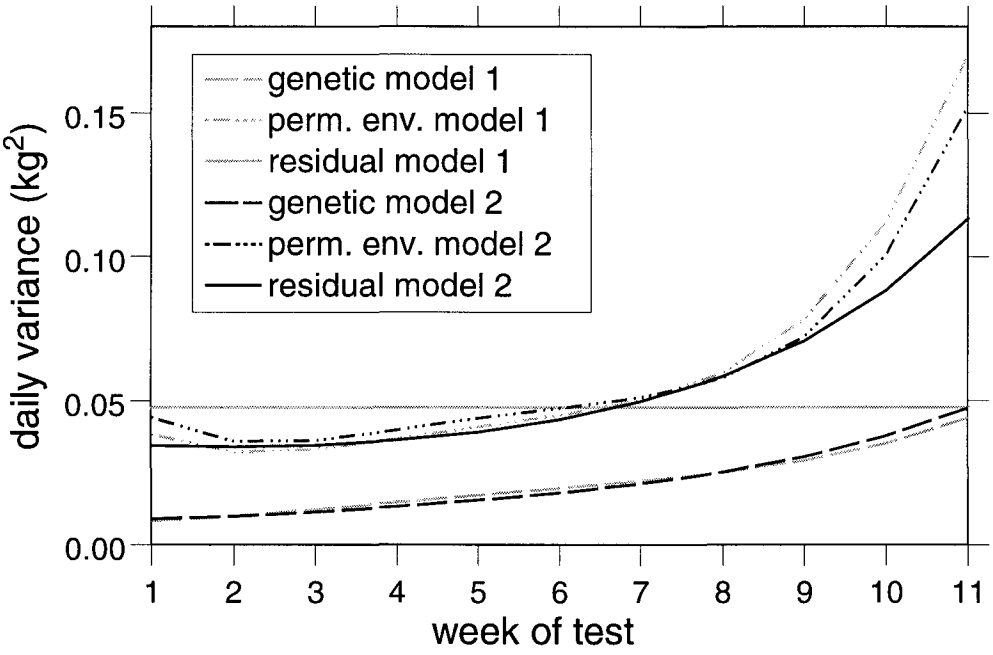


Figure 4: Course of variances for daily feed intake (kg) of Large White growing pigs for models 1 and 2.

Under model 2, lower residual variance in early test weeks is partly compensated by higher permanent environmental variance, and vice versa for late test weeks. While the course of genetic variance is similar for Large White and French Landrace pigs, the pattern shown for the permanent environmental variance is different and less regular than for the genetic variance. This has also an influence on the course of heritabilities for weekly means of feed intake per day (Figure 6), which shows a different pattern for French Landrace than for Large White pigs. The general rise of variance during the testing period may partly be due to the fact that the feed intake capacity of animals increases with age and size, but it may also be influenced by variable length of testing periods, as less (slower growing) animals have records in the last two or three weeks (Table I). Because

length of testing periods of individual pigs depends on body weight gain, fit of polynomials for faster growing pigs is based on less records than for slower growing pigs. As accuracy of polynomial fit can only be guaranteed between the first and last record of an individual pig, daily variance may be overestimated for late weeks because polynomials of fast growing pigs are not accurate any more.

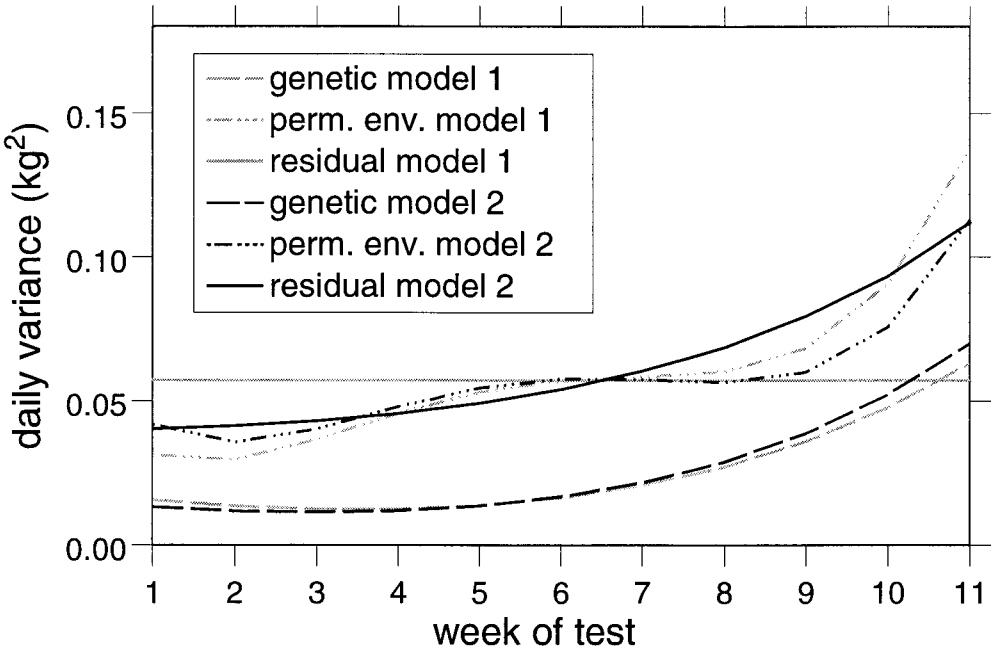


Figure 5: Course of variances for daily feed intake (kg) of French Landrace growing pigs for models 1 and 2.

Heritabilities found in this study are substantially lower than the estimate of 0.42 for average daily feed intake found by Labroue *et al.* [17] for the same data. Most of this difference may be explained by the difference in methodology, as the residual variance (around curves) is reduced by averaging daily feed intake over the whole testing period. Compared to the model with weekly means of daily feed intake that was used here, the residual variance is reduced by a factor equal to the average number of weeks that animals were on test. In her PhD-thesis, Labroue [16] also estimated heritabilities for weekly means of feed intake in weeks 2, 6 and 10 of the testing period based on the same data. These estimates are lower than the estimate for average daily feed intake, but on average still 0.1

higher than our results (Figure 6), except for the slightly lower estimate of heritability in week 6 for French Landrace. These differences can not be explained by reduction of residual variance, as weekly means of daily feed intake were used in both studies. One possible reason are differences in effects included in the models. Labroue [16] used a fixed effect for group size, while group effects were included as random permanent environmental effect of pen in this study. But as variance of permanent environmental effect of pens is small compared to additive genetic variance, this explains only about ten percent of differences in heritability estimates.

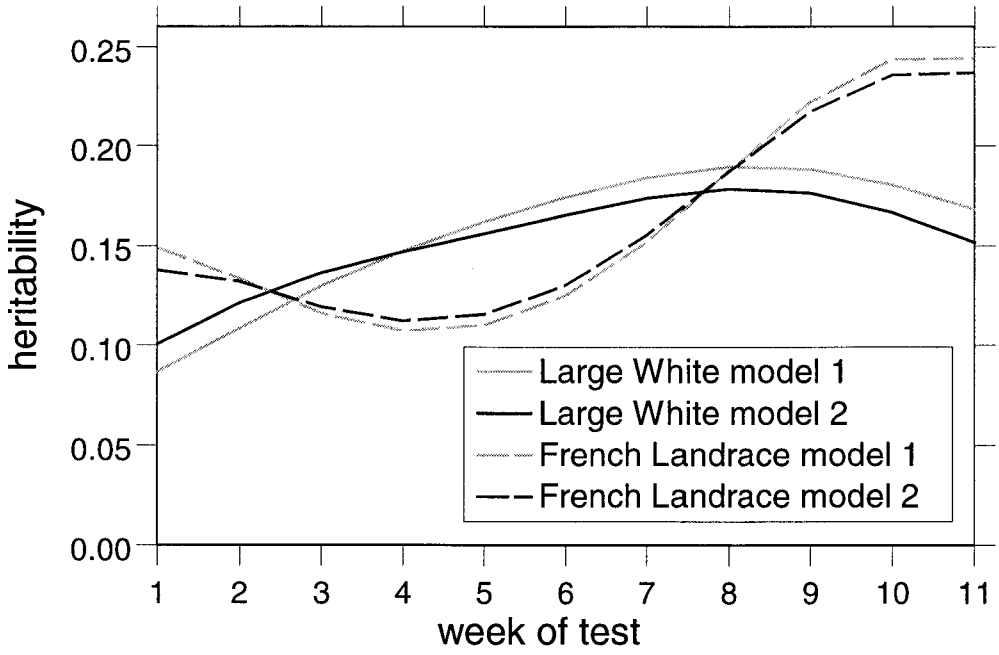


Figure 6: Course of heritabilities for daily feed intake (kg) of Large White and French Landrace growing pigs for models 1 and 2.

Our heritability estimates for weekly means of daily feed intake are slightly lower than the values found by Von Felde *et al.* [25]. Heritability estimates of Hall *et al.* [11] for four biweekly means of daily feed intake lay in between the ones found here and those of Von Felde *et al.* [25]. They are comparable to our results for Large White pigs, if the reduction of the residual variance due to biweekly means (compared to weekly means) is accounted for. The estimate of

de Haer and de Vries [3] for heritability of average daily feed intake lies in the middle range of our estimates for weekly means of daily feed intake, while estimates from other studies are higher [6, 11, 17, 25].

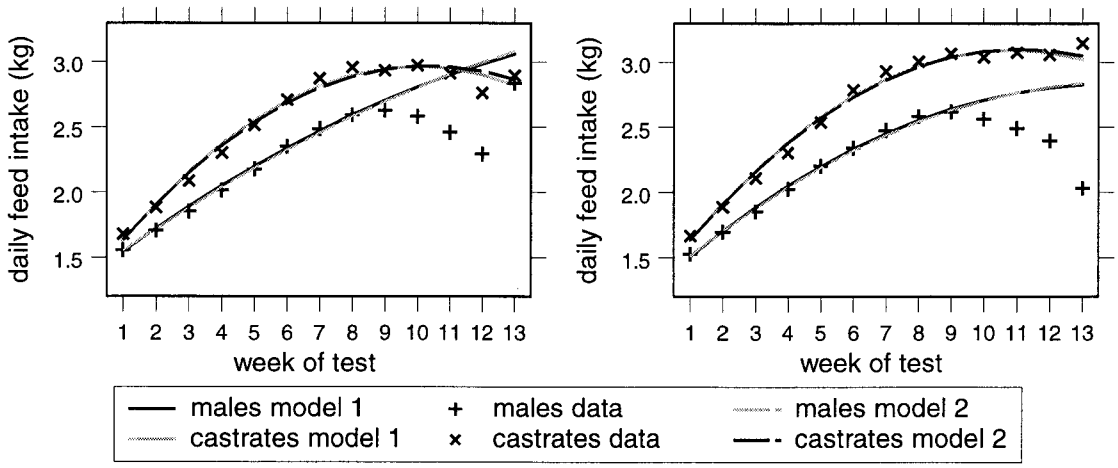


Figure 7: Course of weekly means of daily feed intake (kg) for males and castrates of Large White (left panel) and French Landrace (right panel) growing pigs estimated with models 1 and 2, as well as from data of each test week separately.

3.4. Model fit

Estimates of feed intake curves for fixed effects of sex (Figure 7) are almost identical for both models. For males, fit of polynomials with estimates from single test weeks is very good for the first eight test weeks, while differences get more pronounced as number of animals with records decreases. For castrates, fit of polynomials is better in late test weeks than for males. As castrates grow slower on average than males, a higher proportion of castrates has records in late test weeks, which leads to a better fit of polynomials in late test weeks. Phenotypic variances of weekly means of daily feed intake (Figure 8) are very similar for both methods of correcting data for fixed effects. Therefore only results of the analysis of variance are shown, using a fixed effects model for each test week separately. As expected from the model, estimates from model 2 are closer to estimates from data corrected for fixed effects (sex and batch) for the

first eight weeks of the testing period than estimates from model 1. For the remaining five weeks estimates from model 1 are better, except for the last week of French Landrace. Differences between breeds in phenotypic variance estimated from corrected data for last test weeks occurred only by chance, as a few castrated French Landrace pigs with big differences in weekly means of daily feed intake happened to be paired in two batches. Generally, model estimates of phenotypic variance are too high for later test weeks, where the number of animals with records is reduced (Table I). This supports that polynomials fitted to feed intake records of fast growing pigs may be inaccurate after they finished the test and therefore cause overestimated daily variances for late weeks (see section 3.3 and Figures 4 and 5).

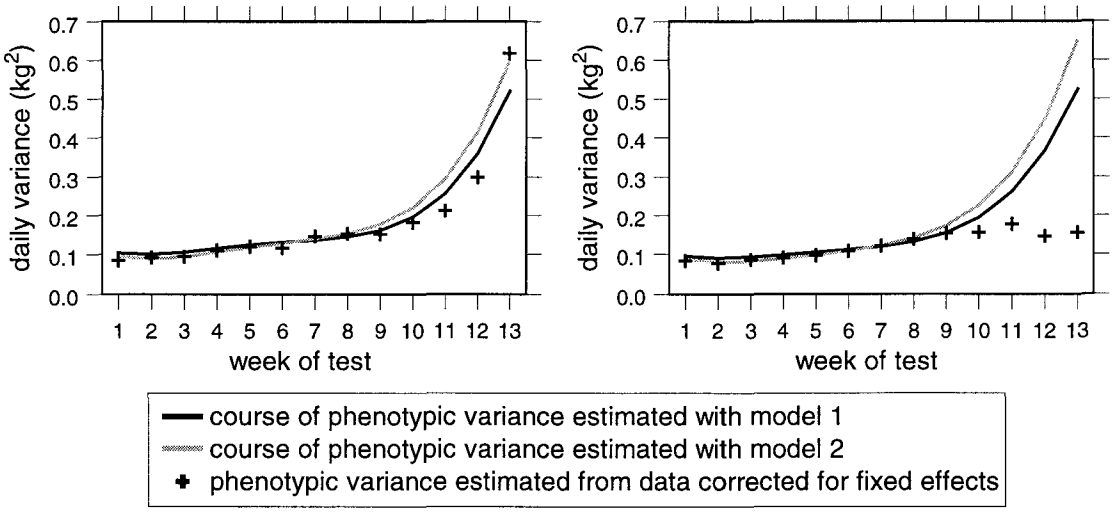


Figure 8: Course of phenotypic variance for weekly means of daily feed intake (kg) of Large White (left panel) and French Landrace (right panel) growing pigs estimated with models 1 and 2, as well as from data corrected for fixed effects (sex and batch) of each test week separately.

For both breeds curves of residual variances of models 1 and 2 intersect between weeks 6 and 7 (Figures 4 and 5). The constant residual variance in model 1 is likely to overestimate the true residual variance in the first and to underestimate it in the second half of the testing period. The quadratic polynomial used to fit the natural logarithm of the residual variance of each test day in model 2

(equation (3)), results in an almost perfect fit of phenotypic test day variance for the first eight weeks of the testing period (Figure 8). Afterwards, the phenotypic test day variance estimated with model 1 fits estimates from data of single test days corrected for fixed effects better. This is just because in late test weeks the underestimated residual variance of model 1 partly compensates for the overestimated daily variance due to random regression coefficients, which is even higher with model 1 than with model 2 (Figures 4 and 5). Therefore model 2 is preferred over model 1.

3.5. Genetic eigenfunctions and eigenvalues

Any conceivable evolutionary change in a populations mean feed intake curve can be written in terms of a weighted sum of the eigenfunctions. The rate at which a population will evolve from its current mean feed intake curve to some new curve favoured by selection is determined by the eigenvalues associated with eigenfunctions responsible for that change. A large eigenvalue indicates that a change corresponding to that eigenfunction will happen rapidly, while a small eigenvalue indicates that the change will be slow [15].

Eigenfunctions calculated from estimates of genetic (co)variance matrices of random regression coefficients do not differ much between models and are also very similar for the two breeds (Figure 9). Between 83 and 90 % of the genetic variance for the course of daily feed intake is explained by the first eigenfunction, without change of signs but increasing difference from zero during the testing period. This means that selection in one direction at any time during the testing period will cause a response into the same direction over the whole period, which would be bigger for last than for first weeks of the testing period. The second eigenfunction changes signs shortly after nine weeks of the testing period, which is when fastest growing pigs already reached the desired slaughter weight. Its response to selection would be bigger in the beginning than towards the end of the testing period, while the opposite applied to the first eigenfunction. The third eigenfunction changes signs earlier in the testing period,

but explains less than one percent of the variance in feed intake curves, which is negligible. Selection for higher feed intake in the beginning of the testing period, and constant or lower feed intake towards the end, would involve the second (for increasing feed intake in the beginning), as well as the first eigenfunction (for decreasing feed intake towards the end of the testing period). Much more weight would have to be placed on the second eigenfunction, as its associated eigenvalue is much smaller than the one of the first eigenfunction. Changing feed intake curves by selection in the desired way thus seems to be difficult, although not impossible.

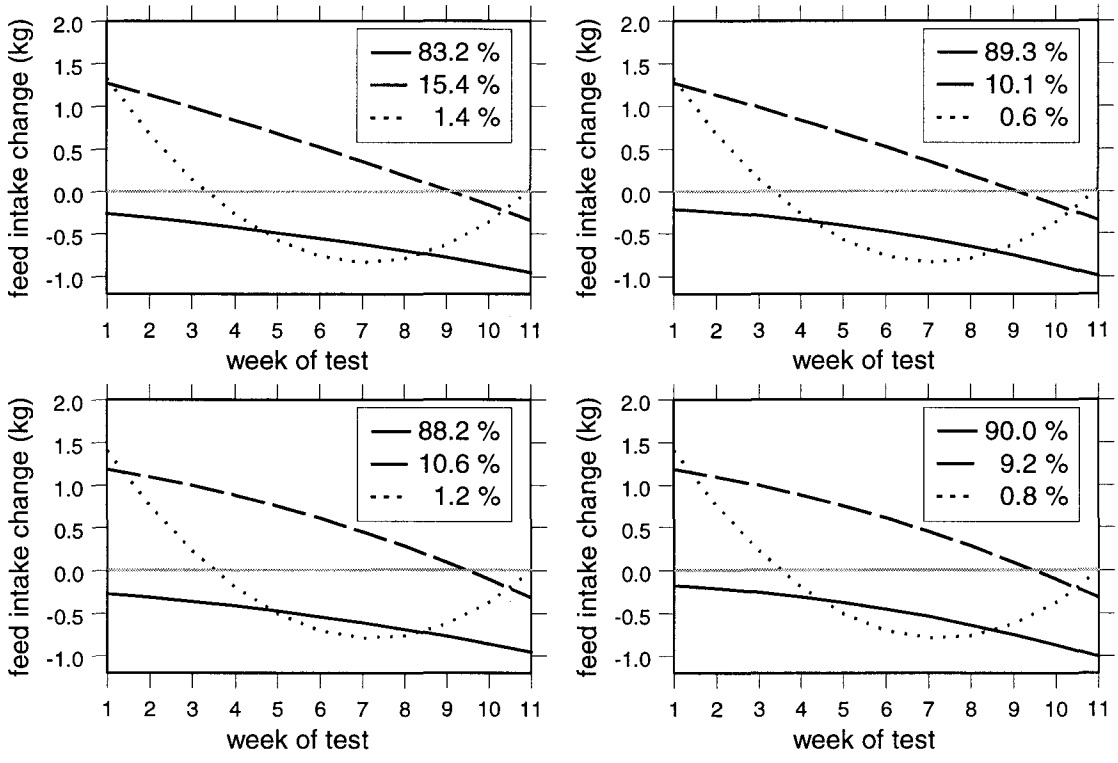


Figure 9: Eigenfunctions for daily feed intake (kg) of Large White (left panels) and French Landrace (right panels) growing pigs for models 1 (upper panels) and 2 (lower panels). Eigenvalues transformed to a percent scale (legend) indicate relative importance of corresponding eigenfunctions.

4. Conclusions

Random regression coefficients provide more information on daily feed intake of growing fattening pigs than a simple mean over the whole testing period. The amount of information is comparable to a multivariate analysis of weekly means of feed intake per day, taken over the whole testing period. The advantage of the random regression model is, that fewer parameters (traits) are needed to describe this information. But it is not sure, whether this additional information can be used to improve efficiency of lean growth. Flat posterior distributions of genetic correlations indicate, that information on genetic regression coefficients (especially linear and quadratic) seems to be limited in the data. This may be because the number of animals with records was quite low compared to the high number of levels of genetic effects to be estimated. This lack of information and the complexity of the random regression model seem to be the main reasons for the slow mixing of Gibbs chains of genetic (co)variances. From heritabilities of random regression coefficients of feed intake curves we conclude, that changes of the overall level are easier to achieve than changes of slope or inflexion of feed intake curves. Genetic eigenfunctions also reveal that an improvement of feed efficiency by selection on the shape of feed intake curves seems difficult. For a final assessment of possible routes of improvement of efficiency of lean growth by means of selection on feed intake curve parameters, correlations with other traits might be helpful, such as daily gain, feed conversion ratio and carcass traits. For this further research is needed.

References

[1] Andersen S., Pedersen B., Growth and food intake curves for group housed gilts and castrated male pigs, Anim. Sci. 63 (1996) 457-464.

[2] Cole D.J.A., Chadd S.A., Voluntary food intake of growing pigs, in: Forbes J.M., Varley M.A., Laurence T.L.J. (Eds.), The voluntary food intake of pigs, Occasional Publication of the British Society of Animal Production 13, 1989, pp. 61-70.

[3] De Haer L.C.M., de Vries A.G., Effects of genotype and sex on the feed intake pattern of group housed growing pigs, Livest. Prod. Sci. 36 (1993) 223-232.

[4] De Vries A.G., Kanis E., A growth model to estimate economic values for food intake capacity in pigs, Anim. Prod. 55 (1992) 241-246.

[5] De Vries A.G., Kanis E., Swine breeding goals, short- and long-term considerations, in: Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada, August 7-12, vol. 17, 1994, pp. 390-397.

[6] Eissen J.J., Breeding for feed intake capacity in pigs, doctoral thesis, Animal Breeding and Genetics Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands, 2000.

[7] Fowler V.R., Biological advances towards genetic improvement in pigs, in: Proc. 3rd World Congress on Genetics Applied to Livestock Production, Lincoln, Nebraska, USA, July 16-22, vol. XI, Genetics of Reproduction, Lactation, Growth, Adaptation, Disease, and Parasite Resistance, 1986, pp. 345-354.

[8] Gelman A., Carlin J.B., Stern H.S., Rubin D.B., Bayesian data analysis, Chapman & Hall, London, 1995.

[9] Geman S., Geman D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on pattern analysis and machine intelligence, PAMI 6 (1984) 721-741.

[10] Geyer C.J., Practical Markov chain Monte Carlo, Stat. Sci. 7 (1992) 473-511.

[11] Hall A.D., Hill W.G., Bampton P.R., Webb A.J., Genetic and phenotypic parameter estimates for feeding pattern and performance test traits in pigs, Anim. Sci. 68 (1999) 43-48.

[12] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97-109.

[13] Johnson V.E., Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths, J. Am. Stat. Assoc. 91 (1996) 154-166.

[14] Kanis E., de Vries A.G., Optimization of selection for food intake capacity in pigs, Anim. Prod. 55 (1992) 247-255.

[15] Kirkpatrick M., Lofsvold D., Bulmer M., Analysis of the inheritance, selection and evolution of growth trajectories, Genetics124 (1990) 979-993.

[16] Labroue F., Aspects génétiques du comportement alimentaire chez le porc en croissance, Thèse de Doctorat, Ensar, Rennes, 1996.

[17] Labroue F., Gueblez R., Sellier P., Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace pigs, Genet. Sel. Evol. 29 (1997) 451-468.

[18] MathSoft, S-Plus, Version 3.4 for UNIX, Supplement, Data Analysis Products Division, MathSoft, Incorporation, Seattle, Washington, USA, 1996.

[19] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller W., Equation of state calculations by fast computing machines, J. Chem. Phys.21 (1953) 1087-1092.

[20] Raftery A.E., Lewis S., How many iterations in the Gibbs sampler?, in: Bernardo J.M., Berger J.O., Dawid A.P., Smith A.F.M. (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, U.K., 1992, pp. 763-773.

[21] Roberts G.O., Markov chain concepts related to sampling algorithms, in: Gilks W.R., Richardson S., Spiegelhalter D.J. (Eds.), Markov chain Monte Carlo in Practice, Chapman & Hall, London, 1996, pp. 45-57.

[22] Schaeffer L.R., Dekkers J.C.M., Random regression in animal models for test-day production in dairy cattle, in: Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada, August 7-12, vol. 18, 1994, pp. 443-446.

[23] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in Threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229-249.

[24] Stahel W.A., Statistische Datenanalyse, Eine Einführung für Naturwissenschaftler, Vieweg, Braunschweig/Wiesbaden, 1995.

[25] Von Felde A., Roehe R., Looft H., Kalm E., Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars, Livest. Prod. Sci. 47 (1996) 11-22.

[26] Webb A.J., Genetic control of growth, composition, appetite and feed utilisation: non-ruminants, Options for genetic change, in: Proc. 3rd World Congress on Genetics Applied to Livestock Production, Lincoln, Nebraska, USA, July 16-22, vol. XI, Genetics of Reproduction, Lactation, Growth, Adaptation, Disease, and Parasite Resistance, 1986, pp. 337-344.

[27] Webb A.J., Future challenges in pig genetics, Animal Breeding Abstracts 63 (1995) 731-736.


Seite Leer /
Blank leaf

Chapter 5:

Multiple trait model combining random regressions for daily feed intake with single measured performance traits of growing pigs

Urs Schnyder[a], Andreas Hofer[a], Florence Labroue[b], Niklaus Künzi[a]

[a] Institute of Animal Science,

Swiss Federal Institute of Technology (ETH),

CH-8092 Zürich, Switzerland

[b] Institut technique du porc,

La Motte au Vicomte, BP 3,

35651 Le Rheu cedex, France

This chapter has been accepted for publication and is reprinted with permission:

Genetics Selection Evolution, Volume 34, Number 1, January/February 2002, pp. xx-xx.

Abstract

A random regression model for daily feed intake and a conventional multiple trait animal model for the four traits average daily gain on test (ADG), feed conversion ratio (FCR), carcass lean content and meat quality index were combined to analyse data from 1 449 castrated male Large White pigs performance tested in two French central testing stations in 1997. Group housed pigs fed *ad libitum* with electronic feed dispensers were tested from 35 to 100 kg live body weight. A quadratic polynomial in days on test was used as a regression function for weekly means of daily feed intake and to describe its residual variance. The same fixed (batch) and random (additive genetic, pen and individual permanent environmental) effects were used for regression coefficients of feed intake and single measured traits. Variance components were estimated by means of a Bayesian analysis using Gibbs sampling. Four Gibbs chains were run for 550 000 rounds each, from which 50 000 rounds were discarded from the burn-in period. Estimates of posterior means of covariance matrices were calculated from the remaining two million samples. Low heritabilities of linear and quadratic regression coefficients and their unfavourable genetic correlations with other performance traits reveal that altering the shape of the feed intake curve by direct or indirect selection is difficult.

1. Introduction

Electronic feeders installed in central testing stations allow for the measurement of individual daily feed intake of performance tested growing pigs. Today's pig selection programs only make use of these data by calculating average daily feed intake as a simple mean of daily feed intake records over the whole testing period. We have shown in a previous study that more information can be retained from these data with a random regression model, using a quadratic polynomial to describe the course of daily feed intake of growing fattening pigs [11]. Genetic eigenfunctions and low heritabilities of linear and quadratic random regression coefficients of daily feed intake indicate that changes of the overall level are easier to achieve than changes of slope or inflexion of feed intake curves. It therefore seems difficult to improve the efficiency of lean growth by selecting for a higher feed intake in the beginning of the fattening period while leaving the feed intake capacity at its present level towards the end [11]. Such an advantage over the use of traditional traits (average daily feed intake, average daily gain and/or the ratio of the two, i.e. feed conversion) for selection of pigs for growth performance would be necessary to justify the use of a random regression model for routine evaluations.

Correlations of random regression coefficients for feed intake with traditional single measured performance traits of growing pigs might help to judge the potential of random regression models for future pig breeding programs. To our knowledge, no attempt has been published to combine a random regression model for a trait with repeated measurements with a conventional multiple trait model for single measured traits in a joint analysis.

The objective of this study was to combine the random regression model previously used for the analysis of daily feed intake data [11] with a multiple trait model for single measured performance traits of growing pigs and to assess

possible routes of improvement of the efficiency of lean growth based on estimates of genetic and phenotypic correlations obtained from this joint analysis.

2. Material and methods

2.1. Data

1 449 castrated Large White pigs were performance tested in two French central testing stations in 1997. Growing pigs were housed in group pens equipped with one electronic feed dispenser each (Acema-48, Acemo, Pontivy, Morbihan, France), where *ad libitum* daily feed intake was recorded. Groups that were on test during the same period of time on the same testing station form a batch. There was a total of 155 groups in 13 batches. After one week of adaptation to the automatic feed dispensers, pigs were put on test with about 35 kg live body weight and slaughtered after end of test with 100 kg live body weight on average. Weekly means of feed intake per day were calculated and saved as the record of the middle day of the test week, in order to reduce the amount of data for the evaluations. Whenever records of more than one day per week were missing, all the records of this week were discarded and the weekly mean was set to missing. This resulted in records for days 4, 11, 18,..., 81, 88, 95 (Table I). Other traits included in this evaluation were average daily gain and feed conversion ratio calculated for the period between start and end of test, as well as carcass lean content and meat quality index determined after slaughtering of tested animals.

Table I: Number (n) and proportion (%) of tested animals with records for weekly means of daily feed intake by test week (or corresponding test day).

Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Day	4	11	18	25	32	39	46	53	60	67	74	81	88	95
n	1423	1444	1443	1441	1442	1435	1423	1407	1378	1213	713	225	51	3
%	98.2	99.7	99.6	99.4	99.5	99.0	98.2	97.1	95.1	83.7	49.2	17.6	3.5	0.2

2.2. Model

The following random regression model, which is a quadratic polynomial in days on test d_m was fitted to weekly means of daily feed intake records:

$$\begin{aligned} y_{ijkm} = & \; batch_{0k} + batch_{1k} * d_m + batch_{2k} * d_m^2 \\ + & \; a_{0i} + a_{1i} * d_m + a_{2i} * d_m^2 \\ + & \; p_{0j} + p_{1j} * d_m + p_{2j} * d_m^2 \\ + & \; e_{0i} + e_{1i} * d_m + e_{2i} * d_m^2 \\ + & \; \varepsilon_{ijkm} \end{aligned} \qquad (1)$$

where $batch_{nk}$ are fixed regressions for the period and station of test; a_{ni} are random regressions for animal additive genetic effects; p_{nj} and e_{ni} are random regressions for permanent environmental effects of pen and the tested individual, respectively; ε_{ijkm} is a random residual error which accounts for daily deviations of feed intake from the expected trajectory of animal i on day d_m. What is called "permanent environmental effect of the tested individual", is a residual for regression coefficients. This random regression model corresponds to the one used in a previous analysis of daily feed intake records of performance tested growing pigs [11]. Fixed regression coefficients due to the gender of the animals as well as random regression coefficients due to litter permanent environmental effects were dropped from the model, since only castrated males were tested, which usually had no litter mates on test.

Daily deviations from the estimated feed intake curve of an animal (residuals ε_{ijkm}) were assumed to be independent of each other. All the animals were assumed to have the same residual variance for feed intake on a given day on test d_m, which was modelled as follows:

$$\sigma^2_{\varepsilon_m} = \exp\left(\gamma_0 + \gamma_1 * d_m + \gamma_2 * d_m^2\right) \qquad (2)$$

In a previous analysis, this model for the residual variance proved to be better than a constant residual variance over the whole testing period [11]. Changes in the magnitude of the residual variance are mainly due to scale effects, since daily

feed intake of pigs increases with stomach and gut size during the growing period.

The model for single measured performance traits average daily gain, feed conversion ratio, carcass lean content and meat quality index contains the same fixed and random effects as for regression coefficients for weekly means of daily feed intake. Additionally, live weight at the end of test (before slaughtering) was included as a covariable for average daily gain and feed conversion ratio:

$$y_{nijk} = \beta_n * weight_i + batch_{nk} + a_{ni} + p_{nj} + e_{ni} + \varepsilon_{nijk} \qquad (3)$$

where y_{nijk} is the record for trait n of animal i in pen j and batch k. β_n is the regression of trait n on the covariable weight at end of test. For the combination of the two models, additive genetic (a_{ni}) and permanent environmental effects of pen (p_{nj}) of single measured traits (n) are assumed to be correlated with the corresponding effects for random regression coefficients for daily feed intake. Since residuals for regression coefficients are fitted explicitly as individual permanent environmental effects in the random regression model for daily feed intake, such individual permanent environmental effects (e_{ni}) were also fitted for single measured traits. Individual permanent environmental effects are assumed to be correlated among single measured traits and regression coefficients for feed intake. The residuals ε_{nijk} of single measured traits correspond to the residuals ε_{ijkm} in equation (1), which account for deviations of daily feed intake from the expected trajectory. Residuals ε_{nijk} of single measured traits are assumed to be normally distributed and independent of each other as well as from residuals of daily feed intake. The two residual terms in model (3) for single measured traits (e_{ni} and ε_{nijk}) were included to reach compatibility with the random regression model (1) for daily feed intake. Explicitly fitting individual permanent environmental effects e_{ni} in a random regression model is necessary for a proper definition of heritabilities of regression coefficients, since they play the role of residuals for these artificial traits [11]. If one desires to allow for correlations between these explicitly fitted residuals of regression coefficients and residuals

of single measured traits in a joint analysis, the only possibility is to fit individual permanent environmental effects explicitly for single measured traits also.

Normal distribution of feed intake data and single measured performance traits is assumed:

$$\mathbf{y} \mid \mathbf{b},\mathbf{a},\mathbf{p},\mathbf{e},\sigma^2_{\varepsilon_{nm}} \sim N\left\{\mathbf{Xb}+\mathbf{Za}+\mathbf{Vp}+\mathbf{We},\mathbf{I}\sigma^2_{\varepsilon_{nm}}\right\} \quad (4)$$

y is a vector containing data for all traits; **b** is a vector containing fixed effects for batch and regressions β_n on the covariable weight at end of test; **a** is the vector of additive genetic effects; **p** and **e** are vectors containing permanent environmental effects; **X**, **Z**, **V** and **W** are incidence matrices; **I** is the identity matrix and $\sigma^2_{\varepsilon_{nm}}$ is the residual variance around feed intake curves for day on test d_m, or the variance of uncorrelated residuals for single measured traits, respectively.

The following assumptions were used for the distributions of fixed and random effects:

$$\begin{aligned}
&\mathbf{b} \sim \text{constant}\\
&\mathbf{a} \mid \mathbf{A},\mathbf{G}_0 \sim N\{\mathbf{0},(\mathbf{A}\otimes\mathbf{G}_0)\}\\
&\mathbf{p} \mid \mathbf{P}_0 \sim N\{\mathbf{0},(\mathbf{I}\otimes\mathbf{P}_0)\}\\
&\mathbf{e} \mid \mathbf{E}_0 \sim N\{\mathbf{0},(\mathbf{I}\otimes\mathbf{E}_0)\}
\end{aligned} \quad (5)$$

A is the numerator relationship matrix, $\mathbf{G_0}$ is the (co)variance matrix of random additive genetic effects and $\mathbf{P_0}$ and $\mathbf{E_0}$ are (co)variance matrices for random permanent environmental effects. All these (co)variance matrices are of dimension 7 x 7 (three regression coefficients plus four single measured traits).

Informative priors with low numbers of degrees of freedom were used for the variance components. For the 7 x 7 (co)variance matrices $\mathbf{G_0}$, $\mathbf{P_0}$ and $\mathbf{E_0}$, inverse Wishart distributions with nine degrees of freedom were used. Scale parameters for inverse Wishart prior distributions (Tables II and III) were chosen such that resulting expected values of covariance matrices corresponded to our expectation. Expected values for (co)variances of feed intake regression coefficients were taken from our results of an earlier study [11], while genetic

Table II: Lower diagonal elements of the symmetric scale matrix $\mathbf{S_G}$ for the inverse Wishart prior distribution of the additive genetic covariance matrix ($\mathbf{G_0}$) between intercept, linear and quadratic regression coefficients for daily feed intake and single measured performance traits average daily gain (ADG), feed conversion ratio (FCR), carcass lean content (CLC) and meat quality index (MQI).

Trait	Intercept	Linear	Quadratic	ADG	FCR	CLC	MQI
Intercept	2.23e-2						
Linear	-3.60e-4	1.40e-5			symmetric		
Quadratic	2.90e-6	-7.00e-8	1.90e-9				
ADG	4.90	0.0	0.0	3386.0			
FCR	1.90e-3	0.0	0.0	-2.186	0.0080		
CLC	-1.562e-1	0.0	0.0	0.0	-0.0990	7.620	
MQI	1.486e-2	0.0	0.0	0.0	0.0141	-0.1451	1.105

and permanent environmental (co)variances for single measured performance traits were derived from Labroue et al [7]. Their results for average daily feed intake were used for genetic correlations between single measured traits and the intercept of feed intake curves. Priors for genetic covariances of single measured performance traits with linear and quadratic regression coefficients of daily feed intake were set to zero (Table II), since no prior information about their true value was available. For simplicity, prior values of all permanent environmental covariances of single measured traits were also set to zero (Table III). Total permanent environmental (co)variance (Table III) was divided into its components pen ($\mathbf{P_0}$) and individual ($\mathbf{E_0}$) permanent environmental (co)variance with a ratio of 1 to 9. Priors for parameters γ_0, γ_1 and γ_2, that describe the course of the residual variance $\sigma^2_{\varepsilon_m}$ for weekly means of daily feed intake, were assumed independent of each other and normally distributed with standard deviations of 1.5 (γ_0), 0.1 (γ_1) and 0.01 (γ_2). These standard deviations represent a relatively

wide range of values, that parameters γ_0, γ_1 and γ_2 might reasonably take. The same values were used in an earlier study [11], where they were chosen to express the low level of knowledge about distributions of these parameters. As the Metropolis-Hastings algorithm performed well with these values, they were not changed for the present study.

Table III: Lower diagonal elements of the symmetric scale matrix $\mathbf{S_{PE}}$ for the inverse Wishart prior distribution of the total permanent environmental covariance matrix (sum of $\mathbf{P_0}$ and $\mathbf{E_0}$) between intercept, linear and quadratic regression coefficients for daily feed intake and single measured performance traits average daily gain (ADG), feed conversion ratio (FCR), carcass lean content (CLC) and meat quality index (MQI).

Trait	Intercept	Linear	Quadratic	ADG	FCR	CLC	MQI
Intercept	3.06e-2						
Linear	-1.14e-3	1.96e-4			symmetric		
Quadratic	1.11e-5	-2.62e-6	3.97e-8				
ADG	0.0	0.0	0.0	5079.0			
FCR	0.0	0.0	0.0	0.0	0.032		
CLC	0.0	0.0	0.0	0.0	0.0	3.267	
MQI	0.0	0.0	0.0	0.0	0.0	0.0	4.418

Unlike residuals for daily feed intake in a random regression model, uncorrelated residuals for single measured traits can not be distinguished from individual permanent environmental effects. To avoid difficulties of distribution of variance between the two environmental effects of single measured traits, the residual variance $\sigma^2_{\varepsilon_n}$ was not estimated, but fixed to a value 10 000 times smaller than the expected phenotypic variance of the trait. This computational trick forced the residual variance of single measured traits to be attributed to the individual permanent environmental (co)variance matrix $\mathbf{E_0}$. This is illustrated below for two traits with repeated and single measurements, respectively. Suppose, the true

permanent environmental and residual (co)variance structures for these two traits are given by:

$$\mathbf{E}_0 = \begin{bmatrix} \sigma^2_{e_1} & \sigma_{e_{1,2}} \\ \sigma_{e_{1,2}} & \sigma^2_{e_2} \end{bmatrix} \quad ; \quad \mathbf{R} = \mathbf{I}\sigma^2_{\varepsilon_n} = \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 \\ 0 & \sigma^2_{\varepsilon_2} \end{bmatrix} \tag{6}$$

If the residual variance can be estimated for the trait with repeated measurements (trait 1) and is fixed to a small value s^2 (smaller than the true value) for the single measured trait (trait 2), above components will be estimated as:

$$\mathbf{E}_0 = \begin{bmatrix} \sigma^2_{e_1} & \sigma_{e_{1,2}} \\ \sigma_{e_{1,2}} & (\sigma^2_{e_2} + \sigma^2_{\varepsilon_2} - s^2) \end{bmatrix} \quad ; \quad \mathbf{R} = \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 \\ 0 & s^2 \end{bmatrix} \tag{7}$$

The major part of residuals of the single measured trait will thus be included in explicitly fitted permanent environmental effects, if the mixed model equations are built with these (co)variance components. As long as the value chosen for s^2 is smaller than the (unknown) true residual variance of the single measured traits, estimates of covariances in (7) will certainly be unbiased. As long as the permanent environmental correlation calculated from $\mathbf{E}_0$ in equation (7) does not reach the limits of the parameter space, even higher values than the true residual variance can be chosen for s^2. The following conditions must always hold:

$$\begin{aligned} & -1 \le \frac{\sigma_{e_{1,2}}}{\sqrt{\sigma^2_{e_1}(\sigma^2_{e_2} + \sigma^2_{\varepsilon_2} - s^2)}} \le 1 \\ \Rightarrow\ & 0 \le \frac{(\sigma_{e_{1,2}})^2}{\sigma^2_{e_1}(\sigma^2_{e_2} + \sigma^2_{\varepsilon_2} - s^2)} \le 1 \\ \Rightarrow\ & 0 < s^2 \le \sigma^2_{e_2} + \sigma^2_{\varepsilon_2} - \frac{(\sigma_{e_{1,2}})^2}{\sigma^2_{e_1}} \end{aligned} \tag{8}$$

The value zero is not allowed for s^2 because **R** in (7) has to be positive definite.

2.3. Variance components estimation

For the estimation of (co)variance components our own programs were used, applying Bayesian methodology using Gibbs sampling. The joint posterior

distribution of the parameters given the data is the product of the likelihood and the prior distributions of all parameters. From there marginal distributions are derived easily, as they only have to be known up to proportionality. This results in normal distributions for solutions of covariables, fixed and random effects and in inverse Wishart distributions for the (co)variance matrices for additive genetic and permanent environmental effects. The parameters γ_0, γ_1 and γ_2, that describe the course of the residual variance $\sigma^2_{\varepsilon_m}$, had to be sampled via a Metropolis-Hastings algorithm, as their distribution is not a standard one. A detailed description of the procedure used can be found in Schnyder *et al.* [11]. Mixed model equations (MME) were processed block wise by means of Cholesky decomposition and backsubstitution when generating new solutions in the Gibbs sampler. The data was analysed including (model 1) and excluding (model 2) "weight at end of test" as a covariable for single measured traits average daily gain and feed conversion ratio, to investigate the influence of this covariable on heritability estimates. For both models four Gibbs chains were run, with 550 000 samples each.

2.4. Post-Gibbs analysis

Burn-in was determined for all (co)variances by the method of Raftery and Lewis [10], using their Fortran program "gibbsit". Additionally, line plots of samples of (co)variance components from every 100th round of Gibbs sampling were used to check convergence of parameters to their stationary distributions. For graphical analysis of Gibbs chains the statistical software package S-Plus [8] was used. Samples from the burn-in period of each chain were discarded, and posterior means calculated from the remaining samples served as estimates of (co)variance components.

Heritabilities, genetic and phenotypic correlations were calculated from samples of (co)variance components. For regression coefficients for feed intake, the phenotypic covariance matrix is defined as the sum of additive genetic ($\mathbf{G_0}$) and

permanent environmental ($\mathbf{P_0}$, $\mathbf{E_0}$) covariance matrices [11]. For single measured traits, the residual variance is also included, i.e. the fixed value s^2 from equation (7) is added to the sum of estimated additive genetic and permanent environmental variances. For heritabilities, genetic and phenotypic correlations, effective sample size [12] and standard errors of posterior means (Monte Carlo errors) were estimated using estimates of Monte Carlo variance obtained by the method of initial monotone sequence estimator [3]. This estimator was preferred by Geyer [3] over the initial positive sequence estimator, because of making large reductions in the worst overestimates while doing little to underestimates. Each Gibbs chain was processed separately, using samples after burn-in only. Estimates of effective sample size were summed over the four Gibbs chains. The variance of an arithmetic mean of n independent values is equal to the original variance of these values divided by n (see e.g. [13]). Therefore, estimates of standard errors of overall estimates of posterior means of (co)variance components, are obtained by averaging estimates of standard errors of posterior means of the four individual chains, dividing this average by two.

(Co)variances between daily feed intake records and single measured performance traits were calculated from posterior means of (co)variance matrices of random regression coefficients for feed intake and single measured performance traits as shown in equation (9) below for additive genetic (co)variances:

$$\mathbf{C_G} = \mathbf{\Phi G_0 \Phi'} \quad ; \quad \mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I_n} \end{bmatrix} \quad ; \quad \mathbf{\Phi_m} = \begin{bmatrix} 1 & d_1 & d_1^2 \\ \vdots & \vdots & \vdots \\ 1 & d_m & d_m^2 \end{bmatrix} \tag{9}$$

$\mathbf{C_G}$ is a matrix containing genetic (co)variances between daily measurements of feed intake and single measured performance traits of dimension (m + n) rows by (m + n) columns, where m is the number of days (weeks) with measurements of feed intake and n is the number of single measured traits; $\mathbf{G_0}$ is the genetic (co)variance matrix between the 3 random regression coefficients for daily feed

intake and the n single measured traits; Φ is a matrix of $(m + n)$ rows by $(3 + n)$ columns consisting of (m by 3) matrix $\mathbf{\Phi_m}$ containing covariables for quadratic polynomials (1, day, day^2) for each day with feed intake records in the upper left corner and the (n by n) identity matrix $\mathbf{I_n}$ in the lower right corner, with zeros everywhere else. If $\mathbf{G_0}$ is split into its submatrices corresponding to (co)variances of regression coefficients for feed intake ($\mathbf{G_{1,1}}$), (co)variances of single measured traits ($\mathbf{G_{2,2}}$) and covariances between regression coefficients and single measured traits ($\mathbf{G_{1,2}}$), $\mathbf{C_G}$ can be written as follows:

$$\mathbf{C_G} = \begin{bmatrix} \mathbf{\Phi_m G_{1,1} \Phi'_m} & \mathbf{\Phi_m G_{1,2}} \\ \mathbf{G_{2,1} \Phi'_m} & \mathbf{G_{2,2}} \end{bmatrix} \tag{10}$$

Residual variances around feed intake curves were calculated for the same m days with measurements of feed intake according to equation (2), using posterior means of parameters γ_0, γ_1 and γ_2. The sum of calculated additive genetic ($\mathbf{C_G}$) and permanent environmental ($\mathbf{C_P}$ and $\mathbf{C_E}$) (co)variance matrices, with residual variances around feed intake curves ($\sigma^2_{\varepsilon_m}$) added to variances of daily feed intake and fixed residual variances ($\sigma^2_{\varepsilon_n}$) added to variances of single measured traits, yields the phenotypic (co)variance matrix $\mathbf{C}$ between weekly means of daily feed intake and single measured performance traits:

$$\mathbf{C} = \mathbf{C_G} + \mathbf{C_P} + \mathbf{C_E} + \begin{bmatrix} \mathbf{I_m}\sigma^2_{\varepsilon_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I_n}\sigma^2_{\varepsilon_n} \end{bmatrix} \tag{11}$$

From additive genetic and phenotypic (co)variance matrices, heritabilities, genetic and phenotypic correlations were calculated. Course of variances and heritability for weekly means of daily feed intake, genetic and phenotypic correlations between weekly means of daily feed intake, as well as their correlations to single measured traits, were plotted for the whole testing period.

3. Results and discussion

3.1. Behaviour of the Gibbs sampler

Burn-in periods estimated with the fortran program "gibbsit" by Raftery and Lewis [10] differed substantially between parameters, chains and specified quantiles of interest. Highest estimates were found for estimation of 50 %-quantiles of genetic variances of single measured traits carcass lean content and meat quality index. Based on these estimates and after graphically checking whether Gibbs chains had converged to a stationary distribution, 50 000 rounds of burn-in were chosen for all parameters of all chains.

Table IV: Estimates of effective sample size (sum over four Gibbs chains) for heritabilities (bold), genetic (above diagonal) and phenotypic (below diagonal) correlations of intercept, linear and quadratic regression coefficients for daily feed intake, and for single measured performance traits average daily gain (ADG), feed conversion ratio (FCR), carcass lean content (CLC) and meat quality index (MQI). Model 1 with covariable "weight at end of test" for ADG and FCR.

Trait	Intercept	Linear	Quadratic	ADG	FCR	CLC	MQI
Intercept	**88**	78	95	104	30	60	63
Linear	2280	**70**	84	71	49	62	75
Quadratic	1520	2532	**59**	67	33	50	45
ADG	513	605	433	**29**	20	34	47
FCR	1273	1388	2376	15	**21**	34	42
CLC	814	355	280	246	179	**29**	24
MQI	1645	1140	994	405	717	147	**21**

Sums of estimates of effective sample size per Gibbs chain (Tables IV and V) were very low compared to the 500 000 rounds of Gibbs sampling run after burn-in for each chain (2 000 000 samples total). Especially surprising was the

estimate of effective sample size for the phenotypic correlation between average daily gain (ADG) and feed conversion ratio (FCR) in model 1 with covariable "weight at end of test" for ADG and FCR (Table IV), which was much lower than for model 2 without covariable for ADG and FCR (Table V). A possible reason for this low estimate of effective sample size for the phenotypic correlation between ADG and FCR may be found in the special interrelations between these traits. FCR is average daily feed intake divided by ADG and ADG is weight at end of test minus weight at start, divided by the number of days on test, i.e. both traits are ratios and the covariable specified for both traits is involved too.

Table V: Estimates of effective sample size (sum over four Gibbs chains) for heritabilities (bold), genetic (above diagonal) and phenotypic (below diagonal) correlations of intercept, linear and quadratic regression coefficients for daily feed intake, and for single measured performance traits average daily gain (ADG), feed conversion ratio (FCR), carcass lean content (CLC) and meat quality index (MQI). Model 2 without covariable for ADG and FCR.

Trait	Intercept	Linear	Quadratic	ADG	FCR	CLC	MQI
Intercept	**39**	22	34	79	35	39	40
Linear	2602	**76**	46	20	39	68	75
Quadratic	3142	2418	**30**	56	26	44	63
ADG	772	835	686	**28**	22	39	17
FCR	1277	1074	1033	197	**21**	33	26
CLC	629	471	352	106	209	**29**	33
MQI	598	978	735	213	415	138	**26**

The reason for the generally slow mixing of Gibbs chains can be found in fixing the residual variance to a small value and explicitly fitting individual permanent environmental effects for single measured traits. With such a model, traits are fitted almost perfectly by the specified effects, which reduces the freedom of the

sampler to change a single effect. This was confirmed by the convergence of a Gauss-Seidel algorithm with a simulated data set. The calculations involved in Gibbs sampling of fixed and random effects are almost identical with calculations used in the Gauss-Seidel algorithm for solving the mixed model equations. Convergence of a Gauss-Seidel algorithm and mixing of the Gibbs sampler for a given model are therefore closely related. A data set was generated according to a model similar to our model for single measured traits (3), assigning relative values of 70 to the individual permanent environmental variance and 30 to the residual variance. The mixed model equations for this data were then set up using values 99.9 and 0.1 for individual p.e. and residual variances, respectively, and solved using the Gauss-Seidel algorithm. The solutions were the same as for the mixed model equations set up using the true values for variance components, but it took many more rounds to reach the convergence criteria. On the contrary, convergence was much faster if individual permanent environmental effects were not fitted explicitly, but only taken account of by assigning a value of 100 to the residual variance, i.e. the sum of the true individual p.e. and residual variances. Such a parameterisation was used by Meyer *et al.* [9] for a joint analysis of two traits with single and repeated measurements, respectively. This would certainly also improve the mixing of the Gibbs sampler for our single measured traits, but does not allow for residual correlations between random regression coefficients and single measured performance traits, as explicitly fitting individual permanent environmental effects for regression coefficients is necessary for a proper definition of heritability for these artificial traits [11]. Fixing residual variances to higher values than the ones used in this study would already improve mixing of the Gibbs sampler. One needs to make sure, though, that estimates of individual permanent environmental covariances are not affected by the choice of fixed residual variances (see equation (8)).

The following strategy is recommended for the analysis of a random regression model combined with single measured traits:

1) run a short Gibbs chain with the residual variance of single measured traits fixed to a small value (s^2 ~ 1-10°% of phenotypic variance) to get an indication of the distribution of variance among effects;

2) if necessary, adjust s^2 based on individual permanent environmental correlations (equation (8), new s^2 higher if correlation close to zero and lower if close to (–1) or 1);

3) then run the Gibbs sampler for as many rounds as needed for the desired accuracy of estimates.

Table VI: Averages (μ) and standard deviations (s.d.) over all tested animals for intercept, linear and quadratic regression coefficients of daily feed intake (coefficients fitted to records of each animal separately) and for single measured performance traits average daily gain (ADG), feed conversion ratio (FCR), carcass lean content (CLC) and meat quality index (MQI), together with estimates of posterior mean of phenotypic standard deviations from the two models with (σ_{ph} 1) and without (σ_{ph} 2) covariable "weight at end of test" for ADG and FCR.

Trait	Intercept	Linear	Quadratic	ADG	FCR	CLC	MQI
Unit	kg	kg/day	kg/day^2	g	kg/kg	%	-
μ	1.344	3.60e-02	-1.47e-04	851.74	2.918	56.67	10.734
s.d.	0.364	2.32e-02	3.33e-04	87.85	0.234	3.37	2.565
σ_{ph} 1	0.250	1.48e-02	1.94e-04	67.82	0.274	2.64	2.449
σ_{ph} 2	0.256	1.49e-02	1.96e-04	83.56	0.217	2.64	2.447

Table VI compares model estimates of phenotypic standard deviations (Gibbs posterior means) with a simple estimate of standard deviation from the raw data (not corrected for fixed effects). For regression coefficients of daily feed intake raw data estimates were obtained by first fitting a quadratic polynomial to feed intake records of each animal separately and then treating the resulting regression

coefficients like single measured traits. Mean values (Table VI) of intercept and linear regression coefficients for daily feed intake are positive, while it is negative for the quadratic regression coefficient. Values for the linear and especially the quadratic regression coefficient are small, because they are multiplied with the day of test and squared day of test, respectively, to yield kilograms of daily feed intake. When comparing estimates of phenotypic standard deviations in Table VI, another problem in the analysis with covariable weight at end of test included in the model was discovered. Like fixed effects, covariables are expected to reduce the variance of random effects. Therefore, estimates of phenotypic standard deviation of ADG and FCR were expected to be smaller for model 1 than for model 2. This was the case for ADG, but the estimate of phenotypic standard deviation of FCR obtained with model 1 was higher than estimates from both, model 2 and raw data. Instead of reducing variances of random effects, fitting "weight at end of test" as a covariable for FCR seemed to introduce additional variance. The reason for this erratic behaviour of the Gibbs sampler was found in fixing residual variances of single measured traits to a very small value and explicitly fitting residuals as individual permanent environmental effects. Apparently, the Gibbs sampling algorithm was not able to react appropriately if too high values were sampled for the solution β_n of the covariable "weight at end of test" for FCR. Explicitly fitted individual permanent environmental effects must have incorporated the changes of residuals caused by the sample for β_n. Since this had no influence on the fixed residual variance used to set up the mixed model equations, the value for β_n was not forced to be reduced in the next round of Gibbs sampling. For model 2 without covariable "weight at end of test" for ADG and FCR, no such erratic effects occurred. Cross-classified fixed and random effects seem to be less affected by the missing feedback from the fixed residual variance, since the impact of a change in the solution for one effect on the resulting "phenotypic fit" is much smaller than for covariables. However, fixing the residual variance of single measured traits to a very small value had an impact on the mixing of the Gibbs

chain for this model too. In the following, only results from model 2 (without covariable "weight at end of test" for ADG and FCR) will be reported, since estimates of heritabilities and correlations were influenced by the erratic behaviour of the Gibbs sampler with model 1.

Table VII: Estimates of posterior means of heritabilities (bold), genetic (above diagonal) and phenotypic (below diagonal) correlations of intercept, linear and quadratic regression coefficients for daily feed intake, and for single measured performance traits average daily gain (ADG), feed conversion ratio (FCR), carcass lean content (CLC) and meat quality index (MQI).

Trait	Intercept	Linear	Quadratic	ADG	FCR	CLC	MQI
Intercept	**0.32**	-0.02	0.83	0.82	0.50	-0.33	-0.04
Linear	-0.40	**0.06**	-0.35	0.38	0.48	-0.55	0.57
Quadratic	0.28	-0.91	**0.03**	0.63	0.16	0.13	-0.24
ADG	0.30	0.29	-0.08	**0.45**	0.33	-0.28	0.29
FCR	0.25	0.11	-0.10	-0.34	**0.21**	-0.65	0.04
CLC	-0.13	-0.24	0.13	-0.09	-0.44	**0.79**	-0.27
MQI	0.01	0.03	0.01	0.02	0.08	-0.11	**0.25**

3.2. **Heritabilities and correlations**

3.2.1 Feed intake curve parameters

The estimate of 0.32 for the heritability of the intercept regression coefficient of daily feed intake (Table VII) is higher than what we have found in an earlier study [11], and is identical with the estimate found by Eissen [2] in a two step approach. Heritabilities for linear and quadratic regression coefficients are in the same range as reported earlier. Phenotypic correlations are very similar to the ones found earlier, but genetic correlations are different (Table VII). The genetic correlation between the intercept and the quadratic regression coefficient is higher than reported earlier for another set of Large White data [11], while the

genetic correlation between linear and quadratic regression coefficients is lower. Genetic correlations among regression coefficients (Table VII) indicate that selection for a higher intercept might lead to flatter feed intake curves. But as heritabilities of linear and quadratic regression coefficients are low, indirect selection responses are expected to be small. This confirms that it is easier to change the overall level than the shape of feed intake curves.

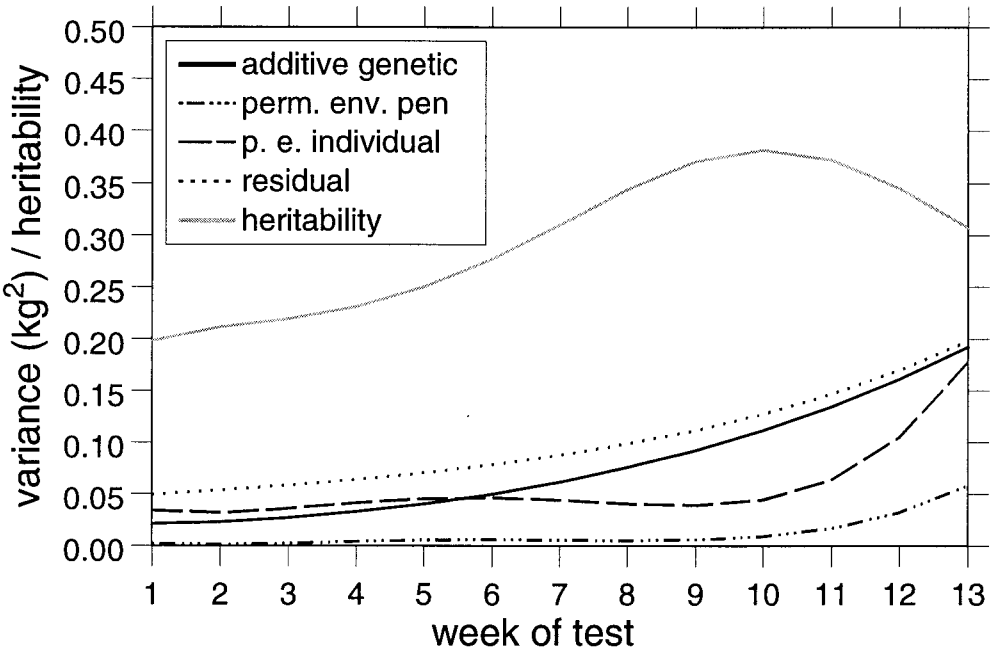


Figure 1: Course of variances and heritability for weekly means of daily feed intake (kg) of Large White growing pigs.

3.2.2 Weekly means of daily feed intake

Figure 1 shows the course of heritability, additive genetic, permanent environmental and residual variances for weekly means of daily feed intake. Week 14 is not shown, as only three animals had records in this last week of test (Table I). Course of variances is similar to what we found earlier for Large White and French Landrace pigs [11], but with increased additive genetic and reduced permanent environmental variance. Consequently, heritability estimates are also higher (Figure 1) than in our previous study [11]. Heritability for weekly means of daily feed intake increases from 0.20 in the first week of the testing period to

0.38 in week 10 (Figure 1), which is in the range of values reported by other authors [4, 5, 6, 14]. Because of the relatively high variation around feed intake curves, the heritability for weekly means of daily feed intake is lower in the first seven weeks of the testing period (Figure 1) than the heritability of the intercept regression coefficient (Table VII), which should represent a very similar information. Selection for higher feed intake in the beginning of the testing period should thus rather be based on the intercept regression coefficient than on weekly means of daily feed intake of early test weeks.

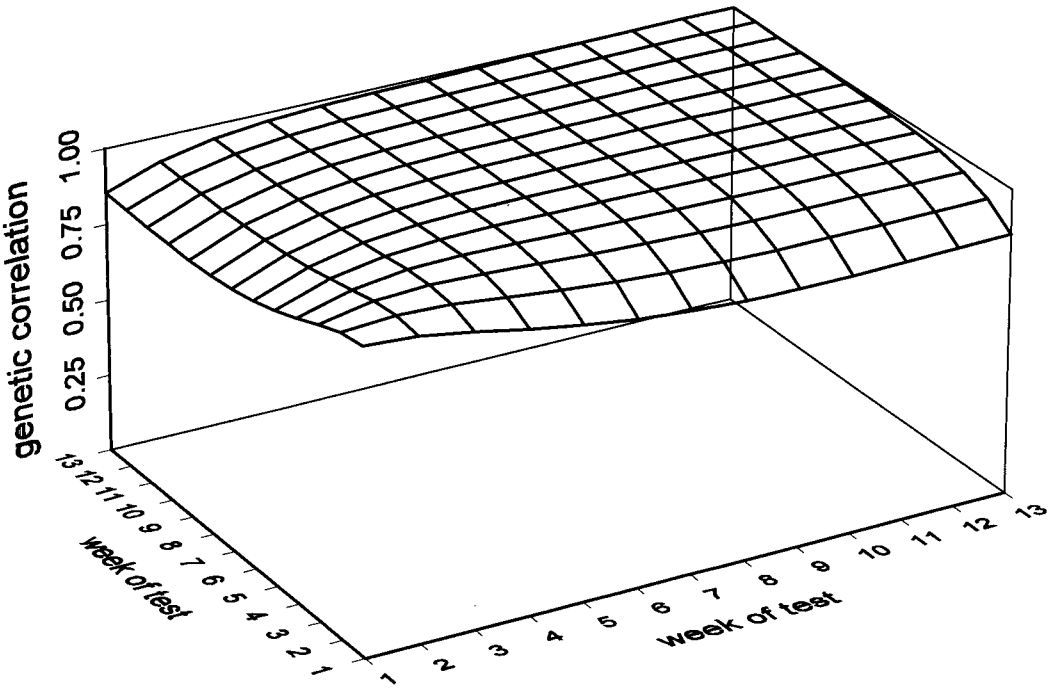


Figure 2: Genetic correlations between weekly means of daily feed intake (kg) of Large White growing pigs.

Genetic correlations between weekly means of daily feed intake (Figure 2) are very high. Lowest estimates were found for genetic correlations between feed intake in week 1 and feed intake in the second half of the testing period, which were still higher than 0.8. These values are higher than estimates of Hall *et al.* [5], who found values between 0.61 and 0.99 using a covariance function model for weekly means of daily feed intake of pigs tested between 45 and 95 kg live

body weight. Estimates of genetic correlations of weekly means of daily feed intake from other studies [6, 14], using conventional multiple trait models, are also lower than our estimates from a random regression model. These high genetic correlations indicate that selection on daily feed intake at any point during the testing period will result in a similar response over the whole period. Phenotypic correlations between weekly means of daily feed intake (Figure 3) are substantially lower than genetic correlations. Because of the influence of the residual variance around feed intake curves, (hypothetical) repeated measures of feed intake for the same test week and the same animal need not be the same.

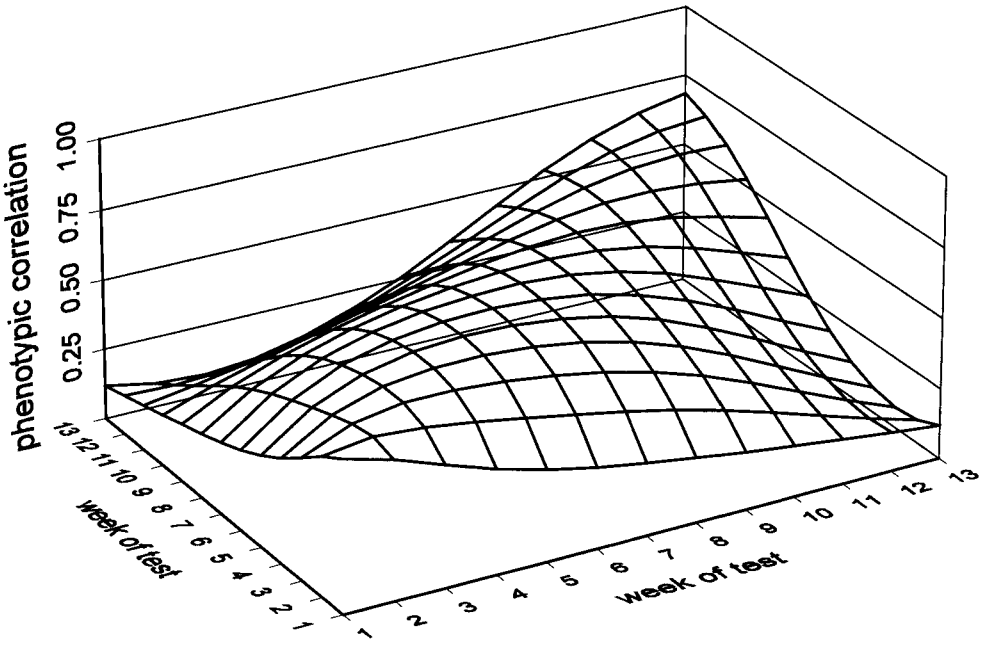


Figure 3: Phenotypic correlations and repeatability (on the diagonal) for weekly means of daily feed intake (kg) of Large White growing pigs.

The "phenotypic correlations" between records of weekly means of daily feed intake of the same test week, shown on the diagonal in Figure 3, thus represent the repeatability for weekly means of daily feed intake (variance explained by random regression coefficients divided by the total variance). Compared to estimates from regression coefficients alone (based on upper left part of $(\mathbf{C_G}+\mathbf{C_P}+\mathbf{C_E})$ in equation (11) only, without residual variances), phenotypic

correlations between feed intake records of different test weeks are also reduced due to residual variances. For consecutive test weeks, our estimates are in the same range as those of Hall *et al.* [5], while they are lower for test weeks further apart. Labroue [6] found similar estimates of phenotypic correlations between weekly means of daily feed intake with a multiple trait model, while Von Felde *et al.* [14] estimated higher phenotypic correlations between records in the second part of the testing period.

3.2.3 Single measured performance traits

Heritability estimates (Table VII) for single measured performance traits are very similar to those found by Labroue *et al.* [7] for Large White pigs. For model 1 with "weight at end of test" included as a covariable for ADG and FCR, lower heritabilities were estimated for ADG (0.33) and FCR (0.09).

Phenotypic correlations between single measured performance traits (Table VII) lie in the range of values found in literature [1, 2, 4, 7, 14]. Estimates of genetic correlations (Table VII) between CLC and other traits are similar to those found by Labroue *et al.* [7], while substantial differences (opposite signs) were found for genetic correlations between ADG, FCR and MQI. Eissen [2] estimated genetic correlations between ADG, FCR and CLC, which are similar to our result (Table VII). Estimates of Hall *et al.* [4] (ADG-FCR) and Von Felde et. al [14] (ADG-FCR, ADG-CLC) are in the same range as those of Labroue *et al.* [7]. Genetic correlations between MQI and other traits were closer to those reported by Labroue *et al.* [7] for French Landrace pigs than for Large Whites.

3.2.4 Correlations between feed intake parameters and single measured performance traits

Estimates of phenotypic correlations between regression coefficients for daily feed intake and single measured performance traits varied between -0.24 and 0.30 (Table VII). Intercept and linear regression coefficients show positive phenotypic correlations with ADG and FCR and negative phenotypic correlations

with CLC, which are similar to those found by Eissen [2]. Phenotypic correlations of the quadratic regression coefficient with these traits have opposite signs and for MQI phenotypic correlations to regression coefficients for daily feed intake are essentially zero. Genetic correlations of all three regression coefficients with ADG and FCR were positive, while genetic correlations of CLC with intercept and linear regression coefficients were found to be negative. Except for his estimate of 0.25 for the genetic correlation between CLC and the intercept regression coefficient of a linear fit to daily feed intake records, Eissen [2] estimated similar genetic correlations for intercept and linear regression coefficients with these performance traits. Because of this difference (negative correlation with CLC) we are not as optimistic as Eissen [2] about possible benefits of the intercept regression coefficient for selection. The estimate of genetic correlation between MQI and the linear regression coefficient was high and positive, while low negative values were found for other regression coefficients. Labroue *et al.* [7] estimated genetic correlations for average daily feed intake and MQI of 0.00 for Large White pigs and 0.21 for French Landrace.

Genetic and phenotypic correlations between single measured performance traits and feed intake regression coefficients result in almost constant genetic (Figure 4) and phenotypic (Figure 5) correlations between performance traits and weekly means of daily feed intake over the whole testing period. Phenotypic and genetic correlations are comparable to values reported in the literature for phenotypic and genetic correlations between average daily feed intake and other performance traits [1, 2, 4, 7, 14]. While phenotypic correlations are situated at the lower end of the range of values reported, genetic correlations tend to be slightly higher. Eissen [2] and Hall *et al.* [4] reported genetic correlations between average daily feed intake and feed conversion ratio similar to our results, while Labroue *et al.* [7] and Von Felde *et al.* [14] estimated genetic correlations close to zero.

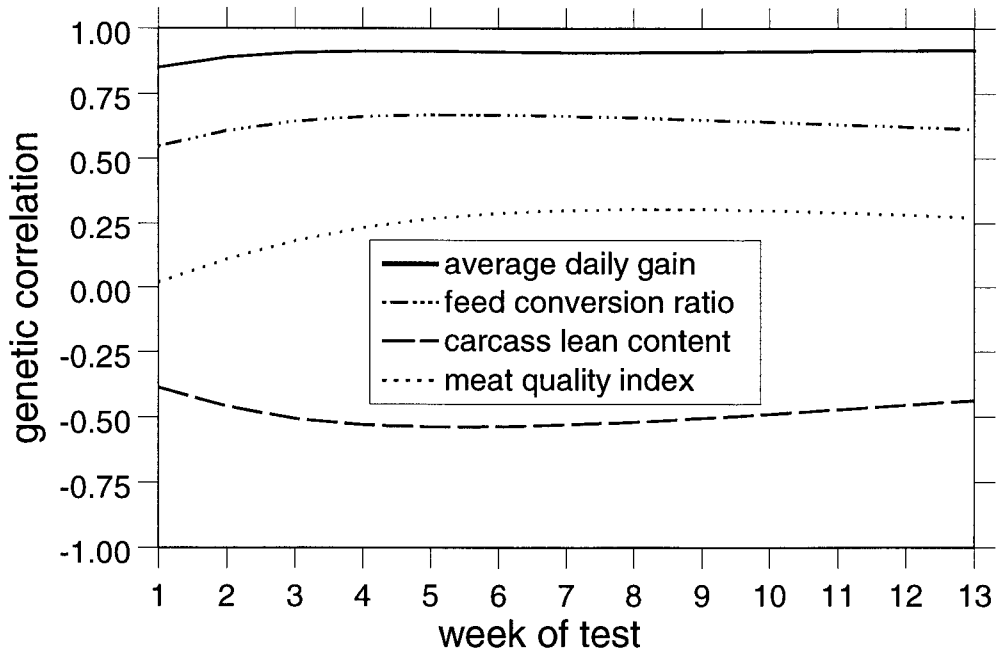


Figure 4: Course of genetic correlations of single measured performance traits with weekly means of daily feed intake (kg) of Large White growing pigs.

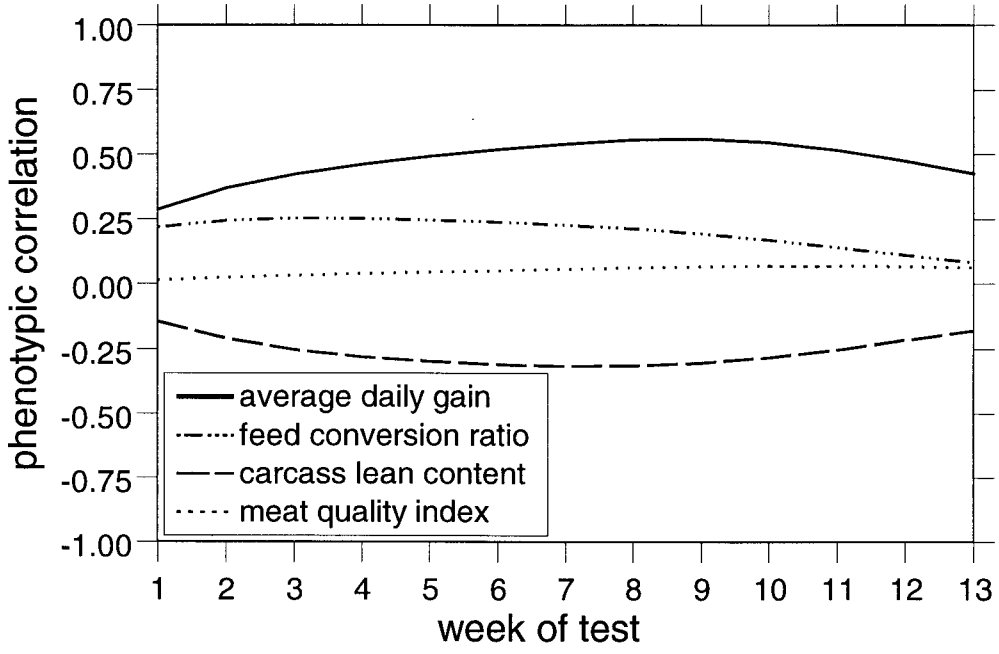


Figure 5: Course of phenotypic correlations of single measured performance traits with weekly means of daily feed intake (kg) of Large White growing pigs.

Selection for higher average daily gain and improved meat quality is expected to result in a higher feed intake over the whole testing period, while selection for improved feed conversion (lower FCR) and leaner carcasses is expected to reduce feed intake over the whole testing period. No big differences in the magnitude of these changes were found during the testing period.

Table VIII: Estimates of standard errors of posterior means (Monte Carlo errors) for heritabilities (bold), genetic (above diagonal) and phenotypic (below diagonal) correlations of intercept, linear and quadratic regression coefficients for daily feed intake, and for single measured performance traits average daily gain (ADG), feed conversion ratio (FCR), carcass lean content (CLC) and meat quality index (MQI).

Trait	Intercept	Linear	Quadratic	ADG	FCR	CLC	MQI
Intercept	**0.039**	0.047	0.036	0.047	0.102	0.052	0.030
Linear	0.001	**0.002**	0.033	0.083	0.033	0.025	0.009
Quadratic	0.003	0.000	**0.004**	0.105	0.084	0.040	0.052
ADG	0.008	0.006	0.007	**0.073**	0.103	0.065	0.048
FCR	0.004	0.001	0.002	0.004	**0.011**	0.017	0.024
CLC	0.002	0.001	0.000	0.005	0.001	**0.046**	0.025
MQI	0.002	0.001	0.001	0.001	0.002	0.003	**0.011**

3.3. Monte Carlo errors

Estimates of Monte Carlo errors, i.e. standard errors of posterior means (Table VIII) were quite low compared to low estimates of effective sample size (Table V). This is due to the high number of samples (2 000 000) included in these posterior means. Generally, estimates of standard error of posterior means are lower for phenotypic than for genetic correlations. This is partly due to higher estimates of effective sample size (better mixing of the Gibbs chain), but also because the interval of values visited by the sampler was quite narrow for phenotypic correlations compared to genetic correlations. Despite the fact that

estimates of Monte Carlo error (standard deviation of posterior mean) are very low, estimates of heritabilities and correlations should be taken carefully due to high autocorrelations between consecutive samples, which result in low estimates of effective sample size (Table V).

4. Conclusions

Explicitly fitting individual permanent environmental effects together with fixing residual variances for single measured traits is a possibility to allow for residual correlations between random regression coefficients and single measured traits in a joint analysis. Estimates of (co)variance components from such models have to be analysed carefully, though, especially if covariables for single measured traits are involved. If no residual correlations between the two types of traits are required, explicitly fitting individual permanent environmental effects for regression coefficients only and allowing for residual correlations between single measured traits should be preferred.

Heritabilities of random regression coefficients of feed intake curves show that reasonable selection responses can only be expected from the intercept regression coefficient. Changes of slope or inflexion of feed intake curves by direct selection are difficult to achieve. Genetic correlations of feed intake curve parameters with other performance traits are very similar to genetic correlations of average daily feed intake with these traits. Therefore no big advantage is expected from using feed intake regression coefficients instead of average daily feed intake in selection programmes.

References

[1] De Haer L.C.M., de Vries A.G., Effects of genotype and sex on the feed intake pattern of group housed growing pigs, Livest. Prod. Sci. 36 (1993) 223-232.

[2] Eissen J.J., Breeding for feed intake capacity in pigs, doctoral thesis, Animal Breeding and Genetics Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands, 2000.

[3] Geyer C.J., Practical Markov chain Monte Carlo, Stat. Sci. 7 (1992) 473-511.

[4] Hall A.D., Hill W.G., Bampton P.R., Webb A.J., Genetic and phenotypic parameter estimates for feeding pattern and performance test traits in pigs, Anim. Sci. 68 (1999) 43-48.

[5] Hall A.D., Bampton P.R., Webb A.J., The effect of stage of growth on the genetic and phenotypic parameter estimates of food intake in pigs using a covariance function model, in: Proceedings of the British Society of Animal Science, Annual Meeting Scarborough, March 2000, p. 40.

[6] Labroue F., Aspects génétiques du comportement alimentaire chez le porc en croissance, Thèse de Doctorat, Ensar, Rennes, 1996.

[7] Labroue F., Gueblez R., Sellier P., Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace pigs, Genet. Sel. Evol. 29 (1997) 451-468.

[8] MathSoft, S-Plus, Version 3.4 for UNIX, Supplement, Data Analysis Products Division, MathSoft, Incorporation, Seattle, Washington, USA, 1996.

[9] Meyer K., Hammond K., Mackinnon M.J., Parnell P.F., Estimates of covariances between reproduction and growth in Australian beef cattle, J. Anim. Sci. 69 (1991) 3533-3543.

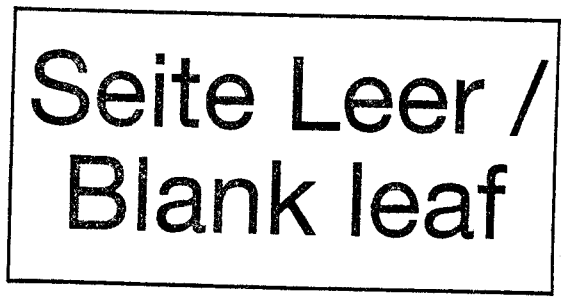
[10] Raftery A.E., Lewis S., How many iterations in the Gibbs sampler?, in: Bernardo J.M., Berger J.O., Dawid A.P., Smith A.F.M. (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, U.K., 1992, pp. 763-773.

[11] Schnyder U., Hofer A., Labroue F., Künzi N., Genetic parameters of a random regression model for daily feed intake of performance tested French Landrace and Large White growing pigs, Genet. Sel. Evol. 33 (2001) xxx-xxx.

[12] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in Threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229-249.

[13] Stahel W.A., Statistische Datenanalyse, Eine Einführung für Naturwissenschaftler, Vieweg, Braunschweig/Wiesbaden, 1995.

[14] Von Felde A., Roehe R., Looft H., Kalm E., Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars, Livest. Prod. Sci. 47 (1996) 11-22.


Seite Leer /
Blank leaf

Chapter 6:

General discussion

1. Introduction

The purpose of this study was to develop a random regression model for the genetic evaluation of daily feed intake data of growing pigs (or other traits with repeated time-dependent observations). After an introductory review of the relevant literature, first experience with feed intake (curves) and random regression models was gained with the analysis of simulated data (chapter 3) before analysing real feed intake data from growing pigs performance tested in French central testing stations (chapter 4). Finally, the random regression model for daily feed intake was combined with a multiple trait model for single measured performance traits (chapter 5). This final chapter discusses some problems and/or features of the model and methodology used in this study for the analysis of daily feed intake data, and presents some final conclusions drawn from the results of the above mentioned chapters.

2. Choice of data

Originally, it was planned to analyse feed intake records from Swiss Landrace and Large White pigs performance tested in the Swiss central testing station for pigs (MLP Sempach). During 1997, 96 electronic feeders were installed in group pens for 10 pigs each, and since November of that year all pigs were performance tested with the new feeding system. Since not enough pigs would have had daily feed intake records available when needed for this project in mid 1998, analyses were started using feed intake data from French Landrace and Large White pigs, performance tested between 1992 and 1994 in three French central testing stations (chapter 4). This data was provided by Florence Labroue (Institut téchnique du porc, Le Rheu, France) who previously analysed it for her Ph.D.-thesis [8]. When enough feed intake data from pigs tested in Switzerland was available in 1999, a problem caused by a change of diet during the test was discovered in the data. The diet was changed for all pigs in a pen after they had

eaten 80 kg of the first feed (with higher nutrient content) on average. This caused a kink in the average feed intake curve where the change of feeds occurred (after 46 days on test on average). Daily feed intake increased almost linearly before the change and showed a much higher daily increase of feed intake afterwards [9]. Immediately before the change, feed intake was most probably limited by metabolic constraints (nutrient content vs. demand), while mechanical constraints (stomach and gut size) must have been limiting afterwards. This resulted in two parts of the feed intake curve, for which separate curves should have been estimated. Alternatively, some kind of correction for nutrient content could have been applied to feed intake records from one of the two parts. The usefulness of such a correction was doubted, since limiting factors for daily feed intake were most probably different before and after the change of diet. Therefore only French data was analysed in this study. For a routine analysis using a random regression model for feed intake data of pigs performance tested in the Swiss testing station, either the feeding regimen should be changed to one feed only for the whole testing period, or only records from one of the two parts should be included.

3. Polynomial random regression model

3.1. Why ordinary quadratic polynomials?

The quadratic polynomial used in this thesis as a regression function for the description of daily feed intake of growing pigs was chosen based on results of Anderson and Pedersen [1], who showed that a cubic polynomial in days on test is sufficient for fitting cumulated feed intake. This corresponds to a quadratic polynomial for daily feed intake, as the function for daily feed intake can be written as the first derivative of the function describing cumulated feed intake. By increasing the order of polynomial fit, the residual variance could be reduced until a perfect fit is reached. Increasing the order of polynomial fit would also substantially increase the number of covariances to be estimated for each random

effect. Since daily feed intake is expected to evolve smoothly with age (almost linear) within the growing period considered, this additional effort seems not to be justified.

The reasons mentioned by Anderson and Pedersen [1] for the choice of polynomials instead of non-linear functions such as Gompertz or logistic functions also apply to our situation. It is much easier to introduce random effects and (co)variance components into linear models than into non-linear models. Furthermore, asymptotic properties of these non-linear models are not needed in our situation, where growing pigs are performance tested between 30 and 100 kg live body weight. Since animals are not taken to adult weight, maximum feed intake will probably not be reached either, and asymptotic values of non-linear models would thus be poorly estimated from the available data [1]. Additionally, usage of polynomials for random regression coefficients allows for the comparison of the resulting covariance matrices to covariance function models of Kirkpatrick *et al.* [7] and the calculation of eigenvalues and eigenfunctions to assess the possibilities for changing the feed intake curve by selection. For this purpose, covariance matrices were transformed to orthogonal polynomial scale (see chapter 4 of this thesis).

Alternatively, orthogonal polynomials could have been used directly as regression functions. An advantage of normalised orthogonal polynomials (e.g. Legendre polynomials) compared to ordinary polynomials is the faster convergence of a classical Gauss-Seidel algorithm for solving mixed model equations and potentially also better mixing of the Gibbs sampler. Since we modified the Gauss-Seidel algorithm as well as the Gibbs sampler to process all equations (intercept, linear and quadratic regression parameters) on each level of a given effect at once (blocking), using Cholesky decomposition and backsubstitution, no difference in convergence was encountered between the two types of quadratic polynomials. Nothing can be said about the convergence of higher order polynomial models, since this was not tested. Since orthogonal

polynomials showed no advantage in terms of convergence in our situation, we continued to use ordinary polynomials, as we think that they are more straightforward to understand.

3.2. Polynomials in days on test

Concerns about the validity of a random regression model using quadratic polynomials in days on test in a situation with different length of testing periods were investigated in a simulation study (chapter 3 of this thesis). As no evidence of bias in estimates of (co)variance components was found, similar models were then used to analyse real feed intake data. With real data, some differences were encountered between model estimates of phenotypic test day variances of late test weeks and phenotypic variances calculated directly from records of the corresponding test days (chapter 4). These differences are due to the reduced number of feed intake records for late test days, pertaining to slower growing pigs. Feed intake in late test weeks is probably less variable if only slower growing animals have records, than it would be if all animals had records for these test days. Another possible reason is, that polynomials fitted to daily feed intake of fast growing pigs, recorded between 30 and 100 kg live weight, are not appropriate for later test weeks, when slower growing pigs are still on test. This might lead to an overestimation of the phenotypic variance due to the model, which could not be encountered in our simulation study (chapter 3), as we used the same model for (co)variance component estimation as for generating the data. Modelling feed intake as a polynomial in weight instead of days on test would remedy the problem of different length of testing periods, as the testing period is defined between 30 and 100 kg live body weight. Unfortunately, this was not possible with the available data, since live weight was not recorded at regularly interspersed intervals, but only in the beginning and towards the end of the testing period.

3.3. Problems in feed intake records

A certain limitation of quadratic polynomials may be their susceptibility to outliers, especially if an animal has only few records (e.g. missing data due to malfunction of the feeders or loss of transponders for identification in the feeders). A possible reason for outliers in feed intake records is illness of tested animals, which may result in drastically reduced feed intake for several days and consequently also in low records of weekly means of daily feed intake. Together with missing records in the beginning of the testing period, very low records of weekly means of daily feed intake due to illness may result in atypical estimates of feed intake curve parameters, e.g. very high intercepts, negative values for linear and positive values for quadratic regression coefficients. If such problems are not recognised and corresponding records are not deleted from the data, resulting estimates of feed intake curves (quadratic polynomials) might thus be implausible. Such inaccurate estimates of feed intake curves also influence estimates of (co)variance components of regression coefficients, therefore feed intake data should be carefully checked for problems before it is used for the estimation of (co)variance components. Alternatively, a regression function with an intercept and a linear term only (first degree polynomial), as proposed by Eissen [3], may be more robust to such outliers and missing data.

For routine evaluations, the problem of missing records may be less severe, as estimating breeding values is computationally feasible without combining daily feed intake records into weekly means. Only the residual variance has to be adapted for an evaluation using records of daily feed intake directly compared to one using weekly means of daily feed intake, if these weekly means were associated with the middle day of the week when estimating (co)variance components as in this study. If an animal had less than six daily records available in a test week, the weekly mean of daily feed intake was set to missing. Animals with missing records of weekly means of daily feed intake might thus have up to five records available in the corresponding week, which could be used for

estimating breeding values in a routine evaluation. If this results in records spread over the whole testing period where records of weekly means were missing, outliers due to illness might have less impact on estimates of regression coefficients.

4. Gibbs sampling for random regression models

One main problem of the present study was the slow mixing of the Gibbs sampling algorithm used for estimation of (co)variance components. This resulted in the need for very long Gibbs chains and consequently a very high amount of CPU-time used. A possible reason for this problem may be the high degree of dependence between the parameters of a model where the same regression function is used for all fixed and random effects. New realisations of solutions for regression coefficients were sampled separately for each level of each effect, conditionally on all other model parameters. Dependence between model parameters limits the freedom of parameters updated in each round of the Gibbs sampler to explore their parameter space. This means that parameters make only very small moves in one round of Gibbs sampling, which consequently results in high autocorrelations between samples from consecutive rounds. In spite of the very large number of Gibbs samples drawn for each model, effective sample size was too small to allow for the estimation of posterior densities. As already mentioned in paragraph 3.1 of this general discussion chapter, using normalised orthogonal polynomials instead of ordinary polynomials in days on test as regression functions would not remedy the problem of slow mixing in our situation. Simultaneous sampling of different effects would probably result in better mixing, but is usually not feasible in high-dimensional parameter spaces [4]. One possibility to improve mixing of MCMC algorithms for normal linear mixed models may be by applying hierarchical centring reparametrisations, as suggested by Gelfand *et al.* [4]. This option was

not incorporated in the present study and further investigations are needed to properly assess its effectiveness.

Jamrozik and Schaeffer [6] and Rekaya *et al.* [10] used considerably shorter Gibbs chains for the estimation of (co)variance components for random regression models for test day yields of dairy cows than we used for our analyses of daily feed intake of growing pigs (chapters 4 and 5). Jakobsen [5] estimated (co)variance components for a model which combined random regressions for a continuous trait (test day yields of dairy cows) and a binary trait (disease liability). Although she used shorter Gibbs chains, her estimates of effective sample size were higher than the ones found in this study. These differences in the mixing of Gibbs chains for different random regression models may be due to differences in the complexity of the models and the structure of the data. Jamrozik and Schaeffer [6] only fitted a regression function for fixed and random animal additive genetic effects, Rekaya *et al.* [10] also for random permanent environmental effects. Both of them probably had better data structures for the estimation of additive genetic regression coefficients, as they used first lactation data from eight and thirteen consecutive years, respectively. This means, that the oldest animals with records already had daughters and granddaughters with records in the data set, which is advantageous for estimating additive genetic effects in an animal model. This was certainly not the case in the data set used for the multivariate analysis in chapter 5, since all animals were performance tested in the same year and slaughtered after the end of test, and only to a much lesser extent in chapter 4. Jakobsen [5] used a sire model, for which estimation of additive genetic effects is easier than for a comparable animal model, since more information is available per level of estimated effect. The amount of information available per level of estimated effect thus seems to influence the mixing of the Gibbs sampler as well as it influences the convergence of a Gauss-Seidel algorithm for solving the mixed model equations.

5. Feed intake and efficiency of lean growth

The reason for applying a random regression model to daily feed intake data was the question whether it is possible to change the average feed intake curve by selection to improve efficiency of lean growth. For this purpose it would be desirable to have measurements of live body weight at regularly interspersed intervals (e.g. weekly weightings) throughout the entire testing period. This would allow for modelling feed intake or time on test as a function of body weight, with the advantage of a fixed range of values for the explanatory variable (e.g. 30 to 100 kg live body weight). With such a model the relationship between body weight, feed intake and time on test could be quantified throughout the growing period, which was not possible based on the data available for this study. We therefore had to concentrate on daily feed intake records, measured by means of electronic feed dispensers on group housed growing pigs, performance tested between 30 and 100 kg live body weight in French central testing stations.

Cameron [2] suggested to improve the efficiency of lean growth by selection for improved lean tissue growth rate instead of for feed efficiency or its inverse feed conversion ratio. The advantage of such a selection strategy is, that the efficiency of lean growth is improved by increasing lean deposition rather than by reducing fat deposition and feed intake capacity. Furthermore, recording of feed intake is not needed to get lean tissue growth rate. According to Eissen [3], feed intake capacity of end product gilts is already limiting for reaching the full potential for lean growth and should therefore be increased. He also states that a higher feed intake capacity of lactating sows will be necessary if selection in dam lines for litter size results in more piglets to be nursed by a sow. Indirect selection in dam lines to increase feed intake of lactating sows via correlated traits (e.g. feed intake and weight gain during the growing period), would also result in a higher feed intake capacity of growing end product pigs [3]. Whether this is desirable depends on the relationship between feed intake capacity and optimum level of

feed intake of end product pigs, which is determined by the growth potential of these usually crossbred animals.

For improving the efficiency of lean growth, daily feed intake should be increased in the beginning of the testing period, while it should remain unchanged towards the end. Possible selection strategies to achieve this goal include direct selection on regression coefficients or selection on feed intake capacity at predefined points during the growing period. Both strategies could be based on results of a polynomial random regression model applied to daily feed intake. Heritabilities of regression coefficients for daily feed intake and genetic eigenfunctions presented in chapter 4 indicate that it will be very difficult to change the shape of average feed intake curves by selection. This impression is supported by correlations between daily feed intake and single measured performance traits presented in chapter 5, which are almost constant throughout the entire testing period. The best way to improve daily feed intake in the beginning of the testing period might be to select for a higher intercept parameter of feed intake curves. But advantages of such a selection scheme compared to selection for average daily feed intake are limited due to the unfavourable genetic correlation of the intercept parameter with carcass lean content. Since parameter estimates were very unfavourable for changing the feed intake curve of growing pigs, we decided not to compare the above mentioned selection strategies in more detail.

The models used for analysis of daily feed intake in this thesis could certainly still be refined. Instead of the fixed regressions on days on test for each batch, a scalar environmental effect could be fitted for each test date within station or batch. The importance of such an effect would mainly depend on the range of start dates within each batch, since environmental conditions are less variable for station tested pigs than e.g. for dairy cows on different farms. In this thesis, the same order of polynomial fit was used for fixed and all random effects to guarantee a proper definition of heritability for regression coefficients. This

corresponds to a hierarchical model, where the same fixed and random effects are fitted for all parameters of the quadratic polynomial production function for daily feed intake. Given the relatively low heritabilities for linear and especially quadratic regression coefficients, a reduction of polynomial fit for random effects might be appropriate. This has the disadvantage, that some information about the variability of feed intake curves is lost with a linear compared to a quadratic polynomial. Taking this reduction of the order of polynomial fit one step further would mean to go back to a fixed regression model or even a simple repeatability model. Retrieving information on the variability of feed intake curves was the reason for applying a random regression model in the first place, without this possibility the advantage over the conventional method of averaging daily feed intake over the whole testing period is small. To decide whether a random regression model should be used for routine evaluations of daily feed intake records in a specific population, (co)variance components specific to this population are necessary. If the potential for changing the shape of feed intake curves of growing pigs is not more encouraging than found in this study, the additional effort for using a random regression model for routine evaluations of daily feed intake is not justified.

6. Final conclusions

The following conclusions can be drawn from this thesis:

- Polynomial random regression models in combination with calculation of eigenfunctions and their associated eigenvalues are a useful tool to analyse longitudinal data and to assess the amount of (genetic) variation available in a population along such a trajectory (e.g. feed intake curve).
- Gibbs sampling for random regression models is very computer intensive because a high number of rounds is needed for reliable estimates, due to high autocorrelations between Gibbs samples.
- Changing the shape of feed intake curves by selection to improve the efficiency of lean growth will be very difficult.

References

[1] Andersen S., Pedersen B., Growth and food intake curves for group housed gilts and castrated male pigs, Anim. Sci. 63 (1996) 457-464.

[2] Cameron N.D., Across species comparisons in selection for efficiency, in: Proc. 6th World Congress on Genetics Applied to Livestock Production, Armidale, NSW, Australia, January 11-16, vol. 18, 1994, pp. 443-446.

[3] Eissen J.J., Breeding for feed intake capacity in pigs, doctoral thesis, Animal Breeding and Genetics Group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands, 2000.

[4] Gelfand A.E, Sahu S.K., Carlin B.P., Efficient parametrisations for normal linear mixed models, Biometrika 82 (1995) 479-478.

[5] Jakobsen J.H., Genetic correlations between the shape of the lactation curve and disease resistance in dairy cattle, Ph.D. thesis, Department of Animal Sciences and Animal Health, The Royal Veterinary and Agricultural University, Grønnegårdsvej 3, DK-1870 Frederiksberg C, Denmark, and Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, Research Centre Foulum, P.O. Box 50, DK-8830 Tjele, Denmark, 2000.

[6] Jamrozik J., Schaeffer L.R., Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins, J. Dairy Sci. 80 (1997) 762-770.

[7] Kirkpatrick M., Lofsvold D., Bulmer M., Analysis of the inheritance, selection and evolution of growth trajectories, Genetics124 (1990) 979-993.

[8] Labroue F., Aspects génétiques du comportement alimentaire chez le porc en croissance, Thèse de Doctorat, Ensar, Rennes, 1996.

[9] Rebsamen A., Lorenz D., Schwörer D., Stationsprüfungen 1998, MLP Quartalsbericht I/99, Suisseporcs-Information Nr. 4, 1999, Beilage 1-3.

[10] Rekaya R., Carabaño M.J., Toro M.A., Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle, Livest. Prod. Sci. 57 (1998) 203-217.

Lebenslauf

Personalien

Name:	Schnyder
Vorname:	Urs
Geburtsdatum:	24. September 1971
Heimatort:	Diessbach bei Büren (BE)
Zivilstand:	verheiratet

Ausbildung

1978-1984	Primarschule Biberist (SO)
1984-1986	Bezirksschule Biberist (SO)
1986-1991	Oberrealschule der Kantonsschule Solothurn (SO) (mathematisch-naturwissenschaftliches Gymnasium)
Abschluss:	Matura Typus C
1991-1997	Studium der Agronomie mit Fachrichtung Nutztierwissenschaften an der Abteilung für Agrar- und Lebensmittelwissenschaften der ETH Zürich
Abschlüsse:	Diplom als Ingenieur-Agronom ETH
	Fähigkeitsausweis für das Lehramt an Berufs- und Fachschulen und anderen höheren Lehranstalten im Bereich Agrarwissenschaften

Berufliche Tätigkeit

1997-2001	Assistent am Institut für Nutztierwissenschaften, Gruppe Tierzucht, ETH Zürich
2001	Wissenschaftlicher Mitarbeiter, applied genetics network, Altendorf