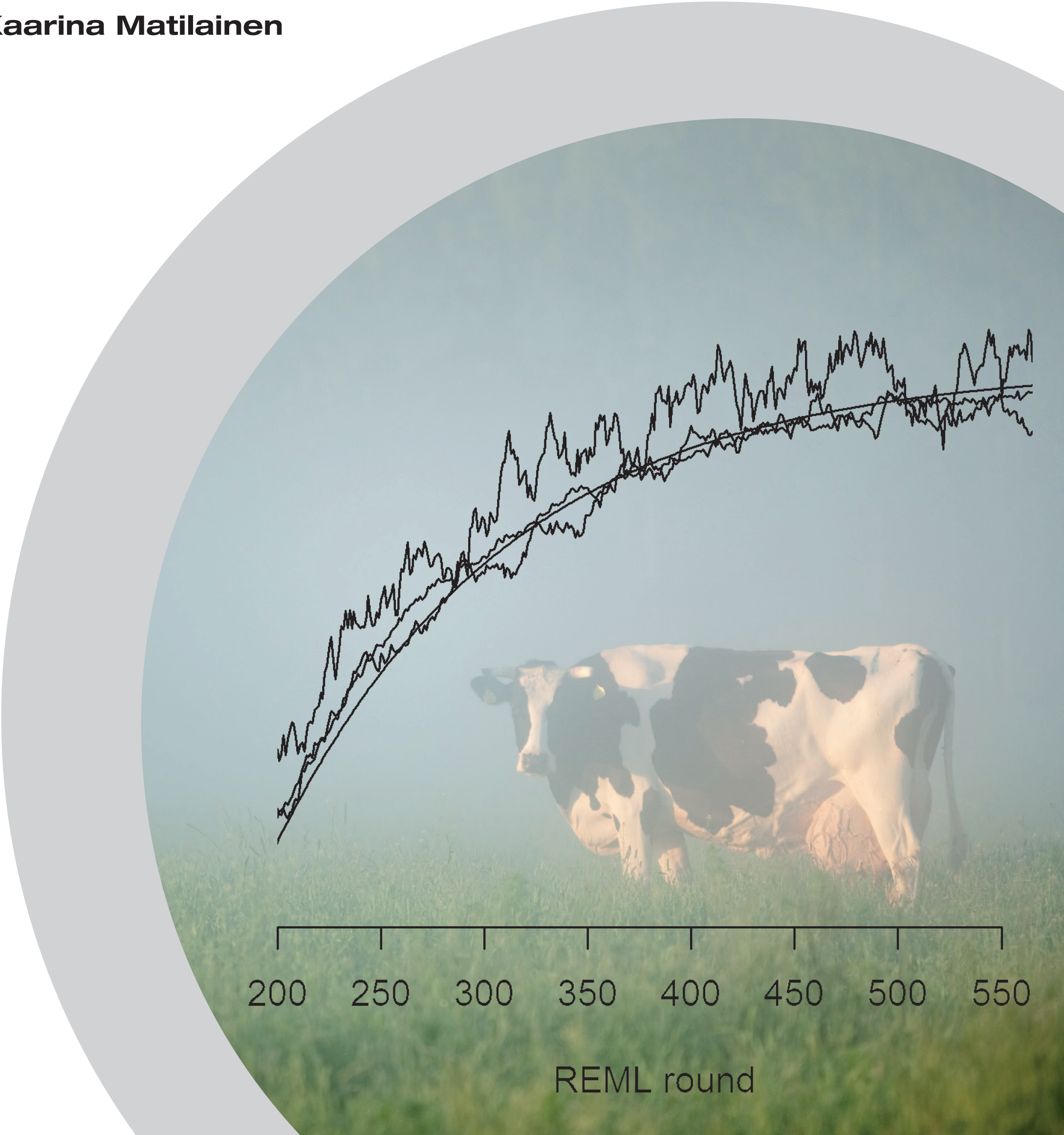
MTT SCIENCE 30

# Variance component estimation exploiting Monte Carlo methods and linearization with complex models and large data in animal breeding

## Doctoral Dissertation

**Kaarina Matilainen**



MTT SCIENCE 30

# Variance component estimation exploiting Monte Carlo methods and linearization with complex models and large data in animal breeding

## Doctoral Dissertation

**Kaarina Matilainen**

Academic dissertation
To be presented, with the permission of the Faculty of Agriculture and Forestry of
the University of Helsinki, for public examination in Auditorium XIV, Main building,
Fabianinkatu 33, on December 5th 2014, at 12 noon.

ISBN 978-952-487-568-4 (Print)
ISBN 978-952-487-569-1 (Electronic)
ISSN 1798-1824 (Printed version)
ISSN 1798-1840 (Electronic version)
URN http://urn.fi/URN:ISBN:978-952-487-569-1
www.mtt.fi/mtttiede/pdf/mtttiede30.pdf

**Distribution and sale**
MTT Agrifood Research Finland,
Media and Information Services,
FI-31600 Jokioinen,
e-mail julkaisut@mtt.fi
**Printing year** 2014
**Cover image** Jaakko Suvala/Rodeo / MTT archive, graph/Kaarina Matilainen
**Printing house** Tampereen Yliopistopaino Juvenes Print Oy

## Custos:

Professor Asko Mäki-Tanila, Department of Agricultural Sciences, P.O. Box 28, University of Helsinki, Finland

## Supervisors:

Docent Ismo Strandén, MTT Agrifood Research Finland, Biotechnology and Food Research, Humppilantie 7, 31600 Jokioinen, Finland

Professor Esa A. Mäntysaari, MTT Agrifood Research Finland, Biotechnology and Food Research, Humppilantie 7, 31600 Jokioinen, Finland

## Reviewers:

Professor Just Jensen, Aarhus University, Department of Molecular Biology and Genetics, Blichers Allé 20, 8830 Tjele, Denmark

Dr Bjørg Heringstad, Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, P.O. Box 5003, 1432 Ås, Norway

## Opponent:

Dr John Hickey, The Roslin Institute, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK

# Variance component estimation exploiting Monte Carlo methods and linearization with complex models and large data in animal breeding

**Kaarina Matilainen**
MTT Agrifood Research Finland, Biotechnology and Food Research, Humppilantie 7,
FI-31600 Jokioinen, Finland, kaarina.matilainen@mtt.fi

## Abstract

Inference for the variance components in linear mixed effects models is theoretically well understood. Many methods have also been presented for nonlinear models. Genetic evaluations in animal breeding are however characterized by the enormous size of the models and data. This means that the methods in estimation have to be computationally efficient. The purpose of this study was to find efficient methods for the estimation of the variance components for large data sets and complex mixed effects models in animal breeding.

The focus of the study was, first, on the restricted maximum likelihood (REML) estimation applied to a linearized model of nonlinear mixed effects model and, second, on the REML estimation of large linear mixed effects models by the Monte Carlo (MC) method. Performance of the methods were mostly studied using simulated data sets, but the feasibility of the MC based expectation maximization (EM) REML was also studied using dairy cattle field data.

The analyses of a data set mimicking pig live weights showed that linearization works moderately well when the data is good, but estimation of parameters related to adult weight becomes unstable when weight observations from the right tail of the animals' growth curve were missing. However, the simulation study showed that having even a small proportion of animals with adult weights improved the results when compared to the estimates based on observations from prematurely slaughtered animals only.

The MC based EM REML method converged to the same solutions as the analytical EM REML, and a small number of MC samples did not introduce systematic bias to the estimates of genetic parameters in the analysis of simulated dairy cattle data set. Furthermore, analyses of field data proved the MC EM REML to be superior to the analytical EM REML both in computing time and in the memory needed. Compared to MC EM REML, the MC Newton-type methods converged faster, but sampling variation of the estimates increased. Sampling variation differed somewhat also between the Newton-type methods.

Developing a fast algorithm for MC based REML estimation requires a convergence criterion that is robust for sampling variation. A stopping rule that can be calculated during the analysis was introduced. The applied convergence criterion monitored the progress of convergence and was only a little influenced by MC noise. It also worked reasonably well with small number of MC samples, which is a property that may be useful for analyzing large scale and complex models.

**Keywords:** *mixed effects model, variance components, linearization, Monte Carlo, REML, EM-algorithm, Newton's method*

# Monte Carlo -menetelmät ja linearisointi varianssikomponenttien arvioinnissa analysoitaessa suuria aineistoja vaativilla jalostusarvostelumalleilla

**Kaarina Matilainen**
MTT, Biotekniikka- ja elintarviketutkimus, Humppilantie 7, 31600 Jokioinen,
kaarina.matilainen@mtt.fi

## Tiivistelmä

Keskeinen osa eläinjalostusta on geneettisistä tekijöistä johtuvan vaihtelun erottaminen havaintojen kokonaisvaihtelusta. Geneettistä vaihtelua arvioidaan tilastotieteestä tutuilla sekamalleilla. Sekamallien sisältämien varianssikomponenttien arviointiin liittyvää teoriaa on tutkittu paljon lineaarisilla sekamalleilla. Myös epälineaarisille malleille on esitetty monia arviointimenetelmiä. Koska jalostusarvosteluissa käytettävät aineistot ja mallit ovat usein suuria, arviointimenetelmien yksi tärkeä ominaisuus on laskennallinen tehokkuus.

Tämän tutkimuksen tarkoituksena oli löytää tehokas menetelmä varianssikomponenttien arviointiin analysoitaessa eläinjalostuksessa käytettäviä suuria aineistoja monimutkaisilla sekamalleilla. Erityisesti tutkimus keskittyi 1) epälineaaristen sekamallien varianssikomponenttien arviointiin mallin linearisoinnilla ja REML-tyyppisellä menetelmällä ja 2) Monte Carlo (MC) –menetelmän hyödyntämiseen sekä EM (expectation maximization) että Newtonin tyyppisissä REML-analyyseissä lineaarisille sekamalleille. Menetelmiä tutkittiin simuloiduilla aineistoilla. MC-menetelmää hyödyntävän EM REML:n soveltuvuutta testattiin myös todellisella lypsykarja-aineistolla.

Linearisointi toimi kohtuullisesti simuloidulla aineistolla, joka kuvasi eläinten painon kehitystä syntymästä aikuispainoon asti. Aikuispainoon liittyvien varianssikomponenttien arvioiden luotettavuus heikkeni, kun aineisto sisälsi havaintoja ainoastaan aikuispainoa edeltävältä ajalta. Pienikin aikuispainohavaintojen lisäys kuitenkin paransi luotettavuutta.

MC-menetelmää käyttävä EM REML konvergoi samoihin varianssikomponenttien arvioihin kuin analyyttinen menetelmä, eikä pieni MC-otosten määrä näkynyt systemaattisena harhana arvioissa. Todellisen aineiston analyysit osoittivat MC EM REML -menetelmän olevan parempi sekä laskenta-ajaltaan että tietokoneen muistitarpeeltaan kuin analyyttinen EM REML -menetelmä. MC-menetelmän soveltaminen Newtonin tyyppisiin menetelmiin sai REML-ratkaisut konvergoimaan nopeammin kuin MC EM REML:llä, mutta arvioiden otosvaihtelu oli suurempaa. Otosvaihtelun suuruus vaihteli käytetyn Newtonin menetelmän mukaan.

Jokaisessa MC REML -analyysissä tarkasteltiin myös uutta konvergenssikriteeriä. Uudella kriteerillä pystyttiin vähentämään MC-menetelmästä johtuvaa vaihtelua konvergenssin seurannassa. Kriteeri toimi melko hyvin myös pienillä MC-otosmäärillä, mikä on toivottava ominaisuus analysoitaessa suuria aineistoja ja monimutkaisia malleja.

---

**Avainsanat:** *sekamalli, varianssikomponentti, linearisointi, Monte Carlo, REML, EM-algoritmi, Newtonin menetelmä*

# Acknowledgements

I wish to express my sincere gratitude to my supervisors Ismo Strandén and Esa Mäntysaari for introducing me into the research field of animal breeding and their expert guidance in this work. Also I wish to express gratitude to co-authors Martin Lidauer, Robin Thompson and Marja-Liisa Sevón-Aimonen for their fruitful collaboration. You contributed to a pleasant working team and I am most deeply grateful for your support and encouragement during my work.

The research work for this thesis has been done at MTT Agrifood Research Finland and financial support was provided by Rothamsted Research. Data was provided by Faba, Hollola, Finland. These organizations are greatly acknowledged.

Thanks to Pekka Uimari for always helping with questions related to my studies. You kept the wheels in motion at the final crunch. Further, my thanks go to thesis reviewers, Just Jensen and Bjørg Heringstad, for valuable suggestions. Thanks also go to my custos Asko Mäki-Tanila.

I warmly thank Kirsi Muuttoranta for advice in so many official and unofficial matters that came up against. My gratitude goes also to current and former colleagues at the biometrical genetics team. In addition to all the advice and assistant I have got, I wish to thank you for a fun-filled moments and shared sport hours. Thanks to all of you. Keep going!

I would like to express my deepest appreciation to my family for providing the requirements for everyday life. To my parents, my sister and brother and their families, for listening if needed but not asking too much about my work; to my mother- , father- and sister-in-law not least for being babysitters on so many nights during the last months; But foremost to my sons Esko and Niilo for their patience when I have been an absent and absent-minded mother and my husband Jukka for his patience, love and support.

Jokioinen, October 2014

Kaarina Matilainen

# List of original publications

This thesis is based on the following publications:

I Vuori K., Strandén I., Sevón-Aimonen M.-L., Mäntysaari E.A. (2006). Estimation of non-linear growth models by linearization: a simulation study using a Gompertz function. Genetics, Selection, Evolution, 38, 343-358. doi:10.1051/gse:2006008

II Matilainen K., Mäntysaari E.A., Lidauer M.H., Strandén I., Thompson R. (2012). Employing a Monte Carlo algorithm in expectation maximization restricted maximum likelihood estimation of the linear mixed model. Journal of Animal Breeding and Genetics, 129, 457-468. doi:10.1111/j.1439-0388.2012.01000.x

III Matilainen K., Mäntysaari E.A., Lidauer M.H., Strandén I., Thompson R. (2013). Employing a Monte Carlo algorithm in Newton-type methods for restricted maximum likelihood estimation of genetic parameters. PLoS ONE, 8, e80821. doi:10.1371/journal.pone.0080821

The publications are referred to in the text by their Roman numerals. Reprints of the original articles are published with the kind permission of BioMed Central Ltd., John Wiley & Sons Ltd. and PLOS.

The author (née Vuori) participated in planning the studies, performed the data simulations and all analyses, interpreted the results with the co-authors and was the main writer of all the articles.

# Abbreviations

| | |
|---|---|
| AI | average information |
| BM | Broyden's method |
| EBLUP | empirical best linear unbiased prediction |
| EM | expectation maximization |
| MC | Monte Carlo |
| MCMC | Markov chain Monte Carlo |
| ML | maximum likelihood |
| MME | mixed model equations |
| NR | Newton-Raphson |
| PEV | prediction error variance |
| PX-EM | parameter expansion version of expectation maximization |
| REML | restricted maximum likelihood |
| SAEM | stochastic approximation expectation maximization |
| TD | test day |
| VC | variance component |

# Contents

# 1 Introduction

## 1.1 Genetic parameters

Estimation of genetic parameters is one of the essential steps necessary for reliable animal evaluations. Genetic parameters reveal the amount of genetic variation in traits, the heritability of traits and possible repeatability of traits as well as correlations between traits. The heritability of a trait indicates the ability to change the appearance of the trait in a population when selective breeding is applied to the trait. The higher the heritability the more accurately breeding values can be estimated with the same number of observations. Because genetic gain is affected by accuracy together with genetic variance, selection intensity and generation interval, more genetic change can be achieved within the same time period using more accurate breeding values given the other parameters are kept unchanged. In addition to heritability, breeding programs pay attention to genetic and phenotypic correlations between traits so that the positive correlation between two traits can be exploited or undesired co-response due to selection can be decreased.

Estimation of genetic parameters is based on correlated phenotypes between relatives. Therefore, analyses require statistical models which describe environmental components and the sources of variation between observations. Estimation of variance components (VC) tries to quantify the extent to which variation between the observations is due to genetic and non-genetic environmental differences. Hereby the values of genetic parameters are population and data dependent. Also the different models used may lead to different estimates about the true state of the population investigated. Therefore the VC need to be re-estimated when some of these components change.

## 1.2 Estimation methods of genetic parameters

One of the most traditional methods to estimate VC is analysis of variance (ANOVA). Because it is applicable for a very limited data structure, Henderson (1953) presented three different methods. The most versatile method is known as Henderson's method III. The disadvantage of the method was its inability to take into account relationships between animals, i.e., it could not account for the effect of selection. Interest in maximum likelihood (ML) approaches increased in 1970's (Harville, 1977). The ML approach uses probability distributions as likelihood functions: While data have probability distribution given the parameters, ML maximizes the likelihood function of parameters given the data. ML is a flexible modeling technique, which allows users to account for more assumptions in the model and patterns of missing traits in multivariate analysis. It is also able to take into account relationships between animals through the numerator relationship matrix. VC estimates by the ML method are, however, dependent on the number of levels within the fixed effects. In theory, when simple models allowed by ANOVA are used, ML estimation leads to biased variance estimates whereas ANOVA estimates are unbiased. This deficiency is corrected in the restricted maximum likelihood (REML) method, which takes into account the loss of information due to simultaneous estimation of fixed effects (Patterson and Thompson, 1971). A competing estimation approach is based on Bayesian inference which accounts for the loss of information due to estimation of all parameters other than the one of interest. In practice, REML and Bayesian approaches often lead to similar inferences.

The REML approach may lead to massive computations and many methods have been presented over the years to calculate REML estimates. The simplest methods use values of the REML likelihood only. More efficient methods use first derivatives of the REML likelihood. These are the so-called first-order methods. Even better is to use both the first and the second derivatives of the REML likelihood. These are called second-order methods. Because the matrix of second derivatives can be difficult to calculate, interesting alternatives are the quasi-Newton methods. They rely on approximations of the second derivatives based on the direction of the most recent step. Quasi-Newton methods usually result in faster convergence compared to the first-order methods but slower convergence compared to the true second-order methods because the information matrix is replaced by an approximation. A good summary of the history of VC estimation is given e.g. by Hofer (1998) and Thompson et al. (2005).

After decades of development work, inference for VC in linear mixed effects models is theoretically well understood. However, animal breeding evaluations are often characterized by a large number of observations and complex models which may lead to computational problems. This underlines the fact that the methods in estimation, especially in model development tools, have to be computationally effective. For VC estimation, existing computer packages are often prohibitively slow (Bayesian approaches) or require huge computer memory capacity (REML approaches). Notable challenges are random regression test day (TD) models, international multivariate comparisons and individual gene models which all require up-to-date statistical and numerical techniques.

Linear models are flexible. They can be used, for example, to model the lactation curves of cows by random regression models. Sometimes animal growth curves are modeled by a sigmoid or S-shaped function. As an example, Gompertz function has been shown to fit pig growth data, such as live weight and protein retention (Wellock et al., 2004; Whittemore and Green, 2002; Whittemore et al., 1988) well. Models using S-shaped function are called nonlinear because the fixed and random effects do not enter the model linearly. An advantage of the Gompertz function is that predictions beyond the observed data range can be made more realistic. This is a useful property that is needed, for example, to predict the adult weight of prematurely slaughtered pigs. Nonlinear models are, however, more complicated to solve than linear models (Davidian and Giltinan, 1995). In animal evaluations, the Bayesian framework has been popular due to the use of Markov chain Monte Carlo (MCMC) methods that allow the solution for the required numerically complicated integration (e.g. Blasco et al., 2003; Rekaya et al., 2001). Another possibility is to approximate the likelihood function using linearization (Davidian and Giltinan, 1995; Pinheiro and Bates, 1995; Wolfinger, 1993; Wolfinger and Lin, 1997) or numerical integration (Pinheiro and

Bates, 1995). However, all these methods are computationally intensive.

### 1.2.1 Exploiting Monte Carlo methods

Monte Carlo (MC) simulation is used to estimate the integral needed for calculation of expectations of a random variable. MC methods are often used when closed-form solutions are difficult or impossible to obtain. Wei and Tanner (1990) introduced the MC expectation maximization (EM) algorithm for cases where maximization for complete data is simple and the expectation can be approximated by MC simulations. This has mainly been used within the classical likelihood framework in the analysis of complex models, like generalized linear mixed models, for which expectations cannot be calculated analytically (e.g. Walker, 1996; McCulloch, 1997; Booth and Hobert, 1999). It can also be utilized in VC estimation of linear mixed effects models via Gibbs sampling to avoid insurmountable computations in the analysis of large data sets and models by the analytical EM REML method (e.g. Guo and Thompson, 1991; Thompson, 1994; García-Cortés and Sorensen, 2001; Harville, 2004). García-Cortés et al. (1992) applied MC EM to VC estimation in a different way. To estimate the prediction error variances (PEV) within each REML round, independent data sets, corresponding to the original data, are generated using the assumed linear model. Fixed and random effects are then estimated from the simulated data sets. This enables the calculation of PEV without inversion or decomposition of the coefficient matrix and leads to memory requirements equal to those used in solving the mixed model equations (MME).

Because EM REML may converge slowly, improvements have been offered by second-order methods. Such methods include for example the Newton-Raphson (NR) method using an observed information matrix and Fisher's method of scoring with observed information matrix replaced by an expected information matrix. Klassen and Smith (1990) presented a sampling scheme for Fisher's method of scoring. Interestingly it does not need any derivatives, but may be inapplicable for large and complex models as is often the case with derivative free methods in general. At the same time as Wei and Tanner (1990) introduced MC EM, they proposed an improvement for convergence via MC approximation of Louis's method (Louis, 1982). Again, the use has mainly been in analyses involving complex likelihoods as is the case e.g. in utilization of MC NR method for generalized linear mixed models in Kuk and Cheng (1997). However, use of the simple sampling method presented in García-Cortés et al. (1992) has not been applied with second-order methods.

### 1.2.2 Exploiting linearization

Analytical solutions for the integral of the nonlinear likelihood function can only rarely be found. Therefore approximation methods based on Taylor-series expansion or Laplacian approximation have been discussed extensively. Early methods used first-order Taylor series expansion of nonlinear functions around expectation of the random effects (Davidian and Giltinan, 1995). Lindstrom and Bates (1990) suggested a more accurate method where the expansion is made around current estimates of the random effects. This is recommended especially for cases where the VCs are large and substantial inter-individual variation exists (Lindstrom and Bates, 1990). Subsequent research has focused more on the second-order Taylor series expansion of integrals invoked by the Laplacian approximation (Pinheiro and Bates, 1995; Wolfinger, 1993; Wolfinger and Lin, 1997).

Wolfinger and Lin (1997) presented an interesting choice of approximation for the integrant because of general and familiar formulation. They gave two alternative approaches to select points of expansion: a zero expansion method using expected values, and an EBLUP-expansion method using the empirical best linear unbiased predictors (EBLUP) of the random effects. Both approximations lead to algorithms that iteratively fit linear mixed effects models to the suitably transformed data using either ML or REML. Therefore, they allow the use of commonly applied methods for linear mixed effects models, and they are called linearization methods here. Because conditions for the functionality of the approximation methods are difficult to identify, Wolfinger and Lin (1997) recommended simulation studies for assessing the performance of the methods in diverse kinds of nonlinear models and data sets. To our knowledge, the appropriateness of the EBLUP REML method has prior to this study not been verified for large animal breeding data sets.

## 1.3 Objectives of the study

The purpose of this study was to find an efficient method for estimation of the VC of large data and complex mixed effects models in animal breeding. The focus was on the REML estimation of large nonlinear mixed effects models utilizing linearization and the REML estimation of large linear mixed effects models by the MC method. More specifically aims were:

1. To study the applicability of a method based on Taylor series expansion (Wolfinger and Lin, 1997) in animal breeding with data simulated and analyzed using a Gompertz function growth model.
2. To show the feasibility of the MC EM REML method proposed in García-Cortés et al. (1992) for large data sets and complex linear models by comparing it with analytical EM REML using simulated and field data.
3. To study the performance of the MC method in different Newton-type methods for VC estimation of linear mixed effects models by analyzing small simulated data.
4. To study a new convergence criterion applicable for the MC methods.

# 2 Materials and methods

In the following, the linear and nonlinear mixed effects models are introduced and the likelihood based estimation is considered. The ML and REML methods are explained, and some methods to maximize the REML likelihood function by analytical and MC methods are described. Also assessment of the convergence in case of the MC based estimation methods is discussed. Finally, data and programs used in the analyses are given.

## 2.1 Likelihood of linear mixed effects model

Consider a simple single trait linear mixed effects model

(1) $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$,

where $\mathbf{y}$ is the vector of $n$ observations, $\mathbf{b}$ is the vector of $p$ fixed effects, $\mathbf{u}$ is the vector of $q$ random breeding values, and $\mathbf{e}$ is the vector of $n$ random residuals. Furthermore, $\mathbf{X}$ and $\mathbf{Z}$ are known design matrices, which relate records to fixed effects and random genetic effects, respectively. Usually, the random effects $\mathbf{u}$ and $\mathbf{e}$ are assumed to be independent of each other and to follow Gaussian distribution: $\mathbf{u} \sim N(\mathbf{0},\mathbf{G})$ and $\mathbf{e} \sim N(\mathbf{0},\mathbf{R})$, where $\mathbf{G} = \mathbf{A}\sigma_u^2$, $\mathbf{A}$ is $q \times q$ numerator relationship matrix, $\mathbf{R} = \mathbf{I}\sigma_e^2$ and $\mathbf{I}$ is $n \times n$ identity matrix. The unknown VC are genetic variance $\sigma_u^2$ in $\mathbf{G}$ and residual variance $\sigma_e^2$ in $\mathbf{R}$. Denote the vector of unknowns by the parameter vector $\boldsymbol{\theta}$.

The assumptions of the model result in Gaussian distribution for observations: $\mathbf{y} \sim N(\mathbf{Xb},\mathbf{V})$, where $\mathbf{V} = \mathbf{ZGZ}^T + \mathbf{R}$. The likelihood function for the Gaussian distributed data in $\mathbf{y}$ is

$$
\begin{aligned}
L(\mathbf{b},\boldsymbol{\theta} \mid \mathbf{y}) = {} & (2\pi)^{-\frac{n}{2}} |\mathbf{R}|^{-\frac{1}{2}} (2\pi)^{-\frac{q}{2}} |\mathbf{G}|^{-\frac{1}{2}} \\
& \int \exp\Big(-\tfrac{1}{2}(\mathbf{y}-\mathbf{Xb})^T \mathbf{R}^{-1}(\mathbf{y}-\mathbf{Xb}) \\
& -\tfrac{1}{2}\mathbf{u}^T\mathbf{G}^{-1}\mathbf{u}\Big)d\mathbf{u},
\end{aligned}
\tag{2}
$$

which leads to the logarithmic likelihood function

$$
\begin{aligned}
l(\mathbf{b},\boldsymbol{\theta} \mid \mathbf{y}) &= \log L(\mathbf{b},\boldsymbol{\theta} \mid \mathbf{y}) \\
&= -\tfrac{n}{2}\ln(2\pi) - \tfrac{1}{2}\ln|\mathbf{V}| \\
&\quad -\tfrac{1}{2}(\mathbf{y}-\mathbf{Xb})^T \mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb}).
\end{aligned}
\tag{3}
$$

This function is used in the ML framework (Searle et al, 1992). The aim is to find parameter values $\boldsymbol{\theta}$ that maximizes the likelihood function given the observed data $\mathbf{y}$. This can be obtained by taking derivatives of the log likelihood function with respect to $\boldsymbol{\theta}$, equating to zero and solving for $\boldsymbol{\theta}$. Often this approach requires the use of so-called profile likelihood of $\boldsymbol{\theta}$ constructed by substituting the current estimates of the fixed effects $\hat{\mathbf{b}}$,

(4) $\hat{\mathbf{b}} = \left(\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$,

for $\mathbf{b}$.

## 2.2 Likelihood of nonlinear mixed effects model by linearization

Consider a single trait nonlinear model. Let the model be $\mathbf{y} = f(\mathbf{X},\mathbf{b},\mathbf{Z},\mathbf{u}) + \mathbf{e}$, where $\mathbf{y}$ is the vector of $n$ observations, $f$ is the nonlinear function and $\mathbf{e}$ is the vector of $n$ random residuals. Assume further that each of the nonlinear model parameters can be modeled by a linear mixed effects model using terms $\mathbf{X}$, $\mathbf{b}$, $\mathbf{Z}$ and $\mathbf{u}$ as defined earlier. For example, the nonlinear function could be a three parametric Gompertz function $f = \alpha \exp(-\beta \exp(-\kappa t))$ where the parameters $\alpha$, $\beta$ and $\kappa$ can be modeled by linear mixed effects model, i.e.,

$$\begin{aligned} \mathbf{y} = (\mathbf{Xb}_\alpha + \mathbf{Zu}_\alpha)\exp(-(\mathbf{Xb}_\beta + \mathbf{Zu}_\beta) \\ \exp(-(\mathbf{Xb}_\kappa + \mathbf{Zu}_\kappa)t)) + \mathbf{e}. \end{aligned} \tag{5}$$

Now the fixed effects are $\mathbf{b}^T = \lfloor \mathbf{b}_\alpha^T \ \ \mathbf{b}_\beta^T \ \ \mathbf{b}_\kappa^T \rfloor$, and the random effects are $\mathbf{u}^T = \lfloor \mathbf{u}_\alpha^T \ \ \mathbf{u}_\beta^T \ \ \mathbf{u}_\kappa^T \rfloor$ and $\mathbf{e}$. Again the random effects $\mathbf{u}$ and $\mathbf{e}$ are assumed to be independent from each other and to follow Gaussian distribution: $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, where $\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$. Because the Gompertz function results in three levels for each fixed and random effect in the model, covariance matrix $\mathbf{G}_0$ is $3\times3$ matrix with unique elements of $\sigma_{u_\alpha}^2$, $\sigma_{u_\beta}^2$, $\sigma_{u_\kappa}^2$, $\sigma_{u_{\alpha\beta}}$, $\sigma_{u_{\alpha\kappa}}$ and $\sigma_{u_{\beta\kappa}}$.

For the nonlinear model considered here, the profile likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta} \mid \mathbf{y}) = (2\pi)^{-\frac{n}{2}} |\mathbf{R}|^{-\frac{1}{2}} (2\pi)^{-\frac{q}{2}} |\mathbf{G}|^{-\frac{1}{2}} \\ \int \exp\Big( -\tfrac{1}{2}\big(\mathbf{y} - f(\mathbf{X}\hat{\mathbf{b}} + \mathbf{Zu})\big)^T \mathbf{R}^{-1} \\ \big(\mathbf{y} - f(\mathbf{X}\hat{\mathbf{b}} + \mathbf{Zu})\big) - \tfrac{1}{2}\mathbf{u}^T \mathbf{G}^{-1}\mathbf{u} \Big) d\mathbf{u}. \end{aligned} \tag{6}$$

Calculation of this likelihood requires calculation of the integral. When the function $f$ is nonlinear, a closed form can only rarely be found and the integral must be solved numerically. One possibility is to approximate the integral by quadratic Taylor-series expansion of the exponent (Wolfinger and Lin, 1997). The second-order expansion with respect to the random effects gives

$$\begin{aligned} l^*(\boldsymbol{\theta} \mid \mathbf{y}) &= \log L(\boldsymbol{\theta} \mid \mathbf{y}) \\ &= -\tfrac{n}{2}\ln(2\pi) - \tfrac{1}{2}\ln\left|\mathbf{V}^*\right| \\ &\quad -\tfrac{1}{2}\big(\mathbf{Y} - \mathbf{X}^*\hat{\mathbf{b}}\big)^T \mathbf{V}^{*-1}\big(\mathbf{Y} - \mathbf{X}^*\hat{\mathbf{b}}\big), \end{aligned} \tag{7}$$

where

$$\begin{aligned} \mathbf{V}^* &= \mathbf{Z}^*\mathbf{G}\mathbf{Z}^{*T} + \mathbf{R} \\ \mathbf{Y} &= \mathbf{y} - f\big(\mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}}\big) + \mathbf{X}^*\hat{\mathbf{b}} + \mathbf{Z}^*\hat{\mathbf{u}} \\ \mathbf{X}^* &= \left.\frac{\partial f}{\partial \mathbf{b}^T}\right|_{\mathbf{b}=\hat{\mathbf{b}}} \\ \mathbf{Z}^* &= \left.\frac{\partial f}{\partial \mathbf{u}^T}\right|_{\mathbf{u}=\hat{\mathbf{u}}} \end{aligned} \tag{8}$$

are calculated at $\hat{\mathbf{b}}$ as the current estimate of the fixed effects and $\hat{\mathbf{u}}$ as the current EBLUP of the random effects. Consequently, linearization leads to a log likelihood function that iteratively fits linear mixed effects models to the suitably transformed data $\mathbf{Y}$. Therefore, it allows the use of commonly applied methods for linear mixed effects models.

## 2.3 REML

The ML estimates tend to be biased downwards for VC (Patterson and Thompson, 1974). Instead of the ML estimates, REML estimates proposed by Patterson and Thompson (1971) are widely preferred in practice. These estimates account for loss in degrees of freedom caused by the estimation of fixed effects $\mathbf{b}$. In principle the REML method is based on linear transformation of the data $\mathbf{y}$: $\mathbf{z} = \mathbf{K}^T\mathbf{y}$, where $\mathbf{K}$ is $n\times(n-p)$ matrix of full rank for which $\mathbf{K}^T\mathbf{X} = \mathbf{0}$, and $n$ and $p$ are numbers of observations and fixed effects, respectively. Elements in $\mathbf{z}$ may be referred to as error contrasts and they follow the multivariate normal distribution $\mathbf{z} \sim N(\mathbf{0}, \mathbf{K}^T\mathbf{VK})$. Log likelihood function for error contrasts will be

$$
\begin{aligned}
l_{REML}(\boldsymbol{\theta} \mid \mathbf{y}) &= -\tfrac{n}{2}\ln(2\pi) - \tfrac{1}{2}\ln\left|\mathbf{K}^T\mathbf{V}\mathbf{K}\right| \\
&\quad - \tfrac{1}{2}\left(\mathbf{K}^T\mathbf{y}\right)^T\left(\mathbf{K}^T\mathbf{V}\mathbf{K}\right)^{-1}\left(\mathbf{K}^T\mathbf{y}\right) \\
&= -\tfrac{n}{2}\ln(2\pi) - \tfrac{1}{2}\ln\left|\mathbf{V}\right| \\
&\quad - \tfrac{1}{2}\ln\left|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right| \\
&\quad - \tfrac{1}{2}\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\right)^T\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\right),
\end{aligned} \tag{9}
$$

and the REML estimates are parameter values $\boldsymbol{\theta}$ that maximize the REML likelihood function given the observed data $\mathbf{y}$.

Similarly, REML for the approximated log likelihood function of the nonlinear model is possible. The error contrasts $\mathbf{z} = \mathbf{K}^T\mathbf{Y}$, where $\mathbf{Y}$ is transformed data and $\mathbf{K}$ is a full rank matrix such that $\mathbf{K}^T\mathbf{X}^* = \mathbf{0}$, follow the multivariate normal distribution $\mathbf{z} \sim N\left(\mathbf{0}, \mathbf{K}^T\mathbf{V}^*\mathbf{K}\right)$. The REML solutions for the approximated log likelihood function are obtained by maximizing the equation

$$
\begin{aligned}
l^*_{REML}(\boldsymbol{\theta} \mid \mathbf{y}) &= -\tfrac{n}{2}\ln(2\pi) - \tfrac{1}{2}\ln\left|\mathbf{V}^*\right| \\
&\quad - \tfrac{1}{2}\ln\left|\mathbf{X}^{*T}\mathbf{V}^{*-1}\mathbf{X}^*\right| \\
&\quad - \tfrac{1}{2}\left(\mathbf{Y} - \mathbf{X}^*\hat{\mathbf{b}}\right)^T\mathbf{V}^{*-1}\left(\mathbf{Y} - \mathbf{X}^*\hat{\mathbf{b}}\right),
\end{aligned} \tag{10}
$$

where transformed data $\mathbf{Y}$, covariance matrix $\mathbf{V}^*$ and the working incidence matrices $\mathbf{X}^*$ and $\mathbf{Z}^*$ are obtained by linearization (Wolfinger and Lin, 1997).

In general, maximization of the REML likelihood requires use of numerical methods. The EM algorithm and three Newton-type methods are considered in the following sections. The approaches are shown for a simple linear mixed effects model. Maximization of the approximated REML likelihood of the nonlinear model corresponds to the maximization of the REML likelihood of a random regression model as presented for the EM algorithm in II. For multivariate models maximization by EM REML is considered e.g. in Henderson (1984) and in Mäntysaari and Van Vleck (1989). There is no single method that will be best for every application and some guidelines were given by Misztal (2008).

### 2.3.1 REML by EM algorithm

The EM algorithm was presented by Dempster et al. (1977) for a variety of examples including the VC estimation. Central to the EM algorithm is that by completing observations with unobserved data the maximization becomes easy. This can be achieved by completing the observations with values of random effects $\mathbf{u}$ and residuals $\mathbf{e}$. The EM algorithm iterates between two steps called the E-step and the M-step. The E-step computes a Q-function, defined as an expectation of the logarithmic likelihood function for the complete data given the current parameters:

$$
\begin{aligned}
&Q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^k\right) = const \\
&\quad - \tfrac{1}{2}\ln\left|\mathbf{R}\right| - \tfrac{1}{2}tr\left(\mathbf{R}^{-1}\left(\hat{\mathbf{e}}\hat{\mathbf{e}}^T + \mathbf{W}\mathbf{C}\mathbf{W}^T\right)\right) \\
&\quad - \tfrac{1}{2}\ln\left|\mathbf{G}\right| - \tfrac{1}{2}tr\left(\mathbf{G}^{-1}\left(\hat{\mathbf{u}}\hat{\mathbf{u}}^T + \mathbf{C}^{uu}\right)\right),
\end{aligned} \tag{11}
$$

where $\mathbf{W} = [\mathbf{X} \;\; \mathbf{Z}]$, $\mathbf{C}$ is the inverse of the full MME coefficient matrix, $\mathbf{C}^{uu}$ is a submatrix of $\mathbf{C}$ corresponding to random effects, and $\hat{\mathbf{u}}$ and $\hat{\mathbf{e}}$ are solutions of the MME using current values of variance components. The M-step maximizes the Q-function. Taking derivatives with respect to $\sigma_u^2$ in $\mathbf{G}_0$ and $\sigma_e^2$ in $\mathbf{R}_0$ and equating them to zero gives the following estimates:

$$
\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}^T\mathbf{A}^{-1}\hat{\mathbf{u}} + tr\left(\mathbf{A}^{-1}\mathbf{C}^{uu}\right)}{q} \tag{12}
$$

and

$$
\hat{\sigma}_e^2 = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}} + tr\left(\mathbf{W}\mathbf{C}\mathbf{W}^T\right)}{n}. \tag{13}
$$

The EM algorithm is based on first derivatives only and it does not give standard errors for parameters as a by-product. The EM algorithm is known to be relatively stable although slow to converge.

### 2.3.2 REML by Newton-type methods

Methods relying on the first and second derivatives of the REML log likelihood function $l_{REML}(\boldsymbol{\theta} \mid \mathbf{y})$ with respect to $\boldsymbol{\theta}$ are called Newton-type methods. The principle in all Newton-type methods is to find a solution vector where the first derivative of the function to be maximized is zero. The first derivatives of the log-likelihood form the gradient vector

$$
\begin{aligned}
\mathbf{J}(\boldsymbol{\theta}) &= \frac{\partial l(\boldsymbol{\theta} \mid \mathbf{y})}{\partial \boldsymbol{\theta}} \\
&= \tfrac{1}{2}\mathbf{y}^T\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\mathbf{P}\mathbf{y} - \tfrac{1}{2}tr\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\right) \\
&= \begin{bmatrix} -\dfrac{1}{2}\left(\dfrac{q}{\sigma_u^2} - \dfrac{\hat{\mathbf{u}}^T\mathbf{A}^{-1}\hat{\mathbf{u}} + tr\left(\mathbf{A}^{-1}\mathbf{C}^{uu}\right)}{\sigma_u^4}\right) \\ -\dfrac{1}{2}\left(\dfrac{n}{\sigma_e^2} - \dfrac{\hat{\mathbf{e}}^T\hat{\mathbf{e}} + tr\left(\mathbf{WCW}^T\right)}{\sigma_e^4}\right) \end{bmatrix}
\end{aligned} \tag{14}
$$

and the second derivatives of the log-likelihood with respect to all parameters yield the observed information matrix

$$
\begin{aligned}
\mathbf{H}(\boldsymbol{\theta}) &= -\frac{\partial^2 l(\boldsymbol{\theta} \mid \mathbf{y})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}} \\
&= \mathbf{y}^T\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\mathbf{P}\mathbf{y} - \tfrac{1}{2}tr\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\right),
\end{aligned} \tag{15}
$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}$. The NR method uses the observed information matrix and the gradient vector in calculating new estimates of parameters $\hat{\boldsymbol{\theta}}^k$ at iteration round $k$:

$$
\hat{\boldsymbol{\theta}}^k = \hat{\boldsymbol{\theta}}^{k-1} - \mathbf{H}\left(\hat{\boldsymbol{\theta}}^{k-1}\right)^{-1}\mathbf{J}\left(\hat{\boldsymbol{\theta}}^{k-1}\right), \tag{16}
$$

where information matrix and gradient vector are computed with current VC estimates $\hat{\boldsymbol{\theta}}^{k-1}$.

A variant of the NR method, named Fisher scoring, replaces the observed information matrix with the expected information matrix:

$$
E(\mathbf{H}(\boldsymbol{\theta})) = -\tfrac{1}{2}tr\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\right). \tag{17}
$$

Another Newton-type method is the average information (AI) method (Johnson and Thompson (1995) and Gilmour et al. (1995)), which utilizes the average of the observed and expected information matrices:

$$
\begin{aligned}
\mathbf{AI}(\boldsymbol{\theta}) &= \tfrac{1}{2}\left(\mathbf{H}(\boldsymbol{\theta}) + E(\mathbf{H}(\boldsymbol{\theta}))\right) \\
&= \mathbf{y}^T\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}}\mathbf{P}\mathbf{y}.
\end{aligned} \tag{18}
$$

The AI method is attractive because computation of the average of the observed and expected information matrices is easier than either of the components. It removes the complicated trace calculation and the AI matrix can be computed by solving the MME once for each VC parameter with data replaced by a suitable working vector (Jensen et al., 1997). AI REML is currently the most common VC estimation method used in animal breeding. Newton-type methods reach fast convergence in the neighborhood of the maximum. Furthermore, the observed information matrix, as well as its good approximations, the expected information matrix and the AI matrix, can be used to calculate standard errors for the parameters.

Approximation of second derivatives may also be based on the direction of the most recent estimation step (e.g. in Nocedal and

Wright, 2006). These quasi-Newton methods usually result in faster convergence compared to linear methods but slower convergence compared to Newton-type methods because the information matrix is replaced by an approximation. Broyden's method (BM) is a quasi-Newton method for numerical solution of nonlinear equations (Broyden, 1965). BM updates the inverse of the information matrix (instead of the information matrix itself) within each round:

$$\begin{aligned}(19)\qquad \mathbf{H}(\boldsymbol{\theta})^{-1} &\approx \mathbf{B}(\boldsymbol{\theta})^{k} \\ &= \mathbf{B}(\boldsymbol{\theta})^{k-1} \\ &\quad + \frac{\left(\Delta\boldsymbol{\theta} - \mathbf{B}(\boldsymbol{\theta})^{k-1}\Delta\mathbf{J}(\boldsymbol{\theta})\right)\Delta\boldsymbol{\theta}^{T}\mathbf{B}(\boldsymbol{\theta})^{k-1}}{\Delta\boldsymbol{\theta}^{T}\mathbf{B}(\boldsymbol{\theta})^{k-1}\Delta\mathbf{J}(\boldsymbol{\theta})},\end{aligned}$$

where $\Delta\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{k-1} - \hat{\boldsymbol{\theta}}^{k-2}$ and $\Delta\mathbf{J}(\boldsymbol{\theta}) = \mathbf{J}\left(\hat{\boldsymbol{\theta}}^{k-1}\right) - \mathbf{J}\left(\hat{\boldsymbol{\theta}}^{k-2}\right)$. Instead of true gradients, round-to-round changes in the EM estimates can be used for both the update of the inverse information matrix and the update of new estimates: $\Delta\mathbf{J}(\boldsymbol{\theta}) \approx \left(\hat{\boldsymbol{\theta}}^{k-1} - \hat{\boldsymbol{\theta}}^{k-1}_{EM}\right) - \left(\hat{\boldsymbol{\theta}}^{k-2} - \hat{\boldsymbol{\theta}}^{k-2}_{EM}\right)$. Apart from scaling, these changes are relative to the original gradients (Jamshidian and Jennrich, 1993). BM leads to superlinear convergence although sequence $\{\mathbf{B}(\boldsymbol{\theta})^{k}\}$ does not converge to the observed information matrix at the maximum (Dennis and Moré, 1974). A disadvantage is that BM does not give estimates of the standard errors for the parameters.

## 2.4 REML utilizing MC

The computational problem in VC estimation for animal breeding is not necessarily the complexity of the model, but rather the need to analyze large-scale data sets to obtain sufficiently accurate genetic parameter estimates. Both EM and Newton-type methods use the first derivative of the likelihood in the maximization. Calculation of the first derivative requires the elements of the inverse of the coefficient matrix of MME. When the data are large and the model has many random effects, it is often impossible to calculate the exact inverse of the coefficient matrix of MME using direct methods. Instead, the inverse can be estimated by MC sampling methods. The MC sampling scheme presented by García-Cortés et al. (1992) is used both in the EM and the Newton-type REML methods in the following.

### 2.4.1 MC EM REML

The idea in García-Cortés et al. (1992) is to estimate the elements of the inverse coefficient matrix $\mathbf{C}$ by generating samples from the same distribution as the original data: $\tilde{\mathbf{y}}^{h} = \mathbf{Z}\tilde{\mathbf{u}}^{h} + \tilde{\mathbf{e}}^{h}$, $h = 1,\ldots,s$. Here $\tilde{\mathbf{u}}^{h}$ denotes the *h*<sup>th</sup> sample drawn, such that $\tilde{\mathbf{u}}^{h} \sim N(\mathbf{0},\mathbf{G})$. After solving the MME for $\hat{\mathbf{u}}^{h}$ using the sampled data as observations, we can calculate average PEV over the samples. As a matter of fact, a variety of methods can be used to calculate the PEV (Hickey et al., 2009). Two of the methods in Hickey et al. (2009) seem to be useful in many models and are here called methods 1 and 2 by García-Cortés et al. (1995). In linear mixed effects model theory, the true variance of random effects is the variance of predicted values plus the variance of prediction errors. On this basis, method 1 by García-Cortés et al. (1995) defines that approximation of the trace part for the genetic random effect in equation (12) can be obtained via

$$(20)\quad SS^{h}_{\sigma_u^2} = q\sigma_u^2 - \left(\hat{\mathbf{u}}^{h}\right)^{T}\mathbf{A}^{-1}\left(\hat{\mathbf{u}}^{h}\right)$$

and similarly approximation of the trace part for the residual effect in equation (13) can be obtained via

$$(21)\quad SS^{h}_{\sigma_e^2} = n\sigma_e^2 - \left(\hat{\mathbf{e}}^{h}\right)^{T}\left(\hat{\mathbf{e}}^{h}\right).$$

In method 2 by García-Cortés et al. (1995) the trace of the inverse of the coefficient matrix is estimated based on the actual prediction errors of the simulated effects and estimated effects solved from the model given the simulated data:

$$(22)\quad SS^{h}_{\sigma_u^2} = \left(\tilde{\mathbf{u}}^{h} - \hat{\mathbf{u}}^{h}\right)^{T}\mathbf{A}^{-1}\left(\tilde{\mathbf{u}}^{h} - \hat{\mathbf{u}}^{h}\right)$$

and

$$(23)\quad SS^{h}_{\sigma_e^2} = \left(\tilde{\mathbf{e}}^{h} - \left(\tilde{\mathbf{y}}^{h} - \hat{\mathbf{y}}^{h}\right)\right)^{T}\left(\tilde{\mathbf{e}}^{h} - \left(\tilde{\mathbf{y}}^{h} - \hat{\mathbf{y}}^{h}\right)\right).$$

Now the new MC EM REML estimates are

$$(24)\quad \hat{\sigma}_u^2 = \tfrac{1}{q}\left(\hat{\mathbf{u}}^{T}\mathbf{A}^{-1}\hat{\mathbf{u}} + \tfrac{1}{s}\sum_{h=1}^{s} SS^{h}_{\sigma_u^2}\right)$$

and

$$(25)\quad \hat{\sigma}_e^2 = \tfrac{1}{n}\left(\hat{\mathbf{e}}^{T}\hat{\mathbf{e}} + \tfrac{1}{s}\sum_{h=1}^{s} SS^{h}_{\sigma_e^2}\right).$$

The larger the $s$, the better is the approximation of the trace.

### 2.4.2 MC Newton-type REML

The simple sampling method to approximate PEV presented in García-Cortés et al. (1992) can be utilized in Newton-type methods as well. Recall that the Newton-type methods are based on first and either exact or approximated second derivatives of the log likelihood function. The first derivatives of the log likelihood function contain the same trace elements as the EM algorithm and the same sampling based approximation can hence be used:

$$(26)\quad \frac{\partial l(\boldsymbol{\theta}\mid\mathbf{y})}{\partial\sigma_u^2} = -\frac{1}{2}\left(\frac{q}{\sigma_u^2} - \frac{\hat{\mathbf{u}}^{T}\mathbf{A}^{-1}\hat{\mathbf{u}} + \frac{1}{s}\sum_{h=1}^{s} SS^{h}_{\sigma_u^2}}{\sigma_u^4}\right)$$

and

$$(27)\quad \frac{\partial l(\boldsymbol{\theta}\mid\mathbf{y})}{\partial\sigma_e^2} = -\frac{1}{2}\left(\frac{n}{\sigma_e^2} - \frac{\hat{\mathbf{e}}^{T}\hat{\mathbf{e}} + \frac{1}{s}\sum_{h=1}^{s} SS^{h}_{\sigma_e^2}}{\sigma_e^4}\right).$$

By definition, the expected information matrix at convergence is $E(\mathbf{H}(\boldsymbol{\theta})) = E\left(\mathbf{J}(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta})^{T}\right)$. Use of the MC method with independent and identically distributed samples enables the approximation of the information matrix by the variances of the gradients over the samples within each NR REML round. However, method 1 using equations (20) and (21) for estimation of trace needs to be used in computation of the sampling variance of the gradients, because method 2 using equations (22) and (23) only gives the variances of PEV. In particular, the information matrix $\mathbf{H}(\boldsymbol{\theta})$ for the single trait model in (1) is approximated by the following elements:

$$(28)\quad Var\left(\frac{\partial l^{h}(\boldsymbol{\theta}\mid\mathbf{y})}{\partial\sigma_u^2}; h=1,\ldots,s\right) = Var\left(\frac{\frac{1}{2}\left(\hat{\mathbf{u}}^{h}\right)^{T}\mathbf{A}^{-1}\left(\hat{\mathbf{u}}^{h}\right)}{\sigma_u^4}; h=1,\ldots,s\right)$$

$$(29)\quad Var\left(\frac{\partial l^{h}(\boldsymbol{\theta}\mid\mathbf{y})}{\partial\sigma_e^2}; h=1,\ldots,s\right) = Var\left(\frac{\frac{1}{2}\left(\hat{\mathbf{e}}^{h}\right)^{T}\left(\hat{\mathbf{e}}^{h}\right)}{\sigma_e^4}; h=1,\ldots,s\right)$$

and

$$Cov\left(\frac{\partial l^h(\boldsymbol{\theta}\,|\,\mathbf{y})}{\partial\sigma_u^2},\frac{\partial l^h(\boldsymbol{\theta}\,|\,\mathbf{y})}{\partial\sigma_e^2};h=1,\ldots,s\right)$$

$$(30)\quad =Cov\left(\frac{\frac{1}{2}\left(\hat{\mathbf{u}}^h\right)^T\mathbf{A}^{-1}\left(\hat{\mathbf{u}}^h\right)}{\sigma_u^4},\frac{\frac{1}{2}\left(\hat{\mathbf{e}}^h\right)^T\left(\hat{\mathbf{e}}^h\right)}{\sigma_e^4};h=1,\ldots,s\right).$$

Interestingly, MC sampling is needed only for the estimation of first derivatives in AI and BM methods. Both the AI matrix in MC AI REML and the approximated inverse of information matrix in MC BM REML can be calculated using the current VC estimates.

### 2.4.3 Convergence criterion

Definition of the convergence criterion is not straightforward in methods based on MC sampling because of the sampling variation. For example, a criterion like the relative round-to-round change in consecutive VC estimates becomes unreliable. Reducing the MC noise by increasing the number of MC samples is impractical because that makes the estimation computationally inefficient (Booth and Hobert, 1999). To reduce the effect of sampling variation, a linear regression on the latest rounds of estimates may be fitted. An idea of this approach is predicting VC estimates by using linear regression on estimates of previous REML rounds and replacing the change in consecutive VC estimates by the changes in linear predictions. The greater the number of rounds used in a linear regression, the smaller is the effect of MC noise on the convergence criterion. For the Newton-type methods, convergence may be checked by a similar method. Hence the approach is the same but the prediction method is applied to the gradients instead of the estimates.

## 2.5 Testing of methods

Two simulated data sets and one field data were used to illustrate the linearization and the MC REML methods described here. The first simulated data contained growth curve data and was used to illustrate the linearization method, whereas the other simulated data and the field data were used to compare the VC estimation by different REML methods. The outlines of the data sets and models are described below and displayed in Tables 1 and 2, respectively. Lastly, the programs used in the analyses are described.

Table 1. Structure of the data sets used in the studies I, II and III.

| Response | No. of observations | No. of ao[1] | No. of ap[2] | No. of eq[3] | To study performance of |
|---|---|---|---|---|---|
| Simulated growth data | Live weight 144,000 | 4,800 | 210 | 15,036 | Linearization for full data (I) |
| | Live weight 54,405 | 4,800 | 210 | 15,036 | Linearization for truncated data(I) |
| | Live weight 58,799 | 4,800 | 210 | 15,036 | Linearization for partly truncated data (I) |
| Simulated 305-day data | Milk yield 3,000<br>Fat yield 3,000 | 3,000 | 3,150 | 6,400 | MC EM REML and new convergence criterion (II) |
| | Milk yield 569<br>Fat yield 569 | 569 | 715 | 1,450 | MC Newton-type methods and new convergence criterion (III) |
| Field TD data | Milk yield 51,004<br>Fat yield 25,316<br>Protein yield 25,316 | 5,399 | 10,822 | 160,221 | MC EM REML and new convergence criterion (II) |

[1] animals with observations
[2] animals in pedigree
[3] equations in MME

Table 2. Description of models used in the studies I, II and III.

| Response | Fixed effects | Random effects | No. of VC | Study |
|---|---|---|---|---|
| Simulated live weights | Sex | Additive genetic sire + Non-genetic animal | 13 | Applicability of linearization method (I) |
| Simulated 305-day milk and fat yield | Herd | Additive genetic animal | 6 | Performance of MC based methods and new convergence criterion (II and III) |
| Field TD data for milk, protein and fat | Herd and TD interaction + lactation curve | Additive genetic animal lactation curve + Non-genetic animal lactation curve | 96 | Feasibility of MC EM REML and new convergence criterion for large data sets (II) |

### 2.5.1 Data sets and analyses

To study the effectiveness of the linearization method, the first simulated data contained live weights based on the nonlinear Gompertz function (I). The simulated data included growth records from 4800 tested animals and the pedigree included 210 sires (Table 1). The statistical model included one fixed effect, random additive genetic sire effects, random non-genetic animal effects other than the sire and random residuals. Hence, the total number of variance components to be estimated was 13 (Table 2). The method was examined through the analysis of full data and two different subsets of the data. The general performance of the linearization method was tested with the full data. The second data was a truncated time trajectory set, and it was used to test the performance of the method for incomplete data (Table 1). In practice incomplete data are common e.g. in pig production, where adult weights are unavailable due to early slaughter ages. The third data combined the first two data sets by having some animals with complete data and most animals with incomplete data (Table 1). The VC estimates were only solved by analytical EM REML because of its stability. The results are based on 50 simulation replicates. The performance was measured by relative bias and relative standard deviation, as percentage from the true value, for the VC estimates.

The second simulated data was used to compare different MC REML and analytical REML methods (II and III). Because large data sets appear mostly in dairy cattle, the data structure resembled the dairy cattle observations for 305-day milk and fat yields. For the EM REML comparisons, records were simulated for 3000 cows with a pedigree comprising 3150 animals (Table 1). A subset of this data, with 569 animals with records and a total of 715 animals in the pedigree, was used in comparisons of the Newton-type REML methods (Table 1). Phenotypic records were simulated by a bivariate linear model with fixed herd effects, random additive genetic animal effects and random residuals. The total number of estimated VC was 6 (Table 2). MC EM REML, as well as each MC Newton-type REML method, was tested with 20, 100 and 1000 MC samples per REML iteration round.

The computing efficiency of MC EM REML was illustrated by a multiple-trait random regression TD model applied to true Finnish Ayrshire TD data (II). A subset from a large data set was also taken to enable calculation of analytical EM REML estimates. The data comprised 5399 animals with records and 10822 animals in the pedigree. There were 51004 TD records for milk yield, of which approximately half were associated with observations for protein and fat yield (Table 1). The multiple-trait model consists of a vector of fixed herd and TD interactions, a vector of fixed lactation curve regression effects, a vector of random non-genetic regression effects, a vector of random additive genetic regression effects, and a vector of random residuals. The total number of estimated VC for this model was 96 (Table 2). Three alternatives for the number of MC samples within an MC EM REML round (20, 5 and 1) were tested.

### 2.5.2 Software

The general computing software MiX99 (Lidauer et al., 2011) and DMU (Madsen and Jensen, 2012) were used for solving large scale MME and VC estimates when possible. New methods required some modifications however. Implementation of the linearization procedure required the Gompertz function formulas to be included in MiX99. Only MME for linearized model was solved by MiX99. The pseudo data and working variables were written to an external file and read by DMU for the estimation of VC. Thus, there was no need to make changes to DMU. Also MC EM REML was implemented in MiX99. New modules needed in the software were data generation and calculation of quadratic forms. The available routines to solve MME were used as they had been implemented in MiX99. However, all three different MC based Newton-type REML methods appeared to be too complicated to be included in MiX99 for this study. These methods were implemented by R software (R Core Team, 2014) instead. Consequently, the analyzed data had to be of moderate size.

# 3 Main results

The results from the linearization study are presented first (I). Then, the efficiency of the MC method in EM REML and applicability of the MC method in Newton-type methods are explored (II and III). Finally, assessment of convergence is considered (II and III).

## 3.1 Linearization

The performance of the linearization method was investigated based on analysis of 50 simulation replicates of pig growth. The summary of the results in Table 1 and Table 2 in Article I are given here. The full data with observations from the entire growth period worked satisfactorily. The least accurate estimates were obtained for the covariance components of genetic effects (the relative bias and the relative standard deviation were on average 12.5% and 61.8% respectively). Truncated time trajectory increased the relative bias and standard deviation of the VC estimates for all effects. Analyses of the truncated data set were most unstable for adult weight parameters ($\alpha$) but they were also unstable for parameters related to the exponential decay of the initial growth rate ($\kappa$). However, having even a small proportion of animals with weights from birth to adult weight improved the results when compared to the estimates produced from completely truncated data. An improvement was especially seen in the estimates of genetic covariance components (relative bias and standard deviation from on average 33.4% and 118.7% to 6.6% and 74.0% respectively), which were the most unstable parameters in the truncated time trajectory data.

## 3.2 MC EM REML

VC estimates for linear mixed effects models by MC EM REML and analytical EM REML were monitored along the iteration process (Figure 1 and Figure 3 in Article II). When the number of MC samples was large, the estimates by MC EM REML followed the corresponding analytical estimates well. Variation around the analytical estimates was larger when the number of MC samples was reduced. Although deviations from the analytical EM REML estimates can be large with one MC sample, on average VC estimates by MC EM REML did stay at the same level as with more samples. Furthermore, heritability as well as genetic and phenotypic correlations corresponded well with the analytical estimates despite the differences in the VC parameter estimates (Table 2 in Article II). For the field data set, MC EM REML proved to be far superior to analytical EM REML both in computing time and memory need. Calculation of 1000 EM REML rounds using the analytical EM REML method needed 56 days while the MC EM REML method required 65, 20 and 7 hours with 20, 5 and 1 MC samples per MC EM REML round, respectively. Although the data used for the TD model were small, the large number of VC led to time-consuming analysis because of the slow convergence of the EM algorithm.

### 3.3 MC Newton-type REML

The MC method in different Newton-type methods for estimation of VC was found to be feasible. The approximated information matrix for MC NR REML was based on variance over MC samples within each REML round. Hence, the sampling variation depended on the number of MC samples per REML round. In contrast, the approximated information matrices in MC AI REML and MC BM REML had sampling variation only through the current estimates and gradients used to update the information matrix. Compared to MC EM REML, convergence of second derivative methods was fast but sampling variation of estimates along the iteration process was large for all Newton-type MC methods (Figure 1 in Article III). The number of MC samples needed was found to be dependent on the Newton-type method used. Based on the analyses with the number of MC samples 20, 100 and 1000, the minimum number of MC samples to obtain sufficiently accurate information matrix estimates for the small simulated data was 100, 20 and 1000 for MC NR REML, MC AI REML and MC BM REML, respectively.

### 3.4 Assessment of convergence

A convergence criterion that takes into account the sampling variation was developed. First the new criterion was tested with the analytical EM REML. The values based on the new convergence criterion stayed above those of the traditional criterion measuring round to round change (Figure 2 and Figure 4 in Article II). When the linear prediction convergence criterion was applied to the MC EM REML estimates calculated from large number of samples, the criterion followed the corresponding values based on the analytical EM REML estimates. A decrease in the number of MC samples increased the fluctuation in the convergence criterion values (Figure 2 and Figure 4 in Article II). The larger MC variation in the Newton-type methods resulted in extra challenges in defining the convergence. However, a convergence criterion similar to MC EM REML was also found to be feasible for MC AI REML estimates (III).

# 4 Discussion

The aim of this study was to find an efficient method for estimation of VC of nonlinear and large linear mixed effects models. In this chapter we discuss our findings concerning linearization, MC EM REML and MC Newton-type REML. Finally, recommendations for future studies are considered.

## 4.1 Linearization

An advantage of the method presented by Wolfinger and Lin (1997) is its generality. Their linearization approach can be used for different types of models because it was developed for general nonlinear mixed effects models. For example, generalization of the simple Gompertz model used in this study on multiple effects and traits is straightforward. As also discussed by Pinheiro and Bates (1995), the main advantage of the linearization is a possibility to use efficient programs that maximize the restricted maximum likelihood for the linear mixed effects models, even though the two-step iterative procedure with each step itself being iterative can be regarded as computationally intensive.

Linearization enables the use of linear mixed effects model procedures if the approximation is valid. In order to arrive at linear mixed model equations, the dependency of the VC on the fixed effects through the working incidence matrices has to be ignored. Both Lindstrom and Bates (1991) and Wolfinger and Lin (1997) justified this by appealing to intrinsic nonlinearity instead of the nonlinearity of the parameters. Either way, asymptotic correlations between the estimators of the VC and the fixed effects were not found in several data sets analyzed by Pinheiro and Bates (1995). The full data set in this simulation study shows that linearization also works moderately well for the Gompertz function. Furthermore, the method was successfully used for real pig growth data (Koivula et al., 2008). However, statistical properties of estimation methods for linear models do not transmit to nonlinear models due to the approximation. For example, in the simulation study of Pinheiro and Bates (1995), the bias correction ability of REML depended on the nonlinear model used.

However, the success of the approximation method to analyze the nonlinear model greatly depends on the amount and nature of available information (Vonesh, 1996; Wolfinger and Lin, 1997). As expected, the truncated data analysis in this study showed that if observations are missing from the tails of growth curves of all the animals, uncertainty increases and the estimation method can become distorted. This distortion diminished considerably when at least some of the animals had observations up to or close to their mature weights. Close proximity of slaughter time and the inflection point is common in pig field data. Due to selection of tested pigs for breeding, there may be pigs that have observations up to their adult weights. Our simulation study showed that even a small fraction of fully observed animals improved the results when compared to the estimates produced from completely truncated data. General recommendations about the proportion of animals with full data were not made, because the results may be influenced by the population structure.

Convergence of the linearization method has been reported to be difficult by Meza et al. (2007) among others. First of all, convergence depends on the starting values for the fixed effects. Also in our study, good starting values were found to be important. General and simple equations to provide good starting values may be difficult to define. One choice could be to fit a simpler model without random effects first (Lindstrom and Bates, 1990). To improve convergence, one possibility is a different parameterization of the Gompertz model. For example, a multiplicative model using log-transformed data can be analyzed (Vuori et al., 2006; Koivula et al., 2008). This takes into account the common nature of the residuals in real growth data, but also removes the dependence of the derivative of the adult weight parameter on the others.

Instead of the approximation of the likelihood function by linearization, Walker (1996) presented MC EM for nonlinear random models. Studies about exact maximization methods for ML and REML involving numerical integration techniques or MC methods have continued (e.g. stochastic approximation EM (SAEM) in Kuhn and Lavielle (2005) and REML SAEM in Meza et al. (2007)). A comparison of the linearization procedure with the SAEM method in Kuhn and Lavielle (2005) revealed that SAEM was better in terms of robustness to starting values for VC and the accuracy of the estimates. Similar to linear mixed effects models, REML SAEM improved the accuracy of the VC estimates even more, especially in unbalanced cases (Meza et al., 2007). In addition, Meza et al. (2007) found REML SAEM to be very stable. The main disadvantage of the linearization methods in their study was nonconvergence for large number of data sets while REML SAEM converged in all cases.

The EM algorithms may be very slow to converge for some problems. For example when variance parameters are small or the dimensions of random effects are huge. The parameter expansion version of EM (PX-EM) accelerates EM while maintaining its stability (Liu et al., 1998; Foulley and van Dyk, 2000). Also MC based PX-EM has been presented for nonlinear mixed effects models. Wu (2004) recommended it for both exact and approximate inferences, although the benefit depends on the model used and the analyzed data. In Lavielle and Meza (2007), a PX version of SAEM (PX-SAEM) for nonlinear model converged in all runs, while plain SAEM failed to converge in 20 out of 100 runs. The PX-SAEM method improved the convergence particularly during the first iterations when the parameters were highly correlated. This would be the case also in the growth curve analysis between the three parameters of the Gompertz function.

The good properties found for MC based methods in analysis of nonlinear mixed effects models make the approach interesting. However, MC EM for nonlinear model is somewhat different than MC EM REML studied in this thesis. The main difference is that there was no need to sample fixed effects in the VC estimation of linear mixed effects models, while it is needed for estimation of nonlinear models. Because of the more complicated simulation step, this is likely to lead to the use of MCMC instead of the MC method (Kuhn and Lavielle, 2005).

## 4.2 MC EM REML

In general, MC EM REML converged to the same solutions as analytical EM REML, and a small number of MC samples did not introduce systematic bias to the estimates of genetic parameters. However, the required number of MC samples to obtain estimates with an acceptably low MC error depended on the size of the data. For the field data, MC EM REML gave reliable estimates with five MC samples per round and reasonable results with just one MC sample within a round. Thus, analysis of the field data demonstrated the potential of MC EM REML for large and complex models. VC estimates were obtained with relatively fast computing times compared to analytical EM REML and also with low memory requirements corresponding to those needed for breeding value estimation.

Recommendations for increasing the number of MC samples presented in the MC EM literature tend to suggest numbers that may become too high for our purposes in the context of large data sets and many VC (e.g. in Booth and Hobert (1999), Ripatti et al. (2002) and Levine and Fan (2004)). Delyon et al. (1999) used SAEM method which eliminates the need of increasing the number of MC samples as it accounts for estimates from previous REML rounds. The method requires specifying a smoothing parameter that however is not simple to choose. The commonly used recommendation (e.g. Jaffrézic et al., 2006) leads to fast reduction of MC noise. Therefore the VC estimates need to be close to the final values when the SAEM method is started – a situation that may not necessarily be viable with only one MC sample within an MC EM REML round.

In addition to the number of MC samples, the time required to solve MME is the most critical point affecting the computing time by the MC EM REML method. An efficient solving of the MME is vitally important for the presented MC EM REML method, because it requires solutions for the location parameters from the data and from a number of simulated samples of data within each MC EM REML round. We used preconditioned conjugate gradient iteration where a block diagonal preconditioner matrix was used to approximate the MME (Strandén and Lidauer, 1999). However, the time needed in simulation of the samples and computing quadratics proved to be negligible.

Typically, in MCMC analyses, it is recommended to check the convergence by using multiple chains and plotting the parameter estimates along the iteration process. In studies of the MC EM REML method, an automated procedure has been suggested to avoid postprocessing of the chain of estimates (e.g. in Booth and Hobert (1999) and Ripatti et al. (2002)). Also in this study a stopping rule that can be calculated during the analysis was introduced. Basically this alternative convergence criterion for MC EM REML was based on the relative differences in the linear regression coefficients from zero. The applied convergence criterion monitored the progress of convergence and was only little influenced by MC noise. The criterion also worked reasonably well with a small number of MC samples – a property that may be useful when analyzing complex models with many VC to be estimated. Applying the same value for the MC EM REML analysis as for the analytical EM REML analysis unnecessarily increased the number of MC EM REML rounds, although it protects against uncertainty due to sampling variation.

### 4.3 MC Newton-type REML

The use of MC Newton-type methods instead of the MC EM algorithm in REML speeded up the convergence, i.e., led to a lower number of iterations until convergence. However, the sampling variation of the estimates increased compared to the MC EM REML analysis. This is due to multiplication of the gradient by the inverse of the information matrix. Due to different styles for forming an approximation of the information matrix, sampling variation differed between the Newton-type methods.

The MC NR REML method is easy to implement, but may require a large number of MC samples to produce sufficiently accurate approximations of the variances of the first derivatives over samples. MC AI REML, in contrast, works better even with small number of MC samples, because the AI matrix has no extra sampling noise as it depends only on variance parameters estimated in the previous round. With the same number of MC samples, iterations in MC AI REML are computationally more demanding than in MC NR REML because the MME system needs to be solved at each MC AI REML round as many times as there are VC parameters to be estimated. An advantage of the MC NR and the MC AI REML methods is the possibility to obtain easily standard errors of estimates. MC BM REML is computationally the least expensive of the considered methods when the number of REML rounds and the number of MC samples are kept the same. To circumvent evaluation of the information matrix, BM REML corrects the approximation of the inverse of information matrix from round to round based on the gradients. Analytical BM REML worked reasonably well with the small data set in our study, whereas the MC BM REML method required a large number of MC samples. This indicates the sensitivity of the MC BM REML method to changes in gradients from round to round. Furthermore, extra computations are needed for standard errors after convergence has been reached.

If each round of iteration in the studied Newton-type methods requires many more samples than in MC EM REML, the overall solving time will only be reduced in cases where the Newton method can enhance the convergence dramatically. Obtaining a fast algorithm for MC based REML estimation requires the development of a practical convergence criterion for the Newton-type methods. Although convergence is the same regardless of the number of MC samples, MC variation affects the values of the convergence criteria. The criterion presented by Booth and Hobert (1999) is based on a change in consecutive parameter estimates relative to their standard errors whereas the criterion by Kuk and Cheng (1997) relies on the gradient vector and its standard errors. Both of these may need an enormous number of MC samples to work properly. Our convergence criterion for the MC Newton-type methods is based on relative differences in the linear predictions of gradients. This seemed to be a suitable and simple approach for large-scale genetic evaluations.

Kuk and Cheng (1997) and Gauderman and Navidi (2001) presented an MC NR method which was based on the use of MC approximation for the observed information matrix used by Wei and Tanner (1990). However, the estimator of the information matrix was often not positive definite. In both studies, modified estimator of the observed information matrix was presented to guarantee positive definiteness. The new estimator in Gauderman and Navidi (2001) approximates the expected information matrix, like the one presented here. It requires only the first derivatives of the log-likelihood. However, they made no comparisons with the MC EM algorithm. Instead in an example by Kuk and Cheng (1997), their improved MC NR method was indeed found to be faster than the MC EM algorithm. One interesting approximation of the information matrix is called the empirical Fisher information matrix (Meilijson, 1989; Scott, 2002), where the idea is to calculate the covariance over individuals' gradients instead of the sampled gradients.

## 4.4 Recommendations for further research

Simulation study about the linearization method was done only using the Gompertz function with a model, data structure and parameters mimicking a pig breeding data set. For other studies, like Gompertz growth curve in chicken, or perhaps for Brody's function, the linearization method may work differently. In all cases the two-step iterative algorithm studied here would need careful examination with regard to the initial values and convergence criteria.

For linear mixed effects models, the MC based EM algorithm is superior only for very large data sets. Therefore, number of MC samples can be chosen to be small, e.g. 1-3. I would suggest further study to focus on detection of convergence instead of increasing the sample size along the iteration. In practice, it would be computationally efficient to use SAEM or another averaging method after some (perhaps loose) convergence is met. Suitable critical values for convergence criterion should be studied with different kinds of models and data sets both for MC EM and MC Newton-type REML methods.

Our results show that the use of MC methods in different Newton-type methods for VC estimation is feasible, although there was variation in efficiency between the implementations. However, analysis of our small simulated data implies that the number of MC samples needed for accurate estimation depends on the method used. This work encourages testing the performance of the presented methods in analysis of large-scale problems. As models grow larger and more complex, the efficiency of different MC Newton-type methods becomes more difficult to predict. Further experience is especially needed on the behavior of MC BM REML in VC estimation of complex models.

The estimates of the MC Newton-type REML analyses presented here were regressed towards corresponding EM REML estimates whenever they were outside the parameter space. Yet, this does not guarantee convergence to the true solutions, especially with respect to the MC BM REML method. One way to increase the robustness of estimation methods is reparameterization of the VC matrices by Cholesky decomposition (Groeneveld, 1994). For some reason however, Cholesky decomposition of the VC matrix did not work in Wolfinger and Lin (1997) for nonlinear mixed effects model. Anyway the performance of this option is worth considering in future VC estimation studies.

In conclusion, all the methods studied gave positive results with respect to efficient VC estimation for large scale linear and nonlinear mixed effects models. A recommendation for further studies about MC EM REML for linear mixed effects models is related to the assessment of convergence. For the MC Newton-type methods and the linearization above all more experience would be needed with different kinds of models and data sets. Therefore modifications in efficient software are needed in order to use MC based Newton-type methods. The literature considered here presents many proposals for stability, like reparameterization and PX-EM, which could be considered in the future.

# References

Blasco A., Piles M. and Varona L. 2003. A Bayesian analysis of the effect of selection for growth curves in rabbits. Genetics Selection Evolution, 35, 21–41.

Booth J.G. and Hobert J.P. 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. Journal of the Royal Statistical Society: Series B, 61, 265–285.

Broyden C.G. 1965. A class of methods for solving nonlinear simultaneous equations. Mathematics of Computation, 19, 577–593.

Davidian M. and Giltinan D.M. 1995. Nonlinear Models for Repeated Measurement Data. Chapman & Hall, London.

Delyon B., Lavielle M. and Moulines E. 1999. Convergence of a stochastic approximation version of the EM algorithm. The Annals of Statistics, 27, 94–128.

Dempster A.P., Laird N.M. and Rubin D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B, 39, 1–38.

Dennis J.E. and Moré J.J. 1974. A characterization of superlinear convergence and its application to quasi-Newton methods. Mathematics of Computation, 28, 549–560.

Foulley J.-L. and van Dyk D.A. 2000. The PX-EM algorithm for fast stable fitting of Henderson’s mixed model. Genetics Selection Evolution, 32, 143–163.

Gauderman W.J. and Navidi W. 2001. A Monte Carlo Newton-Raphson procedure for maximizing complex likelihoods on pedigree data. Computational Statistics & Data Analysis, 35, 395–415.

García-Cortés L.A. and Sorensen D. 2001. Alternative implementations of Monte Carlo EM algorithms for likelihood inference. Genetics Selection Evolution, 33, 443–452.

García-Cortés L.A., Moreno C., Varona L. and Altarriba J. 1992. Variance component estimation by resampling. Journal of Animal Breeding and Genetics, 109, 358–363.

García-Cortés L.A., Moreno C., Varona L. and Altarriba J. 1995. Estimation of prediction-error variances by resampling. Journal of Animal Breeding and Genetics, 112, 176–182.

Gilmour A.R., Thompson R. and Cullis B.R. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics, 51, 1440–1450.

Groeneveld E. 1994. A reparamete-rization to improve numerical optimization in multivariate REML (co)variance component estimation. Genetics Selection Evolution, 26, 537–545.

Guo S.W. and Thompson E.A. 1991. Monte Carlo estimation of variance component models for large complex pedigrees. IMA Journal of Mathematics Applied in Medicine and Biology, 8, 171–189.

Harville D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72, 320–338.

Harville D.A. 2004. Making REML computationally feasible for large data sets: Use of Gibbs sampler. Journal of Statistical Computation and Simulation, 74, 135–153.

Henderson C.R. 1953. Estimation of variance and covariance components, Biometrics, 9, 226–252.

Henderson C.R. 1984. Estimation of variances and covariances under multiple trait model. Journal of Dairy Science, 67, 1581–1589.

Hickey J.M., Veerkamp R.F., Calus M.P.L., Mulder H.A. and Thompson R. 2009. Estimation of prediction error variances via

Monte Carlo sampling methods using different formulations of the prediction error variances. Genetics Selection Evolution, 41, 23.

Hofer A. 1998. Variance component estimation in animal breeding: a review. Journal of Animal Breeding and Genetics, 115, 247–265.

Jaffrézic F., Meza C., Lavielle M. and Foulley J.-L. 2006. Genetic analysis of growth curves using the SAEM algorithm. Genetics Selection Evolution, 38, 583–600.

Jamshidian M. and Jennrich R.I. 1993. Conjugate gradient acceleration of the EM algorithm. Journal of the American Statistical Association, 88, 221–228.

Jensen J., Mäntysaari E.A., Madsen P. and Thompson R. 1997. Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. Journal of the Indian Society of Agricultural Statistics, 49, 215–236.

Johnson D.L. and Thompson R. 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. Journal of Dairy Science, 78, 449–456.

Klassen D.J. and Smith S.P. 1990. Animal model estimation using simulated REML. In: Proceedings 4th World Congress on Genetics Applied to Livestock Production XIII: 472–475. Edinburgh, 23–27 July 1990.

Koivula M., Sevón-Aimonen M.-L., Strandén I., Matilainen K., Serenius T., Stalder K.J. and Mäntysaari E.A. 2008. Genetic (co)variances and breeding value estimation of Gompertz growth curve parameters in Finnish Yorkshire boars, gilts and barrows. Journal of Animal Breeding and Genetics, 125, 168–175.

Kuhn E. and Lavielle M. 2005. Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis, 49, 1020–1038.

Kuk A.Y.C. and Cheng Y.W. 1997. The Monte Carlo Newton-Raphson algorithm. Journal of Statistical Computation and Simulation, 59, 233–250.

Lavielle M. and Meza C. 2007 A parameter expansion version of the SAEM algorithm. Statistics and Computing, 17, 121–130.

Levine R.A. and Fan J. 2004. An automated (Markov chain) Monte Carlo EM algorithm. Journal of Statistical Computation & Simulation, 74, 349–360.

Lidauer M.H., Matilainen K., Mäntysaari E.A. and Strandén I. 2011. Technical reference guide for MiX99. Release VI/2011. MTT Agrifood Research Finland. URL http://www.mtt.fi/BGE/Software/MiX9.

Lindstrom M.J. and Bates D.M. 1990. Nonlinear mixed effects models for repeated measures data. Biometrics, 46, 673–687.

Liu C., Rubin D.B. and Wu Y.N. 1998. Parameter expansion to accelerate EM: The PX-EM algorithm, Biometrika, 85, 755–770.

Louis T.A. 1982. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society: Series B, 44, 226–233.

Madsen P. and Jensen J. 2012. A user's guide to DMU. A package for analysing multivariate mixed models. Version 6, release 5.1. University of Aarhus, Denmark.

McCulloch C.E. 1997. Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association, 92, 162–170.

Meilijson I. 1989. A fast improvement to the EM algorithm on its own terms. Journal of the Royal Statistical Society: Series B, 51, 127–138.

Meza C., Jaffrézic F. and Foulley J.-L. 2007. REML estimation of variance parameters in nonlinear mixed effects models using the SAEM algorithm. Biometrical Journal, 49, 876–888.

Misztal I. 2008. Reliable computing in estimation of variance components. Journal of Animal Breeding and Genetics, 125, 363–370.

Mäntysaari E. and Van Vleck L.D. 1989. Restricted maximum likelihood estimates of variance components from multitrait sire models with large number of fixed effects. Journal of Animal Breeding and Genetics, 106, 409–422.

Nocedal J. and Wright S.J. 2006. Numerical optimization. Springer, New York, 2nd edition.

Patterson H.D. and Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545–554.

Patterson H.D. and Thompson R. 1974. Maximum likelihood estimation of components of variance. Proceedings of the 8th international biometric conference, 197–207.

Pinheiro J.C. and Bates D.M. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics, 4, 12–35.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rekaya R., Weigel K.A. and Gianola D. 2001. Hierarchical nonlinear model for persistency of milk yield in the first three lactations of Holsteins. Livestock Production Science, 68, 181–187.

Ripatti S., Larsen K. and Palmgren J. 2002. Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. Lifetime Data Analysis, 8, 349–360.

Scott W.A. 2002. Maximum likelihood estimation using the empirical Fisher information matrix. Journal of Statistical Computation and Simulation, 72, 599–611.

Searle S.R., Casella G. and McCulloch C.E. 1992. Variance components. John Wiley & Sons, Inc., New York.

Strandén I. and Lidauer M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. Journal of Dairy Science, 82, 2779–2787.

Thompson R. 1994. Integrating best linear unbiased prediction and maximum likelihood estimation. In: Proceedings 5th World Congress on Genetics Applied to Livestock Production XVIII: 337–340. Guelph, 7–12 August 1994.

Thompson R., Brotherstone S. and White I.M.S. 2005. Estimation of quantitative genetic parameters. Philosophical Transactions of the Royal Society B, 360, 1469–1477.

Vonesh E.F. 1996. A note on the use of Laplace's approximation for nonlinear mixed-effects models. Biometrika, 83, 447–452.

Vuori K., Strandén I. and Mäntysaari E.A. 2006. Solving large scale nonlinear mixed models. 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, MG, Brazil.

Walker S. 1996. An EM algorithm for nonlinear random effects models. Biometrics, 52, 934–944.

Wei G. and Tanner M. 1990. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association, 85, 699–704.

Wellock I.J., Emmans G.C. and Kyriazakis I. 2004. Describing and predicting potential growth in pig. Animal Science, 78, 379–388.

Whittemore C.T. and Green D.M. 2002. The description of the rate of protein and lipid growth in pigs in relation to live weight. The Journal of Agricultural Science, 138, 415–423.

Whittemore C.T., Tullis J.B. and Emmans G.C. 1988. Protein growth in pigs. Animal Production, 46, 437–445.

Wolfinger R.D. 1993. Laplace's approximation for nonlinear mixed models. Biometrika, 80, 791–795.

Wolfinger R.D. and Lin X. 1997. Two Taylor-series approximation methods for nonlinear mixed models. Computational Statistics & Data Analysis, 25, 465–490.

Wu L. 2004. Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. Journal of the American Statistical Association, 99, 700–709.

MTT is publishing its research findings in two series of publications:
MTT Science and MTT Growth.

MTT CREATES VITALITY THROUGH SCIENCE

MTT SCIENCE

**www.mtt.fi/julkaisut**

The MTT Science series includes scientific presentations and abstracts from
conferences arranged by MTT Agrifood Research Finland.
Doctoral dissertations by MTT research scientists
will also be published in this series.
The topics range from agricultural and food research
to environmental research in the field of agriculture.

MTT, FI-31600 Jokioinen, Finland.
email julkaisut@mtt.fi