Institut für Nutzpflanzenwissenschaften und Ressourcenschutz

Professur für Pflanzenzüchtung

Prof. Dr. J. Léon

Bayesian Adaptive Markov Chain Monte Carlo Estimation of Genetic Parameters

Inaugural-Dissertation

zur

Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr. agr.)

der

Hohen Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Zu Bonn

vorgelegt am 08-12-11

von

Boby Mathew

aus

Kottayam, Indien

Referent:	**Prof. Dr. Jens Léon**
Korreferent:	**Prof. Dr. Heiko Schoof**
Tag der mündlichen Prüfung:	**23-03-12**
Erscheinungsjahr:	**2012**

Acknowledgments

First and foremost, my utmost gratitude to my supervisor, Prof. Dr. Jens Léon for his encouragement, patience, motivation, enthusiasm, immense knowledge and continuous support throughout my PhD study. I consider it an honor to work with him and he has been my inspiration in the completion of this research work. I owe my deepest gratitude to Prof. Dr. Jens Léon for his excellent supervision.

I also express my deepest gratitude to Dr. Andrea Bauer for her guidance, kind support, valuable suggestions and discussions, that laid the foundations for my PhD Work.

Also I would like to express my sincere gratitude to Dr. Mikko J. Sillanpää and Dr. Petri Koistinen, university of Helsinki for their guidance and valuable suggestions throughout my research study.

Meanwhile I am grateful to Dr. H. Schumann and Mrs Anne Reinders for their technical guidance. Also I would like to thank Dr. A. Ballvora and Dr. A. Naz for the scientific discussions. I offer thanks to Mrs Annette Schneider who helped me to have a nice working atmosphere.

Special thanks to Ana for the valuable suggestions and comments which helped me to improve my thesis. Also I would like to thank Hedda von Quistorp and Karin Woitol for their technical assistance. I am grateful to my colleagues Bong-Song, Tigest, Ismail, Ranya, Mohammed, Wiebke, Melanie, Alexendra, Merle and Karola for their mutual support and friendly working atmosphere. Many thanks to my Indian friends for their kind support. For the financial support I am thankful to the Theodor-Brinkman-Graduate School of the Faculty of Agriculture at the University of Bonn.

Finally my deepest gratitude to my beloved parents, my brother and sister for their love, support and encouragement throughout my study.

List of Tables

List of Figures

Abbreviations

Abbreviation	Explanation
BV	Breeding Value
BLUP	Best Linear Unbiased Prediction
ESS	Effective Sample Size
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
M-H	Metropolis–Hastings
MVN	Multivariate Normal Distribution
REML	Restricted Maximum Likelihood

Zusammenfassung

Eine exakte Schätzung von genetischen Parametern ist entscheidend für ein leistungsfähiges genetisches Evaluierungssystem. Normalerweise werden REML- und Bayes-Verfahren für die Schätzung von genetischen Einflussfaktoren angewendet. Bei der Bayes-Methode werden die Informationen, die über einen Parameter durch A-priori-Wahrscheinlichkeitseinschätzung bekannt sind mit den Daten und Erfahrungen aus aktuellen Studien kombiniert und in eine A-posteriori-Verteilung überführt. In der vorliegenden Arbeit wird ein neuer, schnell anpassungsfähiger Markov Chain Monte Carlo (MCMC) sampling Algorithmus vorgestellt, welcher die Vorteile des Hybrid-Gibbs sampler mit denen des Metropolis-Hastings Algorithmus zur Einschätzung von genetischen Einflussfaktoren in linear mixed models mit mehreren Zufallsvariablen in sich vereinigt. Dieser neue MCMC Algorithmus arbeitet in 2 Stufen: im ersten Schritt wird der Hybrid Gibbs sampler genutzt, um eine effiziente vorgeschlagene Kovarianzstruktur für die Varianzkomponenten zu erlernen, während im zweiten Schritt der M-H Algorithmus zur Aufstellung neuer Werte basierend auf der erlernten Kovarianzstruktur aus Schritt 1 zur Anwendung kommt. Normalerweise verzögern die Abhängigkeiten unter den Zufallsvariablen die Annäherung der Markov-Kette an einen stationären Zustand. Also wurden diese Zufallsvariablen in einem weiteren Schritt von der Wahrscheinlichkeitsschätzung ausgeschlossen, um das Gemisch der Kette zu verbessern. Der neue Algorithmus zeigte gute Mischeigenschaften und war zweimal schneller als der Hybrid-Gibbs sampler, um eine a-posteriori-Verteilung von Varianzkomponenten zu erstellen, außerdem können bei dieser Methode auch mehrere Modes festgestellt werden. Mit der vorgeschlagenen exponentiellen Vorbewertung für Varianzkomponenten ist es weiterhin möglich solche Maximalwerte bei der posterior Verteilung auf den Wert Null

zu schätzen im Falle, dass keine Dominanz besteht. Die Durchführung der Methode wurde mit realen und simulierten Datensätzen veranschaulicht.

Abstract

Accurate estimation of genetic parameters is crucial for an efficient genetic evaluation system. REML and Bayesian methods are commonly used for the estimation of genetic parameters. In Bayesian approach, the idea is to combine what is known about the parameter which is represented in terms of a prior probability distribution together with the information coming from the data, to obtain a posterior distribution of the parameter of interest. Here a new fast adaptive Markov Chain Monte Carlo (MCMC) sampling algorithm is proposed. It combines both hybrid Gibbs sampler and Metropolis-Hastings (M-H) algorithm, for the estimation of genetic parameters in the linear mixed models with several random effects. The new adaptive MCMC algorithm has two steps: in step 1 the hybrid Gibbs sampler is used to learn an efficient proposal covariance structure for the variance components, and in step 2 the M-H algorithm is used to propose new values based on the learned covariance structure from step 1. Normally the dependencies among the random effects slow down the convergence of the MCMC chain. So in the second step of the algorithm those random effects were marginalized from the likelihood to improve the mixing of the chain. The new algorithm showed good mixing properties and was about twice time faster than the hybrid Gibbs sampling to produce posterior for variance components. Also the new algorithm was able to detect different modes in the posterior distribution. Moreover, the new proposed exponential prior for variance components was able to provide estimated mode of the posterior dominance variance to be zero in case of no dominance. The performance of the method was illustrated with field data and simulated data sets.

1 Introduction

The main goal of plant breeding is to change the genetics of the plants to develop new variates with desirable characteristics. To achieve these objectives, plant breeders cross thousands of plants each year and selecting the plants with desired characteristics are always difficult. The science of plant breeding has been changing rapidly with the new development in molecular biology techniques and statistical methods. Molecular biology techniques and statistical methods can remarkably improve the selection process, and since 1920s, statistical methods were applied to analyze gene action and distinguish heritable variation from variation caused by environment.

1.1 Phenotype and Genotype

Phenotype is the observable physical characteristic of a plant, which is determined by both genotype and environmental influences. The genotype of a plant is a function of effects of the genes and hence cannot be observed. Many genes are involved in the inheritance and the environment often plays a crucial role in the expression of the phenotype. Thus, the phenotypic value P_{ijk} of a plant k in a population depends on genotypic g_i and environmental e_j effects:

$$P_{ijk} = \mu + g_i + e_j + \varepsilon_{ijk} \tag{1}$$

where μ is the population mean and ε_{ijk} residual effect.

1.1.1 Phenotypic variation

Phenotypic variation is the degree to which plant varies and it is the fundamental for evolution by natural selection. Both genetic and environmental factors as well as interactions between them contribute to phenotypic variation

in plants. The genetic variation can be further subdivided into three components called additive, dominance and epistatic variances (Lynch and Walsh 1998). Additive genetic variance measures the genetic variation associated with the average effects of substituting one allele for another at a given locus. Dominance variance is due to the interaction between alleles in the same locus whereas epistatic variance is due to the interaction between alleles in different loci. The genetic properties of a population are often expressed in terms of gene frequencies and genotype frequencies. Phenotypic variance within a population is the result of genetic variance and environmental sources. So the total phenotypic variance V_P can be expressed as:

$$V_P = V_A + V_D + V_I + V_E + V_\varepsilon \tag{2}$$

where V_A is the additive genetic variance, V_D is the dominance genetic variance, V_I is the epistatic variance, V_E is the variance due to environmental effects and V_ε is the residual variance. The presence of non-additive effects complicates many formulations in quantitative genetic, but unfortunately it cannot be ignored. Ignoring the dominance effect can lead to biased estimates of additive genetic variance, also the dominance effect is difficult to separate from common environmental effects. The epistasis describes the non-additivity of effect between the loci and is often difficult to compute. The additive variance, which is the variance of breeding values can be expressed as:

$$V_A = 2pq[a + d(p - q)]^2 \tag{3}$$

Similarly the dominance variation can be expressed as:

$$V_D = (2pqd)^2 \tag{4}$$

The total genetic variance, V_G arising from one locus can be expressed as:

$$V_G = V_A + V_D + V_I$$

$$= 2pq[a + d(q - p)]^2 + [2pqd]^2 + ... \quad (5)$$

Here p is the dominant allele frequency and q is the recessive allele frequency in the population. And a and d are the additive and domiance effect respectiely.

1.2 Breeding Value (BV)

Breeding value estimate the ability of a plant to produce superior offspring based on the measurement of performance. Breeding values describe the genetic merit of an individual and hence its ability to produce superior offspring. So considerable effort has been devoted to develop new statistical methods to estimate the breeding values. It is important to consider the performance of the relatives while estimating the breeding values, because all offspring receive a one-half of alleles from each parent. With the help of statistical methods information from the performance of relatives can be considered while predicting the breeding values. This is often done with the use of additive and dominance relationship matrices calculated from the pedigree information. The relationship matrices are commonly calculated based on coefficient of coancestry: it is the probability, that two genes are identical by descent in two individuals. Calculation of coefficient of coancestry is based on several assumptions: 1) pedigree information of parents is accurate, 2) the base population of ancestors are unrelated, 3) effects of selection, whereas mutation and genetic drift are negligible. Piepho *et al.* (2008) has suggested that the additive variance and BV are often biased without the complete pedigree records. Panter and Allen (1995), De Souza *et al.* (2000), Pattee *et*

al. (2001), Bauer *et al.* (2006), Crossa *et al.* (2006) and Oakey *et al.* (2006) have shown that selection based on parental breeding value was superior to normal selection strategies in self-pollinating crops. Hence the estimation of breeding values can improve the selection among parental inbred lines of self pollinating crops. The practical objective of quantitative genetics is to find out how one can use the observations, made on the population as it stands to predict the outcome of any particular breeding method. Best Linear Unbiased Prediction (Henderson 1963, Henderson 1975) methods are commonly used for the prediction of breeding values.

Defined in terms of average effects, the breeding value of an individual is equal to the sum of average effects of the gene it carries. For a single locus with two alleles, the breeding values of the genotypes are:

Genotype	Breeding Value
A_1A_1	$2\alpha_1 = 2q\alpha$
A_1A_2	$\alpha_1 + \alpha_2 = (q-p)\alpha$
A_2A_2	$2\alpha_2 = -2q\alpha$

where α is the average effect of gene substitution, α_1 is the average effect of the gene A_1, α_2 is the average effect of the gene A_2, p and q are the gene frequencies of A_1 and A_2, respectively.

Generally breeding values are calculated either based on the own performance of a line or based on the breeding values of its parents. Most of the traits are controlled by multiple gene and it is often difficult to get exact measure of gene frequencies p and q with out the help of molecular data. So it is more practical to use the performance of the relatives to estimate the breeding values, because all offspring receive a one-half of alleles from each parent. In a random mating population the additive genetic variance is equivalent to the variance of breeding values of individuals (Lynch and Walsh

1998). Wall *et al.* 2005 showed that nonadditive effect play a crucial role on the ranking of breeding values. So it is important to consider dominance effects while estimating the breeding values.

1.3 Inbreeding

Inbreeding is the mating of individuals that are closely related through common ancestry. For breeders, it is a useful way of fixing traits in a breeding population. However, inbreeding holds potential problems, the gene-pool caused by continued inbreeding leads the deleterious genes to become widespread. Inbreeding will lead to the reduction of the mean phenotypic value of a population, called inbreeding depression (Falconer 1989). The response of a population to inbreeding depends primarily on the level of dominance genetic variance. In a study carried out by De Boer and Hoeschele (1993) it was shown that the presence of inbreeding induces nonzero covariances between additive and dominance effects. However, (Bauer *et al.* 2006; Oakey *et al.* 2006; Bauer and Léon 2008) predicted the breeding values (assuming no dominance) for the self-pollinating crops by accounting for inbreeding among the lines. When nonzero covariance exists due to inbreeding, computational procedures for estimation of the variance components are further complicated. However in the current study I considered datasets with inbreeding and without inbreeding.

1.4 Heritability

Quantitative traits are often polygenic (Lynch and Walsh 1998) and they are significantly influenced by environmental effects. The accurate estimation of allele frequencies in a population is often difficult, so it is easy to express genetic influences in terms of heritability. Hence the accurate estimation of

heritability plays a crucial role in selection process. Heritability measures the relative influence of environment on the development of a specific quantitative trait. Estimation of heritability (proportion of phenotypic variance attributable to genetic factors) and breeding values are of primary interest, in order to plan an efficient breeding program for the trait of interest. Heritability is often considered as the first step in unraveling the genetic basis of a trait. Heritability (in the broad sense) is often expressed as the ratio of genetic variance to phenotypic variance:

$$h^2 = \frac{V_G}{V_P} \tag{6}$$

The ratio V_A/V_P is called the heritability in the narrow sense and it expresses the extend to which phenotypes are determined by the genes transmitted from the parents. Accurate heritability estimates are important to identify the genetic variation present in the population. Hsu *et al.* (2005) have shown that pedigree information of reasonable size is one of the important factors affecting the heritability estimates.

1.5 Statistical Modeling

Statistical inference is drawing conclusion about unknown quantities from the observed data. To make inference it is necessary to fit the data with help of a statistical model. A statistical model is a set of mathematical equations which describe the behavior of a system under study. The model can depend on a set of model parameters and the inference of the model parameters, we are interested is called parameter estimation. There are two set of variables associated with a model, response and explanatory variables. Response variables are the outcome of a study and the response variable

are used for the prediction. Response variables are often called dependent variables or predicted variables. Explanatory variables are any variables that explains the response variables and often called independent variables or predictor variables. Explanatory variables can be continuous or categorical, a categorical variables are factors with two or more levels. The objective of statistical modeling is to fit the data to the model and the best model is the model that produces the least unexplained variation (the minimal residual deviance), subjected to the constraint that all the parameters of the model should be statistical significant. The structure of the model is:

$$response\ variable \sim explanatory\ variable(s)$$

Ideally one should include all relevant information in a statistical model. Selecting the important explanatory variable is always demanding in practice. In Bayesian inference statistical conclusions about the unknown quantities are made in terms of probability statements. And the probability statements are conditional on the observed data. In Bayesian concept a statistical model is usually represented as a pair (D,P), where D is the set of possible observations(data) and P the set of possible probability distributions on D.

1.6 Restricted Maximum Likelihood (REML) method

The Maximum Likelihood (ML) estimator of the variance components in a linear model can be biased. Restricted maximum likelihood (REML) accounts this problem by using the likelihood of a set of residual and is generally considered superior to ML. Patterson and Thompson (1971) introduced restricted maximum likelihood estimation (REML) as a method of estimating variance components for unbalanced incomplete block designs. The REML

approach keeps the estimator within the parameter space $(0, +\infty)$, and therefore, REML is a biased procedure. REML is often preferred to maximum likelihood estimation because it takes into account the loss of degrees of freedom in estimating the mean and gives unbiased estimates for the variance parameters. REML estimates are often less biased than the Maximum Likelihood Estimates. The drawback of REML is that the distribution properties of the estimators are not known, except asymptotically.

1.7 Bayesian Methods

Genetic data, that produce the observed data are often the results of complex and stochastic processes, therefore they cannot be studied without the use of probabilistic models. Bayesian inference, based on probability is a convenient way to deal with these sorts of problem. The main difficulty with likelihood methods are optimization problems such as multiple modes, solution of likelihood equations etc, whereas integration problem is more often associate with Bayesian approach. ML methods can be very sensitive to small data perturbations if the model includes two or more explanatory variables, that are hard to disentangle from each other. In Bayesian methods the posterior distribution summarizes uncertainty around the point estimate in a probabilistic form. In Bayesian approach, the idea is to combine what is known about the parameter (this knowledge is represented in terms of a prior probability distribution) with the information coming from the data (likelihood function), to obtain a posterior distribution of the parameter of interest. Bayes theorem, which provide the basis for the Bayesian inference is:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{7}$$

where $P(\theta)$ is the prior probability of the parameter θ, $P(D|\theta)$ is the likelihood of θ, and $P(\theta|D)$ is the posterior of θ given D.
Steps in the Bayesian approach include:

1. Specify distribution for each random variable in the model.
2. Combine the distribution into the joint posterior distributions.
3. Find the conditional marginal distributions from the joint posterior distribution.
4. Implement Markov Chain Monte Carlo(MCMC) method to maximize the joint posterior distribution.

Wang *et al.* (1993) and Sorensen and Gianola (2002) applied Bayesian methods for the prediction of breeding values. In Bayesian methods the standard computational approach is to use Markov chain Monte Carlo (MCMC) methods to draw samples from posterior distributions. Gibbs sampler and Metropolis–Hastings algorithm are the two commonly used Markov chain Monte Carlo (MCMC) methods. M-H algorithm is mainly used for models that are not conditionally conjugate. Gibbs sampler is a special case of Metropolis-Hastings sampling, wherein the random value is always accepted. In Gibbs sampling, the updater samples from the fully conditional posterior distribution, which is proportional to the likelihood function and the prior distribution through Bayes theorem. The Gibbs sampler is very widely applicable to broad class of Bayesian problems, where the direct simulation from the posterior distribution is not possible.

1.8 Prior Distributions

In the Bayesian framework there is no distinction between fixed and random effects, and fixed effect is a random variable for which the prior knowledge is vague. The choice of the prior is often considered as one of the important step in Bayesian analysis. One can use informative and non informative priors based on the amount of information available. If the data is very informative about the quantity being estimated, then an uninformative prior is an easy choice. But if the data are poor, then the posterior will be heavily influenced by the prior. In Bayesian analysis the prior information is combined with the information from the data to generate the posterior distribution.

1.8.1 Non informative priors

The application of Bayesian methodology often uses non informative priors. Non informative priors are used when there is little or no prior information is available. Uniform (Laplace, 1812) prior is one of the most widely used non informative priors. The inverse-gamma (ϵ, ϵ) is also used as a non informative prior in Bayesian analysis. But the resulting inference will be sensitive to ϵ, in case where σ is estimated to be near zero (Gelman, 2006).

1.8.2 Informative priors

An alternative approach is to use an informative prior. The selection of informative priors are based on the careful examination of expert knowledge.

1.9 Markov Chain

A Markov chain is a collection of random variables X_i with the property that the next state depends only on the current state. It is expected that the

Markov chain will converge to some equilibrium distribution, independently of the initial distribution after a number of transitions. This is one of most important property of a Markov chain. The initial probability distribution of the states of the chain and the matrix of transition probabilities are the two components of a Markov chain. These two components together guide the evolution of the Markov chain. The Markov property states that the future state of the system, given its current state depends only on the current state of the system. Thus:

$$P(X_{n+1}|X_1, X_2, .., X_n) = P(X_{n+1}|X_n)$$

1.10 Markov Chain Monte Carlo (MCMC)

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain. MCMC algorithms are based on Markov chains, which evolves in discrete time. MCMC methods have become an important computational tool in Bayesian statistics, because it allows samples to be drawn from complex posterior distribution. With MCMC one can draw simulations from a wide range of distributions. The general MCMC algorithm is as follows:

1. Set initial value x_1 and set counter i=1.

2. Generate next value, conditionally on the previous: $x_{i+1} \sim f(x|x_i)$, set counter $i = i + 1$.

3. Go to step 2, until required sample size is generated.

The Markov chains used in MCMC methods are homogeneous, the conditional distribution of $x^{(i+1)}/x^{(i)}$ does not depend on the index i. Convergence is

the one of the important property associated with a MCMC sampler and it measures whether the chain reached its stationary distributions. Generally the initial 1000 to 5000 (this is called the **burn-in** period) elements are discarded and then one of the various convergence tests are used to assess whether stationary distribution has been reached. There are many different versions of MCMC algorithms such as, slice sampling, Gibbs sampling, Metropolis algorithm and Metropolis-Hastings algorithm. Metropolis-Hastings algorithm and Gibbs sampler are the commonly used MCMC methods. Generally a poor starting value can greatly increase the burn-in period.

1.11 Mixing

Mixing is another important property of MCMC, chain is said to be poorly mixing if it stays in small regions of the parameter space for long period of time. Mixing refers to the dependence of X_i and X_{i+t}. If the chain has good mixing then the dependence between X_i and X_{i+t} decays rapidly as a function of t. If the target distribution is multi-modal then poor mixing can arise and the value can traps near one of the modes.

1.12 Identifiability Problem

General Markov chain Monte Carlo (MCMC) methods are facing a wide range of practical and theoretical issues and parameter identifiability is the one of the main problem faced by MCMC. In linear mixed models the random effects are generally susceptible to identifiability problem. Also identifiability occurs when the posterior have multiple modes and the conventional MCMC samplers will fail to movie between different modes in the posterior. When the random effects or variance components fitted to the model have multiple solutions among their parameter spaces given the observed data, such parameters

are said to be unidentifiable. Recently Wall *et al.* (2005) has shown that non-additive random genetic effects (epistatic interaction and dominance deviation) are important in the estimation of breeding values. Unfortunately, in practice identifiability problems complicate the estimation of non-additive random genetic effects (Misztal 1997; Waldmann *et al.* 2008).

Since the 1980's, the use of Markov Chain Monte Carlo (MCMC) methods have revolutionized the Bayesian analysis of complex statistical models (Robert and Casella 2004). REML and Bayesian methods are widely used in animal breeding programs. Bayesian analysis via Gibbs sampling has some advantages over REML methods. Gibbs sampling can provided the whole posterior distribution for the variance components whereas REML provides the point estimates. But Bayesian methods are computationally demanding and still much focus is given to improve the total computational time. Recently (Bauer et al 2009; Waldmann et al. 2008) applied Bayesian Gibbs sampling for quantitative genetics research studies in plants and the latter developed a fast hybrid Gibbs sampler, which accounted for additive and dominance variances in the mixed model. Still accounting inbreeding while estimating breeding values is one of the major concern in self-pollinating crops. Inbreeding induces non-zero covariance between the additive and dominance effects and which complicates the calculation. Also much focus is given to improving the efficiency and convergence of MCMC samplers. Moreover parameter identifiability due to multi-modality is another major problem arises when the non-additive random genetic effects are included in the model. The efficiency of a MCMC algorithm depends critically on the transition kernel of the Markov chain (Hastings 1970; Roberts and Rosenthal 2001), but the choice of an efficient kernel, which produces a rapidly mixing chain, is often difficult.

1.13 Objectives

Accurate and fast estimation of genetic parameters underlying quantitative traits using mixed linear models with additive and dominance effects is of great importance in both natural and breeding populations. REML and Bayesian methods are commonly used for the estimation of the genetic parameters. However Bayesian methods using MCMC algorithms are usually needs computationally demanding sampling techniques so their use is limited. Moreover conventional MCMC algorithms may suffer from poor mixing and slow convergence rate. In addition poor parameter identifiability is another main problem faced by MCMC methods due to the existence of multiple modes in linear mixed models. So adaptive MCMC algorithms have been proposed which can use the previous history of the chain to "learn" the proposal distribution parameters, which are efficient for exploring the posterior distribution of the model using the data at hand. The adaptive MCMC algorithm provides better convergence rate and mixing properties compared to the conventional MCMC algorithms. Also the learned the proposal distribution parameters will help the algorithm to find different modes in the posterior distribution. The main objectives of the study are

1. To know the impact of adaptation on estimation accuracy of the genetic parameters.

2. To determine the effect of adaptation process on total computational time.

3. To identify how different prior distributions affect the mixing of the MCMC chains.

4. Find the impact of adaptation on the mixing and convergence rate of the MCMC chains.

5. Address the parameter identifiability problem.

2 Models and Methods

Genetic covariances between individuals are an important factor for the prediction of breeding values. These genetic covariances can be calculated from the pedigree informations. The genetic covariance is composed of three components: the additive genetic variance, the dominance variance and the epistatic variance. In the current study I considered additive and dominance relationship matrices for the calculation of breeding values. The additive and dominance relationship matrices were used in the linear mixed model to estimate the breeding values and the variance components. Algorithms to calculate these matrices are explained below.

2.1 Additive Relationship Matrix

The additive relationship matrix, which describes the genetic relationship between individuals, can be calculated from the pedigree informations. Henderson (1976) developed a fast recursive method for the calculation of additive relationship matrix $\mathbf{A}$, from the pedigree information. The matrix is symmetric and its diagonal elements ($\mathbf{a_{ii}}$) is equal to $1+\mathbf{F_i}$ where $\mathbf{F_i}$ is the inbreeding coefficient of the $\mathbf{i^{th}}$ line. Let the pedigree be coded from 1 to $\mathbf{n}$ and ordered in a way that parents precede their progenies. Then the following algorithm is used to compute $\mathbf{A}$. Here a_{ij} is the element of the matrix $\mathbf{A}$ in the i^{th} row and j^{th} column. If both parents sir ($\mathbf{s}$) and dam ($\mathbf{d}$) of a line $\mathbf{i}$ are known

$\mathbf{a_{ij} = a_{ji} = 0.5(a_{js} + a_{jd})} \qquad where\ \mathbf{j = 1}\ to\ \mathbf{(i - 1)}$

$\mathbf{a_{ii} = 1 + 0.5(a_{sd})}$

If only one parent(s) is known and unrelated

$\mathbf{a_{ij} = a_{ji} = 0.5(a_{js})} \qquad where\ \mathbf{j = 1}\ to\ \mathbf{(i - 1)}$

$\mathbf{a_{ii} = 1}$

If both parents are unknown and unrelated

$\mathbf{a_{ij} = a_{ji} = 0}$ *where* $\mathbf{j = 1}$ *to* $\mathbf{(i - 1)}$

$\mathbf{a_{ii} = 1}$

2.2 Dominance Relationship Matrix

The dominance genetic effect results from the interaction of alleles at a locus. If two animals have the same set of parents or grandparents, then it is possible that they posses the pair of alleles in common. The dominance genetic relationship between an individual **x** with parents **s** and **d** and an individual **y** with parents **f** and **m** can be calculated as follows:

$$\mathbf{d_{xy} = 0.25(u_{sf}u_{dm} + u_{sm}u_{df})}$$

where $\mathbf{u_{ij}}$ is the additive genetic relationship between **i** and **j**. Thus the dominance relationship matrix (**D**), which describes the dominance relationship among individuals can be calculated from the additive relationship matrix.

2.3 The Mixed Linear Model

Association models and Mixed models are the two proposed methods for the estimation of genomic breeding values. In the current research genetic parameters were estimated using mixed models. Linear mixed models provide a powerful mean of estimating genetic parameters. The linear mixed model assumes that the relationship between the mean of the dependent variable and the fixed and random effects can be modeled as a linear function. The mixed linear model can include both fixed and random effects. Henderson

(1985a,b) has shown that linear mixed models can be used for the estimation of additive and dominance genetic variances. Consider the mixed linear model (Henderson 1985):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z_1}\mathbf{a} + \mathbf{Z_2}\mathbf{d} + \mathbf{e}, \tag{8}$$

where $\mathbf{y}$ is an $n \times 1$ vector of phenotypic observations, $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed (environmental) effects, $\mathbf{a}$ is a $q \times 1$ vector of random additive genetic effects, $\mathbf{d}$ is a $q \times 1$ vector of random dominance genetic effects, $\mathbf{e}$ is a $n \times 1$ vector of error terms, which are independently normally distributed with mean zero and variance σ_e^2. Moreover, $\mathbf{X}, \mathbf{Z_1}$ and $\mathbf{Z_2}$ are known incidence matrices, where $\mathbf{X}$ associates $\boldsymbol{\beta}$ to the phenotypic observations $\mathbf{y}$. For the simulated datasets $\mathbf{Z_1}$ and $\mathbf{Z_2}$ associates genetic effects $\mathbf{a}$ and $\mathbf{d}$ respectively to the observation vector $\mathbf{y}$. Whereas for the field data $\mathbf{Z_1}$ and $\mathbf{Z_2}$ associates random genetic effects $\mathbf{a}$ and genotype-by-environment interaction (**GxE**) to $\mathbf{y}$. The additive genetic relationship matrix $\mathbf{A}$ (assumed to be nonsingular), which describes additive genetic relationships among lines, was calculated using the available pedigree information. And dominance relationship matrix $\mathbf{D}$ (also assumed to be nonsingular) is the dominance matrix , which describes dominance variances and covariance among lines. Here the total phenotypic variation coming from the observation vector $\mathbf{y}$ can be explained by the summation of variation due to the additive random effects ($\mathbf{a}$), random dominance effects ($\mathbf{d}$) and the error variance ($\mathbf{e}$). In a Bayesian framework, all the unknown parameters are sampled from probability distributions using sampling algorithms. In the current study Gibbs sampler was used to sample the random parameters like additive and dominance effects from their corresponding distributions. In the new approach Gibbs sampler was used in the first step called the learning phase and in the second step, called the adapted phase a metropolis-Hastings (MH) algorithm was used to estimated the variance components. These two

algorithms combined to from the adaptive MCMC method. The hybrid Gibbs sampler and the normal M-H algorithm are explained below.

2.4 Gibbs sampling

Gibbs Sampler (Casella and George, 1992) is a Markov chain Monte Carlo (MCMC) method for generating draws from joint posterior by using draws of the conditional posteriors, and is a special case of Metropolis-Hastings sampling (Chib and Greenberg, 1995). The Gibbs sampling algorithm is one of the commonly used Markov chain Monte Carlo algorithms. Gibbs sampler is useful when the direct simulation from the posterior distribution is not possible. Gibbs sampling is also known as alternating conditional sampling. In the current study a hybrid Gibbs sampler was used to sample the random effects. The hybrid Gibbs sampler is a combination of both single-site Gibbs sampling algorithm (eg, Sorensen and Gianola 2002) and blocked Gibbs sampling algorithm (Garcia-Cortes and Sorensen 1996). The blocked Gibbs sampling has a faster convergence rate and better mixing when the parameters in the data are correlated (Waldmann *et al.* 2008). But in blocked Gibbs sampling the inverse of the coefficient matrix $\mathbf{C}$ is needed, which is computationally challenging. The hybrid Gibbs sampler which uses block updates every 50th iteration is much faster than plain blocked Gibbs sampling and it holds better mixing properties than the single-site Gibbs sampler. In Bayesian analysis, it is needed to assign prior distributions for the hyperparameters. In the current study Gamma prior distribution was assigned for the hyperparameters with parameters k_i and λ_i and mean k_i/λ_i. It was decided to use $k_i = 1$ and $\lambda_i = 0.001$ (*i.e.*, the exponential distribution with mean $1/\lambda_i$) in order to obtain flat priors.

The algorithm of hybrid Gibbs sampling as follows:

1. Initialize ψ_a, ψ_d and ψ_e with some reasonable positive values. Set $k_a^* = k_a + q/2$, $k_d^* = k_d + q/2$, and $k_e^* = k_e + n/2$. Here $\mathbf{q}$ and $\mathbf{n}$ are the number of lines and the number of records respectively.

2. Single-site Gibbs sampling:

   (a) Sample θ_i from $N(\hat{\theta}, 1/(C_{i,i}\psi_e))$, where $\hat{\theta}$=$(\mathbf{W}'\mathbf{y} - \mathbf{C}_{i,-i}\boldsymbol{\theta}_{-i})/C_{i,i}$. Here $\boldsymbol{\theta}_{-i}$ is $\boldsymbol{\theta}$ without its i^{th} component, $\mathbf{C}_{i,-i}$ is the i^{th} row of $\mathbf{C}$ without its i^{th} component, and finally $C_{i,i}$ is i^{th} diagonal element of $\mathbf{C}$.

   (b) Calculate $\lambda_a^* = \lambda_a + (\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})/2$, $\lambda_d^* = \lambda_d + (\mathbf{d}^T\mathbf{D}^{-1}\mathbf{d})/2$, and $\lambda_e^* = \lambda_e + 1/2||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z_1}\mathbf{a} - \mathbf{Z_2}\mathbf{d}||^2$.

   (c) Sample the precision parameters ψ_i from Gamma(k_i^*, λ_i^*) for $i = a, d, e$.

   (d) Calculate $\alpha_a = \psi_a/\psi_e$, $\alpha_d = \psi_d/\psi_e$ and update the coefficient matrix $\mathbf{C}$.

3. Block Gibbs sampling (every 50^{th} iteration):

   (a) Generate $\mathbf{a}^*$ from MVN$(\mathbf{0}, \mathbf{A}/\psi_a)$ and $\mathbf{d}^*$ from MVN$(\mathbf{0}, \mathbf{D}/\psi_a)$.

   (b) Generate $\mathbf{z}^*$ from MVN$(\mathbf{Z_1}\mathbf{a}^* + \mathbf{Z_2}\mathbf{d}^*, \mathbf{I}/\psi_e)$.

   (c) Calculate $\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$.

   (d) Calculate $\boldsymbol{\theta}$ as $[\mathbf{0}', \mathbf{a}^{*\prime}, \mathbf{d}^{*\prime}]' + \mathbf{C}^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$, where $\mathbf{0}$ is zero vector of the size of the fixed effects vector $\boldsymbol{\beta}$.

   (e) Calculate $\lambda_a^* = \lambda_a + (\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})/2$, $\lambda_d^* = \lambda_d + (\mathbf{d}^T\mathbf{D}^{-1}\mathbf{d})/2$, and $\lambda_e^* = \lambda_e + 1/2||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z_1}\mathbf{a} - \mathbf{Z_2}\mathbf{d}||^2$.

   (f) Sample the precision parameters ψ_i from Gamma(k_i^*, λ_i^*), for $i = a, d, e$.

(g) Calculate $\alpha_a = \psi_a/\psi_e$, $\alpha_d = \psi_d/\psi_e$ and update the coefficient matrix $\mathbf{C}$.

(h) go to 2a, repeat the steps until the MCMC chain is converged.

2.5 The Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm is commonly used for MCMC simulation. M-H algorithm requires a proposal generating distribution and the performance is greatly depend on the covariance structure of the kernel. An adaptive M-H algorithm can find the optimal covariance structure for the proposal distribution from the previous history of the MCMC chain. The basic idea depends on the fact that, instead of computing the values for the target distribution $p(x)$, only needed to compute the ratio of the target at two distinct parameter values $p(x)/p(x^*)$, the integral in the Bayes formula cancels out. Unlike Gibbs sampling M-H weighs all draws equally but not all the draws are accepted (this is like accept-reject method).

Let $\mathbf{x^t}$ be the current state then the M-H algorithm generates a Markov chain in which each state $\mathbf{x^{t+1}}$ is depends only on the previous state $\mathbf{x^t}$. The algorithm uses a proposal density $\mathbf{q}(\mathbf{x'};\mathbf{x^t})$, which depends on the current state $\mathbf{x^t}$, to generate a new proposed sample $\mathbf{x'}$. This proposal is accepted as the next value ($\mathbf{x^{t+1}} = \mathbf{x'}$), if α drawn from $\mathbf{U(0,1)}$ satisfies

$$\alpha < \{\frac{\mathbf{p(x')q(x^t;x')}}{\mathbf{p(x^t)q(x';x^t)}}, \mathbf{1}\} \tag{9}$$

If the proposal is accepted then $\mathbf{x^{t+1}} = \mathbf{x'}$, otherwise the current value is retained. Choosing a good proposal distribution is very important, otherwise most of the proposed values will be rejected. In the current study a Gaussian distribution centered on the current state $\mathbf{x^t}$ was used as the proposal

distribution.

2.6 Adaptive MCMC

Recent theoretical developments (Haario *et al.* 2001; Roberts and Rosenthal 2007) have renewed the interest of adaptive MCMC methods in research studies. The adaptive MCMC methods can be used to determine suitable and efficient "proposal distribution" for M-H sampler by looking the data. These methods usually differ in how the learning phase of the MCMC sample is utilized in the final posterior estimates. Here simply omit (through away) part of the MCMC sample used to learn the proposal distribution (i.e. learning phase).

Convergence of the general Bayesian Gibbs sampling algorithms, which use single-site updates for the variance components, can be slow due to posterior dependencies. More efficient sampler is obtained by updating all variance components jointly and removing dependencies within the sample, thus the random walk M-H algorithm was considered. In the current study a fast adaptive MCMC algorithm was developed, combining both hybrid Gibbs sampling and M-H algorithm, for the estimation of variance components. In the new approach the adaptive MCMC runs in two stages. First, run the algorithm to obtain empiric estimate for the posterior covariance structure of log transformed variance components (this part of the MCMC is called learning period). In the second phase of the algorithm, use this covariance structure to formulate an effective proposal distribution for a Metropolis–Hastings algorithm, which uses a likelihood function where the random effects have been integrated out. In the learning phase of the algorithm the hybrid Gibbs sampler was used to sample random (additive genetic and dominance) effects.

2.6.1 Marginalization

The likelihood function is function of all parameters of a statistical model, which is used to fit the observational data. If someone is interested in a particular parameter, it is possible to average over the effect of nuisance parameters from the model, this process is known as marginalization. This process will help to remove the correlations between parameters. The dependencies among breeding values and dominance effects slow down the convergence of the MCMC chain. So the effect of breeding values and dominance effects were marginalized away before computing the posterior probability in the adapted phase. Here the adaptive MCMC was divided into two classes: first class where the effect of breeding values and dominance effects were marginalized away before computing the posterior probability in the adapted phase, and second class those effects were included to calculate the posterior probability in the adapted phase.

In the current study, two different hierarchical models was used, former to be used in the learning phase and the latter in the adapted phase of the estimation algorithm. If all the priors are chosen to be the same, then these two hierarchical models are identical except that most parameters have been integrated out analytically from the latter.

2.6.2 Hierarchical model 1

Let the precision parameters ψ_a, ψ_d and ψ_e be the inverses of the variances σ_a^2, σ_d^2 and σ_e^2 respectively. Here σ_a^2, σ_d^2 and σ_e^2 are the additive, dominance and error variances respectively. Then by model (8), the phenotypic observation for a given trait is modeled as a linear combination of explanatory variables.

For given $\boldsymbol{\beta}$, $\mathbf{a}$, $\mathbf{d}$, and ψ_e, vector $\mathbf{y}$ follows a multivariate normal distribution

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{a}, \mathbf{d}, \psi_e \sim \text{MVN}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z_1}\mathbf{a} + \mathbf{Z_2}\mathbf{d}, \mathbf{I}/\psi_e), \tag{10}$$

where $1/\psi_e$ is the residual variance of the model. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{a}, \mathbf{d})$ be the unknown location parameters and $\boldsymbol{\psi} = (\psi_a, \psi_d, \psi_e)$ be the precision parameters. By Bayes theorem, the joint posterior density of unknown parameters is proportional to

$$p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \propto p(\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi}), \tag{11}$$

where $p(\boldsymbol{\psi}) = p(\psi_a)p(\psi_d)p(\psi_e)$ and $p(\boldsymbol{\theta}|\boldsymbol{\psi}) = p(\boldsymbol{\beta})p(\mathbf{a}|\psi_a)p(\mathbf{d}|\psi_d)$ are the prior distributions and $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})$ is the likelihood from (10). For the Bayesian analysis, one must assign a prior distribution for the unknown model parameters. So $\boldsymbol{\beta}$ was assigned an improper uniform prior distribution.

$$p(\boldsymbol{\beta}) \propto \text{constant}. \tag{12}$$

Conditionally on the precision parameters, the genetic effects were assigned multivariate normal prior distributions with zero mean vector $\mathbf{0}$ (of size q),

$$\mathbf{a}|\psi_a \sim \text{MVN}(\mathbf{0}, \mathbf{A}/\psi_a), \qquad \mathbf{d}|\psi_d \sim \text{MVN}(\mathbf{0}, \mathbf{D}/\psi_d) \tag{13}$$

Before assigning a prior distribution for the precision parameters, the phenotypic observation vector $\mathbf{y}$ was standardized in order to use the same prior for different data sets (which may originally have very different phenotypic scales). After the standardization, the precision parameters ψ_a, ψ_d and ψ_e were assumed to follow a Gamma prior distribution with parameters k_i and

λ_i and mean k_i/λ_i,

$$\psi_i \sim \text{Gamma}(k_i, \lambda_i), \qquad i = a, d, e \tag{14}$$

In the current study $k_i = 1$ and $\lambda_i = 0.001$ (*i.e.*, the exponential distribution with mean $1/\lambda_i$) was used, in order to obtain flat priors. This choice allows the variance components to be shrunken very nearly to zero, if this is warranted by the data. This follows since the prior (14) implies an inverse gamma prior with parameters (k_i, λ_i) for the variance component σ_i^2. The inverse gamma density raises from value zero to its maximum at the mode $\lambda_i/(k_i + 1)$ and then decays slowly. Shrinkage-type priors have been used before, *e.g.*, in variable selection (O'Hara and Sillanpää 2009) and in haplotype estimation (Gasbarra *et al.* 2011) as well as in penalized likelihood estimation of genetic covariance matrices (Meyer and Kirkpatrick 2010).

2.6.3 Hierarchical model 2:

In the adapted phase of the algorithm a model was used, where all the unknown location parameters $\boldsymbol{\theta}$ are integrated out from model (Eq. 8). The joint posterior density of parameters $\boldsymbol{\psi}$ is

$$p(\boldsymbol{\psi}|\mathbf{y}) \propto p(\boldsymbol{\psi})p(\mathbf{y}|\boldsymbol{\psi}). \tag{15}$$

To mimic the improper uniform prior (12), the fixed effects $\boldsymbol{\beta}$ were assigned a normal prior distribution with zero mean vector $\mathbf{0}$ and large covariance matrix $\mathbf{B}\sigma^2_{\boldsymbol{\beta}}$, where $\sigma^2_{\boldsymbol{\beta}}$=$10^6$,

$$\boldsymbol{\beta} \sim \text{MVN}(\mathbf{0}, \mathbf{B}\sigma^2_{\boldsymbol{\beta}}).$$

Here $\mathbf{B}$ is the unscaled prior covariance matrix between fixed effects. The genetic effects $\mathbf{a}$ and $\mathbf{d}$ were assigned the multivariate normal priors (13), and variance components the Gamma priors (14). After these choices it is a simple matter to integrate out the location parameters from the model (cf. pp. 313–314 in Sorensen and Gianola, 2002), namely

$$\mathbf{y}|\boldsymbol{\psi} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{16}$$

where $\boldsymbol{\Sigma} = \mathbf{XBX}'\sigma^2_{\boldsymbol{\beta}} + \mathbf{Z_1AZ}'_{\mathbf{1}}/\psi_a + \mathbf{Z_2DZ}'_{\mathbf{2}}/\psi_d + \mathbf{I}/\psi_e$.

2.6.4 Estimation in the learning phase:

To implement the Gibbs sampler, one needs the fully conditional posterior distributions of all unknown parameters ($\boldsymbol{\theta}$ and $\boldsymbol{\psi}$) of hierarchical model 1. These can be found, e.g., from Waldmann *et al.* (2008). To update $\boldsymbol{\theta}$, samples can be drawn either element-wise or block-wise from the fully conditional posterior distribution

$$\boldsymbol{\theta}|\boldsymbol{\psi}, \mathbf{y} \sim \text{MVN}(\hat{\boldsymbol{\theta}}, \mathbf{C}^{-1}/\psi_e), \tag{17}$$

where $\hat{\boldsymbol{\theta}}$ is the solution to the linear system $\mathbf{C}\boldsymbol{\theta} = \mathbf{W}'\mathbf{y}$. Here

$$\mathbf{C} = \mathbf{W}'\mathbf{W} + \mathbf{V}, \quad \mathbf{W} = [\mathbf{X}, \mathbf{Z_1}, \mathbf{Z_2}], \quad \mathbf{V} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1}\alpha_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}^{-1}\alpha_d \end{bmatrix} \tag{18}$$

with $\alpha_a = \psi_a/\psi_e$, $\alpha_d = \psi_d/\psi_e$. The precision parameters are sampled from their fully conditional posterior distributions,

$$\psi_i|\boldsymbol{\theta}, \mathbf{y} \sim \text{Gamma}(k_i^*, \lambda_i^*), \qquad i = a, d, e$$

where $k_a^* = k_a + q/2$, $\lambda_a^* = \lambda_a + (\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})/2$, $k_d^* = k_d + q/2$, $\lambda_d^* = \lambda_d + (\mathbf{d}^T\mathbf{D}^{-1}\mathbf{d})/2$, $k_e^* = k_e + n/2$, and $\lambda_e^* = \lambda_e + 1/2||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z_1}\mathbf{a} - \mathbf{Z_2}\mathbf{d}||^2$. During the learning phase of the algorithm the hybrid Gibbs sampler with block update every 50^{th} iteration was used to sample the random additive and dominance effects. Section 1.9 describes the details of the sampling algorithm.

2.6.5 Estimation in the adapted phase (class 1)

The history of the chain during the learning phase was used, in order to form the proposal distribution for the parameters of hierarchical model 2. In the second, adapted phase of the algorithm, a M-H algorithm was used to update log-variance components block-wise using putative samples generated from the learned proposal distribution. M-H algorithm uses random-walk proposals: the proposed parameter vector is generated by adding to the current parameter vector an increment from a multivariate normal distribution with zero mean and covariance matrix $\mathbf{S}_p$. The selection of the proposal covariance matrix was based on the theoretical results of Roberts *et al.* (1997) and Roberts and Rosenthal (2001). These authors show that if the posterior distribution is approximately multivariate normal with covariance matrix $\mathbf{S}$, then the optimal choice for the proposal covariance matrix $\mathbf{S}_p$ is approximately $(2.38)^2/d\,\mathbf{S}$, where d is the number of unknown parameters in the posterior distribution. In order to better be able to use this result, the new algorithm works on the logarithmic scale, i.e., the vector $\boldsymbol{\tau} = (\tau_a, \tau_d, \tau_e)$ was used as the new parameter vector, where the τ's are the logarithms of the variance components, $\tau_i = \log(\sigma_i^2) = -\log(\psi_i)$, $i = a, d, e$. This reparametrization eliminates the positivity constraints which are present for the variance components or their inverses. At the same time, it makes the posterior distribution resemble more closely a multivariate normal distribution. Since the posterior covariance

matrix $\mathbf{S}$ of vector $\boldsymbol{\tau}$ is unknown, it was estimated with the sample covariance matrix $\hat{\mathbf{S}}$, which is calculated from the log-transformed variance components simulated during the learning phase.

After the proposed parameter vector $\boldsymbol{\tau}^*$ has been generated by adding a noise vector to the current parameter vector $\boldsymbol{\tau}$, the proposed $\boldsymbol{\tau}^*$ is either accepted or rejected as the new state of the Markov chain based on the value of the Metropolis–Hastings acceptance ratio r, which is now given by

$$r = \frac{p(\boldsymbol{\tau}^*)}{p(\boldsymbol{\tau})} \frac{p(\mathbf{y}|\boldsymbol{\tau}^*)}{p(\mathbf{y}|\boldsymbol{\tau})} \tag{19}$$

Here the likelihood ratio can be evaluated based on Eq. (16), after the log-transformed variance components $\boldsymbol{\tau} = (\tau_a, \tau_d, \tau_e)$ and $\boldsymbol{\tau}^* = (\tau_a^*, \tau_d^*, \tau_e^*)$ have been transformed to precision parameters, using the formulas

$$\psi_i = \mathrm{e}^{-\tau_i}, \quad \psi_i^* = \mathrm{e}^{-\tau_i^*}, \qquad i = a, d, e.$$

For $\boldsymbol{\tau}$, the likelihood is

$$p(\mathbf{y}|\boldsymbol{\tau}) = (2\pi)^{-n/2} \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})}} \exp\{-\frac{1}{2}\mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y}\}, \tag{20}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{y}$ conditionally on the current values of the parameters,

$$\boldsymbol{\Sigma} = \mathbf{X}\mathbf{B}\mathbf{X}'\sigma_{\boldsymbol{\beta}}^2 + \mathbf{Z_1}\mathbf{A}\mathbf{Z_1'}/\psi_a + \mathbf{Z_2}\mathbf{D}\mathbf{Z_2'}/\psi_d + \mathbf{I}/\psi_e.$$

For $\boldsymbol{\tau}^*$, the likelihood $p(\mathbf{y}|\boldsymbol{\tau}^*)$ is obtained from a similar formula where $\boldsymbol{\Sigma}$ is replaced by the covariance matrix of $\mathbf{y}$ conditionally on the proposed values

of the parameters,

$$\boldsymbol{\Sigma}^* = \mathbf{X}\mathbf{B}\mathbf{X}'\sigma^2_{\boldsymbol{\beta}} + \mathbf{Z_1}\mathbf{A}\mathbf{Z_1}'/\psi^*_a + \mathbf{Z_2}\mathbf{D}\mathbf{Z_2}'/\psi^*_d + \mathbf{I}/\psi^*_e,$$

In order to evaluate the prior ratio $p(\boldsymbol{\tau}^*)/p(\boldsymbol{\tau})$ in Eq. (19), it is necessary to take into account the prior formulated for the vector of precision parameters $\boldsymbol{\psi}$. Using the change-of-variables formula for probability densities, the prior ratio can be calculated as

$$\frac{p(\boldsymbol{\tau}^*)}{p(\boldsymbol{\tau})} = \frac{p(\boldsymbol{\psi}^*)}{p(\boldsymbol{\psi})}\frac{|J^*|}{|J|}. \tag{21}$$

Here $p(\boldsymbol{\psi}) = p(\psi_a|k_a, \lambda_a)p(\psi_d|k_d, \lambda_d)p(\psi_e|k_e, \lambda_e)$ is the product of the three gamma densities (14), and similarly $p(\boldsymbol{\psi}^*)$ is the product of the same gamma densities evaluated at the proposed precision parameters. Further, $J = -\exp(-\tau_a - \tau_d - \tau_e)$ is the Jacobian (determinant) arising from expressing $\boldsymbol{\psi}$ in terms of $\boldsymbol{\tau}$, and $J^* = -\exp(-\tau^*_a - \tau^*_d - \tau^*_e)$ is the Jacobian from expressing $\boldsymbol{\psi}^*$ in terms of $\boldsymbol{\tau}^*$. In the actual M–H algorithm, first calculated the logarithm of the M–H ratio r, and then calculated the logarithm of the ratio of the absolute Jacobians,

$$\log\frac{|J^*|}{|J|} = -(\tau^*_a - \tau_a + \tau^*_d - \tau_d + \tau^*_e - \tau_e). \tag{22}$$

The sampling algorithm during the adapted phase is as follows. First estimated the posterior covariance matrix $\mathbf{S}$ of the log-transformed variance components from the output of the learning phase, and calculate the proposal covariance matrix as $\mathbf{S}_p = (2.38)^2\hat{\mathbf{S}}/d$. Then iterated the following steps.

1. Let $\boldsymbol{\tau}$ be the current values in logarithmic scale. Generate new values $\boldsymbol{\tau}^* = \boldsymbol{\tau} + \mathbf{w}$, where $\mathbf{w}$ is simulated from $\text{MVN}(\mathbf{0}, \mathbf{S}_p)$. Transform $\boldsymbol{\tau}$ and

$\boldsymbol{\tau}^*$ to precision parameter vectors $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^*$.

2. Calculate the logarithm of the M–H acceptance ratio $\log(r)$ using Equations (19)–(22).

3. Accept the proposed value $\boldsymbol{\tau}^*$, if a random number drawn from the uniform distribution over $[0, 1]$ is less than r. If the proposal is accepted then the proposed parameter vector is taken as the current vector $\boldsymbol{\tau}=\boldsymbol{\tau}^*$, otherwise the current value is retained.

Since the breeding values and the dominance effects have been integrated out from the likelihood, this sampling algorithm reduces the problems of the Gibbs sampler which arise due to posterior dependences between the random effects and the variance components.

The whole adaptive algorithm consisting of the learning phase and the adapted phase is described more fully in section 2.9. It has been implemented in the Matlab (2007) environment where most of the analyses have been performed.

When the target distribution is multimodal, a random walk may rarely move between modes and this will lead to poor parameter identifiability. Adaptive MCMC methods are useful for such multimodal problems, where the adaptive MCMC methods adapt the transition kernel of the chain, using information obtained from previous iterations. Such adaptation enables the movement of the chain between different modes.

2.7 Adaptive MCMC (Class 2)

Normally posterior dependences between the random effects will affect the convergence rate and mixing properties of the MCMC chain. In order to check the effect of posterior dependences on the convergence rate and mixing

properties of the chain, a model was tested without integrating out those random effects in the adapted phase(class 2) of the algorithm.

The algorithm(adaptive MCMC, class 2) for the proposed sampling is as follows.

1. Calculate the proposal covariance matrix $\mathbf{S}$ from the learned MCMC samples in logarithmic scale.

2. Let $\boldsymbol{\tau}$ be the current values in logarithmic scale. Generate new values $\boldsymbol{\tau}^* = \boldsymbol{\tau} + \mathbf{w}$, where $\mathbf{w}$ is simulated from MVN$(\mathbf{0}, \mathbf{S}_p)$ where $\mathbf{S}_p = (2.38)^2\hat{\mathbf{S}}/d$.

3. Calculate the MH acceptance ratio r as the product of $Gamma$ densities for the proposed and current values, r $= \prod_{i=a,d,e} \frac{Gamma(\psi_i^*|k_i,\lambda_i)}{Gamma(\psi_i|k_i,\lambda_i)}$, where $\psi_i = e^{-2\tau_i}$, $\psi_i^* = e^{-2\tau_i^*}$, i=a,d,e.

4. Accept the proposed value ψ^* with probability $min\ (1, r^*)$, where $r = r^* + J$ and J is Jacobian term. If the proposal is accepted then τ=τ^*, otherwise the current value is retained.

2.8 Calculation of the likelihood ratio

Calculating the determinants of high-dimensional matrices is challenging, since numerical problems arise as the dimension increases. In order to calculate the likelihood ratio $p(\mathbf{y}|\boldsymbol{\psi}^*)/p(\mathbf{y}|\boldsymbol{\psi})$, it was needed to compute the determinants of the covariance matrices of $\mathbf{y}$ conditionally on the proposed and current point. These matrices were scaled to mitigate numerical problems. Scaling was based on the identity $\det(s\boldsymbol{\Sigma}) = s^n \det(\boldsymbol{\Sigma})$ (valid whenever s is a scalar and $\boldsymbol{\Sigma}$ is an $n \times n$ matrix) and the identity $(s\boldsymbol{\Sigma})^{-1} = \boldsymbol{\Sigma}^{-1}/s$ (valid whenever s is a scalar and $\boldsymbol{\Sigma}$ is an invertible matrix). Then set $s = 1/\psi_e$, $s^* = 1/\psi_e^*$ as the

scaling factors for the current ($\boldsymbol{\Sigma}$) and proposed ($\boldsymbol{\Sigma}^*$) values, respectively. Let $\boldsymbol{\psi}^* = (\psi_a^*, \psi_d^*, \psi_e^*)$ be the proposed values of (inverses of) variance components and $\boldsymbol{\psi} = (\psi_a, \psi_d, \psi_e)$ be their current values. The logarithm of the likelihood ratio was calculated as

$$\begin{aligned}\log \frac{p(\mathbf{y}|\boldsymbol{\psi}^*)}{p(\mathbf{y}|\boldsymbol{\psi})} = &-\frac{n}{2}(\log(s^*) - \log(s)) - \frac{1}{2}(\log(\det(\boldsymbol{\Sigma}^*/s^*)) - \log(\det(\boldsymbol{\Sigma}/s))) \\ &- \frac{1}{2s^*}(\mathbf{y}'(\boldsymbol{\Sigma}^*/s^*)^{-1}\mathbf{y}) + \frac{1}{2s}(\mathbf{y}'(\boldsymbol{\Sigma}/s)^{-1}\mathbf{y}).\end{aligned} \tag{23}$$

Here the determinants and quadratic forms were calculated using Cholesky decomposition. If $\mathbf{M} = \boldsymbol{\Sigma}/s$ is a $n \times n$ positively definite symmetric matrix, then its Cholesky decomposition is $\mathbf{M} = \mathbf{L}\mathbf{L}'$, where $\mathbf{L}$ is the lower triangular Cholesky factor. The determinant is calculated as $\log(\det(\mathbf{M})) = 2\sum_i^n \log(L_{i,i})$, where $L_{i,i}$ is the i^{th} diagonal element of $\mathbf{L}$. The quadratic form $\mathbf{y}'(\boldsymbol{\Sigma}/s)^{-1}\mathbf{y} = \mathbf{y}'\mathbf{M}^{-1}\mathbf{y}$ is calculated using the identity $\mathbf{y}'\mathbf{M}^{-1}\mathbf{y} = (\mathbf{L}^{-1}\mathbf{y})'(\mathbf{L}^{-1}\mathbf{y})$, where $\mathbf{L}^{-1}\mathbf{y}$ is calculated by solving $\mathbf{z}$ from the equation $\mathbf{L}\mathbf{z} = \mathbf{y}$.

2.9 Adaptive MCMC algorithm

The complete adaptive MCMC algorithm is as follows:

1. Initialize ψ_a, ψ_d and ψ_e with some reasonable positive values. Set $k_a^* = k_a + q/2$, $k_d^* = k_d + q/2$, and $k_e^* = k_e + n/2$.

2. Single-site Gibbs sampling:

   (a) Sample θ_i from $N(\hat{\theta}, 1/(C_{i,i}\psi_e))$, where $\hat{\theta}$=$(\mathbf{W}'\mathbf{y} - \mathbf{C}_{i,-i}\boldsymbol{\theta}_{-i})/C_{i,i}$. Here $\boldsymbol{\theta}_{-i}$ is $\boldsymbol{\theta}$ without its i^{th} component, $\mathbf{C}_{i,-i}$ is the i^{th} row of $\mathbf{C}$

without its i^{th} component, and finally $C_{i,i}$ is i^{th} diagonal element of $\mathbf{C}$.

(b) Calculate $\lambda_a^* = \lambda_a + (\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})/2$, $\lambda_d^* = \lambda_d + (\mathbf{d}^T\mathbf{D}^{-1}\mathbf{d})/2$, and $\lambda_e^* = \lambda_e + 1/2||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z_1}\mathbf{a} - \mathbf{Z_2}\mathbf{d}||^2$.

(c) Sample the precision parameters ψ_i from Gamma(k_i^*, λ_i^*) for $i = a, d, e$.

(d) Calculate $\alpha_a = \psi_a/\psi_e$, $\alpha_d = \psi_d/\psi_e$ and update the coefficient matrix $\mathbf{C}$.

3. Block Gibbs sampling (every 50^{th} iteration):

(a) Generate $\mathbf{a}^*$ from MVN$(\mathbf{0}, \mathbf{A}/\psi_a)$ and $\mathbf{d}^*$ from MVN$(\mathbf{0}, \mathbf{D}/\psi_a)$.

(b) Generate $\mathbf{z}^*$ from MVN$(\mathbf{Z_1}\mathbf{a}^* + \mathbf{Z_2}\mathbf{d}^*, \mathbf{I}/\psi_e)$.

(c) Calculate $\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$.

(d) Calculate $\boldsymbol{\theta}$ as $[\mathbf{0}', \mathbf{a}^{*\prime}, \mathbf{d}^{*\prime}]' + \mathbf{C}^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$, where $\mathbf{0}$ is zero vector of the size of the fixed effects vector $\boldsymbol{\beta}$.

(e) Calculate $\lambda_a^* = \lambda_a + (\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})/2$, $\lambda_d^* = \lambda_d + (\mathbf{d}^T\mathbf{D}^{-1}\mathbf{d})/2$, and $\lambda_e^* = \lambda_e + 1/2||\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z_1}\mathbf{a} - \mathbf{Z_2}\mathbf{d}||^2$.

(f) Sample the precision parameters ψ_i from Gamma(k_i^*, λ_i^*), for $i = a, d, e$.

(g) Calculate $\alpha_a = \psi_a/\psi_e$, $\alpha_d = \psi_d/\psi_e$ and update the coefficient matrix $\mathbf{C}$.

4. Setting up the adapted MCMC (after the learning period):

(a) Transform the samples from the learning period into logarithmic scale with the formula $\tau_i = -\log(\psi_i)$, for $i = a, d, e$.

(b) Calculate the sample covariance matrix $\hat{\mathbf{S}}$ from the transformed variables τ_i. Calculate the proposal covariance matrix $\mathbf{S}_p = (2.38)^2/d\,\hat{\mathbf{S}}$ where $d = 3$. Initialize the current state $\boldsymbol{\tau}$ from the last state visited during the learning phase.

5. The iterations during the adapted phase:

a) Generate proposed values $\boldsymbol{\tau}^*$ from the Gaussian distribution MVN$(\boldsymbol{\tau}, \mathbf{S}_p)$. Calculate the $\boldsymbol{\psi}$ values and $\boldsymbol{\psi}^*$ values corresponding to the current and the proposed vectors.

b) Calculate logarithm of the M–H acceptance ratio r by calculating the logarithm of the prior ratio $p(\boldsymbol{\tau}^*)/p(\boldsymbol{\tau})$ where the Jacobian ratio was taken into account, and also the logarithm of the likelihood ratio.

c) Draw u from the uniform distribution over $[0, 1]$ and accept the proposed value $\boldsymbol{\tau}^*$, if $u < r$. If the proposal is accepted then assign $\boldsymbol{\tau}$=$\boldsymbol{\tau}^*$, otherwise the current value is retained.

In a random-walk M–H algorithm that used in the adapted phase, the acceptance rate (the ratio between the number of times the proposed value is accepted to the total number of iteration after the learning period) should be between 10% and 50%, but the optimal rate is around 23% (see Roberts and Rosenthal, 2001).

2.10 Chi-square prior:

The scaled inverse chi-square distribution was used as the prior distribution to see the impact of prior on MCMC properties like mixing and effective sample size (ESS). Scaled inverse chi-square distribution used as a prior distribution

for hyperparameters.

$$p(\sigma_i^2|v_i, S_i^2) \propto (\sigma_i^2)^{-(v_i/2+1)} exp(-\frac{v_i S_i^2}{2\sigma_i^2}), \quad i = a, d, e \tag{24}$$

Here v_i is the degree of belief and S_i^2 is the prior value for the hyperparameters(Sorensen and Gianola 2002). It was decided to use v_i as -2 and S_i^2 as 0 to obtain flat priors.

2.11 MCMC convergence diagnostics

For Markov Chain Monte Carlo (MCMC) methods in applications it is important how to determine when it is safe to stop sampling and use the samples to estimate characteristics of the distribution of interest. One of the main problem with MCMC is to check whether the simulation has converged. Convergence can be assessed by starting the simulation from several different initial conditions, and by monitoring when the different simulation chains become sufficiently mixed together. Various MCMC convergence diagnostics tools have been developed over the years. Trace plots of the sampled MCMC values versus iteration number is one of commonly used tool for diagnostics. Trace plots are useful to estimate the degree of mixing in a simulation.

2.12 Effective Sample Size (ESS)

Effective sample size (Waagepetersen *et al.*, 2008; Geyer, 1992) is the approximate number of independent samples which would deliver the same estimation accuracy as the dependent MCMC samples. ESS is based on the central limit theorem (CLT) for Markov chains. Let $x_0, x_1, \ldots$ be the Markov chain (MC) and consider a scalar valued function h defined on the state space.

If the MC satisfies a CLT for this function, then as the sample size increases

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=0}^{n} h(x_i) - E_\pi h(x)\right) \xrightarrow{d} N(0, \tau_h \operatorname{var}_\pi h(x)), \tag{25}$$

where π is the stationary density of the MC, $E_\pi h(x)$ is the expected value of $h(x)$ under π, $\operatorname{var}_\pi h(x)$ is the variance of $h(x)$ under π, and τ_h is the integrated autocorrelation time for estimating $E_\pi h(x)$ for the given MC, defined as

$$\tau_h = 1 + 2\sum_{i=0}^{\infty} \operatorname{corr}_\pi(h(x_i), h(x_{i+k})), \tag{26}$$

Here corr_π is the correlation between the values when the chain is started from the stationary distribution ($x_0 \sim \pi$). On the other hand, if $y_0, y_1, ..., y_n$ are i.i.d samples from the stationary distribution π, then by the central limit theorem for i.i.d. sequences

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=0}^{n} h(y_i) - E_\pi h(x)\right) \xrightarrow{d} N(0, \operatorname{var}_\pi h(x)). \tag{27}$$

Comparing Eq. (25) and Eq. (27) gives ESS $= n/\tau_h$ as the effective sample size, when the expectation $E_\pi h(x)$ was estimated using the arithmetic mean of a large number of values $h(x_1), \ldots, h(x_n)$ based on the history of the MC. There are different methods available for estimating τ_h and ESS, but in the current study the R package `coda` was used.

2.13 Algorithm to calculate breeding value

During the adapted phase of the algorithm, the sampler generates values only from the marginal posterior of the variance components. Even if the new method is primarily intended for the estimation of the genetic variances,

it is possible to generate MCMC samples for the additive and dominance genetic values afterwards, by sampling them block-wise from their fully conditional posterior distribution conditionally on each of the values of the variance components in the MCMC sample, generated by the adaptive MCMC sampler.

Algorithm to calculate the breeding values using blocked Gibbs sampler:

1. Let σ_a^2, σ_d^2 and σ_e^2 be the variance components generated in the adapted phase.

2. Calculate $\alpha_a = \sigma_e^2/\sigma_a^2$, $\alpha_d = \sigma_e^2/\sigma_d^2$ and update the coefficient matrix $\mathbf{C}$ using Equation [12].

3. Generate $\mathbf{a}^*$ from MVN$(\mathbf{0}, \mathbf{A}\sigma_a^2)$ and $\mathbf{d}^*$ from MVN$(\mathbf{0}, \mathbf{D}\sigma_d^2)$.

4. Generate $\mathbf{z}^*$ from MVN$(\mathbf{Z_1}\mathbf{a}^* + \mathbf{Z_2}\mathbf{d}^*, \mathbf{I}\sigma_e^2)$.

5. Calculate $\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$.

6. Calculate $\boldsymbol{\theta}$ as $[\mathbf{0}', \mathbf{a}^{*\prime}, \mathbf{d}^{*\prime}]' + \mathbf{C}^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{z}^*)$, where $\mathbf{0}$ is zero vector of the size of the fixed effects vector $\boldsymbol{\beta}$.

7. Calculate the genetic parameters ($\mathbf{ge}$) of $\mathbf{n}$ individuals corresponding to the current variance components as $\mathbf{ge} = \mathbf{a} + \mathbf{d}$.

Repeat steps 1 to 7 until genetic parameters are sampled using all the variance components from the adapted phase.

2.14 Restricted Maximum Likelihood (REML)

Restricted Maximum Likelihood (REML) is one the commonly used method for the estimation of the genetic parameters in animal breeding programs. To

compare the estimation accuracy of the new method with REML method, a REML analysis was performed using the software package ASreml (Gilmour *et al.* 2006). Both REML and Bayesian analysis were carried out for the same datasets. In general ASreml provided the point estimates for the genetic parameters. REML differs from ML in that the likelihood of the data is maximized only for the random effects, thus REML is a restricted solution. The REML procedure requires that the observation vector $\mathbf{y}$ has a multivariate normal distribution. The corresponding linear model for the REML analysis is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z_1}\mathbf{a} + \mathbf{Z_2}\mathbf{d} + \mathbf{e}, \tag{28}$$

where $\mathbf{y}$ is an $n \times 1$ vector of phenotypic observations, $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed (environmental) effects, $\mathbf{a}$ is a $q \times 1$ vector of random additive genetic effects, $\mathbf{d}$ is a $q \times 1$ vector of random dominance genetic effects, $\mathbf{e}$ is a $n \times 1$ vector of error terms, which are independently normally distributed with mean zero and variance σ_e^2. Moreover, $\mathbf{X}, \mathbf{Z_1}$ and $\mathbf{Z_2}$ are known incidence matrices, where $\mathbf{X}$ associates $\boldsymbol{\beta}$ to the phenotypic observations $\mathbf{y}$. For the simulated datasets $\mathbf{Z_1}$ and $\mathbf{Z_2}$ associates genetic effects $\mathbf{a}$ and $\mathbf{d}$ respectively to the observation vector $\mathbf{y}$. Whereas for the field data $\mathbf{Z_1}$ and $\mathbf{Z_2}$ associates random genetic effects $\mathbf{a}$ and genotype-by-environment interaction ($\mathbf{GxE}$) to $\mathbf{y}$. The additive genetic relationship matrix $\mathbf{A}$ (assumed to be nonsingular), which describes additive genetic relationships among lines, was calculated using the available pedigree information. And dominance relationship matrix $\mathbf{D}$ (also assumed to be nonsingular) is the dominance matrix , which describes dominance variances and covariance among lines.

The corresponding Mixed Model Equation(MME) for the REML analysis

is:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z_1} & \mathbf{X'Z_2} \\ \mathbf{Z_1'X} & \mathbf{Z_1'Z_1} + \mathbf{A}^{-1}\alpha_a & \mathbf{Z_1'Z_2} \\ \mathbf{Z_2'X} & \mathbf{Z_2'Z_1} & \mathbf{Z_2'Z_2} + \mathbf{D}^{-1}\alpha_d \end{bmatrix} * \begin{bmatrix} \beta \\ \mathbf{a} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z_1'y} \\ \mathbf{Z_2'y} \end{bmatrix} \quad (29)$$

with $\alpha_a = \sigma_a^2/\sigma_e^2$, $\alpha_d = \sigma_d^2/\sigma_e^2$. Where σ_a^2 is the additive variance, σ_d^2 is the dominance variance and σ_e^2 is the residual effect.

2.15 QTLMAS XII workshop data

This is the simulated data set obtained from the QTLMAS XII workshop web page,

`http://www.computationalgenetics.se/QTLMAS08/QTLMAS/DATA.html`.

The dataset was generated following an animal breeding protocol, consisting of 5,865 individuals from seven generations. For the first four generations (total 4665 individuals) both pedigree and phenotype information are available, and this subset of data was considered for the analysis. Additive relationship matrix **A** and dominance relationship matrix **D** were calculated from the pedigree information. The main motivation to analyze the QTLMAS dataset was to test how the new method will behave in the absence of dominance effect.

2.16 Simulated data

A C program was developed, which simulates 'virtual' populations for the variance component estimation. Because of the identifiability problems faced during the analysis, two different datasets were considered, one of which resulted in a unimodal posterior distribution of dominance variance and another in a bimodal posterior. To develop the bimodal dataset, a base population of 50 unrelated lines were considered, where each of the 25 females

were mated with 25 males and each crossing resulted in 5 offspring (in total 3175 individuals, including the base population). For the unimodal dataset a base population of 40 lines were considered, 20 females and 20 males and each crossing resulted in 9 offspring (in total 3640 individuals, including the base population).

Additive relationship matrix $\mathbf{A}$ and dominance relationship matrix $\mathbf{D}$ were calculated from the pedigree information as described in the model section. To simulate a quantitative trait $\mathbf{y}$ three factors were generated, additive effect $\mathbf{a}$, dominance effect $\mathbf{d}$ and noise $\mathbf{e}$, and the vector of phenotypic observations was calculated as their sum,

$$\mathbf{y} = \mathbf{a} + \mathbf{d} + \mathbf{e}.$$

Here vectors $\mathbf{a}$, $\mathbf{d}$ and $\mathbf{e}$ were drawn from MVN$(\mathbf{0}, \mathbf{A}\sigma_a^2)$, MVN$(\mathbf{0}, \mathbf{D}\sigma_d^2)$ and MVN$(\mathbf{0}, \mathbf{I}\sigma_e^2)$ respectively. The Cholesky decomposition of the covariance matrices $\mathbf{A}\sigma_a^2$ and $\mathbf{D}\sigma_d^2$ was used to draw samples from these distributions. Hence the random genetic effects $\mathbf{a}$ and $\mathbf{d}$ were calculated as $\mathbf{a} = \mathbf{P}\mathbf{z}_a$ and $\mathbf{d} = \mathbf{T}\mathbf{z}_d$, where $\mathbf{z}_i \sim N(\mathbf{0}, \mathbf{I})$ and $\mathbf{P}$ and $\mathbf{T}$ are the Cholesky factors $\mathbf{P}\mathbf{P}' = \mathbf{A}\sigma_a^2$ and $\mathbf{T}\mathbf{T}' = \mathbf{D}\sigma_d^2$. To validate the new estimation methods, two data sets was generated, one dataset with single mode and another dataset with two mode, using heritability 0.31 (σ_a^2=800, σ_d^2=600, σ_e^2=3025).

2.17 Simulated dataset with finite number of loci

The above mentioned datasets were based on infinite number of loci so it was decided to considered a dataset with finite number of loci. To create the population I considered a base population of 20 lines, which assumed to be unrelated and homozygous with an in breeding coefficient of 0.99. In the first

crossing cycle, 5 lines of the base population were randomly chosen and crossed with another 5 randomly chosen lines, each cross produced 5, **F1** progeny lines. In the second crossing cycle each **F1** lines were selfed to produce 5, **F2** progeny lines and the crossings were carried out in same manner till the **F4** generation. The base population was assigned an inbreeding coefficient of 0.99 whereas **F1**, **F2**, **F3** and **F4** were assigned an inbreeding coefficient of 0.00, 0.50, 0.75 and 0.87 respectively.The simulated trait was controlled by 1000 unlinked loci having two alleles and the dominance was complete. The genotypic value at each locus of the a line was normally distributed with mean 0 and standard deviation 1. The phenotypic value for each line was simulated by adding genotypic effect location effect and residual error. The relationship matrix **A** describes the genetic relationships between individuals of a population and the algorithm as outlined by Henderson to compute the relationship matrix from the pedigree information. In contrast to Henderson (1976) inbreeding was considered for this dataset and the diagonal elements of the matrix A was of the form $1+F_i$ (where F_i is the inbreeding coefficient). Following Jacquard(1974), in an inbred population the dominance relationship between an individual **x** with parents **c** and **d** and an individual **y** with parents **e** and **f** can be calculated as follows:

$$D_{xy} = 1/3 * (1/2 * A_{xy} + 1/4 * A_{ce} * A_{df} + 1/4 * A_{cf} * A_{de}) \qquad (30)$$

Hence the elements of the dominance relationship matrix **D** can be calculated from the additive relationship matrix **A**. Horner and Charles 1956 provided the theoretical framework to calculate the sample variance components for the populations generated from crossing between homozygous lines and subsequent self-fertilization. In their study they showed that in the F2-generation, the relation of additive genetic and dominance variances is

equal. However, due to self-pollination, in the following filial generations the coefficient of additive genetic and dominance variance started to be different. In the current study, these differences were not accounted for the estimation of the variance components. Moreover it is difficult to calculate the variance components for different filial generations separately.

2.18 Field data:

Real data from 82 spring barley (H. *vulgare* L.) lines originating from German North Rhine Westphalia (Bauer *et al.*, 2006, 2008, 2009) core collection were analyzed. These lines were cultivated in randomized complete-block design with three replications in three different years (2001, 2002, and 2003) at the Research Station 'Dikopshof' of University of Bonn, Germany. For the real data, few replications were missing and the missing values were imputed by the average value for non missing replications for the corresponding year. There are a number of alternative ways of dealing with missing data, however as the number of missing values were very few we expect that those methods will not make much differences. Pedigree information was available for all the lines and the phenotypic observations of trait 'thousand kernel mass' was measured for all the lines. For the field data, we considered genotype-by-environment interaction instead of the dominance relationship in the linear mixed model (1) and accounted for the inbreeding among lines. Following Bauer *et al.* (2009), two different covariance structures were applied to the model the genotype-by-environment interaction. In the first approach called Bayes_ID, genotype-by-environment interaction was assumed to be independently and identically normally distributed. Whereas in the second approach (Bayes_A^{ext}) an extended relationship matrix $A^{ext} = A \otimes I$ (here '$\otimes$' is the Kronecker product of two matrices) was used to model the genotype-by-environment

interaction. Moreover the fixed year effect was considered in the X matrix along with the overall mean for the analysis.

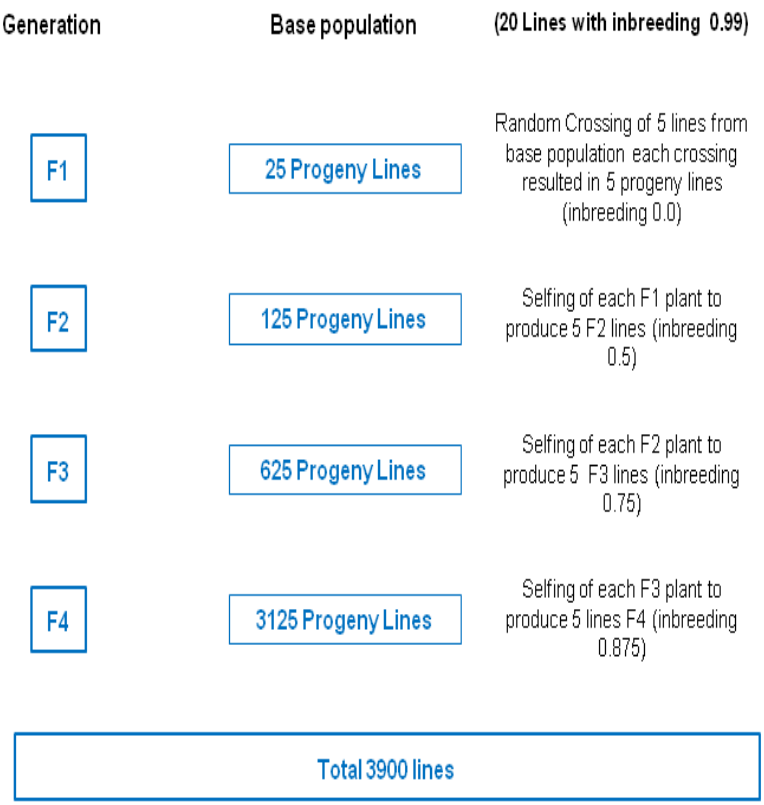

Figure 1: Schematic representation of the crossing of simulated dataset with finite number of loci till F4 generations.

3 Results

To validate the new algorithm I used both simulated data sets, QTLMAS XII workshop data and field data. It was decided to use large datasets for the analysis because of the identifiability problem with the dominance variance detected in test runs. Also I want to ensure that differences in the analysis results are not due to reasons other than real differences in the sampling efficiencies between two algorithms.

3.1 Class 1 adaptive MCMC

In the class 1 algorithm where the effect of breeding values and dominance effects were marginalized away before computing the posterior probability in the adapted phase. The variance components were estimated for both simulated data sets, QTLMAS XII workshop data and the field data using the class 1 adaptive MCMC algorithm.

3.1.1 simulated data

To validate the new algorithm, the analysis was done with two simulated datasets, the unimodal dataset with 3640 individuals and the bimodal dataset with 3175 individuals. The estimates for variance components based on all the individuals of the simulated datasets were calculated using the new adaptive MCMC method and the REML method (Table 1). The REML estimates of the variance components were calculated using the ASReml software (Gilmour *et al.* 2006). True values given in Table 1 are the values used in the simulations. The implemented MCMC had a total chain length of 50000, consisting of a burn-in period of 2000 iterations, a learning phase from iteration number 2000 to 5000, and finishing with the adapted phase from iterations 5000

Table 1: The estimates of variance components and broad-sense heritabilities for the learning and adapted phases from the MCMC analyses of the two simulated datasets. REML estimates and true simulated values are also shown. The names 'unimodal data' and 'bimodal data' are based on the characteristics that these data sets showed during the MCMC analysis.

	Variance components estimated from the learning phase			Variance components estimated from the adapted phase			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
Bimodal data								
σ_a^2	672.57	607.73	573.67	721.49	695.87	679.37	752.99	800
σ_d^2	545.93	510.21	453.20	493.10	522.30	675.40	716.00	600
σ_e^2	3107.70	3143.70	3013.70	3132.10	3105.00	3020.20	2882.60	3025
h^2	0.28	0.26	0.25	0.27	0.28	0.30	0.33	0.31
Unimodal data								
σ_a^2	873.36	820.90	879.80	751.20	744.53	779.80	781.28	800
σ_d^2	619.36	642.23	658.70	591.50	585.77	579.80	571.68	600
σ_e^2	2845.40	2865.00	2894.70	2965.00	2971.90	2960.90	2928.79	3025
h^2	0.34	0.33	0.34	0.33	0.31	0.31	0.31	0.31

to 50000. Acceptance ratios for the bimodal and unimodal data sets were 28% and 26%, respectively. The point estimates, mean and median of the posterior distribution of the variance components, were calculated from the MCMC samples. The point estimates can give an indication that whether the posterior distributions is close to normality or not. To calculate the mode of the posterior distribution a kernel smoothing approach following (Hoti *et al.* 2002), was used.

A properly implemented MCMC sampler should be able to cover all the areas supported by the target distribution, but the existence of multiple modes makes this difficult. A conventional MCMC algorithm usually fail to jump between the different modes and therefore may visit only a single mode. Although running the chain for a very long time is a solution for this problem,

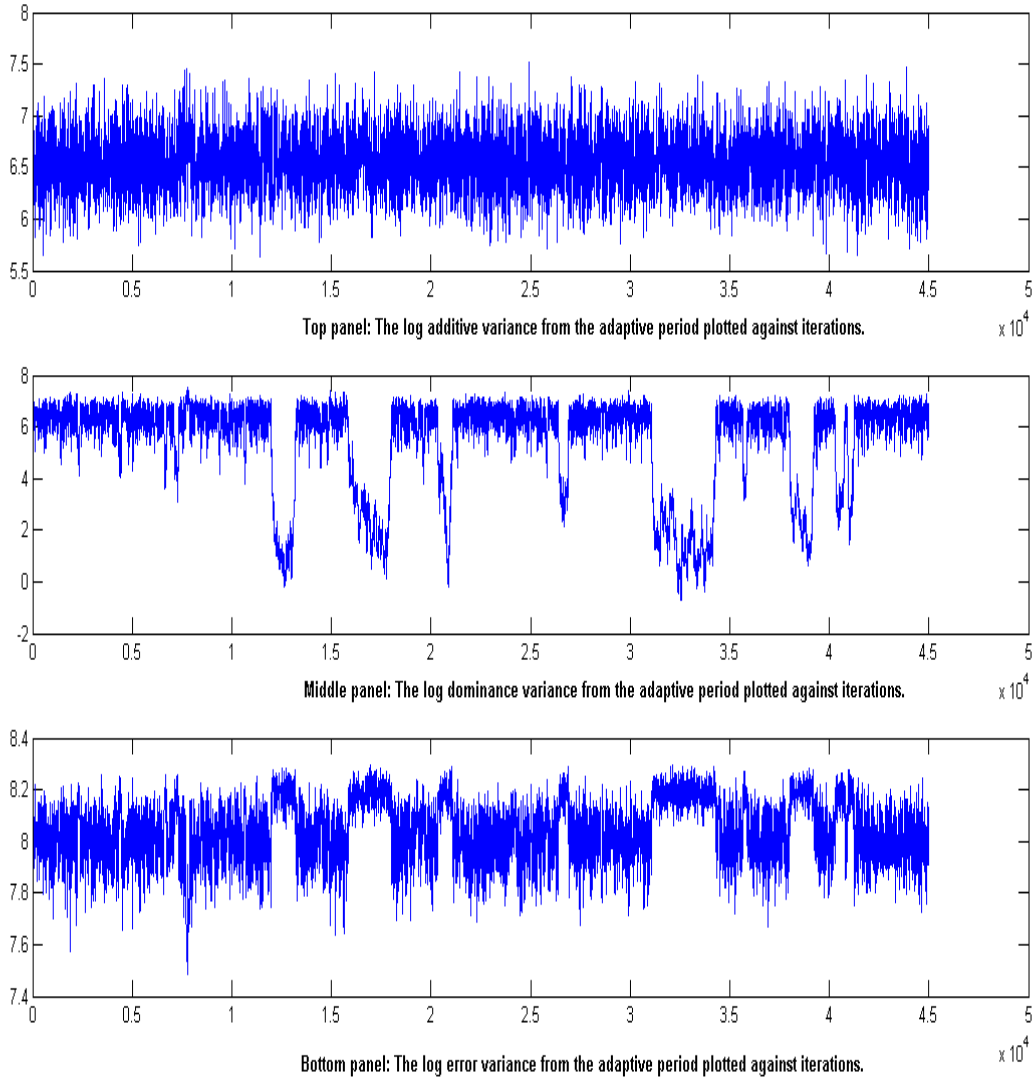

Figure 2: The logarithm of the variance components for the bimodal dataset plotted against MCMC iteration number. The trace plots show 45000 iterations from the adapted phase. From the figure after 12000 iterations the chain moves to a different mode in the posterior distribution. The same mode is again visited after a certain number of iterations.

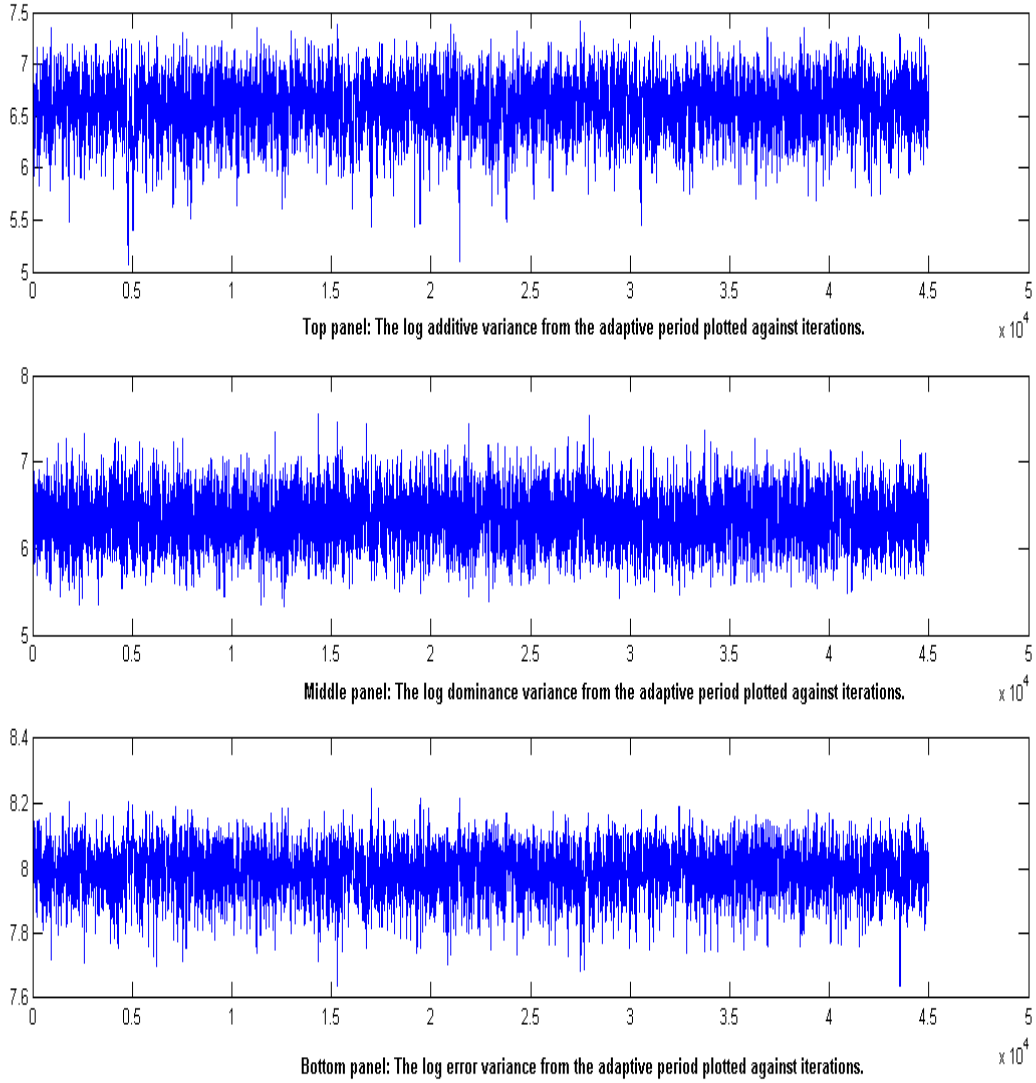

Figure 3: The logarithm of the variance components for the unimodal dataset plotted against MCMC iteration number. The trace plots show 45000 iterations from the adapted phase.

but this is computationally highly demanding. In the new approach, the posterior covariance structure estimated from the learning phase helps the sampler to move freely between the different modes of the target distribution. The new MCMC algorithm was able to detect two different modes in the posterior for dominance and additive variances in the bimodal dataset, whereas REML always returns a single mode (and the identified mode may depend on its starting values). Figure 2, shows that adaptive MCMC algorithm is able to move between the different modes of the posterior of the bimodal dataset. In Figure 2 the X-axis represent the number of iterations and Y-axis represent the the variance components in logarithmic scale. From the trace plots for the dominance and error variance components from Figure 2 it can seen that around 12000th iterations both dominance and error variance components move to a different region in the parameter space. The same region is again visited by the algorithm after certain number of iterations. So it can be concluded that there are different modes in the posterior distribution. From Figures 2 and 4 that the new adaptive MCMC algorithm was able to detect different modes in the distribution with a relatively low number of iterations, whereas the conventional MCMC method like hybrid Gibbs sampler had problems visiting different modes in the test runs (results are not shown). Table 2 summarizes the rough estimates for the two different modes. To estimate the modes, the kernel smoothing approach (Hoti *et al.* 2002) was applied. The mode 1 values are close to the true simulated values. In order to visualize the different modes in the posterior, a histogram with hexagonal bins was drawn for the log-transformed dominance and error variance components (Figure 4) with the aid of the `hexbin` package of R.

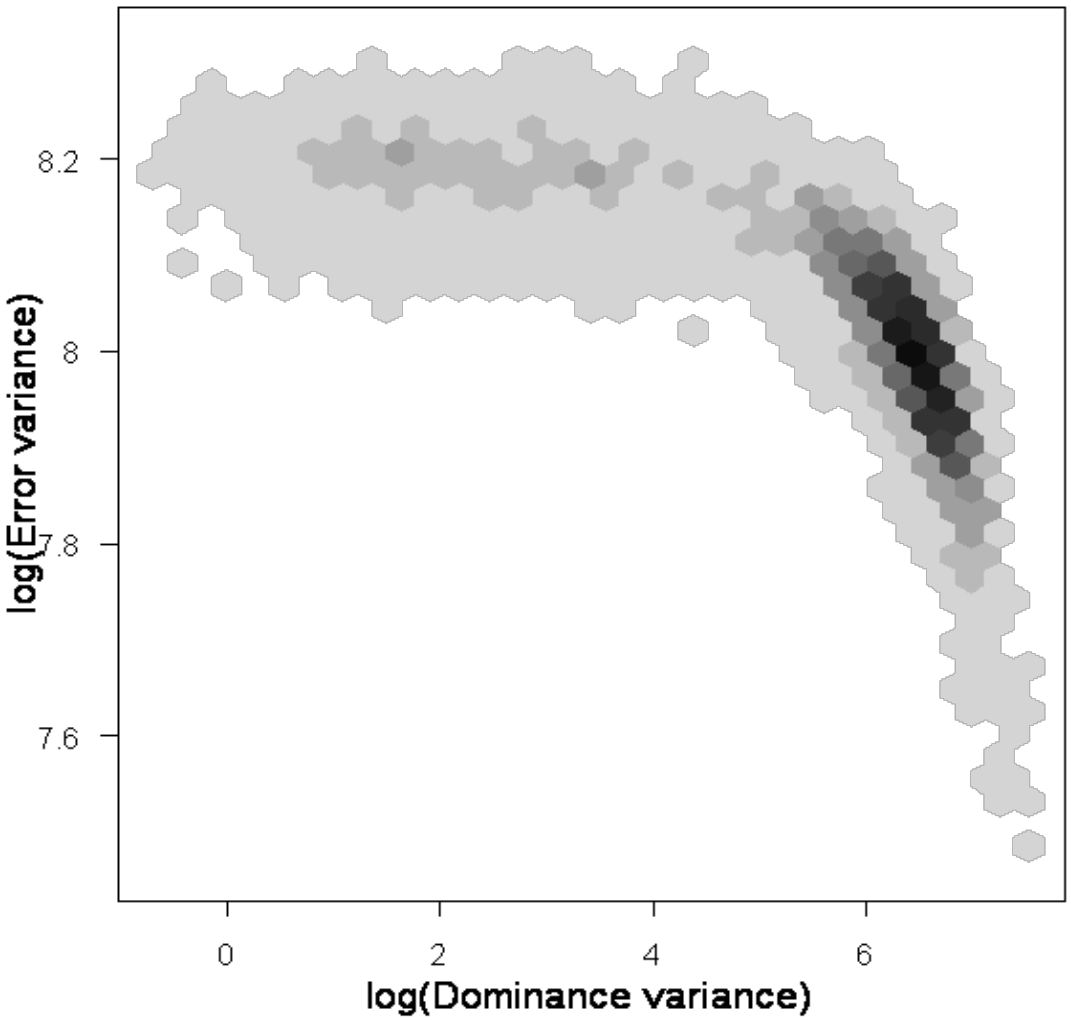

Figure 4: Histogram of the log-transformed dominance and error variance components using hexagonal bins. The plotting plane is divided into a number hexagons and darker the hexagon, the more points it represents.

Table 2: The two different modes of the variance components from the adapted period for the simulated bimodal dataset. The posterior mode estimates are obtained from the adapted phase of the MCMC analysis. REML estimates and true simulated values are also shown.

	Mode 1	Mode 2	REML	True
σ_d^2	675.40	2.67	716.00	600
σ_e^2	3020.20	3505.2	2882.60	3025

3.1.2 QTLMAS XII workshop data

I considered a subset of 4665 individuals (first four generations) from the QTLMAS XII workshop data for the analysis. The pedigree information for the first four generations was available, and hence **A** and **D** matrices were calculated from the pedigree. For further details of data, see Lund *et al.* (2009). The heritability of the QTLMAS XII workshop data was around 0.30 with zero dominance effect. The main motivation to analyze the QTLMAS dataset was to test how the new method behaves in the absence of a dominance effect. The variance components were estimated using the adaptive MCMC and the REML methods (Table 3). The implemented MCMC had a total chain length of 50000 with a burn-in period of 2000 iterations, a learning phase from iteration number 2000 to 5000, and the adapted phase from iterations 5000 to 50000. The acceptance ratio for the data set was 35%. The point estimates were calculated as before. In the analysis I obtained heritability around 0.30. Hallander *et al.* (2010) used a different prior and obtained a heritability point-estimate of 0.34 from a smaller subset of data, using a Bayesian model containing additive polygenic effects only. They used uniform distributions as non-informative choice of priors to the standard deviations.

Table 3: The estimates of the variance components and broad-sense heritabilities for the learning and adapted phases from the MCMC analysis of the QTLMAS XII dataset. REML estimates and true simulated values for entire pedigree are also shown. The true value for the additive variance was calculated as the variance of true genomic breeding values omitting relationships between individuals and the residual variance was calculated accordingly to obtain a heritability around 0.30.

	Variance components estimated from the learning phase			Variance components estimated from the adapted phase			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
σ_a^2	1.33	1.32	1.10	1.34	1.33	1.31	1.35	1.36
σ_d^2	0.09	0.10	0.10	0.01	0.00	0.00	0.00	0.00
σ_e^2	3.06	3.06	2.84	3.13	3.13	3.15	3.12	3.20
h^2	0.46	0.46	0.29	0.30	0.29	0.29	0.30	0.30

3.1.3 Effective Sample Size

Effective Sample Size (ESS) (Waagepetersen *et al.* 2008; Geyer 1992) is a popular diagnostic tool for MCMC methods. A high value of ESS implies that the autocorrelation is low and which is an indication that the mixing of the MCMC chain is good. ESS determines the approximate number of independent samples which would provide the same estimation accuracy as the dependent MCMC samples. The ESS values were calculated with the R package `coda` (Plummer *et al.* 2006).

Adequate mixing of MCMC sampler over different parts of the parameter space is essential for the convergence of MCMC algorithms, but conventional MCMC algorithms may suffer from slow mixing. From the trace plots (Figure 5) for the learning phase and adapted phase, it shows that the adapted MCMC is mixing well compared to the general hybrid Gibbs sampler (used in the learning phase). Thus the adaptation has significantly improved the mixing

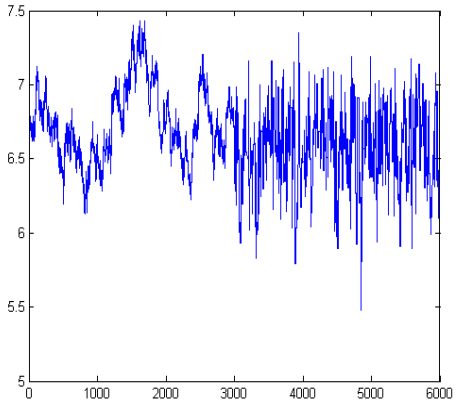

Figure 5: Trace plot of the log-transformed additive variance component for the unimodal simulated dataset. First 3000 samples are taken from the learning phase and the remaining samples are from the adapted phase.

property of the algorithm, by learning an appropriate covariance structure for the proposal distribution. This visual impression is confirmed by Table 4, which summarizes the ESS for the unimodal, bimodal and QTLMAS datasets. To calculate the ESS, an MCMC chain with a length of 3000 from the learning phase and a chain of same length from the beginning of the adapted phase were considered after a burn-in period of 2000 iterations. The ESS values from Table 4 clearly support better mixing properties of variance components in the adapted phase for all the datasets. The new proposed prior allows the chain to mix well, and at the same time it allows a realistic estimate of the dominance variance also in the case of no dominance, because in such a case the prior shrinks the posterior towards zero.

When the target distribution is bimodal, the conventional MCMC algorithm may have difficulties moving between modes. Also the REML method fails to identify different modes of the distribution. The new adaptive MCMC algorithm was able to visit the different modes even after a low number of

iterations and showed good mixing properties.

Table 4: Effective Sample size (ESS) for 3000 iterations of the two MCMC algorithms with the unimodal, bimodal and QTLMAS datasets.

	ESS for the variance components from the learning phase			ESS for the variance components from the adapted phase		
	σ_a^2	σ_d^2	σ_e^2	σ_a^2	σ_d^2	σ_e^2
Unimodal dataset						
ESS	10.36	8.76	29.56	176.71	318.15	191.81
Bimodal dataset						
ESS	12.82	12.10	11.27	240.56	159.68	176.44
QTLMAS dataset						
ESS	41.24	3.58	116.98	242.90	93.37	204.64

3.2 Class 2 adaptive MCMC

Marginalizing the nuisance parameters from the likelihood is generally beneficial to the MCMC mixing. To know the impact of marginalization on the mixing properties a model was considered without marginalizing the random effects in the adapted phase of the algorithm. In the class 2 algorithm where the effect of breeding values and dominance effects were included to compute the posterior probability in the adapted phase. ESS (Table 5) in the adapted phase of the class 2 MCMC algorithm was near to the true values than the learning phase, however the variance components estimates (Table 6) for the class 2 adaptive MCMC in the adapted phase were nowhere near the true values. The acceptance ratio of the dataset was around 67%. Here the additive and dominance variance were summed to the error variance. Posterior correlation between the parameters restricts the free movement of the chain and the chain remains at specific points for longer time. But it is

crucial that the Markov chain has good mixing (fully explore the likelihood surface) properties. By comparing the results for the class 1 and class 2 algorithm one can conclude that marginalization play a crucial role in the mixing properties of the chain.

Table 5: Effective Sample Size (ESS) for 3000 iterations from the learning phase and 3000 iterations from the adapted phase for the class 2 MCMC with the unimodal, bimodal and QTLMAS datasets.

	ESS for the variance components from the learning phase			ESS for the variance components from the adapted phase		
	σ_a^2	σ_d^2	σ_e^2	σ_a^2	σ_d^2	σ_e^2
ESS						
Unimodal data	10.36	8.76	29.56	92.76	239.67	19.93
Bimodal data	12.82	12.10	11.27	51.16	160.35	9.25
Workshop data	41.24	3.58	116.98	12.08	656.36	9.38

Table 6: The estimates of the variance components and broad-sense heritabilities for the learning and adapted phases from the class 2 algorithm of the QTLMAS XII dataset. REML estimates and true simulated values for entire pedigree are also shown.

	Variance components estimated from the learning phase			Variance components estimated from the adapted phase			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
Workshop data								
σ_a^2	1.32	1.32	1.10	0.12	0.02	0.01	1.35	1.36
σ_d^2	0.09	0.10	0.10	0.03	0.06	0.02	0.00	0.00
σ_e^2	3.06	3.06	2.84	5.59	5.01	4.81	3.12	3.20
h^2	0.46	0.46	0.29	0.02	0.00	0.00	0.30	0.30

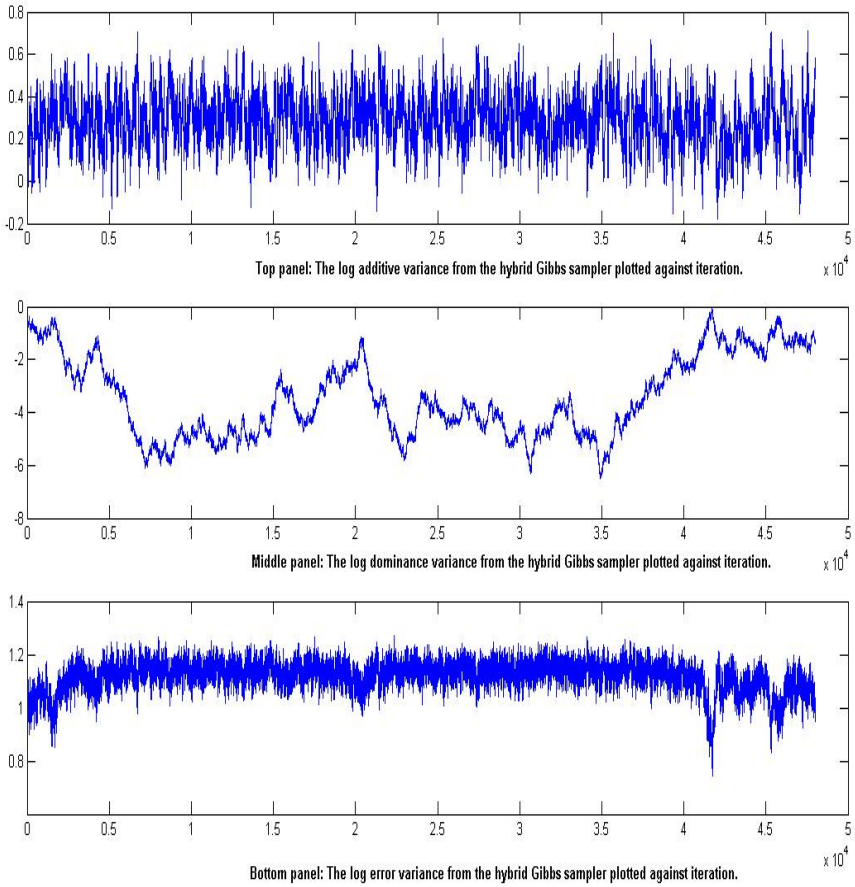

Figure 6: The logarithm of the variance components for the workshop dataset with scaled inverse chi-square prior plotted against MCMC iteration number. The trace plots show 48000 iterations from the normal hybrid Gibbs sampler.

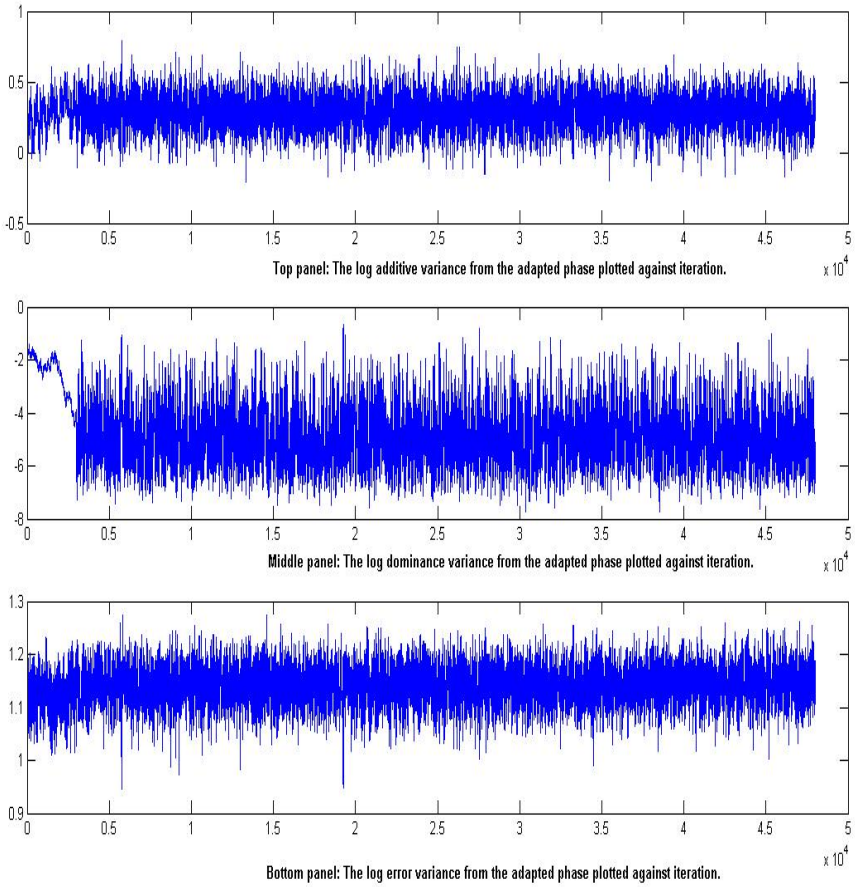

Figure 7: The logarithm of the variance components for the workshop dataset plotted against MCMC iteration number. The trace plots show 3000 iterations from the learning phase and 45000 iterations from the adapted phase.

3.3 Sensitivity analysis

In the current study I was also interested to know the impact of prior distributions on the ESS and estimation accuracy. So the analysis was done with two different prior distribution: 1) Gamma prior distribution 2) Scaled inverse chi-square distribution. For the Gamma prior distribution $\mathbf{k} = 1$ and $\lambda = 0.001$, was assigned to obtain flat prior. For the Scaled inverse chi-square v_i = -2 and $S_i^2 = 0$, was assigned to obtain flat prior. ESS were calculated for two different priors using normal hybrid Gibbs sampler. To calculate the ESS a MCMC chain of length 48000 iterations was considered after a burning period of 2000 iterations and Table 7 summarizes the results. For the Gamma prior distribution the phenotypic observation vector $\mathbf{y}$, was standardized to use same prior for different dataset. Table 7 shows that there is no significant difference between ESS for the two different prior distributions. However I decided to use Gamma prior distribution in the learning phase of the algorithm, because I wanted to use the same prior in both the learning and adapted phase of the algorithm.

Table 7: Effective Sample size (ESS) for the Scaled inverse chi-square prior and Gamma prior distribution for the workshop data. A MCMC chain of length 48000 iterations from the normal Gibbs sampler was considered to calculate the ESS.

	σ_a^2	σ_d^2	σ_e^2
Gamma prior(ESS)			
Workshop data	623.12	16.98	592.14
Bimodal data	258.45	21.29	35.86
Unimodal data	176.92	128.90	169.90
Scaled inverse chi-square prior(ESS)			
Workshop data	601.66	15.02	166.00
Bimodal data	256.6	67.61	99.10
Unimodal data	173.19	147.34	190.62

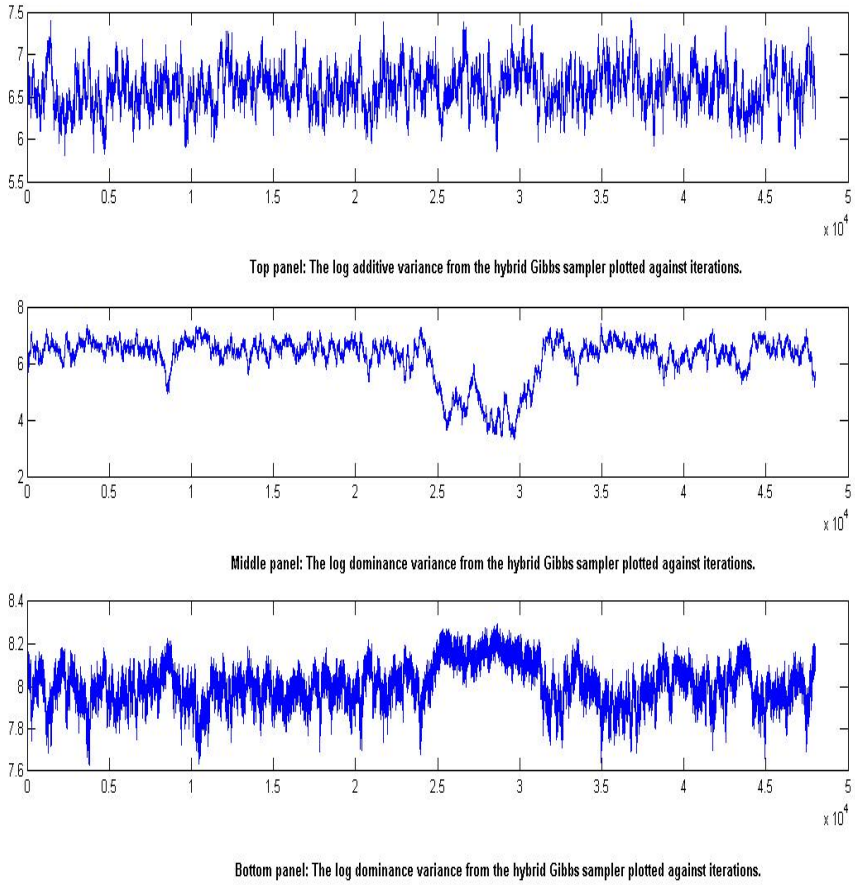

Figure 8: The logarithm of the variance components for the bimodal dataset with scaled inverse chi-square prior plotted against MCMC iteration number. The trace plots show 48000 iterations from the normal hybrid Gibbs sampler.

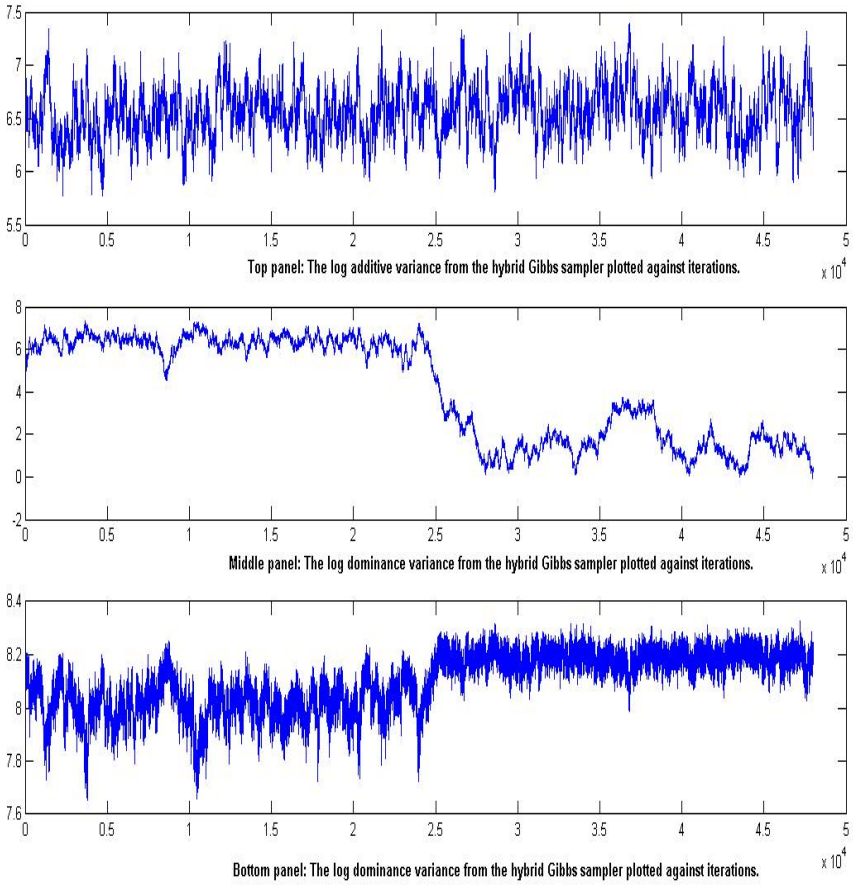

Figure 9: The logarithm of the variance components for the bimodal dataset with Gamma prior plotted against MCMC iteration number. The trace plots show 48000 iterations from the normal hybrid Gibbs sampler.

By comparing the trace plots of the variance components for the bimodal data using Gamma prior (Figure 8) and the scaled inverse chi-square distribution (Figure 9) with the hybrid Gibbs sampler, scaled inverse chi-square prior was able to move rapidly between two different modes. Whereas the movement of the Gamma prior between different modes were slow. The variance components were also calculated for the simulated data sets (Table 8) and the workshop data set (Table 9) for the two different prior distributions using hybrid Gibbs sampler. From Figure 6 and 6 for the workshop dataset the adaptive MCMC algorithm shows better mixing properties than the hybrid Gibbs sampler and it was supported by ESS values, also there was much improvement in the mixing of dominance variance in the adapted phase. From the Figures (6 and 7) one can conclude that the marginalization of the random effects reduces the autocorrelation between the parameters, which eventually improved the mixing of the chain. From Table 8 the estimates for the unimodal dataset using scaled inverse chi-square prior gave better estimates compared to the Gamma prior distribution, because the Gamma prior was stayed in one mode for a long period of time. But in both case the chain is not converged and more number of iterations are needed for the convergence of the MCMC chains. Generally the prior has little influence on the estimated parameters when the analysis is done with large size of pedigree data. But in the case of workshop data (Table 9) with zero dominance the Gamma prior was able to provide values near to the true values with the hybrid Gibbs sampler whereas the chi-square prior failed to do so. Indicating that the new Gamma prior is working well for datasets with zero dominance. Also the ESS values was higher for the workshop data with Gamma prior distribution.

Table 8: The variance components and broad-sense heritability for different prior distributions for the bimodal, unimodal and workshop datasets. A MCMC chain of length 48000 iterations from the hybrid Gibbs sampler was considered to calculate the variance components.

	Variance Components estimated using Gamma prior			Variance Components estimated using scaled inverse chi-square prior			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
Bimodal								
σ_a^2	725.80	699.68	580.90	757.32	730.10	670.12	752.99	800
σ_d^2	317.94	207.87	13.45	605.39	626.68	579.15	716.00	600
σ_e^2	3304.00	3359.00	3608.00	3009.10	2989.00	2059.03	2882.60	3025
h^2	0.24	0.21	0.14	0.31	0.31	0.37	0.33	0.31
Unimodal								
σ_a^2	747.06	743.69	720.90	786.13	781.20	750.15	752.99	800
σ_d^2	578.66	554.43	540.45	620.10	593.20	570.85	716.00	600
σ_e^2	2976.80	2975.70	2965.00	2924.20	2925.60	2920.00	2882.60	3025
h^2	0.30	0.30	0.30	0.32	0.32	0.31	0.33	0.31

Table 9: The variance components and broad-sense heritability for different prior distributions for the workshop data. A MCMC chain of length 48000 iterations from the normal Gibbs sampler was considered to calculate the variance components.

	Variance Components estimated using Gamma prior			Variance Components estimated using scaled inverse chi-square prior			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
σ_a^2	1.33	1.32	0.98	1.34	1.33	1.10	1.35	1.36
σ_d^2	0.02	0.00	0.00	0.09	0.02	0.10	0.00	0.00
σ_e^2	3.12	3.12	2.90	3.06	3.08	3.10	3.12	3.20
h^2	0.30	0.30	0.25	0.32	0.30	0.27	0.30	0.30

In the current study I also used the empirical variance of the phenotypic observation vector $\mathbf{y}$, as the starting value for λ with Gamma prior distribution. But this prior provided non-zero estimate for the dominance variance with workshop data. So it was decided to use flat prior for the analysis, which was able to provide better estimate in the case of zero dominance.

3.4 Estimation using scaled inverse chi-square prior distribution in the learning phase

In the current study I also calculated the variance components and ESS using scaled inverse chi-square prior in the learning phase and Gamma prior in the adapted phase using the class 1 adaptive MCMC algorithm. The acceptance rate was around 20%. Table 10 compare the variance components estimates for the scaled inverse chi-square prior in the learning phase and Gamma prior in the adapted phase (Chi-Gamma) with Gamma prior in both phases (Gamma-Gamma) for the workshop dataset. And Table 11 summarizes the ESS calculated for those two cases (Gamma-Gamma and Chi-Gamma). In the learning phase of the algorithm Gamma prior was able to provide better estimates than the chi-square prior. But in the adapted phase there was no significance difference between the estimated variance components. Both priors used in the learning period were able to provide the optimal covariance matrix for the proposal distribution. However the values of ESS was high for the Gamma prior in both learning phase and adapted phase, so it is recommended to use Gamma prior distribution in the learning period of the algorithm.

Both priors was able to give zero estimate for the dominance variance in the adapted phase of the algorithm. Also the posterior distributions for the variance components were close to normality. The idea to use Gamma (1,

Table 10: The estimates of the variance components and broad-sense heritabilities for the learning and adapted phases from the MCMC analysis of the QTLMAS XII dataset using two different prior distributions in the learning phase. REML estimates and true simulated values for entire pedigree are also shown. The true value for the additive variance was calculated as the variance of true genomic breeding values omitting relationships between individuals and the residual variance was calculated accordingly to obtain a heritability around 0.30.

	Variance components estimated from the learning phase			Variance components estimated from the adapted phase			REML	True
	Mean	Median	Mode	Mean	Median	Mode		
Gamma-Gamma								
σ_a^2	1.33	1.32	1.10	1.34	1.33	1.31	1.35	1.36
σ_d^2	0.09	0.10	0.10	0.01	0.00	0.00	0.00	0.00
σ_e^2	3.06	3.06	2.84	3.13	3.13	3.15	3.12	3.20
h^2	0.46	0.46	0.29	0.30	0.29	0.29	0.30	0.30
Chi-Gamma								
σ_a^2	1.32	1.33	1.18	1.35	1.34	1.30	1.35	1.36
σ_d^2	0.31	0.34	0.05	0.00	0.00	0.00	0.00	0.00
σ_e^2	2.86	2.86	2.72	3.14	3.14	3.12	3.12	3.20
h^2	0.36	0.36	0.31	0.30	0.30	0.29	0.30	0.30

Table 11: Effective Sample size (ESS) for 3000 iterations from the learning phase and 45000 iterations from the adapted phase of the MCMC algorithm using different priors in the learning phase with QTLMAS dataset.

	ESS for the variance components from the learning phase			ESS for the variance components from the adapted phase		
	σ_a^2	σ_d^2	σ_e^2	σ_a^2	σ_d^2	σ_e^2
Gamma-Gamma						
ESS	41.24	3.58	116.98	4000.07	2750.97	3858.53
Chi-Gamma						
ESS	38.58	3.42	23.09	2704.53	3046.28	3335.62

0.001) prior for precision was that variances are shrinked towards zero. This will help model selection automatically meaning that if there is no dominance in the data, this prior will shrink the variance component value towards zero and resulting posterior estimate should also be near zero. Model selection is one of the main problem associated with variance component estimation and the new prior was able to do model selection in the analysis.

3.5 Simulated dataset with finite number of loci

In contrast to the other datasets, for the current study it was decided to simulate a dataset with finite number of loci. The data was simulated in a more piratical frame work considering progenies till **F4** generations. In order to know the impact of inbreeding on the performance of the proposed new algorithm I decided to account inbreeding in the estimation of genetic parameters. Generally in the infinitesimal frame work each trait is assumed to be influenced by an infinite number of additive genes having small effect. To bring the simulated dataset into close proximity with the infinitesimal frame work, I considered around 1000 normally distributed loci for the simulation.

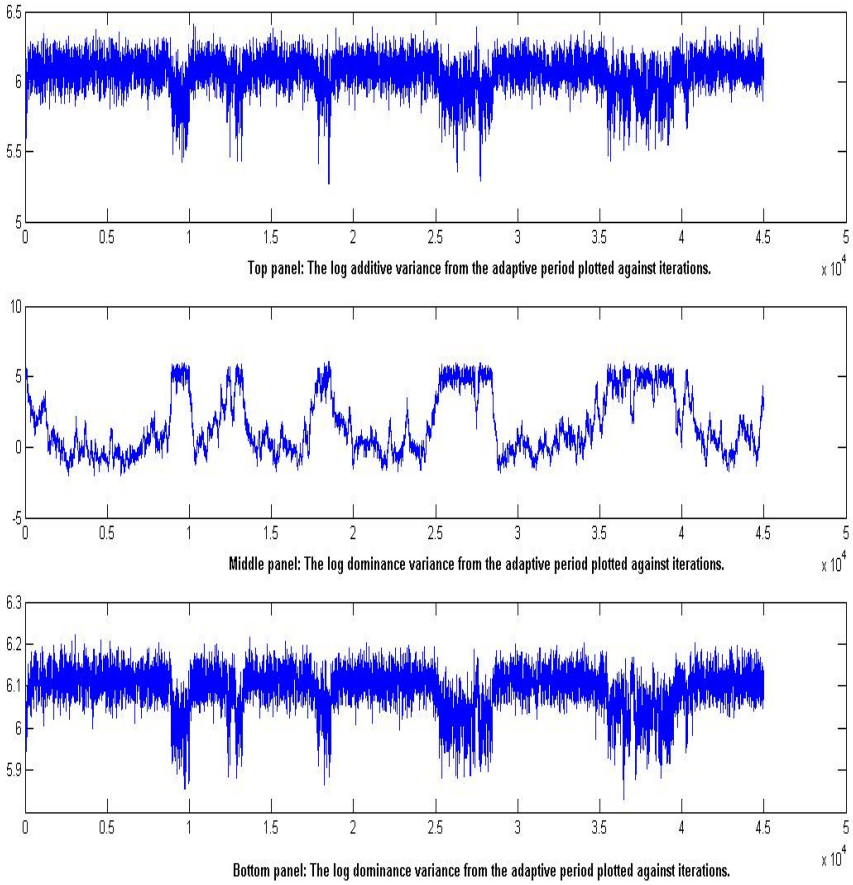

Figure 10: Trace plot of the log-transformed variance component for the simulated dataset with finite number of loci from the adapted phase of the class 1 adaptive MCMC algorithm.

The variance components and the effective sample size were calculated for the simulated dataset with finite number of loci and Table 12 and Table 13 summarizes those results. The variance components estimates from the learning phase were close to the REML estimates, however the variance components estimates from the adapted phase were biased because of the multiple modes (Figure 10) present in the posterior distributions. Moreover it was difficult to get the true variance components for the dataset because of inbreeding and the number of generations (F4) used for the analysis.

Table 12: The estimates of the variance components and broad-sense heritabilities for the learning and adapted phases from the MCMC analysis of the simulated dataset with finite number of loci. REML estimates for entire pedigree is also shown.

	Variance components estimated from the learning phase			Variance components estimated from the adapted phase			REML	True
	Mean	Median	Mode	Mean	Median			
σ_a^2	350.14	350.08	238.46	430.24	433.89	460.39	369.38	
σ_d^2	215.75	214.4	113.27	38.53	2.47	1.85	184.13	
σ_e^2	402.01	401.77	347.47	442.92	446.37	446.99	407.49	
h^2	0.58	0.58	0.43	0.51	0.50	0.51	0.58	

Table 13: Effective Sample size (ESS) for 3000 iterations from the learning phase and adapted phase for the simulated dataset with finite number of loci.

	ESS for the variance components from the learning phase			ESS for the variance components from the adapted phase		
	σ_a^2	σ_d^2	σ_e^2	σ_a^2	σ_d^2	σ_e^2
ESS	30.31	13.34	66.87	328.73	12.07	117.22

3.6 Estimation of breeding values

During the adapted phase of the algorithm, the sampler generates values only from the marginal posterior of the variance components. Primary focus of the method is the estimation of the genetic variances, however it is possible to generate MCMC samples for the additive and dominance genetic values afterwards. In the current study breeding values were estimated by sampling them block-wise from their fully conditional posterior distribution conditionally on each of the values of the variance components in the MCMC sample generated by the adaptive MCMC sampler. In contrast, in the normal hybrid Gibbs sampler, the genetic values are sampled conditionally on each of the values of the variance components. This procedure was tested by calculating the genetic values for the QTLMAS workshop data by using the blocked Gibbs sampler conditionally on every 10^{th} realization (of three variance components) out of 45000 samples from the adapted phase. The linear correlation between the true genetic values (i.e., sum of additive and dominance values) and the estimated genetic values was around 0.71 for the QTLMAS workshop data. Also the REML genetic values showed a correlation around 0.71 with the true genetic values for the same dataset. The adaptive MCMC genetic values showed a strong correlation of around 0.99 with the REML estimates, showing that the adaptive MCMC posterior mean estimates were near to the REML estimates. Also I calculated the correlation for the additive posterior mean estimates as well the dominance estimates from the adapted phase with corresponding REML estimates and both correlations were very high (0.99). Breeding values for the simulated dataset with finite number of loci were also calculated and the correlations were calculated with the corresponding REML breeding values. Breeding values calculated using the adaptive MCMC method showed a strong correlation around 0.83 with

the true simulated genetic values. REML estimates also showed a strong correlation around 0.90 with the true genetic values. Moreover the correlation between the REML estimates and the adaptive MCMC method was around 0.91. For estimation of the breeding values I considered the infinitesimal model. However, Hoeschele *et al.* (1993) proposed a finite locus model for dataset with finite number of loci but in the current research I did not consider the finite locus model.

Table 14: Correlation coefficient (r) calculated between the estimated breeding value and the true genetic value using REML and adaptive MCMC method.

	REML estimated genetic values	Adaptive MCMC estimated genetic values
QTLMAS workshop data		
True genetic values	0.71	0.71
Dataset with finite number of loci		
True genetic values	0.90	0.83

3.7 Field data

The trait 'thousand kernel mass' for 82 spring barley lines from three different years with three replications were considered for analysis with our Calss 1 adaptive MCMC method as well as REML using ASReml software (by assuming same covariance structure for genotype-by-environment interaction as in Bayes_ID). The implemented MCMC had a total chain length of 50000, consisting of a burn-in period of 2000 iterations, a learning phase from iteration number 2000 to 5000, and finishing with the adapted phase from iterations 5000 to 50000. For analysis each year was considered as different location. So

Table 15: The estimates of variance components, heritabilities and the 95% HPD intervals for the field data from the adapted phases of the algorithm using Bayes_ID and Bayes_A^{ext} covariances. REML estimates are also shown.

	Bayes_ID					Bayes_A^{ext}					
	Mean	Median	Mode	2.5	97.5	Mean	Median	Mode	2.5	97.5	REML
σ_a^2	9.27	9.13	9.08	5.15	13.54	9.10	9.15	9.05	5.20	13.69	9.21
σ_h^2	2.45	2.49	2.60	0.00	4.61	2.98	2.70	2.84	0.00	4.77	3.18
σ_e^2	17.67	17.58	17.50	15.33	20.30	17.18	17.23	17.29	15.36	20.25	17.08
h^2	0.76	0.76	0.76			0.76	0.76	0.76			0.75

in order to account the number of locations, heritability was calculated using the formula (Hanson 1963)

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + (\sigma_{g\times e}^2/j) + (\sigma_e^2/j * k)}$$

where σ_a^2 is the additive genetic variance, $\sigma_{g\times e}^2$ is the variance due to genotype-by-environment interactions, σ_e^2 is the error variance, j is the number of years and $k = 3$ is the number of replications. We also calculated the point estimates and 95% highest posterior density intervals for the posterior distribution from the adapted phase of the algorithm using Bayes_ID and Bayes_A^{ext} methods (Table 15). Bauer *et al.* (2009) considered data from two different years (2002 and 2003) for the analysis and in our current study we considered data from three different years (2001, 2002 and 2003). Hence our analysis provided higher heritabilities than in Bauer *et al.* (2009). Both studies showed that the Bayes_A^{ext} estimates were more close to the REML estimates. Moreover, results from both studies indicated that it is important to consider the relationship information between lines while estimating the genotype-by-environment interactions.

4 Discussion

In this study a new adaptive MCMC algorithm was proposed with superior mixing properties compared to the conventional MCMC algorithms. Also the performance of the new algorithm was compared with some of the existing methods. The results obtained from those analysis are discussed below.

4.1 Computational cost (Adaptive MCMC vs hybrid Gibbs sampler)

One of the main problems associated with Bayesian analysis of mixed models with several random effects is that the analysis is computationally demanding. Still much focus is given to improve the computational efficiency of the MCMC algorithms. The single site Gibbs sampling algorithm is faster but suffer with poor mixing of the chain. However the Blocked Gibbs sampler has good mixing properties but computationally demanding. Waldmann *et al.* (2008) proposed a new hybrid Gibbs sampler which was a combination of both single site Gibbs sampler and blocked Gibbs sampling algorithm. The hybrid Gibbs sampler is much faster than the normal blocked Gibbs sampler for estimating additive and dominance genetic variances in the traditional infinitesimal model. In the current study the performance of the hybrid Gibbs sampler was compared with the adaptive MCMC method using simulated pedigree datasets with non-zero additive and dominance genetic variances but no inbreeding, showing that the new adaptive MCMC algorithm was almost two times faster than the hybrid Gibbs sampler. To compare the running times, an adaptive MCMC chain of total length 50000 (burn-in period of 2000 iterations, 3000 iterations in the learning phase and 45000 iterations in the adapted phase) was compared with a hybrid Gibbs sampling chain of same

total length (burn-in period of 2000 iteration and 48000 iterations from the normal hybrid Gibbs sampling). What is more, the adaptive algorithm has superior mixing properties, as shown by the effective sample sizes in Table 4. The speed up is partly due to the fact that, unlike the algorithm of Waldmann *et al.* (2008), the adaptive MCMC does not sample additive and dominance genetic values for individuals at all. In the adaptive MCMC algorithm, the determinants and quadratic forms associated with the covariance matrices at the proposed and current points are needed to calculate the likelihood ratio. Once the proposed value is accepted the determinant and quadratic form at the current point can be replaced by the determinant and quadratic form corresponding to the accepted variance components. This makes the calculation of the likelihood ratio computationally less demanding than the block update of the Gibbs sampler.

4.2 Estimation of breeding values

The initial objective of the study was to develop an efficient MCMC algorithm for the estimation of breeding values for the self pollinating crops. However, because of the practical and theoretical problem faced during the study more focus was given to the accurate estimation of the variance components. Moreover the estimation of true variance component and broad-sense heritability of a trait is important for the calculation of breeding values. In conventional MCMC methods the breeding values and variance components are sampled simultaneously. But the dependences among the breeding values may lead to biased estimates of variance components also one has to run the algorithm for a long period to get proper convergence for the MCMC chain, this is computationally challenging. In the new approach it is possible to take samples from a converged MCMC chain for the variance components in order to calculates

the breeding values. This new approach can provide better estimates for the breeding values with less computational cost. Additionally breeding values estimated using the new approach was supported with a strong correlation of around 0.99 with the REML estimates. Belonsky and Kennedy (1988) has already shown the advantage of using pedigree information during the selection process in animal breeding programs. Piepho *et al.* (2008) shown that in BLUP analysis, without considering the complete pedigree information can lead to biased estimates. So it is important to consider the full pedigree information while estimating the breeding values. In the current study breeding values are estimated from the pedigree and the phenotypic information, but in reality the pedigree information is often incomplete in such case the genetic similarities calculated based on the molecular data can be used for the estimation of breeding values (Bauer *et al.* 2006). It is possible to modify the new adaptive MCMC algorithm to account such information in order to estimate the breeding values.

4.3 Inbreeding and the genetic complexity

From De Boer and Hoeschele (1993) it is known that inbreeding and non additive genetic actions complicates the genetic covariance structure of a population. If someone wants to include inbreeding while estimating breeding values it is necessary to account the covariance between the additive and dominance effect in a inbred population into the model. However I considered a simulated data set with inbreeding to estimated the breeding values using the new adaptive MCMC method without accounting the covariance between the additive and dominance effect. So there is a possibility that the REML and adaptive MCMC methods may provide biased estimates for such datasets with inbreeding. Another type of a model that would suit well to the new estimation

framework is a random effect like genotype-by-environment interaction (Bauer *et al.* 2009) or a Gaussian process model (Crossa *et al.* 2010) instead of dominance effects. Then the dominance relationship matrix would be replaced by the genotype-by-environment interaction covariance matrix or by the covariance function proportional to the evaluations of a reproducing kernel evaluated in marker genotypes. However studies should be done how to include the covariance between the additive and dominance effect in the adaptive MCMC framework for the Bayesian analysis.

4.4 Estimation of Variance components

Finding the best estimation method for the variance components is a primary concern for animal breeders (Lee, 2000). Mixed linear model using REML is widely used in animal breeding During the test runs the scatter plot for the variance components from the hybrid Gibbs sampler showed correlation between dominance and error variance components. This correlation was the main reason to think about using block update for the variance components and adaptive MCMC methods was an effective solution for such problems. Adaptive MCMC methods can learn the optimal covariance structure for the block update of the variance components from the history of the chain (Haario *et al.* 2001; Roberts and Rosenthal 2007). Malve *et al.* (2007) applied adaptive MCMC methods for the Bayesian modeling of algal mass occurrences in the northern hemisphere. In their study they showed that when the parameters are highly correlated the problem can dealt with adaptive schemes. Results from the current study showed that the block update of the variance components improved the mixing properties of the MCMC chain. Also the new method was able to provide better estimates for the variance components than the existing methods like hybrid Gibbs sampler and REML.

Du and Hoeschele (2000) showed that accurate estimation of dominance is important for efficient section strategies. Also Wall et al (2005) showed that non-additive genetic effects play an important role on the ranking of breeding values. Normally large data set is needed for the accurate estimation of non-additive dominance variance (Misztal 1997). In the current study when I did test runs with small datasets there was over estimation of additive variance so I decided to use large data sets for the study. Waldmann *et al.* (2008) showed that the hybrid Gibb sampler with uninformative prior for the dataset with low dominance resulted in considerable over estimation of the dominance variance. But the adaptive MCMC method using non-informative prior was able to return zero dominance with datasets with no dominance. This shows that the adaptive MCMC approach is able to modal selection. For the simulated data set with finite number of loci the estimated variance components were depended on the number of loci simulated.

4.5 Identifiability problem

Identifiability problems arise especially when the dominance relationship matrix **D** is nearly a multiple of the identity matrix. Then certain features of the phenotypic observations can almost as well be attributed to dominance effects as to noise. In such a case the joint marginal posterior of the dominance variance and the error variance should be bimodal. In such a case conventional MCMC samplers may have difficulties moving between the modes. Especially Gibbs samplers can have difficulties to escape from one such mode, but Metropolis-Hastings sampling schemes may behave better. Adding more full-sibs to the pedigree file can improve the multimodality problem to some extent. When only few fulls-sibs are considered the dominance relationship matrix is nearly a multiple of identity matrix. Moreover the error variance

matrix is also an identity matrix. Gelfand and Sahu (1999) and Sorensen and Gianola (2002) suggested that using informative priors can alleviate the identifiability problem. However when I used different informative priors in the analysis there was over estimation of dominance variance for the dataset with zero dominance (QTLMAS dataset).

Conventional MCMC algorithms normal fails to sample from multi-modal target distributions because they only propose small moves, hence the move between different modes become rare and convergence of the chain will slow down. Marinari & Parisi (1992) and Geyer & Thompson (1995) proposed simulated tempering to deal with multi-modal distributions. However in the adaptive MCMC method the algorithm automatically "learn" the proposal covariance matrix from the history of the chain, this helps the algorithm to sample efficiently from multimodal target distributions. The main reason for the identifiability in the current study was the presence of multiple modes in the posterior distribution of the dominance and the error variance components. But in the simulation experiments, the adaptive MCMC algorithm was able to explore the entire parameter space with good mixing properties, and therefore was able to detect different modes in the posterior distribution. In practice non-additive effects are susceptible to identifiability problem. Du and Hoeschele (2000) proposed a finite-locus approximation to infinite-locus modal for the estimation of non-additive parameters, but the estimates are depend on the number of loci used.

4.6 Importance of the optimal proposal covariance structure

From the study it was clear that the proposal covariance matrix play a crucial role in the mixing properties and the acceptance ration of the algorithm.

The proposal covariance matrix $(2.38)^2/d\,\mathbf{S}$ from Roberts *et al.* (1997) and Roberts and Rosenthal (2001) is optimal in a large-dimensional context when the posterior is approximately Gaussian (Roberts and Rosenthal, 2007). In the present study I used different scalings of the posterior covariance matrix, in some cases the acceptance ratio was high and the estimates were bad and in some other cases the acceptance ratio was too low. This scaling factor $(2.38)^2/d$ was also employed in the MCMC sampler of Fang *et al.* (2011), who introduced a new method for QTL mapping. In their sampling scheme they utilized REML estimates in construction of the proposal covariance matrix. If the target distribution is multimodal this approach may fail to move between different modes. In contrast our new adaptive MCMC method use the previous history of the chain to learn the optimal proposal covariance matrix, which enables the algorithm to move between different modes. The theoretical formula turned out to work well enough for my applications with optimal acceptance ratio between 20 % and 50%. The success of adaptive MCMC methods generally depends on how well the proposal covariance structure is learned from the previous history of the chain. Therefore it is important to use a sufficient number of samples in the learning period. In the present study I tried different length of burn-in period and learning period to get an optimal covariance structure. The required sample depends firstly on the dimensionality and on the other characteristics of the posterior distribution and secondly on the mixing properties of the MCMC sampler. Therefore it is impossible to give general prescriptions for it.

4.7 Effects of marginalization and Mixing

In spite of increasing computing power, poor mixing is still one of the main problems with MCMC methods. Poor mixing arises usually due to the high

posterior correlations between parameters. By reducing the autocorrelation between the parameters the mixing can be improved. In the present study it was also tested adaptation in a version of a model where the random effects were not marginalized away (class 2 adaptive MCMC). However, this formulation suffered from poor mixing and slow convergence because of posterior dependencies among the random effects and the variance components. The marginalized model (*i.e.*, hierarchical model 2) which was used in the current study was able to explore the entire parameter space with good mixing properties. The Effective Sample Size (ESS) of a parameter is the number of independent samples from the posterior distribution which the correlated MCMC sample is worth. The ESS of a parameter is one of the commonly used method to assess MCMC mixing (e.g., Carlin & Louis 2000, Chen *et al.* 2000). If the ESSs are low, then the autocorrelations will be high, and that may be an indication of poor mixing of the chain. The adaptive scheme was able to decrease the autocorrelation of the chain to yield much larger effective sample sizes. From the ESSs for the class 1 and class 2 adaptive MCMC algorithms, class 1 showed better mixing properties. Also class 1 adaptive MCMC was able to provided better estimates for the variance components. So marginalization plays a crucial role on the overall performance of the algorithm. Whereas REML can be characterized so that one assumes a uniform distribution for the fixed effects, then integrates the fixed effects and random effects out, and finally maximizes with respect unknown parameters. Moreover convergence of the general Bayesian Gibbs sampling algorithms, which use single-site updates for the variance components can be slow due to posterior dependencies. However in the new proposed method which use the block update of the variance components will help Markov chains to converge reasonably fast to its equilibrium distribution.

4.8 Impact of Prior Distributions and sensitivity analysis

The choice of the prior is one of the important steps in any Bayesian analysis. Generally, the influence of the prior distribution on the posterior is related to the sample size of the data. Zeller *et al.* (1971) gave the framework for two different classes of prior information one based on the data and the other one non-data based. In the first class the prior incorporate information from the previous studies whereas, in the second class the prior information is the result of theoretical consideration. In the present study informative and non-informative priors was use to see the impact of prior distribution on variance component estimation. A sensitivity analysis was carried out using different priors with different degree of belief and most of them seemed to lead to non-zero estimates of dominance variance for the QTLMAS data. However, Gamma$(1, 0.001)$ prior for the precision parameters was able to provide good mixing, while still leading to a realistic estimate of dominance variance in the case of no dominance. The flat prior (Gamma$(1, 0.001)$) was able to give values close to REML estimates. This follows because the prior can then shrink the posterior towards zero. Morita *et al.* (2007) has suggested while fitting a Bayesian model to a data set of 10 observations, an a priori ESS of 1 is reasonable, whereas a prior ESS of 20 implies that the prior dominates posterior inferences.

5 Summary and Conclusion

Breeding value describes the genetic merit of an individual and it is calculated as the deviations from the mean values of the population. The estimated breeding value can improve the selection of favorable parental line. However to ensure the accurate estimation of breeding values it is of great importance to calculate the true genetic variance parameters in the population. But the accurate estimation of them is often difficult because of the complexity in the underlying covariance structure.

REML and Bayesian methods are commonly used for the prediction of breeding values. In the present study I proposed a new fast adaptive MCMC algorithm for the estimation of variance components. The Adaptive MCMC algorithm combines both hybrid Gibbs sampling and M-H algorithm to calculate the breeding values and variance components. In this new approach, the adaptive MCMC runs in two stages. First, the algorithm runs to obtain the empiric estimate for the posterior covariance structure of variance components ,this part of the MCMC is called learning period. Then utilize this estimated covariance structure in the second stage to generate multivariate correlated proposals for variance components in random walk M-H algorithm, in order to improve the mixing properties of the chain. In the second phase of the algorithm the random effects were marginalized from the likelihood. The new proposed algorithm was able to provide better estimates than the existing methods like REML and Gibbs sampling. Moreover the new proposed algorithm was able to detect different modes in the posterior distribution. Additionally, the new proposed exponential prior for variance components was able to provide the estimated mode of the posterior dominance variance to be zero in case of no dominance.

In the current study breeding values are estimated from the pedigree and

the phenotypic information, but in reality the pedigree information is often incomplete. In such a case the genetic similarities calculated, based on the molecular data, can be used for the estimation of breeding values. However it is possible to modify the new adaptive MCMC algorithm to account such information in order to estimate the breeding values. Also it is important to study how to account the covariance between the additive and dominance effect in an inbred population while estimating genetic parameters.

6 References

BAUER, A. M., F. HOTI, T. C. REETZ, W.-D. SCHUH, J. LÉON and M. J. SILLANPÄA, 2009 Bayesian prediction of breeding values by accounting for genotype-by-environment interaction in self-pollinating crops. Genet. Res. 91: 193-207.

BAUER, A. M., T. C. REETZ and J. LÉON, 2006 Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. Crop Sci. 46: 2685-2691.

BAUER, A. M. and J. LÉON, 2008 Multiple-trait breeding values for parental selection in self-pollinating crops. Theor. Appl. Genet. 116: 235-242.

BELONSKY, G. M., and B. W. KENNEDY, 1988 Selection on individual phenotype and best linear unbiased predictor of breeding value in a closed swine nucleus. J. Anim. Sci. 66: 1124-1131.

CASELLA, G. and E. I. GEORGE, 1992 Explaining the Gibbs sampler. AmericanStatistician 46: 167-174.

CARLIN, B. P. and T. A. LOUIS, 2000 Bayes and Empirical Bayes Methods for Data Analysis. 2nd edition. Chapman & Hall Ltd.

CHEN, L., Z. QIN and J. LIU, 2000 Exploring Hybrid Monte Carlo in Bayesian Computation. ISBA 2000 Proceedings.

CHIB, S. and E. GREENBERG, 1995 Understanding the Metropolis-Hastings algorithm. American Statistician 49: 327-335.

CROSSA, J., J. BURGUENO, P. L. CORNELIUS, G. MCLAREN, R. TRETHOWAN, and A. KRISHNAMACHARI, 2006 Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci. 46: 1722-1733.

CROSSA, J., G. DE LOS CAMPOS, P. PÉREZ, D. GIANOLA, J. BURGUEÑO *et al.,* 2010 Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186: 713-724

DE BOER, I. J. M. and I. HOESCHELE, 1993 Genetic evaluation methods for populations with dominance and inbreeding. Theor. Appl. Genet. 86: 245-258.

DE SOUZA, V. A. B., D. H. BYRNE and J. F. TAYLOR, 2000 Predicted breeding values for nine plant and fruit characteristics of 28 peach genotypes. J. Am. Soc. Hort. Sci. 125: 460-465.

DU, F. X., I. HOESCHELE and K. M. GAGE-LAHTI, 1999 Estimation of additive and dominance variance components in finite polygenic models and complex pedigrees. Genetical Research 74: 179-187.

FALCONER, D. S., 1989 Introduction to Quantitative Genetics, Ed. 3. Longmans Green/John Wiley & Sons, Harlow, Essex, UK/New York.

FANG, M., J. LIU, D. SUN, Y. ZHANG, Q. ZHANG, Y. ZHANG and S. ZHANG, 2011 QTL mapping in outbred half-sib families using Bayesian model selection. Heredity 107: 265-276.

GARCIA-CORTES, L. A. and D. SORENSEN, 1996 On a multivariate implementation of the Gibbs sampler. Genet. Sel. Evol. 28: 121-126.

GASBARRA, D., M. PIRINEN, S., KULATHINAL and M. J. SILLANPÄÄ, 2011 Estimating haplotype frequencies by combining data from large DNA pools with database information. IEEE/ACM Transactions on Computational Biology and Bioinformatics 8: 36-44

GELMAN, A., 2006 Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 1: 515-534.

GELFAND, A. E. and S. K. SAHU, 1999 Identifiability, improper priors, and Gibbs sampling for generalized linear models. Journal of the American Statistical Association. 94: 247-253.

GEYER, C. J., 1992 Practical Markov chain Monte Carlo (with discussion). Stat. Sci.7: 473-511.

GEYER, C. J. and E. A. THOMPSON, 1992 Constrained Monte Carlo Maximum Likelihood for Dependent Data. Journal of the RoyalStatistical Society. 54: 657-699.

GILMOUR, A. R., B. J. GOGEL, B. R. CULLIS and R. THOMPSON, 2006 ASReml User Guide, Release 2.0. VSN International, Hemel Hempstead, UK

HAARIO, H., E. SAKSMAN and J. TAMMINEN, 2001 An adaptive Metropolis algorithm Bernoulli 7: 223-242.

HALLANDER, J., P. WALDMANN, C. WANG and M. J. SILLANPÄÄ, 2010 Bayesian inference of genetic parameters based on conditional decompositions of multivariate normal distributions. Genetics 185: 645-654.

HANSON, W. D., 1963 Heritability. In Statistical Genetics And Plant Breed-

ing (ed. W. D. HANSON & H. F. ROBINSON), pp. 125-139. Washington, DC : National Academy of Sciences-National Research Council.

HASTINGS, W. K., 1970 Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57: 97-109.

HENDERSON, C. R., 1963 Selection index and the expected genetic advance, pp. 141-163 in Statistical Genetics and Plant Breeding, edited by W. D. HANSON and H. F. ROBINSON, National Academy of Science, National Research Council Publ. No. 982, Washington, DC.

HENDERSON, C. R., 1975 Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423-447.

HENDERSON, C. R., 1976 A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32: 69-83.

HENDERSON, C. R., 1985a Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J. Anim. Sci. 60: 111-117.

HENDERSON, C. R., 1985b MIVQUE and REML estimation of additive and nonadditive genetic variances. J. Anim. Sci. 61: 113-121.

HOESCHELE, I. and P. M. VANRADEN, 1991 Rapid inverse of dominance relationships matrices for noninbred populations by including sire by dam subclass effects. Journal of Dairy Science 74: 557-569.

HOMER, T. W. and C. R. WEBER, 1956 Theoretical and experimental study of self fertilized populations. Biometrics 12: 404-414.

HOTI, F. J., M. J. SILLANPÄA and L. HOLMSTROM, 2002 A note on estimating the posterior density of a quantitative trait locus from a Markov chain Monte Carlo sample. Genet. Epidemiol. 22: 369-376.

HSU, F. C., D. J. ZACCARO, L. A. LANGE, D. K. ARNETT, C. D. LANGEFELD, L. E. WAGENKNECHT, D. M. HERRINGTON, S. R. BECK, B. I. FREEDMAN, D. W. BOWDEN and S. S. RICH, 2005 The impact of pedigree structure on heritability estimates for pulse pressure in three studies. Hum. Hered. 60: 63-72.

LAPLACE, P. S., 1812 Thèorie analytique des probabilitès. Paris, Veuve Courcier.

LEE, C., 2000 Methods and techniques for variance componentestimation in animal breeding-review. Asian-Aust. J. Anim. Sci. 13(3): 413-422.

LUND, M. S., G. SAHANA, DJ. DE KONING, G. SU and Ö. CARLBORG, 2009 Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. BMC Proceedings, 3(Suppl 1): S1.

LYNCH, M. and B. WALSH, 1998 Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA.

MARINARI, E. and G. PARISI, 1992 Simulated tempering: a new Monte Carlo scheme. Europhys. Lett. 19: 451-458.

MATLAB, 2007 High-performance numeric computation and visualization software, vol 7. The Math Works Inc, Natick, Mass, USA.

MEYER, K. and M. KIRKPATRICK, 2010 Better estimates of genetic covariance matrices by "bending" using penalized maximum likelihood. Genetics 185: 1097-1110.

MISZTAL, I., 1997 Estimation of variance components with large-scale dominance models. J. Dairy Sci. 80: 965-974

MORITA, S., P. F. THALL, and P. MÜLLER, 2008 Determining the effective sample size of a parametric prior. Biometrics 64: 595-602

OAKEY, H., A. VERBYLA, W. PITCHFORD, B. CULLIS and H. KUCHEL, 2006 Joint modelling of additive and non-additive genetic line effects in single field trials. Theor. Appl. Genet. 113: 809-819.

O'HARA, B. R. and M. J. SILLANPÄÄ, 2009 Review of Bayesian variable selection methods: what, how and which. Bayesian Analysis 4: 85-118.

MALVEA, O., M. LAINEB, H. HAARIOC, T. KIRKKALAD and J. SARVALAE, 2007 Bayesian modelling of algal mass occurrences-using adaptive MCMC methods with a lake water quality model. Environmental Modelling and Software 22: 966-977.

PANTER, D. M. and F. L. ALLEN, 1995 Using best linear unbiased predictions to enhance breeding for yield in soybean: I. Choosing parents. Crop Sci. 35: 397-405.

PATTERSON, H.D. and R. Thompson, 1971 Recovery of interblock information when block sizes are unequal. Biometrika 58: 545-554.

PATTEE, H. E., T. G. ISLEIB, D. W. GORBET, F. G. GIESBRECHT, and Z. CUI, 2001 Parent selection in breeding for roasted peanut flavour quality. Peanut Sci. 28: 51-58.

PIEPHO, H. P., J. MOHRING, A. E. MELCHINGER and A. BUCHSE, 2008 BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161: 209-228.

PLUMMER, M., N. BEST, K. COWLES and K. VINES, 2006 coda: Convergence Diagnosis and Output Analysis for MCMC.R News 6(1): 7-11. http://CRAN.R-project.org/doc/Rnews/.

QUAAS, R. L., 1976 Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32: 949-953.

ROBERTS, G. O., A. GELMAN and W.R. GILKS, 1997 Weak convergence and optimal scaling of random walk Metropolis algorithms. Ann. Appl. Prob. 7: 110-120.

ROBERTS, G. O. and J. S. ROSENTHAL, 2007 Coupling and ergodicity of adaptive MCMC. J. Appl. Probab. 44: 458-475.

ROBERTS, G. O. and J. S. ROSENTHAL, 2001 Optimal scaling for various Metropolis- Hastings algorithms. Stat. Sci. 16: 351-367.

ROBERT, C. P. and G. CASELLA, 2004 Monte Carlo Statistical Methods (second edition). New York: Springer-Verlag.

SORENSEN, D. and D. GIANOLA, 2002 Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Springer-Verlag, New York.

WAAGEPETERSEN R., N. IBÁN˜EZ-ESCRICHE and D. SORENSEN, 2008 A comparison of strategies for Markov Chain Monte Carlo computation in quantitative genetics. Genet. Sel. Evol. 40: 161-176.

WALDMANN, P., J. HALLANDER, F. HOTI and M. J. SILLANPÄA, 2008 Efficient MCMC implementation of Bayesian analysis of additive and dominance genetic variances in non-inbred pedigrees. Genetics 179: 1101-1112.

WALL, E., S. BROTHERSTONE, J. F. KEARNEY, J. A. WOOLLIAMS and M. P. COFFEY, 2005 Impact of nonadditive genetic effects in the estimation of breeding values for fertility and correlated traits. J. Dairy Sci. 88: 376-385.

WANG, C. S., J. J. RUTLEDGE and D. GIANOLA, 1993 Marginal inference about variance components in a mixed linear model using Gibbs sampling. Genet. Sel. Evol. 21: 41-62.

ZELLNER, A., 1971 An Introduction to Bayesian Inference in Econometrics. J. Wiley and Sons, New York.