GSE Genetics Selection Evolution

RESEARCH   **Open Access**

Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models

Han A Mulder[1,2*], Lars Rönnegård[3,4], W Freddy Fikse[4], Roel F Veerkamp[1] and Erling Strandberg[4]

Abstract

Background: Genetic variation for environmental sensitivity indicates that animals are genetically different in their response to environmental factors. Environmental factors are either identifiable (e.g. temperature) and called macro-environmental or unknown and called micro-environmental. The objectives of this study were to develop a statistical method to estimate genetic parameters for macro- and micro-environmental sensitivities simultaneously, to investigate bias and precision of resulting estimates of genetic parameters and to develop and evaluate use of Akaike's information criterion using h-likelihood to select the best fitting model.

Methods: We assumed that genetic variation in macro- and micro-environmental sensitivities is expressed as genetic variance in the slope of a linear reaction norm and environmental variance, respectively. A reaction norm model to estimate genetic variance for macro-environmental sensitivity was combined with a structural model for residual variance to estimate genetic variance for micro-environmental sensitivity using a double hierarchical generalized linear model in ASReml. Akaike's information criterion was constructed as model selection criterion using approximated h-likelihood. Populations of sires with large half-sib offspring groups were simulated to investigate bias and precision of estimated genetic parameters.

Results: Designs with 100 sires, each with at least 100 offspring, are required to have standard deviations of estimated variances lower than 50% of the true value. When the number of offspring increased, standard deviations of estimates across replicates decreased substantially, especially for genetic variances of macro- and micro-environmental sensitivities. Standard deviations of estimated genetic correlations across replicates were quite large (between 0.1 and 0.4), especially when sires had few offspring. Practically, no bias was observed for estimates of any of the parameters. Using Akaike's information criterion the true genetic model was selected as the best statistical model in at least 90% of 100 replicates when the number of offspring per sire was 100. Application of the model to lactation milk yield in dairy cattle showed that genetic variance for micro- and macro-environmental sensitivities existed.

Conclusion: The algorithm and model selection criterion presented here can contribute to better understand genetic control of macro- and micro-environmental sensitivities. Designs or datasets should have at least 100 sires each with 100 offspring.

* Correspondence: han.mulder@wur.nl
[1]Wageningen UR Livestock Research, Animal Breeding and Genomics Centre, PO Box 65, 8200 AB, Lelystad, The Netherlands
[2]Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH, Wageningen, The Netherlands
Full list of author information is available at the end of the article

Background

The term "genotype by environment (G × E) interaction" refers to the fact that the best genotype in one environment may not be the best genotype in another environment [1] and that genotypes differ in their response to environmental factors, which means that genetic variance for environmental sensitivity or phenotypic plasticity exists [2]. Some environmental factors (e.g., temperature, soil, diet, etc.) are identifiable and can be categorised (e.g., temperate or tropical climate) or quantified (e.g., temperature) and thus are referred to as macro-environmental factors. Other environmental factors are unknown and referred to as micro-environmental factors [1]. Therefore, genetic variance in macro-environmental sensitivity is the genetic variance due to known environmental factors and can be expressed as the genetic variance in the slope of a reaction norm when environments can be quantified on a continuous scale. If environments are categorized, then phenotypes in different environments are considered as separate traits and the genetic covariances between environments are a measure of genetic variation in macro-environmental sensitivity. Genetic variance in micro-environmental sensitivity is the genetic variance due to unknown environmental factors and can be expressed as differences in environmental variance, sometimes called genetic heterogeneity of environmental variance [3].

Numerous studies in the last 70 years have studied G × E interactions or genetics of macro-environmental sensitivity in animal and plant breeding as well as in evolutionary biology [4-6]. Different modeling approaches have been used [4]. In recent years, reaction norm models have been applied in animal breeding to better understand the environmental factors that determine G × E interactions [7-9]. In evolutionary genetics, many experiments on Drosophila and other laboratory species have been carried out to understand the genetics of macro-environmental sensitivity or phenotypic plasticity [6,10,11]. Other examples of studies on wild life populations include analyses of great tit [12] and butterfly [13] populations that showed the existence of genetic variation in phenotypic plasticity to temperature change.

The genetics of micro-environmental sensitivity or environmental variance have been studied less extensively than the genetics of macro-environmental sensitivity. In evolutionary genetics, several studies have focused on canalization, i.e. selection for reduced variance [14-16]. Recently, there has been renewed interest on this topic due to the development of methods to estimate genetic variance in environmental variance, e.g. Bayesian methods [17,18] and double hierarchical generalized linear models (DHGLM) in a REML setting [19]. Hill and Mulder [20] reviewed 14 studies on this subject and concluded that there is empirical evidence for the existence of genetic variance in environmental variance.

Although there is substantial evidence for genetic variance in macro- and micro-environmental sensitivities, very few studies have studied them together or studied their genetic relationship. Jinks and Pooni [21] postulated the concepts of macro- and micro-environmental sensitivity and studied both in *Nicotiana rustica* but did not investigate the relationships between these two types of environmental sensitivity. However, Perkins and Jinks [22] reported that both types of environmental sensitivity were weakly genetically correlated for most traits in *Nicotiana rustica*. In a study on *Populus*, Wu [23] showed that micro-environmental sensitivity was less heritable than macro-environmental sensitivity and that they were weakly genetically correlated. In another paper, also on *Populus*, Wu and O'Malley [24] detected different sets of genes for macro- and micro-environmental sensitivities. In a mapping study of QTL for macro- and micro-environmental sensitivities in barley, Kraakman et al. [25] identified only a few QTL for micro-environmental sensitivity, but did not detect any QTL for macro-environmental sensitivity. Yampolsky and Scheiner [26] observed a weak genetic correlation between developmental noise, similar to micro-environmental sensitivity, and macro-environmental sensitivity in *Daphnia Magna*. In *Drosophila melanogaster*, both positive and negative genetic correlations between macro- and micro-environmental sensitivities have been reported [27,28]. In farm animals, the genetic correlation between macro- and micro-environmental sensitivities has not been investigated. In addition, to date, no suitable method was available to estimate such correlations in outbred populations and the requirements in terms of number of families and family sizes of the designs necessary to estimate genetic parameters of macro- and micro-environmental sensitivities were unknown.

The objectives of this study were to extend the double hierarchical generalized linear model (DHGLM) of Rönnegård et al. [19] with a reaction norm model to estimate genetic variance for macro- and micro-environmental sensitivities and to investigate bias and precision of estimated variance components in designs resembling dairy cattle populations. In addition, we evaluated Akaike's information criterion (AIC) using approximated h-likelihood to select the best fitting model and studied situations in which true genetic and best fitting statistical models differed. Finally, we applied the model to lactation milk yield in dairy cattle.

Methods

In this section, we describe the assumed quantitative genetic model underlying genetic variance for macro- and micro-environmental sensitivities, the statistical model to

estimate genetic variance in macro- and micro-environmental sensitivities, and the AIC based on an approximation of the h-likelihood that can be used to select the best fitting model. Finally, we present the simulation used to test the statistical model and describe the evaluated scenarios as well as an application to milk yield in dairy cattle.

Quantitative genetic model

Here, we assume that genetic variance for macro-environmental sensitivity is expressed as the genetic variation in the slope of a linear reaction norm. Genetic variance in micro-environmental sensitivity implies that genotypes respond differently to one or several unknown environmental factors and thus, we assume that it is expressed as genetic variance in environmental variance according to an exponential model [29]. Thus, the quantitative genetic model underlying genetic variance in macro- and micro-environmental sensitivities can be formulated as:

$$
\begin{aligned}
P = {} & \mu + A_{int} + A_{sl}x \\
& + \exp\left(0.5\ln(\sigma_E^2) + 0.5A_v\right)\varepsilon
\end{aligned}
\tag{1}
$$

where P is the phenotype, μ is the population mean for the phenotype, x is the environmental parameter (continuous or discrete) that cause genotypes to respond differently, A_{int}, A_{sl} and A_v are the additive genetic or breeding values for the intercept, for the slope of the reaction norm (= macro-environmental sensitivity), and for the environmental variance (= micro-environmental sensitivity), respectively, σ_E^2 is the environmental variance of the exponential model and ε is a scaled environmental deviation with variance one. Note that the average environmental variance is $\overline{\sigma_E^2} = \sigma_E^2\exp\left(0.5\sigma_{A_v}^2\right)$ [3]. The additive genetic values A_{int}, A_{sl} and A_v have variance

$$
\mathbf{G} =
\begin{bmatrix}
\sigma_{A_{int}}^2 & \sigma_{A_{int},A_{sl}} & \sigma_{A_{int},A_v} \\
& \sigma_{A_{sl}}^2 & \sigma_{A_{sl},A_v} \\
& & \sigma_{A_v}^2
\end{bmatrix}
$$

Statistical model

The basis of the statistical model is the DHLGM as presented by Rönnegård et al. [19]. The original DHGLM algorithm in Rönnegård et al. [19] iterates between a linear mixed model for the phenotypic observations: $y_i = P_i =$ phenotypic observation of animal i; and a Gamma GLM for the residual variance ϕ_i, where $\phi_i = \mathrm{E}\left(\frac{\hat{e}_i^2}{1-h_i}\right)$, $\hat{e}_i^2$ is the squared estimated residual of y_i and h_i is the diagonal element of the hat-matrix of $\mathbf{y}$ corresponding to y_i [30]. Felleki et al. [31] extended the model of [19] by allowing

estimation of the genetic covariance (σ_{a,a_v}) between the genetic effect in the mean model and the genetic effect in the variance model (Equation (2)) in the absence of a linear reaction norm model and reported on computational details and a theoretical assessment of this algorithm. In brief, equivalent to using a Gamma GLM, linearizing $\frac{\hat{e}_i^2}{1-h_i}$ around its current fitted value results in vector $\boldsymbol{\psi}$ and yields the following bivariate linear model:

$$
\begin{bmatrix} \mathbf{y} \\ \boldsymbol{\psi} \end{bmatrix} =
\begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X_v} \end{bmatrix}
\begin{bmatrix} \mathbf{b} \\ \mathbf{b_v} \end{bmatrix} +
\begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z_v} \end{bmatrix}
\begin{bmatrix} \mathbf{a} \\ \mathbf{a_v} \end{bmatrix} +
\begin{bmatrix} \mathbf{e} \\ \mathbf{e_v} \end{bmatrix}
\tag{2}
$$

where $\mathbf{X}(\mathbf{X_v})$ is the incidence matrix for fixed effects for $\mathbf{y}$ ($\boldsymbol{\psi}$), $\mathbf{b}$ ($\mathbf{b_v}$) is the vector with solutions for fixed effects for $\mathbf{y}(\boldsymbol{\psi})$, $\mathbf{Z}(\mathbf{Z_v})$ is the incidence matrix for relating observations of $\mathbf{y}(\boldsymbol{\psi})$ to the additive genetic values $\mathbf{a}$ ($\mathbf{a_v}$) for phenotype (environmental variance) $\begin{bmatrix} \mathbf{a} \\ \mathbf{a_v} \end{bmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_a^2 & \sigma_{a,a_v} \\ \sigma_{a,a_v} & \sigma_{a_v}^2 \end{bmatrix} \otimes \mathbf{A}\right)$, where $\mathbf{A}$ is the additive genetic or numerator relationship matrix and σ_a^2 $\left(\sigma_{a_v}^2\right)$ is the genetic variance of $\mathbf{a}(\mathbf{a_v})$. The residuals of $\mathbf{y}$ ($\mathbf{e}$) and $\boldsymbol{\psi}$ ($\mathbf{e_v}$) are assumed to be independent normally distributed $\begin{bmatrix} \mathbf{e} \\ \mathbf{e_v} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{W}^{-1}\sigma_\epsilon^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{W_v}^{-1}\sigma_{\epsilon_v}^2 \end{bmatrix}\right)$, $\mathbf{W} = \mathrm{diag}(\hat{\boldsymbol{\psi}})^{-1}$ and $\mathbf{W_v} = \mathrm{diag}(\frac{1-\mathbf{h}}{2})$, σ_ϵ^2 and $\sigma_{\epsilon_v}^2$ are scaling variances, which are expected to be equal to 1, because $\mathbf{W}$ and $\mathbf{W_v}$ already contain the reciprocals of the residual variances per observation. Estimating σ_ϵ^2 and $\sigma_{\epsilon_v}^2$ allows more flexibility. An iterative estimation procedure is required to obtain estimates for all parameters because of the dependence of the model for $\mathbf{y}$ on the results of the model for $\boldsymbol{\psi}$ and vice versa. The vector $\boldsymbol{\psi}$ and the diagonals of $\mathbf{W}$ and $\mathbf{W_v}$ are updated at each iteration until convergence. The model in Equation (2) can be considered as combining the DHGLM method [19] with the iterative method of Mulder et al. [32].

In order to consider macro- and micro-environmental sensitivity simultaneously, the model for phenotype $\mathbf{y}$ in Equation (2) was extended with a linear reaction norm model to estimate genetic variance in macro-environmental sensitivity, whereas the residual variance model using $\boldsymbol{\psi}$ estimates the genetic variance in micro-environmental sensitivity or environmental variance. Instead of using an animal model, we used a sire model because implementation of the animal model DHGLM resulted in severely biased estimates of variance components with single observations per animal. As an indication of poor model fit, implementation of the animal model DHGLM gave a lower adjusted profile h-likelihood (see next section "Model selection") than an animal model without heterogeneity of residual variance. Sire models are commonly

used in combination with reaction norm models [7,8,33]. In the reaction norm model, we assumed that the environmental parameter x (see Equation (1)) is known without error and does not need to be estimated from the data. The resulting combined macro–micro environmental sensitivity sire model can be formulated as:

$$\begin{bmatrix} \mathbf{y} \\ \boldsymbol{\psi}_\mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_\mathbf{v} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{b}_\mathbf{v} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_\mathbf{s} & \mathbf{Z}_\mathbf{x} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_\mathbf{sv} \end{bmatrix}$$
$$\times \begin{bmatrix} \mathbf{s}_\mathbf{int} \\ \mathbf{s}_\mathbf{sl} \\ \mathbf{s}_\mathbf{v} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_\mathbf{s} \\ \mathbf{e}_\mathbf{sv} \end{bmatrix} \qquad (3)$$

where $\boldsymbol{\psi}_\mathbf{s}$ is a vector with linearized values of transformed squared residuals (see calculation in Appendix), $\mathbf{Z}_\mathbf{s}$ and $\mathbf{Z}_\mathbf{sv}$ are respectively the incidence matrices for the sire effects for the intercept of the reaction norm and for the environmental variance, $\mathbf{Z}_\mathbf{x}$ is the matrix with the environmental parameter x as a covariate for the sire effects for the slope of the reaction norm, and $\mathbf{s}_\mathbf{int}$, $\mathbf{s}_\mathbf{sl}$ and $\mathbf{s}_\mathbf{v}$ are the vectors with the estimated sire effects for intercept, slope, and environmental variance. The sire effects $\mathbf{s}_\mathbf{int}$, $\mathbf{s}_\mathbf{sl}$, and $\mathbf{s}_\mathbf{v}$ are assumed trivariate normally distributed $N\left(\mathbf{0}, \frac{1}{4}\mathbf{G} \otimes \mathbf{A}\right)$, assuming that sire (co)variances are a quarter of the additive genetic variance. The residuals of $\mathbf{y}$ ($\mathbf{e}_\mathbf{s}$) and $\boldsymbol{\psi}_\mathbf{s}$($\mathbf{e}_\mathbf{sv}$) are assumed to be independent normally distributed, because $\text{cov}(e, e^2) = 0$, when e is normally distributed ($\begin{bmatrix} \mathbf{e}_\mathbf{s} \\ \mathbf{e}_\mathbf{sv} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{W}_\mathbf{s}^{-1}\sigma^2_{\epsilon_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_\mathbf{sv}^{-1}\sigma^2_{\epsilon_{sv}} \end{bmatrix}\right)$), where $\sigma^2_{\epsilon_s}$ and $\sigma^2_{\epsilon_{sv}}$ are scaling variances for the residual variances in a sire model. Computations of $\boldsymbol{\psi}_\mathbf{s}$ and of the diagonals of $\mathbf{W}_\mathbf{s}$ and $\mathbf{W}_\mathbf{sv}$ are different due to the use of a sire model compared to an animal model and are explained in the Appendix. The algorithm was implemented by iterating over several runs of ASReml [34]. In each ASReml run, REML-estimates of the variance components were obtained for the current values of $\boldsymbol{\psi}_\mathbf{s}$, $\mathbf{W}_\mathbf{s}$ and $\mathbf{W}_\mathbf{sv}$. The vector $\boldsymbol{\psi}_\mathbf{s}$ and the diagonals of $\mathbf{W}_\mathbf{s}$ and $\mathbf{W}_\mathbf{sv}$ were updated after each ASReml run. On simulated data, 10 ASReml runs were sufficient to obtain converged parameter estimates. ASReml was run with an option to check whether the variance-covariance matrices were positive definite and were forced to be positive definite [34] if they were not.

Model selection

When working with real data, the true genetic model is not known, thus statistical inference can be used to find the statistical model that fits the data best. The h-likelihood concept can be applied when using DHGLM [35], in which the adjusted profile h-likelihood (APHL) is used to assess the significance of variance components.

Because we use an iterative reweighted least square approximation in ASReml, the APHL can be approximated from the log REML likelihood (*logL*) for the bivariate model in Equation (3) fitted in ASReml, after correcting for the fact that the squared estimated residuals of $\mathbf{y}$ are used to compute the response variable $\boldsymbol{\psi}_\mathbf{s}$ (see Appendix for derivation):

$$\text{APHL} = -2logL - \Sigma w_{sv_i}\left(\hat{\sigma}^2_{\epsilon_{sv}}\right)^{-1} - \Sigma \ln\left(\hat{\sigma}^2_{\epsilon_{sv}}\right)/w_{sv_i}, \quad (4)$$

where w_{sv_i} is the weight for the variance model for observation i, i.e. the ith diagonal of $\mathbf{W}_\mathbf{sv}$. Note that Equation (4) can also be used for animal models by replacing elements w_{sv_i} and $\sigma^2_{\epsilon_{sv}}$ with the corresponding elements of Equation (2). To compare nested and non-nested models, we propose to use APHL in combination with Akaike's information criterion (AIC):

$$\text{AIC} = \text{APHL} + 2t, \qquad (5)$$

where t is the number of variance parameters.

Simulation

Monte Carlo simulation was used to evaluate the bias and precision of the estimated genetic parameters of the model and proposed algorithm. Populations of paternal half-sib families were simulated, resembling a simplified population structure in an animal breeding context with large half-sib offspring groups per sire like in dairy cattle. Either the number of sires (50, 100 or 200) was varied while keeping the number of offspring per sire at 100, or the number of offspring per sire (20, 50, 100 or 200) was varied while keeping the number of sires at 100. Sires and dams were assumed unrelated and were not selected. The offspring had phenotypes but the dams or sires did not. Phenotypes were generated for offspring according to the quantitative genetic model in Equation (1). Dams were randomly assigned to herds (herd size = 100 cows) and their offspring (daughters in the case of dairy cattle) were in the same herd. For each herd, the environmental covariate x was sampled from N(0,1). No fixed effects other than a general mean were simulated ($\mu = 0$) but offspring were randomly allocated to contemporary groups with size 10 to investigate the effect of fixed effects on the precision of genetic parameters. Additive genetic effects were sampled from $N(\mathbf{0}, \mathbf{G} \otimes \mathbf{A})$. Different values for genetic variances and genetic correlations were used for simulation. Table 1 gives an overview of the parameter values used in the simulation, both the default and alternative values. In each scenario, only one parameter was varied at a time, i.e. other parameters were kept at their default values. For each set of parameters, 100 replicates were generated and means and standard deviations of estimates were calculated across replicates.

Table 1 Default and alternative parameters values used in Monte Carlo simulation

Parameter	Default	Alternative values
$\sigma^2_{A_{int}}$	0.3	0.1, 0.5
$\sigma^2_{A_{sl}}$	0.05	0.025, 0.10
$\sigma^2_{A_v}$	0.1	0.05, 0.2
Genetic correlations (see text below Equation (1))	0	0.5
Number of offspring per sire	100	20, 50, 200
Number of sires	100	50, 200
Number of replicates	100	

$\sigma^2_{A_{int}}$ is the additive genetic variance for the intercept of the reaction norm; $\sigma^2_{A_{sl}}$ is the additive genetic variance for the slope of the reaction norm or macro-environmental sensitivity and $\sigma^2_{A_v}$ is the additive genetic variance for micro-environmental sensitivity or environmental variance.

Scenarios in which true genetic and statistical models differ

First, we tested situations in which the true genetic and statistical models contained both types of environmental sensitivities for different sets of genetic parameters and designs. In addition, we investigated the power of AIC to select the correct model given the true genetic model. Three situations were simulated with default parameters: presence of genetic variance in (i) macro- or (ii) micro-environmental sensitivity or in (iii) both. Four statistical models were applied to these simulated situations: a combined macro–micro environmental sensitivity model "macro–micro", a macro-environmental sensitivity model "macro", a micro-environmental sensitivity model "micro" and a simple model "simple" with only one additive genetic effect for the phenotype. The "macro–micro" model and the "micro" model were run for 10 ASReml runs to update the weights and the transformed squared residuals, whereas the "macro" model and the "simple" model required only one ASReml run. AIC (Equation 5) was used to select the best fitting model.

Finally, we tested scenarios in which the statistical model deviated from the true genetic model used for simulation, i.e. when model selection failed to select the right model (Table 2). These scenarios were analyzed to explore whether genetic parameters are biased when the statistical model deviates from the true genetic model and whether macro-environmental sensitivity can be detected with a micro-environmental sensitivity model

or vice versa. These scenarios were run with the default set of parameters as listed in Table 1 but with various values of $\sigma^2_{A_{sl}}$ and $\sigma^2_{A_v}$. Scenarios for which the true genetic and statistical models were the same and included either macro or micro-environmental sensitivity were not analyzed since these have been previously investigated, i.e. see [36] for macro-environmental sensitivity and [19] for micro-environmental sensitivity.

Application to milk yield in dairy cattle

The "macro–micro", "macro", "micro" and "simple" models were applied to 305-day lactation milk yield of 142 565 first parity Swedish Holstein cows. Lactation yields were calculated based on test-day data, as described in [37] using the test-day interval method [38]. Seasons were defined as three-month periods (January-March, April-June, July-September, October-December). Herd-year-season (HYS) classes with less than five cows were excluded. Data on a total of 762 sires with on average 187 daughters were available and 213 of these sires had more than 100 daughters. In the model for phenotype (**y**), HYS was a fixed effect and in the model for residual variance (**ψ$_s$**), HYS was a random effect to allow regression to the mean with few observations per HYS and avoid extreme HYS estimates. The model for **y** included a fixed fourth order polynomial for age at calving and a fixed ninth-order polynomial for lactation length, while the model for **ψ$_s$** included a third and a sixth order polynomial for age at calving and lactation length, respectively. The orders of polynomials were determined using a Wald test in univariate models for **y** and **ψ$_s$** and were kept the same for "macro–micro", "macro", "micro" and "simple" models. Herd-year average milk yield standardized to a mean of zero and a variance of one was used as an environmental parameter for macro-environmental sensitivity. The "macro–micro", "macro", "micro" and "simple" models were all run for 50 ASReml runs. After 50 runs, the change in variance components was less than 0.2% of the change between consecutive runs and therefore parameters were considered converged.

Results

Variance components and effect of fixed effects

Estimated genetic variances were close to their true values when simulated genetic correlations were zero, although

Table 2 Scenarios with different combinations of true genetic models and statistical models

True genetic model	Statistical model		
	Macro ES	**Micro ES**	**Macro and micro ES**
Macro ES	Not addressed	A	B
Micro ES	C	Not addressed	D
Macro and micro ES	E	F	G (Default in this study)

Macro ES = macro-environmental sensitivity; *micro ES* = micro-environmental sensitivity.

occasionally the average value deviated slightly from the true value (Table 3). The average of the estimated genetic variance in slope $\left(= \hat{\sigma}^2_{A_{sl}}\right)$ was biased upwards in all cases, but the bias was small (between +2 and +13%). Surprisingly, standard deviations of $\hat{\sigma}^2_{A_{sl}}$ relative to $\sigma^2_{A_{sl}}$ were small (between 21 and 30% of $\sigma^2_{A_{sl}}$). Note that standard deviations across replicates represent standard errors of estimates from a single replicate, e.g. an analysis of one data set. The average of the estimated genetic variance in environmental variance $\left(= \hat{\sigma}^2_{A_v}\right)$ had on average larger deviations from the true value but no direction in bias could be detected. Standard deviations of $\hat{\sigma}^2_{A_v}$ relative to $\sigma^2_{A_v}$ were large (between 31 and 74%), indicating that this parameter had a low precision. In the presence of fixed effects (i.e. contemporary group) for the mean and residual variance, standard deviations of estimated parameters increased on average by 9% and the number of replicates for which the genetic variance-covariance structure was forced to be positive definite also increased. Thus, the conclusion is that estimates of genetic variances are practically unbiased but that the precision of $\hat{\sigma}^2_{A_v}$ is low.

Effect of genetic correlations
In general, estimates of genetic variances remained close to their true values when genetic correlations were varied (<10% difference from the true value) (Table 4). When all genetic correlations were equal to 0.5, the genetic variance of environmental variance was slightly underestimated (−11%). Estimates of genetic correlations were unbiased but had very high standard deviations, especially the genetic correlation between the breeding value for slope and environmental variance (ρ_{A_{sl},A_v}). The number of replicates

in which the variance-covariance structure was forced to be positive definite was in general small but slightly higher when all genetic correlations were equal to 0.5. In addition to the results shown, we simulated an extreme negative genetic correlation ($\rho_{A_{sl},A_v} = -0.9$), as well as an extreme positive genetic correlation ($\rho_{A_{sl},A_v} = 0.9$). In these scenarios, for 40 to 60% of the replicates, the variance-covariance matrix was bended to be positive definite. Due to the large proportion of replicates with bended variance-covariance matrices, the average genetic correlation was biased towards zero. Nevertheless, the general conclusion is that estimates of genetic correlations are largely unbiased but are estimated with low precision.

Effect of different designs
When varying the number of offspring per sire or the number of sire families, means of genetic variances were close to their true simulated values in most cases, except when the number of offspring per sire was 20 for which estimates for $\sigma^2_{A_{sl}}$ and $\sigma^2_{A_v}$ were biased upwards (25% and 20%, respectively) (Table 5). In this case, 51% of the replicates had variance-covariance matrices that were forced to be positive definite, indicating that this is not a suitable design to estimate genetic parameters for macro- and micro-environmental sensitivities. When increasing the number of offspring per sire, the standard deviations of estimates of $\sigma^2_{A_{sl}}$ and $\sigma^2_{A_v}$ decreased substantially, e.g. from 50 to 200 offspring with 54 and 60%, respectively, whereas the standard deviation of estimates of $\sigma^2_{A_{int}}$ decreased only slightly (−10%). When increasing the number of sire families from 50 to 200, standard deviations of estimates of all genetic parameters decreased substantially (from −52 to −58%). Based on these results, we conclude that designs with at least 100 sires with 100

Table 3 Means and standard deviations of estimated genetic parameters across 100 replicates when genetic correlations are zero

Fixed effects[1]	True parameters			Estimated parameters (SD)						Np
	$\sigma^2_{A_{int}}$	$\sigma^2_{A_{sl}}$	$\sigma^2_{A_v}$	$\sigma^2_{A_{int}}$		$\sigma^2_{A_{sl}}$		$\sigma^2_{A_v}$		
No	0.1	0.05	0.1	0.101	(0.020)	0.051	(0.012)	0.095	(0.035)	0
No	0.3	0.05	0.1	0.315	(0.053)	0.054	(0.015)	0.107	(0.046)	0
No	0.5	0.05	0.1	0.507	(0.066)	0.052	(0.013)	0.115	(0.061)	3
No	0.3	0.05	0.05	0.297	(0.047)	0.053	(0.012)	0.053	(0.029)	4
No	0.3	0.05	0.2	0.308	(0.049)	0.053	(0.014)	0.186	(0.065)	0
No	0.3	0.025	0.1	0.298	(0.047)	0.026	(0.008)	0.097	(0.041)	0
No	0.3	0.1	0.1	0.296	(0.052)	0.104	(0.021)	0.083	(0.045)	1
Yes	0.1	0.05	0.1	0.099	(0.022)	0.053	(0.014)	0.091	(0.037)	1
Yes	0.3	0.05	0.1	0.309	(0.044)	0.057	(0.015)	0.104	(0.051)	0
Yes	0.5	0.05	0.1	0.503	(0.084)	0.052	(0.014)	0.104	(0.074)	10

[1]Models have either only a mean as fixed effect (no fixed effects) or have contemporary groups as fixed effects (yes); $\sigma^2_{A_{int}}$ = additive genetic variance of breeding value for the intercept; $\sigma^2_{A_{sl}}$ = additive genetic variance of breeding value for the slope (= macro-environmental sensitivity); $\sigma^2_{A_v}$ = additive genetic variance for environmental variance (= micro-environmental sensitivity); Np = number of replicates with covariance structures forced to be positive definite.

Table 4 Means and standard deviations across 100 replicates of estimated genetic parameters when genetic correlations are not zero

True parameters			Estimated parameters (SD)						
ρ_{A_{int},A_v}	$\rho_{A_{int},A_{sl}}$	ρ_{A_{sl},A_v}	$\sigma^2_{A_{int}}$	$\sigma^2_{A_{sl}}$	$\sigma^2_{A_v}$	ρ_{A_{int},A_v}	$\rho_{A_{int},A_{sl}}$	ρ_{A_{sl},A_v}	Np
0	0	0	0.315	0.054	0.107	−0.005	0.004	0.012	0
			(0.053)	(0.015)	(0.046)	(0.165)	(0.155)	(0.249)	
0.5	0	0	0.303	0.051	0.099	0.554	−0.028	−0.007	1
			(0.047)	(0.012)	(0.048)	(0.155)	(0.136)	(0.203)	
0	0.5	0	0.303	0.053	0.094	−0.010	0.508	−0.014	1
			(0.056)	(0.012)	(0.045)	(0.200)	(0.132)	(0.239)	
0	0	0.5	0.293	0.052	0.092	0.014	0.021	0.558	2
			(0.045)	(0.013)	(0.034)	(0.185)	(0.147)	(0.208)	
0.5	0.5	0.5	0.301	0.053	0.089	0.537	0.517	0.530	6
			(0.051)	(0.013)	(0.036)	(0.171)	(0.138)	(0.192)	

ρ_{A_{int},A_v} = genetic correlation between additive genetic effects for intercept and environmental variance; $\rho_{A_{int},A_{sl}}$ = genetic correlation between additive genetic effects for intercept and slope; ρ_{A_{sl},A_v} = genetic correlation between additive genetic effects for slope (macro-environmental sensitivity) and environmental variance (micro-environmental sensitivity); $\sigma^2_{A_{int}}$ = additive genetic variance of breeding value for intercept (true value = 0.3); $\sigma^2_{A_{sl}}$ = additive genetic variance of breeding value for slope (= macro-environmental sensitivity; true value = 0.05); $\sigma^2_{A_v}$ = additive genetic variance for environmental variance (= micro-environmental sensitivity; true value = 0.10); Np = number of replicates with covariance structures forced to be positive definite.

offspring each are required in order to estimate variance components for macro- and micro-environmental sensitivities with low standard errors.

Model selection

Using AIC to select the best fitting model, the true genetic model was selected as the best statistical model in at least 90% of the 100 replicates when the number of offspring

Table 5 Means and standard deviations across 100 replicates of estimated genetic parameters for different designs

NS	NO	Estimated parameters (SD)						Np
		$\sigma^2_{A_{int}}$	$\sigma^2_{A_{sl}}$	$\sigma^2_{A_v}$	ρ_{A_{int},A_v}	$\rho_{A_{int},A_{sl}}$	ρ_{A_{sl},A_v}	
100	20	0.301	0.063	0.120	−0.008	−0.004	0.083	51
		(0.057)	(0.038)	(0.113)	(0.304)	(0.216)	(0.416)	
100	50	0.309	0.056	0.099	−0.016	−0.013	0.025	8
		(0.051)	(0.020)	(0.060)	(0.279)	(0.165)	(0.376)	
100	100	0.315	0.054	0.107	−0.005	0.004	0.012	0
		(0.053)	(0.015)	(0.046)	(0.165)	(0.155)	(0.249)	
100	200	0.301	0.053	0.093	0.000	0.007	−0.013	0
		(0.046)	(0.009)	(0.024)	(0.135)	(0.116)	(0.168)	
50	100	0.312	0.053	0.107	0.015	0.002	0.036	5
		(0.079)	(0.018)	(0.064)	(0.295)	(0.199)	(0.320)	
200	100	0.301	0.053	0.099	0.009	−0.002	−0.009	0
		(0.033)	(0.009)	(0.028)	(0.142)	(0.108)	(0.163)	

NS = number of sires; NO = number of offspring per sire; $\sigma^2_{A_{int}}$ = additive genetic variance of breeding value for intercept (true value = 0.3); $\sigma^2_{A_{sl}}$ = additive genetic variance of breeding value for slope (= macro-environmental sensitivity; true value = 0.05); $\sigma^2_{A_v}$ = additive genetic variance for environmental variance (= micro-environmental sensitivity; true value = 0.10); $\rho_{A_{int},A_v} = \rho_{A_{int},A_{sl}} = \rho_{A_{sl},A_v}$ = 0; Np = number of replicates with covariance structures forced to be positive definite.

per sire was 100 (Table 6). In only a few cases did the best statistical model differ from the true genetic model. However, when the number of offspring per sire was 50, the correct model was selected for 63% to 95% of the replicates. When both macro- and micro-environmental sensitivity existed and the number of offspring per sire was 50, the macro model was selected rather than the macro–micro model in 27% of the replicates. When only micro-environmental sensitivity existed and 50 offspring per sire, the power to select the correct model was reduced. In conclusion, the power to select the correct model was high in designs with at least 100 offspring per sire.

Scenarios in which true genetic and statistical models differ

It is of interest to determine whether genetic parameters are biased if the wrong statistical model is applied. When either macro or micro-environmental sensitivity was simulated (Table 7), the variance-covariance structures were forced to be positive definite in many replicates, due to estimates of variances at the boundary. Estimates at the boundary were expected since the estimated genetic variance should be close to zero when the true genetic variance is zero. Forcing the genetic variance-covariance structures to be positive definite can bias estimates of parameters. When true macro-environmental sensitivity existed, some micro-environmental sensitivity was detected ($\hat{\sigma}^2_{A_v} > 0$) with the "micro" (scenario A) and the "macro–micro" (scenario B) models, indicating that part of the macro-environmental sensitivity was captured as micro-environmental sensitivity, but $\hat{\sigma}^2_{A_v}$ were small (0.010-0.018). When true micro-environmental sensitivity existed, no macro-environmental sensitivity

Table 6 The best model selected in 100 replicates according to Akaike's information criterion and effect of the number of offspring for different true genetic models

NO	True genetic model	True parameters		Nb of times model selected based on AIC			
		$\sigma^2_{A_v}$	$\sigma^2_{A_{sl}}$	Macro–micro	Macro	Micro	Simple
100	Macro–micro	0.1	0.05	99	1	0	0
	Macro	0	0.05	5	95	0	0
	Micro	0.1	0	3	1	94	2
	Simple	0	0	1	4	5	90
50	Macro–micro	0.1	0.05	63	27	8	2
	Macro	0	0.05	7	87	1	5
	Micro	0.1	0	4	1	71	24
	Simple	0	0	1	1	3	95

NO = number of offspring per sire, "Macro–micro" = model accounting for both macro- and micro-environmental sensitivities; "Macro" = model with only macro-environmental sensitivity; "Micro" = model with only micro-environmental sensitivity; "Simple" = model without macro- and micro environmental sensitivities and only a genetic effect for the phenotype; $\sigma^2_{A_{int}}$ = additive genetic variance of breeding value for intercept (true value = 0.3); $\sigma^2_{A_{sl}}$ = additive genetic variance of breeding value for slope (= macro-environmental sensitivity), $\sigma^2_{A_v}$ = additive genetic variance for environmental variance (= micro-environmental sensitivity), ρ_{A_{int},A_v} = $\rho_{A_{int},A_{sl}}$ = ρ_{A_{sl},A_v} = 0.

was detected ($\hat{\sigma}^2_{A_{sl}} \approx 0$) with the "macro" (scenario C) and "macro–micro" (scenario D) models, but $\hat{\sigma}^2_{A_v}$ were biased downwards (between −16% to −28%), possibly because of the high number of genetic variance-covariance structures that were forced to be positive definite. When both macro and micro-environmental sensitivity existed, the "macro" (scenario E) and "micro" (scenario F) models gave unbiased estimates of either macro-environmental sensitivity or micro-environmental sensitivity and the genetic variance-covariance structure was forced to be positive definite for at maximum three replicates. Thus, the general conclusion is that a discrepancy between the true and statistical models does not lead to large biases in estimated genetic parameters and that contamination between estimates of the two types of environmental sensitivity is very limited.

Application to milk yield in dairy cattle

The "macro–micro", "macro", "micro" and "simple" models were applied to 305-day first lactation milk yield data of Swedish Holsteins (mean = 8693 kg, standard deviation = 1652 kg, skew = 0.18, kurtosis = 0.28). "Macro", "micro" and "macro–micro" models fitted significantly better than the "simple" model. The "micro" model was favoured by AIC (Table 8) and had the best fit. The genetic variance for micro-environmental sensitivity was substantial but lower than for most reported traits [20]: one genetic standard deviation changed micro-environmental sensitivity (= environmental variance) by 21%. The difference in AIC between the "macro–micro" and "micro" models was small and therefore it was interesting to examine the genetic parameters of the "macro–micro" model. The estimated genetic variance for macro-environmental sensitivity was small in comparison to the genetic variance in intercept. For instance, the estimate of the genetic correlation between environments that were −2 and 2 standard

deviations from the overall mean was 0.92, indicating a small level of reranking of sires across the environmental gradient. Estimates of genetic correlations between intercept and macro- and micro-environmental sensitivities were 0.81 and 0.63, respectively, indicating that selection for higher milk yield increases both types of environmental sensitivity. The estimate of the genetic correlation between macro- and micro-environmental sensitivities was 0.76, indicating that they are genetically similar. Standard errors of parameter estimates were smaller than the standard deviations found for the simulations, due to the larger dataset i.e. more sires and more offspring per sire. Thus, macro- and micro-environmental sensitivities may exist for milk yield in dairy cattle and are positively correlated.

Discussion

Model and design

In this study, we developed a model to estimate genetic variances for macro- and micro-environmental sensitivities. The model is an extension of the DHGLM as presented by Rönnegård et al. [19]. Here, we combined a linear reaction norm model to estimate genetic variance for macro-environmental sensitivity with the DHGLM to estimate genetic variance for micro-environmental sensitivity. The animal model in Equation (2) was adapted to a sire model because the animal model produced highly biased estimated variance components because of the high dependence of the estimated breeding values and residuals on the variance ratio used in the mixed model equations, which would differ for each animal. Felleki et al. [31] also reported the presence of bias in variance components with few repeated observations per animal but the bias decreased as the number of repeated observations per animal increased. Furthermore, an animal model with heterogeneous residual variance (DHGLM implementation) gave

Table 7 Means and standard deviations across 100 replicates of estimated genetic parameters when true and statistical models differ

Scenario	True model	True parameters		Statistical model	Estimated parameters (SD)			
		$\sigma^2_{A_{sl}}$	$\sigma^2_{A_v}$		$\sigma^2_{A_{int}}$	$\sigma^2_{A_{sl}}$	$\sigma^2_{A_v}$	Np
A	Macro	0.025	0	Micro	0.303		0.010	51
					(0.045)		(0.020)	
	Macro	0.05	0	Micro	0.303		0.010	48
					(0.044)		(0.013)	
	Macro	0.1	0	Micro	0.308		0.018	37
					(0.050)		(0.019)	
B	Macro	0.025	0	Macro–micro	0.296	0.027	0.012	60
					(0.046)	(0.009)	(0.011)	
	Macro	0.05	0	Macro–micro	0.298	0.052	0.013	51
					(0.048)	(0.011)	(0.013)	
	Macro	0.1	0	Macro–micro	0.307	0.091	0.012	54
					(0.051)	(0.020)	(0.013)	
C	Micro	0	0.05	Macro	0.296	0.002		0
					(0.041)	(0.003)		
	Micro	0	0.1	Macro	0.292	0.003		0
					(0.039)	(0.004)		
	Micro	0	0.2	Macro	0.289	0.002		0
					(0.043)	(0.003)		
D	Micro	0	0.05	Macro–micro	0.303	0.003	0.053	71
					(0.050)	(0.003)	(0.024)	
	Micro	0	0.1	Macro–micro	0.300	0.003	0.084	60
					(0.047)	(0.003)	(0.034)	
	Micro	0	0.2	Macro–micro	0.304	0.002	0.143	66
					(0.049)	(0.003)	(0.058)	
E	Macro–micro	0.05	0.05	Macro	0.297	0.055		0
					(0.047)	(0.014)		
	Macro–micro	0.05	0.1	Macro	0.298	0.053		0
					(0.052)	(0.012)		
	Macro–micro	0.05	0.2	Macro	0.292	0.053		0
					(0.049)	(0.011)		
	Macro–micro	0.025	0.1	Macro	0.306	0.026		0
					(0.050)	(0.008)		
	Macro–micro	0.1	0.1	Macro	0.298	0.106		0
					(0.045)	(0.022)		
F	Macro–micro	0.05	0.05	Micro	0.299		0.050	3
					(0.051)		(0.031)	
	Macro–micro	0.05	0.1	Micro	0.297		0.095	1
					(0.047)		(0.042)	
	Macro–micro	0.05	0.2	Micro	0.304		0.191	0
					(0.045)		(0.070)	
	Macro–micro	0.025	0.1	Micro	0.299		0.105	0
					(0.049)		(0.043)	
	Macro–micro	0.1	0.1	Micro	0.298		0.103	1
					(0.045)		(0.045)	

See Table 2 for schematic overview of scenarios; "Macro–micro" = model accounting for both macro- and micro-environmental sensitivities; "Macro" = model with only macro-environmental sensitivity; "Micro" = model with only micro-environmental sensitivity; $\sigma^2_{A_{int}}$ = additive genetic variance of breeding value for intercept (true value = 0.3); $\sigma^2_{A_{sl}}$ = additive genetic variance of breeding value for slope (= macro-environmental sensitivity); $\sigma^2_{A_v}$ = additive genetic variance for environmental variance (= micro-environmental sensitivity), $\rho_{A_{int},A_v} = \rho_{A_{int},A_{sl}} = \rho_{A_{sl},A_v} = 0$; Np = number of replicates with covariance structures forced to be positive definite.

Table 8 Estimated genetic parameters for macro- and micro-environmental sensitivity of milk yield in dairy cattle

Parameter	Macro–micro		Macro		Micro		Simple	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
$\sigma^2_{A_{int}}$	420 800	27960	420 400	28 004	416 800	27 696	416 000	27 692
$\sigma^2_{A_{sl}}$	11 096	2288	11 116	2320				
$\sigma^2_{A_v}$	0.043	0.008			0.042	0.008		
$\rho_{A_{int},A_{sl}}$	0.808	0.062	0.812	0.063				
ρ_{A_{int},A_v}	0.627	0.073			0.608	0.0751		
ρ_{A_{sl},A_v}	0.765	0.098						
APHL	193 704		194 179		193 692		202 832	
AIC	193 722		194 191		193 704		202 840	

$\sigma^2_{A_{int}}$ = additive genetic variance of breeding value for intercept; $\sigma^2_{A_{sl}}$ = additive genetic variance of breeding value for slope (= macro-environmental sensitivity); $\sigma^2_{A_v}$ = additive genetic variance in environmental variance (= micro-environmental sensitivity); ρ_{A_{int},A_v} = genetic correlation between breeding value for intercept and environmental variance; $\rho_{A_{int},A_{sl}}$ = genetic correlation between breeding values of intercept and slope of reaction norm; ρ_{A_{sl},A_v} = genetic correlation between breeding values of slope and environmental variance; APHL = adjusted profile h-likelihood; AIC = Akaike's information criterion; "Macro–micro" = model accounting for both macro- and micro-environmental sensitivities; "Macro" = model with only macro-environmental sensitivity; "Micro" = model with only micro-environmental sensitivity; "Simple" = model without macro- and micro environmental sensitivities and only a genetic effect for the phenotype; SE = approximate standard error obtained with ASReml.

a poorer adjusted profile h-likelihood than an animal model without homogenous residual variance, which indicates that the former did not produce a better fit than the latter in a scenario that included both macro and micro-environmental sensitivities. Therefore, we decided to use a sire model implementation, because it is more robust than the animal model implementation with DHGLM when animals only have a single observation. Furthermore, sire models are commonly used for reaction norm models, because of their substantially lower computational burden compared to animal models and genetic information about environmental sensitivity typically comes from paternal half-sibs that perform in different environments. As far as we know, this is the first time that a model to estimate genetic variance in macro- and micro-environmental sensitivities suitable for outbred animal populations is presented.

Monte Carlo simulation was used to evaluate bias and precision of estimated genetic parameters. Genetic parameters were unbiased in most situations. The precision was not very high, especially of estimates for $\sigma^2_{A_v}$, as indicated by the high standard deviation of estimates across replicates, particularly in designs with small families. Designs with at least 100 sire families, each with at least 100 offspring, are required to have sufficient precision. Presence of fixed effects, such as contemporary group effects, would increase the required number of sire families and the number of offspring per sire family. These results are in agreement with standard error and power calculations reported by Hill [39], Mulder [40] and Hill and Mulder [20] with respect to estimation of genetic variance for micro-environmental sensitivity or for environmental variance. For instance, Hill and Mulder [20] derived that the

optimal family size to estimate genetic variance for environmental variance with a family design is approximately $2/\gamma^2$, where γ^2 is the square of the coefficient of variation of the within-family variance. Thus, the optimal family size for half-sibs is 137 when $\sigma^2_{A_v} = 0.10$. The only study providing Monte Carlo results for the DHGLM to estimate genetic variance for micro-environmental sensitivity is by Rönnegård et al. [19]. They considered a design with clones, which is more powerful and leads to lower standard deviations of estimates than those in our study.

With respect to macro-environmental sensitivity, the magnitude of the standard deviations of the $\sigma^2_{A_{sl}}$ estimates across replicates was similar to that reported by Calus et al. [36], but larger than those reported by Lillehammer et al. [41], which is explained by the fact that the latter authors simulated more sire families, i.e., 1000 sire families with 100 offspring each. In general, the number of required offspring per family is lower for macro-environmental sensitivity than for micro-environmental sensitivity, as indicated by the lower standard deviations of estimates of genetic variance for the former. This is in agreement with a previous study by Mulder [40], which showed that the power to detect G x E interactions between two environments is greater than the power to detect genetic variance for environmental variance or micro-environmental sensitivity. In the present study, we assumed that the environmental parameter x used for the reaction norm was known without error. This will be true in some cases, e.g. when using temperature or rainfall or other herd characteristics [42]. In other cases, an estimated herd mean is used as the environmental parameter, which is estimated from the data [7,8,33]. Calus et al. [36] showed that genetic variance in macro-environmental sensitivity ($\sigma^2_{A_{sl}}$) was severely underestimated when the environmental parameter was

estimated from the data. This may have also led to the genetic variance in macro-environmental sensitivity to be underestimated in our application to milk yield. Su et al. [43] reported that a Bayesian approach that estimates simultaneously the herd mean and the reaction norm parameters leads to unbiased estimates of $\sigma^2_{A_{sl}}$.

Lillehammer et al. [41] showed that sire models gave upward biased estimates of $\sigma^2_{A_{sl}}$ when heterogeneity of residual variance was ignored, because the unexplained genetic variance in the reaction norm parameters becomes part of the residual variance when using a sire model. In our study, the bias in estimates of $\sigma^2_{A_{sl}}$ was smaller than in Lillehammer et al. [41]. Lillehammer et al. [41] proposed including a dummy animal effect in the model to account for the residual three-quarters of the genetic variance that is not accounted for by the sire effect. This solution was also tested in our model but gave severely biased variance components because of the high dependency of estimated breeding values and residuals on variance ratios that were used in the mixed model equations, which differ by animal when considering heterogeneity of residual variance.

The algorithm developed in this study allowed estimating genetic correlations between the different genetic effects. Standard deviations of estimated genetic correlations were large, especially those of the genetic correlation between macro- and micro-environmental sensitivities (ρ_{A_{sl},A_v}). This large magnitude of the standard deviation of estimates of the genetic correlation was expected considering that the genetic correlation between macro- and micro-environmental sensitivities is mainly based on paternal half-sib information and that both traits have low heritability. Using the equation in Robertson [44] and assuming a heritability of 0.05 for both traits, the standard error is approximately 0.26 when the true genetic correlation between the traits is zero, which is close to the value found here, i.e. 0.25 (Table 4). To increase the precision of estimates of the genetic correlation between macro- and micro-environmental sensitivities, designs with a larger number of families and larger family sizes are required. The application to milk yield data in dairy cattle shows that it should be possible to estimate genetic correlations with standard errors between 0.06 and 0.10, since in most countries datasets with at least 100 bulls each with 100 daughters are easily obtained. Full-sib families or clones would also reduce standard errors of estimates of genetic correlations in comparison to half-sibs.

Here, we showed that the adjusted profile h-likelihood (APHL) can be approximated from REML-output and used in combination with AIC to provide an efficient model selection tool. In addition, we showed that biases in genetic parameters were relatively small when statistical and true models differed. Both results are re-assuring that these models can discriminate between macro- and micro-environmental sensitivities and that the true model of environmental sensitivity can be elucidated using AIC. We also found that the Bayesian information criterion (BIC) was too conservative and favoured the simpler model too often (results not shown). AIC has the advantage that it can be used independent of the order of the fitted models, whereas the likelihood ratio test requires a hierarchical structure such as in a forward selection scheme [45].

Improving biological understanding of environmental sensitivity

The proposed model can contribute to better understand the genetic architecture of environmental sensitivity, e.g. whether macro- and micro-environmental sensitivities are genetically related. Most studies in plants and laboratory species indicate that macro- and micro-environmental sensitivities are weakly correlated [22-24,26-28]. This seems to indicate that selection on one type of environmental sensitivity will hardly affect the other. The first application of the model on milk yield data in dairy cattle, revealed a high genetic correlation between macro- and micro-environmental sensitivities (0.76), suggesting that selection on one type of environmental sensitivity will also affect the other in the same direction. Generally, little is known about these relationships in livestock. Knowledge about these genetic correlations could be used to optimize selection strategies for environmental sensitivity.

In this study, we assumed a linear reaction norm model, but reaction norms can also be non-linear [8]. The model presented here can easily be extended to higher-order polynomials. Furthermore, the genetic basis of micro-environmental sensitivity may not be the same along an environmental gradient and the model for residual variance or micro-environmental sensitivity in Equation (3) could be extended to contain a reaction norm with a known environmental gradient. For instance, in stressful environments, there might be more genetic variance for micro-environmental sensitivity than in less stressful environments. Furthermore, G × E interactions often exist between categorical environments and are often analysed with character state or multivariate models [2]. Character state models do not explicitly estimate breeding values for macro-environmental sensitivity, but these breeding values could be back-calculated by using covariance functions when the environmental parameter responsible for G × E interactions is identified because reaction norm models and character state models are interchangeable [2]. In the case of two environments, a reaction norm model with a dummy environmental parameter with values 0 and 1 would yield results that are identical to a bivariate character state model. Multivariate versions of the DHGLM or reaction norm models with dummy environmental variables could be

used to simultaneously investigate macro- and micro-environmental sensitivities when environments are discrete.

Application to breeding

Taking macro- and micro-environmental sensitivities into consideration is highly relevant in animal breeding. Due to the high level of globalisation in animal breeding programs, it is necessary to breed animals that can perform well in a wide range of environments. Therefore, it may be important to select animals that have limited environmental sensitivity, especially for environments with a higher risk of environmental disturbances. Reduction in environmental sensitivity increases the predictability of performance and reduces risk for farmers [46]. Furthermore, reduction in micro-environmental sensitivity will increase the uniformity of animal products [47], which is a general goal. In plant breeding, application of a model for macro- and micro-environmental sensitivities is also highly relevant since G × E interactions are generally very strong and uniformity of crops is very important. Recent papers by Ordas et al. [48], Makumburage and Stapleton [49] and Kliebenstein [50] show that there is an interest for increased uniformity in plants. Economic values could be derived for micro-environmental sensitivity [47]. The economic value of macro-environmental sensitivity can be determined as a function of the importance of environments along the environmental gradient. Based on Mulder et al. [3,47], progeny testing schemes are more efficient than sib testing schemes to reduce micro-environmental sensitivity since it behaves as a trait with a small heritability. Genomic selection could be an alternative selection strategy with sufficient accuracy and shorter generation intervals.

Conclusions

In this study, a model was developed to estimate genetic parameters of macro- and micro-environmental sensitivities, combining a reaction norm model with a double hierarchical generalized linear model within a REML framework. Simulations showed that the genetic parameters obtained were mostly unbiased, but designs with at least 100 sires, each with 100 half-sib offspring, were required to estimate genetic parameters with sufficient precision. Using AIC, the true genetic model was selected as the best statistical model in at least 90% of replicates when the number of offspring per sire was 100. Application of the model to milk yield data in dairy cattle showed that both types of environmental sensitivity existed. Our model and AIC based on h-likelihood can be used to increase our understanding of the genetic control of environmental sensitivity in livestock populations but more research is needed to test the model in a wider range of situations.

Appendix

DHGLM algorithm for a sire model

The original DHGLM algorithm of Rönnegård et al. [19] was developed for an animal model. Here we describe the estimation algorithm for the sire model used in the current paper, including a few adjustments of the algorithm in Rönnegård et al. [19] to correct for the fact that the residual variance in a sire model (without permanent environmental effects, e.g. with animals with a single observation) contains three quarters of the additive genetic variance in addition to the environmental variance. The adjustments in the algorithm are as follows: adjustment of the squared residuals ($\mathbf{y_v}$) accounting for the fact that the residual variance in a sire model includes three quarters of the additive genetic variance, use of average residual variance $\hat{\sigma}^2_{e_s}$ to calculate $\boldsymbol{\psi_s}$ (instead of predicted individual values), and computations of the diagonal weight matrices $\mathbf{W_s}$ and $\mathbf{W_{sv}}$. These adjustments resulted in a computationally robust algorithm with small or no bias (as presented in Results).

To compute the linearized response $\boldsymbol{\psi_s}$ for the bivariate sire model in Equation (3), first we calculate y_{v_i} as:

$$y_{v_i} = \frac{\hat{e}^2_{s_i}}{(1-h_i)} \times \frac{\overline{\hat{\sigma}^2_{e_s}}}{\hat{\sigma}^2_{e_a}}. \tag{6}$$

This is equivalent to the calculations for the response y_v in [19], except for the multiplication by the ratio of the average estimated residual variance in a sire model ($\overline{\hat{\sigma}^2_{e_s}} = \sigma^2_{e_s}/\bar{w}_s$, where $\bar{w}_s = \text{tr}(\mathbf{W_s})/n$ and n is the total number of records) and the average residual variance in an animal model, which is calculated as $\overline{\hat{\sigma}^2_{e_a}} = \overline{\hat{\sigma}^2_{e_s}} - \frac{3}{4}\hat{\sigma}^2_{A_{int}}$, ($\hat{\sigma}^2_{A_{int}} = 4\hat{\sigma}^2_{s_{int}}$). Because we use a log link function, y_{v_i} is linearized as:

$$\psi_{s_i} = \log\left(\overline{\hat{\sigma}^2_{e_s}}\right) + \frac{y_{v_i} - \overline{\hat{\sigma}^2_{e_s}}}{\overline{\hat{\sigma}^2_{e_s}}}. \tag{7}$$

The diagonals of $\mathbf{W_s}$ are the reciprocals of $\hat{\boldsymbol{\psi}}_\mathbf{s}$ (i.e. $W_{s_i} = 1/\hat{\psi}_{s_i}$), which is the vector of predicted residual variances for each observation based on the previous iteration and is calculated as:

$$\hat{\psi}_{s_i} = \exp\left(\log\left(\overline{\hat{\sigma}^2_{e_s}} - \frac{3}{4}\hat{\sigma}^2_{A_{int}}\right) + s_{v_i}\right) + \frac{3}{4}\hat{\sigma}^2_{A_{int}}. \tag{8}$$

The sire effects for environmental variance s_{v_i} only affect the part of the residual variance which is truly environmental variance and therefore three-quarters of the additive genetic variance is subtracted in the multiplicative part of Equation (8). The diagonals of $\mathbf{W_{sv}}$ are the

reciprocals of the residual variance of ψ_s and were calculated as:

$$W_{sv_i} = \frac{1}{2}(1-h_i)^2 \left(\frac{\hat{\sigma}^2_{e_a}}{\hat{\sigma}^2_{e_s}}\right)^2. \tag{9}$$

This is motivated by the fact that by combining Equations (6) and (7), and by assuming that the estimated residuals are close to the true ones, we have:

$$\text{var}\left(\psi_{s_i}\right) = \text{var}\left(\frac{e^2_{s_i}}{(1-h_i)} \times \frac{\overline{\hat{\sigma}^2_{e_s}}}{\hat{\sigma}^2_{e_a}} \times \frac{1}{\hat{\sigma}^2_{e_s}}\right)$$
$$= \text{var}\left(\frac{e^2_{s_i}}{(1-h_i)} \times \frac{1}{\hat{\sigma}^2_{e_a}}\right) = 2(1-h_i)^{-2}\left(\frac{\hat{\sigma}^2_{e_s}}{\hat{\sigma}^2_{e_a}}\right)^2. \tag{10}$$

since the true residuals are assumed normal and the squared true residuals are therefore Gamma distributed with a variance of $2\left(\hat{\sigma}^2_{e_s}\right)^2$.

The algorithm can be summarized as:

1. Run model on $\mathbf{y}$ in Equation (3) with homogeneous residual variance.
2. Calculate ψ_s, $\mathbf{W_s}$, $\mathbf{W_{sv}}$, where $\mathbf{W_s} = \text{diag}\left(\hat{\sigma}^2_e\right)$ in iteration 1.
3. Run bivariate model in Equation (3).
4. Update ψ_s, $\mathbf{W_s}$, $\mathbf{W_{sv}}$
5. Iterate steps 3 till 4 until convergence

Approximation of the adjusted profile h-likelihood
In our paper, the model selection was based on an approximation of the adjusted profile h-likelihood (APHL), which is defined as [51]:

$$\text{APHL} = \left(2h - \log(\det(\mathbf{H}))\right)\big|_{\tau = \hat{\tau}}, \tag{11}$$

where $\mathbf{H}$ is the Hessian of the h-likelihood and τ is the vector of all fixed and random effects both in the mean and variance parts of the model. For the model in Equation (3), minus two times the h-likelihood ($-2h$) is:

$$-2h = -2\left(l(\mathbf{y}|\mathbf{s_{int}}, \mathbf{s_{sl}}, \mathbf{s_v}) + l(\mathbf{s_{int}}, \mathbf{s_{sl}}, \mathbf{s_v})\right) \tag{12}$$
$$= \sum_{i=1}^{n}\left(\log\left(\hat{\sigma}^2_{e_i}\right) + \frac{\hat{e}^2_i}{\hat{\sigma}^2_{e_i}}\right)$$
$$+ \log\left(\det(\tilde{\mathbf{G}})\right) + \tilde{\mathbf{s}}'\tilde{\mathbf{G}}^{-1}\tilde{\mathbf{s}},$$

where l is log-likelihood, $\hat{\sigma}^2_{e_i}$ ($\hat{\sigma}^2_{e_i} = \hat{\sigma}^2_{\epsilon_s}/w_{s_i}$, with w_{s_i} being the ith diagonal of $\mathbf{W_s}$) is the residual variance of observation i, $\hat{e}^2_i$ is the estimated squared residual of observation i, $\tilde{\mathbf{G}}$ is the covariance matrix of all random

effects ($\tilde{\mathbf{G}} = \frac{1}{4}\mathbf{G} \otimes \mathbf{A}$), $\mathbf{s}$ is a vector of all random sire effects ($\tilde{\mathbf{s}}' = [\mathbf{s_{int}} \quad \mathbf{s_{sl}} \quad \mathbf{s_v}]$) and w_{s_i} is the weight for the mean model for observation i. The minus two log REML likelihood ($logL$) from the bivariate model in Equation (3) is:

$$-2logL = \sum_{i=1}^{n+k}\left(\log\left(\hat{\sigma}^2_{e_i}\right) + \frac{\hat{e}^2_i}{\hat{\sigma}^2_{e_i}}\right)$$
$$+ \log\left(\det(\tilde{\mathbf{G}})\right) + \tilde{\mathbf{s}}'\tilde{\mathbf{G}}^{-1}\tilde{\mathbf{s}} + \log(\det(\mathbf{C})), \tag{13}$$

where the first n residual variances come from the first part of the bivariate model ($\mathbf{y}$) and the next k residual variances come from the second part of the bivariate model (ψ_s) ($k = n$ = number of records), and $\mathbf{C}$ is the Hessian of the bivariate model (i.e. left-hand-side of the mixed model equations). Because log (det($\mathbf{C}$)) is a reasonable approximation of log (det ($\mathbf{H}$)) [31], we can approximate APHL as given in Equation (4) in the main text:

$$\text{APHL} = -2logL - \sum w_{sv_i}\left(\hat{\sigma}^2_{\epsilon_{sv}}\right)^{-1} - \sum \ln\left(\hat{\sigma}^2_{\epsilon_{sv}}\right)/w_{sv_i} \tag{14}$$

Thus, the REML likelihood for the bivariate model is corrected for the fact that the squared residuals are used as "observations" in the bivariate model. Note that equation (14) can also be used for animal models by replacing the elements w_{sv_i} and $\sigma_{\epsilon_{sv}^2}$ with the corresponding elements of Equation (2) or for models with more or fewer random effects in the mean and variance model.

Author details
[1]Wageningen UR Livestock Research, Animal Breeding and Genomics Centre, PO Box 65, 8200 AB, Lelystad, The Netherlands. [2]Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH, Wageningen, The Netherlands. [3]School of Technology and Business Studies, Högskolan Dalarna SE-79188, Falun, Sweden. [4]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, SE-75007, Uppsala, Sweden.

References

1. Falconer DS, Mackay TFC: *Introduction to quantitative genetics.* 4th edition. Essex: Pearson Education Limited; 1996.
2. De Jong G, Bijma P: **Selection and phenotypic plasticity in evolutionary biology and animal breeding.** *Livest Prod Sci* 2002, **78**:195–214.
3. Mulder HA, Bijma P, Hill WG: **Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance.** *Genetics* 2007, **175**:1895–1910.
4. Meyer K: **Factor-analytic models for genotype x environment type problems and structured covariance matrices.** *Genet Sel Evol* 2009, **41**:21.
5. Piepho HP, Mohring J, Melchinger AE, Buchse A: **BLUP for phenotypic selection in plant breeding and variety testing.** *Euphytica* 2008, **161**:209–228.
6. Scheiner SM: **Selection experiments and the study of phenotypic plasticity.** *J Evol Biol* 2002, **15**:889–898.
7. Kolmodin R, Strandberg E, Madsen P, Jensen J, Jorjani H: **Genotype by environment interaction in Nordic dairy cattle studied using reaction norms.** *Acta Agric Scand A* 2002, **52**:11–24.
8. Calus MPL, Veerkamp RF: **Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model.** *J Dairy Sci* 2003, **86**:3756–3764.
9. Bergsma R, Hermesch S: **Exploring breeding opportunities for reduced thermal sensitivity of feed intake in the lactating sow.** *J Anim Sci* 2012, **90**:85–98.
10. Bradshaw AD: **Evolutionary significance of phenotypic plasticity in plants.** *Adv Genet* 1965, **13**:115–155.
11. Scheiner SM: **Genetics and evolution of phenotypic plasticity.** *Annu Rev Ecol Syst* 1993, **24**:35–68.
12. Nussey DH, Postma E, Gienapp P, Visser ME: **Selection on heritable phenotypic plasticity in a wild bird population.** *Science* 2005, **310**:304–306.
13. Windig JJ: **Reaction norms and the genetic basis of phenotypic plasticity in the wing pattern of the butterfly Bicyclus anynana.** *J Evol Biol* 1994, **7**:665–695.
14. Waddington CH: **Canalization of development and the inheritance of acquired characters.** *Nature* 1942, **150**:563–565.
15. Flatt T: **The evolutionary genetics of canalization.** *Q Rev Biol* 2005, **80**:287–316.
16. Debat V, David P: **Mapping phenotypes: canalization, plasticity and developmental stability.** *Trends Ecol Evol* 2001, **16**:555–561.
17. Sorensen D, Waagepetersen R: **Normal linear models with genetically structured residual variance heterogeneity: a case study.** *Genet Res* 2003, **82**:207–222.
18. Ibanez-Escriche N, Sorensen D, Waagepetersen R, Blasco A: **Selection for environmental variation: a statistical analysis and power calculations to detect reponse.** *Genetics* 2008, **180**:2209–2226.
19. Rönnegård L, Felleki M, Fikse F, Mulder HA, Strandberg E: **Genetic heterogeneity of residual variance: estimation of variance components using double hierarchical generalized linear models.** *Genet Sel Evol* 2010, **42**:8.
20. Hill WG, Mulder HA: **Genetic analysis of environmental variation.** *Genet Res* 2010, **92**:381–395.
21. Jinks JL, Pooni HS: **The genetic basis of environmental sensitivity.** In *Proceedings of the Second International Conference on Quantitative Genetics: 31 May - 5 June 1987; Raleigh.* Edited by Weir BS, Eisen EJ, Goodman MM, Namkoong G. North Carolina State University; 1988:505–522.
22. Perkins JM, Jinks JL: **The assessment and specificity of environmental and genotype-environmental components of variability.** *Heredity* 1973, **30**:111–126.
23. Wu RL: **Genetic control of macro- and micro-environmental sensitivities in Populus.** *Theor Appl Genet* 1997, **94**:104–114.
24. Wu RL, O'Malley DM: **Nonlinear genotypic response to macro- and micro-environments.** *Theor Appl Genet* 1998, **96**:669–675.
25. Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA: **Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars.** *Genetics* 2004, **168**:435–446.
26. Yampolsky LY, Scheiner SM: **Developmental noise, phenotypic plasticity, and allozyme heterozygosity in Daphnia.** *Evolution* 1994, **48**:1715–1722.
27. Weber SL, Scheiner SM: **The genetics of phenotypic plasticity. IV. Chromosomal localization.** *J Evol Biol* 1992, **5**:109–120.
28. Scheiner SM, Caplan RL, Lyman RF: **The genetics of phenotypic plasticity. III. Genetic correlations and fluctuating asymmetries.** *J Evol Biol* 1991, **4**:51–68.
29. SanCristobal-Gaudy M, Elsen JM, Bodin L, Chevalet C: **Prediction of the response to a selection for canalisation of a continuous trait in animal breeding.** *Genet Sel Evol* 1998, **30**:423–451.
30. Hoaglin DC, Welsh RE: **The hat matrix in regression and ANOVA.** *Am Stat* 1978, **32**:17–22.
31. Felleki M, Lee D, Lee Y, Gilmour AR, Ronnegard L: **Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models.** *Genet Res* 2012, **94**:307–317.
32. Mulder HA, Hill WG, Vereijken A, Veerkamp RF: **Estimation of genetic variation in residual variance in female and male broiler chickens.** *Animal* 2009, **3**:1673–1680.
33. Windig JJ, Mulder HA, Bohthe-Wilhelmus DI, Veerkamp RF: **Simultaneous estimation of genotype by environment interaction accounting for discrete and continuous environmental descriptors in Irish dairy cattle.** *J Dairy Sci* 2011, **94**:3137–3147.
34. Gilmour AR, Gogel BJ, Cullis BR, Thompson R: *ASReml User Guide Release 2.0.* VSN International Ltd: Hemel Hempstead; 2006.
35. Lee Y, Nelder JA: **Double hierarchical generalized linear models.** *J Roy Stat Soc C* 2006, **55**:139–167.
36. Calus MPL, Bijma P, Veerkamp RF: **Effects of data structure on the estimation of covariance functions to describe genotype by environment interactions in a reaction norm model.** *Genet Sel Evol* 2004, **36**:489–507.
37. Rönnegård L, Felleki M, Fikse WF, Mulder HA, Strandberg E: **Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle.** *J Dairy Sci* 2013, **96**:2627–2636.
38. ICAR: *International agreement of recording practices*; 2012. http://www.icar.org/pages/recording_guidelines.htm
39. Hill WG: **Heterogeneity of genetic and environmental variance of quantitative traits.** *J Ind Soc Agric Stat* 2004, **57**:49–63.
40. Mulder HA: *Methods to optimize livestock breeding programs with genotype by environment interaction and genetic heterogeneity of environmental variance. PhD Thesis.* Wageningen University, Animal Breeding and Genetics Group; 2007.
41. Lillehammer M, Odegard J, Meuwissen THE: **Reducing the bias of estimates genotype by environment interactions in random regression sire models.** *Genet Sel Evol* 2009, **41**:30.
42. Windig JJ, Calus MPL, Veerkamp RF: **Influence of herd environment on health and fertility and their relationship with milk production.** *J Dairy Sci* 2005, **88**:335–347.
43. Su G, Madsen P, Lund MS, Sorensen D, Korsgaard IR, Jensen J: **Bayesian analysis of the linear reaction norm model with unknown covariates.** *J Anim Sci* 2006, **84**:1651–1657.
44. Robertson A: **The sampling variance of the genetic correlation coefficient.** *Biometrics* 1959, **15**:469–485.
45. Burnham KP, Anderson DR: *Model selection and multi-model inference: a practical information-theoretic approach.* New York: Springer; 2002.
46. Eskridge KM, Johnson BE: **Expected utility maximization and selection of stable plant cultivars.** *Theor Appl Genet* 1991, **81**:825–832.
47. Mulder HA, Bijma P, Hill WG: **Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance.** *Genet Sel Evol* 2008, **40**:37–59.
48. Ordas B, Malvar RA, Hill WG: **Genetic variation and quantitative trait loci associated with developmental stability and the environmental correlation between traits in maize.** *Genet Res* 2008, **90**:385–395.
49. Makumburage GB, Stapleton AE: **Phenotype uniformity in combined-stress environments has a different genetic architecture than in single-stress treatments.** *Front Plant Sci* 2011, **2**:12.
50. Kliebenstein DJ: **The quantitative genetics of phenotypic error or uniformity.** *Front Genet* 2011, **2**:59.
51. Lee Y, Nelder JA, Pawitan Y: *Generalized linear models with random effects: unified analysis via H-likelihood.* Chapman & Hall/CRC; 2006.