Use of Structured Antedependence Models to estimate Genotype by Environment Interaction

M. P. L. Calus[1], F. Jaffrézic[2], and R. F. Veerkamp[1]
[1]Animal Sciences Group, Wageningen UR,
P.O. Box 65, 8200 AB Lelystad, The Netherlands
E-mail: mario.calus@wur.nl; *Phone*: +31-(0)320-238265
[2]INRA Quantitative and Applied Genetics, Jouy-en-Josas Cedex, France

Abbreviation key: EP = environmental parameter, **ES** = environmental sensitivity, **RRM** = random regression model, **SAD** = structured antedependence model

INTRODUCTION

The increasing importance of health and fertility traits in breeding programs, has led to a renewed interest in environmental sensitivity (ES), also known as genotype by environment interaction. One aspect of ES is that in a multitrait situation genetic correlations between different traits may depend on the environment in which the traits are expressed. In recent years, ES has mainly been estimated by random regression models, in which traits are estimated as a regression on a parameter reflecting herd environment and management (Calus and Veerkamp, 2003; Kolmodin et al., 2002). These models estimate a covariance function which can be used to calculate genetic variances for a range of environments and genetic correlations of a trait with itself in different environments (Kirkpatrick et al., 1990). Hence, RRM seems a logical method of choice to model ES in a multitrait setting. However, multitrait RRM have a (co)variance structure in which the (co)variance components that model the genetic correlation between traits, might be influenced by the variance components of the separate traits. Similarly, Veerkamp and Goddard (1998) showed that genetic parameters for yield traits are not only influenced by month of lactation and herd production separately, but also by an interaction between those two factors.

An alternative model to model longitudinal data, such as used to model ES, is the structured antedependence model (SAD) (Nunez-Anton and Zimmerman, 2000). In a SAD model random effects are estimated as a function of the same trait in different environments and the (co)variances are modeled by so-called innovation variances and antedependence parameters. One of the major differences between the models is that the SAD allows the genetic correlations between a trait in different environments to be less than unity in situations where the genetic variance is constant across environments, whilst RRM implicitly model the change in covariances and variances simultaneously in one (covariance) function. Also, in a multitrait situation, the genetic correlation between traits across environments can be modeled by fewer parameters in SAD compared to RRM (Jaffrezic et al., 2003). Thus, SAD might be more flexible in estimating genetic correlations between a trait in different environments, and between different traits across environments. The objective of this research was to compare structured antedependence models (SAD) to random regression models (RRM), applied to simulated data sets, in their ability to estimate genotype by environment interaction based on univariate analyses.

MATERIALS AND METHODS

Simulations

The dataset was created by simulating additive genetic and environmental effects following normal distributions. A simulated population contained 250 000 animals, 100 sires and 1 000 herds. Sires had 2 500 daughters and the average number of animals per herd was 250. Only one generation of animals was simulated and no selection was considered. The environmental sensitivity for the additive genetic effect was simulated using i) a random regression with increasing variance across environments depending on environmental values, or ii) a model with five (co)variances with the variances either constant or increasing across five groups of environments. The five discrete groups of environments were formed in the simulations by ordering the herds on simulated herd effects. Environmental values per group were calculated as the average of simulated herd effects within the group. Simulated genetic correlations between groups of environments 1 and 5 were 0.66 for all three simulated data sets.

Ranges of simulated heritabilities of the traits were between 0.3 and 0.6 across environments. Simulated values of other genetic parameters specifically per group of environments are given in the figure and tables in the Results section. Fifty replicates were simulated for all situations considered.

Estimation of Environmental Sensitivity

Variance components were estimated both with a first order RRM and a first order SAD. Herd environments were characterized by environmental parameters (EP), calculated as average phenotypic performance within a herd. Five discrete groups of environments were formed based on increasing EP. ES was modeled in the RRM by applying a random regression for each sire, representing its EBV, on values of an EP for the herds in which his daughters were producing. ES was modeled in the SAD by estimating breeding values for each sire in all five environments. The sire variances were modeled by innovation variances and antedependence parameters depending on the average EP in each of the five discrete groups of herds. The residual variance was in both models estimated separately for each of those five groups. In both models a fixed herd effect was included. The applied RRM was:

$$Y_{klmn} = \mu + herd_k + \sum_{j=0}^{1} \alpha_{jl} P_{jk} + E_{klmn}$$

where

Y_{klmn} is the performance of animal n,
μ is the average performance over all animals,
$herd_k$ is a fixed effect for herd k,
α_{jl} is coefficient j of the random regression on the orthogonal polynomials of all environmental parameters of the daughters of sire l,
P_{jk} is element j of the orthogonal polynomial resembling an environmental parameter of herd k,
E_{klmn} is the residual effect of heifer n in herd k within group of herds m (m = 1, 2, ..., 5).

The applied SAD model was:

$$Y_{klmn} = \mu + herd_k + sire_{lm} + E_{klmn}$$

where $sire_{lm}$ is a random effect for sire l in group of herds m. Following the concept of SAD, $\mathrm{sire}_{lm}=\varepsilon(\mathrm{sire}_{lm})$ for m=1, and $\mathrm{sire}_{lm} = \theta_{m,m-1}{*}\mathrm{sire}_{l(m-1)} + \varepsilon\,(\mathrm{sire}_{lm})$ for $m \geq 2$. The antedependence parameter $\theta_{m,m-1}$ is modeled as $\exp(-\varphi{*}[\mathrm{EP}_m - \mathrm{EP}_{m-1}])$, where φ has to be estimated (Nunez-Anton and Zimmerman, 2000). Definition of both models resulted in sire variances estimated as functions of the EP. All analyses were performed with ASREML (Gilmour et al., 2002). For the SAD models, the OWN function in ASREML was used combined with a program available from the second author.

RESULTS

Estimated sire variances across environments were close to the simulated variances for both models. For the situation with constant simulated variance across environments, the estimates of the SAD model were however closer to the simulated variances (Figure 1) than those of the RRM. Estimated genetic correlations of the trait with itself across environments of the RRM and SAD model were compared to simulated genetic correlations (Table 1). For the simulations with structured variance, genetic correlations were overestimated by both models, but the overestimation was larger with the RRM than with the SAD. For the simulations with variances following a random regression, the genetic correlations were more closely estimated to the simulated values with the RRM and overestimated with the SAD model.

Correlations between simulated and estimated breeding values were calculated within groups of environments (Table 2). For the simulations with structured variance, correlations between simulated and estimated breeding values were slightly higher for the SAD model than for the RRM. For the simulations based on random regression, there were hardly any differences between the SAD model and RRM.

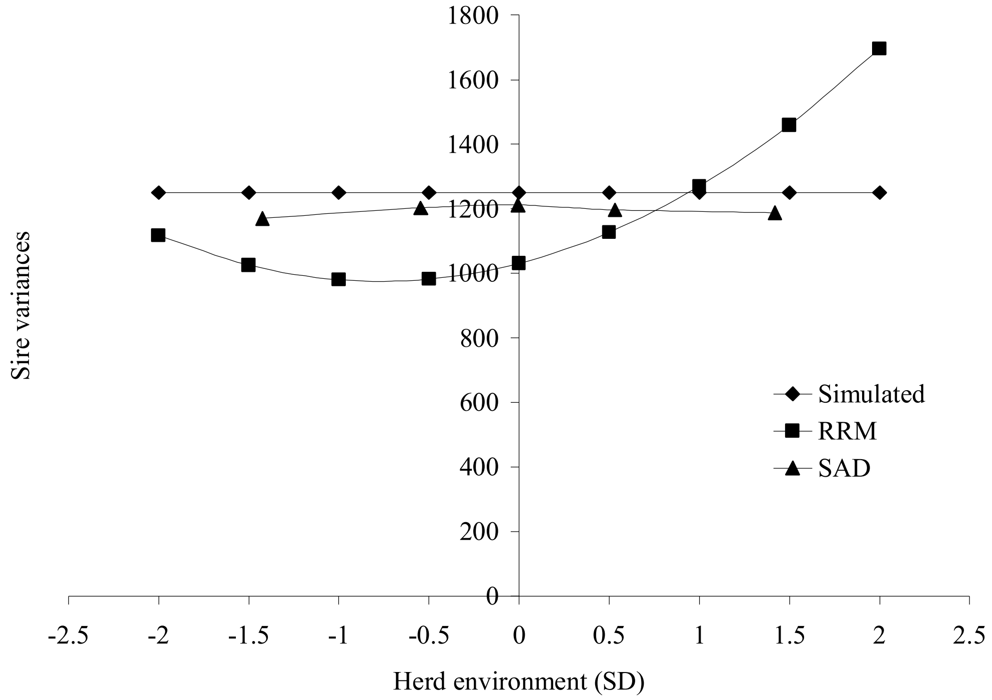

Figure 1. Estimated sire variances across environments with a RRM and SAD model, based on simulations with constant sire variance across environments. The triangles mark the estimates in each of the five discrete environments for the SAD model.

Table 1. Simulated and estimated genetic correlations across groups of environments, based on simulations with structured variances (increasing or constant across environments) or variances based on a random regression (SD ranged from 0.001 to 0.042).

Env		Structured variance (increasing)			Structured variance (constant)			Random regression		
1	2	Simulated	RRM	SAD	Simulated	RRM	SAD	Simulated	RRM	SAD
2	1	0.90	0.96	0.89	0.90	0.97	0.89	0.93	0.93	0.92
3	1	0.81	0.91	0.84	0.81	0.92	0.83	0.85	0.84	0.88
3	2	0.90	0.99	0.94	0.90	0.99	0.94	0.98	0.98	0.96
4	1	0.73	0.83	0.79	0.73	0.85	0.79	0.77	0.77	0.86
4	2	0.81	0.95	0.89	0.81	0.95	0.88	0.95	0.95	0.94
4	3	0.90	0.99	0.95	0.90	0.99	0.94	0.99	0.99	0.98
5	1	0.66	0.70	0.72	0.66	0.70	0.71	0.66	0.66	0.85
5	2	0.73	0.86	0.81	0.73	0.85	0.80	0.89	0.89	0.92
5	3	0.81	0.93	0.86	0.81	0.92	0.86	0.96	0.96	0.96
5	4	0.90	0.98	0.91	0.90	0.97	0.91	0.99	0.99	0.98

Table 2. Correlation between simulated and estimated breeding values (with RRM and SAD models) in a certain group of environments (SD ranged from 0.001 to 0.032).

	Structured variance (increasing)		Structured variance (constant)		Random regression	
Environment	RRM	SAD	RRM	SAD	RRM	SAD
1	0.96	0.97	0.97	0.98	0.98	0.98
2	0.95	0.97	0.97	0.98	0.99	0.99
3	0.95	0.97	0.95	0.98	1.00	0.99
4	0.96	0.98	0.96	0.97	1.00	0.99
5	0.96	0.98	0.96	0.97	0.99	0.99

DISCUSSION and CONCLUSION

Based on the estimated genetic correlations, both RRM (consistent with results of Calus et al. (2004)), and the SAD underestimated genotype by environment interaction. Estimates of genetic correlations across environments of both the RRM and SAD model seemed hardly influenced by the sire variances in the environments, as they were quite similar based on the simulations with structured variance that was either increasing or constant across environments. For the simulations with constant variance across environments, the SAD model did estimate parameters more closely to simulated values than the RRM. However, results of the RRM might be better when using a higher order random regression.

Based on the correlations between simulated and estimated breeding values, SAD models seem to predict breeding values slightly more accurate than RRM. This together with the observation that estimated genetic correlations across environments were less overestimated with the SAD model than the RRM, supports the idea that SAD models, even of a low order, are more flexible than RRM (Jaffrezic et al., 2004).

ACKNOWLEDGEMENTS

This study was financially supported by the Ministry of Agriculture, Nature and Food Quality (Programma 414 “Maatchappelijk verantwoorde veehouderij”).

REFERENCES

Calus, M. P. L., P. Bijma, and R. F. Veerkamp. 2004. Effects of data structure on the estimation of covariance functions to describe genotype by environment interactions in a reaction norm model. Genet. Sel. Evol. 36:489-507.

Calus, M. P. L. and R. F. Veerkamp. 2003. Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model. J. Dairy Sci. 86:3756-3764.

Gilmour, A. R., B. R. Cullis, S. J. Welham, and R. Thompson. 2002. ASREML Reference Manual. New South Wales Agriculture, Orange Agricultural Institute, Orange, NSW, Australia.

Jaffrezic, F., R. Thompson, and W. G. Hill. 2003. Structured antedependence models for genetic analysis of repeated measures on multiple quantitative traits. Gen. Res. 82:55-65.

Jaffrezic, F., E. Venot, D. Laloë, A. Vinet, and G. Renand. 2004. Use of structured antedependence models for the genetic analysis of growth curves. J. Anim. Sci. 82:3465-3473.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124:979-993.

Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. Acta Agric. Scand. Sect. A-Anim. Sci. 52:11-24.

Nunez-Anton, V. and D. L. Zimmerman. 2000. Modeling nonstationary longitudinal data. Biometrics. 56:699-705.

Veerkamp, R. F. and M. E. Goddard. 1998. Covariance functions across herd production levels for test day records on milk, fat, and protein yields. J. Dairy Sci. 81:1690-1701.