

Genotype By Environment Interaction And Prediction Of Genetic Merit

F.F. Cardoso^{*}, M. M. Alencar[†], R. J. Tempelman[§]

Introduction

Genotype by environment interaction ($G \times E$) implies that different genotypes will respond differentially to environmental changes, possibly leading to re-rankings of performance in different environments (Falconer and Mackay, 1996; Lynch and Walsh, 1998). From a selection standpoint, this means that the animals with the best genetic merit in a given environment will not necessarily be the top performers or have the top offspring in a different environment.

In some genetic evaluation models, $G \times E$ has been specified by defining the phenotypic expression of the same measure of performance in different environments as different traits (Falconer, 1952). More recently, this approach has been generalized to an infinite number of environments (or traits) through the use of covariance functions (Kirkpatrick *et al.*, 1990), thus allowing the prediction of reaction norms or environmental sensitivities of animals to gradual variations in the environment (Falconer, 1990; de Jong, 1995).

Extensive evidence of $G \times E$ in several species for livestock production (e.g., Lee and Bertrand, 2002; Pollott and Greeff, 2004; Knap and Su, 2008; e.g., Hammami *et al.*, 2009) highlight the need to consider this effect when predicting genetic merit of populations raised under diverse environmental conditions. The objectives of this work were to review methods for genetic evaluation under the presence of $G \times E$ interaction, to assess evidence of $G \times E$ interaction in livestock performance and to illustrate the prediction of environmentally specific genetic merit using beef cattle data from Brazil.

Material and methods

Statistical Models. The fundamental equation used to consider $G \times E$ in livestock performance is given by (Falconer and Mackay, 1996): $P = G + E + G \times E$, where the phenotype (P) is modeled as function of the genotype (G), the environment (E) and their interaction ($G \times E$). The three main approaches to model $G \times E$ in quantitative genetics have been: 1) to include a sire by environment interaction based on a sire model; 2) to treat the performance in each environment as a different trait using a sire or animal multitrait model, and 3) to use a random regression model specifying the genotype as a continuous function of the environment, *i.e.*, the reaction norms model. Even though the first approach has been historically useful to quantify $G \times E$ (e.g., Bertrand *et al.*, 1985; Notter *et al.*, 1992), its utility to predict environment specific genetic merit is restricted to sire \times environment combinations observed in the data. This is a major limitation given the unbalanced nature of

^{*}Embrapa Southern Region Animal Husbandry, Bagé, Brazil; [†]Embrapa Southeastern Region Animal Husbandry, São Carlos, Brazil; [§]Michigan State University, East Lansing, USA.

most livestock performance records. The other two procedures are more commonly used to predict genetic merit specific to each environment under a broader range of models. The multitrait model is more appropriate to applications where there is limited number of environmental sub-classes such as, for example, countries in international evaluations (Schaeffer, 1994; Costa *et al.*, 2000; de Mattos *et al.*, 2000) or geographical areas in national evaluations (Ferreira *et al.*, 2001; Bryant *et al.*, 2007). In this case, if the genetic correlations between the same trait in different environments are significantly different from unity, $G \times E$ is inferred such that the trait is controlled by partially different sets of genes in each environment (Falconer, 1952) and genetic merit is specific to each environmental sub-class. On the other hand, the reaction norms are obtained by random regression of genotypes performance on the average production level observed in each environment (Calus *et al.*, 2002; Kolmodin *et al.*, 2002) or other proper environmental descriptors (Fikse *et al.*, 2003; Calus *et al.*, 2005; Bohmanova *et al.*, 2008). Even though some modifications may be required for particular implementations such as the multicountry dairy evaluation (Schaeffer, 1994), the multitrait model is well established and straightforward; therefore, we focus our methods presentation hereinafter on the third approach, the reaction norms obtained by random regression. This approach has been widely used to investigate $G \times E$ interaction in livestock production during the last few years, including dairy cattle (Kolmodin *et al.*, 2002; Calus *et al.*, 2006; Strandberg *et al.*, 2009), beef cattle (Correa *et al.*, 2009; Pegolo *et al.*, 2009), sheep (Pollott and Greeff, 2004) and swine (Knap and Su, 2008).

Linear reaction norms models. Consider the following animal model (AM):

$$y_{ij} = \mathbf{x}'_j \boldsymbol{\beta} + \mathbf{w}'_j \mathbf{h} + a_j + e_{ij}, \quad (0)$$

where y_{ij} is a phenotypic record of animal j in the environment i ; $\boldsymbol{\beta}$ is the vector of fixed effects; \mathbf{h} is the vector of random environmental effects; \mathbf{x}'_j and \mathbf{w}'_j are row incidence vectors; a_j is the j th animal additive genetic effect and e_{ij} is the residual term. To obtain the reaction norm or environmental sensitivity of animal j , its genetic value must be expressed as a function of the environment (Falconer, 1990). The most commonly used approach is based on a two step procedure, first obtaining estimates of the environmental variable and then using these estimates as values of a "known" covariate in a linear random regression model (Kolmodin *et al.*, 2002). Let \hat{h}_i be a solution to i th element of \mathbf{h} , which can be used as a covariate in the following reaction norms model:

$$y_{ij} = \mathbf{x}'_j \boldsymbol{\beta} + \phi \hat{h}_i + a_j + b_j \hat{h}_i + e_{ij}. \quad (1)$$

Here a_j corresponds to the random additive genetic intercept or reaction norm level of animal j , and b_j is random additive linear regression coefficient or reaction norm slope of animal j on the environmental level represented by \hat{h}_i . Moreover, ϕ is a fixed regression coefficient of y_{ij} on \hat{h}_i . The specification of the reaction norms model in Equation (1) has a linear form and parameters can be estimated using standard likelihood or Markov Chain Monte Carlo methods (MCMC) procedures. Nevertheless, this specification has the limitation of using a function of the data, \hat{h}_i , in place of an actually unknown covariate, h_i , in the sampling model. These may lead to an understatement of the uncertainty by treating environmental variables as known and may also lead to biased genetic merit prediction, for example due to different

genetic means across herds and to genetic trends (Su *et al.*, 2006). To overcome this limitation, Su *et al.* (2006) proposed the following one step model:

$$y_{ij} = \mathbf{x}'_j \boldsymbol{\beta} + h_i + a_j + b_j h_i + e_{ij}. \quad (2)$$

In this nonlinear model all unknowns, $\boldsymbol{\beta}$, h_i , a_j and b_j , are jointly estimated using a Bayesian approach. If conjugated prior specifications are adopted, this model can be implemented via Gibbs sampling, first sampling the environmental variables h_i 's using adjusted records

$y_{(h)ij} = y_{ij} - \mathbf{x}'_j \boldsymbol{\beta} - a_j = (1 + b_j) h_i + e_{ij}$ and then obtain all other parameters conditioning on sampled values of h_i 's from $y_{(\theta)ij} = y_{ij} - h_i = \mathbf{x}'_j \boldsymbol{\beta} + a_j + b_j h_i + e_{ij}$ (Su *et al.*, 2006).

Residual environmental heteroskedasticity. It is reasonable to expect a scale effect on residual variances of reaction norms models, whereby larger environmental values would be associated with larger environmental variability (Falconer and Mackay, 1996). Therefore, we propose an extension of model (2) to allow for residual variances that are specific to each environmental level, *i.e.*:

$$y_{ij} = \mathbf{x}'_j \boldsymbol{\beta} + h_i + a_j + b_j h_i + e_{ij}, \text{ with } e_{ij} \sim N\left(0, \sigma_{e_i}^2\right) \text{ for all } i. \quad (3)$$

Here $\sigma_{e_i}^2 = \sigma_e^2 \times \eta^{h_i}$, where σ_e^2 is a reference residual variance and η is a residual heteroskedasticity parameter, following the structural model proposed by Cardoso *et al.* (2005). Note that this multiplicative model is equivalent to an additive model relating the environmental level h_i to the log of $\sigma_{e_i}^2$ by $\log \sigma_{e_i}^2 = \log \sigma_e^2 + h_i \times \log \eta$.

Angus data. To illustrate the utility of the reaction norms models to predict genetic merit under G \times E, we studied post weaning gain adjusted for a 345 days period (PWG) of 63,098 Angus calves. These data was collected on herds taking part of the Brazilian Angus Improvement Program run by the National Breeders Association "Herd Book Collares" from 1974 to 2005. There were 95,896 animals in the pedigree file.

Bayesian inference. All models were implemented using MCMC. The prior distributions of the location parameters were uniform for $\boldsymbol{\beta}$ and multivariate normal for \mathbf{h} and for the reaction norms intercepts and slopes. Moreover, we adopted inverted gamma distributions as priors for the environmental and residual variances and for the heteroskedasticity parameter. Furthermore, an inverted Wishart prior distribution was presumed for the additive reaction norms covariance matrix. Gibbs sampling was used to obtain samples of all parameters in Models (0), (1), and (2), but Metropolis-Hastings steps were required to sample from non-recognizable full conditional densities of environmental variables h_i 's and heterogeneity parameter η in Model (3). All MCMC chains were implemented using Intergen software (Cardoso, 2008) with 100,000 cycles after 50,000 cycles of burn-in and their convergence was assessed by the Geweke's Z criterion (Geweke, 1992). Model fit comparisons were based on the Deviance Information Criterion – DIC (Spiegelhalter *et al.*, 2002).

Sire environmental sensitivity. The genetic value of sire j specific to a given environment H can be obtained from models 1, 2 and 3 by $g_j | H = a_j + b_j H$. For values of b_j close to zero, g_j will be relatively constant on H and sire j is said to be a robust genotype to

environmental changes, whereas plastic genotypes are those that substantially change their performance on the environmental gradient. The larger the departure of b_j from zero, the greater the environmental sensitivity (de Jong, 1995). The consequences of G×E on selection across different environments were assessed comparing the slope of the reaction norms of these sires according to their origin, North America (NA) or Brazil (BR), and by rank correlations of genetic prediction for all animals in low, medium and high environmental levels, which were respectively defined by the values representative of the first quartile, median and third quartile of the environmental effects distribution.

Results and discussion

Model comparison. Model fit and convergence results on Angus PWG data are shown in table 1. The Geweke Z criterion was not significant when applied to the Bayesian deviance of all four models, thereby suggesting convergence of their MCMC chains to the stationary posterior distributions. The animal model showed a poorer fit when compared to the three alternative reaction norms specifications, which account for $G \times E$ and genetic heterogeneity of variance on the environmental gradient. Among the reaction norms models, model (2), being the unknown covariate model with homogeneous residual variances, had the best fit followed by models (1) and (3). Jointly estimating all unknown was previously shown to be a better approach than using environmental means to regress genetic effects in reaction norms models (Su *et al.*, 2006). The heteroskedastic model (3) was surprisingly poorer than its homoskedastic counterparts, models (1) and (2). Further research may be needed to identify structural functions other than the one specified in model (3), since other investigators have found heteroskedastic residual variances with reaction norms applications (Calus *et al.*, 2002; Pollott and Greeff, 2004; Knap and Su, 2008) and there is evidence that residual variances show a linear increase with increasing herd production levels (Kolmodin *et al.*, 2002).

Table 1: Average deviance, penalty for effective number of parameter, deviance information criterion (DIC), Geweke convergence Z statistics and p-value obtained for different models¹ applied to Angus data

<i>Model</i>	<i>Deviance</i>	<i>Penalty</i>	<i>DIC</i>	<i>Geweke Z</i>	<i>P-value Z</i>
(0)	581,151	8,270	589,421	0.15	0.8828
(1)	542,509	26,068	568,576	1.15	0.2487
(2)	539,380	22,640	562,020	0.96	0.3346
(3)	561,929	14,424	576,353	-1.08	0.2786

¹Model (0) = Animal Model; Model (1) = Reaction norms on estimated environmental levels; Model (2)= Reaction norms with unknown environmental levels; Model (3) = Reaction norms with unknown environmental levels and residual heteroskedasticity.

The environmental gradient for Angus PWG estimated by model (0) ranged from -90 to +240 kg, with little difference between models (Pearson correlation > 0.99). Heritabilities were lower at harsh environments and increased with the improvement of the production conditions for all reaction norms models (Figure 1). Models (1) and (2) showed larger estimated heritabilities on positive environmental values than Model (3) for which residual variance was also proportional to the environmental gradient. When compared to the traditional animal model, all reaction norms model allocated a greater proportion of the

phenotypic variation to genetic components as opposed to unaccounted environmental factors. Consequently, a greater selection response could be expected, especially in the best environments (Kolmodin *et al.*, 2003).

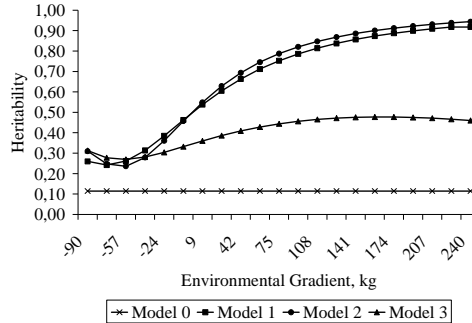


Figure 1: Posterior mean heritabilities as a function of the environmental gradient obtained by different models: (0) animal, (1) reaction norms on estimated environmental levels, (2) reaction norms with unknown covariate, and (3) heteroskedastic residual reaction norms

Genetic merit predictions. Rank correlations among genetic merit predictions obtained by the traditional animal model (0) and the best fitting reaction norms model with unknown covariates (2) under low or high environmental levels (table 2) were below the suggestive biological importance threshold of 0.80 (Robertson, 1959), pointing out substantial $G \times E$ on Brazilian Angus PWG. This was also observed by the crossing of bulls reaction norms (Figure 2). The average environmental sensitivity of the NA bulls (0.077 ± 0.011 kg) was not significantly different from BR bulls with NA sires (0.046 ± 0.019 kg), but exceeded that for BR bulls with local sires (0.011 ± 0.027 kg), which in turn were more robust to environmental changes.

Table 2: Spearman rank correlations among posterior mean genetic values for post-weaning gain of Angus cattle obtained by an animal model and by a reaction norms with unknown covariates at different environmental levels

<i>Model (environmental level)</i>	<i>Model 2 (low)</i>	<i>Model 2 (average)</i>	<i>Model 2 (high)</i>
Model 0 (overall)	0.76	0.80	0.78
Model 2 (low)		0.88	0.81
Model 2 (average)			0.99

Accounting for $G \times E$ in genetic evaluations. The benefits of considering $G \times E$ on genetic evaluations will depend on the livestock specie, on the breeding and genetic structure of the target population and on the observed environmental diversity. Even though some evidence of $G \times E$ has been generated for poultry (Mathur and Horst, 1994; Deeb and Cahaner, 2001) and swine (Knap and Su, 2008), this should not be a main concern issue for these species due to the closed company based structure and highly uniform and controlled conditions of most breeding programs. Cattle and small ruminant populations, especially those raised under

extensive grazing conditions, have greater potential to enhance selection progress by having environment-specific genetic predictions. In this case, seedstock herds owned by individual producers are often combined into large scale genetic evaluations as run by breed associations or international consortiums. The dairy sires' international evaluation (Interbull) is the most noticeable genetic evaluation program that accounts for $G \times E$. On the other hand, a similar initiative for beef cattle, the Pan American Hereford evaluation (Breedplan), assumes absence of $G \times E$ (de Mattos *et al.*, 2000). Nevertheless, several studies with beef breeds under grazing conditions in Brazil (de Alencar *et al.*, 2005; Cardoso, 2009; Correa *et al.*, 2009; Mattar, 2009) demonstrate that genetics suitable for harsh environments is generally quite different from that for average or superior conditions as evident from low genetic correlations between environmental gradients extremes (Fig. 3). These breeds and other similar populations could benefit from implementing reaction norms predictions to select for robust animals or to optimize breeding programs that select seedstock for multiple environments (de Jong and Bijma, 2002; Kolmodin and Bijma, 2004). Safer exchange of genetics across countries or regions could also be achieved by adjusting genetic predictions from the exporting evaluation place by the $G \times E$ with the production conditions at the importing location (Costa *et al.*, 2000; Montaldo *et al.*, 2009).

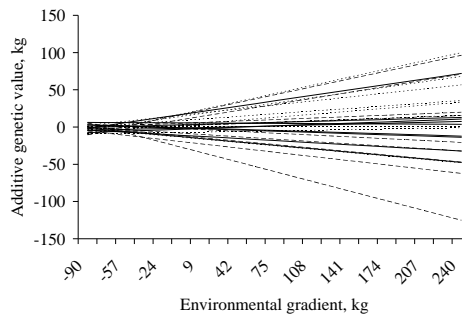


Figure 2: Reaction norms for a random sample of 27 bulls over the post-weaning gain environmental gradient.

Gene by environment interaction. As genomic selection becomes a reality for livestock improvement (Hayes *et al.*, 2009), it may become necessary to consider $G \times E$ at a gene (or marker) by environment level for genetic merit prediction. Reaction norms models for assessing gene by environment interaction are available (Lillehammer *et al.*, 2007) and have been applied to detect environmentally dependent QTL effects on milk yield (Lillehammer *et al.*, 2008), although further work is warranted in this subject.

Conclusion

Evidence of genotype by environment interaction in several livestock species warrant the opportunity to optimize selection progress using environment specific genetic merit predictions. These predictions can be obtained in practice by multitrait models when there is limited number of environmental sub-classes or, preferably, by more general reaction norms models when proper continuous environmental descriptors are available.

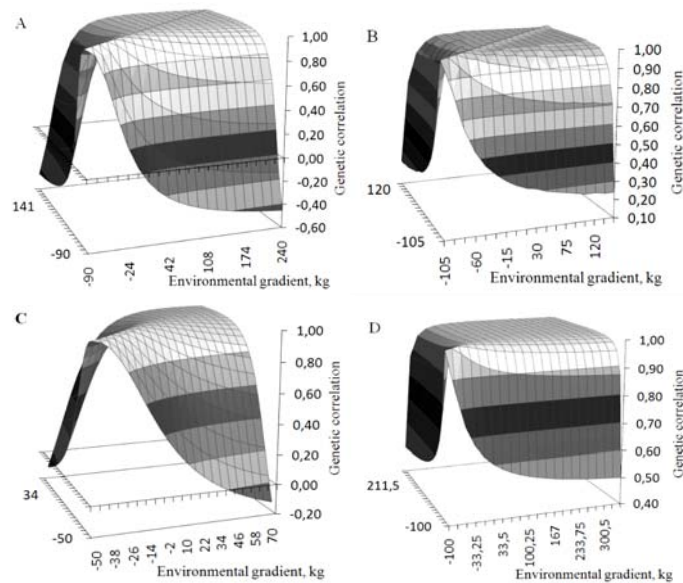


Figure 3: Genetic correlation for post-weaning performance on different environmental conditions for different breeds in Brazil: A = Angus (A), B = Canchim 3/8 Zebu 5/8 Charolais (Mattar, 2009), C = Devon (Correa *et al.*, 2009) and D = Hereford (Cardoso, 2009).

References

- Bertrand, J., Berger, P., and Willham, R. (1985). *J. Anim. Sci.* 60: 1396-1402.
- Bohmanova, J., Misztal, S., Tsuruta, H. *et al.* (2008). *J. Dairy Sci.* 91: 840-846.
- Bryant, J., Lopez-Villalobos, N., Pryce, J. E., *et al.* (2007). *J. Dairy Sci.* 90: 1538-1547.
- Calus, M., Groen, A., and de Jong, G. (2002). *J. Dairy Sci.* 85: 3115-3123.
- Calus, M., Janss, L., and Veerkamp, R. (2006). *J. Dairy Sci.* 89: 4846-4857.
- Calus, M., Windig, J., and Veerkamp, R. (2005). *J. Dairy Sci.* 88: 2178-2189.
- Cardoso, F., (2008). *Intergen Manual version 1.01. Docs 74, Embrapa Pecuária Sul.*
- Cardoso, F., Rosa, G., and Tempelman, R. (2005). *J. Anim. Sci.* 83: 1766-1779.
- Cardoso, L. (2009). *Master's thesis.* UFRGS.
- Correa, M., Dionello, N., and Cardoso, F. (2009). *Rev. Bras. Zootec.* 38: 1468-1477.
- Costa, C., Blake, R. W., Pollak, E. J., *et al.* (2000). *J. Dairy Sci.* 83: 2963-2974.
- de Alencar, M., Mascioli, A., and de Freitas, A. (2005). *Rev. Bras. Zootec.* 34: 489-495.
- de Jong, G. (1995). *Am. Nat.* 145: 493.
- de Jong, G., and Bijma, P. (2002). *Livest. Prod. Sci.* 78: 195-214.

- de Mattos, D., Bertrand, J., and Misztal, I. (2000). *J. Anim. Sci.* 78: 2121-2126.
- Deeb, N., and Cahaner, A. (2001). *Poultry Sci.* 80: 695-702.
- Falconer, D., (1952). *Am. Nat.* 86: 293-298.
- Falconer, D., (1990). *Genet. Res.* 6: 57-70.
- Falconer, D., Mackay, T. (1996). *Introduction to quantitative genetics*. Longman Group Ltd.
- Ferreira, V., Penna, V., Bergmann, J., *et al.* (2001). *Arq. Bras. Med. Vet. Zootec.* 53:385-392.
- Fikse, W., Rekaya, R., and Weigel, K. (2003). *Livest. Prod. Sci.* 82: 223-231.
- Geweke, J. (1992). In *Bayesian statistics*. Oxford University Press.
- Hammami, H., Rekik, B., Gengler, N. (2009). *Biotechnol. Agron. Soc. Environ.* 13: 155-164.
- Hayes, B., Bowman, P., Chamberlain, A., *et al.* (2009). *J. Dairy Sci.* 92: 1313-1313.
- Kirkpatrick, M., Lofsvold, D., and Bulmer, M. (1990). *Genetics* 124: 979-993.
- Knap, P., and Su, G. (2008). *Animal* 2: 1742-1747.
- Kolmodin, R., and Bijma, P. (2004). *Genet. Sel. Evol.* 36: 435-454.
- Kolmodin, R., Strandberg, E., Jorjani, H., *et al.* (2003). *Anim. Sci.* 76: 375-385.
- Kolmodin, R., Strandberg, E., Madsen, P., *et al.* (2002). *Acta Agric. Scand. Section A-Anim. Sci.* 52: 11-24.
- Lee, D., and Bertrand, J., (2002). *J. Anim. Sci.* 80: 330-337.
- Lillehammer, M., Goddard, M. E., Nilsen, H., *et al.* (2008). *Genetics* 179: 1539-1546.
- Lillehammer, M., Odegard, J., and Meuwissen, T. (2007). *Genet. Sel. Evol.* 39: 105-121.
- Lynch, M., Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Ass., Inc.
- Mathur, P., and Horst, P. (1994). *Poultry Sci.* 73: 1777-1784.
- Mattar, M. (2009). *Doctoral dissertation*. Unesp.
- Montaldo, H., Nunez, S., Ruiz, F., *et al.* (2009). *J. Dairy Sci.* 92: 5270-5275.
- Notter, D., Tier, B., and Meyer, K. (1992). *J. Anim. Sci.* 70: 2359-2365.
- Pegolo, N., Oliveira, H., Albuquerque, L., *et al.* (2009). *Genet. Mol. Biol.* 32: 281-287.
- Pollott, G., and Greeff, J. (2004). *J. Anim. Sci.* 82: 2840-2851.
- Robertson, A. (1959). *Biometrics* 15: 469-485.
- Schaeffer, L. (1994). *J. Dairy Sci.* 77: 2671-2678.
- Spiegelhalter, D., Best, N., Carlin, B., *et al.* (2002). *J. Roy. Stat. Soc. B Met.* 64: 583-616.
- Strandberg, E., Brotherstone, S., Wall, E., *et al.* (2009). *J. Dairy Sci.* 92: 3437-3446.
- Su, G., Madsen, P., Lund, M. S., *et al.* (2006). *J. Anim. Sci.* 84: 1651-1657.