Tree Genetics & Genomes
DOI 10.1007/s11295-013-0648-2

ORIGINAL PAPER

An assessor-specific Bayesian multi-threshold mixed model for analyzing ordered categorical traits in tree breeding

Eduardo P. Cappa · Luis Varona

Received: 10 December 2012 / Revised: 30 May 2013 / Accepted: 12 June 2013

Abstract Many traits of biological interest in tree breeding are assessed using more than two ordered discrete categories. These scores have a more or less arbitrary and subjective assignment by the assessors, which could lead to a strong departure from the Gaussian distribution. Different assessors may also use different regions of the available scale. This study describes the use of the multi-threshold mixed model proposed by Varona et al. (J Anim Sci 87:1210–1217, 2009), which allows different thresholds for different assessors on an underlying Gaussian distribution. This method was applied to a six-point score for stem quality in an open-pollinated progeny trial of *Prosopis alba* Griseb. Four mixed models were used: (1) a linear mixed model with observed score (LMM); (2) a linear mixed model with transformed "normal scores" (LMM_NS); (3) a threshold mixed model (TMM); and (4) an assessor-specific multi-threshold mixed model (MTMM). Dispersion parameters were estimated using Bayesian techniques via the Gibbs sampling with a data augmentation step. The proposed MTMM produced higher posterior mean heritabilities (0.096) than the commonly used LMM (0.077). Posterior mean heritabilities from LMM_NS (0.094) and TMM (0.097) were comparable to those obtained using MTMM; however, MTMM yielded slightly more precise estimates than TMM. Although correlations of the estimated breeding values were high between different models (from 0.88 to 0.99), the heterogeneity in the estimated posterior means of the thresholds between the three assessors caused notable changes in the top 10 families between TMM and MTMM. The proposed model is helpful in fitting subjective ordered categorical traits assessed by different assessors in tree breeding.

Keywords Ordered categorical traits · Assessor · Multi-threshold mixed model · Bayesian inference · Tree breeding

Introduction

Many traits assumed to be under polygenic control and of biological interest or economic importance in tree breeding are visually assessed along some hypothetical gradient using more than two ordered discrete categories. Examples of these ordered categorical traits include stem straightness (Bannister 1979; Gwaze et al. 2001; Wu and Matheson 2005); crown form integrated into one score (Raymond and Cotterill 1990) or a separate visual score assigned for each component: branch thickness (Hai et al. 2008), branch angle (Hannrup et al. 2003), branch whorls, ramicorn branches, and forks; resistance to insect and diseases (Yanchuk et al. 1988; Wu et al. 1996); cold acclimation (Wei et al. 2001); frost damage (Codesido and Fernández-López 2009); drought damage (Dutkowski and Potts 2012); foliage color and foliage density (Zas et al. 2008); and flowering intensity (Cané-Retamales et al. 2010). Statistical analysis of these traits with standard linear mixed models (LMM) (Henderson 1984) is the most common method in forest genetic evaluation (e.g., Gwaze et al. 2001; Wei et al. 2001; Wu and Matheson 2005; Codesido and Fernández-López 2009; Dutkowski and Potts 2012).

Communicated by R. Burdon

E. P. Cappa
Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Recursos Biológicos, De Los Reseros y Dr. Nicolás Repetto s/n, 1686 Hurlingham, Buenos Aires, Argentina

E. P. Cappa
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

L. Varona
Unidad de Genética Cuantitativa y Mejora Animal, Universidad de Zaragoza, 50013 Zaragoza, Spain

E. P. Cappa (✉)
Bosques Cultivados, Instituto de Recursos Biológicos, Centro de Investigación en Recursos Naturales, Instituto Nacional de Tecnología Agropecuaria, De Los Reseros y Dr. Nicolás Repetto s/n, 1686 Hurlingham, Buenos Aires, Argentina
e-mail: ecappa@cnia.inta.gov.ar

However, the visual score representing a phenotype expression in a predefined scale is usually arbitrary and subjective, and the output could lead to a strong departure from the Gaussian distribution (Gianola and Foulley 1983). Therefore, statistical genetic analyses based on these mixed models are generally unsuitable for such traits (Gianola 1980).

Alternatively, generalized linear models (Agresti 2007) account for the discrete distribution of the data assuming an underlying continuous distribution (liability). The observed response category is a function of the liability and fixed unknown thresholds (Wright 1934; Dempster and Lerner 1950; Gianola 1982; Gianola and Foulley 1983); the observed categorical phenotype will be expressed when the underlying value is contained in the region defined by the thresholds for that value. The most common liability distributions are the Gaussian (probit approach) and logistic (logit approach). However, biological and statistical reasons favor the use of the Gaussian approach for modeling ordered categorical data in genetic evaluations. From a biological perspective, the parameters are easier to interpret and considered the result of a linear combination of small effects stemming from alleles at a large number of loci, plus random environmental components (Gianola and Foulley 1983). From a statistical perspective, it can be easily implemented (Albert and Chib 1993; McCulloch 1994). The generalized linear mixed models with a Gaussian underlying liability distribution (threshold mixed model (TMM)) have been used in tree breeding for binary traits (e.g., Steane et al. 2006; Li et al. 2006; Cappa et al. 2010; Hamilton et al. 2010) and to a lesser extent in categorical traits (e.g., Cané-Retamales et al. 2010).

Another technique to account for discrete distribution of categorical data and departure from the Gaussian distribution is transforming the data into “normal score” i.e., “optimal score” (Gianola and Norton 1981), before estimating variance components based on the linear mixed model (LMM_NS). This method also assumes an underlying normal distribution of the phenotype and is frequently called “normal score transformation” in the literature. Transformation of categorical data to normal score is routinely employed in the Swedish tree breeding program (e.g., Ericsson et al. 1994; Hannerz et al. 1999; Högberg et al. 2010). Comparisons between LMM_NS and the traditional LMM and TMM of more than two ordered discrete categories from forest genetic trials are limited.

In addition to the discrete distribution of these traits, there are other problems related to these traits. Despite consensus about the subjective score for each tree, as recommended by Cotterill and Dean (1990), different assessors tend to use different regions of the available scale during their observations (Bannister 1979). For example, each individual assessor may use a wider or narrower range of values (Cotterill and Dean 1990). Moreover, the assessor's expertise influences the assessed visual score (Cotterill and Dean 1990; Sierra de Grado et al. 1999).

To account for both the discrete distribution of these traits and the different categorization patterns of each assessor, Varona and Hernández (2006) proposed to fit a set of specific thresholds for each assessor (i.e., panelist) for sensory analysis of panel data, despite a common underlying Gaussian distribution to all panelists. Varona et al. (2009) extended this multi-threshold model to a mixed model approach in an animal breeding context. This assessor-specific multi-threshold mixed model (MTMM) has not been applied to forest genetics yet.

A serious limitation of parameter estimation under threshold and multi-threshold mixed models is that the likelihood or marginal posterior distributions do not have close forms and some approximations have to be used (Wang et al. 1997). Albert and Chib (1993) showed that Markov Chain Monte Carlo (MCMC) Bayesian methods via Gibbs sampling, in conjunction with data augmentation (Tanner and Wong 1987), lead to fully posterior distributions that are easy to sample from. Sorensen et al. (1995) applied this approach to estimate genetic parameters of a threshold mixed model in an animal breeding context with more than two ordered discrete categories. Recently, Bayesian inference using MCMC algorithms has become available for estimating the parameters from a LMM in tree breeding (e.g., Cappa and Cantet 2006; Cappa and Cantet 2008; Waldmann et al. 2008; Apiolaza et al. 2010). However, their use for estimating the parameters of ordered categorical traits from multi-threshold mixed model has not been reported in forest breeding programs.

The objective of this study was to apply the assessor-specific multi-threshold mixed model proposed by Varona et al. (2009) to estimate variance components (or function of them) via Gibbs sampling with a data augmentation step, from mixed model for data with more than two ordered discrete categories from forest genetic trials. The method was applied to data from a six-point subjective score for stem quality trait recorded on a progeny trial of *Prosopis alba* Griseb (hereafter *P. alba*). The resulting estimates of all dispersion parameters and predicted breeding values for the proposed MTMM are compared with those from the conventional LMM, LMM_NS, and TMM.

Materials and methods

Data

As part of a broad *P. alba* tree breeding program of the National Institute of Agricultural Technology (Instituto Nacional de Tecnología Agropecuaria, INTA, projects PNFOR-4234 and PNFOR-44341), more than 400 mature open-pollinated (OP) parent trees were phenotypically selected in wild populations. Candidate trees were healthy and in “good shape” and well developed (Verga et al. 2009). Five open-pollinated progeny trials were established in Salta, Santiago del Estero, Chaco,

Table 1 Six-point subjective score system used to assess stem quality in an open-pollinated progeny trial of *P. alba*

Score	Well-defined or not defined main stem and number and diameter of branches
1	Not defined main stem with several thick branches of similar sizes
2	Without a defined main stem with a separation of the trunk into two stems with similar sizes
3	Poorly defined main stem with a separation of the trunk into two and one of the two stems with fewer size
4	Well-defined main stem with branches of diameters equal or less than the diameter of the main stem
5	Well-defined main stem with fewer thin branches
6	Well-defined main stem without branches or fewer very thin branches

Formosa, and Entre Ríos provinces, Argentina. A progeny trial with 201 OP families of *P. alba* was used in this study. The average progeny per mother was 16 trees varying from 6 to 22. Ten genetic groups were formed according to provenances. The trial site was located in Vivero San Martin (latitude 27°94′56″ S, longitude 64°21′92″ W, elevation 174 m) southeastern Santiago del Estero province. The soil was formed from loessical silt with scarce development; their profiles being characterized by A-AC-C horizons (Galizzini et al. 1999). The experimental design was a randomized complete block, with three replicates of 1 to 10 multiple-tree non-contiguous plots, and the spacing was 4×4 m. Total height (mean, 1.94 m; standard deviation, 0.83 m), diameter at root collar (mean, 3.95 cm; standard deviation, 2.23 cm), and stem quality, the trait analyzed in this study, were assessed at the age of 28 months in the 2,723 surviving trees. The stem quality of each tree was visually assessed using a six-point subjective score, which was divided into two groups for ease of measurement in the field: scores 1, 2, and 3 were trees without a defined main stem or forking, and scores 4, 5, and 6 were trees with a well-defined main stem. However, the six categories resulted mainly from the branch diameter trait. A detailed description of the scoring system is shown in Table 1 and Fig. 1. Each tree was visually assessed by one of a total of three assessors, each person assessing alternating individual columns in a serpentine fashion. Thus, each assessor evaluated the stem quality trait on all families.

Statistical models of analysis

The stem quality trait was analyzed using the following mixed models.

1. LMM and LMM_NS

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{f} + e_i \qquad (1)$$

where y_i is the phenotypic data for each record i on the observed original scale (LMM) or the corresponding transformed normal score (Gianola and Norton 1981) (LMM_NS). The stem quality data were transformed into normal score within each block. In addition, $\boldsymbol{\beta}$ is a $p\times 1$ vector of fixed effects including blocks (3 levels), genetic groups to account for the means of the different origins of parents (10 levels), and assessors (3 levels); $\boldsymbol{f} \sim N(0, \boldsymbol{I}\sigma_f^2)$ is a $f\times 1$ random vector containing the family (genetic) effects, where σ_f^2 is the family variance; and e_i is the random error term included in $[e_1' | \ldots | e_i' | \ldots | e_n'] = \boldsymbol{e}$ and distributed as $\boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma_e^2)$, where σ_e^2 is the error variance. Furthermore, $\boldsymbol{x}'_i$ is a $1\times p$ row vector of the known incidence matrix ($\boldsymbol{X}$), which relates y_i to the elements of $\boldsymbol{\beta}$, and $\boldsymbol{z}'_i$ is a $1\times f$ row vector of the known incidence matrix ($\boldsymbol{Z}$), which relates y_i to the elements of $\boldsymbol{f}$.

The Bayesian approach via Gibbs sampling was used to estimate the marginal posterior distribution of all the parameters in model (1), following Sorensen and Gianola (2002). Conjugate prior densities were chosen for all parameters. To reflect a prior state of uncertainty for the fixed effects while keeping the posterior distribution proper, we selected $\boldsymbol{\beta} \sim N_p(\boldsymbol{0}, \boldsymbol{K})$ with $\boldsymbol{K}$ a diagonal matrix with large elements ($k_{ii} > 10^8$). For the prior distributions of σ_f^2 and σ_e^2, we used a scaled inverted chi-square. Therefore, the joint and conditional posterior densities are Gaussian for $\boldsymbol{\beta}$ and $\boldsymbol{f}$ and scaled inverted chi-square for σ_f^2 and σ_e^2.

2. TMM

In this model, the variable y_i is the expression of an underlying continuous unobserved random variable (l_i)

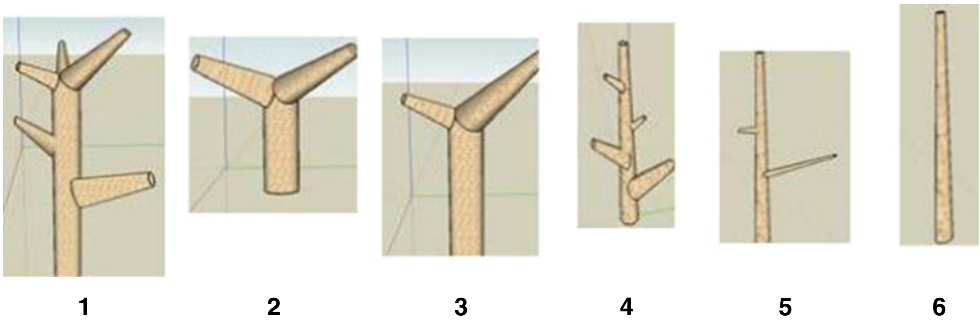

Fig. 1 Detail of the stem quality score

for each record i, called liability. Each y_i represents an assignment into one C mutually exclusive ordered categories more or less arbitrarily delimited by $C+1$ thresholds ($\boldsymbol{t}$), such that $t_{\min}<t_1<t_2<\ldots<t_{C-1}<t_{\max}$ and where $t_{\min}=-\infty$ and $t_{\max}=+\infty$. Then, in matrix notation, the TMM to analyze the liability is:

$$l_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{f} + e_i. \qquad (2)$$

In ensure identifiability of all parameters in (2), two values need to be restricted. The threshold between categories 2 and 3 was set to 0, and the residual variance was set to 1 (i.e., $\boldsymbol{e}\sim N\,(0,\boldsymbol{I})$).

We used a Bayesian approach following Sorensen et al. (1995). To facilitate the development of Gibbs sampling, the unobserved liability $\boldsymbol{l}$ was included as an unknown parameter. This approach, known as data augmentation (Tanner and Wong 1987), yields full conditional posterior distributions that have a standard form and are easy to sample from (Sorensen et al. 1995). Truncated normal prior distribution was assumed for $\boldsymbol{l}$ and uniform prior distribution in the interval $[t_{\min}, t_{\max}]$ for the thresholds $\boldsymbol{t}$. Prior and conditional posterior distributions of $\boldsymbol{\beta}$, $\boldsymbol{f}$, and σ_f^2 are the same as for LMM. Conditional posterior distributions were truncate Gaussian for the underlying variable ($\boldsymbol{l}$) and uniform in the interval $[t_{\min}, t_{\max}]$ for the thresholds.

3. MTMM

This model assumes that a random variable y_{ij} is observed for each tree j for $j=1,\ldots,n_i$ (number of trees assessed by the ith assessor) and for i assessors for $i=1,\ldots,a$ (number of assessors). Then, following the work of Varona and Hernández (2006), the variable y_{ij} is the expression of the underlying continuous unobserved liability (l_{ij}) for the record j and assessor i. Each y_{ij} represents an assignment into one C_i mutually exclusive ordered categories more or less arbitrarily delimited by C_i+1 thresholds for each assessor ($\boldsymbol{t}_i$), such that $t_{\min}<t_{i1}<t_{i2}<\ldots<t_{iC_i-1}<t_{\max}$, and where t_{ij} is the jth threshold for the ith assessor and $t_{\min}=-\infty$ and $t_{\max}=+\infty$. These thresholds for each assessor transform the liability to the observed scale. In matrix notation, the corresponding MTMM to analyze the liability is:

$$l_{ij} = x_i'\boldsymbol{\beta} + z_i'\boldsymbol{f} + e_{ij} \qquad (3)$$

The presence of a specific threshold for each assessor took into account the variation captured by the assessor effect in LMM and TMM; thus, the fixed effect of assessor was not included in the model (3). The MTMM included a vector $\boldsymbol{\beta}$ with 3 blocks and 10 genetic groups. The threshold between categories 2 and 3 was set to 0 in the assessor 3, and the residual variance was set to 1 (i.e., $\boldsymbol{e}\sim N\,(0,\boldsymbol{I})$), in order to achieve identifiability.

Again, a Bayesian approach via Gibbs sampling with a data augmentation step (Tanner and Wong 1987; Sorensen et al. 1995) was used to estimate the parameters in model (3). Appendix provides a detailed description of the prior distributions of all parameters, the likelihood of the data, the joint, and marginal conditional posterior densities for MTMM (3), as well as the Gibbs sampling algorithms.

A set of programs were written in FORTRAN to perform all calculations of the two linear mixed models and the two generalized linear mixed models. The FORTRAN programs are available from the first author on request.

Posterior inference, estimation of heritabilities, and model comparison

The values of the hypervariances for the variance components of LMM_NS, TMM, and MTMM were estimated from the same original data set using an empirical Bayes approach via Gibbs sampling, with a classical mixed model including fixed effects of blocks, genetic groups and assessors, and random family genetic effects (i.e., from LMM). The degrees of belief (υ) were set to 10 to reflect a relatively high degree of uncertainty.

A single Gibbs chain of 1,010,000 samples was drawn, and the first 10,000 iterates were discarded as *burn-in* for the four linear and generalized mixed models of analysis. The 1,000,000 additional samples were used for computing the summary from the marginal posterior distribution. Convergence was monitored by plotting the iterations against the mean of the draws up to each iteration (running mean plots) and using the Z criterion of Geweke (1992) for each parameter. Marginal posterior densities for all parameters were estimated by the Gaussian kernel method (Silverman 1986; Chap. 2). Mean, mode, median, standard deviation, and 95 % high posterior density interval (95 % HPD) were then calculated with *Bayesian Output Analysis* (version 1.1.7; Smith 2003) for all parameters from the individual marginal posteriors, under the free-software R (R Development Core Team 2011). At the end of each iteration, and assuming that all the families were maternal half sibs (Bessega et al. 2009), individual heritability in the original scale for LMM and in the transformed normal score for LMM_NS was calculated as $\widehat{h}^2 = 4\times\widetilde{\sigma}_f^2/(\widetilde{\sigma}_f^2+\widetilde{\sigma}_e^2)$ and in the underlying scale for TMM and MTMM as $\widehat{h}^2 = 4\times\widetilde{\sigma}_f^2/(\widetilde{\sigma}_f^2+1)$, where $\widetilde{\sigma}_f^2$ and $\widetilde{\sigma}_e^2$ are the values of the family and error variances sampled at a given iteration.

The estimates of all dispersion parameters were used to compare the fit between all the linear (LMM and LMM_NS) and the threshold (TMM and MTMM) mixed models.

Additionally, the Deviance Information Criterion (DIC; Spiegelhalter et al. 2002) was computed to compare the fit of the LMM, TMM, and MTMM. The DIC value was not calculated for the LMM_NS since this model used a different set of data, i.e., transformed "normal scores" instead of observed phenotype (i.e., raw data). The DIC criterion is defined as: $DIC = \overline{\mathrm{D}}(\theta_M) + p_D$, where $\overline{\mathrm{D}}$ (θ_M) is the posterior mean of the deviance and p_D the "effective number of parameters." A smaller DIC value indicates a better fit and lower degree of model complexity. Further comparisons were provided by Spearman rank correlations and Pearson product–moment correlation coefficients to compare whether the predicted breeding values differed among models.

Results

Table 2 describes the distribution of stem quality scores assigned by the three assessors and by each assessor. Clearly, it did not follow a Gaussian distribution. Table 2 also shows the heterogeneity in the distribution of the observed phenotype (i.e., raw data) across assessors. For example, assessor 1 assigned category 1 to only 7.3 % of the trees, whereas assessor 2 assigned the same category to about 39.0 %. Moreover, assessor 3 used only five of the six categories available to evaluate the stem quality. These results show that each assessor placed the trees in different regions of the six-point quality score categories and highlight the importance of accounting for differences between assessor distributions in the stem quality trait analysis.

Estimates of variances and heritabilities for the two linear (LMM and LMM_NS) and two generalized (TMM and MTMM) mixed models are summarized in Table 3. In general, posterior means, medians, and modes were quite similar for all parameters, thus indicating that the marginal posterior distributions are nearly symmetric. Estimates of marginal posterior means of σ_f^2 and h^2 were relatively similar for TMM (0.025 and 0.097, respectively) and MTMM (0.026 and 0.096, respectively). However, HPD (95 %) interval using MTMM was slightly narrower than the corresponding interval using TMM, which suggests more certainty of the estimate obtained from the model that accounted for heterogeneity between assessors. Estimated heritabilities from TMM (0.97) and MTMM (0.96) were similar to those estimated from LMM_NS (0.094). As expected from the theory (Gianola 1979), heritability estimates from threshold and multi-threshold mixed models were much higher (0.097 and 0.096, respectively) than those from LMM (0.077).

Table 2 Distribution of stem quality scores generated by all three assessors and by each assessor as a proportion of the total number of trees scored (total)

Assessor	Score						Total
	1	2	3	4	5	6	
1, 2, and 3	18.11	37.97	29.34	8.41	4.52	1.65	2,723
1	7.27	33.30	43.14	6.25	6.45	3.59	976
2	38.97	32.53	12.45	10.15	4.80	1.09	916
3	7.82	49.46	31.77	9.03	1.93	0	831

Table 3 Posterior statistics for family variance (σ_f^2), error variance (σ_e^2), and heritability (h^2) for the four linear and generalized mixed models studied

Model	Parameter	Mean	Median	Mode	SD	95 % HPD
LMM	σ_f^2	0.023	0.022	0.020	0.007	0.013–0.037
	σ_e^2	1.169	1.168	1.167	0.035	0.113–0.227
	h^2	0.077	0.074	0.066	0.025	0.043–0.123
LMM_NS	σ_f^2	0.020	0.019	0.018	0.006	0.012–0.030
	σ_e^2	0.826	0.825	0.820	0.025	0.786–0.867
	h^2	0.094	0.091	0.084	0.027	0.057–0.143
TMM	σ_f^2	0.025	0.024	0.022	0.008	0.014–0.040
	σ_e^2	1.00	–	–	–	–
	h^2	0.097	0.094	0.086	0.030	0.055–0.153
MTMM	σ_f^2	0.026	0.024	0.022	0.008	0.014–0.039
	σ_e^2	1.00	–	–	–	–
	h^2	0.096	0.092	0.086	0.030	0.054–0.151

Abbreviations used for the models were described in the text

SD standard deviation, *HPD* high posterior density interval

Table 4 shows rank and product–moment correlations between predicted breeding values of family genetic effects for stem quality from four different linear and generalized mixed models. Overall, Spearman and Pearson correlations for the stem quality trait were high (from 0.88 to 0.99), but larger between the linear models (i.e., LMM vs. LMM_NS; 0.99 for rank and product–moment correlations, respectively) than between linear and generalized mixed models (from 0.93 to 0.96), followed by the correlations between generalized models (TMM vs. MTMM; 0.88 and 0.89 for rank and product–moment correlations, respectively). Figure 2 shows a scatter plot of the predicted breeding values of family genetic effects from TMM and MTMM.

Although Spearman rank correlations were high, the ranking among families differed across models. Table 5 shows the ranking of the 10 best and 10 worst families based on MTMM for the two linear and two generalized mixed models. As shown above, the largest difference was found between TMM and MTMM. Families 64, 107, and 195 performed very well in MTMM (ranked 7, 9, and 10, respectively); however, they were not among the best 10 families in the other three models. The proportion of common families in the top 10 families (5 %) was 0.7 between MTMM and LMM, 0.7 between MTMM and LMM_NS, and 0.5 between MTMM and TMM.

Table 4 Spearman rank correlations (upper diagonal) and Pearson product–moment correlation coefficients (lower diagonal) between the breeding values predicted using linear and generalized mixed models

Models	LMM	LMM_NS	TMM	MTMM
LMM	1.0	0.988	0.947	0.931
LMM_NS	0.990	1.0	0.950	0.939
TMM	0.950	0.955	1.0	0.884
MTMM	0.929	0.939	0.886	1.0

Abbreviations for the models were described in the text. In all cases, correlations off-diagonal were highly significant from zero ($p < 0.001$)

Figure 3 presents the posterior mean estimates for five threshold parameters (assessors 1 and 2) and four threshold parameters (assessor 3), which indicates a large variation between assessors. For example, the threshold between categories 2 and 3 was set to 0 by assessor 3, whereas posterior mean estimates by assessor 1 and 2 were negative (−0.431) and positive (0.389), respectively, and their 95 % HPD did not include the 0 (−0.529 – −0.335 and 0.294–0.485 for assessors 1 and 2, respectively). Category 2 was delimited by thresholds whose posterior mean estimates were −1.673 and −0.431, −0.479 and 0.389, and −1.643 and 0 for assessors 1, 2, and 3, respectively. These estimates agree with the difference found in the observed phenotype (Table 2 and Fig. 3). For example, 43.1 and 31.8 % of the raw data were assigned to category 3 by assessors 1 and 3, respectively, whereas only 12.5 % was assigned to this category by assessor 2.

Table 5 Ranking of the 10 best and 10 worst families of *P. alba* based on the assessor-specific multi-threshold mixed model (MTMM) for the linear and threshold models studied

Rank	LMM	LMM_NS	TMM	MTMM
Best	9	9	16	1
	1	2	1	2
	2	1	4	3
	7	6	5	4
	4	3	3	5
	5	4	8	6
	16	14	53	7
	8	7	13	8
	11	19	22	9
	27	28	21	10
Worst	187	186	138	192
	172	174	143	193
	171	184	190	194
	164	172	178	195
	186	187	169	196
	193	196	193	197
	198	197	198	198
	197	198	197	199
	201	201	201	200
	200	200	200	201

Abbreviations for the models were described in the text

Discussion

This study proposed an MTMM to fit ordered categorical traits assessed by different assessors in forest genetic evaluations. The precise alignment of the estimated posterior means of the thresholds with the distribution of the individual assessor's score (Fig. 3) suggests that the proposed model is a helpful model-to-fit subjective ordered categorical traits assessed by different assessors in tree breeding. The comparison between the proposed MTMM and the conventional TMM, which fits common thresholds for the three assessors and accounts for the differences between assessors by a systematic fixed effect, showed similar estimates of the posterior means of σ_f^2 and h^2 for the stem quality trait (Table 3). The MTMM displayed a slightly higher DIC than the TMM (i.e., 4,931.78 vs. 4,924.58, respectively), most possibly due to the higher penalty of having a larger "effective number of parameters." However, MTMM yielded slightly more precise estimates (i.e., narrower 95 % HPD interval; Table 3) than TMM. We expect that the advantage of using MTMM will be greater for ordered categorical data sets showing more heterogeneity of the subjective visual scale across assessors. Although there were practically no differences between posterior mean estimates obtained from MTMM and TMM for variance components, the heterogeneity in the estimated posterior means of the thresholds between assessors (Fig. 3) had a higher impact on the predicted breeding values from TMM and MTMM (Table 4 and Fig. 2).

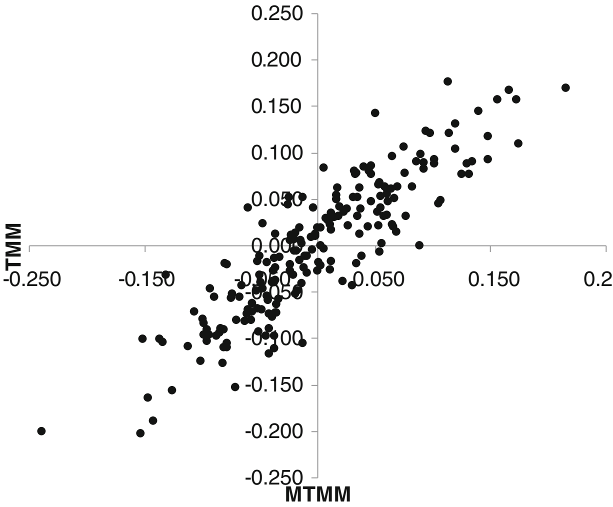

Fig. 2 Scatter plot of predicted breeding values of family genetic effects from the threshold mixed model (TMM) and the assessor-specific multi-threshold mixed model (MTMM)

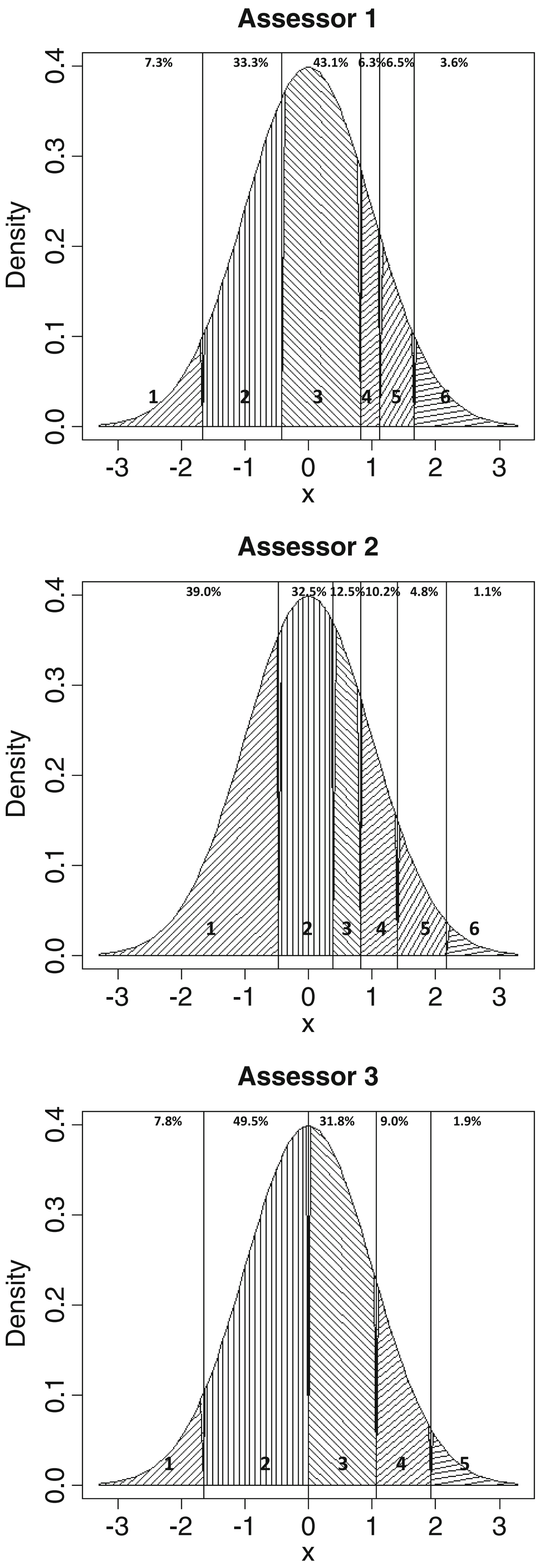

Fig. 3 Posterior mean estimates for thresholds for each assessor in the assessor-specific multi-threshold mixed model. *Numbers on the upper part* of each figure refer to distribution of the assessment generated by a proportion of the total number of trees scored. *Numbers on the lower part* of each figure refer to categories of the stem quality score. Observe that assessor 3 used only five of the six categories available to evaluate the stem quality

Therefore, selecting the 10 best families based on MTMM (Table 5) would have consequences on selection decision and rate of genetic progress. This is the first study applying a multi-threshold model to the analysis of ordered categorical traits from forest genetic trials. Therefore, making comparisons with other studies is not possible. However, in an animal breeding context, Varona et al. (2009) showed that the multi-threshold mixed model with a specific set of thresholds for each slaughterhouse had a better fit (smallest DIC value) than the threshold animal mixed model for the subjective conformation and fat cover score in the Pirenaica beef cattle breed. In addition, they also found slight differences in terms of variance components; the estimated posterior means of the threshold and multi-threshold mixed models of h^2 were 0.24 and 0.23 for conformation and 0.13 and 0.13 for fat cover scores, respectively.

Variance components and heritabilities of the six-point ordered categorical stem quality score were also estimated using a conventional linear model in the observed score (LMM). Given that the stem quality score used in this study has a non-normal distribution (Table 2), the differences in the estimated posterior means of h^2 between the standard LMM and TMM (0.077 vs. 0.097) and between LMM and MTMM (0.077 vs. 0.096) were greater than between heritability estimates from TMM and MTMM (0.097 vs. 0.096, respectively). However, the linear model produces a heritability estimate in observed scale, while the threshold and multi-threshold models provide a heritability estimate in underlying scale (liability). When the LMM estimate of heritabilities on the observed scale (0.077) was transformed into the same underlying scale (Gianola 1979; Eq. (11), the LMM estimate of the heritability was higher (0.085) than in the original scale (0.077), but no greater than those estimated by the direct estimation procedures, i.e., TMM (0.097) and MTMM (0.096). Gianola (1979) described smaller estimated heritability on the linear scale with respect to the underlying scale for ordered categorical traits, which is consistent with the results of previous research (e.g., Olesen et al. 1994; Yang et al. 1998; Varona et al. 1999; Abdel-Azim and Berger 1999). For example, Yang et al. (1998) analyzed six-point score responses of lodgepole pine to infection of western gall rust and found that the estimated heritability on the underlying scale (0.44) was higher than that estimated on the discontinuous scale (0.25). These higher estimates of the heritabilities for both threshold and multi-threshold models should contribute, in part, to greater genetic improvement and are indicative of superiority of the threshold models over the linear model. Additionally, the TMM and MTMM had a better fit (smallest DIC values) than the standard LMM (i.e., 4,924.58 and 4,931.78 vs. 5,117.54, respectively). Instead, strong rank (0.95) and product–moment (0.95) correlations were found between the posterior mean breeding values from LMM and TMM, in keeping with those reported by Varona et al. (2009) for two subjective categorical traits: fat cover (0.95) and conformation score (0.92). These high correlations may be due to the relatively wide range of categories observed in the outward scale. In general, results show superiority of the threshold over the linear models for the analysis of ordered categorical traits (e.g., Abdel-Azim and Berger 1999). In a simulation research, Abdel-Azim and Berger (1999) investigated the goodness of fit of threshold models for analyzing categorical traits with different number of categories (2, 5, and 10), incidence of categories (extreme, moderate, and normal), heritabilities in the underlying scale (0.04, 0.20, and 0.50), and data structure (unbalanced and balanced). The authors reported that threshold models had better fit than all the linear models studied and showed that accuracy of genetic parameters increased significantly with a larger number of categories, a more normal distribution of incidences, increased heritability, and more balanced data.

TMM and MTMM explore a direct estimation procedure of the variance components based on a generalized linear mixed model, while LMM_NS uses a two-step approach, i.e., prior to the analysis with a LMM, the observed scores are transformed into normal score (Gianola and Norton 1981). Interestingly, when we transformed the observed six-point score by creating this new set of scores for the stem quality trait, LMM_NS yielded comparable heritability estimates (0.094) with TMM (0.097) and MTMM (0.096). Furthermore, rank and product–moment correlations were high between LMM_NS and TMM (0.96 and 0.94, respectively) and between LMM_NS and MTMM (0.95 and 0.94, respectively). Canonical scaling was applied in forest genetic studies, but to the best of our knowledge, this is the first research in tree breeding to compare LMM with transformed normal score with the conventional LMM and generalized linear mixed models for ordered categorical traits. In animal breeding, Matilainen et al. (2009) also found comparable correlations (0.97) for the five-point calving ease score between the linear model with a normalized score and the threshold model for the direct genetic effects. Therefore, based on the results of this particular case, it appears that the transformation to normal score is a recommended method to ordered categorical traits in forest genetic analysis.

The assignment of ordered categorical scores to response categories is not trivial. Gianola and Norton (1981) showed that heritability in the observed scale is not invariant with respect to the scores. Yang et al. (1998) confirmed this finding using nine different categories (from 3 to 11) in a simulation study and also in a polychotomous response of lodgepole pine to western gall rust using six- and two-score categories. Therefore, working on the underlying continuous scale using LMM_NS, TMM, or MTMM presents several advantages over working on the observed scale using the traditional LMM. In our study, transforming the data into a common underlying continuous distribution made it possible to correct the heterogeneity of assessors' perceptions. Additionally, it allowed us to compare heritability estimates obtained from different breeding populations to determine whether any difference observed among populations truly reflected unequal scope for genetic improvement or they were simply a consequence of difference in the observed scale (Yang et al. 1998). The proposed model would also be advantageous when estimating genetic parameters of ordered categorical traits of multi-environmental forest genetic trials, especially when sites influence the degree of site-specific categorical trait expression. In this case, assessors usually apply a relative score to the trees in a single test location (i.e., site-specific score) rather than relative to all trees in genetic tests (i.e., standardized absolute score). This methodology may limit the utility of these data in genetic analyses across sites, since the scale of grading may be different according the different site (e.g., Pswarayi et al. 1997; Wu and Matheson 2005). The use of a particular set of thresholds for each site should correct the heterogeneity between sites, thus transforming the data into a common underlying continuous distribution.

A Bayesian procedure coupled with a MCMC technique (i.e., Gibbs sampling with a data augmentation step) has been used to estimate the variance components. An alternative approach for estimating dispersion parameters is the use of an approximation of the marginal density of the variance components, referred to as marginal maximum likelihood (MML; Tempelman 1998). Various approximated MML estimation methods have been used for estimating variance components of threshold mixed models, including approximated expectation-maximization (Harville and Mee 1984; Stiratelli et al. 1984), and Laplacian (Raudenbush et al. 2000) methods. Comparison of frequentist and Bayesian estimators is difficult because central issues related to the comparison of frequentist estimators (such as repeated sampling or bias) do not have the same meaning in the Bayesian school (Gelman et al. 1995, p. 108). There is no small sample distribution for MML approaches, but asymptotic normality under certain conditions on the eigenvalues of the information matrix (see Cressie and Lahiri 1993). Even for generalized linear mixed models with a Gaussian underlying liability distributions (i.e., TMM and MTMM), the calculation of the information matrix using MML methods may be unfeasible with large data sets. Moreover, whereas MML provides only point estimates of the parameters and the asymptotic approximation of their variances, the Bayesian approach allows more general inferences as the exact posterior distribution is available. Therefore, variance, or standard errors, or posterior HPD intervals, or the probability of a parameter being less than a given value can be reported. However, no significant differences between posterior mean, mode, and median of the genetic parameters σ_f^2 and h^2 were found in the two linear and the two threshold models studied; thus, we would not anticipate a substantial difference between the MCMC Bayesian approach applied in our study and the MML methods for these genetic parameters. Hence, the methodological effects of MML and MCMC techniques on parameter estimates should be less important than the effects of different modeling of the ordered categorical traits shown above.

A data augmentation step in the Gibbs sampling algorithm facilitates computations as stated by Sorensen et al. (1995). However, they also noted that threshold parameters have very slow mixing properties, and this is probably related to the data augmentation approach. In spite of the visual inspection of the running mean plots and the results of the Geweke's test, our data showed no evidence that lack of convergence was a problem for the σ_f^2 and h^2. Geweke's test detected a slight lack of convergence (p values less than 0.05 for the Z scores) in 3 of the 13 threshold parameters estimated (p value equal to 0.03, 0.03, and 0.01). The slow convergences for these thresholds could be more critical with an individual additive tree mixed model. Cowles (1996) presented a Metropolis–Hastings algorithm for generating liability data and thresholds jointly, instead of individually from their respective full conditionals. For a simulated seven-point categorical data set with a Gaussian liability, she showed a much faster convergence with this new algorithm.

Conclusions

This study applied an assessor-specific multi-threshold mixed model for the analysis of an ordered categorical trait from a forest genetic trial to estimate family variance and heritability using a Bayesian method via Gibbs sampling with a data augmentation step. This mixed model with heterogeneous thresholds regarding assessors account for both the discrete distribution of these traits and the different categorization patterns of each assessor. It was found to be well suited to the analysis of subjective ordered categorical data of forest genetic evaluations assessed by different assessors. This approach will be most effective for those ordered categorical data sets that have a strong departure from the Gaussian distribution and significant difference between assessors.

Acknowledgments This research was supported by grant from Agencia Nacional de Promoción Científica y Tecnológica (PICT 00321) of Argentina, under the Programa de Modernización Tecnológica III, Contrato de Préstamo BID 1728/OC-AR. The author would like to thank to Dr. Aníbal Verga for kindly providing the data used for the study, to Carla Salto, who assisted with measurements and data processing, and two anonymous referees for their insightful comments on an earlier version of this manuscript.

Appendix A

Prior distributions and joint posterior densities for all parameters of the assessor-specific multi-threshold mixed model

In a conjugate approach, the prior densities for all parameters are chosen to be closed under sampling (Robert and Casella 1999), which means that both prior and posterior belong to the same family of distributions. In order to reflect a prior state of uncertainty for the fixed effects in a mixed linear model, while keeping the posterior distribution proper (Hobert and Casella 1996), $\boldsymbol{\beta}$ is taken to be N_p $(\boldsymbol{0}, \boldsymbol{K})$. The matrix $\boldsymbol{K}$ is diagonal with large elements ($k_{ii} > 10^8$), and the prior density of $\boldsymbol{\beta}$ is then proportional to:

$$p(\beta|\boldsymbol{K}) \propto \left| \prod_{i=1}^{p} k_{ii} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} \frac{\beta_i^2}{k_{ii}} \right\} \tag{4}$$

The vector of family effects is distributed a priori as $\boldsymbol{f} \sim N(0, \boldsymbol{I}\ \sigma_f^2)$, so that:

$$p\left(\boldsymbol{f} \middle| \sigma_f^2\right) \propto \left(\sigma_f^2\right)^{-\frac{f}{2}} exp\left\{ -\frac{\boldsymbol{f}'\boldsymbol{f}}{2\sigma_f^2} \right\} \tag{5}$$

Following Sorensen and Gianola (2002), we chose to use independent scaled inverted chi-square densities as a priori distributions for the variance components σ_f^2 with known parameters υ and δ^2, so that:

$$p\left(\sigma_f^2 \middle| \upsilon, \delta^2\right) \propto p\left(\sigma_f^2\right)^{-\left(\frac{\upsilon}{2}+1\right)} \exp\left\{ \frac{\upsilon \delta^2}{2\sigma_f^2} \right\} \tag{6}$$

Finally, the thresholds for each assessor $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \dots, t_{iCi-1})$ are assumed to be distributed as order statistics from a uniform distribution in the interval $[t_{\min}, t_{\max}]$:

$$p(\boldsymbol{t}_i) = (C_i - 1)! \left(\frac{1}{t_{\min} - t_{\max}} \right)^{C_i - 1} I(t_i \in \mathbf{T}_i) \tag{7}$$

where $I(.)$ is an indicator function and $\mathbf{T}_i = \{(t_{i1}, t_{i2}, \dots, t_{iC_i-1}) | t_{\min} \le t_{i1} \le t_{i2} \le \dots \le t_{iC_i-1} \le t_{\max}\}$.

Likelihood functions of the assessor-specific multi-threshold mixed model

Applying Bayesian theorem, the augmenting joint posterior density, i.e., the density of all unknown parameters, including the liability ($\boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{t}, \boldsymbol{l}, \sigma_f^2$) given the observed data ($\boldsymbol{y}$) and the prior information, is written as the product of the prior distributions (4), (5), (6), and (7), and the likelihood function, which results in:

$$p\left(\beta, \boldsymbol{f}, \boldsymbol{l}, \boldsymbol{t}, \sigma_f^2 \middle| \boldsymbol{y}, \upsilon, \delta^2\right) \propto p\left(\beta\right) p\left(\boldsymbol{f} \middle| \sigma_f^2\right) p\left(\sigma_f^2 \middle| \upsilon, \delta^2\right) p\left(\boldsymbol{l} \middle| \beta, \boldsymbol{f}\right) p\left(\boldsymbol{t}\right) p\left(\boldsymbol{y} \middle| \beta, \boldsymbol{f}, \boldsymbol{l}, \boldsymbol{t}, \sigma_f^2\right) \tag{8}$$

Following Varona and Hernández (2006), last term in (8), the likelihood of the observed data, can be expressed as:

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{l}, \boldsymbol{t}, \sigma_f^2) = & \prod_{i=1}^{a} \prod_{j=1}^{n_i} 1\left(l_{ij} < t_{ij}\right) 1\left(y_{ij} = 0\right) \\ & + 1\left(t_{i1} < l_{ij} < t_{i2}\right) 1\left(y_{ij} = 1\right) + \dots \\ & + +1\left(t_{i1} < l_{ij} < t_{i2}\right) 1\left(y_{ij} = C_i + 1\right) \end{aligned} \tag{9}$$

Conditional posterior densities for all parameters of the assessor-specific multi-threshold mixed model

Inference on any parameter by means of the Gibbs sampler requires conditional posterior densities in close form. Collecting the fourth and last terms to the right of (8), the full conditional posterior distribution of the liability follows a truncate normal distribution with density:

$$p\left(l_{ij} \middle| \boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{l}_{-ij}, \boldsymbol{t}\right) = \frac{\phi\left(\boldsymbol{x}_i' \boldsymbol{\beta} + \boldsymbol{z}_i' \boldsymbol{f}, 1\right)}{\Phi(t_{ik+1} - \boldsymbol{x}_i' \boldsymbol{\beta} - \boldsymbol{z}_i' \boldsymbol{f}) - \Phi(t_{ik} - \boldsymbol{x}_i' \boldsymbol{\beta} - \boldsymbol{z}_i' \boldsymbol{f})} \tag{10}$$

For the thresholds, collecting the last two terms in (8), the full conditional posterior distribution of the jth threshold for the ith assessor t_{ij} (Varona and Hernádez 2006) is:

$$p\left(t_{ij} \middle| \boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{l}, t_{-ij}, \sigma_f^2, \boldsymbol{y}\right) = \frac{1}{\min(\boldsymbol{l}|\boldsymbol{y} = j+1)_i - \max(\boldsymbol{l}|\boldsymbol{y} = j)_i} \tag{11}$$

where $\min(\boldsymbol{l}|\boldsymbol{y} = j+1)_i$ is the maximum liability for the records that failed in the $j+1$ category from the ith assessors, and

$\max(\boldsymbol{l}|\boldsymbol{y}=j)_i$ is the minimum liability for the records that failed in the $j+1$ category for the ith assessors.

The posterior conditional density of the family variance σ_f^2 is:

$$p\left(\sigma_f^2\middle|\beta,\boldsymbol{f},\boldsymbol{l},\boldsymbol{y}\right)\propto p\left(\sigma_f^2\right)^{-\left(\frac{n+\upsilon}{2}+1\right)}\exp\left\{\frac{\widetilde{\upsilon}\widetilde{\delta}^2}{2\sigma_f^2}\right\} \tag{12}$$

which is a scaled inverted χ^2 density with $\widetilde{\upsilon}=n+\upsilon$ degrees of freedom and scale parameter $\widetilde{\delta}^2=\left(\boldsymbol{f}'\boldsymbol{f}+\upsilon\delta^2\right)/(\upsilon+n)$.

Following Sorensen and Gianola (2002), the joint conditional density of β and $\boldsymbol{f}$ is equal to

$$\begin{bmatrix}\widehat{\beta}\\ \widehat{\boldsymbol{f}}\end{bmatrix}\left|\boldsymbol{l},\sigma_f^2,\boldsymbol{y}\sim N\left(\begin{bmatrix}\widehat{\beta}\\ \widehat{\boldsymbol{f}}\end{bmatrix},\begin{bmatrix}\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{Z}\\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z}+\boldsymbol{I}\sigma_f^{-2}\end{bmatrix}^{-1}\right)\right. \tag{13}$$

The vectors $\widehat{\beta}$ and $\widehat{f}$ in (14) are the solutions to the following system of equations

$$\begin{bmatrix}\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{Z}\\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z}+\boldsymbol{I}\sigma_f^{-2}\end{bmatrix}\begin{bmatrix}\widehat{\beta}\\ \widehat{\boldsymbol{f}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}'\boldsymbol{l}\\ \boldsymbol{Z}'\boldsymbol{l}\end{bmatrix} \tag{14}$$

At each iteration, the Gibbs algorithm proceeds by first sampling $\boldsymbol{l}$ from (10), then t_j from (11), the family variance σ_f^2 from (12), and the linear parameters $\widehat{\beta}$ and $\widehat{f}$ of model (3) from (13).

References

Abdel-Azim G, Berger PJ (1999) Properties of threshold model predictions. J Anim Sci 77:582–590

Agresti A (2007) An introduction to categorical data analysis, 2nd edn. Wiley, Hoboken

Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 88:669–679

Apiolaza LA, Chauhan SS, Walker JCF (2010) Genetic control of very early compression and opposite wood in *Pinus radiata* and its implications for selection. Tree Genet Genomes 7:563–571. doi:10.1007/s11295-010-0356-0

Bannister MH (1979) An early progeny trial in *Pinus radiata*. 2. Subjective assessment of crookedness. N Z J For Sci 9:241–261

Bessega C, Saidman BO, Darquier MR, Ewens M, Sánchez L, Rozenberg P, Vilardi JC (2009) Consistency between marker and genealogy-based heritability estimates in an experimental stand of *Prosopis alba* (Leguminosae). Am J Bot 96:458–465

Cané-Retamales C, Mora F, Vargas-Reeve F, Perret S, Contreras-Soto R (2010) Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. Euphytica 178:177–183. doi:10.1007/s10681-010-0292-y

Cappa EP, Cantet RJC (2006) Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. Can J For Res 36:1276–1285

Cappa EP, Cantet RJC (2008) Direct and competition additive effects in tree breeding: Bayesian estimation from an individual tree mixed model. Silvae Genet 57:45–49

Cappa EP, Pathauer PS, Lopez GA (2010) Provenance variation and genetic parameters of *Eucalyptus viminalis* in Argentina. Tree Genet Genomes 6:981–994. doi:10.1007/s11295-010-0307-9

Codesido V, Fernandez-Lopez J (2009) Implications of genotype × site interaction on *Pinus radiata* breeding in Galicia. New Forests 37:17–34

Cotterill PP, Dean CA (1990) Successful tree breeding with index selection, division of forestry and forest products. CSIRO, Australia

Cowles KM (1996) Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. Statist Comput 6:101–111

Cressie N, Lahiri SN (1993) The asymptotic distribution of REML estimators. J Multivar Anal 45:217–233

Dempster ER, Lerner M (1950) Heritability of threshold characters. Genetics 35:212–236

Dutkowski GW, Potts BM (2012) Genetic variation in the susceptibility of *Eucalyptus globulus* to drought damage. Tree Genet Genomes. doi:10.1007/s11295-011-0461-8

Ericsson T, Danell Ö, Andersson B (1994) Genetic variation of *Pinus contorta* var. *latifolia* breeding material in Sweden. Can J For Res 24:723–729

Galizzi FA, Angueira C, Prieto D (1999) Suelos de la Planta Piloto de drenaje del INTA Santiago del Estero. Revista de Ciencias Forestales Quebracho 7:52–60

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman and Hall, New York

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics 4. Oxford University Press, Oxford

Gianola D (1979) Heritability of polychotomous characters. Genetics 93:1051–1055

Gianola D (1980) A method of sire evaluation for dichotomies. J Anim Sci 51:1266–1271

Gianola D (1982) Theory and analysis of threshold characters. J Anim Sci 54:1079–1096

Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. Genet Sel Evol 15:201–224

Gianola D, Norton HW (1981) Scaling threshold characters. Genetics 99:357–364

Gwaze DP, Wolliams JA, Kanowski PJ, Bridgwater FE (2001) Interactions of genotype with site for height and stem straightness in *Pinus taeda* in Zimbabwe. Silvae Genet 50:135–140

Hai PH, Jansson G, Harwood C, Hannrup B, Thinh HH (2008) Genetic variation in growth, stem straightness and branch thickness in clonal trials of *Acacia auriculiformis* at three contrasting sites in Vietnam. For Ecol Manag 255:156–167

Hamilton MG, Potts BM, Greaves BL, Dutkowski GW (2010) Genetic correlations between pulpwood and solid-wood selection and objective traits in *Eucalyptus globulus*. Ann For Sci 67:511–520. doi:10.1051/forest/2010013

Hannerz M, Sonesson J, Ekberg I (1999) Genetic correlations between growth and growth rhythm observed in a short-term test and performance in long-term field trials of Norway spruce. Can J For Res 29:768–778

Hannrup B, Säll H, Jansson G (2003) Genetic parameters for spiral grain in Scots pine and Norway spruce. Silvae Genet 52:215–220

Harville DA, Mee RW (1984) A mixed-model procedure for analyzing ordered categorical data. Biometrics 40:393–408

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph

Hobert JP, Casella G (1996) The effects of improper priors on Gibbs sampling in hierarchical linear models. J Am Stat Assoc 91:1461–1473

Högberg KA, Persson B, Hallingbäck HR, Jansson G (2010) Relationships between early assessments of stem and branch properties and sawn timber traits in a *Pinus sylvestris* progeny trial. Scand J Forest Res 25:421–431

Li H, Ghosh S, Amerson H, Li B (2006) Major gene detection for fusiform rust resistance using Bayesian complex segregation analysis in loblolly pine. Theor Appl Genet 113:921–929

Matilainen K, Mrode R, Strandén I, Thompson R, Mäntysaari EA (2009) Linear–threshold animal model for birth weight, gestation length and calving ease in United Kingdom Limousin beef cattle data. Livest Sci 122:143–148

McCulloch CE (1994) Maximum likelihood variance components estimation for binary data. J Am Stat Assoc 89:330–335

Olesen I, Perez-Enciso M, Gianola D, Thomas DL (1994) A comparison of normal and nonnormal mixed models for number of lambs born in Norwegian sheep. J Anim Sci 72:1166–1173

Pswarayi IZ, Barnes Richard D, Birks JDS, Kanowski PJ (1997) Genotype-environment interaction in a population of *Pinus elliottii* Engelm var. *elliottii*. Silvae Genet 46:35–40

Raudenbush SW, Yang M, Yosef M (2000) Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. J Comput Graph Statist 9:141–157

Raymond C, Cotterill PP (1990) Methods of assessing crown form of *Pinus radiata*. Silvae Genet 39:67–71

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. Accessed 15 Mar 2012.

Robert CP, Casella GC (1999) Monte Carlo statistical methods. Springer, New York

Sierra de Grado R, Diez-Barra R, Alia Miranda R (1999) Evaluación de la rectitud del fuste en seis procedencias de *Pinus pinaster* Ait. Investig Agrar Sistemas y Recur Forestales 8(2):264–278

Silverman B (1986) Density estimation for statistics and data analysis. Chapman and Hall, London

Smith BJ (2003) Bayesian Output Analysis Program (BOA) version 1.0 user's manual. Available from http://www.public-health.uiowa.edu/boa/. Accessed 15 Mar 2012.

Sorensen D, Gianola D (2002) Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer, New York

Sorensen D, Andersen S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. Genet Sel Evol 27:229–249

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). J R Stat Soc Ser Series B 64:583–639

Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM (2006) A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. Tree Genet Genomes 2:30–38

Stiratelli R, Laird N, Ware JH (1984) Random-effects models for serial observations with binary response. Biometrics 40:961–971

Tanner M, Wong W (1987) The calculation of posterior distributions by data augmentation. J Am Stat Assoc 82:528–550

Tempelman RJ (1998) Generalized linear mixed models in dairy cattle breeding. J Dairy Sci 81:1428–1444

Varona L, Hernández P (2006) A multithreshold model for sensory analysis. J Food Sci 71:333–336

Varona L, Misztal I, Bertrand JK (1999) Threshold-linear versus linear-linear analysis of birth weight and calving ease using an animal model: I. Variance component estimation. J Anim Sci 77:1994–2002

Varona L, Moreno C, Altarriba J (2009) A model with heterogeneous thresholds for subjective traits: fat cover and conformation score in the Pirenaica beef cattle. J Anim Sci 87:1210–1217

Verga A, López Lauenstein D, López C, Navall M, Joseau J, Gómez C, Royo O, Degano W, Marcó M (2009) Morphological characterization of algarrobo trees (*Prosopis* sp.) from the phytogeographical regions of Chaco and northern Espinal in Argentina. Revista de Ciencias Forestales Quebracho 17:31–40

Waldmann P, Hallander J, Hoti F, Sillanpää MJ (2008) Efficient MCMC implementation of Bayesian analysis of additive and dominance genetic variances in non-inbred pedigrees. Genetics 179:1101–1112

Wang CS, Quaas RL, Pollak EJ (1997) Bayesian analysis of calving ease scores and birth weights. Genet Sel Evol 29:117–143

Wei RP, Lindgren K, Lindgren D (2001) Parental environment effects on cold acclimation and height growth in lodgepole pine seedlings. Silvae Genet 50:252–257

Wright S (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. Genetics 19:506–536

Wu HX, Matheson AC (2005) Genotype by environment interactions in an Australia-wide radiata pine diallel mating experiment: implications for regionalized breeding. For Sci 51:29–40

Wu HX, Ying CC, Muir JA (1996) Effect of geographic variation and jack pine introgression on disease and insect resistance in lodgepole pine. Can J For Res 26:711–726

Yanchuk AD, Yeh FC, Dancik BP (1988) Variation of stem rust resistance in a lodgepole pine provenance family plantation. For Sci 34:1067–1075

Yang RC, Dhir NK, Yeh FC (1998) Intraclass correlation of polychotomous responses of lodgepole pine to infection of western gall rust: a simulation study. Silvae Genet 47:108–115

Zas R, Sampedro L, Moreira X, Martins P (2008) Effect of fertilization and genetic variation on susceptibility of *Pinus radiata* seedlings to *Hylobius abietis* damage. Can J For Res 38:63–72