Genetic evaluation combining purebred and crossbred data in a pig breeding scheme

N. Ibáñez-Escriche,*2 J. Reixach,† N. Lleonart,† and J. L. Noguera*

*Genètica i Millora Animal, Institut de Recerca i Tecnologia Agroalimentaria (IRTA), Av. Rovira Roure, 191, 25198 Lleida, Spain; and †Selección Batallé S.A., Riudarenes, Spain

ABSTRACT: Genetic evaluations using purebred data alone and combined purebred and crossbred information were performed for lean meat percentage in a pig breeding scheme. One purebred (PB) model and 2 crossbred models (CCPS1 and CCPS2) were used in the analyses. Data were obtained from the Selección Batallé S.A. Company (Riudarenes, Spain) and spanned a period of 4 yr (2006 to 2009). The data corresponded to 3 nuclei of purebred populations, Landrace (LD), Duroc (DU), and Pietrain (PI); 1 multiplying farm with animals from a 2-way cross (TB1; DU × LD); and commercial farms with animals from a 3-way cross (TB2; TB1 × PI). Genetic parameters were similar across the models, with the exception of purebred PI. The DU and LD purebreds presented large heritabilities (0.5 to 0.6) for lean meat percentage, whereas the PI purebred showed a lower heritability (approximately 0.1) for the PB model and moderate heritability for the CCPS1 and CCPS2 models (0.2 to 0.3). The mean reliability of the predicted purebred breeding values was clearly increased when the CCPS1 and CCPS2 models were used. Moreover, a reranking of the animals with important changes in the selection decisions was observed in the PI purebred. In a simulation study, the CCPS1 model achieved a greater response to selection than the PB model for the PI purebred. On another hand, between the CCPS1 and CCPS2 models, CCPS1 was slightly superior in terms of predictive ability, exhibiting a greater robustness. These results illustrate the usefulness of using crossbred models to evaluate lean meat percentage in this pig breeding scheme.

Key words: Bayesian analysis, crossbred, cross-validation, pig, purebred, reliability, response to selection

INTRODUCTION

The breeding system of domestic pigs is pyramidal. Nucleus herds supply genetically improved breeding stock to multiplier herds, which disseminate genetic gain by supplying hybrid stock to commercial herds. As Dekkers (2007) pointed out, one of the main limitations of this system is that purebred performance (PB) in nucleus farms can be a poor predictor of future crossbred performance (CC) on commercial farms [i.e., genetic correlations from 0.4 to 0.7 (Lutaaya et al., 2001; Merks and de Vries, 2002)]. An alternative is to include phenotypic data from crossbred relatives (CCPS) in estimation of breeding values of purebreds, as proposed by Wei and van der Steen (1991) and Lo et al. (1993). The CCPS system can increase the response to selection (Bijma and van Arendonk, 1998) but requires accurate recording of performance and pedigrees at the CC level. Such data are not common; thus, CCPS has mostly been studied using computer simulations.

Lo et al. (1993) presented a theory for estimating the covariance between crossbred animals for an additive model, and Cantet and Fernando (1995) showed how to use the theory to predict breeding values by BLUP. However, the genetic variance components for the multibreed models were difficult to estimate with REML (Elzo, 1994; Birchmeier et al., 2002) or Bayesian (Cardoso and Tempelman, 2004) methods. For these reasons, various approximations have been used. In particular, Spilke et al. (1998) reported a small loss in efficiency when using a multiple-trait approach in which the records from each line were treated as separate traits. Recently, García-Cortés and Toro (2006) decomposed the covariance matrix of the additive values (G) by a source of variability, allowing a simple ANOVA and solving the estimation problem. These authors showed empirically that this model is equivalent to the additive...
model developed by Lo et al. (1993), and Munilla Le-
guizamón and Cantet (2010) presented the formal deri-
vation. However, this model assumes a homogeneous 
residual variance. An extension model with heteroge-
neous residual variance ought to be developed because,
in some cases, as in pig breeding schemes, the purebred 
and crossbred lines do not share the environment. 
The aims of this study were 1) to perform CCPS 
for lean meat percentage (LM%) in a real pig popula-
tion using 2 crossbred models, the model presented by 
Spilke et al. (1998) and the model proposed by García-
Cortés and Toro (2006), which was extended for hetero-
gegeneous residual variances; 2) to compare the predicted 
ability of the 2 crossbreed models; and 3) to compare 
the impact of CCPS with a classical breeding program 
based on PB in the selection criterion.

MATERIALS AND METHODS

The animals were managed under standard intensive 
conditions. The protocols were approved by the Ethical 
and Animal Care Committee at IRTA.

Data

Data for the study were obtained from the Selecció 
Bataished S.A. Company (Riudarenes, Spain) and 
spanned a period of 4 yr, from 2006 to 2009. The data 
corresponded to 3 nucleus herds of PB popula-
tions, Landrace (LD), Duroc (DU), and Pietrain (PI); 
1 multiplying farm with animals from a 2-way cross 
(TB1; DU × LD); and commercial farms with animals 
from a 3-way cross (TB2; TB1 × PI). The trait ana-
lyzed was LM%, measured in the slaughterhouse (be-
tween 170 to 185 d of life) by Autofom. This system 
allows individual records to be obtained automatically 
(see Busk et al., 1999). Table 1 shows the distribution 
of records, means, and SD of LM% for the PB and CC 
used in the analyses. Table 2 shows the characteristics 
of the pedigree file used in the analysis, the distribution 
of the contemporary groups, and the BW at slaughter. 
The contemporary groups were composed of the pen of 
reared pigs and the day of slaughter.

Statistical Models

One PB model and 2 CCPS models were used in the 
genetic evaluation.

Purebred Model. Within-line single-trait analyses 
were carried out using the following PB model:

\[ y = Xb + Za + e, \]

where \( y \) is the vector of observations; \( b, a, \) and \( e \) are 
the vectors of systematic, genetic, and residual effects, 
respectively; and \( X \) and \( Z \) are incidence matrices. The vector of systematic effects included contempor 
ary group (see Table 3), sex (male or female), and BW at the end of a test period nested within sex.

CCPS Models. The first CCPS model used 
(CCPS1), proposed by Spilke et al. (1998), considered 
records from PB (DU, LD, PI) and CC lines (TB1, 
TB2) as 5 correlated traits:

\[ y_i = X_ib_i + Z_ia_i + e_i, \]

where the \( i \) subscript denotes the vectors and matrices 
for the appropriate breed composition (\( i = DU, LD, PI, \)

Table 1. Distribution of records, means, and SD of lean meat percentages for each purebred and crossbred1 population

<table>
<thead>
<tr>
<th>Item</th>
<th>Duroc</th>
<th>Landrace</th>
<th>Pietrain</th>
<th>TB1</th>
<th>TB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>2565</td>
<td>950</td>
<td>1890</td>
<td>965</td>
<td>8555</td>
</tr>
<tr>
<td>Percentage</td>
<td>17.19</td>
<td>6.37</td>
<td>12.66</td>
<td>6.46</td>
<td>57.32</td>
</tr>
<tr>
<td>Mean</td>
<td>48.41</td>
<td>54.96</td>
<td>63.66</td>
<td>51.59</td>
<td>59.53</td>
</tr>
<tr>
<td>SD</td>
<td>4.66</td>
<td>3.28</td>
<td>2.44</td>
<td>4.21</td>
<td>2.34</td>
</tr>
</tbody>
</table>

1TB1 = 2-way cross (Duroc × Landrace); TB2 = 3-way cross (TB1 × Pietrain).

Table 2. Characteristics of the pedigree and data file of the breed groups

<table>
<thead>
<tr>
<th>Item1</th>
<th>Individuals</th>
<th>Boars</th>
<th>Sows</th>
<th>Contemporary group</th>
<th>Mean BW (kg) of female sex</th>
<th>Mean BW (kg) of male sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedigree file</td>
<td>Duroc</td>
<td>7,916</td>
<td>424</td>
<td>4,347</td>
<td>65</td>
<td>83.53</td>
</tr>
<tr>
<td></td>
<td>Landrace</td>
<td>1,615</td>
<td>168</td>
<td>479</td>
<td>23</td>
<td>84.81</td>
</tr>
<tr>
<td></td>
<td>Pietrain</td>
<td>4,661</td>
<td>331</td>
<td>1,314</td>
<td>42</td>
<td>77.40</td>
</tr>
<tr>
<td></td>
<td>TB1</td>
<td>1,907</td>
<td>91</td>
<td>699</td>
<td>14</td>
<td>84.83</td>
</tr>
<tr>
<td></td>
<td>TB2</td>
<td>8,555</td>
<td>296</td>
<td>848</td>
<td>109</td>
<td>84.84</td>
</tr>
<tr>
<td>Total</td>
<td>24,654</td>
<td>1,120</td>
<td>7,568</td>
<td>253</td>
<td>83.77</td>
<td>84.44</td>
</tr>
</tbody>
</table>

1TB1 = 2-way cross (Duroc × Landrace); TB2 = 3-way cross (TB1 × Pietrain).
TB1, TB2). The model extended for all breed composition is

$$\begin{pmatrix} y_{DU} \\ \vdots \\ y_{TB2} \end{pmatrix} = \begin{pmatrix} X_{DU} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & X_{TB2} \end{pmatrix} \begin{pmatrix} b_{DU} \\ \vdots \\ b_{TB2} \end{pmatrix} + \begin{pmatrix} e_{DU} \\ \vdots \\ e_{TB2} \end{pmatrix},$$

and assuming that $y$ conditioning on the unknown parameters follows a multivariate normal distribution,

$$\begin{pmatrix} y_{DU} \\ \vdots \\ y_{TB2} \end{pmatrix} \sim N\left( \begin{pmatrix} b_{DU} \\ \vdots \\ b_{TB2} \end{pmatrix}, \begin{pmatrix} X_{DU}b_{DU} + Z_{DU}a_{DU} \\ \vdots \\ X_{TB2}b_{TB2} + Z_{TB2}a_{TB2} \end{pmatrix} \right),$$

where $y_i$ is the vector of observations for the $i$th breed; $b_i$, $a_i$, and $e_i$ are the vectors of systematic, genetic, and residual effects, respectively; $X_i$ and $Z_i$ are incidence matrices of each breed; and $R_e$ is a diagonal (co)variance residual matrix. In this case, the incidence matrices for each breed were different. To simplify computations, we used data augmentation, a technique that fills the data vector with random imputations, so that incidence matrices were the same for all traits (Sorensen and Gianola, 2002). After data augmentation, the model can be written as

$$\begin{pmatrix} y | b, a, R_e \end{pmatrix} \sim N\left( Xb + Za, R_e \right),$$

where $y$ is a vector of augmented data; $X$ and $Z$ are known incidence matrices; and $R_e$ is the (co)variance residual matrix. If the data are sorted by individual, the residual (co)variance matrix can be written as $R_e = R_o \otimes I$, where $R_o$ is the $5 \times 5$ diagonal (co)variance matrix between breeds, $R_o = \text{diag} \{ \sigma^2_{DU}, \sigma^2_{LD}, \sigma^2_{f\|}, \sigma^2_{TB1}, \sigma^2_{TB2} \};$ and $I$ is the identity matrix, with the same order as the number of individuals. The vector of systematic effects $b$ included contemporary group (253 levels), sex (male or female), and BW at the end of a test period nested within sex.

The second CCPS model (CCPS2) follows the model for crossbreeding schemes described by García-Cortés and Toro (2006). This model is equivalent to the additive model developed by Lo et al. (1993), which described the rules to obtain $G$, the covariance matrix of the additive values, including both the pure breed contributions and the segregation deviations. In our case, in which 3 breeds are involved in the founder generation, and given the cross design of the data (see Table 3), there is segregation variance only between the DU and LD breeds. The $G$ components defined by Lo et al. (1993) are

$$g_a = f_{DU}^t \sigma^2_{DU} + f_{LD}^t \sigma^2_{LD} + f_{f\|}^t \sigma^2_{f\|} + 2 \left( f_{DU}^t f_{LD}^t + f_{DU}^t f_{f\|}^t + f_{LD}^t f_{f\|}^t \right) \sigma_{as}^2 + \frac{1}{2} g_{SD},$$

which can be written as

$$g_a = \sum_{i=1}^3 f_i^t \sigma^2_b + 2 \left( f_{DU}^t f_{LD}^t + f_{DU}^t f_{f\|}^t + f_{LD}^t f_{f\|}^t \right) \sigma_{as}^2 + \frac{1}{2} g_{SD},$$

and

$$g_y = \frac{1}{2} \left( g_{sr}^t + g_{dy}^t \right),$$

where the subscripts $DU$, $LD$, and $PI$ indicate the corresponding breed (DU, LD, and PI, respectively); $f_i^t$ is the proportion of genes of animal $i$ coming from breed $b = (DU, LD, PI)$; $S$ and $D$ are the sire and the dam of animal $i$; $S$ and $D$ are the sire and the dam of animal $j$; $\sigma^2_b$ is the additive variance component of breed $b$; $\sigma^2_{as}$ is the segregation variance between breeds DU and LD; and $g_{SD}$ is the covariance between genotypic values of the parents in animal $i$. Note that for TB2 individuals, the terms $\left( f_{DU}^t f_{LD}^t + f_{DU}^t f_{f\|}^t + f_{LD}^t f_{f\|}^t \right)$ have a value of 0 and the term $\left( f_{DU}^t f_{LD}^t + f_{DU}^t f_{f\|}^t \right)$ has a value of 0.25 (i.e., $f_{DU}^t = 0$, $f_{LD}^t = 1$, $f_{f\|}^t = 0$, $f_{DU}^t = 0.5$, $f_{LD}^t = 0.5$).

This particular definition of $G$ allows a multibreed population model to be analyzed as a standard animal model: $y = Xb + Za + e$. However, in this multibreed model, both the pure breed and the segregation components are mixed into the additive covariance matrix $G$. Therefore, $G$ cannot be expressed as the numerator relationship matrix times the additive variance com-

---

Table 3. Additive variances for the different breed populations

<table>
<thead>
<tr>
<th>Breed population</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purebred</td>
<td>$\sigma^2_{DU}$</td>
</tr>
<tr>
<td>Duroc (DU)</td>
<td>$\sigma^2_{DU}$</td>
</tr>
<tr>
<td>Landrace (LD)</td>
<td>$\sigma^2_{LD}$</td>
</tr>
<tr>
<td>Pietrain (PI)</td>
<td>$\sigma^2_{PI}$</td>
</tr>
<tr>
<td>Two-way cross (TB1) DU × LD</td>
<td>$1/2(\sigma^2_{DU} + \sigma^2_{LD})$</td>
</tr>
<tr>
<td>Three-way cross (TB2) PI × (DU × LD)</td>
<td>$1/2(\sigma^2_{DU} + \sigma^2_{LD}) + 1/2(\sigma^2_{DU,LD})$</td>
</tr>
</tbody>
</table>

1 $\sigma^2_{DU,LD} = \text{segregation variance between breeds DU and LD.}$
ponents as in the standard animal model, and, as a consequence, the estimation of $G$ becomes more complicated (Elzo, 1994). To avoid this problem, García-Cortés and Toro (2006) proposed an equivalent model in which $G$ can be split into several parts, allowing a simple analysis of the variance components. Following García-Cortés and Toro (2006), the $G$ of our multibreed population can be defined as $$G = A_{DU}\sigma_{aDU}^2 + A_{LD}\sigma_{aLD}^2 + A_{PI}\sigma_{aPI}^2 + A_S\sigma_{aSDU-LD}^2,$$

where $A_i$ ($i = DU$, $LD$, $PI$, $S$) are the partial numerator relationship matrices with order $q \times q$ ($q$ = number of animals) corresponding to the breed origin ($DU$, $LD$, $PI$), and $S$ is the segregation term. Afterward, our multibreed model became

$$y = Xb + Z_Du a_{DU} + Z_{LD} a_{LD} + Z_{PI} a_{PI} + Z_{SDU-LD} a_{SDU-LD} + e,$$

where $y$ is the vector of data; $b$ is the vector that contains the systematic effects contemporary group (351 levels), sex (male or female), and the covariables breed group effect and BW at the end of a test period nested within sex; $a_i$ ($i = DU$, $LD$, $PI$, $S_{DU-LD}$) are the vectors of the breeding values split by origin components and the segregation term [see García-Cortés and Toro (2006) and Munilla Leguizamón and Cantet (2010) for more information], and $X$, and $Z_i$ ($i = DU$, $LD$, $PI$, $S_{DU-LD}$) are the corresponding incidence matrices. Note that the breeding values of the individuals that do not correspond to a particular breed origin or segregation term are equal to 0.

The presence of a common-litter effect was not included in the models. Previous analyses performed in the purebred models revealed an influence of the common-litter effect within 3 to 7% of the total variance, which could overestimate the heritabilities. However, their inclusion in crossbred models considerably increases the number of parameters and makes their estimation computationally difficult.

### Statistical Analysis

In the 3 models, a hierarchical Bayes construction was considered. The distribution of the data (LM%) was assumed to be multivariate normal:

$$\left(y|b, a, R\right) \sim N\left(Xb + \sum Z_i a, R\right),$$

where the subscript $j$ denotes the model ($j = PB$, $CCPS1$, $CCPS2$). Improper priors were then assumed for the systematic effects, whereas genetic and residual effects were assumed to be distributed as $N(0, G_j)$ and $N(0, R_j)$, where $G_j$ and $R_j$ are the additive and residual covariance matrices for the model $j$, respectively. Note that $G_{PB} = A_{PB}\sigma_{aPB}^2$, $R_{PB} = I\sigma_e^2$, $G_{CCPS1} = A \otimes G_0$, $G_0 = 5 \times 5$ (co)variance matrix, $R_{CCPS1} = I \otimes R_0$, $R_0 = \text{diag}(\sigma_{r_DU}^2, \sigma_{r_LD}^2, \sigma_{r_PI}^2, \sigma_{r_{PI-B}}^2, \sigma_{r_{PI-B}}^2)$.

$G_{CCPS2} = A_{DU}\sigma_{aDU}^2 + A_{LD}\sigma_{aLD}^2 + A_{PI}\sigma_{aPI}^2 + A_S\sigma_{aSDU-LD}^2$ and $R_{CCPS2}$ is the heterogeneous diagonal residual variance matrix, with order $n \times n$ ($n$ = number of total data), composed of $\sigma_{r_DU}^2$, $\sigma_{r_LD}^2$, $\sigma_{r_PI}^2$, $\sigma_{r_{PI-B}}^2$, $\sigma_{r_{PI-B}}^2$ variances, where $A$ is the additive genetic relationship matrix; $\otimes$ is the Kronecker product; $\sigma_a^2$ is the additive genetic variance; $\sigma_e^2$ is the residual variance; the subscripts $DU$, $LD$, $PI$, $TB1$, and $TB2$ correspond to the pig population; and $S$ corresponds to the segregation terms. The prior distribution of the dispersion parameters was assumed improper, with the exception of the residual variances of $DU$ and $LD$ for the CCPS2 model that were inverted scaled chi-squares with the scale parameters $v = 100$ and $S^2 = 5.8$, and $v = 20$ and $S^2 = 3$, respectively. Bayesian inference via Markov chain Monte Carlo (MCMC) methods was used to analyze the data. The fully conditional distributions of the 3 models for systematic ($b_j$) and additive effects ($a_j$) were multivariate normal distributions. The fully conditional distributions for the residual variances of the 3 models (PB, $CCPS1$, and $CCPS2$) and the additive variances of the PB and $CCPS2$ models were inverted scaled chi-squares. However, inverted Wishart distribution was the fully conditional distributions for the additive genetic (co)variance matrices of the $CCPS1$ model. Marginal posterior distributions of all unknowns were estimated using the Gibbs sampling algorithm (e.g., Geman and Geman, 1984). After exploratory analyses, we used a single chain with a total of 500,000 samples for each analysis, after a burn-in period of 50,000. Convergence was tested separately for all dispersion parameters using the Raftrey and Lewis (1992) algorithm and the $Z$ criterion of Geweke (1992). The effective sample size was evaluated using the method of Geyer (1992), and Monte Carlo sampling errors were computed using the time-series procedures described by Geyer (1992).

### Model Comparisons

#### Reliabilities and Rank Correlations

The comparison between PB and $CCPS$ models was based on reliabilities and rank correlations of estimated additive values within PB individuals in the data and pedigree. The reliabilities were calculated as $r_{ij}^2 = 1 - pev_{ij}/\sigma_{a_{ij}}^2$, where $r_{ij}^2$ is the reliability for animal $i$ and breed $j$, $pev_{ij}$ is the corresponding prediction error variance, and $\sigma_{a_{ij}}^2$ is the additive variance for breed $j$ (Lutaya et al., 2002).

#### K-Fold Cross-Validation

This method was used to evaluate the models based on their ability to predict “future data.” The entire data set was partitioned into 4 disjoint subsets, each with approximately one-
fourth of the records, by taking random samples of data points. The cross-validation procedure used 3 of the 4 subsets for fitting and prediction (training set), and the remaining subsets were used to test predictive ability (testing set). Two different criteria were used to compare the predictive ability of the models.

**Mean Squared Error.** The mean squared error (MSE) was computed as

$$\text{MSE} = \frac{1}{n_{\text{iter}}} \sum_{i=1}^{n_{\text{iter}}} \frac{1}{n_{\text{data}}} \sum_{j=1}^{n_{\text{data}}} (y - \hat{y})^2,$$

where $y$ and $\hat{y}$ correspond to the observed and predicted observations, respectively; $n_{\text{data}}$ is the number of data points in a testing subset; and $n_{\text{iter}}$ is the number of iterations of the MCMC chain. Models with the smallest MSE were regarded as those with the best predictive ability.

**Pearson’s Correlation.** Pearson correlations (PC) between observed and predicted observations were calculated as

$$\rho_{y,\hat{y}} = \frac{1}{n_{\text{iter}}} \sum_{i=1}^{n_{\text{iter}}} \frac{\text{cov}(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}},$$

where $\text{cov}(y, \hat{y})$ is the estimate of the covariance between observed and predicted LM% records; $\sigma_y$ and $\sigma_{\hat{y}}$ are the estimates of SD of observed and predicted records; and, as above, $n_{\text{iter}}$ is the number of iterations of the MCMC chain. The model providing the highest correlation was considered as the one with the best predictive ability of the yet to be observed records.

The predicted observations $\hat{y}$ were computed as follows:

$$\hat{y}_{ij} = x_j^i \hat{b}_i + z_j^i \hat{a}_i,$$

where $\hat{y}_{ij}$ is the predicted value for iteration $i$ and individual $j$, $x_j^i$ and $z_j^i$ are the $j$ rows of the $X$ and $Z$ incidence matrices, respectively, and $\hat{b}_i$ and $\hat{a}_i$ are samples of the systematic and additive vectors for the iteration $i$.

**Response to Selection.** The impact of CCPS vs. a classical breeding program based on PB in the selection criterion was evaluated by comparing the average EBV of the future boars selected using the CCPS1 or PB model. The data analyzed were simulated with the same structure as the real data and under the CCPS2 model. By applying a selection pressure of 30%, from the 45 future selection candidates of the PI line, 15 boars were selected for each model as parents of the next generation. The true average breeding values from the 15 selected boars of each model were computed and compared.

## RESULTS AND DISCUSSION

In this study we estimated the (co)variances of LM% using 1 purebred model in which each PB line was analyzed separately, and 2 CCPS models in which data from PB and CC animals were jointly analyzed. The first CCPS model (CCPS1) considered the LM% records of each PB and CC lines as different traits. This enables a simple multiple-trait analysis. However, this model assumes that purebreds contribute equally to the crossbred variances, and it is known that this model can provide biased estimates (Spilke et al., 1998). Nevertheless, this model allows for nonadditivity and genotype × environment interaction (Wei, 1992). The second model (CCPS2) was the multibreed additive model proposed by García-Cortés and Toro (2006), which is equivalent to the model by Lo et al. (1993). Hence, it takes into account the proper specification of the additive genetic covariance structure, but it does not account for nonadditive effects of genotype × environ-

### Table 4. Posterior means (SD) of the additive and segregation variances for each purebred and crossbred population

<table>
<thead>
<tr>
<th>Model</th>
<th>Duroc (SD)</th>
<th>Landrace (SD)</th>
<th>Pietrain (SD)</th>
<th>TB1 (SD)</th>
<th>TB2 (SD)</th>
<th>$\sigma^{2}_{SD\text{DU-LD}}$ (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td>8.49 (1.33)</td>
<td>6.90 (1.17)</td>
<td>0.62 (0.19)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCPS1</td>
<td>8.32 (0.56)</td>
<td>6.53 (0.94)</td>
<td>0.96 (0.24)</td>
<td>6.65 (0.76)</td>
<td>1.85 (0.15)</td>
<td></td>
</tr>
<tr>
<td>CCPS2</td>
<td>6.45 (0.46)</td>
<td>5.43 (0.51)</td>
<td>0.82 (0.16)</td>
<td>5.94 (0.50)</td>
<td>3.51 (0.27)</td>
<td>0.23 (0.04)</td>
</tr>
</tbody>
</table>

1TB1 = 2-way cross (Duroc × Landrace); TB2 = 3-way cross (TB1 × Pietrain); $\sigma^{2}_{SD\text{DU-LD}}$ = segregation variance between Duroc (DU) and Landrace (LD) purebreds; PB = purebred; CCPS1 = crossbred model 1; CCPS2 = crossbred model 2.

### Table 5. Posterior means (SD) of the heritability for each purebred and crossbred population

<table>
<thead>
<tr>
<th>Model</th>
<th>Duroc (SD)</th>
<th>Landrace (SD)</th>
<th>Pietrain (SD)</th>
<th>TB1 (SD)</th>
<th>TB2 (SD)</th>
<th>$\sigma^{2}_{SD\text{DU-LD}}$ (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td>0.58 (0.07)</td>
<td>0.71 (0.08)</td>
<td>0.12 (0.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCPS1</td>
<td>0.57 (0.03)</td>
<td>0.69 (0.07)</td>
<td>0.18 (0.04)</td>
<td>0.55 (0.04)</td>
<td>0.37 (0.03)</td>
<td></td>
</tr>
<tr>
<td>CCPS2</td>
<td>0.47 (0.03)</td>
<td>0.60 (0.04)</td>
<td>0.27 (0.04)</td>
<td>0.59 (0.03)</td>
<td>0.50 (0.03)</td>
<td>0.08 (0.02)</td>
</tr>
</tbody>
</table>

1TB1 = 2-way cross (Duroc × Landrace); TB2 = 3-way cross (TB1 × Pietrain); $\sigma^{2}_{SD\text{DU-LD}}$ = segregation variance between Duroc (DU) and Landrace (LD) purebreds; PB = purebred; CCPS1 = crossbred model 1; CCPS2 = crossbred model 2.
Inferences of Model Parameters

Tables 4 and 5 present features of the marginal posterior distributions of additive genetic variances and heritabilities, respectively. Posterior mean additive variances and heritabilities within the breeds were similar in the PB and CCPS1 models for DU and LD purebreds, whereas the CCPS2 model had smaller posterior means for these purebreds. This could have been because the models were building up under different genetic assumptions, which could affect the numerical estimations obtained. Note that the CCPS2 model considered different sources of variation, such as the segregation variance. Moreover, the crossbred data had a smaller contribution of the DU and LD breeds, which could make the appropriate separation of the sources of variation of these breeds difficult. In the case of PI, the posterior means of additive genetic variance and heritability did not follow the same pattern. The highest posterior mean of additive genetic variance was for the CCPS1 model, followed by the CCPS2 and PB models. However, their posterior intervals overlapped. The greatest posterior mean heritability was achieved for the CCPS1 model, followed by the CCPS2 and PB models. In this case, the CCPS2 and PB models clearly showed different posterior mean heritabilities. Note that the heritability of the PI purebred increased considerably in the models in which crossbred data were considered. This could imply that using crossbred data could improve the accuracy of the PI evaluation.

Model Comparison

The mean predicted reliabilities of the predicted PB breeding values obtained for each model are presented.
in Table 6. The 2 CCPS models showed greater mean reliabilities than the PB model for all breeds. The improvement in reliability was approximately 16, 8, and 50%, for DU, LD, and PI, respectively, when the breeding values were estimated using the CCPS1 model, and was 21, 12, and 75%, respectively, when the CCPS2 model was used. Reliabilities of EBV were greater for the CCPS models than for the PB model. It was especially noticeable in the PI purebred, which had the largest contribution, with 8,555 records, in the TB2 individuals. Within the CCPS models, in terms of reliability, the CCPS2 model was superior to the CCPS1 model, particularly regarding the PI breed. However, reliability depends on the genetic parameters, and CCPS2 provided a greater heritability than CCPS1 for the PI purebred (Table 5). Thus, using crossbred data in the genetic evaluation considerably increased the reliability of the estimated purebred breeding values. Moreover, it would be in agreement with the results found by Lutaaya et al. (2002) in a joint evaluation of purebreds and crossbreds for lifetime daily BW gain and backfat in swine.

Table 7 shows the rank correlations of the predicted breeding values between the models of PB animals. Rank correlations were greater (>0.80) for all models and breeds. However, the PI breed showed a smaller rank correlation between the models compared with the other purebreds, especially between the PB and CCPS2 models. These features would indicate that different models in the genetic evaluation could give different ranks of the selection candidates. These results are in disagreement with the results obtained by Lutaaya et al. (2002), in which no differences between the rankings of the PB and CCPS models were found. However, the numbers of records from crossbred animals used in that study were small compared with the number of purebred records. To study the prediction ability of the CCPS models, a cross-validation analysis was performed. Tables 8 and 9 show the results from the 4-fold cross-validation analysis based on the Monte Carlo estimates of MSE and PC, respectively, for each breed composition. The MSE for the CCPS1 model was less than that for the CCPS2 model; therefore, accuracy of prediction of EBV was better in all cases except 3 (subsets 1 and 2 for the LD purebred and subset 4 for the PI purebred). Nevertheless, the differences in MSE were very small between the models, especially for the PI purebred and the TB2 crossbred. In the case of the Monte Carlo estimates of PC, with the exception of subset 4 and the PI purebred, the CCPS1 model achieved better results than the CCPS2 model.

In spite of the fact that the differences between the models were not large, especially in terms of MSE, the cross-validation results showed that the CCPS1 model

### Table 8. Monte Carlo estimates of mean squared errors (MSE) from CCPS1 and CCPS2 model subsets 1, 2, 3, and 4

<table>
<thead>
<tr>
<th>Subset</th>
<th>Model</th>
<th>Duroc</th>
<th>Landrace</th>
<th>Pietrain</th>
<th>TB1</th>
<th>TB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE1</td>
<td>CCPS1</td>
<td>14.25</td>
<td>7.97</td>
<td>5.36</td>
<td>9.46</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>15.49</td>
<td>7.72</td>
<td>5.49</td>
<td>10.9</td>
<td>5.11</td>
</tr>
<tr>
<td>MSE2</td>
<td>CCPS1</td>
<td>13.05</td>
<td>7.75</td>
<td>5.92</td>
<td>13.62</td>
<td>4.67</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>14.48</td>
<td>7.25</td>
<td>6.10</td>
<td>14.61</td>
<td>5.37</td>
</tr>
<tr>
<td>MSE3</td>
<td>CCPS1</td>
<td>12.70</td>
<td>9.35</td>
<td>5.22</td>
<td>9.42</td>
<td>4.49</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>14.81</td>
<td>10.6</td>
<td>5.49</td>
<td>12.52</td>
<td>5.33</td>
</tr>
<tr>
<td>MSE4</td>
<td>CCPS1</td>
<td>13.65</td>
<td>7.50</td>
<td>5.48</td>
<td>10.18</td>
<td>4.34</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>16.51</td>
<td>6.46</td>
<td>5.34</td>
<td>11.14</td>
<td>4.96</td>
</tr>
<tr>
<td>Mean</td>
<td>CCPS1</td>
<td>13.41</td>
<td>8.14</td>
<td>5.50</td>
<td>10.67</td>
<td>4.57</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>15.10</td>
<td>8.01</td>
<td>5.61</td>
<td>12.29</td>
<td>5.19</td>
</tr>
</tbody>
</table>

1CCPS1 = crossbred model 1; CCPS2 = crossbred model 2; TB1 = 2-way cross (Duroc × Landrace); TB2 = 3-way cross (TB1 × Pietrain).

### Table 9. Monte Carlo estimates of Pearson correlations (PC) from CCPS1 and CCPS2 model subsets 1, 2, 3, and 4

<table>
<thead>
<tr>
<th>Subset</th>
<th>Model</th>
<th>Duroc</th>
<th>Landrace</th>
<th>Pietrain</th>
<th>TB1</th>
<th>TB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>CCPS1</td>
<td>0.61</td>
<td>0.43</td>
<td>0.32</td>
<td>0.65</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>0.58</td>
<td>0.38</td>
<td>0.31</td>
<td>0.60</td>
<td>0.34</td>
</tr>
<tr>
<td>PC2</td>
<td>CCPS1</td>
<td>0.62</td>
<td>0.52</td>
<td>0.37</td>
<td>0.61</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>0.54</td>
<td>0.36</td>
<td>0.33</td>
<td>0.60</td>
<td>0.34</td>
</tr>
<tr>
<td>PC3</td>
<td>CCPS1</td>
<td>0.60</td>
<td>0.51</td>
<td>0.31</td>
<td>0.60</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>0.51</td>
<td>0.33</td>
<td>0.31</td>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td>PC4</td>
<td>CCPS1</td>
<td>0.64</td>
<td>0.54</td>
<td>0.25</td>
<td>0.65</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>0.58</td>
<td>0.46</td>
<td>0.28</td>
<td>0.51</td>
<td>0.27</td>
</tr>
<tr>
<td>Mean</td>
<td>CCPS1</td>
<td>0.62</td>
<td>0.50</td>
<td>0.31</td>
<td>0.63</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>CCPS2</td>
<td>0.55</td>
<td>0.38</td>
<td>0.30</td>
<td>0.53</td>
<td>0.32</td>
</tr>
</tbody>
</table>

1CCPS1 = crossbred model 1; CCPS2 = crossbred model 2; TB1 = 2-way cross (Duroc × Landrace); TB2 = 3-way cross (TB1 × Pietrain).
had a better prediction of future data (predicted ability) than the CCPS2 model. These results were in contrast to the reliability results, which favored the CCPS2 model. This fact could be explained because, although the CCPS2 model fitted the data better than the CCPS1 model, this model would be more robust than the CCPS2 model. Moreover, the large records required for the CCPS2 model could also have affected the predictive performance of the subsets.

As mentioned previously, using the crossbred records could improve the accuracy of the breeding values, and they could affect the candidate selection ranking and the posterior response to selection. Figure 1 shows the breeding value ranking of 45 PI boar candidate selections using the PB model and the CCPS1 model. The models gave different ranks, and this could influence the candidates selected as parents. When a selection pressure of 30% was applied, which, in this case, meant selecting 15 boars, there was a difference between the models on the selection candidates of 5 boars (Figure 1). This fact could have some implications for the response to selection. However, this could not be tested because the true breeding values were not known. For this reason, a simulation process, via the model by Lo et al. (1993), was carried out using the structure of the population evaluated previously. The simulated data were analyzed using the PB and CCPS1 models, and the average of the true breeding values from the 15 boars selected out of 45 selection candidates was computed for each model. The results from the simulation (Table 10) show that the CCPS1 model would provide a greater response to the selection for LM% (0.21). Regarding the purebred model, it would represent a 10% additional response in the PI purebred. These outcomes are in agreement with previous studies (Comstock et al., 1949; Wei and Van Der Werf, 1994; Bijma and van Arendonk, 1998) that predicted an increase in the response to selection when a crossbred evaluation was performed.

**Implications**

On the whole, an increase in reliability was obtained when crossbred and purebred information was combined to perform the genetic evaluation. As a consequence, an important reranking of the animals with a corresponding change in selection was also shown when the genetic evaluation was based on the PB or CCPS models. The increase in reliability and the changes in ranking were more noticeable for the PI purebred, which had a small heritability and a greater contribution (50%) to the TB2 crossbred. Nevertheless, between the CCPS models, the approximate model (CCPS1) was slightly superior in terms of predictive ability, exhibiting a greater robustness. These results support the idea of using crossbred models to evaluate LM% in this

![Figure 1. The ranking of 45 Pietrain boars candidates to selection for the purebred model (PB) vs. crossbred model 1 (CCPS1). The points to the left of the vertical line are the boars selected using the CCPS1 model, and those underneath the horizontal line are the boars selected using the PB model. The difference in selecting boars using the CCPS1 or PB model is indicated by the points inside the circles. Color version available in the online PDF.](image)

**Table 10. Average of the true breeding values of the 15 Pietrain boars selected from the PB and CCPS1 models**

<table>
<thead>
<tr>
<th>Item</th>
<th>PB</th>
<th>CCPS1</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated</td>
<td>1.05</td>
<td>1.26</td>
<td>0.21</td>
</tr>
</tbody>
</table>

1PB = purebred; CCPS1 = crossbred model 1.
pig breeding scheme, because they could imply an increase in the response to selection.

In our analysis, LM% was modeled under the additive action. It is known that in related growth traits, the magnitude of dominance variance is small (Misztal et al., 1998). However, a model with dominance effects could be implemented to improve the estimation of additive values and the availability of specific combining abilities for each pair of prospective parents.

LITERATURE CITED


