

# Genetic evaluation of age at first calving for Guzará beef cattle using linear, threshold, and survival Bayesian models

Lais C. Brito,<sup>\*,†</sup> Joaquim Casellas,<sup>‡</sup> Luis Varona,<sup>§,¶</sup> Paulo S. Lopes,<sup>\*</sup> Henrique T. Ventura,<sup>§</sup> Maria Gabriela C. D. Peixoto,<sup>&</sup> Sirlene F. Lázaro,<sup>\*</sup> and Fabyano F. Silva<sup>\*,¶</sup>

<sup>\*</sup>Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais 36570-000, Brazil;

<sup>†</sup>Department of Animal Science, Universidade Federal Rural da Amazonia, Capanema, Para 68700-030,

Brazil; <sup>‡</sup>Grup de Recerca en Millora Genetica Molecular Veterinaria, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain; <sup>§</sup>Departamento de Anatomia, Embriologia y Genetica Animal, Universidad de Zaragoza, Zaragoza E-50013, Spain; <sup>¶</sup>Instituto Agroalimentario de Aragon (IA2), Zaragoza E-50013, Spain; <sup>§</sup>Associação Brasileira dos Criadores de Zebu, Uberaba, Minas Gerais 38022-330, Brazil; and <sup>&</sup>Embrapa Gado de Leite, Juiz de Fora, Minas Gerais 36038-330, Brazil

**ABSTRACT:** Age at first calving (AFC) is characterized as a censored trait due to missing values provided by recording mistakes and nonoccurrence or delay in calving communication. In this context, we aimed to compare several statistical methods for genetic evaluation of AFC in Guzará beef cattle under a Bayesian approach. Seven different methods were used for this purpose. The traditional linear mixed model (LM), which considers only uncensored records; the LM with simulated records (SM), which is based on data augmentation framework; the penalty method, in which a constant of 21 d was added to censored records; the bivariate threshold-linear method considering (TLcens) or not (TLmiss) censored information; and the piecewise Weibull proportional

hazards model considering (PWPHcens) or not (PWPH) censored records. Heritability estimates ranged from 0.19 (TLcens) to 0.28 (SM) in non-survival approaches; and 0.40 and 0.46 to PWPH and PWPHcens methods, respectively. In general, breeding values correlations between different methods and the percentage of selected bulls in common indicated reranking, with these correlation ranging from -0.28 (between SM and PWPH) to 0.99 (between TLmiss and LM). The traditional LM, which considers only uncensored records, should be preferred due to its robustness and simplicity. Based on cross-validation analyses, we conclude that the TLmiss could be also a suitable alternative for breeding value prediction, and censored methods did not improve the analysis.

**Key words:** beef cattle, censored data, survival analysis, threshold-linear models

© The Author(s) 2018. Published by Oxford University Press on behalf of the American Society of Animal Science. All rights reserved. For permissions, please e-mail: [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

J. Anim. Sci. 2018.XX:XX-XX

doi: 10.1093/jas/sky112

## INTRODUCTION

Age at first calving (AFC) is highly related to fertility and reproductive efficiency in cattle (Berry and Evans, 2014). Low AFC is typically associated with heifer precocity and longer lifetime productivity (Bormann and Wilson, 2010). Some

females usually do not have AFC records due to nonoccurrence or failure in the calving communication and to inconsistent records in the data editing. These kind of unknown observations are often defined as censored records, which can be exploited by using appropriate statistical methods (Tarrés et al., 2006; Hou et al., 2009; Garcia et al., 2016). Among these methods, stand out the data augmentation (Guo et al., 2001), the penalization of censored records (Johnston and Bunter, 1996), the bivariate linear-threshold model assuming the

<sup>¶</sup>Corresponding author: [fabyanofonseca@ufv.br](mailto:fabyanofonseca@ufv.br)

Received September 4, 2017.

Accepted March 19, 2018.

censoring status as an extra trait (Varona et al., 1999), and the survival analysis techniques (Casellas, 2007). In summary, these methods seek to increase the amount of information by treating the missing records as partial known observations.

The data augmentation replaces the censored values by random samples generated from a positive truncated normal distribution taking into account systematic and random estimated effects. The penalty method (PM) assumes that censoring is due to delayed AFC by adding a constant (number of days, often 21 d based on the assumption that the heifer should be fertile in the subsequent estrous cycle) over contemporary data to replace the censored values. Bivariate analysis uses the censoring status (binary trait) as a correlated trait to improve the accuracy of genetic evaluation. It assumes that a correlation between AFC and the censoring status might reduce the problem of nonrandom censoring. Finally, survival analysis assumes a hazard function that provides the calving probability to estimate missing AFC values. The Weibull distribution is usually preferred because of its flexibility, but other generalizations based on additional time-dependent effects may be desirable to better fit the temporal probability changes (Casellas, 2007).

Toward this orientation, we aimed to compare these mentioned methods under a Bayesian framework for genetic evaluation of AFC records in Guzerá cattle. These comparisons were accessed by predictive performance via cross-validation and goodness-of-fit measures.

## MATERIALS AND METHODS

### Data

Guzerá fertility data were provided by the Brazilian Association of Zebu Cattle (ABCZ), Embrapa Dairy Cattle Research Corporation, and Brazilian Center for the Guzerá Genetic Improvement. Animal Care and Use Committee approval was not necessary for this study because analyses were performed on existing field data obtained under standard herds management from commercial breeders. The animals from these herds are raised under pasture conditions (supplemented with minerals ad libitum) and seasonal mating system (data from embryo transfer or in vitro fertilization techniques were not considered).

The trait AFC was defined as the time interval (in days) between birth and first calving. Initially, the data were composed by 121,352 records, with mean of  $1,277.27 \pm 245$  d, ranging from 740 and

1,860 d. Data from females with AFC out of the range of two SD within contemporary groups (CG) were considered as missing observations. The removal of outliers within CG varying from 2.5 to 3.5 SD has been used for growth traits in Nelore cattle (Silva et al., 2017). However, given the inconsistency of AFC trait in Zebu cattle under tropical conditions, we opted to be more restrictive by using 2 SD. After editing, the dataset included 69,157 AFC records collected between 1991 and 2012. A total of 230 CG remained for the analysis, which were formed as the combination of herd (a total of 10), year, and season of birth (dry season, from April to September; and rainy season, from October to March). The maximum and minimum CG sizes were 4 and 2,493 records, respectively.

The dataset consisted of 69,157 AFC phenotypes (uncensored records) and 7,023 AFC censored records, totalizing 76,180 records. As a whole, 230 CG were formed as the combination of herd, year, and season of birth. Each CG had at least four uncensored records. The pedigree file included a total of 104,588 animals.

### Statistical Methods

Age at first calving data were analyzed using seven different analytical approaches under a Bayesian framework.

### Linear Mixed Model

The linear mixed model (LM) is restricted to uncensored AFC data. Take as starting point the following standard animal LM:

$$\mathbf{y} = \mathbf{l}\mu + \mathbf{W}\mathbf{d} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (1)$$

where  $\mathbf{y}$  is the vector of AFC records;  $\mu$  is the general mean;  $\mathbf{d}$  is the vector of CG (herd-year-season) effects;  $\mathbf{a}$  is the vector of additive genetic effects;  $\mathbf{e}$  is the residual vector; and  $\mathbf{W}$  and  $\mathbf{Z}$  are the incidence matrices associated with  $\mathbf{d}$  and  $\mathbf{a}$ , respectively. It was assumed that  $\mathbf{d} \sim N(0, \mathbf{I}\sigma_d^2)$ ,  $\mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2)$ , and  $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ , being  $\mathbf{A}$  the numerator relationship matrix,  $\sigma_d^2$  the CG permanent environmental variance,  $\sigma_a^2$  the additive genetic variance,  $\mathbf{I}$  an identity matrix, and  $\sigma_e^2$  the residual variance.

### Imputation Method Via Simulation

The simulation models (SM) also assumed equation [1], but included additional records from data augmentation framework for heifers with censored records. Thus,  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_{ur} \\ \mathbf{y}_{cr} \end{bmatrix}$  is a vector in which

$\mathbf{y}_{ur}$  stored uncensored records and  $\mathbf{y}_{cr}$  stored augmented values from censored records. Using Gibbs sampling approach (Sorensen et al., 1998; Guo et al., 2001),  $\mathbf{y}_{cr}$  were sampled from their respective predictive distributions. It was assumed that  $\mathbf{y}_{cr}$  values followed a Gaussian truncated distribution whose lower limit was defined by the maximum values of AFC within the corresponding CG. Thus, augmented data  $\mathbf{y}_{cr}$  were updated within each iteration of the Gibbs sampler as an observation for each censored record (Korsgaard et al., 2003; Donoghue et al., 2004).

### Penalty Method

The PM is equivalent to SM, but censored records were replaced by a set of new records by adding a constant of 21 d over the highest AFC value within each CG (Donoghue et al., 2004; Hou et al., 2009).

### Threshold-Linear Method

The threshold-linear model (TLmiss) is represented by a bivariate analysis where one trait is continuous (being the censored records treated as missing records) and the another one was a threshold binary trait which indicates the censored status. Binary records were associated to liability values representing latent continuous records (Sorensen and Gianola, 2002). At each Markov Chain Monte Carlo (MCMC) iteration, the binary records generate a liability value below or over a given threshold. Considering the linear model [1], this model can be adapted by considering  $\mathbf{y}' = [\mathbf{y}_{ur} \ 0]$ :

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{l} \end{bmatrix} = \mathbf{1}\boldsymbol{\mu} + \begin{bmatrix} \mathbf{W}_y & 0 \\ 0 & \mathbf{W}_l \end{bmatrix} \begin{bmatrix} \mathbf{d}_y \\ \mathbf{d}_l \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_y & 0 \\ 0 & \mathbf{Z}_l \end{bmatrix} \begin{bmatrix} \mathbf{a}_y \\ \mathbf{a}_l \end{bmatrix} + \begin{bmatrix} \mathbf{e}_y \\ \mathbf{e}_l \end{bmatrix} \quad (2)$$

where  $\mathbf{y}$  is the vector of AFC records;  $\mathbf{y}_{ur}$  representing uncensored records;  $\mathbf{l}$  is the vector of liabilities of censored status;  $\mathbf{W}$ ,  $\mathbf{Z}$ ,  $\mathbf{d}$ ,  $\mathbf{a}$ , and  $\mathbf{e}$  are defined as previously, but specific to AFC records or liability predisposition. The following distributions were assumed:

$$\begin{bmatrix} \mathbf{a}_y \\ \mathbf{a}_l \end{bmatrix} \sim N(0, \mathbf{G}_0 \otimes \mathbf{A}), \quad \begin{bmatrix} \mathbf{d}_y \\ \mathbf{d}_l \end{bmatrix} \sim N(0, \mathbf{D}_0 \otimes \mathbf{I}),$$

$$\text{and } \begin{bmatrix} \mathbf{e}_y \\ \mathbf{e}_l \end{bmatrix} \sim N(0, \mathbf{R}_0 \otimes \mathbf{I})$$

where  $\mathbf{G}_0$  and  $\mathbf{R}_0$  are the additive genetic and residual (co)variance matrix, respectively, as proposed by Varona et al. (1999); and,  $\mathbf{D}_0$  is the (co)variance matrix for CG effects.

### Threshold-Linear Censored Method

Represents the same model described in Eq. [2], however, despite of missing values,  $\mathbf{y}' = [\mathbf{y}_{ur} \ \mathbf{y}_{al}]$ , where  $\mathbf{y}_{al}$  are the censored records added to a constant of 21 d over the highest AFC value within each CG.

### Piecewise Weibull Proportional Hazards Method

The piecewise Weibull proportional hazards model (PWPH) was developed within the survival analysis framework. The Weibull distribution was assumed as the baseline hazard function for the observed value. The data assumed for the animal  $i$  are  $(t_i, \delta_i)$ , where  $t_i$  is the vector of uncensored records and  $\delta_i$  is a censor indicator random variable (Casellas, 2007). The baseline hazard function was assumed as:

$$\mathbf{h}_0(\mathbf{t}) = \lambda\rho(\lambda\mathbf{t})^{\rho-1} \quad (3)$$

Considering  $\boldsymbol{\theta}' = [\boldsymbol{\mu}', \mathbf{d}', \mathbf{a}']$  and  $\mathbf{U}' = [\mathbf{W}', \mathbf{Z}']$ , equation [3] can be rewritten conditionally to  $\boldsymbol{\theta}$  as follows:

$$\mathbf{h}(\mathbf{t}|\boldsymbol{\theta}) = \mathbf{h}_0(\mathbf{t})\exp(\boldsymbol{\theta}\mathbf{U}) \quad (4)$$

Where  $\mathbf{h}(\mathbf{t}|\boldsymbol{\theta})$  is the vector of hazard functions which compute the limiting probability of the parturition at time  $\mathbf{t}$  for a heifer given  $\boldsymbol{\theta}$ ;  $\mathbf{h}_0(\mathbf{t})$  is a Weibull baseline hazard function at time  $\mathbf{t}$  with a positive scale parameter  $\lambda$  and a shape parameter  $\rho$ .

This Weibull model was modified allowing the baseline to assume several unknown change points along the parametric space. The number of change points was predefined and they are estimated simultaneously with the other parameters under a Bayesian framework (Yazdi et al., 2002; Tarrés et al., 2005; Casellas, 2007).

The Bayesian inference on PWPH was accessed by multiplying the likelihood function by the prior distributions of all parameters as follows:

$$p(\mathbf{b}, \mathbf{d}, \mathbf{a}, \sigma_a^2, \sigma_d^2, \rho, \lambda, \boldsymbol{\tau} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{b}, \mathbf{d}, \mathbf{a}, \rho, \lambda, \boldsymbol{\tau}) \quad (5)$$

$$p(\mathbf{b})p(\mathbf{d} | \sigma_d^2)p(\sigma_d^2)p(\mathbf{a} | \mathbf{A}, \sigma_a^2)p(\sigma_a^2)p(\rho)p(\lambda)p(\boldsymbol{\tau})$$

where  $\boldsymbol{\tau}$  is the vector of  $c$  change points ( $c \geq 2$ ). The  $\tau_1$  was fixed to 0,  $\tau_c$  equaled the maximum value in the vector  $\mathbf{y}$ , and  $\tau_j < \tau_{j+1}$ .

Following Damgaard and Korsgaard (2006), bounded uniform priors were assigned to  $\rho$ ,  $\lambda$ , and  $\tau_j$ , whereas remaining prior distributions were defined as for the other methods. The PWPH

method was implemented by assuming an increasing number of change points ( $0 \leq c \leq 4$ ) and the most desirable number of  $c$  was selected according to Deviance Information Criterion (**DIC**), defined by Spiegelhalter et al. (2002).

### *Piecewise Weibull Proportional Hazards Censored Method*

The piecewise Weibull proportional hazards censored method (**PWPHcens**) represents the same previously methods, however includes censored data by adding prior information,  $t' = [t_{ur} \ t_{al}]$ , where  $t_{al}$  are the censored records as defined in the PM.

### *Markov Chain Monte Carlo Sampling*

Inferences about all unknown parameters were done from the marginal posterior distributions. Gibbs sampler (Gelfand and Smith, 1990) with Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) steps in PWPH method were used (Casellas, 2007). For remaining methods, only the Gibbs sampling algorithm was used through TM software (Legarra et al., 2008). More specifically, a unique MCMC process with 100,000 iterations was launched for each analysis, and the first 20,000 iterations were discarded as burn-in, keeping every 50th sample for inference of posterior features. Convergence was monitored by graphical inspection and R package *boa* (Smith, 2007).

### *Methods Comparisons*

Due to a difference in scale between estimated breeding values (**EBV**) of survival analysis and other methods, the observed and predicted uncensored phenotypes were used to compare them by cross-validation. Training population was composed by censored and uncensored records. A total of 30% of uncensored records were set to missing and their predicted phenotypes were compared with observed phenotypes. These populations were randomly redefined 10 times within CG. The predicted phenotypic vector for AFC was calculated as:  $\hat{\mathbf{y}} = \mathbf{1}\hat{\mu} + \mathbf{W}\hat{\mathbf{d}} + \mathbf{Z}\hat{\mathbf{a}}$ . Pearson correlations, mean square errors, and regression analyses between observed and predicted uncensored phenotypes were estimated to access the predictive ability of the compared methods.

Spearman's rank correlation coefficients between EBV from different methods were computed to access possible reranking. In addition,

the percentage of sires (with at least one daughter) selected in common among the methods at different percentiles (TOP1% and TOP10%) was also calculated.

## RESULTS AND DISCUSSION

### *Methods Comparison*

The DIC values for PWPH analysis from 0 to 4 change points were 1,247,055, 1,251,001, 1,246,580, 1,248,432, and 1,257,111, respectively. For PWPHcens method, the DIC values from 0 to 4 change points were 1,646,395, 1,659,998, 1,663,677, 1,659,015, and 1,663,660, respectively. When using censored data (PWPHcens approach), the DIC values suggested that no additional change points in the baseline function were required, only the boundaries points (i.e., 0 and 1,860 d). However, for PWPH approach without censored records, a more parametrized model (two change points) was necessary to properly fit the distribution of AFC data.

The predictive ability was validated by correlation coefficients, mean square error, and bias (regression between observed and predicted phenotype) using cross-validation approach (Table 1). Standard deviations across models suggest precise inference for these results.

Higher correlations and lower MSE and bias were found to TLmiss method, indicating that this is recommended to be used in AFC genetic evaluation with censored records in Guzerá cattle. However, the traditional LM method presented quite similar results with TLmiss, and could be recommended as well. The methods (PM and threshold-linear censored model [**TLcens**]) that take into account the imputation of censored records (penalization) did not improve the analysis and presented similar results. The absence of previous information about AFC censored records and the genetic correlation among AFC data and censored status provide a slightly advantage compared to others approaches. According to Allison (2010), missing data should be influenced by observed values and not depend on any prior information, and this relation between real and censored values should be done by a censoring indicator variable, as well as in threshold analysis.

The survival analysis was not the best choice when handling AFC trait in Guzerá cattle, because PWPH models had predictive ability worse than others methods. However, the literature affirms that PWPH model provides a more flexible framework to accommodate both positive and negative

asymmetry in the data (Casellas, 2007; Casellas, and Bach, 2012).

Within-methods Spearman correlation coefficients and percentage of selected bulls in common (considering different percentiles, TOP1% and TOP10%) between-genetic breeding values predicted from survival and nonsurvival method are shown in Table 2.

The use of rank rather Pearson correlation solves the scale issue of PWPB models and EBVs could be compared. However, the comparison of nonsurvival methods with PWPB methods provided negative correlations estimates, due to the inverted interpretation necessary to PWPB methods as consequence of the relative risk scale presented by EBV results. Low values of EBV are desirable for AFC trait in nonsurvival methods and the opposite happens with survival methods.

Despite of observed differences in the predictive abilities, the Spearman correlations among nonsurvival methods were similar, indicating that no major reranking would be expected across these methods. These similarities and another previously results suggests that either approaches (PM and TLmiss) could be used for genetic evaluations of AFC trait. Simulation method appears to be the only one among nonsurvival methods that present some difference in comparison to others nonsurvival methods. The correlations between nonsurvival and PWPB models were ranged from  $-0.70$  to  $-0.55$ . Moderate correlations (for example  $-0.55$ ) indicate a reranking of animals in genetic evaluation.

Casellas and Bach (2012) working with lambing interval in ewes found correlations ranging from  $-0.715$  to  $-0.676$  between PWPB and LM methods.

**Table 1.** Average mean square error (MSE), correlation coefficient and bias with respective SD obtained from cross-validation analysis

Method <sup>1</sup>	MSE	Correlation	Bias
LM	63,087.19 (294.32)	0.43 (0.0033)	0.99 (0.0141)
SM	110,903.6 (961.42)	0.40 (0.0047)	0.43 (0.0047)
PM	65,277.25 (479.21)	0.43 (0.0037)	0.81 (0.0105)
TLmiss	62,987.62 (293.64)	0.44 (0.0034)	1.01 (0.0156)
TLcens	67,816.38 (256.00)	0.39 (0.0047)	0.80 (0.0114)
PWPB <sub>2</sub>	71,176.51 (293.64)	0.40 (0.0031)	0.92 (0.0562)
PWPB <sub>0</sub> cens	81,120.18 (479.09)	0.30 (0.0051)	0.55 (0.0067)

<sup>1</sup>Methods: LM, SM, PM, TLmiss, TLcens, PWPB<sub>2</sub>, PWPB<sub>0</sub>cens: linear, simulation, penalty, threshold-linear, threshold-linear censored, piecewise Weibull proportional hazard (fixed  $c = 2$ ), and piecewise Weibull proportional hazard censored (fixed  $c = 0$ ) methods, respectively.

**Table 2.** Spearman correlation of all animals (above diagonal) and sires (below diagonal) between predicted breeding values of AFC trait and percentage of sires in common between models at 1% (above diagonal) and 10% (below diagonal) selection percentages

Method <sup>1</sup>	LM	SM	PM	TLmiss	TLcens	PWPB <sub>2</sub>	PWPB <sub>0</sub> cens
Spearman correlations							
LM	—	0.88	0.98	0.99	0.96	-0.61	-0.59
SM	0.82	—	0.92	0.90	0.88	-0.55	-0.70
PM	0.96	0.88	—	0.98	0.94	-0.59	-0.61
TLmiss	0.99	0.87	0.98	—	0.96	-0.60	-0.63
TLcens	0.96	0.85	0.93	0.98	—	-0.57	-0.58
PWPB <sub>2</sub>	-0.30	-0.28	-0.29	-0.30	-0.30	—	0.77
PWPB <sub>0</sub> cens <sup>2</sup>	-0.34	-0.35	-0.33	-0.34	-0.36	0.45	—
Percentage of sires in common							
LM	—	37.21	86.04	93.02	67.44	2.32	6.98
SM	65.27	—	39.53	41.86	30.23	0.0	2.32
PM	89.12	69.91	—	90.70	65.11	2.33	6.98
TLmiss	94.44	69.68	90.74	—	67.44	2.32	6.98
TLcens	85.88	62.5	80.79	84.72	—	2.32	9.30
PWPB <sub>2</sub>	26.62	21.76	27.08	25.93	27.55	—	41.86
PWPB <sub>0</sub> cens <sup>2</sup>	19.44	18.52	19.91	18.98	19.44	45.60	—

<sup>1</sup>Method: LM, SM, PM, TLmiss, TLcens, PWPB<sub>2</sub>, PWPB<sub>0</sub>cens: linear, simulation, penalty, threshold-linear, threshold-linear censored, piecewise Weibull proportional hazard (fixed  $c = 2$ ), and piecewise Weibull proportional hazard censored (fixed  $c = 0$ ) methods.

<sup>2</sup>EBV's on original survival scale.

Higher correlations were found by [Pereira et al. \(2007\)](#) studying age at first conception in Nellore cattle population. These authors reported that when considering just the sires the correlation was even lower; and lower the number of animals, more sensitive is the Spearman correlation to slight changes in classification.

The percentages of sires in common were higher between LM and PM, LM and TLmiss, and PM and TLmiss. On the other hand, lower percentages were found among survival (PWP<sub>H</sub> and PPH<sub>cens</sub>) and nonsurvival methods. Since LM and TLmiss outperformed the other methods based on cross-validation prediction analysis ([Table 1](#)), the selected animals based on survival method would be unsuccessful.

### Variance Components and Genetic Parameter Estimates

Posterior means, SD, and highest posterior density (HPD95%) intervals of variance components and genetic parameter for AFC trait under different methods of handling censored data are presented in [Table 3](#). The LM only considers AFC uncensored values and its results should be used as reference (simplest model).

Among nonsurvival approaches (LM, SM, PM, TLmiss, and TLcens), posterior means of the additive variance under LM, PM, and TLmiss were similar, because estimates were within the highest posterior density interval of the other method. The corresponding estimate for SM and

TLcens methods, however, was significantly higher and lower than these estimates, respectively, and was outside the HPD95% interval for both LM, PM, and TLmiss methods. Since the SM method is based on random numbers generated through truncated normal distribution, which can be seen as an important source of variation, it is expected that this method provides higher variance. Although the PM outperformed the SM in the present study, [Donoghue et al. \(2004\)](#) working with data simulation, reported that the lack of significant differences in the genetic ranking of sires between these two methods suggests that either can be used. However, these same authors recommended further research comparing both methods on genetic evaluation of beef cattle field data. Posterior means of the residual variance for LM and TLmiss were the smallest, whereas the estimate under SM was the highest value. These results imply that LM and TLmiss methods provide a better fit to the data than the other nonsurvival methods when censored records are presented. For all parameters presented, LM and TLmiss were similar, representing the correspondence among the simplest method and threshold analysis. The TLmiss and TLcens model had genetic correlation of  $0.42 \pm 0.10$  and  $0.42 \pm 0.03$ , between the trait representing in linear or threshold way. Genetic correlation considered between data and censored status could be influencing the results, providing low estimates of additive genetic variance in the TLcens model.

**Table 3.** Posterior means, SD, and highest posterior density interval (HPD95%) of variance components and genetic parameter estimates for the trait age at first calving (AFC, in days)

Method <sup>1</sup>	$h^2$	$\sigma_a^2$	$\sigma_d^2$	$\sigma_e^2$
LM	0.26 (0.01) [0.25, 0.28]	19,617.46 (568.70) [18,578.95, 20,790.14]	6,277.02 (644.16) [5,067.28, 7,604.48]	48,900.01 (435.65) [48,041.00, 49,712.81]
SM	0.28 (0.01) [0.27, 0.30]	49,370.57 (1,108.85) [47,195.50, 51,540.91]	32,233.76 (2,601.06) [27,527.90, 37,615.27]	91,999.26 (830.59) [90,288.76, 93,588.29]
PM	0.25 (0.01) [0.23, 0.26]	21,299.24 (563.69) [20,119.11, 22,337.36]	12,102.18 (1,140.93) [11,024.03, 15,420.95]	53,370.49 (459.02) [52,491.59, 54,321.55]
TLmiss	0.25 (0.01) [0.24, 0.27]	18,720.01 (567.00) [17,662.69, 19,872.40]	6,019.29 (630.42) [4,809.97, 7,332.29]	49,488.00 (463.42) [48,536.51, 50,428.08]
TLcens	0.19 (0.01) [0.18, 0.20]	15,999.64 (410.55) [14,145.28, 16,757.71]	13,800.89 (1,131.05) [11,431.68, 15,777.33]	54,000.91 (371.43) [53,618.95, 55,086.87]
PWP <sub>H</sub> <sub>2</sub>	0.40* (0.01) [0.38, 0.42]	1.38 (0.04) [1.28, 1.43]	0.26 (0.03) [0.22, 0.32]	—
PWP <sub>H</sub> <sub>0cens</sub>	0.46* (0.01) [0.47, 0.52]	2.30 (0.06) [2.17, 2.42]	1.1 (0.11) [0.91, 1.30]	—

<sup>1</sup>Methods: LM, SM, PM, TLmiss, TLcens, PWP<sub>H</sub><sub>2</sub>, PWP<sub>H</sub><sub>0cens</sub>: linear, simulation, penalty, threshold-linear, threshold-linear censored, piecewise Weibull proportional hazard (fixed  $c = 2$ ), and piecewise Weibull proportional hazard censored (fixed  $c = 0$ ) methods.

<sup>2</sup>Heritability on the original scale calculated as proposed by [Korsgaard et al. \(1999\)](#) as  $h^2 = \sigma_a^2 / \left[ \sigma_a^2 + \sigma_d^2 + \left( \pi^2 / 6 \right) \right]$ ;  $h^2$  = heritability,  $\sigma_a^2$ ,  $\sigma_d^2$ , and  $\sigma_e^2$  = additive genetic, contemporary group, and residual variance, respectively.

The nonsurvival approaches showed similar estimates of heritability, except for TLcens method. Heritability estimates for AFC trait from field data reported in the literature also oscillate as observed in the present study, ranging from 0.10 to 0.37 in Brazilian Zebu cattle (Boligon and Albuquerque, 2011; Barrozo et al., 2012). Van Melis et al. (2010) indicate that the high genetic variability observed appears to be a characteristic of *Bos indicus* cattle population, because the breed has rarely been selected for sexual precocity.

Although some authors (Johnston and Bunter, 1996; Morris et al., 2000; Phocas and Sapa, 2004) recommended TLmiss to estimate genetic parameters for reproductive traits in beef cattle, we understand that the inclusion of the other trait (censored status) would modify the results. Thus, comparisons based on the direct interpretation of heritability estimated from threshold-linear and linear models are not suitable in context of the present work.

The interpretation of heritability in the case of survival analysis has been in discussion (Yazdi et al., 2002; Pereira et al., 2007), because there is no linear decomposition of phenotypic variance in the survival model (Guo et al., 2001). It must be noted that heritability on the original scale for PWPB models proposed by Korsgaard et al. (1999) were incorrect, as they were unreasonably high, mainly to a reproductive trait. Pereira et al. (2007) using a sire model, also find high values of heritability while using survival approach for Nellore cattle in the same trait (0.51). Among PWPB methods, highest value of heritability was observed in PWPB<sub>0</sub>cens method. Pereira et al. (2006) affirms that this is a result of using additional information for censored data.

## CONCLUSIONS

The traditional LM, which considers only uncensored records, should be preferred due to its robustness and simplicity. Based on cross-validation analyses, we conclude that the TLmiss could be also a suitable alternative for breeding value prediction, and censored methods did not improve the analysis.

## LITERATURE CITED

- Allison, P. D. 2010. Survival analysis using SAS: a practical guide. Sas Institute, Cary, North Carolina.
- Barrozo, D., M. E. Buzanskas, J. A. Oliveira, D. P. Munari, H. H. Neves, and S. A. Queiroz. 2012. Genetic parameters and environmental effects on temperament score and reproductive traits of Nellore cattle. *Animal*. 6:36–40. doi:10.1017/S1751731111001169
- Berry, D. P., and R. D. Evans. 2014. Genetics of reproductive performance in seasonal calving beef cows and its association with performance traits. *J. Anim. Sci.* 92:1412–1422. doi:10.2527/jas.2013-6723
- Boligon, A. A., and L. G. Albuquerque. 2011. Genetic parameters and relationships of heifer pregnancy and age at first calving with weight gain, yearling and mature weight in Nellore cattle. *Livest. Sci.* 141:12–16. doi:10.1016/j.livsci.2011.04.009
- Bormann, M. J., and D. E. Wilson. 2010. Calving day and age at first calving in Angus heifers. *J. Anim. Sci.* 88:1947–1956. doi:10.2527/jas.2009-2249
- Casellas, J. 2007. Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points. *J. Anim. Breed. Genet.* 124:176–184. doi:10.1111/j.1439-0388.2007.00658.x
- Casellas, J., and R. Bach. 2012. Comparison of linear, skewed-linear, and proportional hazard models for the analysis of lambing interval in Ripollesa ewes. *J. Anim. Sci.* 90:1788–1797. doi:10.2527/jas.2011-4533
- Damgaard, L. H., and I. R. Korsgaard. 2006. A bivariate quantitative genetic model for a linear Gaussian trait and a survival trait. *Genet. Sel. Evol.* 38:45–64. doi:10.1051/gse:200502
- Donoghue, K. A., R. Rekaya, and J. K. Bertrand. 2004. Comparison of methods for handling censored records in beef fertility data: simulation study. *J. Anim. Sci.* 82:351–356. doi:10.2527/2004.822351x
- Garcia, D. A., G. J. Rosa, B. D. Valente, R. Carvalheiro, and L. G. Albuquerque. 2016. Comparison of models for the genetic evaluation of reproductive traits with censored data in Nellore cattle. *J. Anim. Sci.* 94:2297–2306. doi:10.2527/jas.2016-0273
- Gelfand, A., and A. F. M. Smith. 1990. Sampling based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85:398–409. doi:10.2307/2289776
- Guo, S. F., D. Gianola, R. Rekaya, and T. Short. 2001. Bayesian analysis of lifetime performance and prolificacy in Landrace sows using a linear mixed model with censoring. *Livest. Prod. Sci.* 72:243–252. doi:10.1016/S0301-6226(01)00219-6
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their application. *Biometrika*. 57:97–109.
- Hou, Y., P. Madsen, R. Labouriau, Y. Zhang, M. S. Lund, and G. Su. 2009. Genetic analysis of days from calving to first insemination and days open in Danish Holsteins using different models and censoring scenarios. *J. Dairy Sci.* 92:1229–1239. doi:10.3168/jds.2008-1556
- Johnston, D. J. and K. L. Bunter. 1996. Days to calving in Angus cattle: genetic and environmental effects, and covariances with other traits. *Livest. Prod. Sci.* 45:13–22. doi:10.1016/0301-6226(95)00088-7
- Korsgaard, I. R., A. H. Andersen, and J. Jensen. 1999. Discussion of heritability of survival traits. *Interbull Bulletin*. 21:31.
- Korsgaard, I. R., M. S. Lund, D. Sorensen, D. Gianola, P. Madsen, and J. Jensen. 2003. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet. Sel. Evol.* 35:159–183. doi:10.1051/gse:2003002
- Legarra, A., L. Varona, and E. López de Maturana. 2008. TM: Threshold Model. <http://snp.toulouse.inra.fr/~alegarra> (accessed May 27, 2018)

- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092. doi:10.1063/1.1699114
- Morris, C. A., J. A. Wilson, G. L. Bennett, N. G. Cullen, S. M. Hickey, and J. C. Hunter. 2000. Genetic parameters for growth, puberty, and beef cow reproductive traits in a puberty selection experiment. *New Zeal. J. Agr Res.* 43:83–91. doi:10.1080/00288233.2000.9513411
- Pereira, E., H. N. de Oliveira, J. P. Eler, J. A. Silva, and M. H. Van Melis. 2006. Use of survival analysis as a tool for the genetic improvement of age at first conception in Nellore cattle. *J. Anim. Breed. Genet.* 123:64–71. doi:10.1111/j.1439-0388.2006.00567.x
- Pereira, E., H. N. Oliveira, J. P. Eler, J. A. Silva, and M. H. Van Melis. 2007. Comparison among three approaches for evaluation of sexual precocity in Nellore cattle. *Animal.* 1:411–418. doi:10.1017/S1751731107392732
- Phocas, F., and J. Sapa. 2004. Genetic parameters for growth, reproductive performance, calving ease and suckling performance in beef cattle heifers. *J. Anim. Sci.* 79:41–48. doi:10.1017/S1357729800054515
- Silva, D. A., F. F. Silva, H. T. Ventura, V. S. Junqueira, A. A. Silva, R. R. Mota, and P. S. Lopes. 2017. Contemporary groups in the genetic evaluation of Nellore cattle using Bayesian inference. *Pesqui. Agropecu. Bras.* 52:643–651. doi:10.1590/s0100-204x2017000800010
- Smith B. J. 2007. *boa*: an R package for MCMC output convergence assessment and posterior inference. *J. Stat. Softw.* 21:1–37.
- Sorensen, D., and D. Gianola. 2002. *Likelihood, Bayesian, and MCMC methods in quantitative genetics.* Springer Science & Business Media Inc., New York.
- Sorensen, D. A., D. Gianola, and I. R. Korsgaard. 1998. Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications. *Acta Agric. Scand. A-Anim. Sci.* 48:222–229. doi:10.1080/09064709809362424
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, A. van der Linde. 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B.* 64:583–639. doi:10.1111/1467-9868.02022
- Tarrés, J., J. Casellas, and J. Piedrafitra. 2005. Genetic and environmental factors influencing mortality up to weaning of Bruna dels Pirineus beef calves in mountain areas. A survival analysis. *J. Anim. Sci.* 83:543–551. doi:10.2527/2005.833543x
- Tarrés, J., J. Piedrafitra, and V. Ducrocq. 2006. Validation of an approximate approach to compute genetic correlations between longevity and linear traits. *Genet. Sel. Evol.* 38:65–83. doi:10.1051/gse:2005027
- Van Melis, M. H., H. N. Oliveira, J. P. Eler, J. B. Ferraz, J. Casellas, and L. Varona. 2010. Additive genetic relationship of longevity with fertility and production traits in Nellore cattle based on bivariate models. *Genet. Mol. Res.* 9:176–187. doi:10.4238/vol9-1gmr710
- Varona, L., I. Misztal, and J. K. Bertrand. 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. doi:10.2527/1999.7781994x
- Yazdi, M. H., P. M. Visscher, V. Ducrocq, and R. Thompson. 2002. Heritability, reliability of genetic evaluations and response to selection in proportional hazard models. *J. Dairy Sci.* 85:1563–1577. doi:10.3168/jds.S0022-0302(02)74226-4