

Comparison between linear and proportional hazard models for the analysis of age at first lambing in the Ripollesa breed

J. Casellas[†]

Grup de Recerca en Millora Genètica Molecular Veterinària, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

(Received 20 June 2015; Accepted 12 October 2015; First published online 9 November 2015)

Age at first lambing (AFL) plays a key role on the reproductive performance of sheep flocks, although there are no genetic selection programs accounting for this trait in the sheep industry. This could be due to the non-Gaussian distribution pattern of AFL data, which must be properly accounted for by the analytical model. In this manuscript, two different parameterizations were implemented to analyze AFL in the Ripollesa sheep breed, that is, the skew-Gaussian mixed linear model (sGML) and the piecewise Weibull proportional hazards model (PWPH). Data were available from 10 235 ewes born between 1972 and 2013 in 14 purebred Ripollesa flocks located in the north-east region of Spain. On average, ewes gave their first lambing short after their first year and a half of life (590.9 days), and within-flock averages ranged between 523.4 days and 696.6 days. Model fit was compared using the deviance information criterion (DIC; the smaller the DIC statistic, the better the model fit). Model sGML was clearly penalized (DIC = 200 059), whereas model PWPH provided smaller estimates and reached the minimum DIC when one cut point was added to the initial Weibull model (DIC = 132 545). The pure Weibull baseline and parameterizations with two or more cut points were discarded due to larger DIC estimates (>134 200). The only systematic effect influencing AFL was the season of birth, where summer- and fall-born ewes showed a remarkable shortening of their AFL, whereas neither birth type nor birth weight had a relevant impact on this reproductive trait. On the other hand, heritability on the original scale derived from model PWPH was high, with a model estimate place at 0.114 and its highest posterior density region ranging from 0.079 and 0.143. As conclusion, Gaussian-related mixed linear models should be avoided when analyzing AFL, whereas model PWPH must be viewed as better alternative with superior goodness of fit; moreover, the additive genetic background underlying this reproductive trait supports its inclusion into current genetic selection programs given its economic importance.

Keywords: age at first lambing, Gaussian linear model, proportional hazard model, ripollesa breed, sheep

Implications

Despite age at first lambing plays a key role on the reproductive performance of sheep flocks, its inclusion in genetic selection programs remains scarce. This research focused on the suitability of two different parameterizations for the genetic evaluation of this trait, as well as the magnitude and statistical relevance of both genetic and environmental sources of variation. Results suggested a remarkable advantage for survival analysis techniques in contrast to mixed linear models, and the estimated heritability ($h^2 = 0.114$) encouraged the inclusion of age at first lambing as selection objective in sheep breeding programs.

Introduction

Cost of producing and maintaining breeding ewes can be high in the sheep industry. Indeed, ewe lifespan can be partitioned into two basic time stages, the costly raising period from birth to

first lambing, and the subsequent 'productive life' until ewe disposal. Replacement ewes produce no output of economic relevance (other than wool in some breeds) until they reach sexual maturity and give birth for the first time. Within this context, the length of the time interval between birth and first lambing become of special relevance for profitability of sheep flocks worldwide. Studies about age at first lambing (AFL) are not particularly abundant in the scientific literature, although highly heterogeneous values have been reported for this trait; within-flock averages can fluctuate from less than a year of age (Gallivan *et al.*, 1993) to almost 23 months of age (Gabiña, 1989).

As other lifetime traits in animal species, AFL has a skewed phenotypic distribution with evident over-representation of the right tail, thus discouraging its analysis under standard Gaussian-related mixed linear models (El-Saied *et al.*, 2005). Both the generalization of Gaussian models to non-symmetrical patterns and the implementation of non-linear models on the basis of skew parametric distributions have been suggested in

[†] E-mail: joaquim.casellas@uab.cat

previous studies (Varona *et al.*, 2008; Casellas and Bach, 2012; García *et al.*, 2014). The main objective of this research was to evaluate two main parameterizations for the analysis of AFL in the Ripollesa sheep breed, the skew-Gaussian mixed linear model (sGML, Sahu *et al.*, 2003) and the piecewise Weibull proportional hazards model (Casellas, 2007). In addition, environmental and genetic sources of variation were characterized in order to provide guidelines to shepherds to optimize flock management and to decide whether AFL can be considered as a reliable selection objective in current breeding schemes or not.

Material and methods

Animal Care and Use Committee approval was not necessary for this study because analyses were performed on existing field data obtained under standard farm management from commercial breeders. Both productive and reproductive data were collected by the stockbreeders themselves and administrated by the Ripollesa Breed Society (Associació Nacional de Criadors d'Oví de Raça Ripollesa (ANCRI), Monells, Spain) within the context of its performance recording scheme.

Field data source

The Ripollesa is a medium-sized sheep breed (rams, 75 to 90 kg of live BW; ewes, 50 to 65 kg of BW) reared in north-east Spain and managed under semi-extensive conditions. This breed is a representative example of the meat-type sheep industry from the Mediterranean basin, the 'pascual'-type lamb being its typical commercial product (22 to 24 kg of live BW at slaughter). Ripollesa individuals have dark brown or black marks on the head and legs, and white semi-fine wool with closed fleece (Esquivelzeta *et al.*, 2011).

This research focused on the age at first lambing (AFL) defined as the number of days between the ewe's birth date and the date of its first lambing. After discarding records shorter than 300 days and longer than 1000 days, AFL data from 10 235 Ripollesa ewes became available. These ewes were born between 1972 and 2013 in 14 purebred Ripollesa flocks with pedigree connection. All flocks located in the provinces of Girona and Barcelona (north-east Spain), they were active members of ANCRI, and contributed data to the official performance recording scheme; these data (e.g. ewe identification number, lambing date, lamb identification number, lamb sex and birth weight, among others) were registered by the shepherds themselves and further administrated by ANCRI. Within-flock summary statistics for AFL data are shown in Table 1, and Figure 1 illustrated the distribution pattern of raw AFL records. Note that all Ripollesa flocks were managed under continuous mating with several rams and, as usual in sheep flocks, ewe-lambs were kept apart from rams until they reached at least 4 to 5 months of age.

The pedigree file included 12 399 individuals, 142 of them being rams contributing between one and 45 daughters each. Indeed, dam was known for all ewes with AFL data, whereas sire information was restricted to 984 ewes (9.6%). Given that several rams served at the same time, paternity

Table 1 Summary of the age at first lambing (days) data from the Ripollesa breed

	<i>n</i>	Mean ± SE	Quantiles (%)		
			10	50	90
Herd 1	660	569.5 ± 6.9	366	569	816
Herd 2	61	607.5 ± 18.9	447	598	830
Herd 3	1180	581.2 ± 3.9	414	565	763
Herd 4	1018	696.0 ± 4.2	505	717	869
Herd 5	1218	523.4 ± 4.0	385	468	744
Herd 6	1484	609.7 ± 3.8	416	608	819
Herd 7	47	534.0 ± 26.6	431	473	736
Herd 8	223	619.3 ± 8.3	464	625	785
Herd 9	2312	561.6 ± 2.7	397	557	741
Herd 10	126	588.2 ± 8.0	422	550	680
Herd 11	125	654.2 ± 13.6	435	634	863
Herd 12	302	554.7 ± 7.5	427	517	776
Herd 13	490	599.5 ± 6.1	402	573	754
Herd 14	989	626.4 ± 4.6	421	646	809
Overall	10 235	590.9 ± 1.5	400	517	794

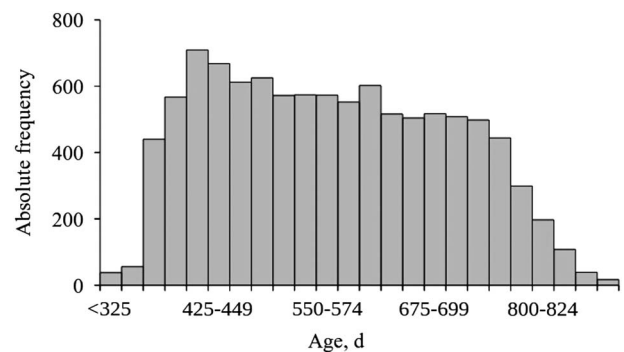


Figure 1 Histogram of age at first lambing data from the Ripollesa sheep breed.

information was recovered by genotyping replacement individuals for a custom 48-single nucleotide polymorphisms panel (Casellas *et al.*, 2014).

Bayesian models

Given the skewed distribution pattern of AFL data shown in Figure 1, analyses were performed on the basis of two different parameterizations characterized by Gaussian-related (Gianola and Fernando, 1986) and proportional hazard-related (Ducrocq *et al.*, 1988) hierarchical models (see below). Both approaches are able to fit data under unsymmetrical distributions, although the realized goodness of fit can differ substantially (Casellas and Bach, 2012). Without loss of generality, all analyses accounted for three independent sources of variation, that is systematic (**b**), permanent environmental (**p**), and additive genetic (**a**) effects. Vector **b** included season of birth of the ewe with four levels (winter, 34.9%; spring, 13.7%; summer, 15.0%; fall, 36.4%), birth type of the ewe with two levels (single, 58.1%; multiple, 41.9%), and birth weight of the ewe as a linear covariate (mean ± SEM, 3.92 ± 0.01 kg), whereas **p** accounted for appropriate flock-year-season effects with 120 levels.

sGML model. Sahu *et al.* (2003) modified the standard Gaussian process in order to accommodate skew patterns by the inclusion of a single additional parameter (Λ). This extra parameter accounts for right- ($\Lambda < 0$) and left-tail over-expression ($\Lambda > 0$), as well as for symmetric Gaussian patterns ($\Lambda = 0$). The following hierarchical structure was assumed to generalize the sGML model,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{p} + \mathbf{Z}_2\mathbf{a} + \mathbf{e}$$

where \mathbf{y} was the vector of phenotypic records, \mathbf{X} , \mathbf{Z}_1 and \mathbf{Z}_2 were appropriate incidence matrices, and \mathbf{e} was the residual term. The Bayesian implementation of this hierarchical structure was assumed as follows:

$$\rho(\mathbf{b}, \mathbf{p}, \mathbf{a}, \sigma_a^2, \sigma_p^2, \sigma_e^2, \Lambda) \propto \rho(\mathbf{y}|\mathbf{b}, \mathbf{p}, \mathbf{a}, \sigma_e^2, \Lambda) \rho(\mathbf{b}) \rho(\mathbf{p}|\sigma_p^2) \rho(\sigma_p^2) \times \rho(\mathbf{a}|\mathbf{A}, \sigma_a^2) \rho(\sigma_a^2) \rho(\sigma_e^2) \rho(\Lambda)$$

where σ_e^2 was the residual variance, σ_a^2 was the additive genetic variance, σ_p^2 was the permanent environmental variance, and \mathbf{A} was the numerator relationship matrix (Wright, 1922). Lack of sire information was accommodated into \mathbf{A} by defining appropriate yearly within-flock genetic groups of unknown parents (Westell *et al.*, 1988). The $\rho(\mathbf{y}|\mathbf{b}, \mathbf{p}, \mathbf{a}, \sigma_e^2, \Lambda)$ term was skew-Gaussian (Sahu *et al.*, 2003; Varona *et al.*, 2008),

$$\rho(\mathbf{y}|\mathbf{b}, \mathbf{p}, \mathbf{a}, \sigma_e^2, \Lambda) = \prod_{i=1}^n \frac{2}{\sqrt{\sigma_e^2 + \Lambda^2}} \phi\left(\frac{e_i}{\sqrt{\sigma_e^2 + \Lambda^2}}\right) \Phi\left(\frac{\Lambda e_i}{\sigma_e \sqrt{\sigma_e^2 + \Lambda^2}}\right),$$

n is the number of phenotypic records in \mathbf{y} , ϕ and Φ the density function and the distribution function of a standard normal distribution with kernel as defined between parentheses, respectively, and $e_i = y_i - (\mathbf{x}_i\mathbf{b} + \mathbf{z}_{1,i}\mathbf{p} + \mathbf{z}_{2,i}\mathbf{a})$; note that \mathbf{x}_i , $\mathbf{z}_{1,i}$ and $\mathbf{z}_{2,i}$ were the i th rows of matrices \mathbf{X} , \mathbf{Z}_1 and \mathbf{Z}_2 , respectively. Priors for \mathbf{p} and \mathbf{a} were stated as $\rho(\mathbf{p}|\sigma_p^2) = MVN(\mathbf{0}, \mathbf{I}_p\sigma_p^2)$ and $\rho(\mathbf{a}|\mathbf{A}, \sigma_a^2) \sim MVN(\mathbf{0}, \mathbf{A}\sigma_a^2)$, \mathbf{I}_p being an appropriate identity matrix with dimensions equal to the number of levels in \mathbf{p} . Prior distributions for \mathbf{b} , σ_a^2 , σ_p^2 and σ_e^2 were assumed scaled inverted χ^2 distributions with hyper-parameters $\nu = -2$ and $S^2 = 0$. Following Varona *et al.* (2008), $\rho(\Lambda)$ was assumed improper uniform between $-\infty$ and ∞ .

Piecewise Weibull proportional hazards model (PWPH). This model (PWPH) was adapted to the context of animal breeding by Casellas (2007) on the basis of an original idea by Nguyen *et al.* (1984) and Noura and Read (1990). The parametrization relied on a standard Weibull model with parameters ρ and λ (Ducrocq *et al.*, 1988), and included a time-dependent effect which changed in magnitude at given change points along the parametric space. Following the Bayesian development by Casellas (2007),

$$\rho(\mathbf{b}, \mathbf{p}, \mathbf{a}, \sigma_a^2, \sigma_p^2, \rho, \lambda, \tau, \xi) \propto \rho(\mathbf{y}|\mathbf{b}, \mathbf{p}, \mathbf{a}, \rho, \lambda, \tau, \xi) \rho(\mathbf{b}) \rho(\mathbf{p}|\sigma_p^2) \rho(\sigma_p^2) \times \rho(\mathbf{a}|\mathbf{A}, \sigma_a^2) \rho(\sigma_a^2) \rho(\rho) \rho(\lambda) \rho(\tau) \rho(\xi)$$

τ was the vector of c change points ($c \geq 2$) and ξ was the vector of $c-1$ regression parameters for the time-dependent

effect. Focusing on the i th phenotypic record (y_i), its *a priori* distribution was stated as $\rho(y_i|\mathbf{b}, \mathbf{p}, \mathbf{a}, \rho, \lambda, \tau, \xi) = [h_0(t)\beta] [S_0(t)^\beta]$, where $h_0(t) = \lambda\rho(\lambda y_i)^{\rho-1} \xi_k$ was the baseline hazard function derived from a piecewise Weibull density,

$$S_0(t) = \exp\left\{-\left[\lambda^\rho \xi_k (y_i^\rho - \tau_k^\rho) + \sum_{j=1}^{k-1} \lambda^\rho \xi_j (\tau_{j+1}^\rho - \tau_j^\rho)\right]\right\}$$

was the piecewise baseline survival function with $y_i > \tau_k$, and $\beta = \exp(\mathbf{x}_i\mathbf{b} + \mathbf{z}_{1,i}\mathbf{p} + \mathbf{z}_{2,i}\mathbf{a})$ was a scalar accounting for systematic, permanent and additive genetic effects. Following Damgaard and Korsgaard (2006), improper uniform priors were assigned to ρ , λ , τ and ξ , whereas remaining *a priori* distributions were defined as for the sGML model.

Sampling from joint posterior distributions. Age at first lambing data were independently analyzed under models sGML and PWPH. Moreover, the number of change points of the Weibull baseline function must be chosen in advance for PWPH model and thus, seven independent analyses were performed from $c = 2$ to $c = 8$ (PWPH2 to PWPH8). All the unknown parameter in models were sampled from their marginal posterior distributions by appropriate Gibbs sampling (Gelfand and Smith, 1990) or Metropolis-Hastings (Metropolis *et al.*, 1953). Three independent Monte Carlo Markov chains (MCMC) with 1 000 000 iterations were run for each analysis, and the first 100 000 iterations were discarded as burn-in. The suitability of the length of the burn-in period was evaluated by both visual inspection and the method of Raftery and Lewis (1992) on the sampling path of σ_a^2 . A total of 18 000 samples of model parameters were saved from each chain with a lag interval of 50 iterations. All analyses were performed in a server with two Intel® Xeon® E5-2400 processors with a CentOS 6.6 Linux operating system. In terms of central processing unit (CPU) time, models sGML and PWPH2 spent less than eight hours for each independent 1 000 000-iteration MCMC. Computing demand rose for models PWPH3 (~18 h) to PWPH8 (~52 h) as the number of change points increased.

Model comparison

The two parameterizations implemented in this manuscript (sGML and PWPH models) where compared by means of the deviance information criterion (DIC; Spiegelhalter *et al.*, 2002). Note that DIC measures two different sources of model performance in terms of the posterior expectation of the Bayesian deviance (model fit) and the effective number of parameters (model complexity). This statistic can be computed from each MCMC and allows for a straightforward comparison between models where smaller DIC values indicate a better fit and a lesser degree of model complexity. Differences of at least three to five DIC units are generally considered as statistically relevant (Spiegelhalter *et al.*, 2002). On the other hand, inference on predicted breeding values should play a key role for genetic selection programs. Within this context, Pearson's correlation coefficients between modal estimates of predicted breeding values from

different model parameterizations were calculated at the end of the MCMC process.

Results and discussion

Age at first lambing in the Ripollesa breed

Our analyses focused on 14 purebred Ripollesa flocks and 10 235 ewes that averaged 591 days at first lambing, and within-flock averages ranged from 523 days to 696 days. These averages were free from upward biases due to infertile ewes (i.e. ewes incapable of getting pregnant that joined the flock until detected and culled by the shepherd) because only ewes with at least one registered lambing contributed phenotypic data, and censored records (Cox, 1972) were not considered. This must not be viewed as a limitation of the parameterizations used in this research (both PWPH (Casellas, 2007) and sGML models (Guo *et al.*, 2001) accommodate right-censored data) but an effort to guarantee field data quality. Age at first lambing in the Ripollesa breed seemed to be placed at the upper boundary of this reproductive trait, although the removal of all right-censored AFL records (i.e. ewes culled before their first lambing) may have biased downwards the average AFL in the Ripollesa breed. Nevertheless, less than 200 ewes were removed from the data set because they were culled at an age of two or more years and without lambing records. Other populations provided similar performances such as the Rasa Aragonesa (528 days to 688 days; Gabiña, 1989) and the Churra breed (604 days; El-Saied *et al.*, 2005), which are also autochthonous breeds from Spain, the Yankasa breed from Nigeria (533 days; Osuhor *et al.*, 1997), or the Awassi breed reared in Spanish farms (~550 days; de la Fuente *et al.*, 2006). Nevertheless, age at first lambing in the Ripollesa breed was clearly delayed when compared with other dairy sheep breeds reared in Spain like Assaf (~450 days), Churra (~420 days; de la Fuente *et al.*, 2006) and Lacaune (433 days; Hernandez *et al.*, 2011). Smaller ages at first lambing were also reported in Finnish Landrace × Targhee (377 days) and Romanov × Targhee (365 days) flocks (Gallivan *et al.*, 1993). As a whole, these estimates for age at first lambing revealed a large scope for AFL improvement in the Ripollesa sheep breed.

Although the average of AFL data gave us an idea about the onset of the reproductive activity in the Ripollesa breed in a broad sense, both the histogram shown in Figure 1 and flock-by-flock specific data about more extreme quantiles

(10% and 90%; Table 1) provided a more detailed characterization of the phenotypic dispersion pattern of AFL. Modal estimate (411 days; Figure 1) was clearly smaller than average estimate in all flocks, this leading to the overextension of the right-tail of the AFL distribution. The absolute frequency for AFL slowly decreased after the modal peak and until ewe-lambs aged ~2 years, showing an accentuated decline after that. The low incidence of first lambing from 2-year-old ewes (or even older individuals; Figure 1) could be linked with the preferential culling of ewes that failed to give birth by first time at these ages. As a whole, this depicted an atypical and non-Gaussian distribution patterns that must be properly accounted for by the analytical model.

Comparison between competing models

Deviance information criterion estimates regarding both Gaussian-related and proportional hazard-related parameterizations for the analysis of the AFL data are shown in Table 2. Although model sGML provided a more flexible framework when both positive and negative asymmetry in the data must be accommodated (i.e. over-expression of the right and left tail of the distribution, respectively), model PWPH was favored. Analyses did not support model sGML with an average DIC of 200 059 units when compared with PWPH models (<136 000 units); model fit under linear mixed model parametrizations was even worse when restricting the analysis to scenarios under perfect symmetry ($L = 0$; DIC = 245 703), although this could be anticipated given the modal and highest posterior density region at 95% (HPD95) obtained for L under model sGML (Table 3). Our estimates about goodness of fit suggested that linear models must be ruled out from AFL data analysis, whereas PWPH models were an appealing alternative with preferable statistical performances. This conclusion agreed with previous studies that advocated the superiority of survival analysis techniques when handling time interval data such as lambing interval in sheep (Casellas *et al.*, 2012), age at first calving (García *et al.*, 2014), or fertility- and longevity-related traits in cattle (Caraviello *et al.*, 2004; Schneider *et al.*, 2005, 2006; MacNeil and Vukasinovic, 2011). Within this context, the implementation of proportional hazards models seems a reasonable alternative for AFL data. To the best of our knowledge, this methodological approach had not been previously implemented for the analysis of birth-to-first parturition data in sheep, whereas few studies were performed in other ruminant species (García *et al.*, 2014).

Table 2 Deviance information criterion (Spiegelhalter *et al.*, 2002) for age at first lambing in the Ripollesa sheep breed under a skew-Gaussian mixed linear model (sGML) and piecewise Weibull hazard model with i change points in the baseline function (PWPH $_i$)

	sGML	PWPH ₂	PWPH ₃	PWPH ₄	PWPH ₅	PWPH ₆	PWPH ₇	PWPH ₈
Chain 1	200 075	134 238	132 537	134 343	134 638	134 716	134 944	135 388
Chain 2	200 054	134 245	132 549	134 340	134 649	134 720	134 956	135 400
Chain 3	200 049	134 228	132 550	134 355	134 638	134 692	134 942	135 392
Mean	200 059	134 237	132 545	134 346	134 642	134 710	134 947	135 393
SE	8	5	4	5	4	8	4	4

Table 3 Mode (highest posterior density region at 95%) for model-specific parameters and variance components from the analysis of age at first lambing in the Ripollesa sheep breed under a skew-Gaussian mixed linear model (sGML) and a piecewise Weibull proportional hazard model with three cut points (PWPH3)

	Model	
	PWPH3	sGML
PWPH3-specific parameters ¹		
ρ	8.496 (8.057 to 8.795)	– ²
$\lambda \times 1000$	1.252 (1.219 to 1.277)	–
τ_1	0	–
ξ_1	1	–
τ_2	351 (336 to 368)	–
ξ_2	7.996 (6.564 to 9.616)	–
τ_3	900	–
sGML-specific parameters ³		
Λ	–	223.6 (217.8 to 230.2)
σ_e^2	–	760.4 (534.8 to 1018.0)
Variance components ⁴		
σ_p^2	1.349 (1.004 to 1.784)	2.914 (2.242 to 3.837)
σ_a^2	0.299 (0.255 to 0.332)	1.034 (0.874 to 1.191)
h^2	0.114 (0.079 to 0.143)	0.045 (0.037 to 0.053)

¹The baseline function was modeled within the context of a Weibull density with appropriate shape (ρ) and scale parameters (λ). Moreover, additional flexibility was gained by including a time-dependent effect defined between $\tau_1 = 0$ and $\tau_3 = 900$, and with magnitude $\xi_1 = 1$ and ξ_2 that changes at τ_2 .

²Not applicable.

³The Gaussian mixed linear model included an extra parameter to accommodate skew-distributed residuals (Λ); note that σ_e^2 was the residual variance.

⁴Under model PWPH3, heritability (h^2) was calculated on the original scale by applying $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_p^2 + 1)$ as described by Yazdi *et al.* (2002). On the other hand, the formula $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_p^2 + \sigma_e^2 + \Lambda^2(1-2I/n))$ can be applied to the sGML model (Sahu *et al.*, 2003; Varona *et al.*, 2008).

Focusing on PWPH parameterizations, model PWPH3 reached the smallest DIC estimate (Table 2), this being almost 1700 DIC units smaller than the following model (PWPH2). Moreover, models PWPH4 to PWPH8 revealed a progressive increase of DIC estimates, adding on average 262 DIC units for each extra change point in the model. This progressive penalization of more parameterized models suggested that marginal benefits of additional change points did not counterbalance model over-parameterization, and PWPH3 was a more parsimonious choice for AFL data from the Ripollesa breed.

Capturing skewness from AFL data

Both sGML and PWPH3 parameterizations evidenced the right-tail over-expressed distribution of AFL data in the Ripollesa sheep breed (Figure 1). Whereas sGML model provided a positive modal estimate of the skewness parameter (Λ) with its HPD95 clearly discarding the null estimate (217.8 to 230.2; Table 3), model PWPH3 smoothed out the slope of the Weibull baseline function after 351 days of age (HPD95, 336 to 368). This accommodated the rapid accumulation of ewes lambing short before their first birthday (Figure 1), accentuating the slope of the Weibull distribution characterized by the initial

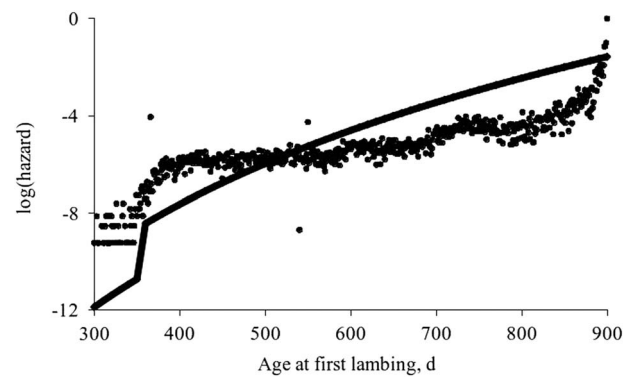


Figure 2 Logarithm of the hazard computed under modal estimates of baseline parameters from model PWPH3 (solid line) and the non-parametric Kaplan–Meier approach (Kaplan and Meier, 1958; black dots).

ρ and λ parameters; note that the magnitude of the time-dependent effect (ξ) on the baseline function raised from 1 ($t \leq \tau_2$) to 7.996 ($\tau_2 < t$), and larger estimates from survival analysis must be interpreted as an increased risk of failure or, in our case, a higher probability of lambing in short. A graphical representation of the hazard function is shown in Figure 2 and compared with the non-parametric Kaplan–Meier estimate (Kaplan and Meier, 1958). Note that both approaches provide similar estimates and departures must be partially attributable to the absence of systematic, permanent and genetic factors when computing the baseline hazard function under model PWPH3; the Kaplan–Meier estimate is computed from raw AFL data and all additional sources of variation are implicitly included. The need for this kind of additional modeling of the standard Weibull baseline function has been previously demonstrated on longevity (Tarrés *et al.*, 2005; Casellas *et al.*, 2008) and fertility data (Casellas and Bach, 2012), although this should not be viewed as a weakness of the survival analysis techniques but a remarkable advantage in terms of flexibility of the PWPH (Casellas, 2007) model.

Systematic and genetic influences on AFL

Although three systematic effects were included in the analytical model, only the season of birth of the ewe had relevant influences on the AFL under PWPH3 parameterization. Both birth type (single or multiple) and birth weight provided close-to-zero regression coefficients with their HPD95 clearly including the null estimate. This agreed with previous analyses in the Yankasa breed of sheep from Nigeria (Osuhor *et al.*, 1997), where the same systematic effects were also discarded. Focusing on the season of birth, spring-born ewes showed the longest AFL although regression coefficient for winter-born ewes did not differ ($\beta = 0.074$; HPD95 = -0.028 to 0.164) from spring-born ewes (this was fixed to 0 as reference). Summer- ($\beta = 0.159$; HPD95 = 0.060 to 0.242) and fall-born ewes ($\beta = 0.394$; HPD95 = 0.280 to 0.513) were more predisposed to shorter AFL. These estimates agreed with the pattern shown by the raw averages of AFL data from our data set for winter- (590 ± 2 days), spring- (627 ± 2 days), summer- (586 ± 2 days) and fall-born ewes (580 ± 2 days).

It could be suggested that spring- and winter-born ewes were slightly penalized in terms of AFL because they were expected to give birth as a first time at counter-season; summer- and fall-born ewes-lambs took advantage of the natural breeding season, this providing better reproductive performances in terms of AFL.

Models sGML and PWP3 also showed remarkable departures from each other when capturing genetic and environmental random sources of variation. Whereas the mixed linear parameterization suggested that flock-year-season variance was almost three times that of the additive genetic variance, model PWP3 provided even larger differences with a modal estimates of 0.299 for the additive genetic variance and 1.394 for the flock-year-season variance (Table 3). Heritability estimates from both parameterizations differed, this being clearly larger from PWP3 ($h^2 = 0.114$) than from sGML ($h^2 = 0.045$). Unfortunately, there is shortage of heritability estimates for AFL in sheep in the scientific literature, and our estimate from the sGML model was slightly higher than the one obtained by Lôbo *et al.* (2009) in a multibreed population ($h^2 = 0.04$). Nevertheless, heritabilities from linear and proportional hazard models cannot be comparable at all. The advantage of proportional hazard models when capturing additive genetic variance from time-interval traits was previously reported by Caraviello *et al.* (2004) in dairy cattle and Casellas and Bach (2012) in sheep, this potentially leading to more rapid genetic progress (Korsgaard *et al.*, 2002). It is important to note that predicted breeding values from mixed linear and proportional hazard models cannot be directly compared because they are

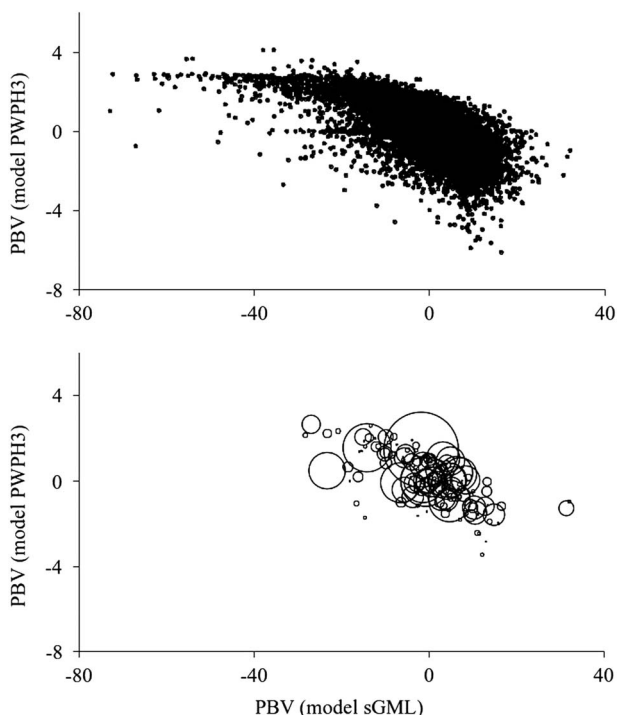


Figure 3 Plot of predicted breeding values (PBV) by the skew-Gaussian mixed linear model (sGML; x-axis) against PBV by piecewise Weibull proportional hazard model with three cut points in the baseline function (PWP3) for all animals (upper panel) and sires (lower panel; the diameter of each spot characterized the number of evaluated daughters, from 1 to 45).

diametrically opposed. Shorter AFL intervals must be obtained by selecting those individuals with large and positive (model sGML) or large and negative (model PWP3) predicted breeding values. Moreover, their magnitude can also differ because the sGML model assumes an additive structure for systematic and random effects (Sahu *et al.*, 2003), whereas PWP3 models rely on an exponential parameterization (Cox, 1972). Within the context of the Ripollesa data set, departures in terms of predicted breeding values from sGML and PWP3 are shown in Figure 3; as expected, the plot suggests a negative correlation of -0.636 , even higher if only sires were considered (-0.694), and departures increased for genetic background predisposing to longer AFL. Within this context, differences between sGML and PWP3 models in terms of response to selection must be larger if selecting for longer AFL than for shorter AFL. Although fertility-related traits have been historically considered as barely heritable (Fogarty, 1995), the heritability under model PWP3 endorses AFL as a candidate trait for current selection programs in sheep.

Acknowledgments

The author is indebted to the Ripollesa breed society (*Associació Nacional de Criadors d'Oví de Raça Ripollesa*, Monells, Spain) for contributing field data. The research contract of J. Casellas was partially financed by Spain's *Ministerio de Ciencia e Innovación* (program *Ramón y Cajal*, reference RYC-2009-04049).

References

- Caraviello DZ, Weigel KA and Gianola D 2004. Comparison between a Weibull proportional hazards model and a linear model for predicting the genetic merit of US Jersey sires for daughter longevity. *Journal of Dairy Science* 87, 1469–1476.
- Casellas J 2007. Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points. *Journal of Animal Breeding and Genetics* 124, 176–184.
- Casellas J and Bach R 2012. Comparison of linear, skewed-linear, and proportional hazard models for the analysis of lambing interval in Ripollesa ewes. *Journal Animal Science* 90, 1788–1797.
- Casellas J, Fina M, Bach R and Piedrafita J 2014. Null genetic differentiation among flocks contributing to the Ripollesa sheep herdbook. *Proceedings of the World Congress on Genetics applied to Livestock Production*, Vancouver, Canada.
- Casellas J, Varona L, Ibáñez-Escriche N, Quintanilla R and Noguera JL 2008. Skew distribution of founder-specific inbreeding depression effects on the longevity of Landrace sows. *Genetics Research* 90, 499–508.
- Cox DR 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 34, 187–220.
- Damgaard LH and Korsgaard IR 2006. A bivariate quantitative genetic model for a linear Gaussian trait and a survival trait. *Genetics, Selection, Evolution* 38, 45–64.
- de la Fuente LF, Gabiña D, Carolino N and Ugarte E 2006. The Awassi and Assaf breeds in Spain and Portugal. 57th Annual Meeting of the European Federation of Animal Science, Antalya, Turkey.
- Ducrocq W, Quaas RL, Pollak EJ and Casella G 1988. Length of productive life of dairy cows. 1. Justification of a Weibull model. *Journal of Dairy Science* 71, 3061–3070.
- El-Saied UM, de la Fuente LF, Carriedo JA and San Primitivo F 2005. Genetic and phenotypic parameter estimates of total and partial lifetime traits for dairy ewes. *Journal of Dairy Science* 88, 3265–3272.
- Esquivelzeta C, Fina M, Bach R, Madruga C, Caja G, Casellas J, Piedrafita J 2011. Morphological analysis and subpopulation characterization of Ripollesa sheep breed. *Animal Genetic Resources* 49, 9–17.
- Fogarty NM 1995. Genetic parameters for live weight, fat and muscle measurements, wool production and reproduction in sheep: A review. *Animal Breeding Abstracts* 63, 101–143.

- Gabiña D 1989. Improvement of the reproductive performance of Rasa Aragonesa flocks in frequent lambing systems. I. Effects of management system, age of ewe and season. *Livestock Production Science* 22, 69–85.
- Gallivan C, Kemp RA, Berger YM and Young LD 1993. Comparison of Finnish Landrace and Romanov as prolific breeds in a terminal-sire crossbreeding system. *Journal of Animal Science* 71, 2910–2918.
- Garcia DA, Carvalheiro R, Rosa GJM, Valente BD and Albuquerque LG 2014. Analysis of age at first calving dealing with censored data. In *Proceedings, of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada*.
- Gelfand A and Smith AFM 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Gianola D and Fernando RL 1986. Bayesian methods in animal breeding theory. *Journal of Animal Science* 63, 217–244.
- Guo S, Gianola D, Rekaya R, Short T 2001. Bayesian analysis of lifetime performance and prolificacy in Landrace sows using a linear mixed model with censoring. *Livestock Production Science* 72, 243–252.
- Hernandez F, Elvira L, Gonzalez-Martin JV, Gonzalez-Bulnes A and Astiz S 2011. Influence of the age at first lambing on reproductive and productive performance of Lacaune dairy sheep under an intensive management system. *Journal of Dairy Research* 78, 160–167.
- Kaplan EL and Meier P 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457–481.
- Korsgaard IR, Andersen AH and Jensen J 2002. Prediction error variance and expected response to selection, when selection is based on the best predictor for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits. *Genetics, Selection, Evolution* 34, 307–333.
- Lôbo AM, Lôbo RN, Paiva SR, de Oliveira SM and Falcó O 2009. Genetic parameters for growth, reproductive and maternal traits in a multibreed meat sheep population. *Genetics and Molecular Biology* 32, 761–770.
- MacNeil MD and Vukasinovic N 2011. A prototype national cattle evaluation for sustained reproductive success in Hereford cattle. *Journal of Animal Science* 89, 1712–1718.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH and Teller E 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21, 1087–1092.
- Nguyen HT, Rogers GS and Walker EA 1984. Estimation in change-point hazard rate models. *Biometrika* 71, 299–304.
- Noura AA and Read KLQ 1990. Proportional hazards change-point models in survival analysis. *Applied Statistics* 39, 241–253.
- Osuhor CU, Osinowo OA, Nwagu BI, Denny FO and Abdullahi-Adee A 1997. Factors affecting age at first lambing in Yankasa ewes. *Animal Reproduction Science* 47, 205–209.
- Raftery AE and Lewis SM 1992. How many iterations in the Gibbs sampler?. In *Bayesian Statistics IV*. (ed. Bernardo JM, Berger JO, Dawid AP and Smith AFM), pp. 763–773. Oxford University Press, Oxford, UK.
- Sahu SK, Dey DK and Branco MD 2003. A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics* 31, 129–150.
- Schneider Mdel P, Strandberg E, Ducrocq V and Roth A 2005. Survival analysis applied to genetic evaluation for fertility in dairy cattle. *Journal of Dairy Science* 88, 2253–2259.
- Schneider Mdel P, Strandberg E, Ducrocq V and Roth A 2006. Genetic evaluation of the interval from first to last insemination with survival analysis and linear models. *Journal of Dairy Science* 89, 4903–4906.
- Spiegelhalter DJ, Best NG, Carlin BP and van der Linder A 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* 64, 583–639.
- Tarrés J, Casellas J and Piedrafita J 2005. Genetic and environmental factors influencing mortality up to weaning of Bruna dels Pirineus beef calves in mountain areas. A survival analysis. *Journal of Animal Science* 83, 543–551.
- Varona L, Ibáñez-Escriche N, Quintanilla R, Noguera JL and Casellas J 2008. Bayesian inference of quantitative traits using skewed distributions. *Genetics Research* 90, 179–190.
- Westell RA, Quaas RL and Van Vleck LD 1988. Genetic groups in an animal model. *Journal of Dairy Science* 71, 1310–1318.
- Wright S 1922. Coefficient of inbreeding and relationship. *The American Naturalist* 56, 330–338.
- Yazdi MH, Visscher PM, Ducrocq V and Thompson R 2002. Heritability, reliability of genetic evaluations and response to selection in proportional hazard models. *Journal of Dairy Science* 85, 1563–1577.