

# Calving Ease Evaluation of French Dairy Bulls with a Heteroskedastic Threshold Model with Direct and Maternal Effects

**Vincent Ducrocq**

*Station de Génétique Quantitative et Appliquée,  
Institut National de la Recherche Agronomique,  
78352 Jouy en Josas, France*

---

## Abstract

Calving ease scores (1 = no assistance, 2 = easy calving, 3 = difficult calving) from two French dairy breeds (NO = Normande and MO = Montbéliarde) were analysed using a heteroskedastic threshold model as proposed by Foulley and Gianola (1996, *Genet. Sel. Evol.* 2:249). A satisfactory model was first selected to describe the effects of environmental factors on the location and dispersion parameters of the underlying normal variable. Four random effects were then added: sire of calf, sire of dam, dam (within sire of dam) and herd-year-season, assuming homogeneous ratios of variance components across environments. Relatively low estimates of heritabilities were obtained on the underlying scale: 5.4% (resp., 5.4% for MO) for the direct effect and 3.1% (resp., 2.7%) for the maternal effect, with a slightly negative correlation between the two. A routine evaluation of French dairy bulls based on this statistical model is now operational.

---

## 1. Introduction

Reduction of dystocia is desirable in dairy cattle because calving disorders induce important economic losses via increased mortality, decreased fertility and higher labour costs. Such a reduction can be achieved through the identification of extreme sires and their proper use in A.I., for example avoiding risky matings (e.g., with heifers). This has led to the development of routine genetic evaluations on dystocia in many countries.

A proper analysis of calving conditions recorded on farm on a subjective scale (with 3 to 5 classes usually) has to account for the discrete nature of the measure, as well as its complex genetic background, combining direct effects of the calf with maternal effects of its dam. In France, a first feasibility study was conducted in dairy cattle at the end of the eighties in the PhD work of Manfredi (1990; see also Manfredi et al. (1991), and Foulley and Manfredi (1991)). Using the methodology developed by Gianola and Foulley (1982) and Harville and Mee (1984),

Manfredi applied a sire- maternal grandsire- dam within maternal grandsire threshold model (S-MGS-D-TM) to two large regional datasets for the Normande and Holstein breeds. He clearly showed that the S-MGS-D-TM was particularly convenient but stressed the importance of data quality. In particular, the necessity of an exhaustive recording system, with complete information about birth conditions of males and dead calves –two very informative sub-populations often overlooked – was emphasised in order to avoid potential evaluation biases. The regional data set that Manfredi analysed was of good quality, but it took several years before an exhaustive recording system could be set up on a national basis. This has been the case since September 1998 and a routine national evaluation is now possible. This paper describes how such an evaluation is being implemented, including all data since 1998 as well as previous data considered satisfactory. In this study, Manfredi's model was extended to incorporate heteroskedastic residual variances, a flexible extension of the threshold model proposed by Foulley and Gianola (1996).

## 2. Material and methods

### 2.1 Data

For computational reasons, the search for a proper statistical model was undertaken in the Normande breed only. Extra results from the other main French dual purpose breed, the Montbéliarde breed, are also presented here.

Calving conditions scores are expressed on a 1 to 5 subjective scale: 1: no assistance; 2: easy calving; 3: difficult calving; 4: caesarean; 5: embryotomy. Given the very low incidence of the last two categories, they were grouped with code 3 and therefore, only 3 categories were analysed.

Data collected in France since 1990 were obtained from the SIG (Système d'Information Génétique) database in Jouy-en-Josas. Data quality and exhaustivity greatly differed between administrative regions (called “départements”). All records from region  $x$  year combinations with fewer than 20% recorded male calves or with more than 95% records scored 1 were discarded. Other edits included dams in lactation 10 or above or with age at calving less than 22 months, regions with less than 1000 records and herds with less than 10 recorded calves. Eventually, the data file for model selection in the Normande breed included 777559 records. For genetic parameter estimation, further edits included sires with less than 10 recorded calves and maternal grand-sires with less than 3 recorded calves. The characteristics of the resulting data sets are reported in Table 1.

**Table 1. Characteristics of the data files analysed**

Breed	Normande (1)	Normande (2)	Montbéliarde (2)
records	777559	577342	409220
sires	4325	2877	4193
herd-year-season	110328	59898	41156
dams	409784	286247	193532
calves/dam	1.90	2.02	2.11
% code 1	79.3	79.1	58.7
% code 2	17.2	17.4	35.9
% code 3	3.5	3.5	5.4

(1) Data set used for model selection

(2) Data sets used for genetic parameter estimation

### 2.2 Methods

Table 1 clearly shows that the incidence of difficult calving is very low in both breeds. A classical analysis based on a linear description of the calving score  $Y_i$  is therefore not proper, in particular because it ignores the strong resulting relationship between the mean and the variance of the trait. The threshold model of S. Wright, made well-known among animal breeders by Foulley, Gianola, Harville and Mee is preferred. It assumes the existence of an underlying non observable, normally distributed variable  $y_i$  (for, say, animal  $i$ ) controlling the observed value  $Y_i$  through a set of  $J-1$  thresholds  $\tau_j$  (for  $J$  categories – here,  $J=3$ ). A linear model is defined to describe the location parameters of the distribution of the  $y_i$ 's, while a probit link is used to relate  $y_i$ 's and  $Y_i$ 's. In practice, let  $\Pi_{ij}$  be the probability to observe  $Y_i$  in category  $j$ . We have:

$$\begin{aligned}\Pi_{ij} &= \text{Prob}(\tau_{j-1} < y_i < \tau_j) \\ &= \Phi\left(\frac{\tau_j - \eta_i}{\sigma_i}\right) - \Phi\left(\frac{\tau_{j-1} - \eta_i}{\sigma_i}\right)\end{aligned}\quad [1]$$

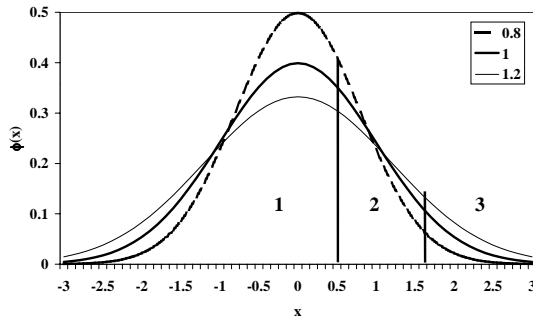
where  $\Phi$  is the cumulative distribution function of the standard normal distribution,  $\eta_i$  and  $\sigma_i^2$  are the mean and the (residual) variance of the underlying variable  $y_i$ . In presence of a set of explanatory variables, we write:

$$\eta_i = \mathbf{x}_i' \boldsymbol{\beta} \quad [2]$$

(or  $\eta_i = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{u}$  if random effects  $\mathbf{u}$  are included). In regular threshold models,  $\sigma_i^2$ , the variance of the underlying, non observable variable, is arbitrarily set to 1. Foulley and Gianola (1996) showed that a better fit of the data can often be obtained when flexibility is added to the threshold model by allowing  $\sigma_i^2$  to vary as a function of environmental covariates  $\mathbf{p}_i$ . This is achieved using a loglink function:

$$\ln(\sigma_i^2) = \mathbf{p}_i' \boldsymbol{\delta} \quad [3]$$

The added flexibility of heterogeneous residual variances on the threshold model is illustrated in figure 1 for 3 values of  $\sigma_i$  (0.8, 1.0 and 1.2) assuming a normal distribution with fixed location parameters. Given the (arbitrary) position of the thresholds on the figure, larger residual standard deviations lead to larger incidence of category 3 while maintaining (in this particular case) incidence of category 2 at a relatively constant level.



**Figure 1. Illustration of the effect of residual standard deviation heterogeneity on observed categories in a threshold model.**

About 40 fixed effect models defined by [1-3] were applied to the initial Normande dataset in order to select the environmental factors to be included in  $\mathbf{x}_i$  and  $\mathbf{p}_i$ . The “candidate” environmental factors considered were: sex of calf, parity of dam, age at calving of dam, region, year, month and their possible combinations. Likelihood ratio tests were performed but given the large data set used, they were almost systematically highly significant. In order to favour more parsimonious models, Akaike’s information criterion was used:

$$AIC = -2 LL + k p \quad [4]$$

where LL is the log-likelihood function, p the number of parameters estimated and  $k=2$ . The best model is supposed to be the one leading to the minimum AIC value. However, AIC still tends to overfit the model when sample size is very large. Atkinson (1981, quoted by McCullagh and Nelder (1983), p 70) suggested the use of values of k in [4] between 2 to 6, larger k’s corresponding to more conservative choices of variables. A value of  $k = 2\ln(\ln(N-r)) \approx 5.2$  was taken, where N is the sample size and r the rank of the coefficient matrix in the

analysis. Fagain, for this criterion, known as Hannan and Quinn’s criterion (HQC). small values are desirable.

A usual goodness-of-fit test was performed for each model based on Pearson’s  $\chi^2$  statistic, as suggested in Foulley and Gianola (1996):

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^3 \left( \text{obs}_{ij} - n_i \hat{\Pi}_{ij} \right)^2 / n_i \hat{\Pi}_{ij} \quad [5]$$

where the file is divided into I sub-populations;  $\text{obs}_{ij}$  is the observed number of observations in category j for sub-population i and  $n_i \hat{\Pi}_{ij}$  is the corresponding expected number, given the model of analysis. A small value of  $\chi^2$  indicates a better agreement between observed and expected values.

Based on this selection criteria, a final model was retained and was extended to a mixed model including the four random effects of Manfredi (1990), namely the effects of sire of calf, sire of dam (=maternal grandsire (mgs) of calf), dam within mgs and hear-year-season. For the heteroskedastic model, this is done by writing:

$$\eta_i = \mathbf{x}_i' \boldsymbol{\beta} + \sum_m \sigma_{mi} \mathbf{z}_{im}' \mathbf{u}_m^* \quad [6]$$

where  $\sigma_{mi}$  is the standard deviation of the  $m^{\text{th}}$  random effect and  $\mathbf{u}_m^*$  the  $m^{\text{th}}$  standardised random effect ( $\text{var}(\mathbf{u}_m^*)=1$ ). As in Foulley and Gianola (1996), an homogeneous intraclass correlation is assumed, i.e.,

$$\sigma_{mi} / \sigma_i = \text{constant} \quad \text{for all } m \quad [7]$$

Variance components were calculated using an approximate EM-REML algorithm. These variance components were then considered as known and EBVs as well as estimates of fixed effects were obtained as MAP estimates in a full scale evaluation involving 827200 records for the Normande breed, (resp. 701428 for the Montbéliarde), 3475 sires (resp., 6853), 446811 dams (resp., 367048). All calculations were performed with a flexible FORTRAN program using sparse matrix techniques.

### 3. Results

#### 3.1. Model selection

Table 2 presents the characteristics and goodness-of-fit tests of 11 models among >40 fitted), used to analyse the initial Normande data set. Models 1 and 2 assume a unit residual variance. Model 3 is the model finally retained. Models 4 and 5 simplify model 3 by replacing interactions by main effects. Model 6 increases the complexity of model 3 by defining an interaction of region x year with season. Models 7 to 11 illustrate the effect of not including each particular source of residual variation.

Accounting for heterogeneity of the residual variance drastically improved all criteria (LL, AIC, HQC or  $\chi^2$  statistics) However, in contrast with the example of Foulley and Gianola (1996) with calving difficulty scores in Simmental, it did not prevent the need for the interaction terms between sex and parity, month and year, region and year. Extending the latter interaction term to season improved some criteria but at a high cost in number of parameters, which is heavily penalised by the HQ criterion. Among the sources of variance heterogeneity, the influences of region and, to a lesser extent, sex were considerable. The improvements related to the inclusion of month or parity heterogeneity did not appear very large (although significant), but it will be seen that this seems to be rather specific to the Normande breed (see section 3.3 and figures 5 and 7).

The final model (model 3) included the interactions of sex of calf by parity-age class (5 age classes in parity 1, 2 in parity 2 and 1 for each subsequent parity), region by year and year

by month. Residual variances were influenced by sex of calf, region, year of calving, month of calving and parity-age class.

#### 3.2. Variance parameters

Estimates of variance components based on the extended version of model 3 indicated in [6] are reported on table 3. Estimates were similar for both breeds but were substantially lower than those obtained by Manfredi (1990). Direct heritabilities were 5.4% in both breeds, while maternal heritabilities were very low (around 3%). This disappointing result still remains to be explained. Two potential explanations are: 1) a difference in the estimation procedure: Manfredi used an approximation of the “tilde-hat approach” of Van Randen and Jung (1988) while a probably more correct EM-REML algorithm was used here. 2) a difference in data quality: a much larger heterogeneity, especially before 1998, in the exhaustivity of data collection is an inherent characteristic of the datasets analysed here.

Based on these estimates, it is possible to predict approximate reliabilities of sire EBVs using a formula proposed by Foulley and Im (1989). Including pedigree information (but not the indirect information due to the correlation between direct and maternal effects), a reliability of 0.50 can be obtained with calving scores on 56 (resp. 52) progeny for sire EBVs on birth conditions (direct effects) in the Normande (resp. Montbéliarde) breed. The corresponding figure for calving conditions (maternal effects) is 72 (resp. 80).

**Table 2. Model selection and goodness-of-fit tests**

Model	Fixed effect part <sup>(a)</sup>	Log(Residual variance) part <sup>(a)</sup>	p <sup>(b)</sup>	-2LL <sup>(c)</sup>	AIC <sup>(c)</sup>	HQC <sup>(c)</sup>	$\chi^2$ (ry) <sup>(d)</sup>	$\chi^2$ (sp) <sup>(d)</sup>	$\chi^2$ (ry*sp) <sup>(d)</sup>
1	sp +ym+ry		<b>270</b>	11712	12252	13118	3901	771	9078
2	sp +ym+rys		395	11197	11987	13254	3885	763	8987
<b>3<sup>(e)</sup></b>	<b>sp +ry+ym</b>	<b>s+r+y+m+p</b>	<b>332</b>	<b>8048</b>	<b>8712</b>	<b>9777</b>	<b>577</b>	<b>153</b>	<b>5185</b>
4	sp +r +ym	s+r+y+m+p	<b>224</b>	9018	9466	10185	1334	<b>128</b>	5831
5	s+p +ry+ym	s+r+y+m+p	319	8202	8840	9863	525	268	5178
6	sp +rys+ym	s+r+y+m+p	457	<b>7491</b>	<b>8405</b>	9871	531	134	<b>5040</b>
7	sp +ry +ym	r+y+m+p	331	8296	8958	10020	546	353	5271
8	sp +ry +ym	s+ y+m+p	304	10782	11390	12365	3059	<b>102</b>	8149
9	sp +ry +ym	s+r+ m+p	323	8255	8901	9937	731	153	5234
10	sp +ry +ym	s+r+y+ p	321	8112	8754	<b>9784</b>	<b>514</b>	136	<b>5052</b>
11	sp +ry +ym	s+r+y+m	319	8132	8770	9793	<b>523</b>	163	5127

**boldface:** best two models in the column

(a) s = sex; p = parity-age-class; r = region; m = month; y = year

sp = sex-parity-age class; ym = year-month; ry(s)= region-year-(season)

(b) number of fitted parameters

(c) LL = log-likelihood (-constant); AIC: Akaike's information criterion; HQC: Hannan & Quinn's criterion (see text)

(d) Pearson's  $\chi^2$  statistics when data are grouped by region-year (rp – 130 classes ), sex-parity-age (sp – 28 classes) or rp\*sp (1783 classes).

(e) final model

**Table 3. Estimates of variance components**

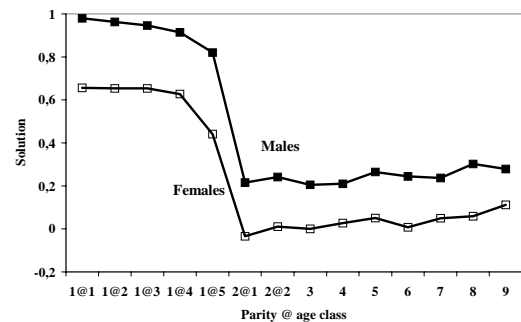
Breed	Montbé liarde	Normande
$\sigma^2$ (sire)	0.021	0.020
$\sigma^2$ (mgs)	0.013	0.016
$\sigma^2$ (dam within mgs)	0.059	0.036
$\sigma^2$ (hys)	0.497	0.387
Cov(sire-mgs)	0.008	0.009
Total $\sigma^2$	1.590	1.458
$h^2$ direct (%)	5.4	5.4
$h^2$ maternal(%)	2.6	3.1
$\rho$ sire-mgs	0.49	0.53
$\rho$ direct-maternal	-0.16	-0.04

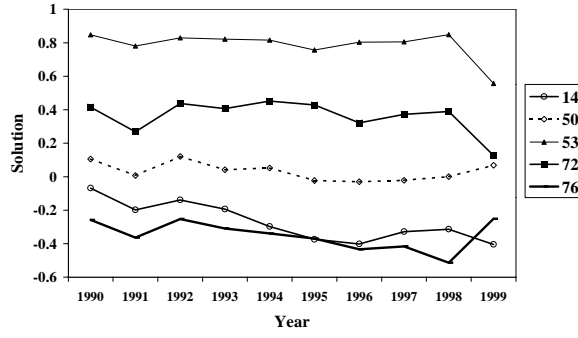
### 3.3 Solutions for fixed effects and residual variance

Figure 2 presents solution estimates for sex and parity-age class in the full scale genetic evaluation. As in Manfredi (1990) and others, it was found that solutions for male calves and first parity cows are much higher than for females and later parities. However, differences between sexes were larger in parity

1 than for later parities, with a decreasing trend with age class, at least for male calves.

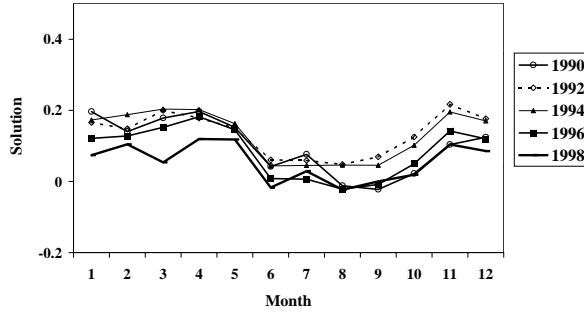
Figure 3 displays the region x year solutions for 5 large regions in the Normande breed. These solutions were surprisingly variable across regions. Differences between extreme regions were larger than between first and later parities! Differences in management, data quality, exact definition of each subjective score may at least partly explain this variability. On the other end, solutions were consistent from year to year, even though trends existed (e.g., in region 14).

**Figure 2. Solutions for sex x parity-age class effects in the Normande breed.**



**Figure 3. Solutions for region x year effects for 5 regions in the Normande breed.**

Figure 4 presents solutions of year x month effects, again for the Normande breed, only for even years for clarity. Typically, summer months corresponded to lower solutions, i.e., more calvings without assistance. The interaction with the year effect was apparent in 1998, with fewer calving problems “than usual” in winter and spring.



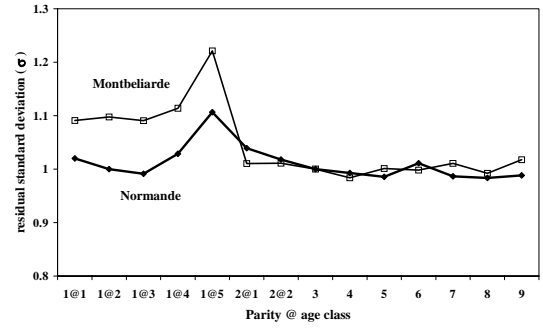
**Figure 4. Solutions for year x month effects in the Normande breed.**

From [3], the residual standard deviation for animal  $i$  is:

$$\hat{\sigma}_i = \exp(\mathbf{p}_i' \hat{\delta}) \quad [8]$$

The effect of a particular factor on the residual standard deviation was depicted by plotting  $\hat{\sigma}_i$  for different levels of the factor. The reference ( $\hat{\sigma}_i = 1$ ) in figures 5 to 7 was the residual variability observed in calving codes of female calves born in September 1998 in region “50” from third parity cows.

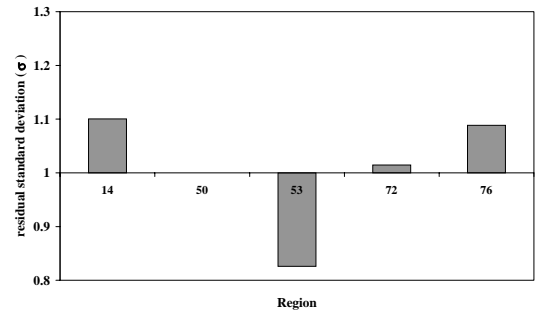
Residual variability estimates were substantially higher in males ( $\hat{\sigma}_i = 1.07$  and 1.18 in the Normande and Montbéliarde breeds, respectively) than in females ( $=1$ ). In both breeds, the residual standard deviations tended to increase for later calvers in first parity (figure 5). For young first calvers, they were similar to later parities in the Normande breed but were at least 10% larger in the Montbéliarde breed.



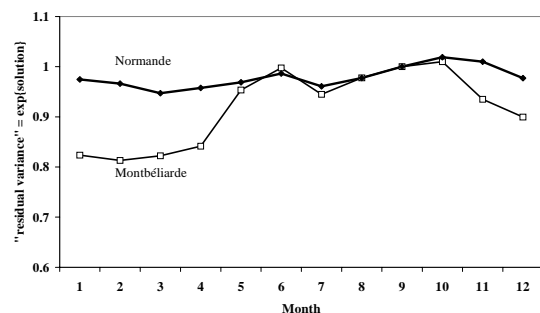
**Figure 5. Residual deviation as a function of parity x age class.**

Residual variability was low in regions with large region x year fixed effect solutions (e.g., region “53” in figures 3 and 6) and high in regions with smaller region x year effects (e.g., regions “14” and “76”).

Finally, residual variability was about 15% lower in winter than in summer in the Montbéliarde breed. As for the effects of sex and parity-age class, this month effect on residual standard deviation was much less pronounced in the Normande breed.



**Figure 6. Residual standard deviation as a function of the region, for 5 regions in the Normande breed.**



**Figure 7. Residual standard deviation as a function of month of calving.**

### 3.4 Genetic evaluation

Assuming the variance components in table [3] as exact, a full scale genetic evaluation was implemented with the same model. With a

very strict convergence criterion, computing time was of about 1 hour and a half for both breeds on an IBM Risc 595 AIX 4.21 computer. Solutions on the underlying scale for sires of calf and sires of dam were first deviated from a rolling base defined as for other traits in France (i.e., for June 2000, the base is the mean EBV of all sires born between 1990 and 1993). To facilitate interpretation, these solutions were then expressed in genetic standard deviation and in expected proportion of easy births (resp., easy calvings) (codes 1 and 2) from (resp. among) first parity cows in winter 1998-1999 assuming a sex ratio of 0.5. Table 4 illustrates the range of EBVs obtained in the Normande breed.

**Table 4. Examples of sire EBVs obtained in the genetic evaluation of the Normande breed**

trait	bull	progeny (c)	EBV (d)	expected easy calvings among first parity cows	observed easy calvings (e)	proportion of progeny from first parity cows (f)
birth conditions (a)	A	6404	-4.7	75.6%	92.0%	2.5%
	B	4015	-2.8	83.0%	92.0%	24.2%
	C	23857	+1.8	94.4%	97.7%	28.4%
calving conditions (b)	D	865	-3.3	82.4%	81.7%	95.7%
	E	17088	-2.1	86.0%	92.5%	30.5%
	F	6842	+2.0	94.3%	96.2%	67.1%

(a) EBV as sire of the calf

(b) EBV as sire of the dam

(c) number of calves for birth conditions or number of daughters for calving conditions

(d) in genetic standard deviation

(e) in the whole population, i.e., including later parities

(f) observed

Bull A was known as an extreme sire for birth conditions, and therefore was almost never used on heifers (2.5%). He got the worst EBV for birth conditions, even though the observed frequency of easy calvings among his calves was similar to the one of bull B. Excluding this sire, the range of expected easy births and calvings with first parity cows was 81 to 94% in the Normande breed and 78 to 91% in the Montbéliarde breed.

### 4. Conclusion

The proposed model will be used for routine evaluations for both birth conditions and calving conditions in France with a first official release in June 2000. It was clearly found that the heteroskedastic threshold model is computationally feasible for national genetic evaluations. Modelling residual variability increased the model flexibility,

improved goodness-of-fit and probably (at least partly) accounted for differences in data quality. Heritability estimates were low. They will have to be estimated again as better quality data accumulate. However, they may not be completely surprising given the low incidence of difficult calvings in both breeds. The more heterogeneous variability observed in the Montbéliarde breed is also an interesting result that deserves attention.

## Acknowledgments

The work of Marie Mathevon (Institut de l'Élevage) and Michèle Briend (INRA) for the preparation of the datasets is gratefully acknowledged.

## References

- Foulley, J.L., Gianola, D. & Thompson, R. 1983. Prediction of genetic merit from data on categorical and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genet. Sel. Evol.* 15, 401-424.
- Foulley, J.L. & Gianola, D. 1996. Statistical analysis of ordered categorical data via a structural heteroskedastic threshold model. *Genet. Sel. Evol.* 28, 249-273.
- Foulley, J.L. & Im, S. 1989. Probability statements about the transmitting ability of progeny-tested sires for an all-or-none trait with an application to twinning in cattle. *Genet. Sel. Evol.* 21, 359-376.
- Foulley, J.L. & Manfredi, E. 1991. Approches statistiques de l'évaluation génétique des reproducteurs pour des caractères binaires à seuils. *Genet. Sel. Evol.* 23, 309-338.
- Gianola, D. & Foulley, J.L. 1983. Sire evaluation of ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15, 201-224.
- Harville, D.A. & Mee, R.W. 1984. A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408.
- Manfredi, E. 1990. Analyse génétique des conditions de naissance chez les bovins par le modèle à seuils. *Thèse de Doctorat*, Université Paris-Sud, Orsay, France.
- Manfredi, E., Ducrocq, V. & Foulley, J.L. 1991. Genetic analysis of dystocia in cattle. *J. Dairy Sci.* 74, 1715-1723.
- McCullagh, P. & Nelder, J.A. 1983. *Generalized linear models*. Chapman and Hall, London, UK.
- VanRaden, P.M. & Jung, Y.C. 1988. A general purpose approximation to restricted maximum likelihood: the tilde-hat approach. *J. Dairy Sci.* 71, 187-195.