Character process model for semen volume in AI rams: evaluation of correlation structures for long and short-term environmental effects

Ingrid David, Loys Bodin, Gilles Lagriffoul, Non Renseigné, Christèle Robert-Granié

Genet. Sel. Evol. 39 (2007) 55–71

DOI: 10.1051/gse:2006033

Original article

Character process model for semen volume in AI rams: evaluation of correlation structures for long and short-term environmental effects

Ingrid DAVID[a]*, Loys BODIN[a], Gilles LAGRIFFOUL[b], Eduardo MANFREDI[a], Christèle ROBERT-GRANIÉ[a]

[a] Station d'amélioration génétique des animaux, Institut national de la recherche agronomique, BP 52627, 31326 Castanet-Tolosan Cedex, France
[b] Institut de l'élevage – ANIO, BP 42118, 31321 Castanet-Tolosan Cedex, France

(Received 24 May 2006; accepted 11 October 2006)

Abstract – The objective of this study was to build a character process model taking into account serial correlations for the analysis of repeated measurements of semen volume in AI rams. For each ram, measurements were repeated within and across years. Therefore, we considered a model including three environmental effects: the long-term environmental effect, which is a random year*subject effect, the short-term environmental effect, which is a random within year subject*collection effect, and the classical measurement error. We used a four-step approach to build the model. The first step explored graphically the serial correlations. The second step compared four models with different correlation structures for the short-term environmental effect. We selected fixed effects in the third step. In the fourth step, we compared four correlation structures for the long-term environmental effect. The model, which fitted best the data, used a spatial power correlation structure for the short-term environmental effect and a first order autoregressive process for the long-term environmental effect. The heritability estimate was 0.27 (0.04), the within year repeatability decreased from 0.56 to 0.44 and the repeatability across years decreased from 0.43 to 0.37.

serial correlations / longitudinal data / semen / sheep

1. INTRODUCTION

Artificial insemination (AI) is an important tool applied in programs of genetic improvement of dairy and meat sheep. In France, AI is used essentially with fresh semen, during summer, a period when semen production units need to produce large amounts of useful semen per day from a limited number

* Corresponding author: Ingrid.David@toulouse.inra.fr

of rams. For a given conservation technique, the number of doses produced per ejaculate depends on the volume, sperm concentration and sperm motility. Each ram in an AI centre is collected repeatedly and frequently during the summer season over several years. Semen records of an adult ram may be considered as repeated measures of a single trait. The classical approach for repeated measures of a single trait is the simple repeatability model. This is widely used for the genetic analysis of semen characteristics in domestic species [1, 10, 20, 23, 24, 33]. However, the simple repeatability model assumes that residual effects are independent and that permanent environmental effects are constant over the productive life of each subject, which may be too restrictive. This may result in biased estimates of variance components [31] and may invalidate inferences about the mean response profile in longitudinal data [9]. Several general models that consider a smooth change over time in the mean as well as in the variance structures are given in the literature: the random regression model, which attempts to model the forms of the functions of time for each component of the phenotype (mean, genetic additive value, permanent environment) [14], the structured antedependence model, which defines the observation at time t as a function of the previous observation [17] and the character process model, which focuses on modelling the covariance functions themselves [27]. The aim of this study was to find the character process model that fits best the semen volume data recorded on adult rams of one French AI centre.

2. MATERIALS AND METHODS

2.1. Data

Data on the ejaculates used in this study originate from rams collected during the 1996–2004 period in a French AI centre. We have focussed our study on data recorded between May and August, which corresponds to the period of intensive semen collection. A total of 36 480 ejaculates from 974 adult males, sons of 230 different sires, was analysed. The final pedigree file consisted of 20 761 animals. The rams were between 2 and 7 years old. The intervals between successive collections within one year varied from 1 to 28 days. The rams were collected over 1 to 5 successive years. The number of records per animal was on average 36 and ranged from 1 to 183. Three traits (volume, sperm concentration and mass motility) were measured at each collection. The character of interest in the present study was the volume, which was read directly in millilitres from a graduated collection tube.

2.2. Model

The assumptions of the simple repeatability model are not compatible with the analysis of this semen data. Firstly, the assumption of independence between residual effects could not be fulfilled. Indeed, uncontrolled daily environmental (temperature) or biological (animal health) influences on semen production on a particular day may be similar to those on neighbouring days. Secondly, the random animal effect may not be constant over years. For instance, animals are assigned to different pens each year and thus a ram may be dominant one year and not be dominant another year, which can affect its semen production. Thus, it is more realistic to update yearly the permanent environmental effect. As a result, three random environmental components influencing each semen record may be considered. Following notations used by Carvalheira *et al.* [4], the first is the long-term environmental effect (LTE), which includes all uncontrolled events that permanently influence the semen production of a particular ram during one year. This LTE effect includes past events specific to each animal that occurred during its first months of life (health, rearing conditions) and within year conditions such as feeding management, relations with other animals in the pen and health events. The second component is the short-term environmental effect (STE), which comprises all other sources of unaccounted temporary variations (correlated) that affect semen production. The third is the classical measurement error.

To fit the data with these new assumptions, the following character process model was used:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{Wp} + \boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 \tag{1.1}$$

where: $\boldsymbol{y}$ is a vector of ejaculate volumes of order K ($K = \sum_{i=1}^{N} n_i$, N being the number of animals, n_i the number of measurements per animal). $\boldsymbol{\beta}$ is a vector of fixed environmental effects of order q. $\boldsymbol{a}$ is a vector of random genetic additive effect of order s ($s = 20\,761$). $\boldsymbol{p}$ is a vector of random LTE effects of order r ($r = \sum_{i=1}^{N} c_i$, where c_i is the number of years of collections for subject i). $\boldsymbol{\varepsilon}_1$ is a vector of random correlated STE effects of order K. $\boldsymbol{\varepsilon}_2$ is a vector of independent residuals of order K. $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{W}$ are incidence matrices of order $K \times q$, $K \times s$, $K \times r$, respectively.

The distribution of random effects are the following: $\boldsymbol{a} \sim N(0, \sigma_a^2 \boldsymbol{A})$ where $\boldsymbol{A}$ is the numerator relationship matrix, $\boldsymbol{p} \sim N(0, \sigma_p^2 \boldsymbol{H}_2)$, $\boldsymbol{\varepsilon}_1 \sim$

$N(0, \tau^2 \boldsymbol{H}_1)$ and $\boldsymbol{\varepsilon}_2 \sim N(0, \sigma_e^2 \boldsymbol{I}_K)$. $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are

$$\boldsymbol{H}_1 = \bigoplus_{i=1}^{N} \bigoplus_{c=1}^{c_i} \boldsymbol{H}_{1ic} \text{ with } \boldsymbol{H}_{1ic} = \begin{bmatrix} 1 & h_{ic,12} & \cdots & h_{ic,1L_{ci}} \\ h_{ic,12} & 1 & \cdots & h_{ic,2L_{ci}} \\ \vdots & \vdots & \cdots & \vdots \\ h_{ic,1L_{ci}} & h_{ic,2L_{ci}} & \cdots & 1 \end{bmatrix} \text{ and } h_{ic,jl} = g_1\left(x_{ic,jl}\right)$$

$$\boldsymbol{H}_2 = \bigoplus_{i=1}^{N} \boldsymbol{H}_{2i} \quad \text{with} \quad \boldsymbol{H}_{2i} = \begin{bmatrix} 1 & s_{i12} & \cdots & s_{i1c_i} \\ s_{i12} & 1 & \cdots & s_{i2c_i} \\ \vdots & \vdots & \cdots & \vdots \\ s_{i1c_i} & s_{i2c_i} & \cdots & 1 \end{bmatrix} \quad \text{and} \quad s_{i,mc} = g_2\left(d_{i,mc}\right)$$

where g_1 and g_2 are unknown correlation functions, $\oplus$ represents the direct sum. L_{ci} is the number of observations made during year c for subject i. $x_{ic,jl}$ is a measurement of "distance" between collections j and l of subject i during year c. d_{imc} is a measurement of "distance" between collections performed during year m and c for subject i.

The vector of observations $\boldsymbol{y}$ is assumed to be normally distributed with $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $var(\boldsymbol{y}) = \sigma_a^2 \boldsymbol{ZAZ}' + \sigma_p^2 \boldsymbol{WH}_2 \boldsymbol{W}' + \tau^2 \boldsymbol{H}_1 + \sigma_e^2 \boldsymbol{I}_K$.

Under this model, a heritability estimate is $\hat{\sigma}_a^2 / \hat{\sigma}_T^2$ where $\hat{\sigma}_T^2 = \hat{\sigma}_a^2 + \hat{\sigma}_p^2 + \hat{\tau}^2 + \hat{\sigma}_e^2$. The within year repeatability estimate is, for all subjects i, a function of x of the forms: $\left(\hat{\sigma}_a^2 + \hat{\sigma}_p^2 + \hat{\tau}^2 g_1(x)\right) / \hat{\sigma}_T^2$ and the repeatability estimate across years is, for all subjects i, a function of d of the forms $\left(\hat{\sigma}_a^2 + \hat{\sigma}_p^2 g_2(d)\right) / \hat{\sigma}_T^2$.

2.3. Model selection

Selection of appropriate correlation structures for LTE and STE is a non-trivial step in the model selection process. The method used is inspired from methods proposed by Verbeke and Molenberghs [34], Verbeke *et al.* [35] and Lesaffre *et al.* [21] and involves four steps.

The first step consisted in exploring graphically the correlation between collections using the empirical semi-variogram [9, 34]. To detect this correlation, the following model was used:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Mf} + \boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 \tag{1.2}$$

where $\boldsymbol{y}$, $\boldsymbol{\beta}$, $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are the same as in model (1.1), $\boldsymbol{f}$ is the random subject intercept ($\boldsymbol{f} \sim N(0, I_N \sigma_f^2)$) with the corresponding incidence matrix $\boldsymbol{M}$,

we draw the scatter plot of $\frac{1}{2}E\left(e_{ij} - e_{il}\right)^2 = \sigma_e^2 + \tau^2\left(1 - g_1\left(x_{ijl}\right)\right)$ *vs.* the corresponding "distance" x_{ijl} where e is the vector of independent residuals of model (1.3):

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Mf} + \boldsymbol{e} \tag{1.3}$$

where the notations are the same as in model (1.2). The total process variance $(\sigma_e^2 + \tau^2 + \sigma_f^2)$ was estimated by $\frac{1}{2}E\left(e_{ij} - e_{i'k}\right)^2 \forall i \neq i'$. For the exploration of the function g_1, the distance considered was the interval in days between observations. For the exploration of g_2, we used residuals obtained with a model similar to (1.3) where $\boldsymbol{y}$ corresponded to the subject average trait per year and the distance considered was the interval in years between mean records. In this step, $\boldsymbol{\beta}$ is the vector of all the potential fixed effects and all the two-way interactions are described in detail in David *et al.* [7].

The second step consisted in choosing g_1 fitting and comparing several serial correlation structures for the model (1.2). We considered four different structures for $\boldsymbol{H}_1$. The first model ($C0_S$) assumed independent STE ($g_1(x) = 0$), the second model (AR_S) assumed a first-order autoregressive correlation structure for $\boldsymbol{H}_1$ ($g_1(x) = \rho^x$ where x is the difference between ranks of collection), the third (SP_S) and fourth (SG_S) models assumed a spatial power ($g_1(x) = \rho^x$) and a spatial Gaussian ($g_1(x) = \rho^{x^2}$) where x is the interval between collections in days, respectively. The models were fitted using the mixed procedure of SAS® 8.2 [30] and the restricted maximum likelihood method to estimate variance components. We used restricted likelihood ratio tests to compare the $C0_S$ model to the three other models (nested models). The corresponding null and alternative hypotheses were $\rho = 0$ and $\rho \neq 0$, respectively. Under the null hypothesis, the test statistic is expected to be distributed as a chi-square with two degrees of freedom. We used the Akaike information criteria (AIC) to compare the non-nested models (AR_S, SP_S, SG_S). The best model will be the one that has the lowest value of −2AIC.

The third step selected fixed model terms in (1.2) based on the model retained in step 2. This has been done step by step by comparing nested models with the likelihood ratio test. For this step, models were fitted using the mixed procedure of SAS® 8.2 [30] and the maximum likelihood method.

The fourth step selected an appropriate correlation structure for the LTE according to the correlation structure for $\boldsymbol{H}_1$ selected in step 2 and the mean structure retained in step 3. We compared four different correlation structures for $\boldsymbol{H}_2$ using the model (1.1). The first model assumed no correlation between LTE ($C0_L$, $g_2(d) = 0$), the second model assumed a correlation of 1 between LTE ($C1_L$, $g_2(d) = 1$), the third model assumed a first-order autoregressive correlation structure for $\boldsymbol{H}_2$ (AR_L, $g_2(d) = \eta^d$, where d is the absolute

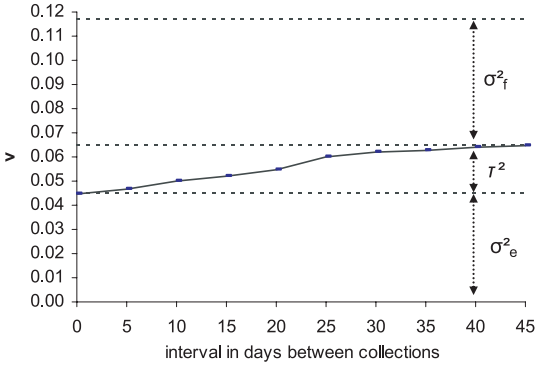

Figure 1. Ram semen volume: observed semi-variogram for the short-term environmental effect. σ_e^2, τ^2 and σ_f^2 represent the variability of the measurement error, the serial correlation component and the random intercept in (1.2), respectively.

difference between ranks of year of collection) and the fourth, a uniform correlation structure (UC_L, $g_2(d)$ = constant). Restricted likelihood ratio tests were used to compare $C0_L$ or $C1_L$ to AR_L or UC_L structures while AIC criteria were used to compare the other models. For this step, data were analysed with an animal model using the ASReml software [13]. Variance components and genetic parameters were estimated at the end of this step using the restricted maximum likelihood method.

Finally, to clarify the relative role of the various model components, we compared the four models obtained by fitting the AR_L or $C1_L$ structure to the LTE variation and the SP_S or $C0_S$ structures for the STE variation. Thereafter in the text, the various sub-models (1.1) are labelled according to the structure of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, for instance AR_L/SP_S.

3. RESULTS

3.1. Semi-variogram

The empirical semi-variograms drawn for the exploration of g_1 and g_2 are presented in Figures 1 and 2, respectively. In both cases, this suggests a serial correlation dependent on time that may be described by an increasing function of the interval between observations. Nevertheless, in both cases, the semi-variogram did not give enough information to choose among the potential $g(x)$ functions. In Figure 1, the empirical estimates of variance components for σ_f^2, τ^2 and σ_e^2 were 0.052, 0.02 and 0.045, respectively.

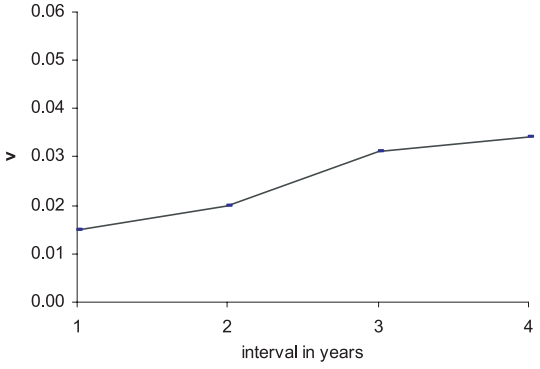

Figure 2. Ram semen volume: observed semi-variogram for the long-term environmental effect.

Table I. Variance component estimates, minus twice restricted log-likelihood and minus twice Akaïke information criteria for different correlation structures for short term environmental effect in model (1.2).

STE[a]	$C0_S$	AR_S	SP_S	SG_S
$-2RL$[1]	4905	2767	2560	2644
$\Delta(-2RL)$[2]	-	2138	2345	2261
$-2AIC$[3]	4909	2775	2568	2652
σ_e^2	0.059	0.049	0.048	0.050
τ^2	-	0.018	0.019	0.013
ρ	-	0.93	0.97	0.96
σ_f^2	0.050	0.039	0.040	0.043

[1] $-2RL$ = minus twice restricted maximum likelihood.

[2] $\Delta(-2RL)$ = likelihood ratio statistics of the corresponding model *vs.* the $C0_S$ model.

[3] $-2AIC$ = minus twice Akaïke information criteria.

[a] Correlation process for the STE: $C0_S \Leftrightarrow g_1(x) = 0 \mid SP_S \Leftrightarrow g_1(x) = \rho^x \mid SG_S \Leftrightarrow g_1(x) = \rho^{x^2}$, where x is the absolute interval in days between observations, $AR_S \Leftrightarrow g_1(x) = \rho^x$, where x is the absolute interval between ranks of observation.

3.2. Selection of short-term environmental effect correlation matrix

Variance component estimates, minus twice-restricted log-likelihood and AIC criteria ($-2AIC$), obtained with the different correlation structures to model STE are presented in Table I. When we compared the $C0_S$ structure to AR_S, SP_S and SG_S structures, the smallest likelihood ratio statistic of two models was 2138, indicating that the $C0_S$ structure fitted the data less well than the other. Concerning the comparison among SP_S, SG_S and AR_S structures, the $-2AIC$ criteria were smaller for the SP_S and all differences in the $-2AIC$ criteria were higher than 80, indicating that the SP_S structure was the

best. The estimates of variance components with the SP_S structure were close to those estimated with the empirical semi-variogram 0.048, 0.019, 0.040 *vs.* 0.052, 0.02 and 0.045 respectively. The correlation estimate was very high between two STE separated by one day ($\rho = 0.97$).

3.3. Fixed effects

The fixed effects included in the final model for the mean structure were the age at collection, the week, the year of collection, the interval with previous collection, the daily variation (AM/PM) and the interaction week*year. The number of collections during the previous year, the rank of the within season collection, the breeding value for milk production in quartile and all other two-way interactions were discarded. Minus twice the difference of maximum log-likelihood values between the full fixed effect model and the reduced models equals 210. The associated number of degrees of freedom is 180, implying that the reduced model was acceptable at the 5% level of significance ($P = 0.0623$). Estimates of the selected fixed effects are described in detail in David *et al.* [7].

3.4. Selection of long-term environmental effect correlation matrix

Variance components, heritability estimates, minus twice AIC criteria and restricted log-likelihood obtained with the different correlation structures for the LTE are presented in the first four columns of Table II. The smallest value of the restricted likelihood ratio tests corresponding to the comparison of AR_L/SP_S model to the $C0_L/SP_S$ or the $C1_L/SP_S$ model was 55, and the AR_L/SP_S model presenting a lower −2AIC criteria than the UC_L/SP_S model. These results indicate that the AR_L/SP_S model has the best fit. The correlation estimate for LTE was $\eta = 0.88$ and was significantly different from 1 (by the LRT test with $C1_L/SP_S$).

Repeatability within and across years, estimated with the AR_L/SP_S model, were decreasing functions of the interval between collections. Within year repeatability varied from 0.56 to 0.44 (Fig. 3) and repeatability across years varied from 0.43 to 0.37 (Fig. 4).

The four last columns in Table II contains the results for the four models obtained by fitting the AR_L or $C1_L$ structure to the LTE variation and the SP_S or $C0_S$ structures for the STE variation. The log-likelihood ratio values indicate that models with the AR_L structure were significantly better than the models with the $C1_L$ structure, and that including the STE component (SP_S *vs.* $C0_S$) was also significantly better. The simple repeatability model ($C1_L/C0_S$) just

Table II. Variance component, heritability estimates and minus twice restricted log-likelihood for different correlation structures for long and short term environmental effects in model (1.1).

	1	2	3	4[d]	5[c]	6
LTE[a]	$C0_L$	UC_L	$C1_L$	AR_L	$C1_L$	AR_L
STE[b]	SP_S	SP_S	SP_S	SP_S	$C0_S$	$C0_S$
−2RL[1]	−64 882	−64 897	−64 896	−64 951	−62 357	−64 337
−2AIC[2]	−64 872	−64 885	−64 886	−64 938	−62 351	−64 329
σ_e^2	0.0481	0.0481	0.0481	0.0479	0.0596	0.0540
τ^2	0.0162	0.0161	0.0176	0.0131	-	-
ρ	0.97	0.97	0.97	0.96	-	-
σ_p^2	0.0018	0.0143	0.0126	0.0182	0.0190	0.0210
η	-	0.88	-	0.88	-	0.66
σ_a^2	0.0508	0.0319	0.0321	0.0306	0.0327	0.0367
h^2	0.43 (0.04)	0.29 (0.04)	0.29 (0.04)	0.27 (0.04)	0.29 (0.04)	0.33 (0.04)

[1] −2RL = minus twice restricted log likelihood.

[2] −2AIC = minus twice Akaïke information criteria.

[a] Correlation process for the LTE:

$C0_L \Leftrightarrow g_2(d) = 0 \mid C1_L \Leftrightarrow g_2(d) = 1 \mid UC_L \Leftrightarrow g_2(d) = \eta \mid AR_L \Leftrightarrow g_2(d) = \eta^d$, where d is the absolute difference between ranks of year of observation.

[b] Correlation process for the STE: $SP_S \Leftrightarrow g_1(x) = \rho^x \mid C0_S \Leftrightarrow g_1(x) = 0$, where x is the absolute interval between collections in days.

[c] Simple repeatability model.

[d] Retained model.

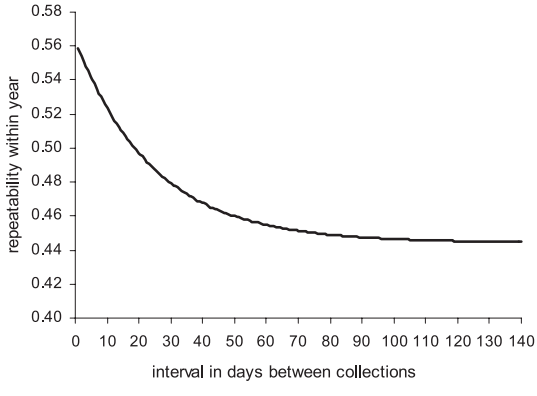

Figure 3. Within year repeatability for ram semen volume in a model with spatial power and autoregressive correlation structures for the short and long-term environmental effect, respectively.

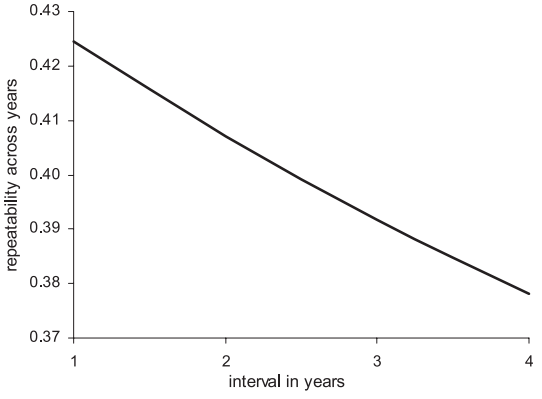

Figure 4. Repeatability across years for ram semen volume in a model with spatial power and autoregressive correlation structures for the short and long-term environmental effect, respectively.

fitting the subject effects inflated the proportion of variance in the residual relative to the other models. A major improvement in fit was observed when either the AR_L correlation to LTE or a correlation structure to the STE was added to the simple repeatability model. Both of these models effectively allow a lower correlation between records from different years than for records within the same year. Having both terms in (AR_L/SP_S) gives a small, though significant, further gain, and results in a lower heritability estimate than the other models. The correlation structures between observations (overall repeatability) modelled with these four models are presented in Figure 5. The simple repeatability model ($C1_L/C0_S$) assumes a constant correlation between observations, which is quite different from the correlations modelled with the three other models. The $AR_L/C0_S$ model, which considers that the correlation between observations made during the same year is constant, tended to underestimate the correlation between observations separated by less than 15 days, and to overestimate the correlation between observations separated by more than 15 days and less than 2 years. The correlations modelled with the AR_L/SP_S and $C1_L/SP_S$ models were quite similar but may differ significantly when the number of years between observations increases.

4. DISCUSSION

4.1. Model

The model proposed here, for the analysis of repeated measurements of semen volume, is a special case of the parametric character process models that

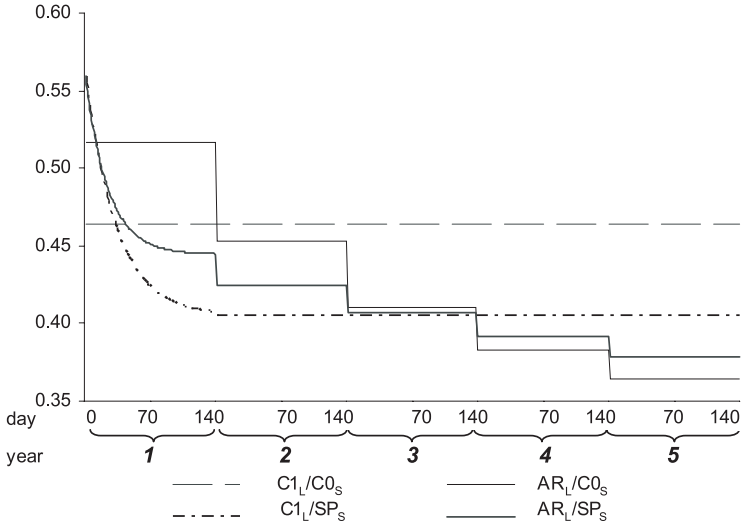

Figure 5. Evolution of the correlation between observations depending on the model $C1_L \Leftrightarrow g_2(d) = 1 \mid AR_L \Leftrightarrow g_2(d) = \eta^d \mid C0_S \Leftrightarrow g_1(x) = 0 \mid SP_S \Leftrightarrow g_1(x) = \rho^x$, where g_1 and g_2 are the functions of the correlation processes for the STE and the LTE, respectively, d is the absolute difference between ranks of year of observation and x is the absolute interval between collections in days.

model the covariance functions [27]. Such models also accommodate a correlation other than 1, over time, for the genetic component. However, we did not use this approach since we checked that the genetic correlations between traits, in a multiple trait analysis with one trait per age, were all high (>0.95). Similar models to the one proposed here have been used to study cow lactations [4,31] but the modelling process differed from ours by omitting the formal justification of the choice of the correlation structures. Other models can be used for the analysis of repeated and correlated data. For example, random regression or structured antedependence models have been proposed for the evaluation of cattle [17–19,25,26]. Nevertheless, Jaffrézic and Pletcher [15] have shown that character process models perform well in comparison to alternative methods, often providing a better fit to the covariance structure.

A first attempt to extend the character process model to the multiple-trait case has been proposed by David *et al.* [6] considering cross-correlations equal to 0, identical ρ and identical η for all traits. Other models that do not ignore cross-correlations have been proposed by Jaffrézic *et al.* [16] using a method that involves an eigen transformation of the variance covariance matrices while Gilmour and Thompson [12] proposed to model the covariance matrix on the

inverse scale. However, in both cases these methods assume that the correlation process has the same form for all traits. These assumptions may be too restrictive and cannot be extended in the case of different correlation processes for each single trait. Exploring graphically the cross-correlation could be an interesting first step to this extension. If we consider two traits A and B, the method for exploring the cross-correlation process for STE would be to draw $\frac{1}{2}E\left(e_{ij}^{A} - e_{ik}^{B}\right)^2$ *vs.* the corresponding distance$_{jk}$, where e_{ij}^{A} and e_{ik}^{B} are the residuals of model (1.3) extended to the multiple trait case where only $\text{cov}(f^A, f^B) \neq 0$. However, further analyses are needed to validate and confirm this approach.

4.2. Model selection

We used a graphical technique to detect serial correlation. This first step in the analysis has three advantages: it indicates whether a correlation structure exists or not, it can help the user in determining an initial form of the covariance matrix, and it gives initial values of variance components that may help convergence to a maximum likelihood solution. However, this technique cannot be used to select an appropriate correlation function. Another way to explore the correlation structure would be to fit a mixed model with an unstructured covariance structure and to use the resulting estimated covariance matrix to suggest a more parsimonious structure. Nevertheless, when the number of time points is large, the REML algorithm may not converge [8]. This was the case for semen production records and the graphical technique was then more suitable.

Fitting linear models implies that an appropriate mean structure as well as a covariance structure be specified, but these structures are not independent of each other [34]. For our analysis, the covariance structure for the STE was determined first, using a full fixed model, while the fixed effects were tested later. Nevertheless, to confirm the choice of the correlation structure for the STE, the different correlation structures were also compared with the reduced mean model and the spatial power structure was still "optimal". A robust procedure for choosing the fixed model was proposed by Liang and Zeger [22]. It is relatively insensitive to the structure of the variance covariance assumed for the data. Verbeke and Molenberghs [34] showed that this estimator is consistent as long as the mean is correctly specified in the model. In agreement with Robert-Granié *et al.* [29], the results we obtained with this approach showed that, in most cases, the robust standard error was smaller than the naïve one.

Concerning the selection of $g_1(x)$, our analysis clearly showed an STE correlation of sequential measurements. Rams were collected with unequal time lags, and models using a function of time-lag in days between collections (SP_S and SG_S structures) fitted the data better than the AR_S model based on the order of collections rather than the actual time interval. The daily correlation of the STE in the final model (AR_L/SP_S) is high: 0.96. However, this correlated component is one-third the size of the uncorrelated component ($\tau^2/\sigma_e^2 \simeq 0.27$). These results indicate that an important component of unexplained and uncorrelated daily variation remained.

With regards to the selection of $g_2(x)$, the assumption of independent LTE was discarded ($C0_L/SP_S$ model) indicating that some influences on semen production persist over years. However, not all influences remain constant since the $C1_L/SP_S$ model with a correlation of 1 was not as good as the AR_L/SP_S and UC_L/SP_S model with a smaller correlation. The autoregressive correlation was better than the uniform correlation. Hypotheses similar to those proposed by Carvalheira *et al.* [3] for cow milk production may explain this result: many events that affect a ram's capacity for semen production occur during productive life, and impart a correlation structure between years of production that decays with time. Several authors [4, 31] have already used the first order autoregressive correlation structure for the analysis of LTE in cow lactations. Similar to the result obtained when omitting the permanent environmental effect in the case of repeated measurements [5], assuming independent LTE ($C0_L/SP_S$ model) yielded higher genetic variance and heritability (0.43 *vs.* 0.27, 0.29) estimates than considering a correlation between LTE because the genetic component picks up the covariance omitted from the environmental component. The results found concerning the correlation processes for STE and LTE mean that there may be some uncontrolled factors, which modify over years the so called "permanent" environmental effect (*i.e.* change in the location of the pen in the shed, modification of social behaviour when changing pen...) and other factors, which induce within year correlations between residuals of the simple repeatability model (health status, body weight change...). Therefore, one has to identify and control these factors to improve semen production and increase heritability and repeatability of semen traits.

4.3. Variance components and genetic parameters estimates

Once the appropriate correlation structures for LTE and STE were selected, we compared the component sub-models with the classic repeatability model. The results showed that the AR_L/SP_S model fitted the data the best. The

heritabilities of these four models ranged from 0.27 to 0.33 and the correlations between estimated breeding values for semen volume estimated with these models were high (>0.98). These results may be specific to the data and the consequence of the high correlations for LTE and STE, indeed the results obtained for similar comparisons in other studies differed a little [3, 4, 31, 32]. However, the results obtained with the AR_L/SP_S model *vs.* the three other models relaxed the assumptions of permanent environmental effect and independent residuals, which are the usual assumptions used in the study of repeated semen trait [2, 10, 11, 20, 23, 24].

The estimate of heritability with the AR_L/SP_S model was moderate, higher than that reported in the literature for ram semen volume (0.07, 0.11 [28], 0.15, 0.20 [10]) and in the middle of the range of heritability reported on all species (from 0.07 to 0.58 [28, 33]).

5. CONCLUSION

We used a character process model to analyse semen volume in AI rams that has previously been used by several authors [3, 4] for test-day records of cows, but with a different modelling process. An empirical semi-variogram gives an informal check for serial correlation. The spatial power correlation and the first order autoregressive structure gave the best fit to the correlation between STE and LTE, respectively. Even if variance components were not markedly affected in this study by the introduction of STE and a first order autoregressive correlation structure for LTE *vs.* the classical simple repeatability model, the proposed model insured an appropriate selection of fixed effects, identified uncontrolled factors and gave more information about the decline of repeatability with time. The heritability estimate was moderate indicating that selection of AI rams on their ability to produce semen volume could lead to a substantial improvement in the number of doses produced per animal. However, the impact of this selection on other traits must be evaluated.

ACKNOWLEDGEMENTS

The authors thank the *ministère de l'Agriculture* for supporting this study in the frame of a "*BELIA* action", directed by the *ANIO* and *INRA*, and the AI centres who provided the data: BMC, CCDEO, Confédération générale de Roquefort, Insem-Ovin, OVI-TEST, CIA Verdilly. Two anonymous reviewers are thanked for numerous constructive criticisms and for improving the clarity of this paper.

REFERENCES

[1] Basso B., Fritz S., Druet T., Guillaume F., Rossignil M.N., Amigues Y., Gabriel R., Sellem E., Salas-Cortes L., Humblot P., Druart X., Estimation de paramètres génétiques et détection de QTL liés à des caractères de fertilité mâle, de production de semence et de qualité de la semence chez le taureau laitier, in: Rencontres Recherches Ruminants, 7–8 décembre 2005, Paris, pp. 145–148.

[2] Brandt H., Grandjot G., Genetic and environmental effects of male fertility of AI-boars, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Armidale, pp. 527–530.

[3] Carvalheira J.G., Blake R.W., Pollak E.J., Quaas R.L., Duran-Castro C.V., Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in United States Holstein herds, J. Dairy Sci. 81 (1998) 2738–2751.

[4] Carvalheira J., Pollak E.J., Quaas R.L., Blake R.W., An autoregressive repeatability animal model for test-day records in multiple lactations, J. Dairy Sci. 85 (2002) 2040–2045.

[5] Clement V., Estimation des paramètres génétiques des petits ruminants en milieu d'élevage traditionnel au Sénégal. Importance de la structure des données et du choix du modèle d'analyse, 1999, Institut national agronomique Paris-Grignon, Paris, p. 117.

[6] David I., Bodin L., Lagriffoul G., Manfredi E., Robert-Granié C., Genetic parameters of ram semen traits using an animal model accounting for serial correlations, in: Proceedings of the 8th World Congress on Genetic Applied to Livestock Production, 13–18 August 2006, Belo Horizonte, Brazil.

[7] David I., Druart X., Lagriffoul G., Manfredi E., Robert-Granié C., Bodin L., Genetic and environmental factors affecting semen traits in Lacaune and Manech tête rousse AI rams, Genet. Sel. Evol. (2007) (in press).

[8] Dawson K.S., Gennings C., Carter W.H., Two graphical techniques useful in detecting correlation structure in repeated measures data, Am. Stat. 51 (1997) 275–283.

[9] Diggle P.J., An approach to the analysis of repeated measurements, Biometrics, 44 (1988) 959–971.

[10] Duval P., Belloc J.P., Albaret M., Girou P., Barillet F., Etude de facteurs de variation de la fonction sexuelle de béliers Lacaune lait et de la fertilité des brebis en insémination artificielle, in: Rencontres Recherches Ruminants, December 1995, Paris, pp. 429–434.

[11] Garcia-Tomas M., Sanchez J., Rafel O., Ramon J., Piles M., Variability, repeatability and phenotypic relationships of several characteristics of production and semen quality in rabbit, Anim. Reprod. Sci. 93 (2006) 88–100.

[12] Gilmour A.R., Thompson R., Statistical models for multidimensional (longitudinal/spatial) data, in: Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, 13–18 August 2006, Belo Horizonte, Brazil.

[13] Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., Thompson R., ASReml User Guide Release 1.0, VSN International Ltd, 2002, Hemel Hempstead, HP11ES, UK.

[14] Henderson C.R., Jr., Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions, Biometrics 38 (1982) 623–640.
[15] Jaffrézic F., Pletcher S.D., Statistical models for estimating the genetic basis of repeated measures and other function-valued traits, Genetics 156 (2000) 913–922.
[16] Jaffrézic F., Thompson R., Pletcher S.D., Multivariate character process models for the analysis of two or more correlated function-valued traits, Genetics 168 (2004) 477–487.
[17] Jaffrézic F., Venot E., Laloë D., Vinet A., Renand G., Use of structured antedependence models for the genetic analysis of growth curves, J. Anim. Sci. 82 (2004) 3465–3473.
[18] Jensen J., Genetic evaluation of dairy cattle using test-day models, J. Dairy Sci. 84 (2001) 2803–2812.
[19] Kettunen A., Mäntysaari E.A., Strandén I., Pösö J., Lidauer M., Estimation of genetic parameters for first lactation test day milk production using random regression models, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Armidale, New South Wales, Australia, pp. 307–310.
[20] Langford G.A., Shrestha J.N.B., Marcus G.J., Repeatability of scrotal size and semen quality measurements in rams in a short-day light regime, Anim. Reprod. Sci. 19 (1989) 19–27.
[21] Lesaffre E., Todem D., Verbeke G., Flexible modelling of the covariance matrix in a linear random effects model, Biometrical J. 42 (2000) 807–822.
[22] Liang K.Y., Zeger S.L., Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.
[23] Manfredi E., Leboeuf B., Bodin L., Boue P., Humblot P., Source de variations génétiques et non génétiques des caractéristiques de production de semence chez le bouc, in: Rencontres Recherches Ruminants, 2–3 décembre 1998, pp. 37–39.
[24] Mathevon M., Buhr M.M., Dekkers J.C., Environmental, management, and genetic factors affecting semen production in Holstein bulls, J. Dairy Sci. 81 (1998) 3321–3330.
[25] Meyer K., Estimating covariance functions for longitudinal data using a random regression model, Genet. Sel. Evol. 30 (1998) 221–240.
[26] Meyer K., Random regressions to model phenotypic variation in monthly weights of Australian beef cows, Livest. Prod. Sci. 65 (2000) 19–38.
[27] Pletcher S.D., Geyer C.J., The genetic analysis of age-dependent traits: modeling the character process, Genetics 153 (1999) 825–835.
[28] Rege J.E., Toe F., Mukasa-Mugerwa E., Tembely S., Anindo D., Baker R.L., Lahlou-Kassi A., Reproductive characteristics of Ethiopian highland sheep. II. Genetic parameters of semen characteristics and their relationships with testicular measurements in ram lambs, Small Rumin. Res. 37 (2000) 173–187.
[29] Robert-Granié C., Heude B., Foulley J.L., Modelling the growth curve of Maine-Anjou beef cattle using heteroskedastic random coefficients models, Genet. Sel. Evol. 34 (2002) 423–445.
[30] SAS®, Cary NC: SAS® Institute Inc., SAS®/STAT Software, version 8, 1999.

[31] Sawalha R.M., Keown J.F., Kachman S.D., Van Vleck L.D., Evaluation of autoregressive covariance structures for test-day records of Holstein cows: estimates of parameters, J. Dairy Sci. 88 (2005) 2632–2642.
[32] Sawalha R.M., Keown J.F., Kachman S.D., Van Vleck L.D., Genetic evaluation of dairy cattle with test-day models with autoregressive covariance structures and with a 305-d model, J. Dairy Sci. 88 (2005) 3346–3353.
[33] Smital J., Wolf J., De Sousa L.L., Estimation of genetic parameters of semen characteristics and reproductive traits in AI boars, Anim. Reprod. Sci. 86 (2005) 119–130.
[34] Verbeke G., Molenberghs G., Linear mixed models for longitudinal data, Ed. Springer Series in Statistics, Hardcover, 1997.
[35] Verbeke G., Lesaffre E., Brant L.J., The detection of residual serial correlation in linear mixed models, Stat. Med. 17 (1998) 1391–1402.