MODELLING ANIMAL SYSTEMS RESEARCH PAPER

Comparison of growth curves of two genotypes of dairy goats using nonlinear mixed models

J. G. L. REGADAS FILHO[1]*, L. O. TEDESCHI[2], M. T. RODRIGUES[1], L. F. BRITO[1] AND T. S. OLIVEIRA[1]

[1] *Departamento de Zootecnia, Universidade Federal de Viçosa, MG, Brazil*
[2] *Department of Animal Science, Texas A&M University, College Station, TX 77843-2471, USA*

SUMMARY

The objective of the current study was to assess the use of nonlinear mixed model methodology to fit the growth curves (weight *v.* time) of two dairy goat genotypes (Alpine, +A and Saanen, +S). The nonlinear functions evaluated included Brody, Von Bertalanffy, Richards, Logistic and Gompertz. The growth curve adjustment was performed using two steps. First, random effects u_1, u_2 and u_3 were linked to the asymptotic body weight (β_1), constant of integration (β_2) and rate constant of growth (β_3) parameters, respectively. In addition to a traditional fixed-effects model, four combinations of models were evaluated using random variables: all parameters associated with random effects (u_1, u_2 and u_3), only β_1 and β_2 (u_1 and u_2), only β_1 and β_3 (u_1 and u_3) and only β_1 (u_1). Second, the fit of the best adjusted model was refined by using the power variance and modelling the error structure. Residual variance (σ_e^2) and the Akaike information criterion were used to evaluate the models. After the best fitting model was chosen, the genotype curve parameters were compared. The residual variance was reduced in all scenarios for which random effects were considered. The Richards (u_1 and u_3) function had the best fit to the data. This model was reparameterized using two isotropic error structures for unequally spaced data, and the structure known in the literature as SP(MATERN) proved to be a better fit. The growth curve parameters differed between the two genotypes, with the exception of the constant that determines the proportion of the final size at which the inflection point occurs (β_4). The nonlinear mixed model methodology is an efficient tool for evaluating growth curve features, and it is advisable to assign biologically significant parameters with random effects. Moreover, evaluating error structure modelling is recommended to account for possible correlated errors that may be present even when using random effects. Different Richard growth curve parameters should be used for the predominantly Alpine and Saanen genotypes because there are differences in their growth patterns.

INTRODUCTION

Studies on the growth curves for livestock have been used extensively to examine how body weight and other characteristics of interest (fat deposition, organ size, etc.) develop over time and in relation to the environment, feed, genotype and other factors. Normally, this relationship (weight *v.* time) is assessed using nonlinear functions, such as $y_i = f(x_i, \beta_i) + e_i$, in which y_i is the average body weight of the animal at time x_i, y_i is estimated using a nonlinear function f with β_i parameters, whose estimates are obtained by using ordinary least-squares regression, and e_i represents the unexplained error. The errors are assumed to be normally, identically and independently distributed with a mean zero and a constant variance, or NIID $(0, \sigma^2)$. On the basis of these assumptions, the nonlinear regression model is considered a fixed-effects model (Craig & Schinckel 2001).

However, when multiple observations are recorded in the same experimental unit, a parameter or coefficient that varies from one unit to another can be considered random (Peek *et al.* 2002). This proposition is based on the concept of random effects, in which an effect is considered random if its levels represent values of a larger population with a probability distribution (Littell *et al.* 2006), which is the case, for example, for the asymptotic body weight

* To whom all correspondence should be addressed. Email: gilsonagro@yahoo.com.br

Table 1. *Summary of the data set assessed*

Genetic composition	No. of animals	No. of weights	Max. obs. per goat	Min. obs. per goat	Max. age (days)
+S	498	4292	21	4	2666
+A	658	5707	18	4	3180
Total	1156	9999	–	–	–

(a common parameter in most of the models used in livestock growth studies).

In addition, measurements taken on the same experimental unit are more closely related than those on other experimental units; on the same experimental unit, measures at short intervals are more correlated than those at longer intervals (both within the same unit) (Littell *et al.* 2006). These relationships are inherent for this type of data, which may disrupt the basic assumptions of statistical analysis, such as independence of errors and variance homogeneity (Craig & Schinckel 2001; Littell *et al.* 2006; Strathe *et al.* 2010). Thus, if random effects are not added to the model, the estimated standard error of the parameters may be biased because the assumption of independence of errors might be violated (Peek *et al.* 2002).

Another factor that supports the use of the nonlinear mixed model methodology is the type of data normally obtained for use in growth studies. In general, the data are unbalanced, with different numbers of body weight measurements for different animals and a tendency for the number of animals in the study to decrease over time due to death, slaughter, disposal and other factors. The unbalanced nature of this type of data can lead to bias in the estimated parameters when using the conventional method (Craig & Schinckel 2001; Wang & Zuidhof 2004).

However, there is still controversy about which growth curve parameters normally used in the literature should be considered as random. This decision should be based on the biological interpretation of the parameters, the significance of the estimated variance components and, above all, common sense.

Given this context, the objectives of the study described in the current paper were the following: (1) to identify the best commonly used nonlinear function to describe dairy goat growth, (2) to determine which growth curve parameters should be linked to random coefficients, (3) to compare the curves obtained from the nonlinear mixed model methodology to curves obtained with traditional methodology

and (4) to compare the curve parameters obtained for two genotypes of dairy goats.

MATERIALS AND METHODS

Animals and data set descriptions

The data used in the current study were collected from the goat herd of the Goat Sector at the Federal University of Viçosa (Universidade Federal de Viçosa-UFV), Viçosa, MG, Brazil. A total of 14 003 weighing records from female goats were collected between 1992 and 2010. Only those goats with four or more recorded weight measurements were included. Data from animals without recorded birth dates were excluded. Then the data were divided according to the genotype of the animals: Saanen (+S) and Alpine (+A). A graphical analysis was conducted to assess the consistency of the data and to identify the excluded end points, including data from animals of all other genotypes. The database is summarized in Table 1.

Animal management consisted of a free-stall milk production system in which animals had access to maize silage and concentrate *ad libitum*. There are normally two calving seasons per year for goats; however, each animal was allowed just one calving per year. During the growth phase, the animals were weaned, on average, when they were 18 kg (110–140 days old) and the first pregnancy occurred, on average, at 35 kg body weight (340–370 days old). The goats lactated for *c*. 300 days, and their average milk production was 2·6 kg milk/day. The number of lactations ranged from zero to seven.

The number of weighings by animal ranged from 4 to 21 (Table 1) and the minimum and maximum differences between two consecutive weighings were 1 and 1190 days, respectively. These features make the data set extremely unbalanced, spatially and numerically. Furthermore, the initial graph analysis suggested that there was an increase in the variance of the data over time (the independent parameter). Thus, a special approach to modelling the growth curves was required.

Table 2. *Evaluated growth curve functions*

Model	General function
Brody	$\hat{y} = \beta_1(1 - \beta_2\,e^{-\beta_3 t})$
Von Bertalanffy	$\hat{y} = \beta_1(1 - \beta_2\,e^{-\beta_3 t})^3$
Richards	$\hat{y} = \beta_1(1 - \beta_2\,e^{-\beta_3 t})^{\beta_4}$
Logistic	$\hat{y} = \beta_1/(1 + \beta_2\,e^{-\beta_3 t})$
Gompertz	$\hat{y} = \beta_1\,e^{-\beta_2 e^{-\beta_3 t}}$

Growth curve functions

Five nonlinear models commonly used in the literature for describing animal growth curves were evaluated. In these models, shown in Table 2, y is the animal weight at time t_i; β_1 the estimated body weight of the animal when $t \to \infty$ or asymptotic body weight; β_2 the constant of integration; β_3 the rate constant, which determines the spread of the curve along the time axis; and β_4 is the constant that determines the proportion of the final size at which the inflection point (IP) occurs. The IP can be determined using the equation $IP_{(\%\beta_1)} = \beta_4^{1/(1-\beta_4)}$, and the instantaneous absolute growth rate can be calculated using the equation $IAGR = \beta_3 \times Y_{est}[(\beta_1/Y_{est})^{1-\beta_4} - 1]/(1 - \beta_4)$, where Y_{est} is the body weight estimated at a given time (Richards 1959).

Nonlinear mixed model

A nonlinear mixed model methodology was used to obtain the growth curve parameters according to the following expression:

$$\mathbf{y}_i = f(\mathbf{x}_{ij}, \boldsymbol{\beta}, \boldsymbol{u}_i) + \mathbf{e}_i \qquad (1)$$

where $\mathbf{y}_i$ is a vector $(n_i \times 1)$ of recorded body weights for the subject $i = 1, 2, \ldots, m$; m is the number of goats; f is a nonlinear function of the covariate matrix $\mathbf{x}_{ij}$; $\mathbf{x}_{ij}$ is a matrix $(n_i \times 2)$ of independent variables, in which the first column contains the ith age of measurement and the second column contains the jth subject (goat) whose body weight was measured; $\boldsymbol{\beta}$ is a vector $(p_i \times 1)$ of unknown fixed-effect parameters; $\boldsymbol{u}_i$ is a vector $(q_i \times 1)$ of unknown random-effect parameters unique to the subject goat i and assumed to follow a multivariate distribution with a mean of zero and a conditional unstructured covariance matrix, $n \times n$ ($\mathbf{G}$); and $\mathbf{e}_i$ is a vector $(n_i \times 1)$ of the error term assumed to follow a multivariate random normal distribution with zero mean and conditional covariance matrix structure ($\mathbf{R}$).

The dependent variable (y) in this case is influenced by quantitative (time) and qualitative (genotype) variables. The effects of the quantitative variables are analysed via nonlinear regression. However, it is also necessary to assess the effects of the qualitative variable on the dependent variable and the regression parameters; thus, *dummy* variables z_1 and z_2 were created, where for the +S genotype, $z_1 = 0$ and $z_2 = 1$, and for the +A genotype, $z_1 = 1$ and $z_2 = 0$. With this approach, it was possible to estimate growth curve parameters (β_1, β_2, β_3 and β_4) independently for each genotype.

The growth curve modelling consisted of two steps, as described below. First, a set of candidate nonlinear mixed models were formulated that were preliminarily analysed to choose the function with the best fit. This first step was used to evaluate which curve parameters should have a random effect component and compared them with the traditional fixed-effect model. However, despite the high hardiness with which nonlinear mixed model methodology treats the correlated errors, it is important to note that heterogeneity and correlated errors can occur even with the inclusion of random effects (Meng & Huang 2010; Yang & Huang 2011). Thus, a refinement of the model chosen in the first step was necessary to accommodate the violated model assumptions (second step).

First step

A central phase in the model-building of mixed-effects models is to decide which of the parameters in the nonlinear model require random effects to account for their between-subject variations and which can be treated as purely fixed effects (Pinheiro & Bates 2000). Thus, the random parameters u_1, u_2 and u_3 were created to account for between-subject variations for the β_1, β_2 and β_3 parameters, respectively, enabling curve parameters to be estimated for each goat (Littell et al. 2006). No random parameter was created for the β_4 parameter in the Richards function because it was difficult to achieve the convergence criterion and the high correlations between the estimated parameters (overparameterized model). The following four combinations of random effects were evaluated: all of the parameters related to random effects (u_1, u_2 and u_3), only parameters related to β_1 and β_2 (u_1 and u_2), only parameters related to β_1 and β_3 (u_1 and u_3), and only that related to β_1 (u_1). To avoid problems with floating-point errors and overflows, β_3 was rescaled ($\times 1000$) in

the models due to the difference in magnitude of this parameter in relation to the others.

First, the equations were fitted to each model and previously described combinations of random effects through PROC NLMIXED (SAS 9.3, see Appendix). The FIRO method was used to achieve convergence. The FIRO method uses the First-Order Method described by Beal & Sheiner (1982) to compute the integral over the random effects to the marginal maximum likelihood. The fixed-effects model was also estimated using PROC NLMIXED, except that, in this case, variance components were not included. The FIRO method does not support the model without random effects; thus, the Adaptive Gaussian Quadrature described by Pinheiro & Bates (1995) was used. The estimated parameters, standard error and residual variance for the fixed-effects model were similar to those of the model obtained using PROC NLIN.

The following criteria were adopted to select the function that best described the growth curve of the goats: (1) convergence (the interactive process in NLMIXED converges at the nth interaction when $(g_k H_k^{-1} g_k)/(|l_k| + 1 e^{-6}) < 1 e^{-8}$ (SAS Institute Inc. 2008)); (2) the final Hessian matrix is non-singular and positive definite; (3) the approach described by Burnham & Anderson (2002) and Vieira *et al.* (2012) using the Akaike information criterion (AIC), the difference among AIC values (Δ_r), the Akaike weights or likelihood probabilities (w_r) and the evidence ratio or relative likelihood (ER$_r$), which can be computed using the following equations:

$$\text{AIC} = 2f(\widehat{\theta}) + 2p$$

$$\Delta_r = \text{AIC}_r - \min \text{AIC}_r$$

$$w_r = \frac{\exp(-\Delta_r/2)}{\sum_{r=1}^{R} \exp(-\Delta_r/2)}$$

$$\text{ER}_r = \frac{\max w_r}{w_r}$$

where $f()$ is the negative of the marginal log-likelihood function, $\widehat{\theta}$ is the vector of parameter estimate, and p is the number of parameters (SAS Institute Inc. 2008); and (4) a graphical analysis of the Pearson residuals against the predicted values was used to evaluate the model assumptions. The Pearson residuals were obtained for the fixed model and random model without a correlation structure as $r_i = e_i/\sqrt{\widehat{\sigma_e^2}}$ and for the model with a correlation structure (second step described below) as $r_i = \hat{C}'^{-1}(y_i - \hat{y}_i)$, where $\hat{C}$ denotes the Cholesky root of the estimated **R** matrix (SAS Institute Inc. 2008).

Second step

The model with the best goodness-of-fit (combination between nonlinear equation and random parameters) was selected for further analysis.

Despite there being no sign of heteroscedasticity in the residuals of the chosen model, it was easy to identify high Pearson residuals obtained even when using random effects in the model. The higher values found for the Pearson residuals, many of them above three, were an indication of problems in the model. The first remedy would be to exclude the data from the data set and reparameterize the model, which would solve the problem but cause a loss of information.

Therefore, it was decided to use two tools to address this problem and the remaining correlated errors. The residual variance was modelled using the power-of-the-mean variance (PV) function. Using the notation described by Littell *et al.* (2006), the variance matrix **R** is assumed to be of the form $\boldsymbol{R} = diag(\sigma^2 |x_i' \beta^*|^\theta)$, where θ is the power to be estimated. In order to choose the value of θ, the power that minimized the approximate -2log-likelihood was calculated (Littell *et al.* 2006); this θ value was estimated to be 2·91.

As previously noted, correlated errors can occur even when random effects are used in the model. These errors primarily occur if there are parameters in the function that do not have a random parameter to account for their between-subject variability. An error structure matrix can be modelled to account for this interrelation. When the model is fitted without modelling the serial correlation, the independently and identically distributed error structure is assumed to be $Var[e_i] = \sigma_e^2$ and $cov[e_i, e_j] = 0$. However, there is one intrinsic characteristic of the current data set that must be considered, namely, the unequally spaced measurement intervals for the animal subjects.

There is a range of structures available in the literature to address this issue. Structures that accept unequally spaced data have been described by Littell *et al.* (2006) and the SAS Institute Inc. (2008). Normally, the covariance is assumed to be a function of the distance between locations. If d_{ij} denotes the distance between the measurements made on the same animal, the covariance models have the general form $COV[e_i, e_j] = \sigma_e^2 [f(d_{ij})]$ (Littell *et al.* 2006).

In the second step, all possible isotropic candidate structures described in the literature were examined;

however, only two structures met the convergence criterion: the spatial power function (SP(POW)) and the MATERN function (SP(MATERN)).

The spatial power structure provides a direct generalization of the auto-regressive structure for equally spaced data and is assumed to be $f(d_{ij}) = \rho^{d_{ij}}$. The MATERN isotropic covariance function is given as $f(d_{ij}) = \dfrac{(d_{ij}/2V)^V 2K_v(d_{ij}/\rho)}{\Gamma(V)}$ where K_V is a modified Bessel function of the second kind of order V; Γ is the gamma function; ρ is a scalar parameter controlling the spatial range of correlation; and V is the 'smoothness' parameter, which allows greater flexibility for modelling the local spatial covariance. When V is small ($V \to 0$), the spatial process is assumed to be rough, and when V is large ($V \to \infty$), the process is assumed to be smooth (Littell *et al.* 2006; Minasny & McBratney 2007; SAS Institute Inc. 2008).

The NLMIXED procedure does not support the modelling of error structure directly; thus, the SAS macro %NLINMIX was used to refine the model to account for the chosen error structure. The restricted maximum likelihood was used along with the two expansion methods available in the SAS macro %NLINMIX (ZERO and EBLUP) to attempt to fit the model. The model adjusted in this way was compared with the fixed model and the random parameter model that was estimated using the previously described approach.

After selecting the model that best described the growth curves and included the previously estimated variance and covariance matrices, the hypotheses regarding one or more growth curve parameters could be tested. Thus, the difference between the growth curve parameters of the genotypes were tested, with the objective of simplifying the model ($\alpha = 0.05$).

Cross-validation

A cross-validation was performed to evaluate the goodness-of-fit of the chosen function. The original data set was randomly partitioned into two sub-data sets (training and test data), each containing 0·50 of the animals per genotype. The division of the original data set was conducted using PROC SURVEYSELECT from the SAS software (version 9.3). The first sub-data set (training data set) was used to fit the previously chosen best model (steps 1 and 2), and the second sub-data set (test data set) was used to evaluate the robustness of the model. Equation precision was measured with the

coefficient of determination (R^2) between the observed and predicted values and the simultaneous F-test of the intercept and slope (intercept = 0 and slope = 1), whereas the accuracy of the equation was determined based on the concordance correlation coefficient (CCC) and the root-mean-square error prediction (RMSEP) and its decomposition into mean bias, systematic bias and random errors (Tedeschi 2006). The model evaluation was implemented using the Model Evaluation System v. 3.1.13 (http://nutritionmo-dels.tamu.edu/mes.html; verified 28 May 2013). Moreover, the parameters of the curve fit from the training data set were compared with the parameters of the curves fit from the original data set using 95% confidence intervals ($IC_{95\%}$).

Impact of the number of weights on random parameters

One of the major advantages of the nonlinear mixed model methodology is the possibility of having individual growth parameters adjusted for each animal and testing to determine how some parameters differ from the population mean. Using the chosen model and the individual deviation of asymptotic and constant rate parameters, it was possible to evaluate the effect of the number of weights on the parameters estimated for each goat using descriptive graphic analysis.

RESULTS

One function did not achieve the convergence criterion (Richards u_1, u_2 and u_3). The Richards (u_1 and u_2) model presented the singular Hessian matrix, which prevents its unique inversion. The Hessian matrix rendered by the Brody model was not positive definite, and the Gompertz (u_1, u_2 and u_3) model had at least one negative eigenvalue.

Table 3 shows the estimated parameters for all of the evaluated models and combinations of random variables. Although represented by the same Greek letters, except for asymptotic body weight (β_1), the estimated parameters cannot be compared among nonlinear functions (Forni *et al.* 2009). The only comparisons that are possible are between combinations of random variables or between assessed genotypes within each function. Among all of the scenarios that were assessed, the +S animals consistently had higher asymptotic body weights. Estimates of asymptotic body weight varied from 43·20 to 58·11 kg for the

Table 3. *Estimated parameters for the different models, random variable combinations and genotypes assessed*

Model	Random variables	+A β_1	β_2	$\beta_{3\ (x\ 1000)}$	β_4	+S β_1	β_2	$\beta_{3\ (x\ 1000)}$	β_4
Brody	u_1	54·6	0·93	2·7	–	60·3	0·93	2·3	–
	u_1, u_2	54·6	0·93	2·7	–	60·0	0·93	2·3	–
	u_1, u_3	52·6	0·93	2·8	–	54·7	0·94	2·7	–
	u_1, u_2, u_3	52·9	0·93	2·7	–	54·3	0·94	2·7	–
	Fixed	54·6	0·92	2·6	–	58·6	0·92	2·4	–
Von Bertalanffy	u_1	51·6	0·51	4·2	–	55·9	0·51	3·7	–
	u_1, u_2	51·6	0·52	4·2	–	56·5	0·52	3·7	–
	u_1, u_3	48·5	0·51	4·4	–	50·7	0·52	4·1	–
	u_1, u_2, u_3	49·3	0·84	3·8	–	54·0	0·86	3·5	–
	Fixed	52·4	0·50	3·9	–	55·8	0·50	3·6	–
Logistic	u_1	48·5	4·9	7·5	–	52·1	4·9	6·7	–
	u_1, u_2	48·1	5·7	7·8	–	52·2	5·6	6·8	–
	u_1, u_3	45·8	4·8	7·7	–	48·2	4·9	7·0	–
	u_1, u_2, u_3	45·6	4·9	7·6	–	48·0	4·9	6·9	–
	Fixed	50·3	4·5	6·5	–	53·3	4·5	6·0	–
Gompertz	u_1	50·6	2·0	4·9	–	54·6	2·0	4·4	–
	u_1, u_2	50·5	2·1	5·0	–	55·2	2·1	4·4	–
	u_1, u_3	47·5	2·0	5·2	–	49·7	2·0	4·8	–
	u_1, u_2, u_3	43·2	4·5	5·6	–	48·1	4·7	5·3	–
	Fixed	51·7	1·9	4·5	–	54·9	1·9	4·1	–
Richards	u_1	57·4	0·98	1·9	0·73	67·4	0·99	1·3	0·67
	u_1, u_2	53·9	0·98	2·2	0·75	55·5	0·99	1·9	0·70
	u_1, u_3	56·4	0·98	2·0	0·77	57·9	0·98	2·0	0·79
	u_1, u_2, u_3			Did not converge					
	Fixed	58·1	1·0	1·6	0·65	64·7	0·99	1·3	0·64

+A genotype and from 48·00 to 67·38 kg for the +S genotype.

The calculated birth weight (calculated as time = 0) for the Logistic and Gompertz (all combinations of random effects) models produced unrealistic estimated birth weight values. Reasonable estimates were made by the Brody and Richards (all combination of random effects) models.

High β_3 parameter values indicate higher precocity, i.e. a higher fractional rate at which the animal approaches asymptotic body weight (Brown *et al.* 1976). The +A animals presented higher values (except for the Richards (u_1 and u_3)) for this parameter than the +S animals, possibly indicating earlier development.

Table 4 shows the values of the criteria used to select the model that best describes goat growth in addition to the estimated variance and covariance components. The fixed-effect models were found to always display higher residual variance than the random-effect models did. The Richards (u_1 and u_3) function had the best fit. The value of Δ_r to the nearest model was 86,

which, according to Burnham & Anderson (2002), indicates essentially no support. The large Δ_r values obtained became a calculation of the likelihood probability and evidence ratio unnecessary, as these criteria are equal to 1 for the Richards (u_1 and u_3) model.

The random effects of the Richards (u_1 and u_3) model were significant ($P<0.001$). The value found for the component that measures the population variability of parameter β_1 was 209·3, and for β_3, this value was 1·20. However, high Pearson residuals were observed when the Richards (u_1 and u_3) model (Figs 1 (*b*) and (*c*)) was used. Thus, this model was chosen for further analysis.

In the second step, the variance power and the modelled error structure matrix were included. Only the SP(POW) and SP(MATERN) structure achieved the convergence criterion. The use of these tools produced low values of the AIC when compared with the previous approach. The SP(MATERN) structure yielded better adjustments, as determined by the fact that $\Delta_r=1126$ (Table 5).

Table 4. *Variance and covariance components, residual variances and information criteria of the fitted growth curves*

Model	Random variables	σ^2_{u1}	σ_{u1u2}	σ_{u1u3}	σ^2_{u2}	σ_{u2u3}	σ^2_{u3}	σ^2_e	AIC	Δ_r
Brody	u_1	74·9 (P<0·001)						12·9	57040	2354
	u_1, u_2	80·0 (P<0·001)	0·08 (P<0·001)		$-4·0\times10^{-05}$ NS			13·0	56999	2313
	u_1, u_3	156·5 (P<0·001)		$-13·3$ (P<0·001)			1·8 (P<0·001)	8·7	54992	306
	u_1, u_2, u_3	160·5 (P<0·001)	0·12 (P<0·001)	$-13·2$ (P<0·001)	$-3·6\times10^{-04}$ (P<0·001)	$-7·6\times10^{-03}$ (P<0·001)	1·8 (P<0·001)	8·8	54772	86
	Fixed							33·5	63490	8804
Von Bertalanffy	u_1	64·8 (P<0·001)						16·6	59293	4607
	u_1, u_2	78·1 (P<0·001)	0·17 (P<0·001)		$2·4\times10^{-03}$ (P<0·001)			15·1	59083	4397
	u_1, u_3	119·2 (P<0·001)		$-16·9$ (P<0·001)			4·1 (P<0·001)	10·8	57063	2377
	u_1, u_2, u_3	167·7 (P<0·001)	$-6·7$ (P<0·001)	$-19·6$ (P<0·001)	0·73 (P<0·001)	1·52 (P<0·001)	3·9 (P<0·001)	7·7	57287	2601
	Fixed							37·69	64679	9993
Logistic	u_1	56·3 (P<0·001)						24·11	62627	7941
	u_1, u_2	69·4 (P<0·001)	8·5 (P<0·001)		5·0 (P<0·001)			19·9	62005	7319
	u_1, u_3	93·1 (P<0·001)		$-22·0$ (P<0·001)			9·7 (P<0·001)	17·0	60983	6297
	u_1, u_2, u_3	106·5 (P<0·001)	7·2 (P<0·001)	$-20·6$ (P<0·001)	0·37 (P<0·01)	$-1·9$ (P<0·001)	7·8 (P<0·001)	16·7	60843	6157
	Fixed							45·7	66613	11927
Gompertz	u_1	61·8 (P<0·001)						18·7	60323	5637
	u_1, u_2	76·1 (P<0·001)	1·1 (P<0·001)		0·11 (P<0·001)			16·3	59980	5294
	u_1, u_3	110·1 (P<0·001)		$-18·5$ (P<0·001)			5·4 (P<0·001)	12·3	58207	3521
	u_1, u_2, u_3*	208·5	$-99·1$	$-54·9$	70·1	35·4	19·1	8·9	58134	3448
	Fixed							39·9	65245	10559
Richards	u_1	83·6 (P<0·001)						11·9	56287	1601
	u_1, u_2†	30·7	$-1·0\times10^{-02}$		$-1·7\times10^{-04}$			13·5	56900	2214
	u_1, u_3	209·3 (P<0·001)		$-13·0$ (P<0·001)			1·2 (P<0·001)	8·5	54686	0
	u_1, u_2, u_3				Did not converge					
	Fixed							31·7	62960	8274

* At least one eigenvalue was negative.
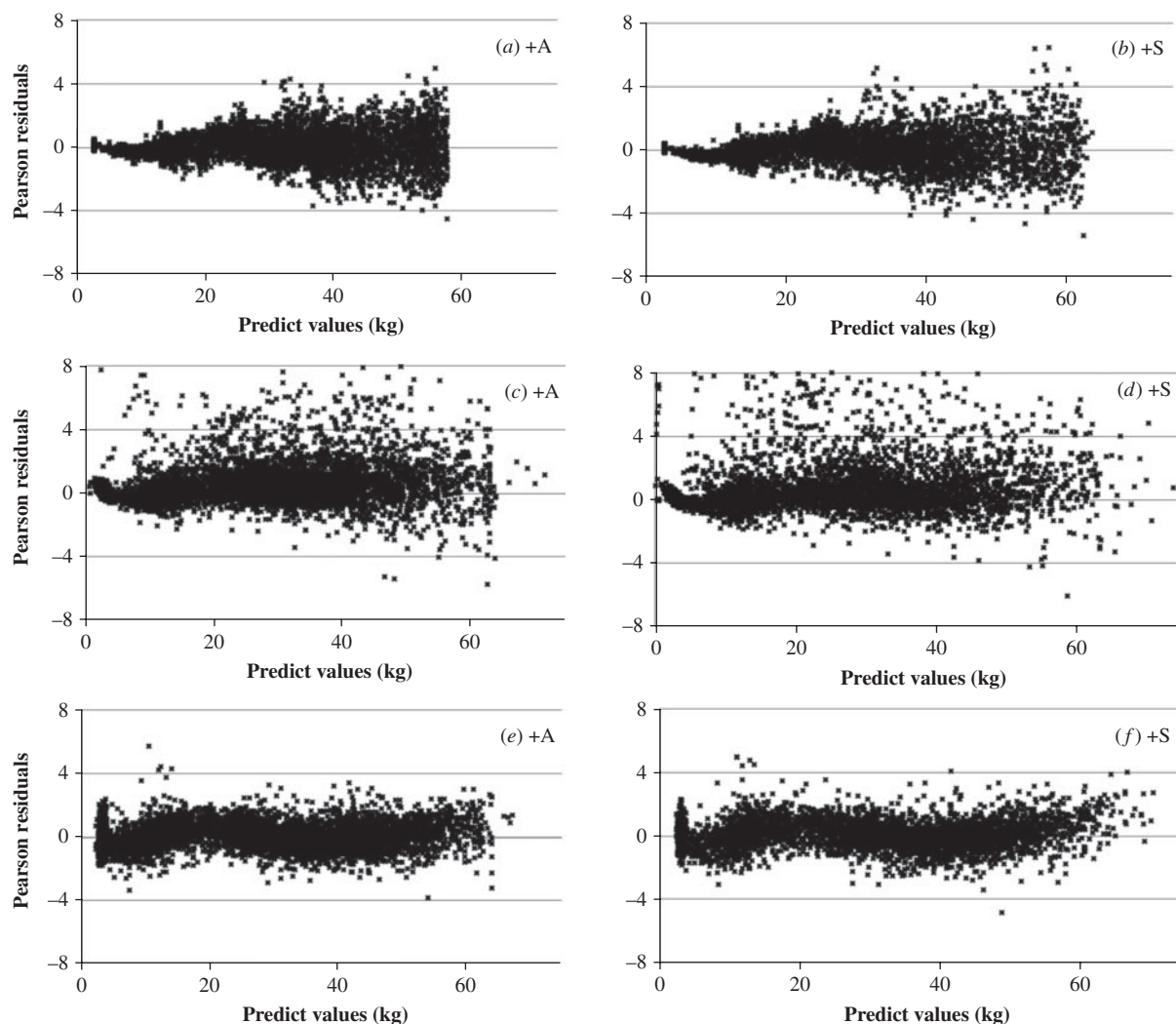† Singular Hessian matrix.
NS: Not significant.

Fig. 1. Pearson residuals of the Richards functions *v.* predicted values (kg) using the following approaches: (*a, b*) Fixed-effects model; (*c, d*) random-effects model (u_1 and u_3); (*e, f*) random (u_1 and u_3) + VP + SP(MATERN).

Using the SP(MATERN) error structure, there were differences between estimated parameters β_1 ($P<0.001$), β_2 ($P=0.003$) and β_3 ($P=0.003$) of the genotypes evaluated; however, there was no difference in the β_4 parameter (Table 5). Therefore, the model was reparameterized to account for just one β_4 parameter. The final parameters adjusted for each genotype were as follows: $\beta_1=52.8$ (s.e. = 0.46), $\beta_2=0.94$ (s.e. = 0.002) and $\beta_3=2.83$ (s.e. = 0.066) for the +A animals, and $\beta_1=56.0$ (s.e. = 0.59), $\beta_2=0.95$ (s.e. = 0.002) and $\beta_3=2.63$ (s.e. = 0.079) for the +S animals. The common β_4 parameter was 0.96 (s.e. = 0.013). The variance components were $\sigma_{u1}^2=47.5$, $\sigma_{u3}^2=0.46$ and $\sigma_{u1u3}=-2.0$ ($P<0.001$) (Table 6). The growth curves are shown in Fig. 2.

The IP was determined as a function of β_4 (Richards 1959). This parameter did not differ between the

evaluated genotypes; the IP at the same proportions of asymptotic body weight for both genotypes (0·36 of β_1) yielded 19·1 and 20·2 kg for genotypes +A and +S, respectively.

Figure 1 shows the Pearson residuals for the Richards model plotted against the predicted values (kg) by genotype for the Fixed model (Figs 1(*a*) and (*b*)), the Random model (u_1 and u_3) (Figs 1(*c*) and (*d*)) and the Random (u_1 and u_3) + VP + SP(MATERN) model (Figs 1(*e*) and (*f*)). When the function was fitted using the traditional approach, heteroscedasticity and the correlated errors occurred. In contrast, when random effects were added to the model, there was an improvement in the quality of the fitting; however, there were high Pearson residuals in this case, and correlated errors are easy to identify by the shape of the plotted residuals. This increase in residuals occurred

Table 5. *Reparameterized Richard model parameters with error structure*

Model	+A				+S			
	β_1	β_2	$\beta_{3 \ (x \ 1000)}$	β_4	β_1	β_2	$\beta_{3 \ (x \ 1000)}$	β_4
SP(POW)	52·9	0·95	2·7	0·90	57·3	0·97	2·3	0·86
SP(MATERN)	52·5	0·94	3·0	0·98	56·5	0·95	2·5	0·94

	σ^2_{u1}	σ^2_{u3}	σ_{u1u3}	σ^2_e	AIC	Δ_r
SP(POW)	71·8 ($P<0·001$)	0·66 ($P<0·001$)	$-3·9$ ($P<0·001$)	0·072	50 740	1126
SP(MATERN)	47·6 ($P<0·001$)	0·45 ($P<0·001$)	$-2·0$ ($P<0·001$)	0·099	49 613	0

Hypothesis test SP(MATERN)	D.F.	T value	P-value
$\beta_{1+S} = \beta_{1+A}$	7683	4·9	<0·001
$\beta_{2+S} = \beta_{2+A}$	7683	3·0	0·003
$\beta_{3+S} = \beta_{3+A}$	7683	$-3·0$	0·003
$\beta_{4+S} = \beta_{4+A}$	7683	$-1·6$	0·143

Table 6. *Reparameterized Richard model with error structure and just one β_4 parameter*

Model	+A			+S			
	β_1	β_2	$\beta_{3 \ (x \ 1000)}$	β_1	β_2	$\beta_{3 \ (x \ 1000)}$	B_4*
SP(MATERN)	52·8	0·94	2·8	56·0	0·95	2·6	0·96

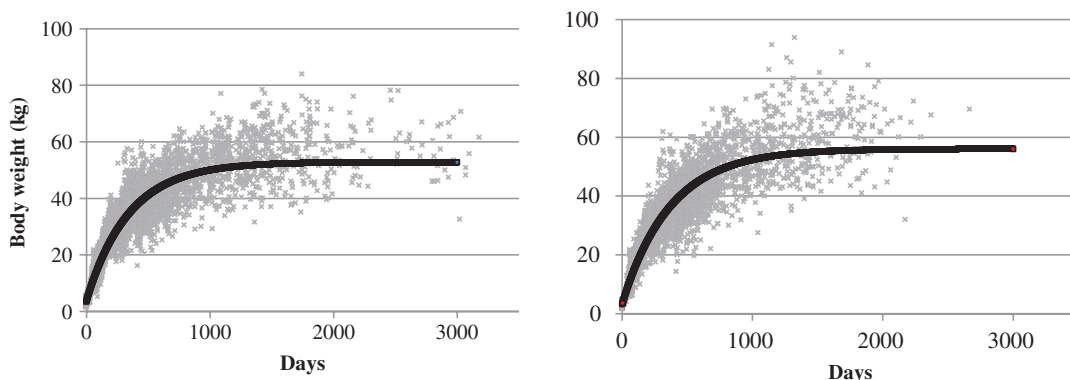|	σ^2_{u1}	σ^2_{u3}	σ_{u1u3}	σ^2_e	AIC
SP(MATERN)	47·8 ($P<0·001$)	0·46 ($P<0·001$)	$-2·0$ ($P<0·001$)	0·099	49 609

* Common β_4 parameter.

Fig. 2. Adjusted growth curve for the two genotypes evaluated.

once there were two common parameters (β_2 and β_4) in the functions, even when the errors were assumed to be independent. The variance power and error structure modelling yielded a better fit, and the Pearson residuals were within an acceptable range.

The Richards (u_1 and u_3) + VP + SP(MATERN) modelling approach was used to fit the training data set in the cross-validation analysis. The parameters fitted (data not shown) did not differ from the parameters fitted from the original data set, indicating that the chosen approach was effective even when using a reduced data set. When using the test data set to evaluate the model, the simultaneous F-test for the intercept ($1·18 \pm 0·14$) and slope ($0·98 \pm 0·0045$)
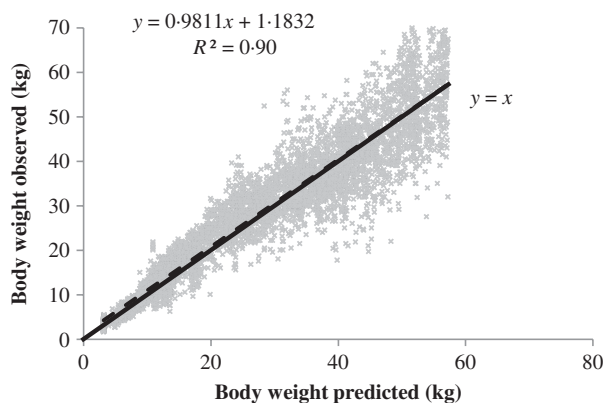
Fig. 3. Linear regression between observed and predicted body weights (dashed line). The continuous line is the $Y=X$ line.

rejected H_0 ($R^2=0.90$). The CCC obtained was 0.95 (ranging from 0 to 1), indicating high accuracy. The RMSEP was 5.48 kg, and its decomposition indicated a high contribution from the random errors (98.37%), whereas the contributions of the mean bias (1.47%) and systematic bias (0.32%) were negligible. Figure 3 shows the relationship between the observed and predicted body weight values from the cross-validation.

Fig. 4 shows that there was no apparent effect on the random parameter estimates from the number of weight measurements.

DISCUSSION

Growth curves and random effects

The nonlinear mixed model has as its main feature the partitioning of error variation into within- and between-subject variations. The fixed effects β_1, β_2, β_3 and β_4 represent the mean values of the parameters in the population of individuals. The individual deviations are represented by the random effects u_1, u_2 and u_3, which are assumed to be distributed normally with mean 0 and unstructured covariance matrix **G**.

For models with only the variance component linked to asymptotic body weight, the error variation was partitioned into variation within goats (σ_e^2) and the variation between goats (σ_{u1}^2). When two variance components were added to the variation between goats, the error variation was partitioned into variation due to asymptotic body weight, the sigmoidal-shaped curve of each animal and the covariance between

these terms ($\sigma_{u1}^2, \sigma_{u2}^2$ and σ_{u1u2}), or due to the asymptotic body weight, rate constant and covariance between these terms ($\sigma_{u1}^2, \sigma_{u3}^2$ and σ_{u1u3}), and so on, when three random effects were added; however, when three variance components were added in the growth curve model, there were indications of overparameterization and more highly correlated parameters, causing problems with convergence and the Hessian matrix.

In consequence, the estimate of the random effects (u_1, u_2 and u_3) of each goat represents the deviation of a determined parameter from its corresponding parameter for its population average. For example, the mean asymptotic body weight of the population of goats with +A genotype is 52.7 kg (Table 6), and the specific estimate of random effects for a specific goat is 5.9 kg (empirical best linear unbiased prediction – EBLUP). Therefore, the mean asymptotic body weight estimated for this specific animal is 58.6 kg, independent of short-term fluctuations in weight due to extraneous environmental effects such as climate and food supply as well as lactation and pregnancy. Thus, the mean estimated parameters are not greatly affected by these types of weight fluctuations. In fact, these parameters represent the mean value of the growth pattern of the goats on the production system.

It is possible that including biologically significant parameters with random variables in the model would reduce the residual variation, given that this tool reduces variability due to the parameters that have probability distributions and that can vary greatly from animal to animal, such as asymptotic body weight or rate constant. The random effect estimate measures the difference between the value assigned to each individual and the average population value (Aggrey 2009).

The high values observed for σ_{u1}^2 in most of the estimated models indicate the considerable contribution of this parameter in the residual variance component when this random variable is not added to the general model. Thus, it is important to incorporate this effect when using this methodology to estimate growth curves.

The σ_{u3}^2 component was almost always found to be significant. In addition, adding the random effects linked to the rate constant improved the estimated model fit, a fact that was validated by better selection criteria (residual variance and AIC) when including this component. Moreover, the significance of this parameter indicates that the rate constant varies among the goat population and is thus subject to breeding programme selection.
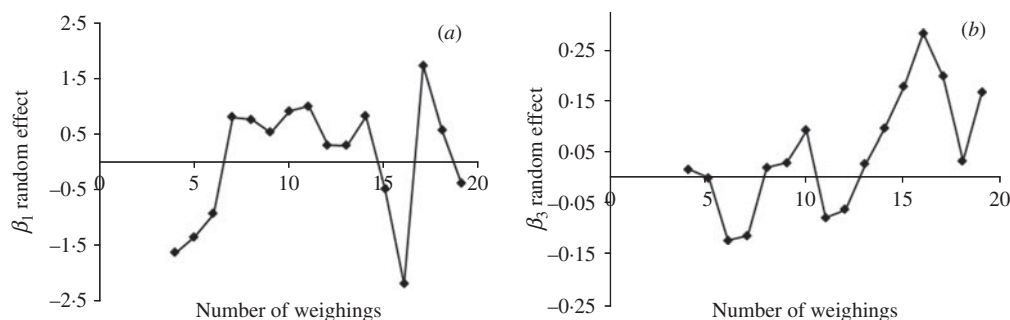
Fig. 4. Effect of the number of weight measurements on the random effects for the (a) u_1 random parameter and (b) u_3 random parameter.

A major advantage of selecting a nonlinear mixed model methodology to fit livestock growth curves is the possibility of including population variation measurements in stochastic models to predict animal performance. Normally, models predicting animal performance are deterministic, static and empirical, and for the same *input*, there is only one *output* that is modelled from the averages, with little or no emphasis on population variation (Pomar *et al.* 2003). Therefore, when small samples are used, such as in the case of small farms, the errors become greater, especially when the model parameters vary greatly, as is the case for asymptotic body weight.

Understanding the variation among animals for the more widely variable parameters of the livestock system (dry matter intake, growth parameters, nutrient use efficiency, etc.) using stochastic components can be essential for understanding the mechanisms involved in population response to certain conditions (Pomar *et al.* 2003). Linking these variations to dynamic models, for which the phenomena are understood over time, and to mechanistic models, for which the biological principles are understood and explained, can make the predictions of such models more credible.

The significant differences in the asymptotic body weight and constant rate parameters of the data are indicative of differences in the growth pattern between the genotypes. Alpine goats probably achieve their mature weight earlier than Saanen goats do, as indicated by the higher rate constant of growth and lower asymptotic body weight estimated for this genotype. This knowledge can be used by farmers to choose between genotypes depending on their interest. In addition, the different asymptotic body weights might be useful in mechanistic models, such as the Small Ruminant Nutrition Systems (SRNS) (Tedeschi

et al. 2010), which makes use of this parameter in its set of equations to predict the degree of maturity for goats. This information is already available for cattle and some sheep genotypes; however, the data set pertaining to goats needs to be expanded.

The Von Bertalanffy, Logistic and Gompertz models have a fixed IP relative to asymptotic body weight, which limits the biological interpretation of these functions due to the lack of flexibility in the estimation of the trend in the instantaneous absolute growth rate. In the case of the Richards model, the IP is variable and is a function of the β_4 parameter. The best model had common parameter $\beta_4 = 0.9642$ [$IC_{95\%}$ 0.9394–0.9890], which yields a transitional function in form between the Brody ($\beta_4 = 0$) and the Gompertz ($\beta_4 = 1$) functions (Richards 1959).

The flexibility of the β_4 parameter is the most important advantage of the Richards function. The IP in goats (0.36 of asymptotic body weight) may differ from the fixed IP adopted by the other functions. The IP occurs where the estimated instantaneous absolute growth rate changes from an increasing to a decreasing function. This information might be useful for strategic plans for goat feeding; information regarding the timing for the greatest capacity of the animal growth could be critical for plan optimization. Moreover, it is expected that from this point (*c.* 20 kg) onward, there is a decrease in growth rate, indicating a necessity for a change in feeding strategy.

Despite the fact that the genotypes have the same origin and are often considered similar in terms of productivity (only visibly different in their coats), differences have been found in their lactation curves (Guimarães *et al.* 2006) and now, in the present study, in their growth patterns. This result indicates the need to use separate models (growth, lactation, etc.) for these two genotypes to predict animal response.
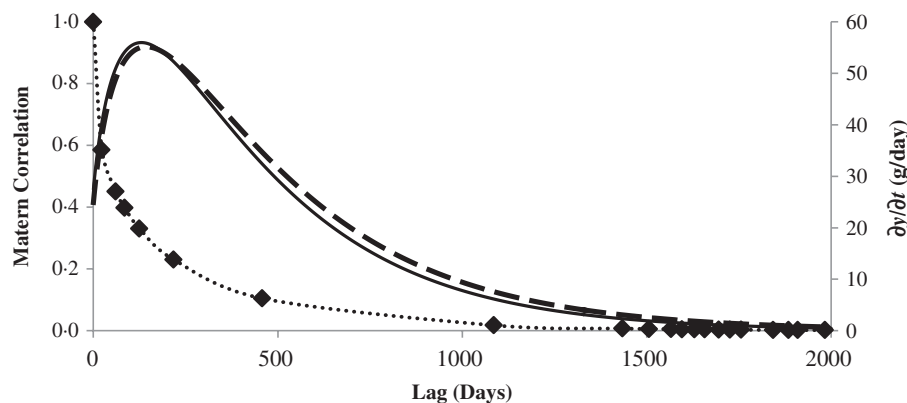
Fig. 5. The left axis shows the MATERN correlation function *v*. Lag (days) (point line). Black diamonds (◆) correspond to the values for animals with 21 weight measurements during their lifetime. The right axis shows the instantaneous absolute growth rate (g/day) for +S animals (continuous line) and +A animals (dashed line).

Modelling of the error structure

As noted by Littell *et al*. (2006), the first tool to be chosen when using the SAS software to fit a nonlinear mixed model is PROC NLMIXED, as it is more general because it accepts other distributions for the dependent variable; however, when it is necessary to model the error structure, the SAS macro %NLIMIX becomes more useful. Additional differences between the two approaches with relation to the estimation method can be found in Littell *et al*. (2006) and Vonesh (2012).

The SP(MATERN) structure has been historically used in spatial studies and has two parameters: $\rho = 411 \cdot 02$ (s.e. $= 52 \cdot 27$) and $V = 0 \cdot 1471$ (s.e. $= 0 \cdot 006$) ($P < 0 \cdot 001$). The lower value found for the V parameter implies that the correlation decreases abruptly between measurements. To illustrate this behaviour, the correlation was plotted between the measurements made on the animal with the maximum number of weight values (21 data points, with the first measure made at birth and the last at 1984 days old) (Fig. 5). The correlation decreases more rapidly over time, so that the correlation between two measurements with an interval of 23 days was 0·59. However, this finding is an indication that the correlation between two measurements was relatively well controlled.

The possible reason that the SP(MATERN) structure adjusted better to this data set is the difference between the measurements found here. The current data included 1–1190 days of difference between two consecutive measurements (Lag). Obviously, two measurements taken at a close interval are typically more highly correlated than measurements taken at more distant time points (Littell *et al*. 2006). In the current data set, 0·61 of the measurements were

made before the animals were 1 year old, which means that the data are temporally close; the IP calculated in the Richards models occurs at 19·05 and 20·22 kg for the +A and +S genotypes, respectively, which occurs at *c*. 131 days old. Measurements taken close to these points will show a lower correlation than measurements taken far from them, because the higher instantaneous absolute growth rate observed around these inflection points. Figure 5 shows this relationship.

Cross-validation

Despite the fact that the simultaneous *F*-test for the intercept and slope rejected the null hypothesis, the high coefficient of determination indicates that the approach chosen is robust enough to estimate the body weight of goats. Furthermore, the RMSEP presented a relatively low value (5·48 kg) with a great contribution from random errors in the total error prediction. This finding indicates that effects that are not controlled are the main factors affecting the predictions. As shown in Fig. 3, the predictions are more credible in the early phases of growth, at lower body weights. Once there is a small variation around the average prediction, an increase in body weight leads to an occurrence of factors that are not controlled and operate to expand the variation around the estimated body weight; however, the average body weight estimated remains without bias, as indicated by the almost total overlap between the dashed line and the continuous lines.

Historically, the growth curve parameters for goats have not been evaluated using adequate statistical

approaches, which may have created biases for the estimated values. The equations presented in the current paper are expected to yield more credible predictions for estimated mean values and population variations for some parameters.

Effects of number of weight values on the random effect estimates

Nonlinear mixed models are a strong tool for modelling and understanding population variability by incorporating random effects to account for between-subject variations. Apparently, the lack of influence of the numbers of weight measurements on random effects is another advantage of this method. Thus, this approach can be used to select animals that are different from the population average with regard to their rate of growth and those with asymptotic body weight.

CONCLUSION

The Richards model is adequate for describing the growth curve of dairy goats. The nonlinear mixed model methodology is an efficient tool to address growth curve features such as heteroscedasticity and correlated errors. Assigning biologically significant parameters with random effects is suggested, thus reducing the residual variance and making the estimates more credible. Moreover, evaluation of the modelling structure error is recommended to account for possible errors correlated even using random effects. Different Richard growth curve parameters should be used for the predominantly Alpine and Saanen genotypes because there are differences in their growth patterns.

REFERENCES

Aggrey, S. E. (2009). Logistic nonlinear mixed effects model for estimating growth parameters. *Poultry Science* **88**, 276–280.

Beal, S. L. & Sheiner, L. B. (1982). Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering* **8**, 195–222.

Brown, J. E., Fitzhugh, H. A. Jr & Cartwright, T. C. (1976). A comparison of nonlinear models for describing weight-age relationships in cattle. *Journal of Animal Science* **42**, 810–818.

Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.

Craig, B. A. & Schinckel, A. P. (2001). Nonlinear mixed effects models for swine growth. *Professional Animal Scientist* **17**, 256–260.

Forni, S., Piles, M., Blasco, A., Varona, L., Oliveira, H. N., Lobo, R. B. & Albuquerque, L. G. (2009). Comparison of different nonlinear functions to describe Nelore cattle growth. *Journal of Animal Science* **87**, 496–506.

Guimarães, V. P., Rodrigues, M. T., Sarmento, J. L. R. & Rocha, D. T. D. (2006). Utilização de funções matemáticas no estudo da curva de lactação em caprinos. *Revista Brasileira de Zootecnia* **35**, 535–543.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. & Schabenberger, O. (2006). *SAS for Mixed Models*, 2nd edn. Cary, NC: SAS Institute Inc.

Meng, S. X. & Huang, S. (2010). Incorporating correlated error structures into mixed forest growth models: prediction and inference implications. *Canadian Journal of Forest Research* **40**, 977–990.

Minasny, B. & McBratney, A. B. (2007). Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma* **140**, 324–336.

Peek, M., Russek-Cohen, E., Wait, D. A. & Forseth, I. N. (2002). Physiological response curve analysis using nonlinear mixed models. *Oecologia* **132**, 175–180.

Pinheiro, J. C. & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**, 12–35.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus*. New York: Springer.

Pomar, C., Kyriazakis, I., Emmans, G. C. & Knap, P. W. (2003). Modeling stochasticity: Dealing with populations rather than individual pigs. *Journal of Animal Science* **81** (Suppl. 2), E178–E186.

Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany* **10**, 290–301.

SAS Institute Inc (2008). *SAS/STAT(r) 9.2 User's Guide*. Cary, NC, USA: SAS Institute Inc.

Strathe, A. B., Danfaer, A., Sorensen, H. & Kebreab, E. (2010). A multilevel nonlinear mixed-effects approach to model growth in pigs. *Journal of Animal Science* **88**, 638–649.

Tedeschi, L. O. (2006). Assessment of the adequacy of mathematical models. *Agricultural Systems* **89**, 225–247.

Tedeschi, L. O., Cannas, A. & Fox, D. G. (2010). A nutrition mathematical model to account for dietary supply and requirements of energy and other nutrients for domesticated small ruminants: The development and evaluation of the Small Ruminant Nutrition System. *Small Ruminant Research* **89**, 174–184.

Vieira, R. A. M., Campos, P. R. D. S. S., da Silva, J. F. C., Tedeschi, L. O. & Tamy, W. P. (2012). Heterogeneity of the digestible insoluble fiber of selected forages *in situ*. *Animal Feed Science and Technology* **171**, 154–166.

Vonesh, E. F. (2012). *Generalized Linear and Nonlinear Models for Correlated Data: Theory and Applications Using SAS*. Cary, NC, USA: SAS Institute Inc.

Wang, Z. & Zuidhof, M. (2004). Estimation of growth parameters using a nonlinear mixed Gompertz model. *Poultry Science* **83**, 847–852.

Yang, Y. & Huang, S. (2011). Comparison of different methods for fitting nonlinear mixed forest models and for making predictions. *Canadian Journal of Forest Research* **41**, 1671–1686.

APPENDICES

SAS code:

```
    Fixed model
    PROC NLMIXED; /*Gaussian quadrature algor-
ithm*/
    /* Initial parameter estimates from PROC NLIN to
start iterations;*/
    parms b01 = 56 b11 = 59
    b02 = .9 b12 = .97
    b03 = 1.2 b13 = 1.5
    b04 = .7 b14 = .8
    s2e = 20;
    /*coding dummy variables for the goat genotypes*/
    b1 = b01*z1 + b11*z2;
    b2 = b02*z1 + b12*z2;
    b3 = b03*z1 + b13*z2;
    b4 = b04*z1 + b14*z2;
    /*nonlinear Richards model*/
    pred = b1*(1 − b2*exp((− b3/1000)*x))**b4;
    model y ~ normal (pred,s2e);
    run;
    Random model
    PROC NLMIXED method = firo; /*FIRO algorithm*/
    parms b01 = 52 b11 = 53
    b02 = .9 b12 = .97
    b03 = 1.9 b13 = 2
    b04 = .7 b14 = .8
    s2u1 = 30 c13 = .99 s2u3 = 1.5 s2e = 10;
    b1 = b01*z1 + b11*z2+u1; /*variance components
u1 and u3 added*/
    b2 = b02*z1 + b12*z2;
    b3 = b03*z1 + b13*z2+u3;
    b4 = b04*z1 + b14*z2;
    pred = b1*(1 − b2*exp((− b3/1000)*x))**b4;
    model y ~ normal (pred,s2e) ;
```

```
    /*Specifying random parameters with normal
distribution */
    random u1 u3 ~ normal ([0,0],[s2u1,c13,s2u3])
subject = animal ;
    run;
    Random model + Variance Power + Error Structure
Model
    %nlinmix(data = cresc, /*Specifies the SAS data set*/
    /*Coding dummy variables for the goat genotypes
and request the nonlinear Richards model*/
    model = %str(
    b1 = b01*z1 + b11*z2+u1;
    b2 = b02*z1 + b12*z2;
    b3 = b03*z1 + b13*z2+u3;
    predv = b1*(1-b2*exp((-b3/1000)*x))**b4;
    ),
    /*Initial parameters estimated*/
    parms = %str(b01 = 53 b11 = 57 b02 = 0.9 b12 = 0.97
b03 = 2.7 b13 = 2.2 b4 = 0.9),
    /*Modelling the variance power statement*/
    derivs = %str(
    wt = 1/predv**(2.91*0.5);
    ),
    /*Specifies the MIXED procedure to be executed for
each iteration*/
    stmts = %str(
    class animal;
    model pseudo_y = d_b01 d_b11 d_b02 d_b12
d_b03 d_b13 d_b4 / noint notest solution cl residual
outp = pred_matern;
    /*Specifies the random parameters and matrix G*/
    random d_u1 d_u3 / subject = animal type = un
solution;
    /*Specifies the matrix R*/
    repeated / subject = animal type = sp(matern)(x);
    weight wt;
    ),
    /*Request the expansion method*/
    expand = zero,
    /*Request the REML method and test the variance
components parameters*/
    procopt = %str(method = reml) covtest
    )
    run;
```