STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **12**, March 2024, pp 298–309.
Published online in International Academic Press (www.IAPress.org)

Stochastic Differential Equations Mixed Model for Individual Growth with the Inclusion of Genetic Characteristics

Nelson T. Jamba [1,*], Patrícia A. Filipe [3], Gonçalo Jacinto [2], Carlos A. Braumann [2]

[1] *Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora, Portugal and Liceu nº 918 do município dos Gambos, Chiange, Gambos, Angola; d39830@alunos.uevora.pt*

[2] *Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora and Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora, Évora, Portugal; gjcj@uevora.pt, braumann@uevora.pt*

[3] *Iscte-Instituto Universitário de Lisboa, Iscte Business School, Quantitative Methods for Management and Economics Department, Lisboa, Portugal and Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora, Évora, Portugal; patricia.filipe@iscte-iul.pt*

Abstract In early work we have studied a class of stochastic differential equation (SDE) models, for which the Gompertz and the Bertalanffy-Richards stochastic models are particular cases, to describe individual growth in random environments, and applied it to model cattle weight evolution using real data. We have started to work on these type of models considering that the model parameters are fixed, i.e. the same for all the animals. Aiming to incorporate variability among individuals, we consider that the model parameters can be random variables, resulting in SDE mixed models. In additon, here we consider SDE mixed models, allowing the parameters to be random and propose to incorporate each animal's genetic characteristics considering the transformed animal's weight at maturity to be a function of its genetic values. The main objective is to extend the SDE mixed model to the more realistic case where the individual's genetic value becomes an important component in the estimated growth curve. For the estimation of the model parameters we have used maximum likelihood estimation theory. Estimates and asymptotic confidence intervals of the parameters are presented. A comparison with SDE non-mixed model and SDE mixed model without the inclusion of genetic characteristics is shown with the conclusion that the incorporation of some genetic characteristics in the model parameters improves the estimation of the animal's growth curve.

Keywords Genetic traits, Individual growth, Mixed models, Stochastic differential equations

Mathematics Subject Classification: 60H10, 60E05, 65D15, 62F10, 92D99

DOI: 10.19139/soic-2310-5070-1829

1. Introduction

In the last four decades, Stochastic Differential Equations (SDE) have been studied and applied to individual growth modeling by several authors using SDE non-mixed models [9, 10, 11, 12, 13, 19, 20, 22, 23] where, for instance, authors have modeled the growth dynamic of animals, trees or glucose, and where the parameters of the SDE have fixed, although unknown, values that need to be estimated from data. When the model parameters are considered random, SDE mixed models are used to model individual growth, with applications in various fields, such as animal growth and pharmacokinetics, among others [5, 6, 7, 14, 17, 18, 21, 25].

Considering M animals, we have used the following general SDE model

*Correspondence to: Nelson T. Jamba (Email: d39830@alunos.uevora.pt). Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora, Portugal and Liceu nº 918 do município dos Gambos, Angola.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2024 International Academic Press

$$dY_i(t) = \beta\,(\alpha - Y_i(t))\;dt + \sigma\,dW_i(t), \quad Y_i(t_{i,0}) = y_{i,0}, \quad i = 1, \ldots, M, \tag{1}$$

where $Y_i(t) = h(X_i(t))$ is the modified size by the transformation h, a monotonous continuously differentiable function (which we assume known) of the real size $X_i(t)$ at age t of the i^{th} individual ($i = 1, \ldots, M$). We have $Y_i(t_{i,0}) = y_{i,0} = h(x_{i,0})$, where $x_{i,0}$ is the size observed at age $t_{i,0}$ (initial observation) for individual i, and $\alpha = h(A)$, where A is the asymptotic size or size at maturity. The growth parameter β is the rate of approach to maturity, σ measures the strength of environmental fluctuations on growth, and $W_i(t)$ ($i = 1, \ldots, M$) are independent standard Wiener processes. Specific choices of the function h lead to stochastic versions of some models commonly used to describe growth, such as the monomolecular model for $h(x) = x$, the Bertalanffy-Richards model for $h(x) = x^c$ ($c > 0$), the Gompertz model for $h(x) = \ln x$ and the logistic model for $h(x) = x^{-1}$ [9, 11]. We refer to model (1) as SDE non-mixed model, when the parameters α, β and σ are fixed values common to all individuals. However, it is natural to think that different individuals might have different values for these parameters, due to their individual characteristics. To take this into account we have studied SDE mixed models, where in model (1) we have considered that the parameters α and/or β were random variables [14, 15]. The maximum likelihood methods are the commonly used methods to estimate the model parameters. Unlike the SDE non-mixed models, where it is usually possible to obtain a closed-form expression for the likelihood function, that is not always the case when SDE mixed models are used. Recently we have made some developments in estimation methods for SDE mixed models, when trend parameters can be considered random, using approximation methods to solve estimation difficulties in SDE mixed models. In our recent work, we have introduced the Delta approximation method (DA) [14], a new approach inspired by the classical statistical delta method, and evaluated its performance with existing approximation methods, such as the Laplace approximation method (LA) [21, 26] or other approximation methods available in *MsdeParEst* and *mixedsde* R packages [4, 16]. The DA method has an advantage over the available R packages by not requiring a common time vector of observations for all individuals.

The inclusion of genetic values in animal growth allows breeders to develop targeted breeding objectives and selection strategies. By prioritizing growth-related traits based on their estimated genetic values, breeders can focus on enhancing specific aspects of animal growth. This approach ensures that breeding efforts are directed towards maximizing growth potential, efficiency, and overall productivity. In this work, we will include a genetic value of the animals in the mixed model, allowing that each animal has a different growth curve influenced by its genetic information. As far as we are aware, letting each animals's growth curve be influenced by their individual characteristics, in particular the animal's genetic information, is a new approach for this type of models. Here, we will consider the case where the asymptotic weight at maturity is random, allowing us to study the influence of genetic values on this parameter of the SDE mixed model, particularly whether they provide a significant contribution to the precision of the growth curve. Comparing the parameters estimation using the SDE non-mixed model (parameters are fixed values, common to all individuals), the SDE mixed model without the inclusion of genetic values (parameters, in this case α, vary randomly from animal to animal) and the SDE mixed model with inclusion of the genetic values (parameter α is considered random and writen as a linear function of the animal's genetic values), we can evaluate if a random effect on the asymptotic weight at maturity is needed and, in that case, if the animal's specific genetic information would improve the parameters estimation. A more precise parameters estimation, and as a consequence, a more precise growth curve, will improve the farmers profit in livestock production.

We illustrate the results using real cattle weight (in kg) data of 10843 animals provided by the *Associação de Criadores de Bovinos Mertolengos* (ACBM, https://www.mertolenga.com), which performs the growing and finishing phases of young Mertolengo breed males, and by associated breeders, whose agricultural holdings are located in the Alentejo region in Portugal. Each animal has weight measurements taken at different age instants, varying from animal to animal, contrary to what is usually assumed in estimation methodologies.

The data regarding the genetic information of the animals results from the genetic evaluation of the Mertolengo cattle breed that was carried out at the *National Institute of Agricultural and Veterinary Research* (INIAV - https://www.iniav.pt), a research center in the veterinary and agronomic field, based in Portugal. Seven genetic traits were considered: growth capacity, maternal capacity, calving interval, average daily weight gain,

carcass weight per day of age, feed conversion index and reproductive longevity. Seven models were fitted, each including a different genetic trait. We then evaluate which genetic trait results in an increasing performance on the model parameter estimates of the individuals.

This paper is organized as follows, in section 2, we begin by describing the SDE mixed models with random α and discuss their main properties. Then, in section 3, we extend the SDE mixed models to incorporate the genetic values. In section 4, we apply our model to real cattle weight data and compare its performance with other SDE models. Finally, in section 5, we summarize the main findings and conclusions of our study.

2. Stochastic differential equations models

Let us consider M individuals, where $X_i(t)$ denotes the size of the i^{th} individual $(i = 1, \ldots, M)$ at age t. In our application, the size refers to the animal's weight, but in other cases, it can denote another growth measure of the individual. We will consider a modified size $Y_i(t) = h(X_i(t))$, $(i = 1, \cdots, M)$, and describe the evolution of individual growth through an SDE of the form (1). For the real cattle weight data we used here for illustration, the transformation $h(x) = \ln x$, that leads to the Stochastic Gompertz model revealed to be the most appropriate to our type of data [8]. The solution of the autonomous SDE (1) is an ergodic diffusion process with drift and diffusion coefficients $\beta\,(\alpha - y_i)$ and σ^2, respectively. These coefficients have bounded continuous derivatives and therefore they satisfy a Lipschitz condition which guarantees the existence and uniqueness of the solution (see, for instance, [1]), given by

$$Y_i(t) = \alpha + e^{-\beta(t-t_{i,0})}(y_{i,0} - \alpha) + \sigma e^{-\beta t}\int_{t_{i,0}}^{t} e^{\beta s} dW(s), \tag{2}$$

which follows a Gaussian distribution with mean $\alpha + e^{-\beta(t-t_{i,0})}(y_{i,0} - \alpha)$ and variance $\frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-t_{i,0})})$ (see, for instance, [1, 9]).

Let $t_{i,j}$ $(i = 1, ..., M,\ j = 1, ..., n_i)$ be the age of the j^{th} weight measurement of the i^{th} animal and let $Y_{i,j} = Y_i(t_{i,j}) = h(X_i(t_{i,j}))$ be the corresponding modified weight according to model (1). For each animal i $(i = 1, \ldots, M)$, denote by $\mathbf{t}_i = (t_{i,0}, t_{i,1}, \ldots, t_{i,n_i})$ its age vector of observations (which may differ from individual to individual), by $\mathbf{Y}_i = (Y_{i,0}, Y_{i,1}, ..., Y_{i,n_i})$ the corresponding vector of modified sizes and by $\mathbf{y}_i = (y_{i,0}, y_{i,1}, ..., y_{i,n_i})$ the observed value of $\mathbf{Y}_i$. We assume $t_{i,j-1} < t_{i,j}$ and make $E_{i,j} = e^{-(t_{i,j}-t_{i,j-1})}$. We see that for $Y_{i,j}$ conditioned on $Y_{i,j-1} = y_{i,j-1}$, the transition distribution for animal i is Gaussian,

$$Y_{i,j} | \left(Y_{i,j-1} = y_{i,j-1}\right) \sim N\left(\alpha + (y_{i,j-1} - \alpha)\, E_{i,j}^{\beta},\ \frac{\sigma^2}{2\beta}(1 - E_{i,j}^{2\beta})\right), \tag{3}$$

with $i = 1, \cdots, M,\ j = 1, \cdots, n_i$. We have applied the maximum likelihood estimation method, [9, 14, 15], to estimate the parameter vector $\mathbf{p} = (\alpha, \beta, \sigma)$. From (3), using the fact that $Y_i(t)$ is a Markov process, we know that,

given $Y_{i,0} = y_{i,0}$ (assumed known), the $\mathbf{Y_i}$ joint probability density function for individual i takes the form

$$\begin{aligned} p_{\mathbf{Y_i}}(\mathbf{y}_i|\alpha,\beta,\sigma) &= \prod_{j=1}^{n_i} \frac{\exp\left(-\frac{1}{2}\frac{\left(y_{i,j}-\alpha-(y_{i,j-1}-\alpha)E_{i,j}^{\beta}\right)^2}{\frac{\sigma^2}{2\beta}(1-E_{i,j}^{2\beta})}\right)}{\sqrt{2\pi\frac{\sigma^2}{2\beta}(1-E_{i,j}^{2\beta})}} \\ &= \frac{\exp\left(-\frac{1}{2}\sum_{j=1}^{n_i}\frac{\left(y_{i,j}-\alpha-(y_{i,j-1}-\alpha)E_{i,j}^{\beta}\right)^2}{\frac{\sigma^2}{2\beta}(1-E_{i,j}^{2\beta})}\right)}{\prod_{j=1}^{n_i}\sqrt{2\pi\frac{\sigma^2}{2\beta}(1-E_{i,j}^{2\beta})}}, \quad i=1,\ldots,M, \end{aligned} \tag{4}$$

and by independence among individuals we obtain the likelihood function for the M animals

$$L(\alpha,\beta,\sigma) = \prod_{i=1}^{M} p_{\mathbf{Y}_i}(\mathbf{y}_i|\alpha,\beta,\sigma). \tag{5}$$

The maximum likelihood estimate of the parameter vector $\mathbf{p}$ is obtained by maximization of (5) or of the log-likelihood function

$$\begin{aligned} LL_Y(\alpha,\beta,\sigma) &= \ln L(\alpha,\beta,\sigma) = \sum_{i=1}^{M} -\frac{n_i}{2}\ln(2\pi) \\ &\quad -\frac{n_i}{2}\sum_{i=1}^{M}\ln\left(\frac{\sigma^2}{2\beta}\right) - \sum_{i=1}^{M}\frac{1}{2}\sum_{j=1}^{n_i}\ln\left(1-E_{i,j}^{2\beta}\right) \\ &\quad -\frac{\beta}{\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{n_i}\frac{\left(y_{i,j}-\alpha-(y_{i,j-1}-\alpha)\,E_{i,j}^{\beta}\right)^2}{1-E_{i,j}^{2\beta}}. \end{aligned} \tag{6}$$

We have just described the general SDE model (1) for the complete growth curve of the animals where the model's parameters α, β, and σ are assumed common to all individuals.
In [14], we consider a generalization to SDE mixed models to take into account that different animals, due to their individual characteristics, may have different values of the parameters. We have denoted by $\mathbf{b}$ the d-dimensional vector of parameters that vary randomly among animals and assumed that the distribution of $\mathbf{b}$ among animals has probability density function (p.d.f.) $p_B(\mathbf{b}|\mathbf{\Psi})$, where $\mathbf{\Psi}$ is the parameter vector that characterizes this distribution and needs to be estimated. Assuming independence among the animals, the M parameter vectors $\mathbf{b}_i$ of the different animals i ($i = 1,\ldots,M$) are independent identically distributed random vectors with common p.d.f. p_B and we assume that $\mathbf{b}_i$ ($i = 1\ldots,M$) are also independent of the Wiener processes that characterize the environmental conditions under which the animals are growing. Let $\mathbf{\Lambda}$ be the vector of the remaining model parameters (those not involved in p_B), assumed to be common to all animals. The likelihood function for M trajectories (animals) is given by

$$\begin{aligned} L(\mathbf{\Lambda},\mathbf{\Psi}) &= \prod_{i=1}^{M} p_{\mathbf{Y}_i}(\mathbf{y}_i|\mathbf{\Lambda},\mathbf{\Psi}) \\ &= \prod_{i=1}^{M}\int_{\mathbb{R}^d} p_{\mathbf{Y}_i}(\mathbf{y}_i|\mathbf{b}_i,\mathbf{\Lambda})p_B(\mathbf{b}_i|\mathbf{\Psi})\,d\mathbf{b}_i, \end{aligned} \tag{7}$$

The case of the single random parameter $\mathbf{b}_i = (\alpha_i)$ with $\alpha_i \sim N(\mu,\theta^2)$ has already been studied by several authors. In [17] the special situation of an age vector of observations $\mathbf{t_i} \equiv \mathbf{t} = (t_0,t_1,\ldots,t_n)$, $i = 1,\ldots,M$

common to all animals is presented. For the general situation, with no such restrictions, it is possible to explicitly compute the integral in the likelihood function, resulting in a final closed-form expression for this function. This is shown in [14, 15], where the exact log-likelihood function for all animals is given by

$$\begin{aligned} LL_Y(\mu,\theta,\beta,\sigma) = \sum_{i=1}^{M}\Bigg(&-\frac{n_i}{2}\ln(2\pi) - \frac{n_i}{2}\ln\left(\frac{\sigma^2}{2\beta}\right) - \frac{1}{2}\sum_{j=1}^{n_i}\ln\left(1-E_{i,j}^{2\beta}\right) - \frac{1}{2}\ln(D_i) \\ &- \tfrac{\beta}{D_i\sigma^2}\sum_{j=1}^{n_i}\frac{\left(y_{i,j}-\mu-(y_{i,j-1}-\mu)\,E_{i,j}^{\beta}\right)^2}{1-E_{i,j}^{2\beta}} + \frac{2\beta^2\theta^2}{D_i\sigma^4}\left(\sum_{j=1}^{n_i}\frac{\left(y_{i,j}-y_{i,j-1}E_{i,j}^{\beta}\right)}{1+E_{i,j}^{\beta}}\right)^2 \\ &- \tfrac{\beta(D_i-1)}{D_i\sigma^2}\left(\sum_{j=1}^{n_i}\frac{\left(y_{i,j}-y_{i,j-1}E_{i,j}^{\beta}\right)^2}{1-E_{i,j}^{2\beta}}\right)\Bigg), \end{aligned} \tag{8}$$

with $D_i = \frac{2\beta\theta^2}{\sigma^2}\sum_{j=1}^{n_i}\frac{1-E_{i,j}^{\beta}}{1+E_{i,j}^{\beta}} + 1$.

For the random β case, $\mathbf{b}_i = (\beta_i)$, it is not possible to obtain a closed-form expression for the likelihood function, and we need to solve the integral that appears in the likelihood function (7) through approximation methods. In [14] we proposed a method to approximate the likelihood function when both α and β are random variables, the delta method. This approach gives simple and closed-form approximation formulas, making it easy to apply. This method also has the advantage of not being restricted to the cases where the age vector of the observations is the same for all trajectories nor does it require equidistant ages of observation, assumptions quite common in most estimation techniques. In [15] the delta method is also used but for the particular cases where just one of the parameters α or β is considered random, and a comparison of our new proposed method with the known Laplace approximation method is performed.

Here, we consider a generalization of SDE mixed models proposed in [14], by incorporating the genetic value of each animal in the mean value of the transformed random asymptotic weight. We consider that the transformed random asymptotic weight is a linear function of the animal's genetic value, allowing a more realistic model and an improvement of the estimates of the model parameters at the individual level. As far as we are aware, this is the first time that the dependence on the animal's genetic values of the model parameters of an SDE mixed model is considered.

3. Mixed models with genetic values

The growth of an animal may depend of a variety of factors, which can be reduced to the three most important: genetic information, the environment and the food intake regime. We will develop the particular case where the parameter α, which represents the transformed asymptotic maturity weight of the animal, is considered random and can incorporate the animals genetic values. Consider that the genetic characterization of each animal is determined by K genetic values. In the first stage, we are interested in identifying the genetic values that can improve the estimation of the animal's weight at maturity and can lead to a significantly better model. So, for simplicity, we will consider incorporating one genetic value at a time ($K = 1$).

Denoting by g_i the genetic value of animal i, $(i = 1, \ldots, M)$, let us consider α_i to be defined as a linear function of the genetic value in the form of

$$\alpha_i = c_0 + c_1 g_i + \delta_i, \tag{9}$$

where δ_i are independent and identically distributed with $\delta_i \sim \mathcal{N}(0, \sigma_\delta^2)$, $(i = 1, \ldots, M)$. Then α_i follows a Gaussian distribution with mean $c_0 + c_1 g_i$ and variance σ_δ^2.

Using the parameter α_i written as a linear function of the genetic value, we can adapt the model presented in the previous section, but instead of $\alpha_i \sim N(\mu, \theta^2)$, we will have $\alpha_i \sim N(c_0 + c_1 g_i, \sigma_\delta^2), (i = 1, \ldots, M)$. In fact, using the same simple math used to derive expression (8), i.e. replacing in (7) the Gaussian density with mean $\mu_i = c_0 + c_1 g_i$ and standard deviation σ_δ for animal i and solving the resulting integral, we end up obtaining a closed-form expression for the log-likelihood for all animals, similar to the expression (8), which will be given by

$$\begin{aligned} LL_Y(c_0, c_1, \sigma_\delta, \beta, \sigma) = \sum_{i=1}^{M} \Bigg(& -\frac{n_i}{2} \ln(2\pi) - \frac{n_i}{2} \ln\left(\frac{\sigma^2}{2\beta}\right) - \frac{1}{2} \sum_{j=1}^{n_i} \ln\left(1 - E_{i,j}^{2\beta}\right) \\ & - \tfrac{1}{2} \ln(D_i) - \tfrac{\beta}{D_i \sigma^2} \sum_{j=1}^{n_i} \frac{\left(y_{i,j} - (c_0 + c_1 g_i) - (y_{i,j-1} - (c_0 + c_1 g_i))\, E_{i,j}^{\beta}\right)^2}{1 - E_{i,j}^{2\beta}} \\ & + \tfrac{2\beta^2 \sigma_\delta^2}{D_i \sigma^4} \left(\sum_{j=1}^{n_i} \frac{\left(y_{i,j} - y_{i,j-1} E_{i,j}^{\beta}\right)}{1 + E_{i,j}^{\beta}} \right)^2 - \tfrac{\beta (D_i - 1)}{D_i \sigma^2} \left(\sum_{j=1}^{n_i} \frac{\left(y_{i,j} - y_{i,j-1} E_{i,j}^{\beta}\right)^2}{1 - E_{i,j}^{2\beta}} \right) \Bigg), \end{aligned} \tag{10}$$

with $D_i = \frac{2\beta\sigma_\delta^2}{\sigma^2} \sum_{j=1}^{n_i} \frac{1 - E_{i,j}^{\beta}}{1 + E_{i,j}^{\beta}} + 1$, where the models parameters are now c_0, c_1, σ_δ, β and σ.

In order to effectively verify if a genetic value can improve the estimation of the animal's weight, a comparison between the SDE mixed model and the SDE mixed model with the inclusion of the genetic value is made (using a likelihood ratio test, the Bayesian information criterion (BIC) and the Akaike information criterion (AIC)). If the two models show no significant differences, the genetic value does not have a significant effect and we may conclude that c_1 is not significantly different from 0, i.e. the mean value of the parameter α_i is common to all animals and equal to $\mu = c_0$. On the other hand, if a significant difference between models is detected, it reveals that the inclusion of the genetic value has a significant effect on the estimation of the animal's weight at maturity, i.e., c_1 is significantly different from 0 and the mean value of α_i is different for each animal and given as a linear function of the specific genetic value g_i as $\mu_i = c_0 + c_1\, g_i$.

4. Real data application

In this section, we use the maximum likelihood estimation method to estimate the parameters for the stochastic Gompertz model using real cattle weight data, for the case where the parameter α is considered random and a function of the genetic values. We compare the results with the ones obtained when no genetic value is included, and using the likelihood ratio test (LRT) we compare both models. In this way, we infer if the inclusion of the genetic values improves the estimation of the model parameters. Since we have a different set of animals for each genetic value, we also present the AIC and the BIC values to compare different models. The database provided by ACBM has information on the weights and ages of 10843 animals, totaling 69782 observations, with a minimum of 3 weight observations and a maximum of 33 observations.

4.1. The genetic information

Genetic evaluation of the Mertolenga cattle breed was carried out at INIAV, from all the data collected by ACBM, namely, records of genealogies, births, weights and carcass information, taking into account the following characteristics: growth capacity until weaning (GC), maternal capacity until weaning (MC), calving interval (CI), average daily weight gain (WG), carcass weight per day of age (CW), feed conversion index (FC) and reproductive longevity (RL). The estimation of the genetic values of each animal for these seven types of traits, takes into account the animal's performance, if known, and the performances of all its relatives (ascendants, descendants and collateral's), taking into consideration the various environmental effects that affect the respective trait. Currently,

at an international level and in several livestock species (cattle, sheep, pigs poultry, goats, horses, etc.) the use of BLUP - Animal Model for the genetic evaluation is widespread [2, 3, 24].

Not all animals have the same age at the available genetic evaluations. Most of the animals had genetic evaluations taken at least at two different instants: at birth and at entrance on ACBM for the finishing phase. We have compared the estimation results considering, for each animal and each genetic characteristic, the first genetic value available, the genetic value closest to 7 months of age (mean age of entrance in ACBM for the growing and finishing phases), and also the mean of the mother and father genetic values. Comparing the estimates obtained using these three different genetic information, we have concluded that, using the values from the genetic evaluation closest and prior to the time of entrance in ACBM results in the best choice in terms of the likelihood of the stochastic differential equation mixed model. That is also the most coherent decision, since the final objective is to give a tool to ACBM to better estimate the animal's growth and accurately estimate the profit obtained by raising and selling the animal. Using the most recent animal's genetic information available by the time of entrance on ACBM will accomplish better that purpose.

Therefore, in our analysis, we will use for each animal and genetic trait, the animal's available genetic value closest and prior to the mean entrance age at ACBM (approximately 7 months). Not all animals had the seven genetic values available. For that reason, depending on the genetic value used in the model, the number of observations used to estimate the parameters vary.

The genetic value of an animal for a given trait represents the value of that animal as a breeder (expressed in the respective units of measurement, i.e. kg, days, %, etc.) and should be interpreted as the genetic superiority or inferiority for the trait relative to the population average. The accuracy of the genetic value estimate is also available and gives us an idea of how confidently the genetic value of the animal for a given characteristic was estimated. However, it is not an indicator of the genetic potential of the animal. The more information about the animal (e.g., various records of birth intervals) and about their relatives (mother, sisters, daughters, grandparents, etc.) there are, the more accurate the estimate of their breeding value will be.

Regarding the seven genetic values in terms of genetic potential:

- For MC, RL, GC, WG and CW, the higher their values are, preferably positive, the better.
- For CI and FC, the lower their values are, preferably negative, the better.

In Table 1, we present the minimum, the 25^{th} quantile (Q1), the median, the mean, the 75^{th} quantile (Q3), the maximum, and the standard deviation (s.d.), as well as the number of animals with the respective genetic value available (M) and the total number of weight observations available for that group of animals (n). Since, to achieve a high genetic potential, some of the genetic values should be high and positive while others should be small and negative, and all genetic values have a very high range (positive and negative) values, it is clear that some animals would have a greater genetic potential that others.

The first five genetic values presented in Table 1 are the ones that should be as high as possible and positive. The last two genetic values should be as small as possible and negative. At the moment of the evaluation of the genetic values, they are usually centered to obtain an evaluated population with a zero mean and making it easy to determine which animals have higher potential and which have lower potential. In the results presented in Table 1, the mean is not 0 because the animals used in the parameter estimation had their genetic values evaluated at a wide variety of evaluation times and so belong to different evaluated populations. The CW, WG and CI values have higher variability.

4.2. Parameters estimates

In Table 2, we can observe the parameter estimates using the stochastic differential equations non-mixed model (NMSDE) - obtained by maximization of (6) - and the mixed model with random α without the genetic characters (MSDE) - obtained by maximization of (8). We can observe that the estimates of all parameters are quite similar in both models. The approximate 95% confidence bands based on the inverse of the empirical Fisher information matrix in both models are also very similar. We present the maximum values of the log-likelihood in terms of the

Table 1. Descriptive statistics for the seven genetic values of the genetic evaluation: minimum (min), 25^{th} quantile (Q_1), median, mean, 75^{th} quantile (Q_3), maximum (max), standard deviation (s.d.), number of animals (M) and total number of weight observations (n).

Genetic value	min	Q_1	median	mean	Q_3	max	s.d.	M	n
MC	-22.82	-3.45	0.63	0.24	4.56	19.92	6.17	7981	45014
GC	-39.44	-5.07	-0.02	-0.17	5.08	24.56	7.78	5186	34526
RL	-11.66	-1.52	0.99	1.04	3.32	19.88	3.82	4234	21317
WG	-45.00	-1.00	8.63	11.54	21.51	106.50	17.30	8302	45424
CW	-99.27	-18.94	-6.18	-6.48	6.53	69.77	21.23	5184	26742
CI	-67.15	-9.27	-0.75	-0.28	8.89	78.48	16.14	7960	44860
FC	-0.73	-0.18	-0.07	-0.10	0.00	0.47	0.15	6952	36415

actual weight, LL_X. Comparing the two models through the LRT, we can conclude that they differ significantly ($p \leq 0.001$). This result shows that assuming the parameter α as a random parameter should be considered as a better model for the growth data of the Mertolengo cattle breed.

Table 2. Estimation results for the NMSDE and for the MSDE models assuming random α. The table shows the values of the maximum likelihood estimates and corresponding approximate 95% asymptotic confidence bands, as well as the corresponding value of the log-likelihood. Besides showing the results for the modified weight α, we present them for the weight $A = h^{-1}(\mu)$ (for the NMSDE model $\mu = \alpha$). The AIC and BIC values are also presented. The magnitude of the p-values of the LRT test between the NMSDE and MSDE are shown: ^{ns} ($p > 0.05$); * ($0.01 < p \leq 0.05$); ** ($0.001 < p \leq 0.01$); *** ($p \leq 0.001$).

	NMSDE	**MSDE**
$\boldsymbol{\mu}$	6.5867 ± 0.0142	6.5889 ± 0.0148
A	725.36 ± 10.28	726.98 ± 10.74
$\boldsymbol{\theta}$	$-$	0.1428 ± 0.0092
$\boldsymbol{\beta}$	1.2914 ± 0.0138	1.3002 ± 0.0143
$\boldsymbol{\sigma}$	0.3078 ± 0.0018	0.3010 ± 0.0019
$\mathbf{LL_X}$	-264443.2	-264281.0^{***}
AIC	528892.4	528570.0
BIC	528919.4	528606.0

Let us now explore the mixed model with random α with the inclusion of genetic values (MSDEG). We fit seven MSDEG models, one for each genetic trait, and compared them with the MSDE and NMSDE models using the same datasets (for each trait, the number of animals M and the number of weight observations n of the corresponding dataset can be seen on Table 1). The parameter estimates were obtained by maximization of (10), (8) and (6), respectively. First, we will compare the NMSDE models with the equivalent MSDE models, to evaluate the relevance of considering a random effect on the parameter α, and then we compare the MSDE models with the MSDEG models, to evaluate the effect of adding that specific genetic trait. Table 3 and 4 present the estimates obtained through the maximum likelihood estimation method and the 95% approximate confidence bands based on the inverse of the empirical Fisher information matrix. The AIC and BIC are also presented to compare the models.

In Table 3, we can observe that, for all the genetic traits, the differences between NMSDE and MSDE were significant at $p \leq 0.01$ or $p \leq 0.001$ and the AIC and BIC values were lower on these MSDE models when compared to the NMSDE models. This shows that a random effect should be present in all groups of animals.

When comparing the MSDEG and MSDE models of Table 3 for the genetic trait MC, we can observe that c_1 (the parameter that indicates the importance of the genetic trait on explaining the variability in α) has a small estimate not significantly different from 0. But we should be aware that such lack of significance was based on an approximation of the confidence interval of c_1 obtained from the inverse of empirical Fisher information matrix. The LRT, however, confirms the lack of significance of the maternal capacity until weaning. For the cases of RL and CI, a significance exists with a p-value of the LRT lower than 0.1% (the AIC values are also lower for these two genetic traits). The maximum likelihood estimates from the NMSDE, the MSDE and the MSDEG models are very similar for all common parameters. For the models with random α, it is interesting to compare the standard deviation of the random effect, i.e. to compare the θ of the MSDE model with the σ_δ of the MSDEG model. With the exception of the non-significant MC trait, we see that the incorporation of the genetic trait, RL or CI, reduces the variability of the random effect.

Table 3. Estimation results for the NMSDE, MSDE and MSDEG models. The table shows the values of the maximum likelihood estimates and corresponding approximate 95% asymptotic confidence bands, as well as the corresponding value of the log-likelihood. Estimates for the actual asymptotic weight (in kg) are shown with $A = h^{-1}(\alpha)$, $A = h^{-1}(\mu)$ and $A = h^{-1}(c_0 + c_1 g_1)$, for NMSDE, MSDE and MSDEG, respectively. The AIC and BIC values are also presented. The magnitude of the p-values of the LRT test between the NMSDE and MSDE and between the MSDE and MSDEG are shown: ^{ns} ($p > 0.05$); * ($0.01 < p \leq 0.05$); ** ($0.001 < p \leq 0.01$); *** ($p \leq 0.001$).

	NMSDE		
	MC	RL	CI
$\boldsymbol{\alpha}$	6.5826 ± 0.0157	6.5892 ± 0.0281	6.2829 ± 0.0157
$\mathbf{A}$	722.40 ± 11.32	727.17 ± 20.43	722.68 ± 11.36
$\boldsymbol{\beta}$	1.2938 ± 0.0154	1.3099 ± 0.0272	1.2930 ± 0.0154
$\boldsymbol{\sigma}$	0.3060 ± 0.0020	0.3395 ± 0.0033	0.3062 ± 0.0020
$\mathbf{LL_X}$	-202064.0	-96107.3	-201397.3
AIC	404134.0	192220.5	402800.7
BIC	404160.1	192244.4	402826.8
	MSDE		
$\boldsymbol{\mu}$	6.5814 ± 0.0163	6.5872 ± 0.0284	6.5816 ± 0.0164
$\mathbf{A}$	721.55 ± 11.78	725.75 ± 20.61	721.69 ± 11.82
$\boldsymbol{\theta}$	0.1452 ± 0.0102	0.0707 ± 0.0313	0.1450 ± 0.0102
$\boldsymbol{\beta}$	1.3052 ± 0.0160	1.3143 ± 0.0277	1.3046 ± 0.0160
$\boldsymbol{\sigma}$	0.2989 ± 0.0021	0.3381 ± 0.0035	0.2991 ± 0.0021
$\mathbf{LL_X}$	-201924.8^{***}	-96104.7^{**}	-201259.7^{***}
AIC	403857.6	192217.3	402527.3
BIC	403889.2	192230.2	402514.0
	MSDEG		
$\mathbf{c_0}$	6.5826 ± 0.0164	6.5951 ± 0.0287	6.5799 ± 0.0163
$\mathbf{c_1}$	0.0010 ± 0.0011	-0.0059 ± 0.0026	0.0016 ± 0.0004
$\mathbf{A}$	722.24 ± 11.84	727.18 ± 20.68	720.14 ± 11.74
$\boldsymbol{\sigma_\delta}$	0.1453 ± 0.0102	0.0679 ± 0.0325	0.1410 ± 0.0104
$\boldsymbol{\beta}$	1.3045 ± 0.0160	1.3134 ± 0.0277	1.3076 ± 0.01610
$\boldsymbol{\sigma}$	0.2989 ± 0.0021	0.3381 ± 0.0035	0.2993 ± 0.0021
$\mathbf{LL_X}$	-201923.2^{ns}	-96095.2^{***}	-201235.6^{***}
AIC	403856.3	192200.3	402481.1
BIC	403899.9	192240.2	402524.7

Table 4. Estimation results for the NMSDE, MSDE and MSDEG models. The table shows the values of the maximum likelihood estimates and corresponding approximate 95% asymptotic confidence bands, as well as the corresponding value of the log-likelihood. Estimates for the actual asymptotic weight (in kg) are shown with $A = h^{-1}(\alpha)$, $A = h^{-1}(\mu)$ and $A = h^{-1}(c_0 + c_1 g_1)$, for NMSDE, MSDE and MSDEG, respectively. The AIC and BIC values are also presented. The magnitude of the p-values of the LRT test between the NMSDE and MSDE and between the MSDE and MSDEG are shown: ^{ns} ($p > 0.05$); * ($0.01 < p \leq 0.05$); ** ($0.001 < p \leq 0.01$); *** ($p \leq 0.001$).

	NMSDE			
	GC	WG	CW	FC
$\boldsymbol{\alpha}$	6.5826 ± 0.0157	6.5884 ± 0.0165	6.6007 ± 0.0240	6.6241 ± 0.0210
$\mathbf{A}$	722.40 ± 11.32	726.61 ± 12.01	735.60 ± 17.65	753.00 ± 15.84
$\boldsymbol{\beta}$	1.2938 ± 0.0154	1.2998 ± 0.0161	1.2835 ± 0.0226	1.2553 ± 0.0192
$\boldsymbol{\sigma}$	0.3060 ± 0.0020	0.3156 ± 0.0021	0.3308 ± 0.0028	0.3254 ± 0.0024
$\mathbf{LL_X}$	-202050.7	-204171.3	-120058.1	-163493.7
AIC	404107.3	408348.7	240122.2	326993.5
BIC	404133.4	408374.8	240146.8	327019.0
	MSDE			
$\boldsymbol{\mu}$	6.5814 ± 0.0163	6.5846 ± 0.0170	6.5932 ± 0.0247	6.6216 ± 0.0216
$\mathbf{A}$	721.55 ± 11.78	723.86 ± 12.33	730.11 ± 18.01	751.15 ± 16.23
$\boldsymbol{\theta}$	0.1452 ± 0.0102	0.1256 ± 0.0117	0.1241 ± 0.0175	0.1181 ± 0.0157
$\boldsymbol{\beta}$	1.3052 ± 0.0160	1.3107 ± 0.0167	1.2988 ± 0.0237	1.2644 ± 0.0198
$\boldsymbol{\sigma}$	0.2989 ± 0.0021	0.3106 ± 0.0022	0.3265 ± 0.0030	0.3216 ± 0.0025
$\mathbf{LL_X}$	-201911.4^{***}	-204100.8^{***}	-120028.9^{***}	-163461.3^{***}
AIC	403830.9	408209.6	240065.7	326930.6
BIC	403864.9	408189.3	240081.6	326897.9
	MSDEG			
$\mathbf{c_0}$	6.5818 ± 0.0164	6.5667 ± 0.0175	6.5987 ± 0.0248	6.6040 ± 0.0218
$\mathbf{c_1}$	0.0004 ± 0.0008	0.0015 ± 0.0004	0.0010 ± 0.0005	-0.2223 ± 0.0534
$\mathbf{A}$	721.64 ± 11.79	723.44 ± 12.31	729.4 ± 17.95	753.89 ± 16.34
$\boldsymbol{\sigma_\delta}$	0.1450 ± 0.0102	0.1215 ± 0.0120	0.1217 ± 0.0178	0.1131 ± 0.0163
$\boldsymbol{\beta}$	1.3051 ± 0.0160	1.3087 ± 0.0166	1.2991 ± 0.0236	1.2608 ± 0.0198
$\boldsymbol{\sigma}$	0.2989 ± 0.0021	0.3106 ± 0.0022	0.3265 ± 0.0030	0.3215 ± 0.0025
$\mathbf{LL_X}$	-201911.0^{ns}	-204073.1^{***}	-120020.4^{***}	-163427.9^{***}
AIC	403832.1	408156.1	240050.8	326865.9
BIC	403875.6	408199.7	240091.8	326908.4

In Table 4, for all genetic traits, the differences between NMSDE and MSDE were significant at $p \leq 0.001$. Also the AIC and BIC values were lower on the MSDE models, allowing us to conclude that for these groups of animals we should consider a random asymptotic weight at maturity. The maximum likelihood estimates from the NMSDE, the MSDE and the MSDEG models are very similar for all parameters. For the models with random α, it is interesting to compare the standard deviation of the random effect, i.e. to compare the θ of the MSDE model with the σ_δ of the MSDEG model. With the exception of the non-significant GC trait, we see that the incorporation of the genetic trait WG, CW or FC reduces the variability of the random effect. Regarding the AIC values we can see that the MSDE models provides lower values than the NMSDE equivalent models, which in turn provides lower values that the equivalent MSDEG models.

In summary, comparing models with the same number of observations, the comparison between NMSDE and MSDE models shows that considering the parameter α to randomly vary among animals provides a very significant improvement, and the comparison between MSDE and MSDEG models shows a very significant improvement for

the incorporation of some genetic traits (which help explain part of the variability in α), but not for others (which seems to have no or low influence on α). The genetic traits that are not significant are MC and GC. The genetic traits RL, CI, WG, CW and FC do have a very significant effect, RL and FC with a negative effect $c_1 < 0$ (i.e., α is high for low genetic values, preferably negative) and CI, WG and CW with a positive effect $c_1 > 0$ (i.e., α is high for high genetic values, preferably positive).

5. Conclusion

Incorporating genetic values in the animal growth process is a valuable approach that can lead to significant improvements in livestock production. By considering the genetic potential of animals, breeders can select individuals with desirable traits, resulting in enhanced growth, productivity, and overall performance of the population.

We have studied a general class of SDE mixed models with the inclusion of genetic values to describe individual growth in a randomly varying environment with a real application to the weight of male Mertolenga cattle. A general model can be written as a variant of the Ornstein-Uhlenbeck model, where the parameter α (asymptotic modified weight) is assumed to be random and where the random effect is written as a linear function of the genetic values of the animal. We have used, in the numerical application to Mertolengo cattle data, the weight of the animals and the logarithmic transformation, which reveals to be the most appropriate to that data, leading to the stochastic Gompertz model. Our interest in using SDE mixed models with genetic values comes from the reasonable idea that model parameters may vary from animal to animal, which, for instance, occurs due to different individual characteristics of the animals such as genetic differences.

We apply the maximum likelihood estimation method to obtain the parameter estimates of the SDE mixed models with inclusion of genetic values (MSDEG) and compare each model with the correspondent ones without the inclusion of the genetic values (MSDE) and with the fixed effects model (NMSDE). We compare the models using the likelihood ratio ratio test and also present the AIC and BIC values. The comparison between the SDE mixed model and the SDE non-mixed model clearly shows that, just by assuming the asymptotic weight at maturity of each animal to be a random variable, results in a highly significant improvement. The further importance of considering a genetic value was evaluated by comparing the corresponding MSDEG model with the MSDE model. The results revealed that the incorporation of the genetic values of reproductive longevity, calving interval, average daily weight gain, carcass weight per day of age or feed conversion index, provides a highly significant improvement, while the incorporation of the maternal capacity until weaning and the growth capacity until weaning genetic values shows non-significant improvement.

This study enables the identification of genetic values that contribute the most to describe the animal growth asymptotic size. By incorporating these genetic values, breeders can improve the efficiency and productivity of animal production systems, offering significant benefits to livestock production.

In future work, we intend to implement MSDE and MSDEG models to improve the optimization of the breeder's profit by raising an animal. We should estimate the optimal selling age and optimize the profit using SDE mixed models instead of SDE non-mixed models. We also intend to extend the SDE mixed model by letting the random growth parameter β depend on the genetic values. A combination of several genetic values simultaneously instead of just one at a time should also be addressed, either for the SDE mixed effect model with a random α and the SDE mixed model with a random β.

Acknowledgements

The authors belong to the research center CIMA (Centro de Investigação em Matemática e Aplicações, Universidade de Évora), supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal), project UID/MAT/04674/2020, https://doi.org/10.54499/UIDB/04674/2020. This work was developed within the Operational Group PDR2020-1.0.1-FEADER-031130 - Go BovMais - Productivity improvement in the system of bovine raising for meat, funded by PDR 2020. We are grateful to José Pais from ACBM and Nuno Carolino

from INIAV for providing the weight data and the genetic data and for their continuous support. We thank to RuralBit for the help in extracting the data from Genpro database.

REFERENCES

1. C. A. Braumann, *Introduction to stochastic differential equations with applications to modelling in biology and finance*, John Wiley & Sons, 2019.
2. N. Carolino and L. T. Gama, *Indicators of genetic erosion in an endangered population: The Alentejana cattle breed in Portugal*, Journal of Animal Science, vol. 86, no. 1, pp. 47-56, 2008.
3. N. Carolino, A. Vitorino, I. Carolino, J. Pais, N. Henriques, M. Silveira and A. Vicente, *Genetic diversity in the Portuguese Mertolenga cattle breed assessed by pedigree analysis*, Animals, vol. 10, no. 11, pp. 1990, 2020.
4. D. Charlotte, S. Adeline and H. Simone, *Mixedsde: estimation methods for stochastic differential mixed effects models*, R package version 5.0, 2018.
5. M. Delattre, V. Genon-Catalot and A. Samson, *Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient*, ESAIM: Probability and Statistics, vol. 19, pp. 671–688, 2015.
6. M. Delattre, V. Genon-Catalot and C. Larédo, *Parametric inference for discrete observations of diffusion processes with mixed effects*, Stochastic Processes and Their Applications, vol. 128, no. 6, pp. 1929–1957, 2018.
7. M. Delattre, V. Genon-Catalot and A. Samson, *Maximum likelihood estimation for stochastic differential equations with random effects*, Scandinavian Journal of Statistics, vol. 40, no. 2, pp. 322–343, 2013.
8. P. A. Filipe, C. A. Braumann and C. J. Roquete, *Modelos de crescimento de animais em ambiente aleatório*. In: M. E. Ferrão, C. Nunes and C. A. Braumann (Editores), Estatística Ciência Interdisciplinar. Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística, Edições SPE, pp. 401–410, 2007.
9. P. A. Filipe, C. A. Braumann, N. M. Brites and C. J. Roquete, *Modelling animal growth in random environments: an application using nonparametric estimation*, Biometrical Journal, vol. 52, no. 5, pp. 653–666, 2010.
10. P. A. Filipe, C. A. Braumann, N. M. Brites and C. J. Roquete, *Prediction for individual growth in a random environment*. In: I. Morlini, P. E. Oliveira, M. G. Temido, C. Henriques, and M. Vichi (eds), Recent Developments in Modeling and Applications in Statistics. Studies in Theoretical and Applied Statistics, Springer, Berlin, Heidelberg, pp. 193–201, 2022.
11. O. Garcia, *A stochastic differential equation model for the height growth of forest stands*, Biometrics, vol. 39, no. 4, pp. 1059–1072, 1983.
12. S. Infante, L. Sanchez and A. Hernández, *Stochastic models to estimate population dynamics*, Statistics, Optimization & Information Computing, vol. 7, no. 2, pp. 311–328, 2019.
13. G. Jacinto, P. A. Filipe and C. A. Braumann, *Profit optimization of cattle growth with variable prices*, Methodology and Computing in Applied Probability, pp. 1–36, 2021.
14. N. T. Jamba, G. Jacinto, P. A. Filipe and C. A. Braumann, *Likelihood function through the delta approximation in mixed SDE models*, Mathematics, vol. 10, no. 3, pp. 385, 2022.
15. N. T. Jamba, G. Jacinto, P. A. Filipe and C. A. Braumann, *Estimation for stochastic differential equations mixed models using approximation methods*, Submitted, 2023.
16. D. Maud and D. Charlotte, *MsdeParEst: parametric estimation in mixed-effects stochastic differential equations*, R package version 1.7, 2017.
17. R. V. Overgaard, N. Jonsson, C. W. Tornøe and H. Madsen, *Non-Linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm*, Journal of Pharmacokinetics and Pharmacodynamics, vol. 32, pp. 85–107, 2005.
18. U. Picchini and S. Ditlevsen, *Practical estimation of high dimensional stochastic differential mixed-effects models*, Computational Statistics & Data Analysis, vol. 55, no. 3, pp. 1426–1444, 2011.
19. U. Picchini, S. Ditlevsen and A. D. Gaetano, *Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations*, Journal of Mathematical Biology, vol. 53, pp. 771–796, 2006.
20. U. Picchini, S. Ditlevsen and A. D. Gaetano, *Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics*, Mathematical Medicine and Biology: A Journal of the IMA, vol. 25, no. 2, pp. 141–55, 2008.
21. U. Picchini, A. D. Gaetano and S. Ditlevsen, *Stochastic differential mixed-effects models*, Scandinavian Journal of Statistics, vol. 37, no. 1, pp. 67–90, 2010.
22. Q. Lv and J. W. Pitchford, *Stochastic von Bertalanffy models, with applications to fish recruitment*, Journal of Theoretical Biology, vol. 244, no. 4, pp. 640–655, 2007.
23. K. Rennolls, *Forest height growth modelling*, Forest Ecology and Management, vol. 71, no. 3, pp. 217–225, 1995.
24. J. Ruane and J. J. Colleau, *Marker assisted selection for genetic improvement of animal populations when a single QTL is marked*, Genetics Research, vol. 66, no. 1, pp. 71–83, 1995.
25. J. Soto and S. Infante, *Ensemble Kalman filter and extended Kalman filter for state-parameter dual estimation in mixed effects models defined by a stochastic differential equation*. In: A. Basantes-Andrade, M. Naranjo-Toro, M. Zambrano Vizuete, and M. Botto-Tobar (eds), Technology, Sustainability and Educational Innovation (TSIE), Springer, Warsaw, Poland, pp. 285–300, 2020.
26. R. Wolfinger *Laplace's approximation for nonlinear mixed models*, Biometrika, 80(4), pp. 791–795, 1993.