ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2023-01-12

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Jacinto, G., Filipe, P. A. & Braumann, C. A. (2022). Weighted maximum likelihood estimation for individual growth models. Optimization. 71 (11), 3295-3311

Further information on publisher's website:
10.1080/02331934.2022.2075745

Publisher's copyright statement:
This is the peer reviewed version of the following article: Jacinto, G., Filipe, P. A. & Braumann, C. A. (2022). Weighted maximum likelihood estimation for individual growth models. Optimization. 71 (11), 3295-3311, which has been published in final form at https://dx.doi.org/10.1080/02331934.2022.2075745. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

ARTICLE TEMPLATE

Weighted Maximum Likelihood Estimation for Individual Growth Models

Jacinto, G.[a,b] and Filipe, P. A.[a,c] and Braumman, C. A.[a,b]

[a] Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora, Évora, Portugal
[b] Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora, Rua Romão Ramalho 59, 7000-671 Évora, Portugal
[c] Instituto Universitário de Lisboa (ISCTE - IUL), ISCTE Business School, Quantitative Methods for Management and Economics Department, Av. das Forças Armadas, 1649-026, Lisboa, Portugal

ARTICLE HISTORY
Compiled November 17, 2021

ABSTRACT
We apply a class of stochastic differential equations to model individual growth in a randomly fluctuating environment using cattle weight data. We have used maximum likelihood theory to estimate the parameters. However, for cattle data, it is often not feasible to obtain animal's observations at equally spaced ages nor even at the same ages for different animals and there is typically a small number of observations at older ages. For these reasons, maximum likelihood estimates can be quite inaccurate, being interesting to consider in the likelihood function a weight function associated to the elapsed times between two consecutive observations of each animal, which results in the weighted maximum likelihood method. We compare the results obtained from both methods in several data structures and conclude that the weighted maximum likelihood improves the estimation when observations at older ages are scarce and the observation instants are unequally spaced, whereas the maximum likelihood estimates are recommended when animals are weighted at equally spaced ages. For unequally spaced observations, a bootstrap estimation method was also applied in order to correct the bias of the maximum likelihood estimates; it revealed to be a more precise alternative, except when the available data only has young animals.

KEYWORDS
bootstrap estimation; cattle growth; stochastic differential equations; weighted maximum likelihood estimation.

AMS CLASSIFICATION
60H10; 62M05; 92B99

1. Introduction

Individual growth models represent changes in the size of a single individual over time. To describe and predict individual growth, deterministic models are frequently applied and it is usual to treat random variations of the data using classical regression models. These type of models can be written in the form of ordinary differential

---

CONTACT Jacinto, G.. Email: gjcj@uevora.pt

equations that describe the average dynamics behind the growth patterns. Although our study is exclusively dedicated to individual growth, many of the models applied to population growth are also used to model individual growth.
The most commonly used deterministic growth models are the Misterlich model or monomolecular model, the logistic model, the Gompertz model and the Bertalanffy-Richards model. These type of models have been commonly used in the most diverse areas [2–4,12,16–18,23,25].

In order to take into account the random fluctuations in the internal and external conditions that affect the growth process (hereby named environmental conditions), stochastic versions of deterministic models, in the form of stochastic differential equations (SDE) have been proposed ([11,13,19,20,22]).

In early work (see, for instance, [7–10]), we have studied stochastic versions of a class of deterministic models where an adequate transformation of the size allows us to work with a general SDE model taking the form of a variant of the Ornstein-Uhlenbeck model. That class includes the monomolecular, the logistic, the Gompertz and the Bertallanffy-Richards models, among others. We have applied such class of models using real data on the evolution of bovine weight. This type of models might be useful in animal breeding, aquaculture or forestry in order to optimize the exploitation of such resources.

For the class of SDE models studied, focused on our application, we have seen that one of the most adequate transformations of the animals weight was the logarithm, which lead us to the Stochastic Gompertz Model ([10]). For this model, we have obtained the log-likelihood function and computed the maximum likelihood (ML) estimates of the parameters and the limits of the asymptotic confidence intervals based on the asymptotic variance-covariance matrix (inverse of the empirical Fisher information matrix). The parameters of the model are quite important on farmers' optimization issues as shown in [15]. However the real data on the evolution of bovine weight is not obtained at the same ages for each animal and since the animals are raised for the meat market, there are very few observations of the animal's weight (particularly at older ages), which turns the parameter estimation through the ML method inaccurate. To cope with this problem, we adapted a weighted maximum likelihood (WML) estimation method to obtain the WML estimates and compare with the results obtained by ML estimation. For the WML method the weights are built such that the time elapsed between observations is considered in the likelihood function ([14,21,26]).

We have performed simulation studies to compare both estimation methods under several scenarios for the observed data. We have worked with simulated data for scenarios corresponding to weights measured at equally spaced ages using different spacings (15 and 30 days) and different maximum ages (2, 4 and 6 years). The "real" ages scenario was also considered with simulation of weights at the effective observed ages of a group of animals randomly selected from the 16029 animals on the database.

For small population sizes, ML estimates may have non-negligible biases and the confidence intervals obtained from the Fisher information matrix may be quite unreliable. In such case, bootstrap methods are recommended ([5]). Since that was the case of the "real" ages scenario, a parametric bootstrap estimation method was also applied to the corresponding simulated data set, allowing us to correct the bias and improve the standard error of the ML estimates. A comparison among the 3 methods is presented.

This work was developed within the project GOBov+: *Productivity improvement in the system of bovine raising for meat*, an Operational Group that involves the participation of several institutions (University of Évora, Veterinary Research Institute, several producer associations). Data was provided by the Associação de Criadores de Bovinos Mertolengos (ACBM), which performs the growing and finishing phases of young Mertolengo males, and by associated breeders, whose agricultural holdings are located in the Alentejo region in Portugal. The ACBM action takes place at its Testing and Breeding Center at Herdade dos Currais e Simalhas (CTR), located 15 km from the city of Évora. ACBM enables breeders to breed and finish their cattle, allowing them to obtain a higher economic value than is normally achieved in weaning sales and, at the same time, helping to solve the problem of breeding and finishing when farms do not have technical and/or economic conditions for that purpose. The Mertolengo cattle is, at the moment, considered by many as the Portuguese cattle breed with higher progression in terms of population increment and market potential.

This paper is organized as follows. In Section 2 we describe the stochastic growth models used. In Section 3 the maximum likelihood and the weighted maximum likelihood estimation methods are presented. In Section 4 the results of the application of the methods to the bovine growth are obtained. In Section 5 we present the main conclusions reached.

2. Stochastic differential equations growth models

Most of the classical deterministic models used in the literature to describe the individual growth of an animal with size X_t (weight, volume, height, length, etc.) at age t follow a differential equation of the form

$$\frac{dY_t}{dt} = \beta(\alpha - Y_t), \quad Y_{t_0} = y_0,$$

where $Y_t = h(X_t)$ is the modified size, obtained by rescaling the actual size using a monotonous continuously differentiable scaling function h (a model specific function we assume to be known). Then $y_0 = h(x_0)$ and $\alpha = h(A)$, where x_0 is the observed size at time t_0 (the first observation) and A represents the asymptotic size or the size at maturity of the individual. The parameter β, always positive, is the growth coefficient and represents the rate of approach to maturity. The choice of h leads to some well known classic models. For instance, when $h(x) = x$ we get the monomolecular model; the Bertalanffy-Richards model [3,23] corresponds to the case $h(x) = x^c$ $(c > 0)$, the Gompertz model to the case $h(x) = \ln x$ (can be considered the limiting case of the Bertalanffy-Richards model when $c \to 0$), and the logistic model [25] to the case $h(x) = x^{-1}$.

In the determinist individual growth models, random variations in data have been treated by classical regression models. The traditional assumption of regression models is that observed deviations from the regression curve are independent, which would be quite appropriate if the deviations were due to measurement errors (negligible in our case) but is totally unrealistic when the deviations are due to random changes on growth rates induced by environmental random fluctuations. Stochastic differential equation models are the most appropriate since they incorporate the dynamics of the growth process and the effect environmental random fluctuations have on such dynamics. Thus, we have considered that individual growth in a random environment can be described by stochastic differential equations of the form

$$dY_t = \beta(\alpha - Y_t)dt + \sigma dW_t, \quad Y_{t_0} = y_0, \tag{1}$$

where σ is an environmental noise intensity parameter and W_t is the standard Wiener process. This model is a variant of the Ornstein-Uhlenbeck model and is known in the financial literature as the Vasicek model [24] (used to study interest rate dynamics).

The solution of (1), Y_t, is easily obtained

$$Y(t) = \alpha + e^{-\beta(t-t_0)}(y_0 - \alpha) + \sigma e^{-\beta t}\int_{t_0}^{t} e^{\beta s} dW(s)$$

and, for ages $t_{k-1} < t_k$, the transition distribution is

$$Y(t_k)|(Y(t_{k-1}) = y_{k-1}) \sim N\left(\alpha + (y_{k-1} - \alpha)e^{-\beta(t_k - t_{k-1})}, \frac{\sigma^2(1 - e^{-2\beta(t_k - t_{k-1})})}{2\beta}\right) \tag{2}$$

3. Parameter estimation

Unlike other applications, we assume that the observation ages of each animal are not equidistant nor equal among different animals. Therefore, assume we have data on m animals and, for animal j ($j = 1, 2, \ldots, m$), we observe at the age instants $t_{j,0}, t_{j,1}, \ldots, t_{j,n_j}$, the weights $X_{t_{j0}}, X_{t_{j,1}}, \ldots, X_{t_{j,n_j}}$. The transformed weights are given by $Y_{j,k} = Y(t_{j,k}) = h(X(t_{j,k}))$ ($j = 1, 2, \ldots, m$, $k = 0, 1, \ldots, n_j$). We assume the trajectories of the different animals correspond to different independent realizations of the Wiener process.

We have used maximum likelihood (ML) estimation theory to estimate the parameter vector $\mathbf{p} = (\alpha, \beta, \sigma)$. We describe details on this procedure, for instance in [8,10] and [9].

In [10], we have seen that one of the best models for our cattle data was the stochastic Gompertz model (SGM), which we are going to use along this paper. So, from now on the transformed weight is $Y(t) = \ln X(t)$. Notice that α is the transformed asymptotic weight in the deterministic case (i.e., when $\sigma = 0$) and the corresponding untransformed asymptotic weight is $A = e^{\alpha}$.

For animal j, since $Y(t)$ is a Markov process, knowing the initial weight $X(t_{j,0})$, the likelihood function for that animal is obtained by the product of the transition densities between consecutive observation ages, so the log-likelihood for animal j is given by

$$\begin{aligned} L_{\mathbf{Y_j}}(\mathbf{p}) &= -\frac{n_j}{2}\ln(2\pi) - \frac{n_j}{2}\ln\left(\frac{\sigma^2}{2\beta}\right) - \frac{1}{2}\sum_{k=1}^{n_j}\ln\left(1 - E_{j,k}^{2\beta}\right) \\ &- \frac{\beta}{\sigma^2}\sum_{k=1}^{n_j}\frac{\left(y_{j,k} - \alpha - (y_{j,k-1} - \alpha)E_{j,k}^{\beta}\right)^2}{1 - E_{j,k}^{2\beta}}, \end{aligned} \tag{3}$$

with $E_{j,k} = e^{-(t_{j,k} - t_{j,k-1})}$. By independence, the global log-likelihood function for the

m animals is given by:

$$L_{\mathbf{Y}}(\mathbf{p}) = L_{\mathbf{Y_1},\mathbf{Y_2},\ldots,\mathbf{Y_m}}(\mathbf{p}) = \sum_{j=1}^{m} L_{\mathbf{Y_j}}(\mathbf{p}). \tag{4}$$

Maximizing (4), we obtain the ML estimator of $\mathbf{p}$.

We now present an alternative method to estimate parameters for the individual growth of bovines when the weights of the animals are not taken at equally spaced age instants. In [21] it was suggested to use the weighted maximum likelihood (WML) estimation method. This method uses a known non-decreasing weight function λ and the weights $\Lambda_{j,k} = \lambda(t_{j,k}) - \lambda(t_{j,k-1})$, which take into account the time elapsed between observations. The weighted log-likelihood function is given by

$$\begin{aligned} L_{\mathbf{Y}}(\mathbf{p}) &= \ln\left(\prod_{j=1}^{m}\prod_{k=1}^{n_j}\left(f_{Y_{j,k}|Y_{j,k-1}=y_{j,k-1}}(y_{j,k})\right)^{\Lambda_{j,k}}\right) \\ &= \left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\frac{\sigma^2}{2\beta}\right)\right)\sum_{j=1}^{m}\sum_{k=1}^{n_j}\Lambda_{j,k} - \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{n_j}\Lambda_{j,k}\ln\left(1 - E_{j,k}^{2\beta}\right) \\ &\quad - \frac{\beta}{\sigma^2}\sum_{j=1}^{m}\sum_{k=1}^{n_j}\Lambda_{j,k}\frac{(Y_{j,k} - \alpha - (Y_{j,k-1} - \alpha)E_{j,k}^{\beta})^2}{1 - E_{j,k}^{2\beta}}. \end{aligned} \tag{5}$$

The weight function λ gives more leverage to animal's weights taken at higher ages and at higher interval between observations. In this paper we consider a function of the form $\lambda(t) = t^w$, and we are particularly interested in the effect of the power w on the quality of the weighted maximum likelihood estimators. Note that if $\lambda(t) = t$ and the observations are taken at the same equally spaced ages for all animals, we are in the case of the usual log-likelihood function (3).

Maximizing (5) with $\lambda(t) = t^w$, we obtain the WML estimator of $\mathbf{p}$. Measurability, consistency and asymptotic normality of WML estimators are assured under conditions that are met in our case and can be seen in [21]. It is possible to find closed formulas for the estimators of the parameter α as a function of the WML estimator of β, as follows,

$$\hat{\alpha} = \frac{\sum_{j=1}^{m}\sum_{k=1}^{n_j}\Lambda_{j,k}\frac{(Y_{j,k} - Y_{j,k-1}E_{j,k}^{\hat{\beta}})}{1 + E_{j,k}^{\hat{\beta}}}}{\sum_{j=1}^{m}\sum_{k=1}^{n_j}\Lambda_{j,k}\frac{1 - E_{j,k}^{\hat{\beta}}}{1 + E_{j,k}^{\hat{\beta}}}} \tag{6}$$

and then obtain $\hat{\sigma}^2$ as a function of $\hat{\alpha}$ and $\hat{\beta}$

$$\hat{\sigma^2} = 2\hat{\beta}\frac{\sum_{j=1}^{m}\sum_{k=1}^{n_j}\Lambda_{j,k}\frac{(Y_{j,k} - \hat{\alpha} - (Y_{j,k-1} - \hat{\alpha})E_{j,k}^{\hat{\beta}})^2}{1 - E_{j,k}^{2\hat{\beta}}}}{\sum_{j=1}^{m}\sum_{k=1}^{n_j}\Lambda_{j,k}}. \tag{7}$$

If we replace expressions (6) and (7) in (5), we can obtain the maximum log-likelihood as a function of $\hat{\beta}$, which then can be maximized with respect to $\hat{\beta}$ through numerical methods. Replacing this estimate for $\hat{\beta}$ in the expression of $\hat{\alpha}$ and $\hat{\sigma^2}$, we obtain the WML estimate of the whole parameter vector $\mathbf{p}$. The asymptotic confidence bands can also be obtained based on the empirical Fisher information matrix.

The dataset contains information on the weight, in kg, of 16029 Mertolengo cattle males, where each animal has several observations with a minimum of 3 and a maximum of 33 weights at ages varying between birth and a maximum age that ranges between 0.2 and 16 years, totalling 96204 observations. Since we have a very high number of animals, for the purpose of our analysis and to obtain the simulation results in an acceptable running time we need a smaller number of animals. For this reason we consider a baseline random sample of 100 animals from the entire dataset of 16029 animals. For these randomly selected 100 animals, we obtain the ML estimates of the parameter vector $\mathbf{p}$ obtained for the SGM, which we consider as ours true parameters values for the purpose of the simulation studies we are going to undertake in order to compare the different estimation methods. These parameters are shown in Table 1. Note that also in this sample we have a high heterogeneity between animals in terms of animal's ages and number of observations per animal.

Table 1. ML estimates and asymptotic standard deviations for the baseline random sample of 100 animals (taken from the entire database of 16029 animals).

	ML estimate	std. dev.
β $(year^{-1})$	1.479	0.061
σ^2 $(year^{-1})$	0.086	0.005
α	6.415	0.048
$A = e^{\alpha}$ (kg)	611.16	29.218

The goal is to assess the performance of the WML estimators compared with the ML estimators, considering the estimation of the whole parameter vector $\mathbf{p}$. Usually, when estimating the parameters to describe animal growth, the true parameters are not known and maximum likelihood estimation is used with an inherent statistical uncertainty and possible biases due to heterogeneities among animals and ages at which weight measurements are taken.

To evaluate the performance of the methods, we will assume from now on that the true parameter values (on which all simulations for the ML and the WML methods are based) are known and are exactly $A = e^{\alpha} = 611.16$kg (corresponding to $\alpha = 6.415$), $\beta = 1.479$ $year^{-1}$ and $\sigma^2 = 0.086$ $year^{-1}$.

Using such true parameter values and the SGM model, we do a simulation study consisting of $N = 100$ Monte Carlo independent simulations of age/weight data sets (N could be any number of our choice, not necessarily 100), taking advantage of the Markov property of the model solution and using its Gaussian transition distributions (obtained in (2)). In each of the N data sets we have the simulated weights of 100 animals at the same age instants $t_{j,k}$, $j = 1, 2, \ldots, m, k = 1, 2, \ldots, n_j$ as the real observation ages of the baseline random sample (let us call those age instants the "real" age scheme). Notice that we know the true parameter values of the N simulated data sets. Finally, we can use the different estimation methods on each simulated data set and see how well they behave.

Notice that the "real" age scheme of the 100 baseline sample is heterogeneous, differing among the 100 animals $j = 1, 2, \ldots, m = 100$ in number of observations n_j and/or in the observation ages $t_{j,k}$ $(k = 1, 2, \ldots, n_j)$. At this stage, we have a simulated study with $N = 100$ simulated replications, each consisting of a data set of 100 animals. What all N data sets have in common is the "real" age scheme they all use, but the weights at such ages of the 100 animals of each data sample vary from data sample to data sample (since they are Monte Carlos simulated).

In order to study the performance of the estimation methods under homogeneous observation ages, we repeat the simulation study, again consisting of a Monte Carlo simulation of $N = 100$ data sets but, instead of using the "real" age scheme, we use a homogeneous age scheme (with equidistant and common observation ages for the animals in a data set, starting at birth). Of course, we use the same homogeneous age scheme for all the N data sets in a simulation study. We will do 6 simulation studies (of $N = 100$ data sets) for 6 homogeneous age schemes, with two choices of age intervals between consecutive weight observations (intervals of 15 days and 30 days) and three choices of the last observation age (2 years old, 4 years old and 6 years old).

The main goal with these different 7 simulated studies (one with the "real" age scheme and the other 6 with the different homogeneous age schemes just mentioned) is to evaluate the influence of having different data structures in the parameter estimates: the influence of having equally spaced ages versus non-equally spaced ages; the influence of having a higher versus a smaller elapsed time between two consecutive ages; the influence of having younger ages versus having older ages. Remember that our "real" age scheme is a little bit of everything: not equally spaced age instants, animals with a small number of weights and a large portion of animals at younger ages.

For each simulated dataset we estimate the parameters through the ML method and the WML methods. Since we wish also to optimize the weight function, we search for the optimal weight function, more precisely, the optimal w within the family of functions given by $\Lambda^w_{j,k} = \lambda(t_{j,k}) - \lambda(t_{j,k-1}) = t^w_{j,k} - t^w_{j,k-1}$, $j = 1, 2, \ldots, m, k = 1, 2, \ldots, n_j$. So, for each replica of the simulated data for whom we know the true parameters, we obtain, for each value w, the WML estimates of the parameter vector $(\alpha, \beta, \sigma^2)$ for the SGM. The WML estimates obtained for each w were then compared with the true values. Notice that for equally spaced weight instants (with spacing δ), when $w = 1$, the WML estimates correspond to the ML estimates, since we have the constant weights $\Lambda^1_{j,k} = \lambda(t_{j,k}) - \lambda(t_{j,k-1}) = t^1_{j,k} - t^1_{j,k-1} = \delta$. Then, for the simulation studies with homogeneous age schemes, when $w = 1$ we are in fact using the ML estimates, and when $w \neq 1$ we are using the WML estimates. For the simulated study with the "real" age scheme, the ML estimates have to be obtained separately.

4. Results

To select the best value w, we consider an error measurement given by the root mean squared relative error (RMSRE)

$$RMSRE = \sqrt{\frac{1}{3}\left(\frac{\sum_{i=1}^{N}(\hat{\alpha}_i - \alpha)^2}{N\alpha^2} + \frac{\sum_{i=1}^{N}(\hat{\beta}_i - \beta)^2}{N\beta^2} + \frac{\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma^2)^2}{N(\sigma^2)^2}\right)}, \tag{8}$$

where the $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{\sigma}_i^2$ are the estimates obtained on simulation number i ($i = 1, \ldots, N$) of the parameters α, β and σ^2. This is the square root of the mean value for the three parameters of their estimated mean square errors expressed in relative terms (i.e., each divided by their squared values). We then choose the value w that minimizes the statistic $RMSRE$.

In Table 2 we present the mean and standard deviation of the WML estimates (based on (5)) of the parameters α, β and σ^2, only for the best value w used in the weight function according to the $RMSRE$. We do it for the 6 simulated studies with homogeneous age schemes and for the simulation study with the "real" age scheme. For this latter scheme, we also present the results using the ML method (4).

Table 2. Mean estimated parameters, empirical standard deviations and $RMSRE$ for the WML method for the best choice of w. With the exception of the "real" age scheme, results coincides with the ML method when $w = 1$. For the "real" age scheme, the table also presents results for the ML method.

	2y15d	2y30d	4y15d	4y30d	6y15d	6y30d	Real Age WML	Real Age ML
w	1.0	0.8	1.0	1.0	1.0	1.0	3.2	–
$RMSRE$	0.015	0.021	0.013	0.010	0.010	0.011	0.176	0.213
β ($year^{-1}$)	1.476	1.470	1.480	1.481	1.481	1.486	1.478	1.454
std. dev.	0.021	0.022	0.021	0.020	0.020	0.017	0.301	0.062
σ^2 ($year^{-1}$)	0.086	0.086	0.086	0.086	0.086	0.086	0.091	0.116
std. dev.	0.002	0.003	0.001	0.001	0.001	0.001	0.019	0.009
α	6.417	6.422	6.414	6.415	6.415	6.412	6.383	6.421
std. dev.	0.018	0.021	0.012	0.011	0.011	0.007	0.157	0.046
$A = e^{\alpha}$ (kg)	612.6	615.4	610.4	611.2	611.0	609.1	599.0	614.9
std. dev.	10.579	12.809	7.461	6.509	6.509	4.238	94.567	28.080

The results on table 2 show that the estimation of the parameter vector $\mathbf{p}$ increases its precision and decreases its variability when we have homogeneous age schemes with animals with older ages and more regular equally spaced observations. This is even more clear observing the $RMSRE$. Only the estimate of the parameter σ^2 does not seem to be influenced by the animals age or the interval between observations, when having a homogeneous age scheme of observations. It is also clear that, when we observe the results for the simulated study with the "real" age scheme, the estimates have much lower precision and higher variability and the choice of an accurate weight function through the parameter w could be a good approach to obtain more accurate estimates.

It is interesting also to note that, for homogeneous age schemes, the best w obtained for the WML method is always around $w = 1$, which gives a real evidence about the ML method as the best estimation method for these age schemes.

According to the $RMSRE$ for the simulation study with the "real" age scheme, the use of the WML method is slightly better for the estimation of the whole parameter vector $\mathbf{p}$, having and error 17% lower. If we envisage only the estimation of the two most influential parameters in applications, α and β, then the ML method not only provides a low bias but also has a lower $RMSRE$ by about 77%, showing the influence of the precision estimation of σ^2 in the ML method.

In fact, we have concluded that with lower ages and non equidistant observations, the ML method can have a higher error than the error observed with homogeneous age schemes of observation. For that reason we forced a simulation study with the same true parameters but using a new more realistic baseline age scheme, hereby called "realistic" age scheme. This age scheme is the same age scheme of a new random sample of 100 animals extracted from our database, but restricting the choice to animals that have at least 4 weight observations with the last observation happening at the maximum age of 18 months. Notice that this is in fact the scenario we face in ACBM, since the purpose of the Association is to raise cattle to sell in the meat market, and according to the dataset the mean selling age of a bovine male is about 14 months old. For this new simulation study, we also generate $N = 100$ replications of 100 animals with the "realistic" age scheme of observations, and obtained the ML and the WML estimates for each w. The results are presented in Table 3.

Table 3. Mean estimated parameters, empirical standard deviations and $RMSRE$ for the WML (with the best choice of w) and ML methods for the simulation study with the "realistic" data scheme (at least 4 observations and last observation at maximum age of 18 months).

	Realistic ages WML	Realistic ages ML
w	0.4	–
$RMSRE$	0.095	0.977
β $(year^{-1})$	1.535	1.661
std. dev.	0.086	0.059
σ^2 $(year^{-1})$	0.094	0.216
std. dev.	0.010	0.008
α	6.352	6.304
std. dev.	0.075	0.071
$A=e^{\alpha}$ (kg)	575.5.0	545.3
std. dev.	43.671	39.213

Observing the results in Table 3 we can conclude that the lack of data at higher ages does not allow the ML method to estimate accurately the model parameters. However, the use of weights in the maximum likelihood allows to recoup the lack of information. The $RMSRE$ has a decrease of about 90%. Note, though, that the best value of w is only 0.4. However, if we use the value seen in previous results of $w = 3.2$ for the weight function, the $RMSRE$ increases (the same happens to the $RMSRE$ of the parameter α and of the parameter β), but it is still 80% lower that the error of the MLE. If we envisage only the estimation of the two most influential parameters, α and β, the WML method is still a better method than ML method, having a $RMSRE$ about 52% lower. We also tested the same results but allowing animals to have a maximum age of 30 months. The allowance of this additional information about animal's weights at higher ages is enough to make the ML method perform better than the WML method.

4.1. Optimization of the weight function

In this section, we evaluate the influence of the power w in the weight function $\Lambda^w_{j,k} = \lambda(t_{j,k}) - \lambda(t_{j,k-1}) = t^w_{j,k} - t^w_{j,k-1}$, $j = 1, 2, \ldots, m, k = 1, 2, \ldots, n_j$. We wish to better understand the range of values w that improves the estimation of the parameter vector $\mathbf{p}$ along the 7 simulation studies, 6 with homogeneous age schemes and one with the "real" age scheme.

In Figures 1, 2 and 3 we present the comparison of the WML estimates of the parameters α, β and σ^2, respectively, for a wide range of w values in the weight function $\Lambda^w_{j,k} = \lambda(t_{j,k}) - \lambda(t_{j,k-1}) = t^w_{j,k} - t^w_{j,k-1}$. The 95% confidence bands for the parameter estimate for the "real" age scheme simulation study are also plotted.

Looking at the three plots, we see that, for homogeneous age schemes of observations, the choice of w has a mild influence, but the best estimates occur when we have an age scheme with older animals. For these homogeneous age schemes, the ML (corresponding to $w = 1$), which was the best (or almost the best) method for the overall estimation of the three parameters, may be slightly outperformed on the estimation of individual parameters by the WML method with appropriate w, particularly when older ages data are not available.

The panorama is quite different for the heterogeneous "real" age scheme of observations. We have seen that an appropriate choice of w was relevant for the overall estimation of the three parameters, but we can see now that it is also very influential on the estimation of each parameter individually. The major problem is that the best w values for individual parameter estimation differ from one parameter to another, a good compromise occurring for w values between 3 and 4.

4.2. Bootstrap Estimation

For the simulation study with the "real" age scheme, for which we know the true parameters, we are going to compare the ML and the WML approaches used above with a new estimation method, which we call for short Bootstrap method and consists in applying a bootstrap bias correction to the parameter estimates of the ML method. In fact, the bootstrap methodology is often used to correct the bias of the maximum likelihood estimates and also to obtain better confidence intervals for the parameter estimates, since the asymptotic confidence intervals based on the Fisher information matrix may be quite unreliable for small sample sizes.

Notice that, when one applies the bootstrap bias correction, one does not know the true parameter values and so the correction is based on simulations using the ML parameter estimates as proxies. To obtain the estimates using the parametric bootstrap method, we proceed as follows:

- For each simulated data set $i = 1, 2, \ldots, N = 100$ (consisting of simulated weights of 100 animals at the ages of the "real" age scheme), we obtain the ML estimate $\hat{\mathbf{p}}^{*\mathbf{i}} = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}^2_i)$ of the parameter vector;
- The estimate $\hat{\mathbf{p}}^{*\mathbf{i}}$ is an approximation of the true parameter vector. Now, we use $\hat{\mathbf{p}}^{*\mathbf{i}}$ as a proxy (of the now assumed unknown true parameter vector) to simulate 500 new data sets (of 100 animal each and with the same "real" age scheme)

$$\mathbf{y}^{*\mathbf{i},\mathbf{k}}_{\mathbf{j}} = (y^{*i,k}_{j,0}, y^{*i,k}_{j,1}, ..., y^{*i,k}_{j,n_j}), k = 1, 2, \ldots, 500, i = 1, ..., N = 100, j = 1, ..., m = 100.$$

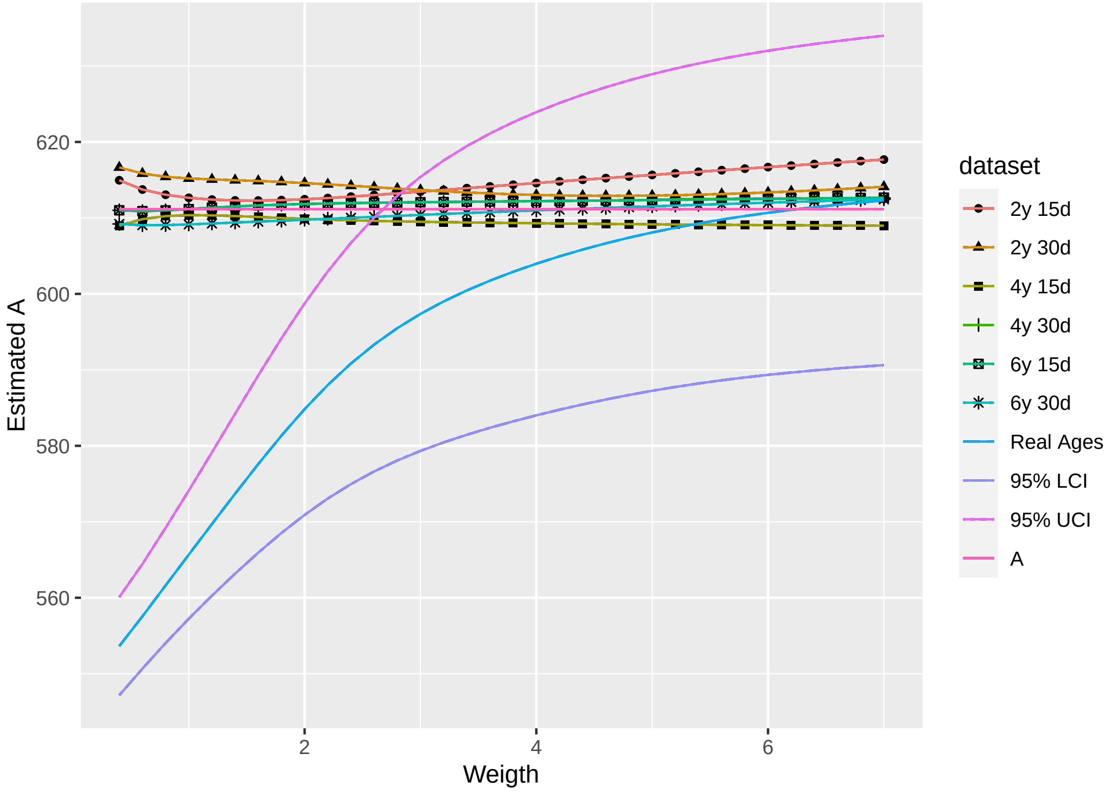

Figure 1. WML estimates of the parameter $A = e^{\alpha}$ as a function of w for the simulation studies with the indicated age schemes. For the "real" age scheme, the 95% confidence intervals are also shown.

For that we use the Markov property of the Y process and the transition distributions given by (2).

- For each $k = 1, 2, \dots, 500$, a new parameter vector is estimated by the ML method and the sample mean of the 500 obtained values, denoted by $\bar{\mathbf{p}}^{*\mathbf{i}}$, is used as an approximation of the expectation of the ML estimate of $\hat{\mathbf{p}}^{*\mathbf{i}}$, thus showing that the ML method tends to give a bias of around $\bar{\mathbf{p}}^{*\mathbf{i}} - \hat{\mathbf{p}}^{*\mathbf{i}}$;
- So, the original ML estimate $\hat{\mathbf{p}}^{*\mathbf{i}}$ should suffer the corresponding bias correction, resulting in a bias corrected ML estimate $\mathbf{p}^{*\mathbf{i},\mathbf{B}} = \hat{\mathbf{p}}^{*\mathbf{i}} - (\bar{\mathbf{p}}^{*\mathbf{i}} - \hat{\mathbf{p}}^{*\mathbf{i}})$;
- Finally the sample mean $\mathbf{p}^{\mathbf{B}}$, the standard deviation and the empirical 2.5%, 50% and 97.5% empirical quantiles of the 100 bias corrected ML estimates $\mathbf{p}^{*\mathbf{i},\mathbf{B}}$ $(i = 1, \dots, N = 100)$ are also obtained.

In Table 4 are shown the true parameter vector, the mean of the ML and of the WML (with $w = 3.2$) estimates of the 100 replicates and the estimates obtained from the bootstrap method and respective 2,5%, 50% and 97,5% quantiles.

Observing the results in the Table 4, we conclude that the parametric Bootstrap method is the most precise method to estimate the growth parameters of bovine weight for the "real" age scheme of observations. The method was able to reduce the error of the estimates (measured through $RMSRE$) in 66% w.r.t. the ML method and in about 60% w.r.t. WML method. The estimates obtained are not only more precise than the estimates obtained by the other methods, but also have lower variability.

As seen in section 4, the MLE performs badly for the "realistic" age scheme, where

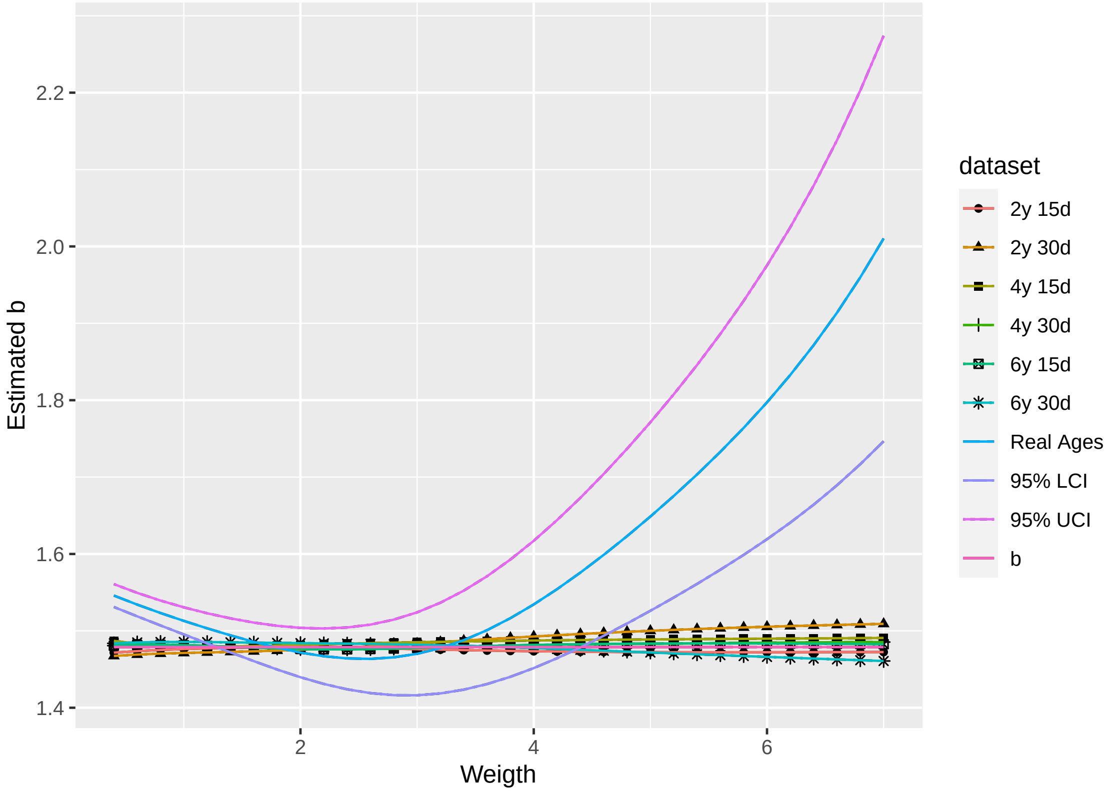

Figure 2. WML estimates of the parameter β (in the plot, b stands for β) as a function of w for the simulation studies with the indicated age schemes. For the "real" age scheme, the 95% confidence intervals are also shown.

Table 4. True parameter values, mean estimated parameters, empirical standard deviations and $RMSRE$ for the ML method, the WML method (with $w = 3.2$) and the Bootstrap method for the simulation study with the "real" age scheme. For the Bootstrap method, the empirical 2,5%, 50% and 97,5% quantiles are also shown.

				Parametric Bootstrap Method			
	True	MLE	WMLE	Bootstrap	q2.5%	q50%	q97.5%
$RMSRE$	-	0.213	0.176	0.072	-	-	-
β $(year^{-1})$	1.479	1.454	1.478	1.478	1.368	1.475	1.596
std. dev.	-	0.062	0.301	0.059	-	-	-
σ^2 $(year^{-1})$	0.086	0.116	0.090	0.083	0.070	0.082	0.098
std. dev.	-	0.009	0.019	0.008	-	-	-
α	6.415	6.421	6.383	6.416	6.322	6.417	6.504
std. dev.	-	0.046	0.152	0.046	-	-	-
$A = e^{\alpha}$ (kg)	611.16	614.92	599.0	610.99	556.28	611.25	666.79
std. dev.	-	28.08	94.68	27.79	-	-	-

animals have observations at most at the age of 18 months. In Table 5, we present the results of the Bootstrap method for the simulation study with the "realistic" age scheme. Observing the results in Table 5, we can conclude that the Bootstrap method largely corrects the estimates of the ML method, reducing the error in 60%. However, the WML method accomplishes a much lower error, showing an error reduction of 90%. The ML and the Bootstrap methods mainly fail in the estimation of the parameter σ^2

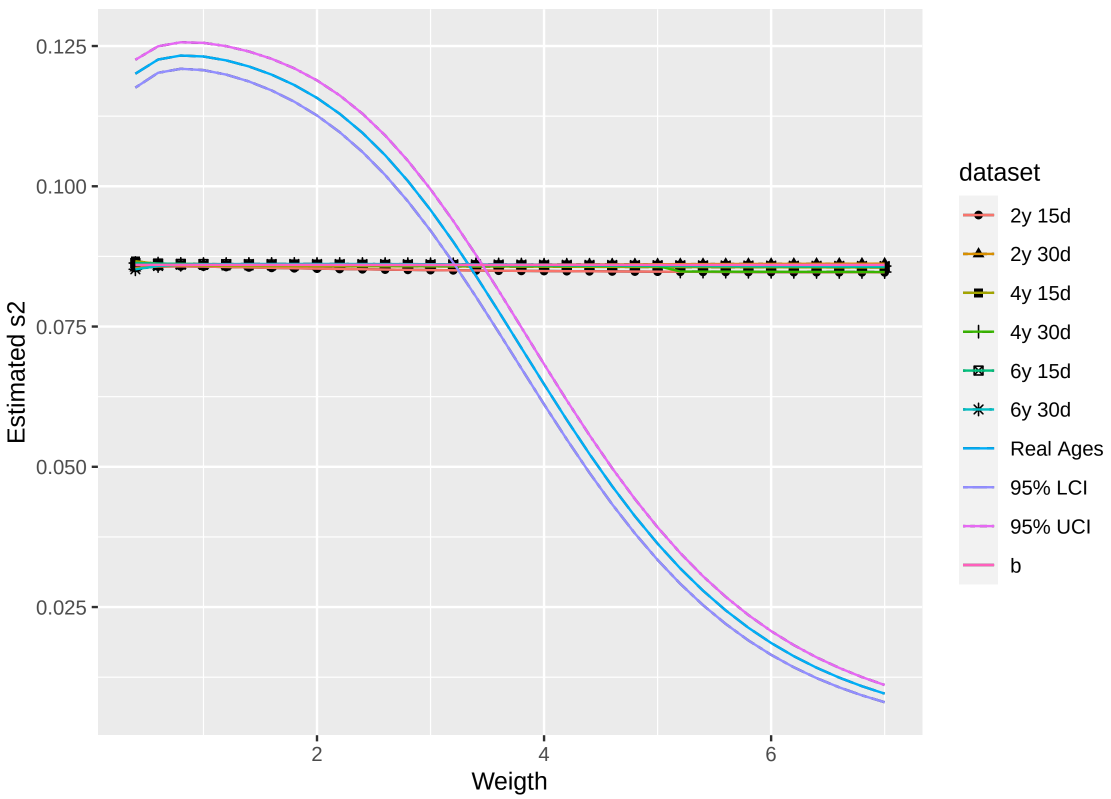

Figure 3. WML estimates of the parameter σ^2 (in the plot, $s2$ stands for σ^2) as a function of w for the simulation studies with the indicated age schemes. For the "real" age scheme, the 95% confidence intervals are also shown.

that measures the intensity of random environment fluctuations. If we observe the error measured just for the parameters α and β, the most influencial in many applications, the error of the Bootstrap method is 52% less than the error of the ML method and only 1% lower than the error of the WML method.

Table 5. True parameter values, mean estimated parameters, empirical standard deviations and $RMSRE$ for the ML method, the WML method (with $w = 0.4$) and the Bootstrap method for the simulation study with the "realistic" age scheme (observations at most at the age of 18 months). For the Bootstrap method, the empirical 2,5%, 50% and 97,5% quantiles are also shown.

				Parametric Bootstrap Method			
	True	MLE	WMLE	Bootstrap	q2.5%	q50%	q97.5%
$RMSRE$	-	0.977	0.095	0.388	-	-	-
β $(year^{-1})$	1.479	1.661	1.535	1.433	1.269	1.428	1.619
std. dev.	-	0.114	0.086	0.091	-	-	-
σ^2 $(year^{-1})$	0.086	0.216	0.094	0.052	<0.001	0.040	0.163
std. dev.	-	0.064	0.010	0.046	-	-	-
α	6.415	6.304	6.352	6.411	6.266	6.415	6.541
std. dev. .	-	0.071	0.075	0.077	-	-	-
$A = e^{\alpha}$ (kg)	611.16	548.30	575.45	602.07	520.25	602.92	682.23
std. dev.	-	39.243	43.671	45.942	-	-	-

5. Conclusion

In this work we model individual growth in a randomly varying environment through a general and flexible class of SDE models. Using the stochastic Gompertz model, we estimate the model parameters through the maximum likelihood (ML) method. However, in the literature most applications are in other areas where available observations are equidistant in time, are taken at the same instant for all trajectories and there are large time observations, sufficient for a reasonable knowledge of the complete growth curve. For our application, regarding the real age schemes of observation data on the evolution of bovine weight of Mertolengo cattle breed, the observations are not taken at equidistant time ages nor are the animals weighted at the same ages and, since the animals are raised for the meat market, there are very few observations of animal weights at older ages. Also, most animals have a very few number of observations.

Due to the structure of the real data, estimating the growth model parameters through the ML method can be inaccurate. We propose estimating the parameters using instead the weighted maximum likelihood (WML) estimation method, in an attempt to improve the estimation when there are a small number of observations at older ages. We also compared the estimation methods for different homogeneous age schemes of observation, using equidistant observations, taken at every 15 or every 30 days, and until a maximum age of 2, 4 or 6 years.

We conclude that, when the age scheme of observations is homogeneous with equidistant observation ages common to all animals and there are a reasonable number of observations, the maximum likelihood estimates are recommended and the accuracy of the estimates improves as the range of observations reaches older ages. For heterogeneous age schemes of observation, like the ones available in our application data for the evolution of bovine weight of Mertolengo cattle breeds, the estimation methods become much more imprecise. A sample of animals with an heterogeneous age scheme of observations, but with a large number of observations and some of them for animals with older ages, is enough for a similar performance of the ML and WML. However, if in the dataset all the observations are for animals at younger ages, the WML method performs a lot better than the ML method, being the recommended method in those situations.

We have also compared the influence in the estimation of the power w on the weight function $\Lambda^w_{j,k} = t^w_{j,k} - t^w_{j,k-1}$. We have concluded that the value of $w = 3.2$ can be the one that, for the heterogeneous "real" age scheme in our case, provides a better estimate of the whole parameter vector $\mathbf{p}$. However, in real datasets with very young animals ("realistic" age scheme), we can minimize the parameter estimation error by using a very small value of $w = 0.4$.

Of course, in each application, the best w value should be sought performing a study similar to the one we did here.

Having noticed that, when we are dealing with real data, the ML and the WML methods can estimate the parameters with bias and high variability, we applied the Bootstrap estimation method (which is the ML method followed by a parametric bootstrap bias correction of the parameter estimates). The results showed that the estimation error decreases significantly using this method when dealing with heterogeneous "real" data schemes where some animals have older observation ages. When facing the heterogeneous "realistic" age scheme with only very young animals, the

Bootstrap method improved significantly the ML estimation method, but, compared with the WML method, the estimates have similar or even worse precision.

Acknowledgements

The authors belong to the research centre CIMA (Centro de Investigação em Matemática e Aplicações, Instituto de Investigação ee Formação Avançada, Universidade de Évora), supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal), project UID/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 ACBM and José Pais (ACBM head engineer) for providing the data and for continuous support. We are also grateful to RuralBit for the help in extracting the data from Genpro database.
We also gratefully acknowledge the Editor and the Referees for valuable comments that have contributed to the improvement of this paper.

References

[1] Braumann, C.A., Carlos, C., Filipe, P.A. and Roquete, C.J. Growth of individuals in randomly fluctuating environments, In: Proceedings of the 2009 International Conference in Computational and Mathematical Methods in Science e Engineering (Vigo-Aguiar J., Alonso P., Oharu S., Venturino E. and Wade B., Eds.), Gijón. 2009; 201-212.

[2] Braumann, C.A. Introduction to Stochastic Differential Equations with Applications to Modelling in Biology and Finance. John Wiley & Sons. 2019.

[3] Bertalanffy, L. von. Quantitative laws in metabolism and growth. The Quarterly Review of Biology. 1957; 34:786–795.

[4] Darmani Kuhi, H. , Porter, T., López, S., Kebreab, E., Strathe, A.B., Dumas, A., Dijkstra, J. and France, J. A review of mathematical functions for the analysis of growth in poultry, World's Poultry Science Journal. 2010; 66:2, 227-240.

[5] Efron, B. and Tibshirani, R.J. An Introduction to the Bootstrap. Chapman & Hall, New York. 1993.

[6] Filipe, P.A., Braumann, C.A. and Carlos, C. Profit optimization for cattle growing in a randomly fluctuating environment. Optimization: A Journal of Mathematical Programming and Operations Research. 2015; 64(6):1393–1407.

[7] Filipe, P.A., Braumann, C.A. and Roquete, C.J. Multiphasic individual growth models in random environments. Methodology and Computing in Applied Probability. 2012; 14(1):49–56.

[8] Filipe, P.A., Braumann, C.A., Brites, N.M. and Roquete, C.J. Modelling animal growth in random environments: an application using nonparametric estimation. Biometrical Journal. 2010; 52(5):653–666.

[9] Filipe, P.A. and Braumann, C.A. Modelling individual animal growth in random environments. In: Proceedings of the 23^{rd} International Workshop on Statistical Modelling (Eilers, P.H.C., Ed.), Utrecht. 2008; 232–237.

[10] Filipe, P.A., Braumann, C.A. and Roquete, C.J. Modelos de crescimento de animais em ambiente aleatório. In: Estatística Ciência Interdisciplinar, Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística (Ferrão, M.E., Nunes, C. and Braumann, C.A., Eds.), Edições SPE. 2007; 401–410.

[11] Garcia, O. A stochastic differential equation model for the height of forest stands. Biometrics. 1983; 39:1059–1072.

[12] Gompertz, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transaction of the Royal Society of London. 1825; 115:515–585.

[13] Ghosh, H., Prajneshu. Optimum fitting of Richards growth model in random environment. Journal of Statistical Theory and Practice. 2019; 13, 6.

[14] Hu, F., Rosenberger, W. F. and Zidek, J. V. Relevance weighted likelihood for dependent data, Metrika. 2000; 51(3): 223–243.

[15] Jacinto, G., Filipe, P.A. and Braumann, C.A. Profit Optimization of Cattle Growth with Variable Prices. Methodol Comput Appl Probab (2021). `https://doi.org/10.1007/s11009-021-09889-z`

[16] Kebreab, E., Schuli-Zeuthen, M., López, S., Dias, R.S., De Lange, C.F.M. and France, J. Comparative evaluation of mathematical functions to describe growth and efficiency of phosphorus utilization in growing pigs. Journal of Animal Science. 2007; 85: 2498–2507.

[17] Kozusko, F. and Bajzer, Z. Combining gompertzian growth and cell population dynamics. Mathematical Biosciences. 2003; 185:153–167.

[18] Ohnishi, S. and Akamine, T. Extension of von Bertallanffy growth model incorporating growth patterns of soft and hard tissues in bivalve molluscs. Fisheries Science. 2006; 72(4):787–795.

[19] Prajneshu and Ghosh, H. Stochastic differential equation models and their applications to Agriculture: an overview. Statistics and Applications. 2019; 17(1):73–83.

[20] Qiming, Lv. and Pitchford, J. Stochastic von Bertalanffy models, with applications to fish recruitment. Journal of Theoretical Biology. 2007; 244:640–655.

[21] Rao, B.L.S.P. Asymptotic theory of weighted maximum likelihood estimation for growth models In: Statistical Inference for Stochastic Processes (Prabhu, N. and Vasawa, I. V., Eds.), Dekker, New York. 1991; 183-208.

[22] Rennolls, K. Forest height growth modelling. Forest Ecology and Management. 1995; 71:217–225.

[23] Richards, F. A flexible growth function for empirical use. Journal of Experimental Botany. 1959; 10:290–300.

[24] Vasicek, O. An equilibrium characterization of the term structure. Journal of Financial Economics. 1977; 5:177–188.

[25] Verhulst, P.F. Notice sur la loi que la population poursuit dans son accroissement. Correspondance Mathématique et Physique. 1838; 10:113–121.

[26] Wang, X., Eeden, C. V., Zidek, J. V. Asymptotic properties of maximum weighted likelihood estimators, Journal of Statistical Planning and Inference. 2004; 119 (1): 37–54.