Stochastic Differential Equations General Models of Individual Growth in Uncertain Environments and Application to Profit Optimization in Livestock Production

Braumann, Carlos A.
Centro de Investigação em Matemática e Aplicações, Universidade de Évora
Apartado 94, 7002-554 Évora, Portugal
E-mail: braumann@uevora.pt

Filipe, Patrícia A.
Centro de Investigação em Matemática e Aplicações, Universidade de Évora
Apartado 94, 7002-554 Évora, Portugal
E-mail: pasf@uevora.pt

Carlos, Clara
Centro de Investigação em Matemática e Aplicações, Universidade de Évora
Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal
Rua Américo da Silva Marinho
2839-001 Lavradio, Portugal
E-mail: clara.carlos@estbarreiro.ips.pt

Roquete, Carlos J.
Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora
Apartado 94, 7002-774 Évora, Portugal
E-mail: croquete@uevora.pt

Individual animal growth in a randomly varying environment is modeled using stochastic differential equation models. These models are generalizations of the classical deterministic growth models used in regression methods, but incorporate a random dynamical term describing the effects of environmental and other random fluctuations on the growth process. We describe parameter estimation and prediction methods, illustrating with data on cow growth of the Mertolengo breed raised in Alentejo (Portugal) under natural conditions. We first show that these models outperform the traditional regression models in predictive power for they take into account the dynamical nature of the growth process and its interaction with environmental fluctuations.

We then apply the models to profit optimization in livestock production, taking into account production costs and sales revenues.

Assuming the animal is to be sold when it reaches some prescribed age, we determine the optimal age at which an animal should be sold in order to maximize profit. Another possibility is to sell the animal when it reaches a prescribed size. The first passage time distribution through a prescribed size is studied and used to determine the optimal size at which the animal should be sold. We then determine which policy (selling at a fixed age or selling at a fixed size) is preferable in terms of profit.

Individual growth models

Most commonly classical deterministic growth models used to describe individual growth of an animal (or plant) in terms of its size X_t (weight, volume, height, lenght, etc.) at instant (age) t can be written in the form of an autonomous differential equation given by

$$(1) \qquad \frac{dY_t}{dt} = \beta(\alpha - Y_t), \, Y_{t_0} = y_0,$$

where Y_t can be considered as a modified size, *i.e.*, $Y_t = h(X_t)$, where h is a known strictly increasing continuously differentiable function. Here $y_0 = h(x_0)$ and $\alpha = h(A)$, where x_0 is the observed size at t_0 (initial observation) and A represents the asymptotic size or size at maturity. The parameter $\beta > 0$ is the growth coefficient and represents the rate of approach to maturity. According to the choice of h in (1) we obtain well known deterministic models. For instance, when $h(x) = x$ we obtain the monomolecular model; if $h(x) = -x^{-1}$ we are in the case of the logistic model; when $h(x) = \ln x$ we have the Gompertz model and for $h(x) = x^c$, $c > 0$, we obtain the Bertalanffy-Richards model.

Individual growth of organisms must take into account environmental random fluctuations that affect the growth rate and autonomous stochastic differential equations (SDE) are quite adequate to describe this phenomenon. The SDE models we present are generalizations of the classical deterministic growth models (1) but incorporate a random dynamical term, describing the effects of environmental and other random fluctuations on the growth process. We use the model

$$(2) \qquad dY_t = \beta \left(\alpha - Y_t \right) dt + \sigma dW_t, \ \ Y(t_0) = y_0.$$

The intensity of the environmental random fluctuations is measured by the parameter $\sigma > 0$ and W_t is the standard Wiener process.

Here we will consider that the asymptotic size A is fixed (the same for all individuals) and the growth coefficient β is constant throughout the complete growth curve. This is a variant of the Ornstein-Uhlenbeck model also known in finance as the Vasicek model.

Some developments on the case where we assume that the parameter A varies randomly from individual to individual can be found in Braumann *et al.* (2009), resulting in a mixed-efects SDE model. In Filipe *et al.* (2010a) we have studied the generalization of model (2) to the multiphasic case, where we have assumed that the growth coefficient β takes different values for different phases of the individual's life.

The solution of (2), for $t \geq t_0$,

$$(3) \qquad Y_t = \alpha + e^{-\beta(t-t_0)}(y_0 - \alpha) + \sigma e^{-\beta t} \int_{t_0}^t e^{\beta s} dW_s$$

is an ergodic diffusion process with drift coefficient $\beta \left(\alpha - y \right)$ and diffusion coefficient σ^2, and follows a gaussian distribution with mean $\alpha + e^{-\beta(t-t_0)}(y_0 - \alpha)$ and variance $\sigma^2(1 - e^{-2\beta(t-t_0)})/(2\beta)$, which converges, as $t \to +\infty$, to a gaussian distribution with mean α and variance $\sigma^2/(2\beta)$. For $t_{k-1} < t_k$ and $\delta_{k-1} = t_k - t_{k-1}$, given $Y_{t_{k-1}} = y_{k-1}$, we see that $Y_{t_k} = \alpha + e^{-\beta \delta_{k-1}}(y_{k-1} - \alpha) + \sigma e^{-\beta t_k} \int_{t_{k-1}}^{t_k} e^{\beta s} dW_s$ and the transition distribution is gaussian with mean and variance given by $\alpha + (y_{k-1} - \alpha) e^{-\beta \delta_{k-1}}$ and $\sigma^2(1 - e^{-2\beta \delta_{k-1}})/(2\beta)$, respectively.

Fitting and Prediction

Consider we wish to model Y as a function of t having observed (t_1, y_1), (t_2, y_2),..., (t_n, y_n), where y_k is the observed value of Y_{t_k}.

A classic regression model can be written using the expression $y_i = f(t_i, \theta) + \varepsilon_i$, where $f(t_i, \theta) = \alpha + e^{-\beta(t_i - t_0)}(y_0 - \alpha)$ is the solution of (1) with $\theta = (\alpha, \beta, y_0)$, and ε_i are *i.i.d.* $\mathcal{N}(0, \sigma_\varepsilon^2)$.

For fitting of Y values, as well as for prediction of future values of Y, we use the expression $\hat{y}_t = \hat{\alpha} + e^{-\hat{\beta}(t-t_0)}(\hat{y}_0 - \hat{\alpha})$, where $\hat{\alpha}$, $\hat{\beta}$ and $\hat{y}_0$ are least square parameters estimates.

The traditional assumption of regression models that observed deviations from the regression curve are independent at different ages is unrealistic when the deviations are due to environmental random fluctuations. SDE models are built to incorporate the dynamics of the growth process and the effect environmental random fluctuations have on such dynamics.

For the SDE models, we can not really talk about fitting the curve for ages $t = t_k$, because Y_{t_k} is the exact value given by (3) of the individual modified size and not, as in the regression case, a value measured with error for which we wish to estimate the true value. Here, for convenience, we will refer as fitting the estimate of the deterministic curve that we would obtain in the absence of random environmental fluctuations ($\sigma = 0$) if we start with size y_0 at time t_0. In our case, this is equivalent to the curve of the expected values of Y_t. So, the fitting of the growth curve can be given by $\hat{Y}_t = \hat{E}\left[Y_{t_0} \mid y_0\right] = \hat{\alpha} + (y_{t_0} - \hat{\alpha})\,e^{-\hat{\beta}(t-t_0)}$, where $\hat{E}$ is the estimate of the mathematical expectation and $\hat{\alpha}$ and $\hat{\beta}$ are maximum likelihood estimates based on the full set of available observations of Y.

For prediction, given the values of the process until instant t_k, Y_{t_1}, Y_{t_2},..., Y_{t_k}, we intend to predict the value Y_t for $t > t_k$

(4)
$$Y_t = \alpha + (Y_{t_k} - \alpha)\,e^{-\beta(t-t_k)} + \sigma e^{-\beta t} \int_{t_k}^{t} e^{\beta s} dW(s).$$

The exact observed sizes until instant t_k should be used to predict future sizes. Since Y_t is a Markov process, $E\left[Y_t \mid Y_{t_1}, \ldots, Y_{t_k}\right] = E\left[Y_t \mid Y_{t_k}\right]$. Conditional on $Y_{t_k} = y_k$, one can see from (4) that Y_t follows a gaussian distribution with mean $\alpha + (y_k - \alpha)\,e^{-\beta(t-t_k)}$ and variance $\frac{\sigma^2}{2\beta}\left(1 - e^{-2\beta(t-t_k)}\right)$. For prediction, we can use $\hat{Y}_t = \hat{E}\left[Y_t \mid Y_{t_k} = y_k\right] = \hat{\alpha} + (Y_k - \hat{\alpha})\,e^{-\hat{\beta}(t-t_k)}$, where $\hat{\alpha}$ and $\hat{\beta}$ are maximum likelihood estimates based on the first k observations. Considering that $\hat{Y}_t - Y_t$ is approximately gaussian, the 95% confidence interval for Y_t can be determined by $\hat{Y}_t - E\left[\hat{Y}_t - Y_t\right] \pm 1.96\sqrt{Var\left[\hat{Y}_t - Y_t\right]}$.

For illustration of the previous results, we have worked with data on the weight of one mertolengo cow randomly selected between a set of 97 animals. For this animal we had available 51 observations from birth till approximately 12 years of age. We have started by comparing the nonlinear regression (NLR) model with the SDE model in terms of fitting. Table 1, shows estimation results based on the data of the full trajectory. Comparing the values of the root of the mean square error (RMSE), shown in Table 2, we can see that the NLR model is better than the SDE model. This result was expected, since the least square method was used to obtain the estimates of the NLR model parameters, which minimizes the sum of the square errors, consequently the RMSE. However, there is only a slight diference between the two models and the NLR model has an additional parameter, y_0, while in the SDE model we use the actual known value $y_0 = h(x_0)$, with $x_0 = 26$ kg for our trajectory.

Table 1: Regression *vs* SDE: parameters estimates based on full data of the chosen animal

		y_0	A	β
Gompertz	NLR	26.9	409.44	1.40
	SDE		406.13	1.49
B-R	NLR	30.75	416.41	1.04
	SDE		407.36	1.22

In terms of prediction, one can see that the SDE model is much better than the NLR model. For further analysis of the quality of prediction, we used for estimation a subset of the data, leaving out the last 15 observations of the trajectory, (t_{36}, y_{36}), (t_{37}, y_{37}), ..., (t_{50}, y_{50}). We present the results for long-term (LT) prediction and step-by-step (SS) prediction at the ages left out of the estimation procedure. In the LT case, we predict future weights of the animal using estimates obtained based on the weights observed until age t_{35}. In each step of the SS case, we just predict the animal size at next age using the current size as a starting point and using maximum likelihood updated parameter estimates based on all observations up to and including the present time. The RMSE values obtained are shown in Table 2 for two models that have proved to be appropriate (see, for instance, Filipe and Braumann, 2007): Gompertz ($h(x) = \ln x$) and Bertalanffy-Richards with $c = 1/3$ ($h(x) = x^{1/3}$).

Table 2: NLR *vs* SDE: values of the root of the mean square error

RMSE		Fitting	LT prediction	SS prediction
Gompertz	NLR	45.4	68.9	59.9
	SDE	46.2	42.8	27.5
B-R	NLR	42.7	64.3	56.3
	SDE	44.9	38.9	27.3

Profit optimization

We study optimization issues for the mean profit from selling an animal, comparing two methodologies, one that consists in selling the animal when a certain optimal age is reached (independently of weight) and the other that consists in selling the animal when an optimal weight is achieved (independently of age). Let us start by the first case - optimization of the mean profit through the animals age.

The profit from selling an animal to the cattle market, L, can be computed as $L = V - C$ where V represents the selling price and C the costs of acquisition and animal raising. Let $X_{t_0} = x_0$ be the weight of the animal at the age of purchase t_0 (assumed known) and t the age at which the animal is to be sold. If we denote by P the price per kg of carcass, R the dressing proportion (reflects the amount of carcass in relation to the animals live weight = carcass weight/live weight, for mertolengo cattle typically about 0.5), C_1 the fixed costs (*e.g.*, purchase price of the animal, veterinary, transportation and commercialization costs) and $C_2 = c_2(t - t_0)$ the variable costs, supposed proportional to the time of animal raising, we can write the profit as $L_t = PRX_t - C_1 - C_2$. We consider the cases of the stochastic Gompertz model (SGM) and the stochastic Bertalanffy-Richards (SBRM), $X_t = h^{-1}(Y_t) = e^{Y_t}$ and $X_t = h^{-1}(Y_t) = Y_t^3$, respectively. Consequently, the profit can be expressed as a function of Y_t, $L_t = l_t(Y_t) = PRe^{Y_t} - C_1 - c_2(t - t_0)$ for the SGM and $L_t = l_t(Y_t) = PRY_t^3 - C_1 - c_2(t - t_0)$ for the SBRM. Since it is known the probability distribution of Y_t, putting $y_0 = \ln x_0$ for the SGM and $y_0 = x_0^{1/3}$ for the SBRM, the L_t probability density function can be obtained as follows

$$(5) \qquad f_{L_t}(u) \quad = \quad f_{Y_t}(l_t^{-1}(u)) \left| \frac{dl_t^{-1}(u)}{du} \right| =$$

$$(6) \qquad = \quad \frac{1}{\sqrt{2\pi \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-t_0)})}} \exp\left(-\frac{\left(l_t^{-1}(u) - \alpha - (y_0 - \alpha)e^{-\beta(t-t_0)} \right)^2}{2\frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-t_0)})} \right) \left| \frac{dl_t^{-1}(u)}{du} \right|$$

with $l_t^{-1}(u) = \ln\left(\frac{u + C_1 + c_2(t-t_0)}{PR} \right)$ for the SGM and $l_t^{-1}(u) = \left(\frac{u + C_1 + c_2(t-t_0)}{PR} \right)^{1/3}$ for the SBRM. For the particular case of SGM, we can easily verify that L_t follows a shifted log-normal distribution.

The mean and variance of L_t are given by

$$(7) \qquad E[L_t] = PRE[X_t] - C_1 - c_2(t - t_0) \text{ and } Var[L_t] = P^2 R^2 Var[X_t].$$

Assuming the initial weight x_0 known, using the properties of the log-normal distribution for the SGM we have obtained

$$E[X_t] = E\left[e^{Y_t}\right] = e^{E[Y_t] + \frac{Var[Y_t]}{2}} \text{ and } Var[X_t] = Var\left[e^{Y_t}\right] = e^{2E[Y_t] + 2Var[Y_t]} - e^{2E[Y_t] + Var[Y_t]},$$

and applying Stein's Lemma for the SBRM case we get

$$E[X_t] = E\left[Y_t^3\right] = 3E[Y_t] Var[Y_t] + E[Y_t]^3$$

and
$$Var\left[X_t\right] = Var\left[Y_t^3\right] = 36E^2\left[Y_t\right]Var^2\left[Y_t\right] + 9E^4\left[Y_t\right]Var\left[Y_t\right] + 15Var^3\left[Y_t\right].$$

Let us consider the situation where a mertolengo cow raised with the mother is bought by a producer for 200 euros, at 7 months (0.58 years) of age (approximate weaning age) and 160kg, to be sold at age t. The usual t for market sale is 16 months (1.33 years). What is the expected profit of this producer? We must consider, in the case of mertolengo cattle breed, that the dressing proportion is 50% of live weight ($R = 0.5$); the usual raising costs (in euros) for an animal from the age of 7 months to age t are: 18.85 for commercialization and transportation, 26.68 for feeding/month, 7.25 for sanitation costs and 1.55 for other costs; and we consider typical selling prices P (euros/kg) of the animal to be 3.25, 3.50 or 3.75 euros. The mean and variance of the profit from selling the animal at age t can then be determined using (7), considering now $C_1 = 200 + 18.85 + 7.25 + 1.55 = 227.45$ and $c_2 = 26.68/\text{month} \times 12\text{months/year} = 320.16/\text{year}$.

Table 3 shows the optimal age (t_{opt}) for selling the animal in order to obtain a maximum mean profit. The correspondent standard-deviation of the profit is shown. To obtain these results we have used maximum likelihood estimates of the parameters, based on the complete data set available.

Table 3: Maximum mean profit (in euros), correspondent optimal selling age (in years) and standard-deviation (in euros)

P	3.25			3.50			3.75		
	t_{opt}	$E[L_{t_{opt}}]$	$sd[L_{t_{opt}}]$	t_{opt}	$E[L_{t_{opt}}]$	$sd[L_{t_{opt}}]$	t_{opt}	$E[L_{t_{opt}}]$	$sd[L_{t_{opt}}]$
SGM	1.05	62.19	109.6	1.13	96.94	127.00	1.19	133.26	143.72
SBRM	0.86	39.73	47.7	0.97	68.49	61.21	1.07	99.63	72.99

Using the previous approach we can determine the best age at which the animal is to be sold in order to optimize the mean profit. However, market demands may be others, such as the demand for animals with a certain specific weight. In this case, it is important to be able to determine the average time required for the animal to reach that weight. For this, the theory of first passage times will find here a very interesting application.

Denote by Q^* the upper threshold for the size X_t of the animal. Since $Y_t = h(X_t)$, the time required for an animal to reach a certain size Q^* is equivalent to the first passage time $T_Q = \inf\{t > 0 : Y_t = Q\}$ of Y_t by $Q = h(Q^*)$. Assume $-\infty < y_0 < Q < +\infty$, with Q in the interior of the state space of Y. In Braumann *et al.* (2009), we present details on obtaining, for our class of SDE models, the expressions for the mean and variance of T_Q:

(8)
$$E[T_Q | Y(0) = y_0] = \frac{1}{\beta} \int\limits_{\sqrt{2\bar{\beta}(y_0 - \alpha)/\sigma}}^{\sqrt{2\bar{\beta}(Q - \alpha)/\sigma}} \frac{\Phi(y)}{\phi(y)} dy$$

and

(9)
$$Var[T_Q | Y(0) = y_0] = \frac{2}{\beta^2} \int\limits_{\sqrt{2\bar{\beta}(y_0 - \alpha)/\sigma}}^{\sqrt{2\bar{\beta}(Q - \alpha)/\sigma}} \frac{1}{\phi(z)} \int\limits_{-\infty}^{z} \frac{\Phi^2(x)}{\phi(x)} dx dz,$$

where Φ and ϕ are the distribution function and the probability density function of a standard normal random variable. The values of the mean and variance of T_Q are obtained by numerical integration of (8) and (9).

The profit from selling the animal when a certain weight Q^* is achieved, is $L_{Q^*} = PRQ^* - C_1 - c_2 T_Q$. Consequently, the mean and variance for the profit is given by $E[L_{Q^*}] = PRQ^* - C_1 - c_2 E[T_Q|Y(0) = y_0]$ and $Var[L_{Q^*}] = c_2^2 Var[T_Q|Y(0) = y_0]$.

We have considered the same situation described above, but now, we have computed the optimal selling weight Q^*_{opt} of the animal, in order to maximize the profit. Results are shown on Table 4.

Table 4: Maximum mean profit (in euros), correspondent optimal selling weight (in kg) and standard-deviation (in euros)

P	3.25			3.50			3.75		
	Q^*_{opt}	$E[L_{Q^*_{opt}}]$	$sd[L_{Q^*_{opt}}]$	Q^*_{opt}	$E[L_{Q^*_{opt}}]$	$sd[L_{Q^*_{opt}}]$	Q^*_{opt}	$E[L_{Q^*_{opt}}]$	$sd[L_{Q^*_{opt}}]$
SGM	292	68.32	68.79	309	106.00	79.85	324	145.40	91.13
SBRM	232	41.74	51.22	256	72.29	64.62	275	105.5	76.37

We can compare the two methodologies presented for optimization of the mean profit by comparing the values of Tables 3 and 4. We can observe that, for the typical market values the second methodology is preferable, since it allows a higher optimal profit than the first one and, in the SGM case, even with a lower standard deviation of that optimal profit (for the SBRM the standard-deviation is slightly higher in the second methodology).

REFERENCES

Filipe, P.A., Braumann, C.A. and Roquete, C.J. (2010a). Multiphasic individual growth models in random environments. Methodology and Computing in Applied Probability (accepted; online 2010.03.27, DOI 10.1007/s11009-010-9172-0)

Filipe, P.A., Braumann, C.A., Brites, N.M. and Roquete, C.J. (2010b). Modelling Animal Growth in Random Environments: An Application Using Nonparametric Estimation. Biometrical Journal, 52(5): 653-666.

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

Filipe, P. A. and Braumann, C. A. (2007). Animal growth in random environments: estimation with several paths. Bull. International Statistical Institute LXII: 5806-5809.