

Animal Growth in Random Environments: Estimation with Several Paths

Patrícia A. Filipe  
Carlos A. Braumann

Bulletin of the International Statistical Institute, LXII  
(Proc. 56<sup>th</sup> Session of the International Statistical Institute)  
In press

# Animal Growth in Random Environments: Estimation with Several Paths

Filipe, Patrícia A.

*Universidade de Évora, CIMA (Centro de Investigação em Matemática e Aplicações)*  
*Rua Romão Ramalho 59*  
*7000-671 Évora, Portugal*  
*E-mail: pasf@uevora.pt*

Braumann, Carlos A.

*Universidade de Évora, CIMA (Centro de Investigação em Matemática e Aplicações)*  
*E-mail: braumann@uevora.pt*

## 1. Introduction

Classical models for the growth of an individual animal (or plant) in terms of its size  $X(t)$  (some measure of weight, volume or height) at time  $t$  have assumed the form of a differential equation

$$(1) \quad dY(t) = b(A - Y(t)) dt, \quad Y(t_0) = y_0,$$

with  $Y(t) = g(X(t))$ , where  $g$  is a strictly increasing function. We have  $y_0 = g(x_0)$ , where  $x_0$  is the size at birth, and  $A = g(a)$ , where  $a$  is the asymptotic size or size at maturity of the individual. The parameter  $b > 0$  is a rate of approach to maturity.

The *Bertalanffy-Richards model*, proposed by von Bertalanffy (1957) and also studied by Richards (1959), has been extensively used and corresponds to  $g(x) = x^c$  for  $c > 0$  and to  $g(x) = \ln x$  for  $c = 0$  (for  $c = 0$ , is also known as the Gompertz model, appropriate if one assumes growth to be basically a multiplicative process). The special case  $c = 1$  is also known as the *Mitscherlich model* and has been used in agriculture (see, for instance, Goldsworthy and Colegrove, 1974), particularly for linear measurements like length or height. If one, however, considers size measured as a volume or a weight,  $c = 1/3$  is quite a common choice (making  $Y(t)$  a kind of "length"). Other choices of  $c$  (including the choice providing the best adjustment) have been proposed.

The Bertalanffy-Richards model has been applied to animal growth data extensively. See, for instance: Freitas (2005), Mazini *et al.* (2003), Ohnishi and Akamine (2006), Oliveira, Lôbo, and Pereira (2000). For an application to tumor growth, see Kozusko and Bajzer (2003).

When one considers the effects of environmental random fluctuations on the growth process, it is natural to propose (see Garcia, 1983) the *stochastic differential equation* (SDE) model

$$(2) \quad dY(t) = b(A - Y(t)) dt + \sigma dW(t), \quad Y(t_0) = y_0,$$

where  $W(t)$  is a standard Wiener process and  $\sigma > 0$  measures the intensity of the effect of environmental fluctuations (internal and external) on growth. The solution is a homogeneous diffusion process with drift  $b(A - y)$  and diffusion coefficient  $\sigma^2$ . The solution is (see, for instance, Braumann, 2005)

$$(3) \quad Y(t) = A + e^{-bt}(y_0 - A) + \sigma e^{-bt} \int_0^t e^{bs} dW(s).$$

The distribution of  $Y(t)$  is Gaussian with mean  $A + e^{-bt}(y_0 - A)$  and variance  $\frac{\sigma^2}{2b}(1 - e^{-2bt})$ .

Usually, random variations in data have been treated by classical regression models. Regression models assume that the observed deviations from a deterministic curve are independent at different times. This would be a realistic assumption if the deviations were due to measurement errors, but it

is totally unrealistic when they are due to random changes on growth rates induced by environmental random fluctuations. For instance, in a regression model, a delay in growth at a certain time has no repercussions on future weights, making regression models inappropriate to model growth in a random environment. The SDE model (2) does not have these shortcomings.

In Patrícia, Braumann, and Roquete (2007), we have considered, for a single path (a single animal), the statistical problems of parameter estimation and of prediction of future population sizes for model (2) and have illustrated the methods with data on the weight of bovine growth. Section 2 gives a brief summary of the estimation part. Here, we extend (see Section 3) the estimation methods to the case of several paths (several animals of the same type raised under similar conditions), assumed to be independent, and also illustrate with bovine data.

The data in the illustrations, provided by Carlos Roquete (ICAM-University of Évora), is from "mertolengo" cattle of the "rosilho" strand raised in the "Herdade da Abóboda" in the Serpa region, at the left margin of the Guadiana river. The animals were raised in pasture, together with their mothers during nursing and later supplemented with silage when pasture is in shortage (Autumn and Winter).

## 2. Parameter estimation for a single path

Assume we observe the evolution of the size of one animal (one path) by measuring its size at the times (counted from birth)  $0 = t_0 < t_1 < \dots < t_n$  and want to estimate  $\mathbf{p} = (A, b, \sigma)$ . Let  $X_k = X(t_k)$  be the animal size at time  $t_k$  ( $k = 1, 2, \dots, n$ ) and let  $Y_k = Y(t_k) = g(X(t_k))$ . Let  $\mathbf{x} = (x_0, x_1, \dots, x_n)$  be the vector of observed values of  $\mathbf{X} = (X_0, X_1, \dots, X_n)$  and let  $\mathbf{y} = (y_0, y_1, \dots, y_n)$ , with  $y_k = g(x_k)$  ( $k = 1, 2, \dots, n$ ). We assume that  $g$  is a known function, so that we can compute the  $y_k$  ( $k = 1, 2, \dots, n$ ).

For  $k = 1, 2, \dots, n$ , one can see from (3) that  $Y_k = A + e^{-b(t_k - t_{k-1})}(Y_{k-1} - A) + \sigma e^{-bt_k} \int_{t_{k-1}}^{t_k} e^{bs} dW_s$ . Therefore, conditioned on having  $Y_{k-1} = y_{k-1}$ , the probability density function (p.d.f.) of  $Y_k$  is

$$f_{Y_k|Y_{k-1}=y_{k-1}}(y_k) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2b}(1 - e^{-2b\Delta t_k})}} \exp \left\{ -\frac{(y_k - A - (y_{k-1} - A)e^{-b\Delta t_k})^2}{2\frac{\sigma^2}{2b}(1 - e^{-2b\Delta t_k})} \right\},$$

where  $\Delta t_k = t_k - t_{k-1}$ . Since  $Y(t)$  is a Markov process, the joint density of  $Y_1 = Y(t_1), \dots, Y_n = Y(t_n)$  (given  $Y_0 = y_0$ , assumed known) is the product of the above conditional p.d.f. for  $k = 1, 2, \dots, n$  and so the log-likelihood function in terms of the  $Y$  variables is given by

$$(4) \quad L_Y(\mathbf{p}) = -\frac{n}{2} \ln \left( \frac{2\pi\sigma^2}{2b} \right) - \frac{1}{2} \sum_{k=1}^n \ln \left( 1 - e^{-2b\Delta t_k} \right) - \frac{b}{\sigma^2} \sum_{k=1}^n \frac{(y_k - A - (y_{k-1} - A)e^{-b\Delta t_k})^2}{1 - e^{-2b\Delta t_k}}.$$

In terms of the  $X$  variables, the log-likelihood is  $L_X(\mathbf{p}) = L_Y(\mathbf{p}) + \sum_{k=1}^n \ln \left( \frac{dY}{dX} \Big|_{x=x_k} \right)$ .

The *maximum likelihood estimator*  $\hat{\mathbf{p}} = (\hat{A}, \hat{b}, \hat{\sigma})$  is obtained by maximization of  $L_Y$  (equivalent to the maximization of  $L_X$ ), using numerical techniques (in the applications we have used the `nlminb` routine from the software S-PLUS). The estimators are asymptotically Gaussian with mean vector  $\mathbf{p}$  and variance-covariance matrix  $\Sigma = \mathbf{F}^{-1}$ , where  $\mathbf{F}$  is the Fisher information matrix with elements  $F_{ij} = -E[\partial^2 L_Y / \partial p_i \partial p_j]$ . The expressions of the  $F_{ij}$  can be explicitly obtained using the properties of the process  $Y(t)$ .  $\mathbf{F}$ , and therefore  $\Sigma$ , can be estimated by replacing  $\mathbf{p}$  by  $\hat{\mathbf{p}}$  on those expressions, thus allowing the construction of approximate confidence intervals for the parameters.

In Filipe, Braumann, and Roquete (2007), we have applied the stochastic Bertalanffy-Richards model, for the particular cases  $c = 0$  and  $c = 1/3$ , to the weight in Kg of a single animal for which we had 79 observations since birth till about 5 years of age. We have considered also the cases  $c = 1$  (not very appropriate for weight data) and  $g(x) = x^c$  with  $c$  unknown ( $\sigma$  also to be estimated from data); the last case is much more cumbersome (we can not use  $L_Y$  because we do not know  $g$ , and so we have to maximize  $L_X$ ) and the improvement over the cases  $c = 0$  and  $c = 1/3$  was not significant.

The estimated asymptotic variance-covariance matrices for  $\hat{\mathbf{p}}$  are

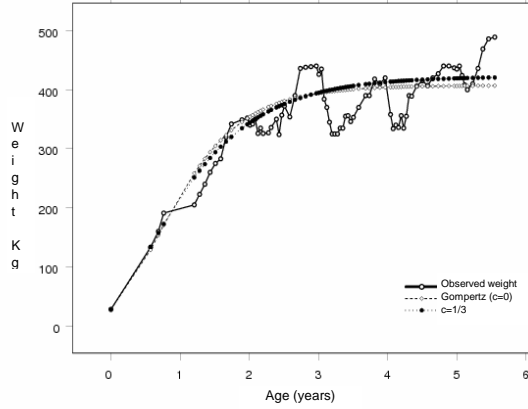
$$V_{c=0} = \begin{bmatrix} 0.00570 & -0.00680 & -0.00005 \\ -0.00680 & 0.03266 & 0.00024 \\ -0.00005 & 0.00024 & 0.00033 \end{bmatrix} \quad V_{c=1/3} = \begin{bmatrix} 1731.7 & -4.9989 & -0.0882 \\ -4.9989 & 0.0457 & 0.0008 \\ -0.0882 & 0.0008 & 0.0018 \end{bmatrix}.$$

Table 1 shows the maximum likelihood estimates (and the corresponding value of  $L_X$ ) together with the approximate 95% confidence intervals. We use the parameter  $a = g^{-1}(A)$  (average weight at maturity) so that we may compare the two cases; the other parameters,  $b$  and  $\sigma$ , are not comparable.

Figure 1 shows the graphs of the adjusted models  $c = 0$  and  $c = 1/3$  in the absence of environmental fluctuations ( $\sigma = 0$ ). Filipe, Braumann, and Roquete (2007) also studied the adjustment of the models in terms of their ability to predict future weights of the animal under study.

**Table 1. Maximum likelihood estimates and 95% confidence intervals (one animal)**

	for $a$	for $b$	for $\sigma$	$L_X$
$c = 0$ (Gompertz)	$407.1 \pm 60.5$	$1.472 \pm 0.354$	$0.2259 \pm 0.0355$	$-338.12$
$c = 1/3$	$422.4 \pm 81.6$	$1.096 \pm 0.525$	$0.5248 \pm 0.0827$	$-337.88$



**Figure 1. Estimated curves for  $c=0$  (Gompertz) and  $c=1/3$ , when  $\sigma = 0$**

### 3. Parameter estimation for several independent paths

Let us consider now several paths of the stochastic process corresponding to different animals of the same type and raised under similar conditions. Assume we have data on  $m$  animals. The size of animal number  $j$  ( $j = 1, 2, \dots, m$ ) is observed at the times (counted from birth)  $0 = t_{j0} < t_{j1} < \dots < t_{jn_j}$  and is, respectively,  $X_{j0} = X(t_{j0}), X_{j1} = X(t_{j1}), \dots, X_{jn_j} = X(t_{jn_j})$ . Let  $Y_{jk} = Y(t_{jk}) = g(X_{jk})$ . Let  $\mathbf{x}_j = (x_{j0}, x_{j1}, \dots, x_{jn_j})$  be the vector of observed values of  $\mathbf{X}_j = (X_{j0}, X_{j1}, \dots, X_{jn_j})$  and let  $\mathbf{y}_j = (y_{j0}, y_{j1}, \dots, y_{jn_j})$ , with  $y_{jk} = g(x_{jk})$  ( $j = 1, 2, \dots, m; k = 1, 2, \dots, n_j$ ). Assume that  $g$  is known.

For animal (trajectory) number  $j$  we can obtain its log-likelihood  $L_{Y_j}$  by proceeding as in (4):

$$(5) \quad L_{Y_j}(\mathbf{p}) = -\frac{n_j}{2} \ln \left( \frac{2\pi\sigma^2}{2b} \right) - \frac{1}{2} \sum_{k=1}^{n_j} \ln \left( 1 - e^{-2b\Delta t_{jk}} \right) - \frac{b}{\sigma^2} \sum_{k=1}^{n_j} \frac{(y_{jk} - A - (y_{j,k-1} - A) e^{-b\Delta t_{jk}})^2}{1 - e^{-2b\Delta t_{jk}}},$$

with  $\Delta t_{jk} = t_{jk} - t_{j,k-1}$ . From the independence, the overall log-likelihood for the  $m$  animals is

$$(6) \quad L_{Y_1, \dots, Y_m}(\mathbf{p}) = \sum_{j=1}^m L_{Y_j}(\mathbf{p}).$$

The *maximum likelihood estimator*  $\hat{\mathbf{p}}$  is obtained by maximization of  $L_{Y_1, \dots, Y_m}$ , again using numerical techniques. The estimators are asymptotically Gaussian with mean vector  $\mathbf{p}$  and variance-covariance matrix  $\Sigma = \mathbf{F}^{-1}$ . The Fisher information matrix  $\mathbf{F}$  is the sum of the Fisher information

matrices of the individual trajectories, which expressions we already know. Replacing  $\mathbf{p}$  by  $\hat{\mathbf{p}}$ , we can again obtain an estimation of  $\mathbf{\Sigma}$  and, therefore, approximate confidence intervals for the parameters.

We have applied the procedure for the stochastic Bertalanffy-Richards model, for the particular cases  $c = 0$  and  $c = 1/3$ , to  $m = 5$  animals of the same strand raised under similar conditions. One of them was the animal considered in the previous section (with 79 observations) and the other four animals have 38 observations each.

The estimated asymptotic variance-covariance matrices for  $\hat{\mathbf{p}}$  were

$$V_{c=0} = \begin{bmatrix} 0.00201 & -0.00257 & -0.00002 \\ -0.00257 & 0.00965 & 0.00008 \\ -0.00002 & 0.00008 & 0.00014 \end{bmatrix} \quad V_{c=1/3} = \begin{bmatrix} 555.65 & -1.7451 & -0.0311 \\ -1.7451 & 0.0115 & 0.0002 \\ -0.0311 & 0.0002 & 0.0006 \end{bmatrix}.$$

Table 2 shows the maximum likelihood estimates (and the corresponding value of  $L_{X_1, \dots, X_m}$ ) together with the approximate 95% confidence intervals.

**Table 2. Maximum likelihood estimates and 95% confidence intervals (five animals)**

	for $a$	for $b$	for $\sigma$	$L_{X_1, \dots, X_5}$
$c = 0$ (Gompertz)	$352.4 \pm 28.3$	$1.708 \pm 0.193$	$0.2534 \pm 0.0234$	-958.84
$c = 1/3$	$384.1 \pm 46.2$	$1.147 \pm 0.211$	$0.5062 \pm 0.0468$	-941.85

#### 4. Conclusions

Stochastic differential equation models for the growth of individual animals were considered and parameter estimation procedures were developed for the case of several trajectories (several animals), where observations of the animal size may be made at different times for different animals. An illustration is shown for cattle data provided by C. Roquete using the stochastic Bertalanffy-Richards model with  $c = 0$  (Gompertz model) and  $c = 1/3$ . Estimation procedures for the case where different animals may have different randomly chosen  $A$  parameters will appear shortly in an upcoming paper.

**Acknowledgment.** This work was developed in CIMA, a research center financed by FCT (Fundação para a Ciência e a Tecnologia) within its 'Programa de Financiamento Plurianual' (under FEDER funding).

#### REFERENCES

- Bertalanffy, L. von (1957). Quantitative laws in metabolism and growth. *The Quarterly Review of Biology*, 34 (3): 786-795.
- Braumann, C. A. (2005). *Introdução às Equações Diferenciais Estocásticas*. Edições SPE.
- Filipe, P. A., Braumann C. A., and Roquete, C. J. (2007). Modelos de crescimento de animais em ambiente aleatório. *Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística* (submitted).
- Freitas, A. (2005). Curvas de crescimento na produção animal. *Revista Brasileira de Zootecnia*, 35: 1843-1851.
- Garcia, O. (1983). A stochastic differential equation model for the height of forest stands. *Biometrics*, 39: 1059-1072.
- Goldsworthy, P. and Colegrove, M. (1974). Growth and yield of highland maize in Mexico. *J. Agriculture Science*, 83: 213-221.
- Kozusko, F. and Bajzer, Z. (2003). Combining gompertzian growth and cell population dynamics. *Mathematical Biosciences*, Vol. 185, p. 153-167.
- Mazini, A., Muniz, J., Aquino, L., and Silva, F. (2003). Análise da curva de crescimento de machos Hereford. *Ciência Agrotécnica*, 27 (5): 1105-1112.
- Ohnishi, S. and Akamine, T. (2006). Extension of von Bertalanffy growth model incorporating growth patterns of soft and hard tissues in bivalve molluscs. *Fisheries Science*, 72 (4): 787-795.
- Oliveira, H., Lôbo, R., and Pereira, C. (2000). Comparação de modelos não-lineares para descrever o crescimento de fêmeas da raça Guzerá. *Pesquisa Agropecuária Brasileira*, 35: 1843-1851.
- Richards, F. (1959). A flexible growth function for empirical use. *J. Experimental Botany*, 10: 290-300.