

INDIVIDUAL GROWTH MODELLING WITH SDE

Patrícia A. Filipe, Carlos A. Braumann, Nuno M. Brites, Gonçalo Jacinto



cima

PEst-OE/MAT/UI0117/2014



UNIVERSIDADE DE ÉVORA
ESCOLA DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE MATEMÁTICA



10 July 2018 | Bragança, Portugal

Model individual growth in random environments using stochastic differential equations (SDE).

Describe the evolution of bovine weight.

Previous research project using cattle weight data from several females of Mertolengo cattle breed of the rosilho phenotype.

Recent research project **GoBov+** - *Productivity improvement in the system of bovine raising for meat* - headed by INIAV and has the participation of the Universidade de Évora (ICAAM, CEFAGE, and CIMA), animal producers associations and other entities.

Data from males of Mertolengo (ACBM - www.mertolenga.com) and Alentejana (ACBRA - <http://www.bovinoalentejano.pt>) cattle breeds .

goBOV+



Why SDE to model individual growth?

- In Individual growth models, random variations in data are often treated using classical regression models
 - ▶ errors are independent. Static models. Prediction is based on the average curve;
- Model individual growth in a random fluctuating environment - Stochastic differential equations (SDE)
 - ▶ Incorporates the dynamics of the growth process and the effect that the environmental random fluctuations have on such dynamics. Predictions are based on the actual size.

X_t individual size (weight, height, length, etc.) at age t

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

- $Y_t = h(X_t)$ with h a strictly increasing C^1 function;
- $y_0 = h(x_0)$ where x_0 being the size at age t_0 (initial age);

Deterministic models

X_t individual size (weight, height, length, etc.) at age t

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

- $Y_t = h(X_t)$ with h a strictly increasing C^1 function;
- $y_0 = h(x_0)$ where x_0 being the size at age t_0 (initial age);

$$h(x) = -x^{-1}: \text{Logistic}$$

$$h(x) = x: \text{Monomolecular}$$

$$h(x) = \ln x: \text{Gompertz}$$

$$h(x) = x^c, c > 0: \text{Bertalanffy-Richards}$$

Deterministic models

X_t individual size (weight, height, length, etc.) at age t

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

- $Y_t = h(X_t)$ with h a strictly increasing C^1 function;
- $y_0 = h(x_0)$ where x_0 being the size at age t_0 (initial age);
- $\alpha = h(A)$, where A is the asymptotic size or size at maturity;
- $\beta > 0$ is the growth coefficient;

Deterministic models

X_t individual size (weight, height, length, etc.) at age t

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

- $Y_t = h(X_t)$ with h a strictly increasing C^1 function;
- $y_0 = h(x_0)$ where x_0 being the size at age t_0 (initial age);
- $\alpha = h(A)$, where A is the asymptotic size or size at maturity;
- $\beta > 0$ is the growth coefficient;

$$Y_t = \alpha + (y_0 - \alpha)e^{-\beta(t-t_0)} \xrightarrow{t \rightarrow +\infty} \alpha$$

$$X_t \xrightarrow{t \rightarrow +\infty} h^{-1}(\alpha) = A$$

$$\frac{dY_t}{dt} = \beta(\alpha - Y_t) + \sigma \varepsilon_t, \quad Y_{t_0} = y_0 \text{ known}$$

or

$$dY_t = \beta(\alpha - Y_t) dt + \sigma dW_t$$

- σ is an environmental noise intensity parameter;
- ε_s is a standard white noise;
- $W_t = \int_0^t \varepsilon_s ds$ is the standard Wiener process.

SDE models

Obtained the solution, distribution and transition density for the following models:

- Monophasic model with constant α and β ;
- Multiphasic model with constant α and different values of β for different phases of the animals life;
- Monophasic model with random α and constant β ;
- Monophasic model with random α and constant β , with the inclusion of genetic values.

The model solution is a homogeneous diffusion process with

drift coefficient $a(y) = \beta(\alpha - y)$
(mean speed of growth described by Y_t)

and

diffusion coefficient $b(y) = \sigma^2$
(a measure of the local magnitude of the fluctuations)

The explicit solution is given by

$$Y_t = \alpha + e^{-\beta(t-t_0)}(y_0 - \alpha) + \sigma e^{-\beta t} \int_{t_0}^t e^{\beta s} dW_s$$
$$\sim \mathcal{N}(\alpha + e^{-\beta(t-t_0)}(y_0 - \alpha), \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-t_0)})) \xrightarrow{t \rightarrow +\infty} \mathcal{N}(\alpha, \frac{\sigma^2}{2\beta}).$$

Transition density between s and t , ($s < t$), is gaussian with mean $\alpha + (y_s - \alpha) e^{-\beta(t-s)}$ and variance $\frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-s)})$

Multiphasic model

Assuming two different growth phases in the evolution of the growth curve

$$dY_t = \beta_t (\alpha - Y_t) dt + \sigma dW_t, \quad Y_{t_0} = y_0, \quad \text{com } \beta_t = \begin{cases} \beta_1, & t \leq u \\ \beta_2, & t > u. \end{cases}$$

where u is the phase switching age.

β_1 se $t \leq u$ e β_2 se $t > u$. u : idade de transição de fase.

The solution is given by:

$$Y_t = \begin{cases} \alpha + e^{-\beta_1(t-t_0)}(y_0 - \alpha) + \sigma e^{-\beta_1 t} \int_{t_0}^t e^{\beta_1 s} dW_s, & t \leq u \\ \alpha + e^{-\beta_2(t-u)}(Y_u - \alpha) + \sigma e^{-\beta_2 t} \int_u^t e^{\beta_2 s} dW_s, & t > u \end{cases}$$

Y_t is now a non-homogeneous diffusion process with drift coefficient $a(t, y) = \beta_t(\alpha - y)$ and diffusion coefficient $b(t, y) = \sigma^2$

- For $t \leq u$

$$Y_t \sim \mathcal{N}(\alpha + e^{-\beta_1(t-t_0)}(y_0 - \alpha), \frac{\sigma^2}{2\beta_1}(1 - e^{-2\beta_1(t-t_0)}))$$

- For $t > u$, the distribution of Y_t is Gaussian with mean

$$\alpha + e^{-\beta_1(u-t_0) - \beta_2(t-u)}(y_0 - \alpha)$$

and variance

$$\frac{\sigma^2}{2\beta_2}(1 - e^{-2\beta_2(t-u)}) + e^{-2\beta_2(t-u)} \frac{\sigma^2}{2\beta_1}(1 - e^{-2\beta_1(u-t_0)}).$$

The asymptotic distribution of Y_t is therefore Gaussian with mean α and variance $\frac{\sigma^2}{2\beta_2}$.

Monophasic model with random α and constant β

$$\alpha \sim \mathcal{N}(\mu, \theta^2)$$

$$Y_t | \alpha \sim \mathcal{N} \left(\alpha + e^{-\beta(t-t_0)}(y_0 - \alpha), \frac{\sigma^2}{2\beta} \left(1 - e^{-2\beta(t-t_0)}\right) \right)$$

Thus Y_t follows a gaussian distribution with mean

$$\mu + e^{-\beta(t-t_0)}(y_0 - \mu)$$

and variance

$$\underbrace{\frac{\sigma^2}{2\beta} \left(1 - e^{-2\beta(t-t_0)}\right)}_{\text{contribution of environmental random fluctuations}} + \underbrace{\theta^2 \left(1 - e^{-\beta(t-t_0)}\right)^2}_{\text{contribution of the asymptotic size differences}} .$$

$$\alpha = c_0 + c_1 VG_1 + c_2 VG_2 + \dots + c_k VG_k + \varepsilon$$

follows a gaussian distribution with mean
 $c_0 + c_1 vg_1 + c_2 vg_2 + \dots + c_k vg_k$ and variance σ_ε^2

The results are similar to the previous case replacing

$$\mu \text{ by } c_0 + c_1 vg_1 + c_2 vg_2 + \dots + c_k vg_k$$

and
 $\theta \text{ by } \sigma_\varepsilon^2$

Genetic Values for growth capacity, maternal capacity, interval between births, average daily weight gain, residual feed intake, feed conversion index and reproductive longevity.

For each model:

- Determined the log-likelihood function;
- Computed:
 - ▶ the maximum likelihood estimates of the parameters;
 - ▶ the limits of the asymptotic confidence intervals based on the asymptotic variance-covariance matrix (inverse of the empirical Fisher information matrix).

MLE of the parameters vector $\mathbf{p} = (\alpha, \beta, \sigma)$. The log-likelihood function for individual j

$$L_{\mathbf{Y}_j}(\mathbf{p}|\mathbf{y}_j) = -\frac{n_j}{2} \ln(2\pi) - \frac{n_j}{2} \ln\left(\frac{\sigma^2}{2\beta}\right) - \frac{1}{2} \sum_{k=1}^{n_j} \ln\left(1 - E_{j,k}^{2\beta}\right) - \frac{\beta}{\sigma^2} \sum_{k=1}^{n_j} \frac{\left(y_{j,k} - \alpha - (y_{j,k-1} - \alpha) E_{j,k}^\beta\right)^2}{1 - E_{j,k}^{2\beta}}$$

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

$$L_{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m}(\mathbf{p}|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m) = \sum_{j=1}^m L_{\mathbf{Y}_j}(\mathbf{p}|\mathbf{y}_j).$$

Maximum likelihood estimation - Biphasic model

MLE of the parameters vector $\mathbf{p}^\diamond = (\alpha, \beta_1, \beta_2, \sigma)$. The global log-likelihood function

$$L_{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m}^\diamond(\mathbf{p}^\diamond | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m) = \sum_{j=1}^m L_{\mathbf{Y}_j}^\diamond(\mathbf{p}^\diamond | \mathbf{y}_j), \text{ where}$$

$$L_{\mathbf{Y}_j}^\diamond(\mathbf{p}^\diamond | \mathbf{y}_j) = \sum_{k=1}^{n_{j,u}} L_{\mathbf{Y}_j}(\mathbf{p}^{\beta=\beta_1} | \mathbf{y}_j) + L_{\mathbf{Y}_j}^{1 \curvearrowright 2}(\mathbf{p}^\diamond | \mathbf{y}_j) + \sum_{n_{j,u}+2}^{n_j} L_{\mathbf{Y}_j}(\mathbf{p}^{\beta=\beta_2} | \mathbf{y}_j)$$

with

$$L_{\mathbf{Y}_j}^{1 \curvearrowright 2}(\mathbf{p}^\diamond | \mathbf{y}_j) = -\frac{1}{2} \left[\ln(\pi\sigma^2) + \ln \left(\frac{G_{j,n_{j,u}+1}^2(1-F_{j,n_{j,u}}^2)}{\beta_1} + \frac{1-G_{j,n_{j,u}}^2}{\beta_2} \right) \right] \\ - \frac{1}{\sigma^2} \frac{(y_{j,n_{j,u}+1} - \alpha - (y_{j,n_{j,u}} - \alpha)F_{j,n_{j,u}+1}G_{j,n_{j,u}+1})^2}{\left(\frac{G_{j,n_{j,u}+1}^2(1-F_{j,n_{j,u}+1}^2)}{\beta_1} + \frac{1-G_{j,n_{j,u}+1}^2}{\beta_2} \right)}$$

where $F_{j,k} = e^{-\beta_1(u-t_{j,k-1})}$ and $G_{j,k} = e^{-\beta_2(t_{j,k}-u)}$.

MLE of the parameters vector $\mathbf{p}^\bullet = (\mu, \theta, \beta, \sigma)$. The global log-likelihood function

$$L_{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m}(\mathbf{p}^\bullet | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m) = \sum_{j=1}^m L_{\mathbf{Y}_j}(\mathbf{p}^\bullet | \mathbf{y}_j), \text{ where}$$

$$\begin{aligned} L_{\mathbf{Y}_j}(\mathbf{p}^\bullet | \mathbf{y}_j) &= -\frac{n_j}{2} \ln(2\pi) - \frac{n_j}{2} \ln\left(\frac{\sigma^2}{2\beta}\right) - \frac{1}{2} \sum_{k=1}^{n_j} \ln(1 - E_{j,k}^{2\beta}) - \frac{1}{2} \ln(D_j) \\ &\quad - \frac{\beta}{D_j \sigma^2} \sum_{k=1}^{n_j} \frac{(y_{j,k} - \mu - (y_{j,k-1} - \mu) E_{j,k}^\beta)^2}{1 - E_{j,k}^{2\beta}} \\ &\quad + \frac{2\beta^2 \theta^2}{D_j \sigma^4} \left(\sum_{k=1}^{n_j} \frac{(y_{j,k} - y_{j,k-1} E_{j,k}^\beta)^2}{1 + E_{j,k}^\beta} \right)^2 - \frac{\beta (D_j - 1)}{D_j \sigma^2} \left(\sum_{k=1}^{n_j} \frac{(y_{j,k} - y_{j,k-1} E_{j,k}^\beta)^2}{1 - E_{j,k}^{2\beta}} \right) \end{aligned}$$

$$\text{with } D_j = 1 + \frac{2\beta\theta^2}{\sigma^2} \sum_{k=1}^{n_j} \frac{1 - E_{j,k}^\beta}{1 + E_{j,k}^\beta}.$$

Maximum likelihood estimation: results

$Y_t = h(X_t) = \ln X_t$ Stochastic Gompertz model (SGM)

$Y_t = h(X_t) = X_t^{1/3}$ Stochastic Bertalanffy-Richards model (SBRM)

MLE and approximate 95% confidence bands.

F	97 ACBM	
	SGM	SBRM
A	411.2 ± 8.1	425.7 ± 9.5
β	1.676 ± 0.056	1.181 ± 0.056
σ	0.302 ± 0.009	0.597 ± 0.019
L_X	-9758.0	-9565.2

M	16021 ACBM		9479 ACBRA	
	SGM	SBRM	SGM	SBRM
A	630.068 ± 6.375	886.532 ± 14.069	748.350 ± 5.841	935.176 ± 9.836
β	1.422 ± 0.012	0.811 ± 0.011	1.569 ± 0.011	0.964 ± 0.010
σ	0.340 ± 0.002	0.622 ± 0.031	0.336 ± 0.002	0.700 ± 0.004
L_X	-368198.124	-359433.329	-333272.32	-326207.67

Maximum likelihood estimation: results - Monophasic vs Biphasic ACBM Females

$u = 0.65$ (approx. 8 months)

	SGM	SBRM
A	418.4 ± 11.12	432.1 ± 11.98
β_1	1.773 ± 0.061	1.239 ± 0.063
β_2	1.176 ± 0.113	0.964 ± 0.101
σ	0.286 ± 0.009	0.584 ± 0.019
$L_{\mathbf{x}}^{\diamond}$	-9721.8	-9555.3

MLE and approximate 95% confidence bands.

Maximum likelihood estimation: results - Monophasic vs Biphasic ACBM Females

$u = 0.65$ (approx. 8 months)

	SGM	SBRM
A	418.4 ± 11.12	432.1 ± 11.98
β_1	1.773 ± 0.061	1.239 ± 0.063
β_2	1.176 ± 0.113	0.964 ± 0.101
σ	0.286 ± 0.009	0.584 ± 0.019
$L_{\mathbf{x}}^{\diamond}$	-9721.8	-9555.3

MLE and approximate 95% confidence bands.

Maximum likelihood estimation: results - Monophasic vs Biphasic ACBM Females

$u = 0.65$ (approx. 8 months)

	SGM	SBRM
A	418.4 ± 11.12	432.1 ± 11.98
β_1	1.773 ± 0.061	1.239 ± 0.063
β_2	1.176 ± 0.113	0.964 ± 0.101
σ	0.286 ± 0.009	0.584 ± 0.019
$L_{\mathbf{x}}^{\diamond}$	-9721.8	-9555.3

MLE and approximate 95% confidence bands.

Maximum likelihood estimation: results - Monophasic vs Biphasic ACBM Females

$u = 0.65$ (approx. 8 months)

	SGM	SBRM	SGM	SBRM
A	418.4 ± 11.12	432.1 ± 11.98	411.2 ± 8.1	425.7 ± 9.5
β_1	1.773 ± 0.061	1.239 ± 0.063	1.676 ± 0.056	1.181 ± 0.056
β_2	1.176 ± 0.113	0.964 ± 0.101		
σ	0.286 ± 0.009	0.584 ± 0.019	0.302 ± 0.009	0.597 ± 0.019
$L_{\hat{x}}^{\diamond}$	-9721.8^{**}	-9555.3^{**}	-9758.0	-9565.2

MLE and approximate 95% confidence bands.

Maximum likelihood estimation: results - Monophasic vs Biphasic Males

$u = 0.65$ (approx. 8 months)

	SGM	SBRM	SGM	SBRM
A	619.220 ± 8.860	924.145 ± 27.755	889.112 ± 13.365	1072.391 ± 20.230
β_1	1.391 ± 0.013	0.750 ± 0.015	1.469 ± 0.013	0.882 ± 0.0121
β_2	1.507 ± 0.027	0.807 ± 0.025	1.138 ± 0.022	0.790 ± 0.019
σ	0.341 ± 0.002	0.620 ± 0.003	0.326 ± 0.002	0.694 ± 0.004
$L_{\mathbf{x}}^{\diamond}$	-368977.23	-358652.22^{**}	-333307.27	-325989.22^{**}

MLE and approximate 95% confidence bands.

Maximum likelihood estimation: results - $\alpha \sim \mathcal{N}(\mu, \theta^2)$

F	SGM	SBRM
$h^{-1}(\mu)$	396.5 ± 13.5	423.3 ± 11.5
β	1.747 ± 0.071	1.192 ± 0.062
σ	0.299 ± 0.009	0.596 ± 0.019
θ	0.082 ± 0.033	0.082 ± 0.101
$L_{\mathbf{x}}^{\bullet}$	-9749.8^{**}	-9564.8^{ns}

M	ACBM		ACBRA	
	SGM	SBRM	SGM	SBRM
$h^{-1}(\mu)$	632.037 ± 6.791	906.958 ± 15.788	736.811 ± 6.423	949.157 ± 11.34
β	1.433 ± 0.012	0.807 ± 0.012	1.614 ± 0.012	0.966 ± 0.011
σ	0.331 ± 0.002	0.603 ± 0.003	0.322 ± 0.002	0.680 ± 0.004
θ	0.152 ± 0.007	0.492 ± 0.023	0.166 ± 0.006	0.449 ± 0.022
$L_{\mathbf{x}}^{\bullet}$	-367859.71^{**}	-359059.65^{**}	-332574.15^{**}	-325817.63^{**}

Maximum likelihood estimation:results - inclusion of genetic values

Growth capacity

	ACBM	ACBRA
c0	6,4647 ± 0,0109	6,6212 ± 0,0092
c1	0,0035 ± 0,0004	0,0030 ± 0,0005
se	0,1475 ± 0,0070	0,1623 ± 0,0063
b	1,4440 ± 0,0122	1,6170 ± 0,0120
s	0,3306 ± 0,0018	0,3220 ± 0,0018
L_X	-367738,4915**	-332490,7225**

Maternal capacity

	ACBM	ACBRA
c0	6,4497 ± 0,0108	6,5922 ± 0,0088
c1	0,0013 ± 0,0007	0,0051 ± 0,0009
se	0,1517 ± 0,0070	0,1625 ± 0,0063
b	1,4314 ± 0,0122	1,6132 ± 0,0119
s	0,3306 ± 0,0018	0,3220 ± 0,0018
L_X	-367853,0026**	-332511,6821**

- Modelling individual growth in a random fluctuating environment through a general and flexible class of SDE models;
- Application to real data from weight of Mertolengo and Alentejana cattle breeds;
- Besides the presented work, in the female case, we have also:
 - ▶ studied prediction issues;
 - ▶ Optimization of the mean profit;
- GoBov+: use of producers associations database on individual animals and their genetic values and weight observations, and of raising costs and selling prices, to adjust the models to male cattle of the mertolengo and the alentejana breeds. The use of the animal's available genetic values is expected to explain part of the observed variability among the animals' parameters and to provide individualized predictions with higher accuracy.

References



Braumann, C.A., Carlos, C., Filipe, P.A. e Roquete, C.J. (2009)

Growth of individuals in randomly fluctuating environments, Em *Proceedings of the 2009 International Conference in Computational and Mathematical Methods in Science e Engineering*, (Vigo-Aguiar J., Alonso P., Oharu S., Venturino E. e Wade B., eds.), Gijón, 201-212.



Casela G. e Berger, R. (2002)

Statistical Inference. Duxbury Resource Center



Filipe, P.A., Braumann, C.A e Carlos, C. (2015)

Profit optimization for cattle growing in a randomly fluctuating environment. *Optimization: A Journal of Mathematical Programming and Operations Research*, Volume 64, issue 6, pp. 1393-1407.



Filipe, P.A., Braumann, C.A, Brites, N.M. e Roquete, C.J. (2013).

Prediction for Individual Growth in a Random Environment. *Recent Developments in Modeling and Applications in Statistics, Studies in Theoretical and Applied Statistics*. Oliveira, P.E., Graça Temido, M., Henriques, C. e Vichi, M. (Eds.), Springer (ISBN: 978-3-642-32418-5), pp. 193-201.



Filipe, P.A., Braumann, C.A. e Roquete, C.J. (2012)

Multiphasic individual growth models in random environments. *Methodology and Computing in Applied Probability*, Volume 14, issue 1, pp. 49-56. DOI:10.1007/s11009-010-9172-0.

Some references...



Filipe, P.A., Braumann, C.A., Brites, N.M. e Roquete, C.J. (2010)

Modelling Animal Growth in Random Environments: An Application Using Nonparametric Estimation. *Biometrical Journal* 52(5):653-666.



Garcia, O. (1983)

A stochastic differential equation model for the height of forest stands, *Biometrics*, Vol. 39, 1059-1072.



Karlin, S. e Taylor, H.M. (1981)

A Second Course in Stochastic Processes. Academic Press.



Linetsky, V. (2004).

Computing Hitting Time Densities for OU and CIR Processes: Applications to Mean-reverting Models. *Journal of Computational Finance* 7, 1–22.



Yi, C. (2010).

On the first passage time distribution of an Ornstein-Uhlenbeck process. *Journal of Quantitative Finance*. 9 (10), 957–960.

OBRIGADA!
THANK YOU!