gSDE: a new software for stochastic growth models

Nuno M. Brites [1] — d7488@alunos.uevora.pt

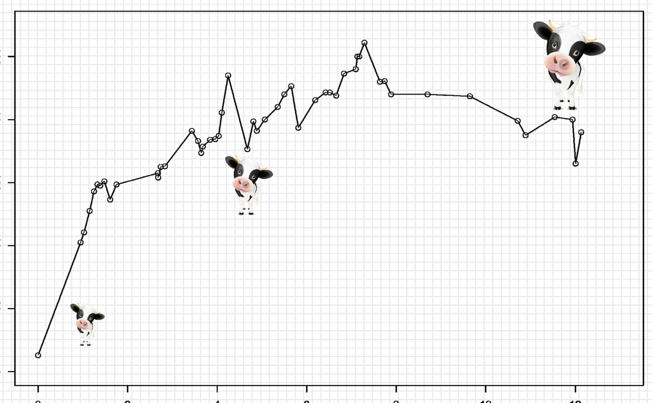
Carlos A. Braumann [1] — braumann@uevora.pt

Patrícia A. Filipe [1] — pasf@uevora.pt

Clara Carlos [1] — clara.carlos@estbarreiro.ips.pt

[1] Research Centre in Mathematics and Applications — University of Évora — Portugal

Model of animal growth in a random environment



$$dX_t = b(a - X_t)\,dt + s\,dW_t\ ,\quad X(0) = x_0$$

$$X(t) = h^{-1}(Y(t))$$

$Y_t = h(X_t)$

$$dY_t = b(A - Y_t)dt + s\,dW_t\ ,\quad Y(0) = y_0$$

$$Y(t) = A - (A - y_0)e^{-b(t-t_0)} + s e^{-bt}\int_{t_0}^{t} e^{bz}\,dW_z$$

X_t is the individual size (weight, length, height) at age t;
x_0 is the individual size at instant t_0 (initial age);
a is the asymptotic size of the individual;
b is the rate of approach to asymptotic size;
W_t is the standard Wiener process.

s measures the intensity of the environmental fluctuations effect on growth;
Y_t is the transformed individual size;
y_0 is the transformed initial size;
A is the mean asymptotic transformed size of the individual.

h is a strictly increasing continuosly differentiable function. Examples:

Gompertz model: $h(x) = \ln(x)$

Bertalanfly-Richards model: $h(x) = x^{1/3}$

User defined function: $h(x) = \ldots$

Monophasic model

ML Estimates

$$dY_t = b(A - Y_t)dt + s\,dW_t$$

$x_{j,t}$ = size of individual j at age t; $y_{j,t} = h\left(x_{j,t}\right)$;

Log-likelihood function for individual j:

$$L_{Y_j}(y_j;p) = \frac{n_j}{2}\ln\left(\frac{b}{\pi s^2}\right) - \frac{1}{2}\sum_{k=1}^{n_j}\ln\left(1-e^{-2b\delta_{j,k}}\right) - \frac{b}{s^2}\sum_{k=1}^{n_j}\frac{(y_{j,k} - A + (A - y_{j,k-1})e^{-b\delta_{j,k}})^2}{1-e^{-2b\delta_{j,k}}};$$

Log-likelihood function for m individuals:

$$L_{Y_1,\ldots,Y_m}(y_1,\ldots,y_m;p) = \sum_{j=1}^{m} L_{Y_j}(y_j;p),\ \text{where } \delta_{j,k} = t_{j,k} - t_{j,k-1} \text{ and } p = (A,b,s).$$

$$\hat{p} = (\hat{A},\hat{b},\hat{s})$$

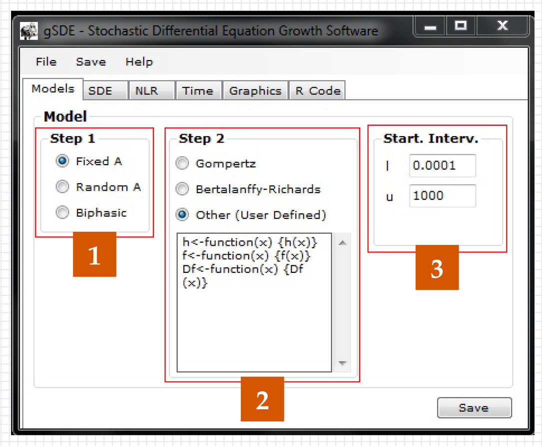
Fitting and Prediction

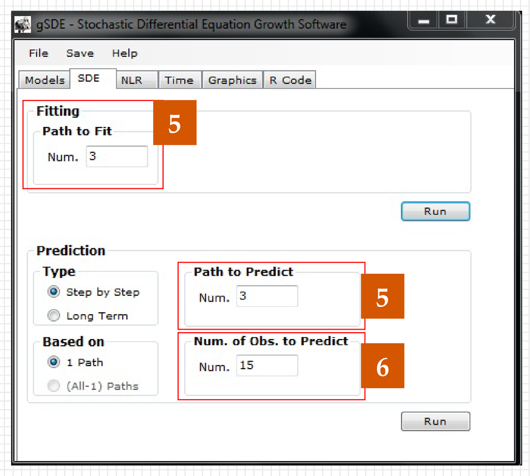
$$\hat{Y}_t = \hat{A} + (Y_{t^*} - \hat{A})e^{-\hat{b}(t-t^*)}$$

$$\text{for } t \geq t^* = \begin{cases}\text{initial age, for fitting;}\\ \text{age of last observation, for prediction.}\end{cases}$$

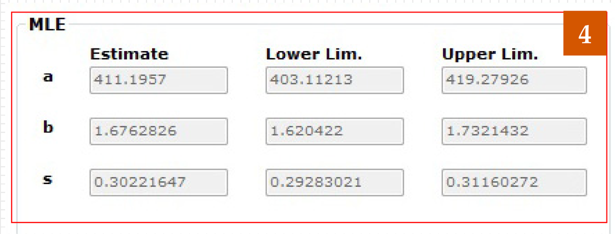
Features:
- Model fitting;
- Long-term predictions:
  - Based on 1 path;
  - Based on (All - 1) paths;
- Step-by-step prediction:
  - Based on 1 path.

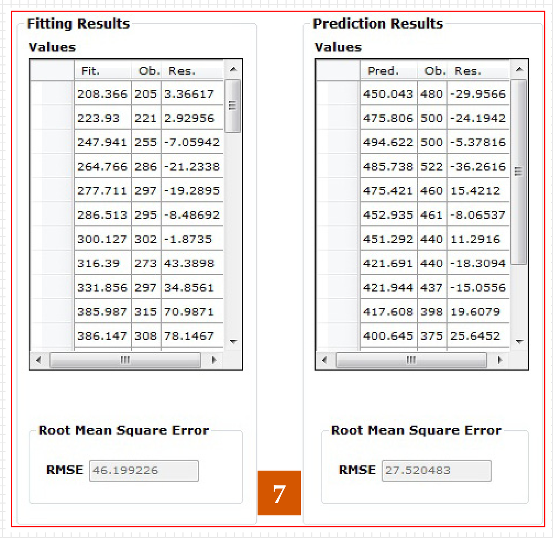
gSDE input





gSDE output





Legend

1 Main model choice;
2 Choice of h function;
3 Minimization starting points;
4 ML Estimates and 95% C.I.
5 Path to fit/predict;
6 Number of observations to fit/predict;
7 Fitted/predicted values and root mean square error.

Random A model

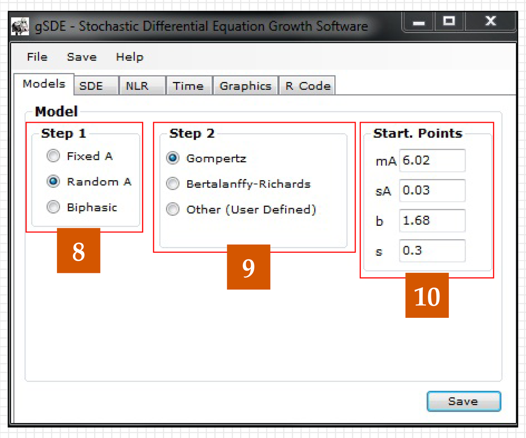
ML Estimates

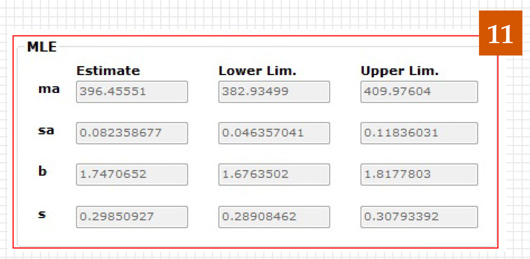
$$dY_t = b(A - Y_t)d_t + s\,dW_t$$

$$A \sim N(mA, sA)$$

- Assuming that diferente individuals have diferente average asymptotic sizes (due to genetic diferences or other factors).

$$\hat{p} = (\widehat{mA},\widehat{sA},\hat{b},\hat{s})$$





8 Random A model choice;
9 Choice of h function;
10 Starting points for mA, sA, b and s;
11 ML Estimates and 95% C.I.

Biphasic model

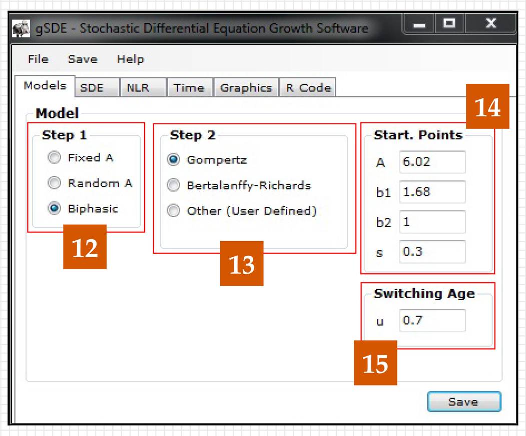
ML Estimates

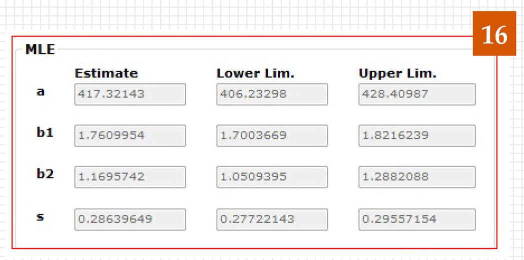
$$dY_t = b_t(A - Y_t)d_t + s\,dW_t$$

$$b_t = \begin{cases} b_1, t \leq u\\ b_2, t > u\end{cases}$$

- b_1 and b_2 are two diferente growth phases;
- u is the phase switching age.

$$\hat{p} = (\hat{A},\hat{b}_1,\hat{b}_2,\hat{s})$$





12 Biphasic model choice;
13 Choice of h function;
14 Starting points for A, b_1, b_2 and s;
15 Choice of phase switching age;
16 ML Estimates and 95% C.I.

Non-linear Regression

$$y_i = f(t_i \mid \xi) + \varepsilon_i;\quad \varepsilon_i \sim N(0, s_\varepsilon)\ \text{i.i.d.}$$

$$f(t_i \mid \xi) = A + e^{-b(t_i - t_0)}(y_0 - A);\ \xi = (A, b, y_0)$$

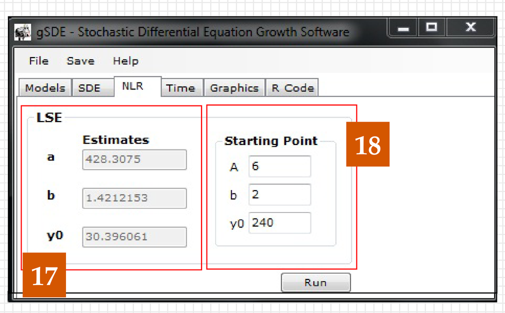
$$\hat{y}_i = f(t_i \mid \hat{\xi}) = \hat{A} + e^{-\hat{b}(t_i - t_0)}(\hat{y}_0 - \hat{A}),$$

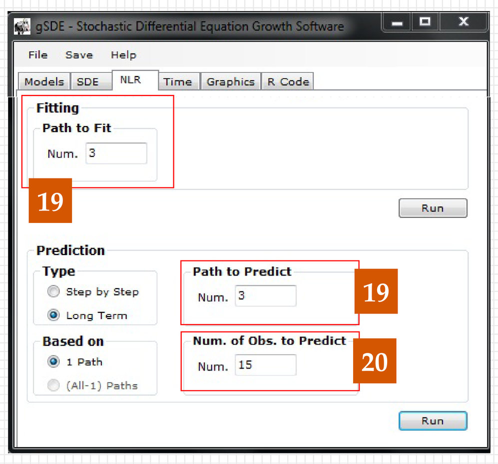
with

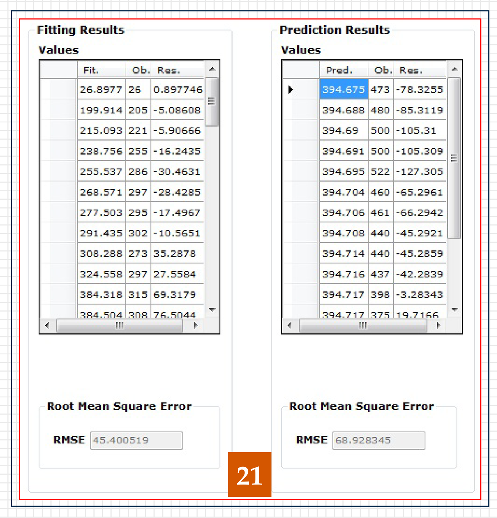
$$\hat{\xi} = (\hat{A},\hat{b},\hat{y}_0)$$

obtained by NLS.

- Traditional regression models are appropriate to model observational errors;
- They are inadequate to model random environmental variations;
- They do not keep memory of past sizes. The basis for prediction is the mean curve, not the current size.







17 Least square estimates;
18 Starting points for a, b and y_0;
19 Path to fit/predict;
20 Number of observations to fit/predict;
21 Fitted/predicted values and root mean square error.

Time to reach a given size

$$T_q = \inf\{t > 0 : X_t = q\} = \inf\{t > 0 : Y_t = Q := h(q)\}$$

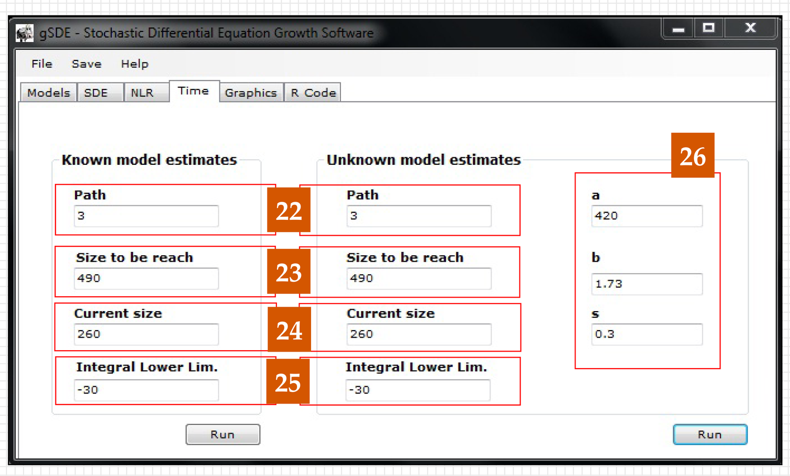
$$E[T_q \mid Y(0) = y_0] = \frac{1}{b}\int_{\xi}^{\eta}\frac{\Phi(z)}{\phi(z)}dz$$

$$Var[T_q \mid Y(0) = y_0] = \frac{1}{b^2}\int_{\xi}^{\eta}\frac{1}{\phi(z)}\int_{-\infty}^{z}\frac{\Phi^2(y)}{\phi(y)}dy\,dz$$

$\xi = \frac{\sqrt{2b}}{s}(y_0 - A)$ and $\eta = \frac{\sqrt{2b}}{s}(Q - A)$;

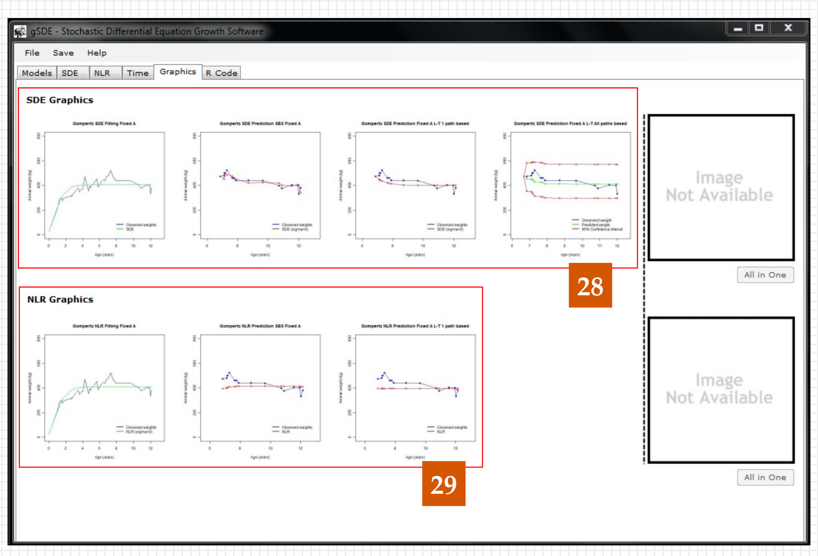
$\phi(.)$ is the probability density function of a standard Gaussian random variable;

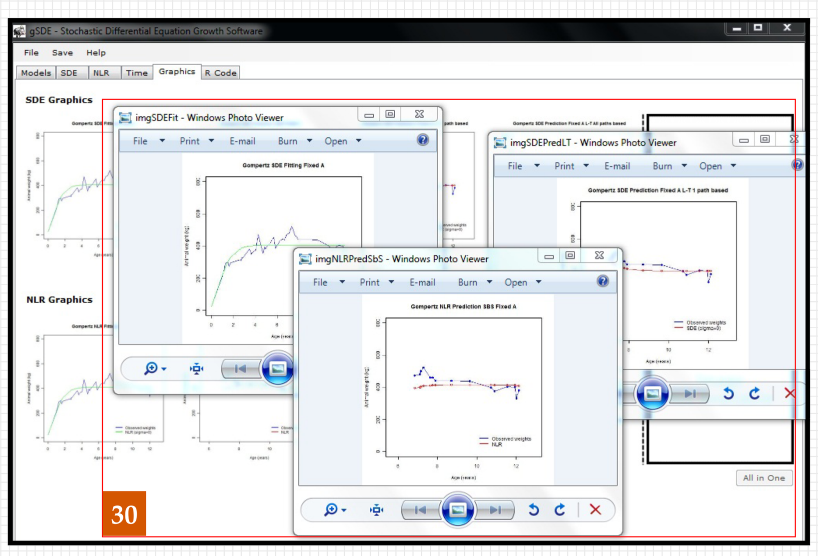
$\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

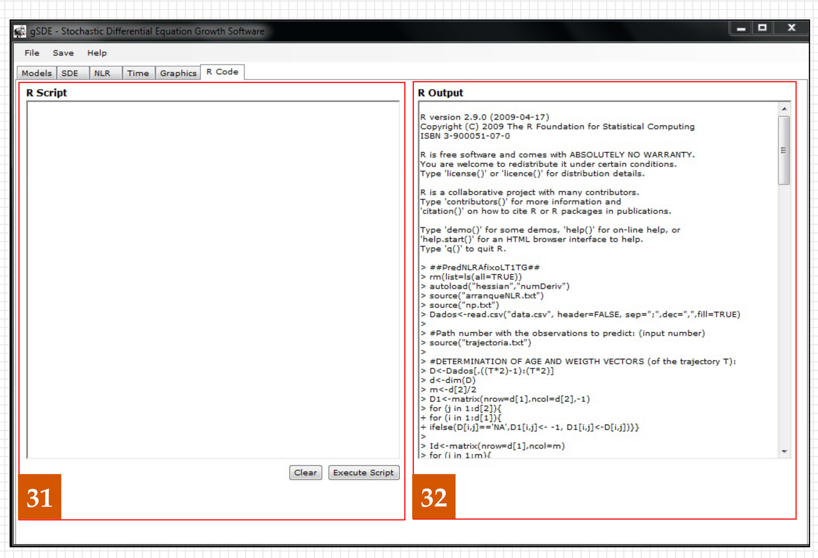


22 Individual to analyse;
23 Size to reach;
24 Individual current size;
25 Lower limit of SD expression;
26 Initial estimates;
27 Mean and SD of T_q.

Graphics and R code







28 SDE fitting/predition graphics;
29 NLR fitting/predition graphics;
30 Separated Windows graphics;
31 User R script;
32 R output.

Acknowledgments

The authors are members of the *Centro de Investigação em Matemática e Aplicações*, of the *Universidade de Évora*, financed by FCT (*Fundação para a Ciência e Tecnologia*).