Alicja Szabelska[1], Michał Siatkowski[2], Teresa Goszczurna[1], Joanna Zyprych[1]

[1]Department of Mathematical and Statistical Methods
Poznań University of Life Sciences
[2]Department of Agricultural Engineering
Poznań University of Life Sciences

COMPARISON OF GROWTH MODELS IN PACKAGE R

Summary. In this paper we present five growth models (growth curves; growth curve models): exponential, Gompertz, logistic, log-logistic and Weibull. For each growth curve the formula and graphical representation are described. Next, we introduce in package R methods for estimation of the model parameters and goodness of fit to the data criteria: Akaike Information Criterion and Bayesian Information Criterion. This criteria permit comparing and choosing the best model. In further part an example illustrating the process of estimating parameters and choosing the best model for a given dataset is presented.

Key words: Akaike Information Criterion, Bayesian Information Criterion, growth curve models, package R

Introduction

The growth curve models are a useful tool for modelling natural events that involve investigations of changes of response in time. In H~unt~ (1982) and H~anusz~ et al. (2008) are described the exemplary applications of the chosen growth curves in agricultural issues. For a given model we can consider time as a continuous variable instead of discrete one. The basic premise in growth curve models is that there exists a functional relationship between response and time. This relationship may be modelled. The models considered in this paper differ from each other by the form and number of parameters. This results in different shapes of the curves. Two questions should be answered when modelling such type of the process. Which growth curve should be used and what will be the values of parameters in the chosen model? These questions can be answered using mathematical methods, such as estimation.

Five models are presented explicitly in this paper. The goal is to provide the reader with various functions in package R for this kind of analysis and a ready-made code for

identification of the best model for specific sets of data. The utilization of growth curves in package R is supported by the example.

Material and methods

First, five growth curves are presented with graphical representation: exponential, Gompertz, logistic, log-logistic and Weibull. To stress the flexibility of the models we display graphs for two exemplary sets of parameters for each model.
Exponential model (Fig. 1):

$$f(x) = \alpha + (\beta - \alpha)\exp\left(\frac{-x}{\gamma}\right)$$
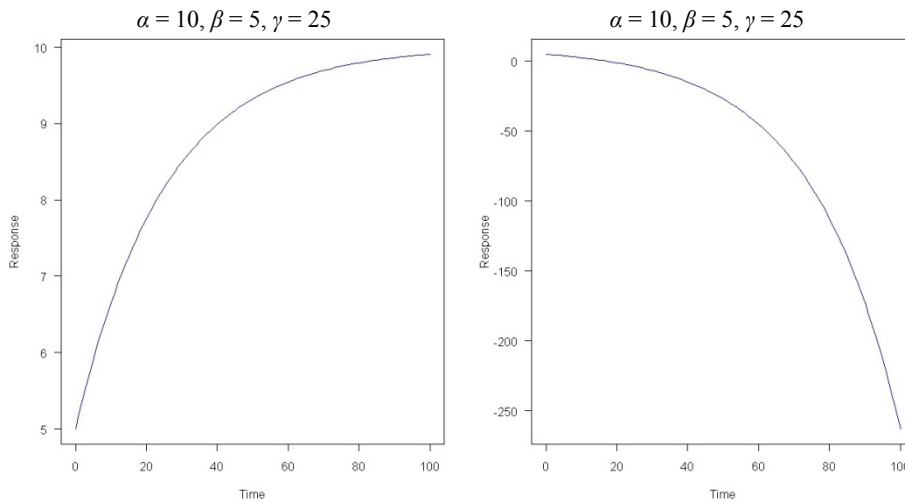

Fig. 1. Plots of exponential model for example two sets of parameters
Rys. 1. Wykres modelu wykładniczego dla dwóch przykładowych zestawów parametrów

Gompertz model (Fig. 2):

$$f(x) = \alpha + (\beta - \alpha)\exp(-\exp(\gamma(x - \delta)))$$

Logistic model (Fig. 3):

$$f(x) = \alpha + \frac{\beta - \alpha}{(1 + \exp(\gamma(x - \delta)))^{\varepsilon}}$$

Log-logistic model (Fig. 4):

$$f(x) = \alpha + \frac{\beta - \alpha}{(1 + \exp(\gamma(\ln x - \ln \delta)))^{\varepsilon}}$$

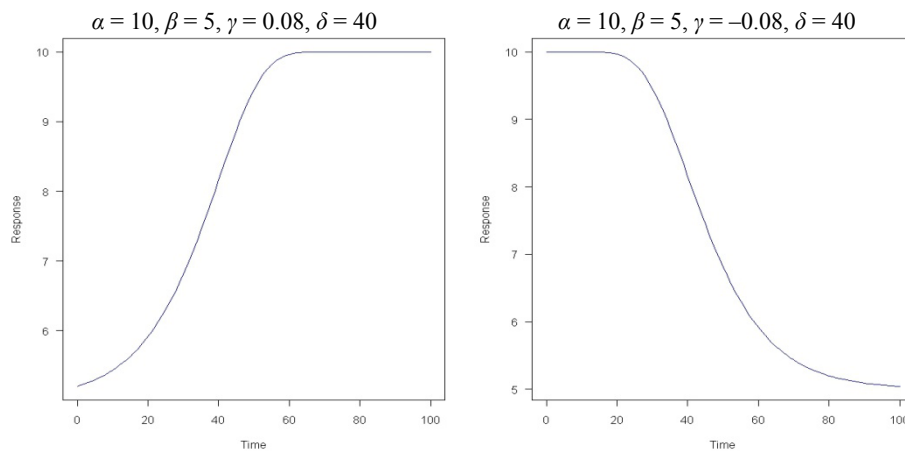$\alpha = 10, \beta = 5, \gamma = 0.08, \delta = 40$ $\alpha = 10, \beta = 5, \gamma = -0.08, \delta = 40$

Fig. 2. Plots of Gompertz model for example two sets of parameters
Rys. 2. Wykres modelu Gompertza dla dwóch przykładowych zestawów parametrów

$\alpha = 10, \beta = 5, \gamma = 0.15, \delta = 35, \varepsilon = 0.5$ $\alpha = 10, \beta = 5, \gamma = -0.15, \delta = 35, \varepsilon = 0.5$
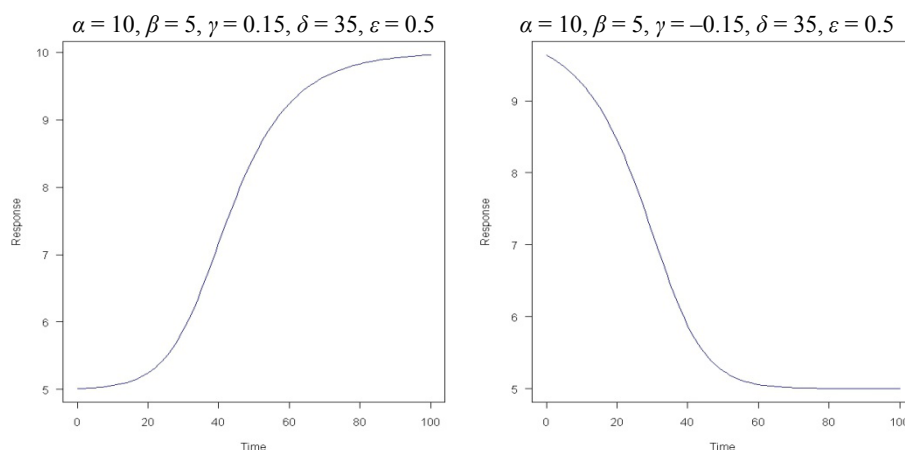

Fig. 3. Plots of logistic model for example two sets of parameters
Rys. 3. Wykres modelu logistycznego dla dwóch przykładowych zestawów parametrów

Weibull model (Fig. 5):

$$f(x) = \alpha + (\beta - \alpha)\exp(-\exp(\gamma(\ln x - \ln \delta)))$$

More information about the models can be found in works by SEBER and WILD (1989) and FINNEY (1979). To be able to use these growth curves it is required to find the values of the parameters in the model. For this purpose statistical package R (R: A LANGUAGE... 2009) can be used. In package R there is package `drc` (RITZ and STREIBIG 2005). There exists function `drm` that estimates parameters for all five models introduced above. The estimation is based on Nelder-Mead, quasi-Newton and conjugate-

$\alpha = 10, \beta = 5, \gamma = 7, \delta = 65, \varepsilon = 1.5$ $\alpha = 10, \beta = 5, \gamma = -7, \delta = 65, \varepsilon = 1.5$
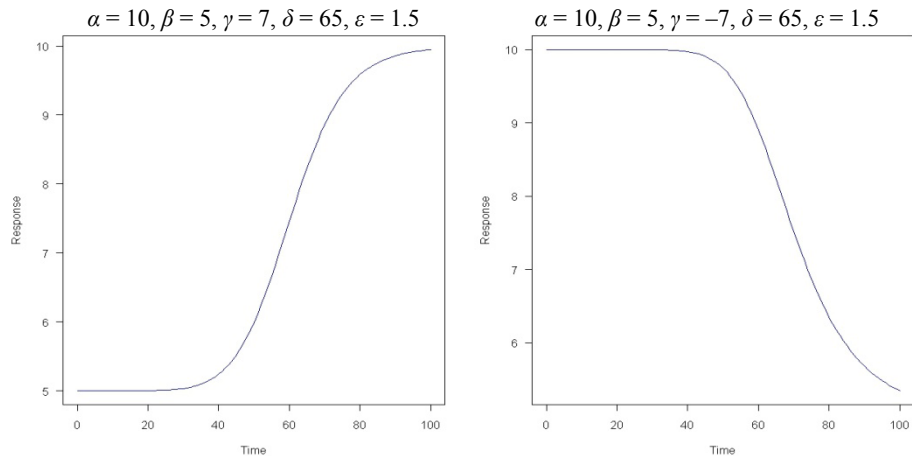
Fig. 4. Plots of log-logistic model for example two sets of parameters
Rys. 4. Wykres modelu log-logistycznego dla dwóch przykładowych zestawów parametrów

$\alpha = 10, \beta = 5, \gamma = 2.5, \delta = 60$ $\alpha = 10, \beta = 5, \gamma = -2.5, \delta = 60$
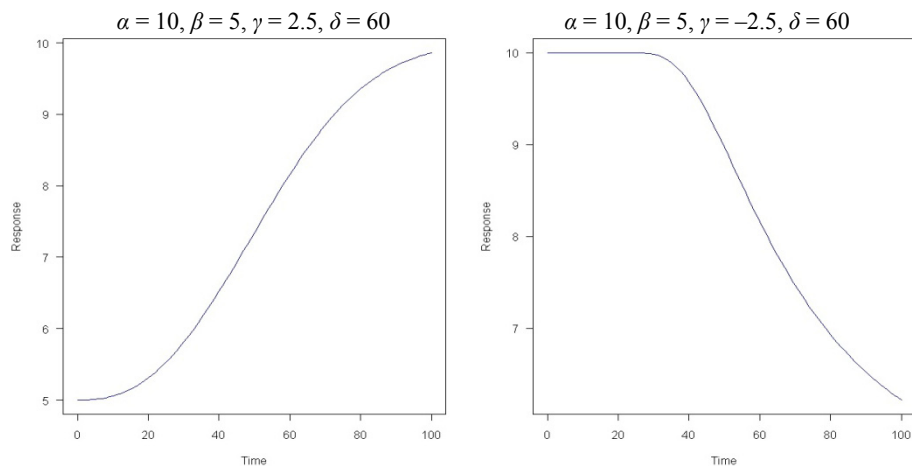
Fig. 5. Plots of Weibull model for example two sets of parameters
Rys. 5. Wykres modelu Weibulla dla dwóch przykładowych zestawów parametrów

-gradient algorithms (NOCEDAL and WRIGHT 1999). The utilization of this function was illustrated using two artificial datasets presented in Table 1 and Table 2.

Usage of `drm` function can be obtained in R by writing a following code:

```
library(drc)
dataset1<-read.table("dataset1.txt")
result.m_k<-drm(response~time, data = dataset1, fct = model_k)
summary(result.m_k)
plot(result.m_k)
```

Table 1. The first set of data
Tabela 1. Pierwszy zbiór danych

Time	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
Response	3	4	6	6	9	16	20	24	35	42	47	64	80	93	120
Time	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300
Response	133	161	174	178	192	196	206	211	213	221	224	229	229	230	232

Table 2. The second set of data
Tabela 2. Drugi zbiór danych

Time	1	5	9	13	17	21	25	29	33	37	41	45	49
Response	100	100	100	100	100	100	100	100	100	99	97	96	93
Time	53	57	61	65	69	73	77	81	85	89	93	97	101
Response	88	85	80	73	67	64	62	58	57	55	52	52	52

The first two lines of the code switch on the `drc` package and include data that will be used. In the third line the `drm` function is performed. The expression *model_k* written with italic font should be replaced with the appropriate name of a chosen growth curve model function from the `drc` package. The names of functions of the considered models are given in Table 3.

Table 3. Names of functions of growth curve models in `drc` package
Tabela 3. Nazwy funkcji modeli wzrostu w pakiecie `drc`

Name of model	Name of function
Exponential	`EXD.3()`
Gompertz	`G.4()`
Logistic	`L.5()`
Log-logistic	`LL.5()`
Weibull	`W1.4()`

The fourth line of the code presents values of estimated coefficients with additional information about the precision of the estimation procedure. The last line generates the plot of the model with estimated values, as well as with the data points. It is also worth mentioning that `drc` package contains more growth curves than the presented ones. Mostly, they are similar to one of the presented models, but with the reduced number of parameters. Readers can find more about other models in help of `drm` function in R.

When parameters for every model are specified we have to decide which model will be used as the mathematical description of the studied process in the experiment. Since all the models can be considered from statistical point of view, we can make the decision based on some statistical criteria. We use Akaike Information Criterion – AIC (AKAIKE 1974) and Bayesian Information Criterion – BIC (SCHWARZ 1978) as a tool of comparison between the models. Both of these values are standard measures of goodness of fit to nonlinear data. It is considered that the model with the lowest value of AIC or alternatively BIC is the most suitable one.

The AIC can be calculated in R by using of the function AIC from stats package as follows:

```
AIC(result.m_1, result.m_2, result.m_3, result.m_4, result.m_5)
```

The BIC can be derived with usage of the same function. However, the parameters of this function need to be used in the following form:

```
AIC(result.m_1, result.m_2, result.m_3, result.m_4, result.m_5,
k=log(nrow(dataset1))).
```

Results

The estimation of parameters for the five considered models and choice of the most suitable model was undergone for the sets of data from Table 1 and Table 2 separately. Table 4 gathers the values of the estimated parameters based on the first dataset.

Table 4. The estimates of parameters for the dataset 1
Tabela 4. Estymatory parametrów dla danych nr 1

Model				
exponential	Gompertz	logistic	log-logistic	Weibull
$\alpha = 1\,738.86$	$\alpha = -9.39$	$\alpha = -0.03$	$\alpha = -6.24$	$\alpha = -2.72$
$\beta = -39.19$	$\beta = 119.93$	$\beta = 2.17$	$\beta = 5.66$	$\beta = 12.00$
$\gamma = 1\,628.74$	$\gamma = -380\,969.66$	$\gamma = 231.86$	$\gamma = 234.66$	$\gamma = 267.75$
	$\delta = 78\,732.23$	$\delta = 141.66$	$\delta = 178.73$	$\delta = 140.47$
		$\varepsilon = 1.20$	$\varepsilon = 0.52$	

In Table 5 the results of estimation of parameters for the second dataset are presented.

Five so-called empirical models were obtained using estimated parameters, which we would like to compare. Thus, we calculated AIC and BIC to find the model with the best fit. The obtained values are presented in Table 6 and Table 7.

Table 6 and Table 7 reveal that the order of AIC and BIC values are the same for both datasets. This means that both coefficients indicate the same choice of the model. In the case of first dataset the lowest values of AIC, equal to 169.76, and BIC, equal to 178.17, has logistic model. However, for the second dataset log-logistic model indicates the best fit with values 53.01 for AIC and 60.56 for BIC. In Figure 6 the plots of the most accurate model with estimated parameters are presented. In addition, the plots contain the points from datasets 1 and 2, respectively.

Table 5. The estimates of parameters for the dataset 2
Tabela 5. Estymatory parametrów dla danych nr 2

Model				
exponential	Gompertz	logistic	log-logistic	Weibull
$\alpha = -13\ 964.57$ $\beta = 112.59$ $\gamma = 23\ 448.18$	$\alpha = -0.07$ $\beta = 99.90$ $\gamma = 48.82$ $\delta = 59.06$	$\alpha = 0.15$ $\beta = 49.01$ $\gamma = 100.18$ $\delta = 54.71$ $\varepsilon = 0.42$	$\alpha = 7.18$ $\beta = 48.59$ $\gamma = 100.07$ $\delta = 61.23$ $\varepsilon = 0.78$	$\alpha = 5.05$ $\beta = 52.97$ $\gamma = 100.50$ $\delta = 68.10$

Table 6. The values of AIC and BIC coefficients for the dataset 1
Tabela 6. Wartości współczynników AIC i BIC dla danych nr 1

Criterion	Model				
	exponential	Gompertz	logistic	log-logistic	Weibull
AIC	268.92	363.79	**169.76**	170.71	204.91
BIC	274.52	370.80	**178.17**	179.12	211.92

Table 7. The values of AIC and BIC coefficients for the dataset 2
Tabela 7. Wartości współczynników AIC i BIC dla danych nr 2

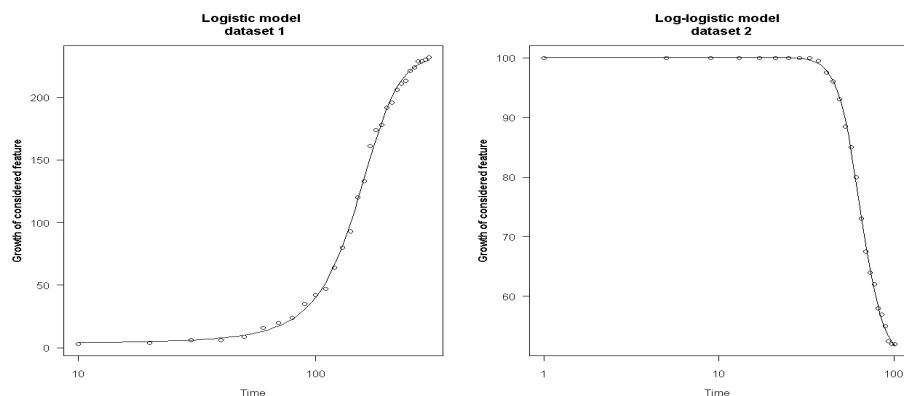| Criterion	Model				
	exponential	Gompertz	logistic	log-logistic	Weibull
AIC	177.98	57.98	54.01	**53.01**	78.89
BIC	183.02	64.27	61.56	**60.56**	85.18

Fig. 6. Models with the best fit for the dataset 1 and the dataset 2
Rys. 6. Najlepiej dopasowane modele dla danych nr 1 i danych nr 2

Conclusions

Considered five growth curves showed differential flexibility in representation of the data. The results showed that when modelling data from Table 1 and Table 2, logistic or log-logistic model should be used. For both of these models there is small difference in values of AIC and BIC. This result suggests that both of the curves could be chosen as satisfactory. Gompertz and Weibull models could also be acceptable in some cases. However, simple structure of the exponential model can lead to weak fit to data compromising its practical usefulness in this kind of problems.

The methods introduced in "Material and methods" section show the procedure of finding the most suitable model (with the appropriate parameters) to the data. Furthermore, the statistical issues of estimation of parameters and the choice of the model can be successfully and easily overtaken by usage of package R. In our opinion the proposed algorithm for detecting the description of the growth process with suitable growth curve model may find practical application in automating such kinds of analysis.

Acknowledgments

References

AKAIKE H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19: 716-723.

FINNEY D.J., 1979. Bioassay and the practise of statistical inference. Int. Stat. Rev. 47: 1-12.

HANUSZ Z., SIARKOWSKI Z., OSTROWSKI K., 2008. Zastosowania modelu Gompertza w inżynierii rolniczej. Inż. Roln. 7: 105.

HUNT R., 1982. Plant growth curves: the functional approach to plant growth analysis. Arnold, London.

NOCEDAL J., WRIGHT S.J., 1999. Numerical optimization. Springer, New York.

R: A LANGUAGE and environment for statistical computing. 2009. R Development Core Team. R Foundation for Statistical Computing, Vienna, Austria. [http://www.r-project.org].

RITZ C., STREIBIG J.C., 2005. Bioassay analysis using R. J. Stat. Softw. 12, 5: 1-22.

SCHWARZ G., 1978. Estimating the dimension of a model. Ann. Stat. 6: 461-464.

SEBER G.A.F., WILD C.J., 1989. Nonlinear regression. Wiley, New York.

PORÓWNANIE KRZYWYCH WZROSTU W PAKIECIE R

Streszczenie. W pracy przedstawiono pięć modeli krzywych wzrostu: wykładniczy, Gompertza, logistyczny, log-logistyczny i Weibulla. Korzystając z ogólnodostępnego pakietu statystycznego R, w szczególności z pakietu drc, przeprowadzono estymację parametrów rozpatrywanych modeli krzywych wzrostu. Zaprezentowano w ten sposób funkcje dopasowujące do danych. Ponadto wyznaczono wartości współczynników AIC oraz BIC, określające miarę dopasowania modeli do

danych empirycznych, umożliwiając wskazanie modelu najlepszego. Otrzymane wyniki przed-stawiono w postaci graficznej.

Słowa kluczowe: Informacyjne Kryterium Akaike, Bayesowskie Kryterium Informacyjne Schwar-za, krzywe wzrostu, pakiet R

Corresponding address – Adres do korespondencji:
Alicja Szabelska, Katedra Metod Matematycznych i Statystycznych, Uniwersytet Przyrodniczy w Poznaniu, ul. Wojska Polskiego 28, 60-637 Poznań, Poland, e-mail: aszab@up.poznan.pl