



Bioassay Analysis using R

Christian Ritz
Royal Veterinary and
Agricultural University

Jens C. Streibig
Royal Veterinary and
Agricultural University

Abstract

We describe an add-on package for the language and environment R which allows simultaneous fitting of several non-linear regression models. The focus is on analysis of dose response curves, but the functionality is applicable to arbitrary non-linear regression models. Features of the package is illustrated in examples.

Keywords: dose response data, multiple curves, non-linear regression.

1. Introduction

Bioassays are experiments with biologically active compounds. In herbicide selectivity studies it is common to run suites of bioassays with dose response curves for different plant species and/or different herbicide preparations. The same principles apply to the study of compounds in toxicology and pharmacology. In herbicide research and development, potencies of compounds usually are compared at some *a priori* response levels, say 50% reduction (ED50) in biomass or other response variables. For example, ED10 denotes 10% effect, that is 90% of the untreated control. To describe herbicide selectivity we compare tolerance of crops, say ED10, and sensitivity of weeds, say ED90. These ways of comparing compound potencies and/or species sensitivity also are widely used in ecotoxicology of xenobiotics.

Therefore, it is relevant to develop software which is capable of carrying out simultaneous non-linear regression analysis on several bioassays. In this paper we will demonstrate the use of the software package **drc** for analysis of multiple dose response curves. The functions in **drc** provide a convenient means for specifying models, controlling the minimization and retrieving relevant results, including comparisons of parameters of interest.

The package **drc** is an add-on package for the language and environment R (R Development Core Team 2004) which is open source and freely available (see <http://www.R-project.org>). R is an implementation of the language S.

The approach taken is to illustrate the functionality through several examples, where the function calls in R and the resulting output are interwoven into explanatory text. In this manner we present applications of the main functions contained in **drc**. We do not give an exhaustive description of the numerous facilities in **drc**. However, we wish to emphasize that various non-linear regression models can be fit using **drc**.

Section 2 introduces the relevant statistical models and some aspects related to the parameters in the models. An overview of the package **drc** is given in section 3. In sections 4, 5, 6, 7 and 8 we illustrate some of the features and functionality of **drc** for various aspects of fitting non-linear regression models with emphasis on bioassay analysis.

2. Non-linear regression models for bioassay data

We consider a suite of n dose response curves and we assume that for each dose response curve the response follows a non-linear curve specified by the function f , which is known apart from a parameter vector that may be different for different curves. Thus, the model for the i th dose response curve is

$$y_{ij} = f(x_{ij}, \alpha_i) + \varepsilon_{ij} \quad j = 1, \dots, m_i, \quad i = 1, \dots, n,$$

where x_{ij} denotes the j th dose value in the i th dose response curve and y_{ij} is the resulting response value, α_i is the unknown parameter vector for dose response curve i and ε_{ij} is the measurement error for the response y_{ij} . The measurement errors $\varepsilon_{11}, \dots, \varepsilon_{nm_n}$ are assumed mutually independent and normally distributed $N(0, \sigma^2)$. In particular, all observations have the same variance (variance homogeneity).

The parameters are estimated using non-linear least squares which amounts to minimizing the sum of squares

$$\sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - f(x_{ij}, \alpha_i))^2$$

with respect to the parameters $(\alpha_1, \dots, \alpha_n)$.

2.1. Commonly used models

Two models for sigmoidal dose response curves are widely used: The logistic model and the Gompertz model. We will briefly discuss these models.

The four-parameter logistic function is given by the formula

$$f(x, (b, c, d, e)) = c + \frac{d - c}{1 + \exp\{b(\log(x) - \log(e))\}} \quad (1)$$

with 4 parameters b, c, d, e . The parameter e is also denoted ED50 and it is the dose producing a response half-way between the upper limit, d , and lower limit, c . The parameter b denotes the relative slope around e . The interpretation of the parameters is discussed by [Streibig, Rudemo, and Jensen \(1993\)](#). The logistic function is symmetric around e . The three-parameter logistic model with the lower limit equal to 0 has the form:

$$f(x, (b, d, e)) = \frac{d}{1 + \exp\{b(\log(x) - \log(e))\}}$$

and the five-parameter logistic model (Finney 1979) is given by the formula

$$f(x, (b, c, d, e, f)) = c + \frac{d - c}{(1 + \exp\{b(\log(x) - \log(e))\})^f}. \quad (2)$$

Letting the parameter f be equal to 1 in (2) yields the four-parameter logistic model (1).

The four-parameter Gompertz model is given by the formula

$$f(x, (b, c, d, e)) = c + (d - c) \exp\{-\exp\{b(\log(x) - e)\}\}. \quad (3)$$

The parameters c and d are the lower and upper limits, as for four-parameter logistic model, b is the relative slope around e , and the e parameter is the logarithm of the inflection point. The Gompertz model is not symmetric around any point. The three-parameter Gompertz model with the lower limit equal to 0 is then

$$f(x, (b, d, e)) = d \exp\{-\exp\{b(\log(x) - e)\}\}.$$

Besides the Gompertz model and the logistic model, Brain-Cousens' model (Brain and Cousens 1989) is used in situations where *hormesis* is present: For small doses the herbicide has the adverse effect resulting in response values above the level of the control group (Calabrese and Baldwin 2001, 2003). Brain-Cousens model is obtained by modifying the four-parameter logistic model

$$f(x, (b, c, d, e)) = c + \frac{d + fx - c}{1 + \exp\{b(\log(x) - \log(e))\}} \quad (4)$$

adding the linear term fx in the numerator. Again we could also consider the modification where the lower limit is set equal to 0.

Notice that the b and e parameters are different in (1), (2), (3) and (4), respectively. In this paper we assume that the response is a decreasing function of the dose (from maximum response to lower limit), corresponding to positive b , but the models also apply to situation where the response is increasing with the dose.

2.2. Initial parameter values and self starter functions

Contrary to linear regression, estimation of parameters in non-linear regression requires the specification of initial parameter values. The choice of the values may influence on the convergence of the estimation algorithm, in the worst case yielding no convergence and in the best case convergence in few iterations.

For many non-linear functions it may be possible to obtain sensible initial parameter values using the interpretation of the parameters. For the four-parameter logistic function this is done by using the maximum (initial value of d) and minimum (initial value of c) value of the ys and then making the transformation

$$\log\left(\frac{d-y}{y-c}\right) = b(\log(x) - \log(e)). \quad (5)$$

Now linear regression can be used to obtain initial values of b and e . The procedure described can be captured in a function which for a given data set returns starting values. We will refer to such a function as a self starter function. A similar approach can be used for the Gompertz model and Brain-Cousens' model. For another example see section 7.

2.3. Comparison of parameters

Having several fitted dose response curves, it may be of interest to compare parameters across curves, for instance comparing lower limits for different curves.

Typically, in bioassay analysis, the main interest lies in comparing quantities that are functions of the parameters. The common approach has been to re-parameterize and re-fit the model for each function of interest (Schabenberger, Tharp, Kells, and Penner 1999). This approach is relatively computer-intensive, requires some skill in manipulating mathematical expressions and moreover is vulnerable to lack of convergence of the non-linear least squares algorithm, as the re-parameterization could result in strongly correlated parameters depending upon the distribution of the responses between the lower and upper limits. Therefore we implement a single-model approach where a single model is fit and the delta method (van der Vaart 1998, chap. 3), is used to calculate approximate standard errors for functions of the parameters.

The quantities effective dosage (ED) and selectivity index (SI) are commonly used to compare different herbicides. Both ED and SI are functions of the parameters: ED_y is defined as the dose that yields a response which is (100-y)% of the maximal response d (a reduction of y%). For instance, ED_y can be expressed by means of the parameters b and e in the four-parameter logistic model

$$\text{ED}_y = e(y/(100 - y))^{1/b}.$$

The Gompertz model yields a similar formula. For Brain-Cousens' model there are no closed-form solutions, but numerical methods can be applied; we use the bisection method. SI(x,y) is the ratio between ED_x for one curve and ED_y for another curves.

$$\text{SI}(x,y) = \text{ED}_x/\text{ED}_y$$

In section 5 we will illustrate the interpretation of ED and SI based on an example.

3. About the package

The **drc** package consists entirely of interpreted R lines. This paper describes version 0.3-3. The current version as well as future versions can be found at <http://www.bioassay.dk>.

The main function is `multdrc` which carries out the estimation of parameters and returns a model fit, an object of class "drc". The default optimisation method is a variant of the Newton algorithm, but this setting can be changed in `multdrc`. The functions described in subsection 2.1 are built-in in `multdrc` and available by specifying the `fct` argument. Table 1 provides a list of some of the built-in functions available.

Function	Model
<code>g3()</code>	Three-parameter Gompertz
<code>g4()</code>	Four-parameter Gompertz
<code>l3()</code>	Three-parameter logistic
<code>l4()</code>	Four-parameter logistic (default)
<code>l5()</code>	Five-parameter logistic
<code>bc13()</code>	Brain-Cousens three-parameter logistic
<code>bc14()</code>	Brain-Cousens four-parameter logistic

Table 1: Built-in functions in the package **drc**.

The default value is the four-parameter logistic model, that is `l4()` (which need not be specified). It is possible to specify initial parameter values manually (the `startVal` argument), but as it may be difficult to guess at values of the parameters for the less experienced researcher, the built-in functions in **drc** have self starter functions. User-defined functions with a user-defined self starter function can be specified (see section 7).

Once a model fit is obtained using `multdrc`, the following methods are available for objects of class "drc", for extracting information

- `anova`: lack-of-fit test or test for reduction between two models
- `coef`: parameter estimates
- `fitted`: fitted values
- `logLik` log likelihood value
- `plot`: plot of the fitted curves
- `residuals`: raw residuals
- `summary`: summary of the model fit
- `vcov`: estimated variance-covariance matrix

In addition the function `plotdrc` produces a plot of the observations, all curves in a single plot or a plot for each curve. The functions `compParm`, `ED` and `SI` provide comparisons of parameters, ED values and SI values, respectively. The functions `ED` and `SI` work with the built-in nonlinear functions. User-defined functions can also be made to work with `ED` and `SI`, but it requires that they provide additional formulas for the calculation of ED values and their standard errors. The function `diagnostics` supplies brief information about convergence of the estimation algorithm.

4. Fitting a single dose response curve

To get started we need to load the package **drc**. This is done using the `library` function

```
> library(drc)
```

The data sets that we will use in this paper are all part of the package. In this section we use the data set `FA` consisting of a single dose response curve (use `data(FA)` to load the data set). The first 5 lines (rows) of the data set are displayed by typing `FA[1:5,]`

```
> data(FA)
> FA[1:5, ]

      MEANLR MM
1 7.580000  0
2 8.000000  0
3 8.328571  0
4 7.250000  0
5 7.375000  0
```

The variable `MM` is the dose of ferulic acid in mM and `MEANLR` is the root length of perennial ryegrass (Inderjit, Streibig, and Olofsdotter 2002).

As already mentioned the main function in `drc` for fitting dose response curves is `multdrc` which can be used to fit data from one or more dose response curves. By default a four-parameter logistic model is fitted to the data. In order to fit this model to the data set `FA` we write

```
> modelex1 <- multdrc(FA)
```

The argument to the function `multdrc` is a data frame which is a collection of columns of the same length, usually forming a data set or part of a data set. The above R call, `multdrc(FA)`, produces no output. All relevant information of the fit of the model to the data is stored in the object `modelex1`. The package `drc` provides *extractors* for extracting various types of information from `modelex1`. We will now introduce some of these extractors.

The `anova` function can be used to obtain a lack-of-fit test, comparing the four-parameter logistic model to a one-way ANOVA model

```
> anova(modelex1)
```

ANOVA table

	ModelDf	RSS	Df	F value	p value
One-way ANOVA	17	5.1799			
DRC model	20	5.4002	3	0.2411	0.8665

The test is not significant, implying that the four-parameter logistic model provides as good a fit as the one-way ANOVA. A summary of the fit, including the type of model fitted, the parameter estimates and their standard deviations, is obtained using the `summary` function

```
> summary(modelex1)
```

A 'logistic' model was fit.

Parameter estimates:

	Estimate	Std. Error	t-value	p-value
b:(Intercept)	2.98224	0.46507	6.41244	2.961e-06
c:(Intercept)	0.48141	0.21219	2.26875	0.0345
d:(Intercept)	7.79296	0.18857	41.32714	3.822e-21
e:(Intercept)	3.05796	0.18573	16.46441	4.268e-13

Estimate of residual variance: 0.2700107

Box-Cox transformation was not performed.

The t -statistics and corresponding p -values are for testing the null hypotheses that the parameters are equal to 0 (not necessarily relevant hypotheses to consider). The estimate of the common variance parameter σ^2 is 0.27. The last line in the `summary` output tells us that no Box-Cox transformation was performed, that is the default setting. A short version of the output only containing the parameter estimates is obtained typing `coef(model1)` (using the function `coef`).

In cases where the assumption of variance homogeneity is violated, a Box-Cox transform-both-sides approach may help. The optimal Box-Cox transformation is calculated and applied if the argument `boxcox=TRUE` (short: `boxcox=T`) is specified. For the above example the call to `multdrc` would become

```
> model1.boxcox <- multdrc(FA, boxcox = T)
> summary(model1.boxcox)
```

A 'logistic' model was fit.

Parameter estimates:

	Estimate	Std. Error	t-value	p-value
b:(Intercept)	2.499631	0.373305	6.695956	1.621e-06
c:(Intercept)	0.366582	0.094803	3.866787	0.001
d:(Intercept)	7.900259	0.338956	23.307625	6.184e-16
e:(Intercept)	2.989496	0.229546	13.023543	3.159e-11

Estimate of residual variance: 0.07634711

Estimated lambda parameter in Box-Cox transformation: 0.4
P-value for test of null hypothesis that lambda=1: 0.00294

Except for the slope, b:(Intercept), none of the parameters changed dramatically, and the test for lack of fit is still not significant as seen from the output from `anova` below.

```
> anova(model1.boxcox)
```

ANOVA table

	ModelDf	RSS	Df	F value	p value
One-way ANOVA	17	1.1778			
DRC model	20	1.5269	3	1.6801	0.2089

Apparently, the optimal parameter λ for the Box-Cox transformation, which is 0.4, is significantly different from 1.00, that is no transformation. This means that the variance of the responses is not homogeneous and that the Box-Cox transformation is required. A residual plot (not shown) using the call `plot(fitted(modelex1.boxcox), residuals(modelex1.boxcox))` showed that the variance inhomogeneity had been removed by the transformation. The fit of the model to the data can be seen in Figure 1.

```
> plot(modelex1.boxcox, xlab = "Dose (mM)", ylab = "Root length(mm)")
```

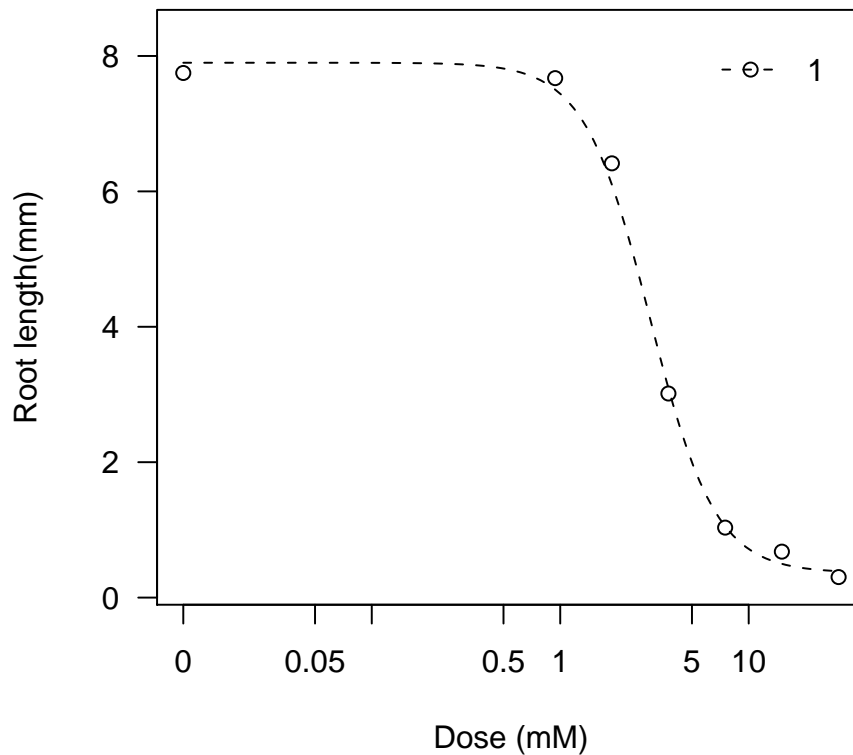


Figure 1: Four-parameter logistic fit using Box-Cox transformation.

5. Simultaneous fitting: Summarising the results

The data set `PestSci` consists of 5 curves each with 7 doses and in 3 replications. The variables are `CURVE`, `DOSE` and `SLOPE`, containing the curve number, the dose values and the response values, respectively. The response is rate of change of Oxygen consumption of chloroplast membranes versus the dose of a herbicide. Relevant references are [Streibig, Dayan, Rimando, and Duke \(1999\)](#); [Nielsen, Ritz, and Streibig](#). To load the data set `PestSci` type `data(PestSci)`.

To get an overview an initial plot of the data set is useful. The function `plotdrc` plots the dose values against the response values with different plot symbols for the different dose response curves (Figure 2). The plot may help getting a first impression of the data, in particular with respect to the ranges of dose values and the range of response values.

```
> plotdrc(PestSci, ylab = "Rate of Oxygen Evolutions", colour = TRUE)
```

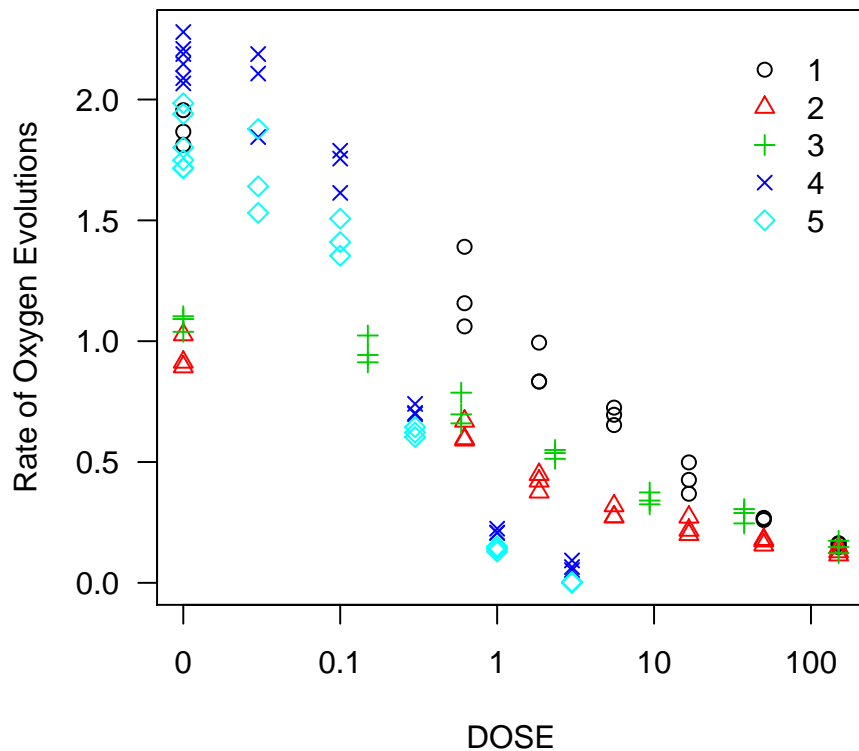


Figure 2: Plot of the data in the data set `PestSci`.

We consider simultaneous fitting of the 5 dose response curves in `PestSci`, assuming each curve follows a four-parameter logistic curve. Furthermore, we assume the parameters differ among curves (in total $4 \cdot 5 = 20$ parameters). This model is fitted using the `multdrc` function

```
> modelx2 <- multdrc(SLOPE ~ DOSE, CURVE, data = PestSci)
```

The formula `SLOPE~DOSE` in the first position relates response to dose. The variable `CURVE` in the second position is the grouping variable, uniquely assigning each observation to a curve. The argument `data=PestSci` specifies the data frame where the variables `CURVE`, `DOSE` and `SLOPE` are found.

The lack-of-fit test comparing the simultaneous four-parameter logistic model to the alternative two-way ANOVA model is obtained applying the function `anova` to the object `modelx2`

```
> anova(modelx2)
```

ANOVA table

	ModelDf	RSS	Df	F value	p value
Two-way ANOVA	70	0.38635			
DRC model	85	0.45955	15	0.8841	0.5842

The test is not significant, meaning that the non-linear regression model provides an acceptable description of the data. Subsequently, we proceed to look at the parameter estimates. The estimates and their standard errors are obtained using the function `summary`

```
> summary(modelx2)
```

A 'logistic' model was fit.

Parameter estimates:

	Estimate	Std. Error	t-value	p-value
b:1	0.5195260	0.0763599	6.8036530	1.346e-09
b:2	0.8009531	0.2257107	3.5485829	0.0006
b:3	0.6818983	0.1285630	5.3040019	8.851e-07
b:4	1.8447607	0.1663488	11.0897146	1.698e-18
b:5	1.6507990	0.1758393	9.3881105	8.863e-15
c:1	-0.0165855	0.1078225	-0.1538223	0.8781
c:2	0.1326173	0.0471817	2.8107768	0.0061
c:3	0.1463932	0.0604363	2.4222722	0.0175
c:4	0.0795360	0.0394604	2.0155892	0.0470
c:5	-0.0090699	0.0443533	-0.2044913	0.8385
d:1	1.8795509	0.0423711	44.3592666	7.412e-61
d:2	0.9459940	0.0422671	22.3813367	1.084e-37
d:3	1.0903222	0.0405608	26.8811926	1.268e-43
d:4	2.1535763	0.0281853	76.4078098	2.017e-80
d:5	1.8062724	0.0292458	61.7617568	1.052e-72
e:1	1.7949368	0.4782218	3.7533561	0.0003
e:2	0.9455148	0.2494531	3.7903514	0.0003
e:3	1.3731181	0.4527616	3.0327617	0.0032

```
e:4 0.1973281 0.0101893 19.3661821 3.231e-33
e:5 0.2107954 0.0138242 15.2482485 2.379e-26
```

Estimate of residual variance: 0.005406425

Box-Cox transformation was not performed.

From the summary it is seen that the lower limit of dose response curve no. 1 and no. 5 are negative which, strictly speaking, is not meaningful, but these two lower limits are not significantly different from zero (right-most column). Comparison of parameters between curves can be accomplished using the `compParm` function.

A plot of the original observations and the fitted dose response curves is obtained using the `plot` function. The resulting plot is displayed in Figure 3, showing reasonable agreement between observations and fitted curves.

```
> plot(modelex2, xlab = "Dose", ylab = "Rate of Oxygen Evolution",
+      conName = "Control", colour = TRUE)
```

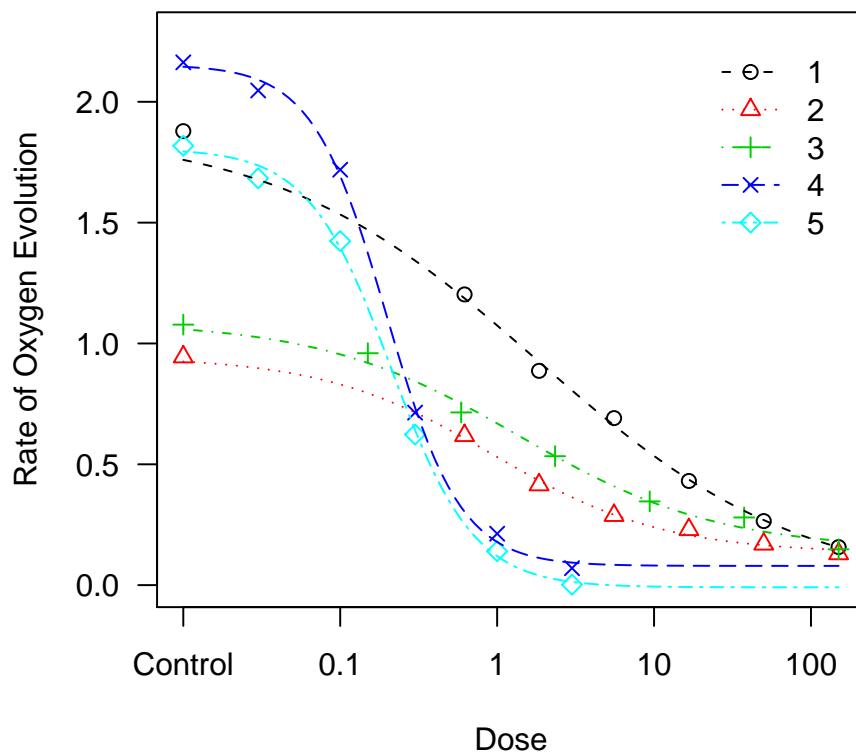


Figure 3: Observed data and fitted dose response curves for the data set `PestSci`.

The call `plot(modelex2, obs="none")` produces a similar plot, but without the observations. Also the argument `conName` specifies the name of the tick mark for the dose level that can be considered as dose 0. The usual graphical parameters in R can also be supplied in the function call; for instance `xlab` and `ylab` to change the default labels of the axes (which are the names of the variables in the data frame). For more details on the `plot` function see `?plot.drc`.

As already mentioned in Section 2, the package **drc** provides functions `ED` and `SI` to compute ED_x values and selectivity indices `SI(x,y)`, respectively, for the built-in models like the four-parameter logistic model. As an example we calculate estimates of the parameters ED₁₀, ED₅₀ and ED₉₀ and their standard errors for all curves in `PestSci`. The call to `ED` is

```
> ED(modelex2, c(10, 50, 90))
```

	Estimate	Std. Error
1:10	0.02613925	1.266207e-02
1:50	1.79493682	4.782218e-01
1:90	123.25517064	1.016492e+02
2:10	0.06085337	4.795887e-02
2:50	0.94551481	2.494531e-01
2:90	14.69102314	1.240795e+01
3:10	0.05474148	2.946568e-02
3:50	1.37311812	4.527616e-01
3:90	34.44286207	2.810508e+01
4:10	0.05996769	6.824301e-03
4:50	0.19732814	1.018932e-02
4:90	0.64932290	8.071697e-02
5:10	0.05569450	8.054704e-03
5:50	0.21079544	1.382424e-02
5:90	0.79782946	1.332322e-01

Notice that the estimates for the parameter ED₅₀ already were displayed above in the `summary` output as the standard parameterization of the four-parameter logistic model involves ED₅₀. The selectivity indices between two curves for ED₁₀ and ED₉₀, `SI(90,10)`, are obtained using the call

```
> SI(modelex2, c(90, 10))
```

	Estimate	Std. Error	t-value	p-value
1/2:90/10	2025.44532	2310.472561	0.8762040	3.829193e-01
1/3:90/10	2251.58622	2217.410800	1.0149613	3.124573e-01
1/4:90/10	2055.35957	1711.128319	1.2005877	2.326127e-01
1/5:90/10	2213.05804	1852.972354	1.1937890	2.352507e-01
2/3:90/10	268.37093	268.783022	0.9947464	3.221465e-01
2/4:90/10	244.98230	208.780349	1.1686076	2.452081e-01
2/5:90/10	263.77869	226.028483	1.1625910	2.476310e-01
3/4:90/10	574.35697	473.206155	1.2116431	2.283687e-01
3/5:90/10	618.42479	512.493908	1.2047456	2.310099e-01
4/5:90/10	11.65865	2.223371	4.7939161	5.413344e-06

The output provides standard errors of the estimates and p -values for testing the null hypothesis that the indices are equal to 1.

The use of SI is best illustrated by using the methods chemical companies use to quantify selectivity in their research and development of new herbicides. Since herbicides have an effect on any plant, be it crops or weeds, we sometimes have to accept a small decrease in the crop yield, for example a 10% decrease might be tolerated (ED10), while a 90% control of the weed (ED90) is considered a reasonable control level. The larger the SI the more selective is one herbicide as compared to another herbicide. Obviously, the above SI(90,10) values are all much larger than 1.00, but only for curve no 4 and 5 the value is significantly larger than 1.00.

The function `SI` can also be used to compare potencies among absolute response levels. The ED x for a dose response curve is a relative response level depending upon the upper and lower limits of the curve. For example in `PestSci` the response for the dose ED50 for the 2nd curve is 0.54 $((0.94+0.13)/2)$ whilst for the 4th curve the response level is 1.38 $((2.15+0.08)/2)$. In fact 1.38 exceeds the maximum response level of the 2nd curve. Consequently, we sometimes want to compare two curves at a certain absolute response level and it can easily be done with `SI`. Let us assume that for various reasons we want to compare curves 2 and 4 at an absolute response level of 0.5. For the 2nd curve it means at the dose ED46 $(0.46 = (0.5-0.13)/(0.94-0.13))$ and for 4th curve at the dose ED21. To obtain only the comparison between the 2nd curve and the 4th curve we use `SI` with an extra vector argument giving the curves to be compared

```
> SI(modelx2, c(46, 21), c(2, 4))

      Estimate Std. Error  t-value    p-value
2/4:46/21 8.043639    2.21059 3.186316 0.001898055
```

6. Simultaneous fitting: Model reduction

In this section we illustrate how to reduce a model using significance tests.

The data set `TM` comprises of 7 response curves, measured at a number of positive dose values, and an additional group of measurements at dose zero. A separate control group frequently occurring case in bioassay analysis. There are three variables: `dose` is the dose, `pct` the curve number and `rgr` the response. The responses are growth rates of duckweed and the treatments are mixtures of two herbicides with different modes of action (Cedergreen 2004). There are 180 observations.

We define a simultaneous model, assuming that each dose response curve follows a four-parameter Gompertz model (3) and with different parameters for different assays. This model is specified as follows

```
> modelx3.1 <- multdrc(rgr ~ dose, pct, data = TM, fct = g4())
```

As `pct` take 8 different values (7 curves + control group), we are specifying a model with 8 curves, where the curve corresponding to the control group only is defined as dose equal to

0 and thus not constitutes an entire dose response curve. The asymmetric four-parameter Gompertz model is specified with the argument `fct=g4()`. The lack-of-fit test against the two-way ANOVA using `anova` (not displayed) confirms that the Gompertz model fits the data.

A summary of the fit `modelex3.1` is

```
> summary(modelex3.1)
```

A 'gompertz' model was fit.

Parameter estimates:

	Estimate	Std. Error	t-value	p-value
b:100	1.0655191	0.2170291	4.9095672	2.344e-06
b:83	0.9288037	0.2498459	3.7175067	0.0003
b:67	0.7840124	0.1500974	5.2233574	5.737e-07
b:50	1.7389206	0.2908936	5.9778581	1.566e-08
b:33	0.8050234	0.1412846	5.6978859	6.163e-08
b:17	0.9136431	0.1806474	5.0576035	1.215e-06
b:0	1.1056319	0.2084593	5.3038276	3.963e-07
c:100	0.0109434	0.0099491	1.0999415	0.2731
c:83	0.0179145	0.0167275	1.0709619	0.2859
c:67	-0.0089369	0.0141037	-0.6336573	0.5273
c:50	-0.0040568	0.0094046	-0.4313600	0.6668
c:33	-0.0189114	0.0122196	-1.5476233	0.1238
c:17	-0.0039410	0.0118115	-0.3336561	0.7391
c:0	-0.0099106	0.0095277	-1.0401952	0.2999
d:999	0.2814899	0.0068124	41.3199574	1.670e-84
d:100	0.3497876	0.0245492	14.2484165	4.906e-30
d:83	0.3166707	0.0237328	13.3431502	1.265e-27
d:67	0.3668345	0.0310219	11.8250075	1.495e-23
d:50	0.2798449	0.0103144	27.1314743	3.096e-60
d:33	0.3749350	0.0316003	11.8649171	1.168e-23
d:17	0.3301507	0.0254194	12.9881540	1.129e-26
d:0	0.3059784	0.0211905	14.4394434	1.530e-30
e:100	4.8721705	0.1126764	43.2403908	2.980e-87
e:83	7.8423011	0.1611212	48.6733107	1.697e-94
e:67	8.2910662	0.1547290	53.5844382	1.858e-100
e:50	8.9175008	0.0811260	109.9216722	2.602e-146
e:33	8.8503914	0.1470690	60.1785064	9.934e-108
e:17	9.0866222	0.1369517	66.3491119	6.726e-114
e:0	9.2302599	0.1146251	80.5256594	3.003e-126

Estimate of residual variance: 0.0005569125

Box-Cox transformation was not performed.

By default only the d parameter (the upper limit at dose zero) is estimated for the control group, represented in the summary by the term `d:999`.

Figure 4 clearly shows that allowing individual upper limits for the individual curves does not produce a fit that agrees with the control group.

```
> plot(model3.1, ylim = c(-0.05, 0.4), conLevel = 1)
```

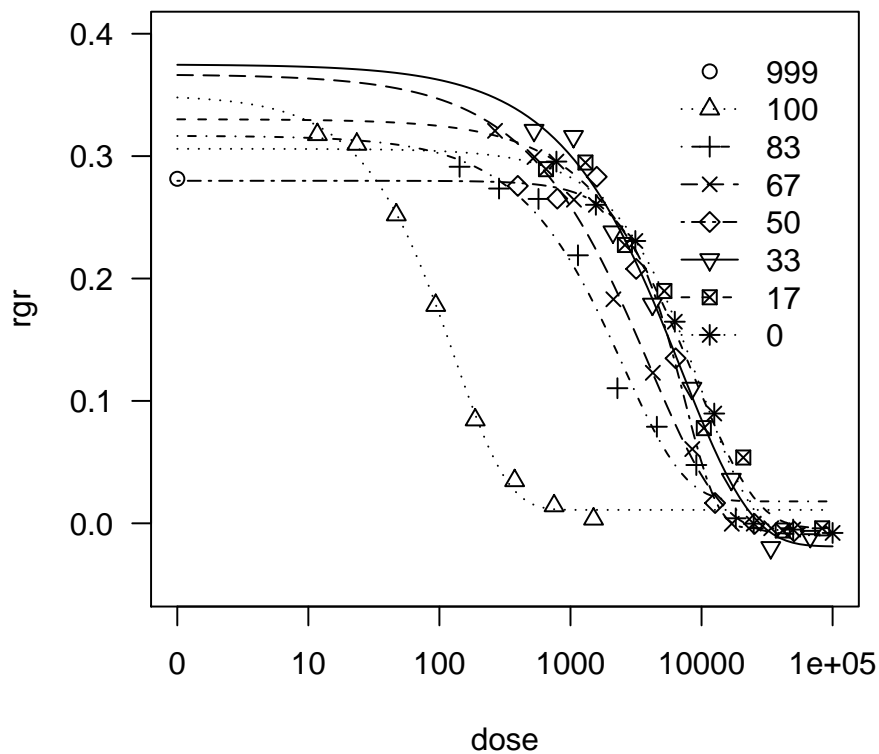


Figure 4: Observed data and fitted dose response curves for the data set TM.

A more reasonable initial model would only have a single d parameter common to all curves, as the control group determines the common upper limit for all 7 curves. This can be specified in `multdrc` using the `collapse` argument specifying which parameters should be collapsed across assays. This argument may be specified using a data frame or a list as argument. If no `collapse` argument is given there are different parameters for different curves (used in sections 4 and 5). The two types of argument are overlapping in their functionality, but not entirely identical, both have strong points. The data frame specification is better for collapsing parameters for arbitrary curves, without requiring the corresponding grouping variable to be defined. The list specification allows more general structures involving more than one variable per parameter (see section 8). Specification of the `collapse` argument by means of a list of


```

modelex3.2 <- multdrc(rgr~dose, pct,
collapse=list(~factor(pct), ~factor(pct), ~1, ~factor(pct))
data=TM, fct=g4())

```

In order to compare `modelex3.2` to `modelex3.1` the `anova` function can be used. The `anova` function with the two model objects as arguments produces an approximate F-test (or a likelihood ratio test) for test of reduction of the larger model to the smaller model. For the models `modelex3.1` and `modelex3.2` `anova` gives an F-test for reduction from the model with different upper limits to the model with a common upper limit.

```
> anova(modelex3.1, modelex3.2)
```

```

Model 1
  fct:      g4()
collapse: pct (for all parameters)
Model 2
  fct:      g4()
collapse: TM[, 2], TM[, 2], colFct(TM[, 2], 1:8), TM[, 2]

```

ANOVA table

	Df	Sum Sq	Mean Sq	F value	p value
Model 2	158	0.107814	0.003389		
Model 1	151	0.084094	0.000557	6.0847	2.869e-06

The test is highly significant and the model with common upper limit is rejected. Notice that the order of the two arguments in `anova` does not matter. The main problem with the data set `TM` is the presence of hormesis (described in section 5), thus a more appropriate approach would be to use Brain-Cousens' model (`fct=bc14`).

Another aspect of `modelex3.1` is that it seems that all lower limits are equal. The model with a common lower limit is

```

> modelex3.3 <- multdrc(rgr ~ dose, pct, data.frame(pct, 1,
+          pct, pct), data = TM, fct = g4())

```

The test for model reduction is not significant

```
> anova(modelex3.3, modelex3.1)
```

```

Model 1
  fct:      g4()
collapse: pct, 1, pct, pct
Model 2
  fct:      g4()
collapse: pct (for all parameters)

```

ANOVA table

	Df	Sum Sq	Mean Sq	F value	p value
Model 1	157	0.087083	0.000498		
Model 2	151	0.084094	0.000557	0.8945	0.5006

Thus `modelelex3.3` provides as good a fit to the data as does `modelelex3.1`. Next we fit a model with the common lower limit equal to 0. This can be done using the built-in function `g3()` which is a three-parameter Gompertz model (lower limit set to 0).

```
> modelelex3.4 <- multdrc(rgr ~ dose, pct, data.frame(pct, pct,
+ pct), data = TM, fct = g3())
```

The F-test for reduction from `modelelex3.3` to `modelelex3.4` is

```
> anova(modelelex3.4, modelelex3.3)
```

```
Model 1
fct:      g3()
collapse: pct, pct, pct
Model 2
fct:      g4()
collapse: pct, 1, pct, pct
```

ANOVA table

	Df	Sum Sq	Mean Sq	F value	p value
Model 1	158	0.087804	0.000721		
Model 2	157	0.087083	0.000555	1.2998	0.2560

Thus we can reduce the initial model to a model with common lower limit equal to 0.

7. A user-defined function

Consider the data set `Puromycin` available in R issuing the command

```
> data(Puromycin)
```

In the help page `?Puromycin`, a Michaelis-Menten model is suggested. This model has the form

$$y = f(x, (Vm, K)) = \frac{Vm x}{K + x} \quad (6)$$

with two parameters, Vm and K . Define this non-linear function in R as follows

```
> MMfct <- function(x, parm) {
+   parm[, 1] * x/(parm[, 2] + x)
+ }
```

where `parm` is matrix consisting of two columns: `parm[,1]` and `parm[,2]` corresponding to V_m and K , respectively. A self starter function could be defined taking the maximum of the response values as initial estimate of V_m and, using this initial estimate of V_m , the equation (6) is then solved in K for an arbitrary pair (x, y) (we take the first pair in the data set). Thus the self starter function looks like

```
> MMssfct <- function(data) {
+   Vm <- max(data[, 2])
+   K <- Vm * data[1, 1]/(data[1, 2] - data[1, 1])
+   return(c(Vm, K))
+ }
```

where `data` is a data frame containing x values in the first column and y values in the second column. Finally we set the names of the parameters (V_m and K)

```
> MMnames <- c("Vm", "K")
```

Now we are ready to use the function `multdrc`. Fitting a simultaneous Michaelis-Menten model to the two states is accomplished using the call

```
> modelex4 <- multdrc(rate ~ conc, state, data = Puromycin,
+   fct = list(MMfct, MMssfct, MMnames))
```

A summary of the model fit containing parameter estimates and estimated standard errors and a comparison of the two V_m parameters is given below.

```
> summary(modelex4)
```

A 'drc' model was fit.

Parameter estimates:

	Estimate	Std. Error	t-value	p-value
Vm:treated	2.1270e+02	6.8097e+00	3.1234e+01	4.268e-18
Vm:untreated	1.6030e+02	7.2397e+00	2.2141e+01	4.930e-15
K:treated	6.4139e-02	8.2831e-03	7.7433e+00	2.711e-07
K:untreated	4.7733e-02	8.9549e-03	5.3303e+00	3.822e-05

Estimate of residual variance: 108.1608

Box-Cox transformation was not performed.

```
> compParm(modelex4, "Vm", "--")
```

	Estimate	Std. Error	t-value	p-value
Vm:treated-Vm:untreated	52.39933	9.939061	5.27206	2.38141e-05

8. An additive model

The data set `PestSci2` consists of 4 dose response curves. Dose values and response values are denoted `DOSE` and `SLOPE`, respectively. The grouping variable is `CURVE` and in addition there are two two-level factors A (levels: a, b) and B (levels: 1, 2), measured for each observation. We assume a four-parameter logistic model. A model with an additive model of the form `A+B` for the parameter d and interaction structure `(A:B)` for the remaining parameters can be specified using a list of formulas

```
modelex5.1 <- multdrc(SLOPE~DOSE, CURVE,
collapse=list(~A:B, ~A:B, ~A+B, ~A:B),
data=PestSci2)
```

A summary of the fit is

```
> summary(modelex5.1)
```

A 'logistic' model was fit.

Parameter estimates:

	Estimate	Std. Error	t-value	p-value
b:(Intercept)	0.4046567	0.4155826	0.9737094	0.3336
b:Ab	1.4466370	0.8650661	1.6722850	0.0990
b:B2	0.4110364	0.7057720	0.5823926	0.5622
b:Ab:B2	-0.1367747	1.4389090	-0.0950544	0.9245
c:(Intercept)	-0.0784325	0.7074487	-0.1108668	0.9120
c:Ab	0.3487946	0.7179824	0.4857983	0.6286
c:B2	0.2288862	0.7385790	0.3099008	0.7576
c:Ab:B2	-0.1969272	0.7566467	-0.2602631	0.7954
d:(Intercept)	1.4193874	0.1422696	9.9767418	5.121e-15
d:Ab	0.4623971	0.1507881	3.0665362	0.0031
d:B2	-0.0066858	0.1432294	-0.0466790	0.9629
e:(Intercept)	1.7574201	4.4084108	0.3986516	0.6914
e:Ab	-1.5558570	4.4080232	-0.3529603	0.7252
e:B2	-0.4686343	4.4797946	-0.1046107	0.9170
e:Ab:B2	0.4350737	4.4781948	0.0971538	0.9229

Estimate of residual variance: 0.09081889

Box-Cox transformation was not performed.

9. Conclusions

We have described the package **drc** for fitting multiple non-linear regression models, with special focus on applications in bioassay analysis. Thus the package **drc** is an attempt to develop a collection of functions specifically designed for bioassay analysis. Relevant interpretations of a model fit is easy using the built-in models.

At the same time we would like to emphasize that **drc** provides functionality for simultaneous fitting of arbitrary non-linear regression models under the assumption of independent and normally distributed measurement errors, such functionality was not previously available in R.

References

- Brain P, Cousens R (1989). “An Equation to Describe Dose Responses where there is Stimulation of Growth at Low Dose.” *Weed Research*, **29**, 93–96.
- Calabrese EJ, Baldwin LA (2001). “Hormesis: U-shaped Dose Responses and their Centrality in Toxicology.” *TRENDS in Pharmacological Sciences*, **22**, 285–291.
- Calabrese EJ, Baldwin LA (2003). “Toxicology Rethinks its Central Belief.” *Nature*, **421**, 691–692.
- Cedergreen N (2004). “Bioassay Data.” Unpublished, Royal Veterinary and Agricultural University, Copenhagen.
- Finney DJ (1979). “Bioassay and the Practice of Statistical Inference.” *International Statistical Review*, **47**, 1–12.
- Inderjit, Streibig JC, Olofsdotter M (2002). “Joint Action of Phenolic Acid Mixtures and its Significance in Allelopathy Research.” *Physiologia Plantarum*, **114**, 422–428.
- Nielsen OK, Ritz C, Streibig JC (????). “Nonlinear Mixed-Model Regression to Analyze Herbicide Dose-Response Relationships.”
- R Development Core Team (2004). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL <http://www.R-project.org>.
- Schabenberger O, Tharp BE, Kells JJ, Penner D (1999). “Statistical Tests for Hormesis and Effective Dosages in Herbicide Dose Response.” *Agronomy Journal*, **91**, 713–721.
- Streibig JC, Dayan FE, Rimando AM, Duke SO (1999). “Joint Action of Natural and Synthetic Photosystem II Inhibitors.” *Pesticide Science*, **55**, 137–146.
- Streibig JC, Rudemo M, Jensen JE (1993). “Dose-Response Curves and Statistical Models.” In P Kudsk, JC Streibig (eds.), “Herbicide Bioassays,” pp. 29–55. CRC Press, Boca Raton.
- van der Vaart AW (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Affiliation:

Christian Ritz
Department of Natural Sciences
Royal Veterinary and Agricultural University
Thorvaldsensvej 40
DK-1871 Frederiksberg C, Denmark
E-mail: ritz@bioassay.dk
URL: <http://www.bioassay.dk>

Jens C. Streibig
Department of Agricultural Sciences (Crop Science)
Royal Veterinary and Agricultural University
Thorvaldsensvej 40
DK-1871 Frederiksberg C, Denmark
E-mail: streibig@bioassay.dk
URL: <http://www.bioassay.dk>