An introduction to the medrc package

Daniel Gerhard

November 14, 2013

Contents

1 Motivation

The analysis of dose-response experiments using nonlinear models allows inference about the model parameters and the prediction of new response values. A common way to summarize the effect of an increasing dose level on the response is deriving parameters from the dose-response curve, like the effective dose $ED(p)$. Also the comparison of several curves by selectivity indices, like the relative potency, is available. For risk estimation in toxicology the estimation of benchmark dose (BMD) levels is a further important topic that is based on the nonlinear modeling of dose-response curves.

The drc package [Ritz and Streibig, 2005] allows the simultaneous fitting of several non-linear regression models, providing a common parameterization for several models and searching automatically for starting values. Several functions are available for model-averaging and inference of derived parameters.

When analyzing dose-response curves, the observed data is often obtained from experiments with hierarchical designs, where the responses can be assigned to several known clusters. Instead of just assuming a single residual error in the dose-response model, the variability within and between clusters can be modeled. Either additional correlation parameters can be introduced to structure the residual error by generalized nonlinear least squares estimation or a distribution function for the cluster effects can be assumed separately to the distribution of the residuals in a mixed model framework.

The nlme package [Pinheiro and Bates, 2000] provides functions to estimate parameters and variance components in nonlinear mixed models by (restricted) maximum likelihood. Some functions to parameterize the non-linear curve are already available, but it is also possible to provide your own modeling function.

The medrc combines the automated dose-response modeling framework of the package drc with the nonlinear mixed estimation framework of the package nlme. Thereby, additional random effects can be introduced to the dose-response models with the unified parameterization of drc, with the availability of inference for derived parameters also for hierarchical models.

2 Broccoli example data

The effect of drought stress on Brassica oleracea is investigated, selecting drought stress resistant varieties out of a population of different DH genotypes. The screening study was carried out on 48 DH lines developed from F1 plants of a cross between the rapid cycling chinese kale (*Brassica oleracea* var. *alboglabra* (L.H. Bailey) Musil) and broccoli (*Brassica oleracea* var. *italica* Plenck). Two stress treatments (not watered and a watered control) are randomly assigned to 4 plants per genotype (2 per treatment) resulting in 192 plants in total. For the genotypes 5, 17, 31, 48, additional 12 plants (6 per treatment) are included into the completely randomized design, which results in a total of 240 plants. For each plant the length of the youngest leaf at the beginning of the experiment is measured daily for a period of 16 days. For the additional 12 plants of the 4 genotypes the leaf water potential was measured as a secondary endpoint (omitted here); due to these destructive measurements some dropouts occur.

```
data(broccoli)
str(broccoli)

## 'data.frame': 3689 obs. of  5 variables:
##  $ LeafLength: num  1.4 1.2 2.5 1.8 1.7 1.4 5.1 5.5 3.1 1.8 ...
##  $ ID        : Factor w/ 240 levels "1","2","3","4",..: 38 62 35 91 76 108 149 237 223 239 ...
##  $ Stress    : Factor w/ 2 levels "control","drought": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Genotype  : Factor w/ 48 levels "1","2","3","4",..: 17 17 17 17 17 17 31 31 31 31 ...
##  $ Day       : int  1 1 1 1 1 1 1 1 1 1 ...

ggplot(broccoli, aes(x=Day, y=LeafLength, group=ID, colour=Stress)) +
  geom_line() +
  facet_wrap(~ Genotype, ncol=8)
```
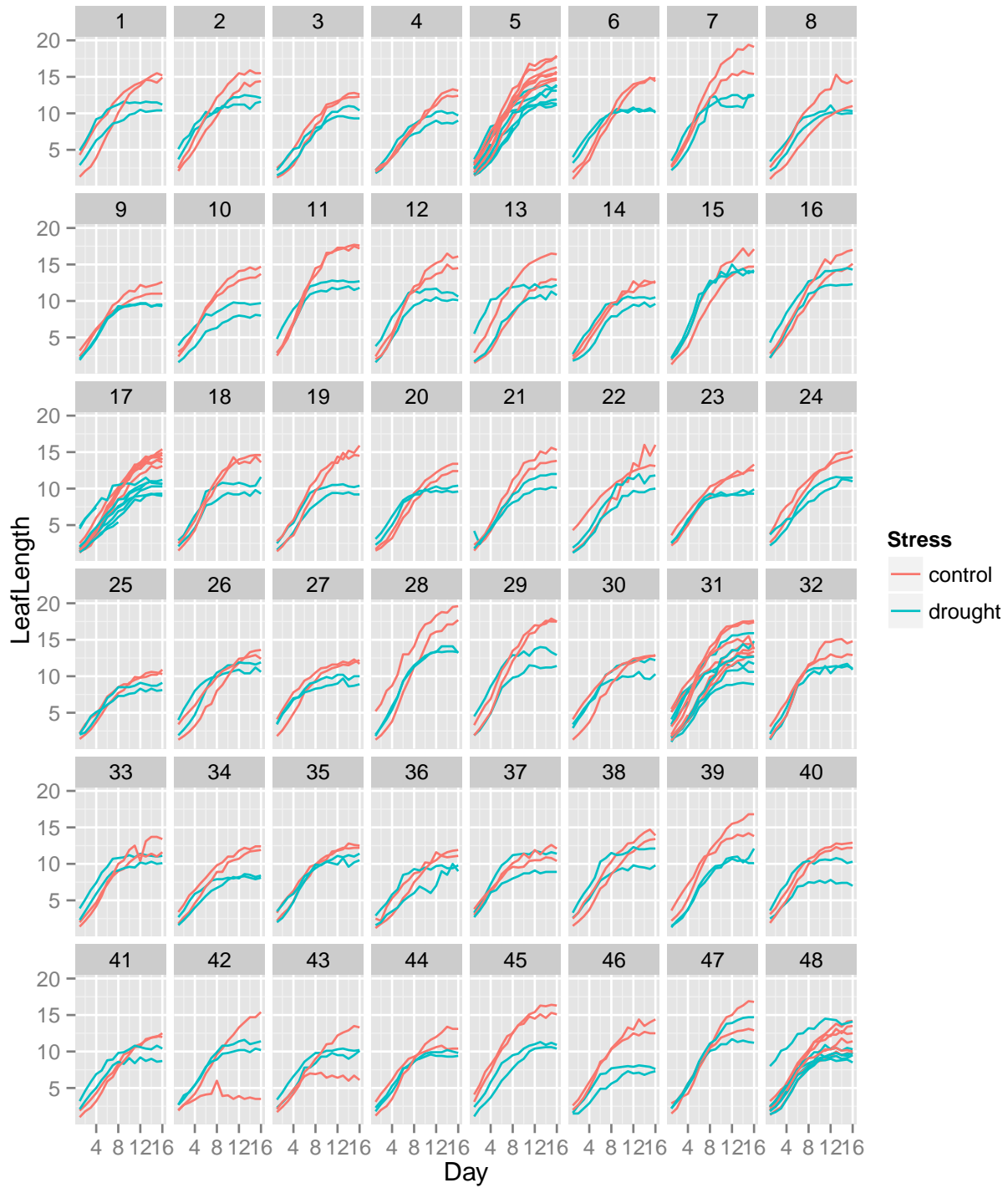
There are two individuals for genotypes 42 and 43 with growth curves, which are not representative for the population of interest, as their leaf growth has stopped in favor of early flowering. These curves are removed from the dataset.

```
bro <- droplevels(subset(broccoli, ID != "110" & ID != "125"))
```

3   Nonlinear mixed effects models

Following the notation of Davidian and Giltinan [1995], a nonlinear model with a single hierarchical level can be defined in two stages, parameterizing the model for treating the variability within the ith subject, and combining the subject specific models in stage 2.

For $i = 1, \ldots, m$ individuals, following model can be assumed

$$y_{ij} \sim f(x_{ij}, \boldsymbol{\beta}_i) + \epsilon_{ij}$$

where y_{ij} are the response vectors of length $j = 1, \ldots, n_i$ with the corresponding vectors of dose levels x_{ij} for each individual. The dose-response curve is described by the functions $f(x_{ij}, \boldsymbol{\beta}_i)$ with an individual-specific $(p \times 1)$ vector of parameters $\boldsymbol{\beta}_i$. The residual vectors $\epsilon_{ij} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{\Lambda}_i)$ are assumed to be normal distributed with a correlation structure defined by the elements of $\boldsymbol{\Lambda}_i$.

In the second stage the inter-individual effects are described by modeling the $\boldsymbol{\beta}_i$. The individual effects are separated into a fixed and random component

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{B}_i \boldsymbol{b}_i$$

where $\boldsymbol{\beta}$ is a stacked vector of fixed effects parameters with one set of model parameters for each curve defined in the design matrix $\boldsymbol{A}_i$. The between individual effects and further between cluster effects can be defined by a random effects vector $\boldsymbol{b}_i$ and the corresponding design matrix $\boldsymbol{B}_i$.

The random effects can be assumed to follow a normal distribution

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{G})$$

where $\boldsymbol{G}$ is the between individual variance-covariance matrix.

3.1 The medrm function

To fit a hierarchical dose-response model with medrc, the medrm function can be used with following main arguments:

form A formula with the name of the response on the left and the name of the dose variable on the right hand side

curveid A formula with the name of a factor, which divides the dataset into several clusters. For each cluster a different fixed effect curve is assumed. On the left hand side, the names of parameters can be given, separated by a + symbol, which are assumed to be different across the curves. Hence, only a subset of parameters may describe the difference between fixed effect curves. If these parameter names are omitted, or curveid is set to NULL, only a single curve is fitted.

data A data.frame object with the dose-response data

fct A model function of package drc.

random The definition of random effects, similar to the random argument in the function nlme. This can be, for example, a formula with parameter names on the left hand side, like in the curveid argument, and 1|name of clusters defining factor on the right hand side. The 1 can also be substituted by an additional covariate. For a more flexible definition, a list of different pdClasses objects of nlme can be given.

start Starting values for the nlme function. If NULL, the initial values for the fixed effects are found automatically, using the self start functionality of package drc.

3.2 Dose-response curve parameterization

In drc a number of different dose-response models are available. Each model can be simplified, by fixing a parameter to a specific values instead of estimating it from the data.

The parameters are defined in a unified way with

b steepness of the curve

c lower asymptote

d upper asymptote

e location of the inflection point

f asymmetry parameter

5-parameter logistic L.5()

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

5-parameter log-logistic LL.5()

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

4 parameter Weibull W1.4() and W2.4()

$$f(x) = c + (d - c)\exp(-\exp(b(\log(x) - \log(e))))$$

or

$$f(x) = c + (d - c)(1 - \exp(-\exp(b(\log(x) - \log(e)))))$$

4 parameter log-Normal LN.4()

$$f(x) = c + (d - c)(\Phi(b(\log(x) - \log(e))))$$

gompertz()

$$f(x) = c + (d - c)(\exp(-\exp(b(x - e))))$$

3 parameter Michaelis-Menten MM.3()

$$f(x, (c, d, e)) = c + \frac{d - c}{1 + (e/x)}$$

Further functions

More functions are available, like fractional polynomial-logistic models fplogistic(p1, p2), or the Brain-Cousens hormesis model braincousens().

3.3 Broccoli data: 5 parameter logistic model

Let us assume that the growth curves for each individual follow a 4 parameter log-logistic model with different parameters for each stress treatment except the lower asymptote. The model function can be conveniently defined as a predefined function of package drc. The parameterization of different fixed effect effect vectors per treatment level can be defined by a curveid argument, with the name of a fixed effect factor in the dataset on the right hand side of a formula, and the parameters to change for different curves on the left hand side. The definition of random effects follow in a similar way. The random genotype effects are parameterized by a variance component for the upper asymptote, and the location of the inflection point. To model the dependency of observations measured at the same plant over time, a second set of random effects is assumed within the genotype effects to model the variance between individuals within genotypes.

For a simplification we ignore the real aim of the study to detect genotype specific stress effects and just model an additive stress effect for each genotype. The same lower asymptote (parameter c) is assumed for both stress treatment levels, as at the beginning of the experiment, all plants are watered at the same level.

```
m5pl <- medrm(LeafLength ~ Day, data=bro,
              fct=L.5(),
              curveid=b + d + e + f ~ Stress,
              random=d + e ~ 1|Genotype/ID)

print(m5pl)

##
## A 'drc' model.
##
## Call:
## medrm(form = LeafLength ~ Day, curveid = b + d + e + f ~ Stress,    data = bro, fct = L.5(), random
##
## Coefficients:
## b.control  b.drought  c.control  c.drought  d.control  d.drought
##    -0.322     -0.491      0.473      0.473     14.462     10.707
## e.control  e.drought  f.control  f.drought
##     2.624      4.364      2.101      0.881
```

With the summary method, the summary output of the nlme component can be accessed. There are also some methods, like VarCorr, to access the specific nlme slots. For any special information about the nlme estimation, the nlme object is stored in a list slot named fit.

```
VarCorr(m5pl)

##                Variance                      StdDev Corr
## Genotype =     pdLogChol(list(d ~ 1,e ~ 1))
## d.(Intercept) 1.86689                        1.3663 d.(In)
## e.(Intercept) 0.06088                        0.2467 0.726
## ID =           pdLogChol(list(d ~ 1,e ~ 1))
## d.(Intercept) 1.82951                        1.3526 d.(In)
## e.(Intercept) 1.28191                        1.1322 0.131
## Residual      0.10657                        0.3265

# same as
VarCorr(m5pl$fit)

##                Variance                      StdDev Corr
## Genotype =     pdLogChol(list(d ~ 1,e ~ 1))
## d.(Intercept) 1.86689                        1.3663 d.(In)
## e.(Intercept) 0.06088                        0.2467 0.726
## ID =           pdLogChol(list(d ~ 1,e ~ 1))
## d.(Intercept) 1.82951                        1.3526 d.(In)
## e.(Intercept) 1.28191                        1.1322 0.131
## Residual      0.10657                        0.3265
```

The predicted random effects can be directly accessed by the function ranef().
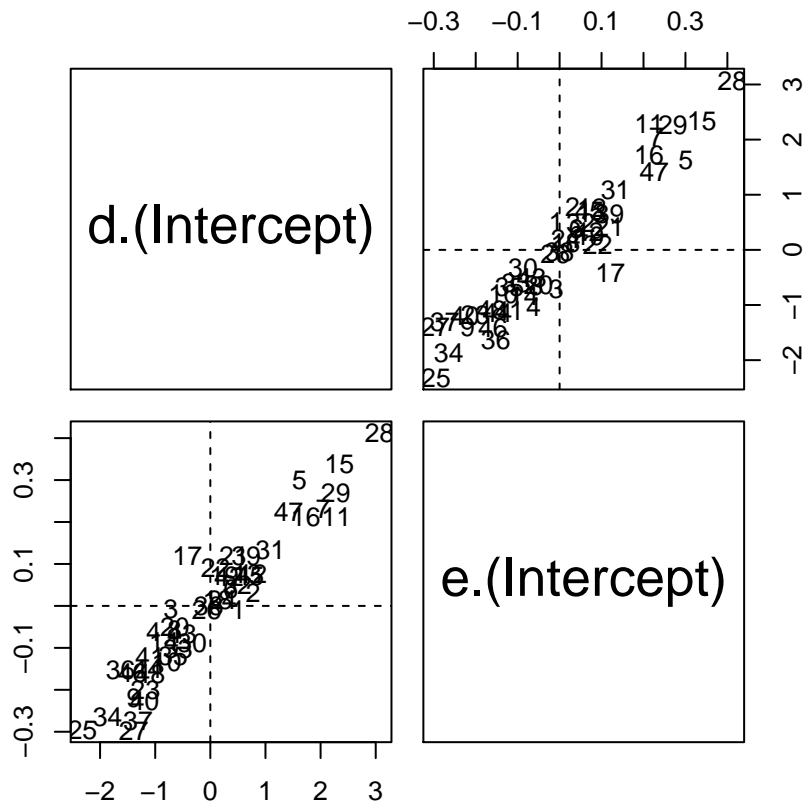
```
re <- ranef(m5pl)[[1]]
head(re)

##   d.(Intercept) e.(Intercept)
## 1        0.5049     -0.007898
## 2        0.7811      0.031048
## 3       -0.7137     -0.007855
## 4       -1.0287     -0.061984
## 5        1.6206      0.301007
## 6        0.3744      0.039826

panellab <- function(x, y, ...){
  abline(h=0, lty=2)
  abline(v=0, lty=2)
  text(x, y, rownames(re))
}
pairs(re, panel = panellab)
```
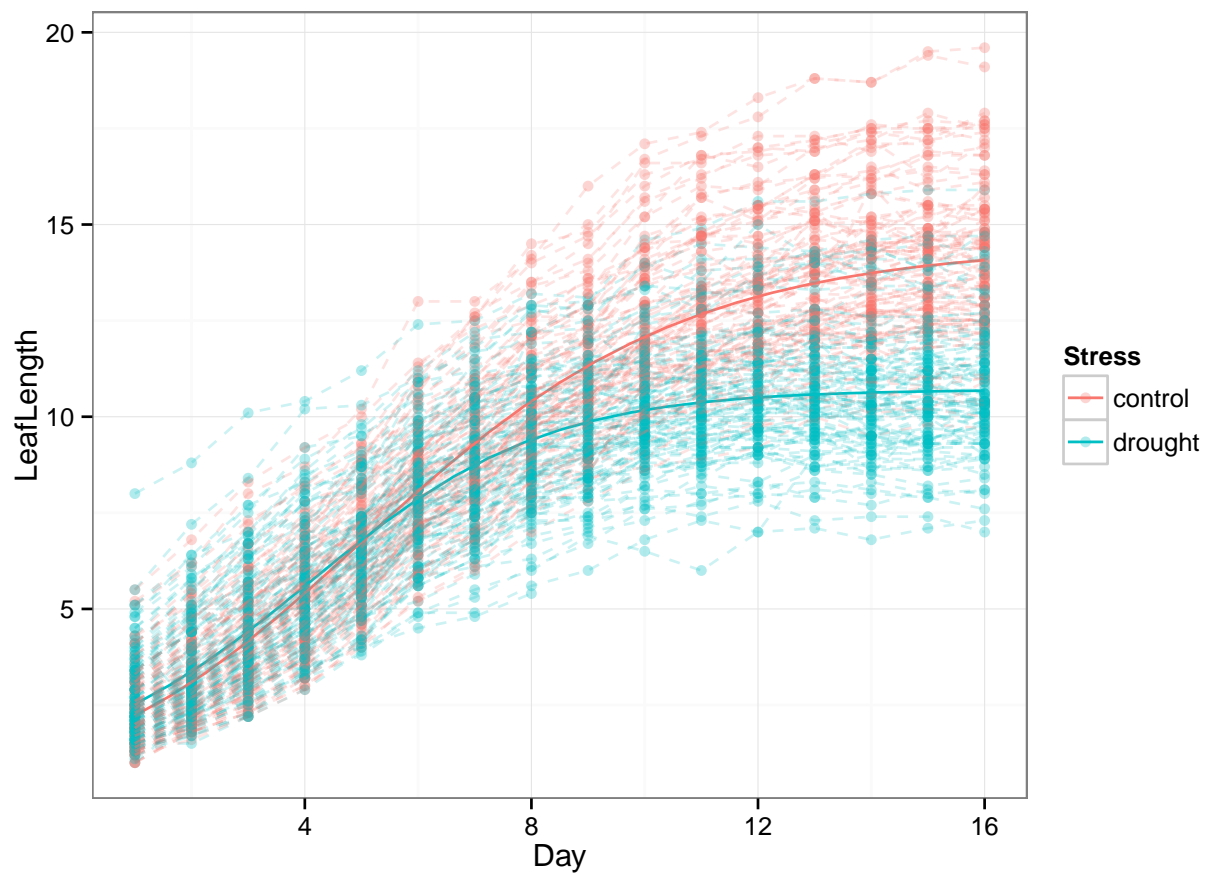
The growth curves based just on the fixed effects of the model can also be plotted. This ggplot object can be extended by some further ggplot2 functions. A plot of curves, conditional on the random effects, is not yet implemented when several fixed effects curves are present.

```
plot(m5pl) +
  geom_line(data=bro, aes(group=ID), linetype=2, alpha=0.2) +
  theme_bw()
```

Other diagnostic graphics, like residuals vs. fitted values, are also easily available.

```
plot(residuals(m5pl) ~ fitted(m5pl))
abline(h = 0, lty = 2, col = "red3")
```
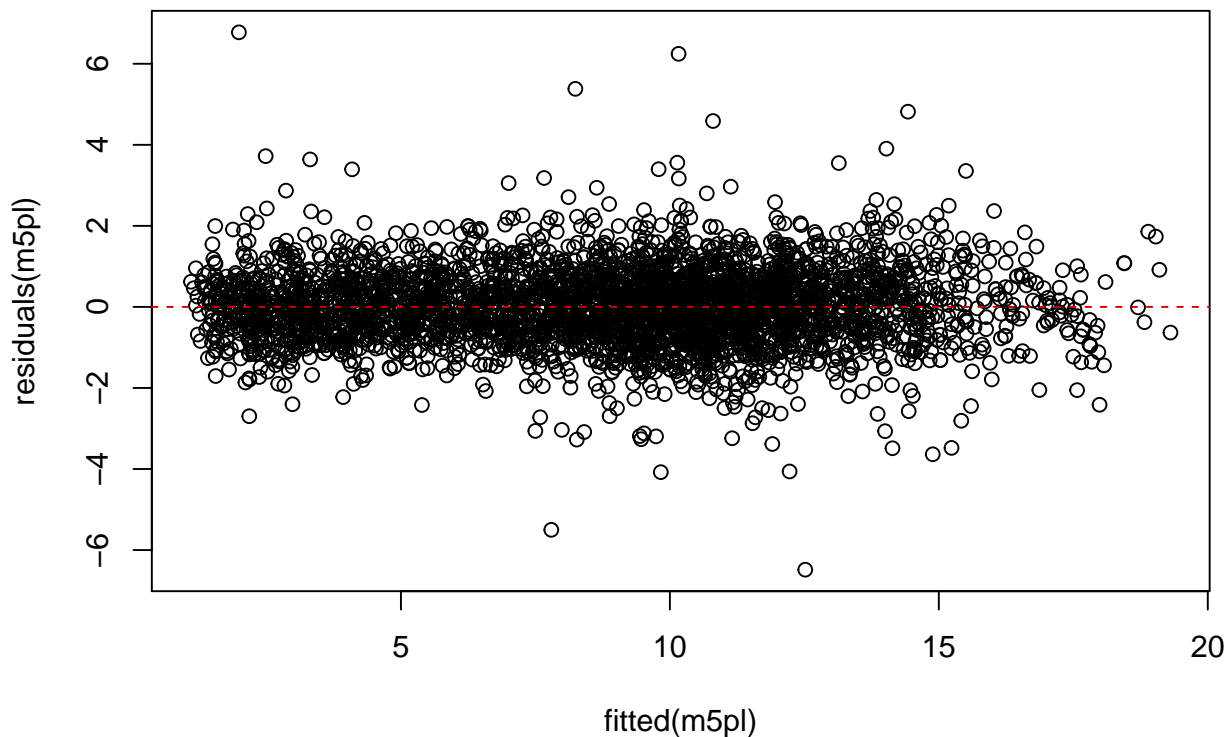
4 Sets of dose-response models

By specifying the nonlinear function $f(x_{ij}, \boldsymbol{\beta}_i)$ the principle shape of the curve is treated as known. Without any prior knowledge about the progress of the dose-response curve it is reasonable to choose from a larger set of dose-response models instead of assuming a single fixed function. With a set of candidate models a specific dose-response relationship can be selected based on the data, or model averaging approaches can be used to incorporate the model uncertainty into the parameter inference.

The framework of package drc allows to fit several dose-response models to the same data in an automated fashion, with several, predefined dose-response curves, all of them composed of a similar set of similar defined parameters with lower and upper asymptotes, steepness, inflection points, etc. As these model functions can be directly used in medrc, the formula interfaces enable the composition of a set of models with different fixed and random effect parameterizations and different dose-response shapes.

For example, a set of different log-logistic and Weibull models can be fitted for the broccoli data.

```
# 3 parameter logistic with lower asymptote fixed at 0
mod1 <- medrm(LeafLength ~ Day, data=bro,
              fct=L.3(),
              curveid=b + d + e ~ Stress,
              random=d + e ~ 1|Genotype/ID)
# 4 parameter logistic
# with the same lower asymptote for both stress treatments
mod2 <- medrm(LeafLength ~ Day, data=bro,
              fct=L.4(fixed=c(NA, 5, NA, NA)),
              curveid=b + d + e ~ Stress,
              random=d + e ~ 1|Genotype/ID)
# 4 parameter Weibull model
mod3 <- medrm(LeafLength ~ Day, data=bro,
              fct=W1.4(),
              curveid=b + d + e ~ Stress,
```

```
                  random=d + e ~ 1|Genotype/ID)
# 2nd parameterization of 4 parameter Weibull model
mod4 <- medrm(LeafLength ~ Day, data=bro,
              fct=W2.4(),
              curveid=b + d + e ~ Stress,
              random=d + e ~ 1|Genotype/ID)
# even a 4 parameter logistic with same parameters
# for both stress treatments is available with a onesided curveid formula
mod5 <- medrm(LeafLength ~ Day, data=bro,
              fct=L.4(),
              curveid= ~ Stress,
              random=d + e ~ 1|Genotype/ID)
# or a 4p-log-logistic model with a lower asymptote fixed at 5
mod6 <- medrm(LeafLength ~ Day, data=bro,
              fct=L.4(fixed=c(NA, 5, NA, NA)),
              curveid=d + e ~ Stress,
              random=d + e ~ 1|Genotype/ID)
```

The different curves can be displayed by the function mmplot().

```
mmplot(mod1, mod2, mod3, mod4, mod5, mod6, ndose=50)
```

5   Parameter inference

As the medrc class provides a coef and vcov method, simultaneous confidence intervals and multiple tests for the fixed effects are directly provided by the add-on packages multcomp. By multiple contrast tests, the comparison of drought stress to the control can be performed for each fixed effect parameter (except the lower asymptote).

```
library(multcomp)
K <- rbind("drought-control | b"=c(-1, 1,  0, 0,  0, 0,  0, 0,  0, 0),
           "drought-control | d"=c( 0, 0,  0, 0, -1, 1,  0, 0,  0, 0),
           "drought-control | e"=c( 0, 0,  0, 0,  0, 0, -1, 1,  0, 0),
           "drought-control | f"=c( 0, 0,  0, 0,  0, 0,  0, 0, -1, 1))
gg <- glht(m5pl, linfct=K)
summary(gg)

##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: medrm(form = LeafLength ~ Day, curveid = b + d + e + f ~ Stress,
##     data = bro, fct = L.5(), random = d + e ~ 1 | Genotype/ID)
##
## Linear Hypotheses:
##                          Estimate Std. Error z value Pr(>|z|)
## drought-control | b == 0  -0.1694     0.0113  -14.95  < 1e-05 ***
## drought-control | d == 0  -3.7550     0.1835  -20.46  < 1e-05 ***
## drought-control | e == 0   1.7395     0.4559    3.82  0.00041 ***
## drought-control | f == 0  -1.2206     0.2467   -4.95  < 1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

6   Inference for derived parameters

Instead of interpreting the model parameters directly, we can set focus on derived parameters, like the effective dose at a specific level $ED(p)$ or the ratio of two effective doses, known as relative potency. Inference for

the $ED(p)$ parameters are available by using the Delta-method to approximate the variance-covariance of the derived parameters.

The growth curve of the broccoli leaf lengths can be summarized, for example, by the effective dose, or in our case by the estimated day, at which 50% or any other percentage of the full length of a leaf is reached. Inference based on the fixed effects can be directly obtained by using the function ED in the drc package.

```
ED(m5pl, respLev = c(25, 50, 75), interval = "delta")

##
## Estimated effective doses
## (Delta method-based confidence interval(s))
##
##             Estimate Std. Error Lower Upper
## control:25    2.835      0.147 2.547  3.12
## control:50    5.542      0.125 5.297  5.79
## control:75    8.584      0.128 8.333  8.83
## drought:25    1.632      0.164 1.312  1.95
## drought:50    3.997      0.125 3.751  4.24
## drought:75    6.298      0.120 6.064  6.53
```

Comparisons of these ED parameters can be made with the drc function SI.

```
SI(m5pl, percVec = c(25, 25), interval = "delta")

##
## Estimated ratios of effect doses
## (Delta method-based confidence interval(s))
##
##                        Estimate Lower Upper
## control/drought:25/25      1.74  1.45  2.03
```

When a set of models are available, the model-averaged ED estimates can be estimated by the medrc function mmaED(). The model specific estimates are weighted by the corresponding information criteria for each model, e.g. using Akaike weights. It is recommended to use maximum likelihood estimation instead of REML, and not mixing simultaneous changes of fixed and random effect structures between the models.

```
mmaED(mod1, mod2, mod3, mod4, mod5, mod6, respLev = c(25, 50, 75), interval = "kang")
```

7 Marginal vs. subject specific predictions

Other than in linear models, the fixed effect parameter in nonlinear mixed effect models can not be interpreted independently from the random effects as a vector of population parameters. Instead only a subject specific interpretation conditional on the random effects is generally available, as

$$E\{f(x_{ij}, \boldsymbol{\beta}_i))\} \neq f(x_{ij}, \boldsymbol{A}_i \boldsymbol{\beta}).$$

To obtain the marginal mean response nevertheless from a hierarchical nonlinear model, a more or less complex integral has to be solved. The marginal expectation for a model with a single fixed effect curve and uncorrelated random effects is equal to

$$E\{f(x_{ij}, \boldsymbol{\beta}_i)\} = \int \cdots \int f(x_{ij}, (\beta_1 + b_1, \dots, \beta_p + b_p)) \, \Phi(b_1) \cdots \Phi(b_p) \, db_1 \cdots db_p$$

where $\Phi()$ denotes the univariate Gaussian density with mean 0 and the variance equal to the diagonal entries in $\boldsymbol{G}$. The multiple integral can be approximated by the weighted sum

$$E\{f(x_{ij}, \boldsymbol{\beta}_i)\} \approx \sum_{n=1}^{N} w_n f(x_{ij}, (\beta_1 + \xi_{1n}, \dots, \beta_p + \xi_{pn})), \quad \text{with} \quad w_n = \prod_{r=1}^{p} w_{rn}$$

using numerical quadrature, with a $(p \times N)$ grid of nodes ξ_{rn} and corresponding weights w_{rn}. For correlated random effects the nodes can be rotated by a matrix of eigenvectors, transforming the matrix of random effects into a set of uncorrelated variables.

The medrc package uses package statmod to calculate the nodes and weights for Gauss-Hermite quadrature.

7.1 Marginal effective dose estimation

We might want to marginalize the ED of the broccoli model conditional on the estimated variance components by numerical integration methods. The medrc package provides functions EDmarg and SImarg, but assuming only simple random effect structures.

```
EDmarg(m5pl, respLev = c(25, 50, 75), interval = "delta", nGQ = 3)

##
## Estimated effective doses
## (Delta method-based confidence interval(s))
##
##           Estimate Std. Error Lower Upper
## control:25   2.762      0.111 2.544  2.98
## control:50   5.599      0.125 5.353  5.84
## control:75   8.743      0.203 8.346  9.14
## drought:25   1.522      0.116 1.294  1.75
## drought:50   4.034      0.131 3.776  4.29
## drought:75   6.490      0.204 6.090  6.89

SImarg(m5pl, percVec = c(25, 25), interval = "delta", nGQ = 3)

##
## Estimated ratios of effect doses
## (Delta method-based confidence interval(s))
##
##                       Estimate Lower Upper
## control/drought:25/25     1.81  1.52  2.11
```

In this case there are only small changes compared to the fixed effects estimates; a difference can be observed, when the variance component for the steepness, the location of the inflection point increases, or a variance component for the asymmetry parameter is introduced into the model.

References

M. Davidian and D.M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall, 1995. ISBN 9780412983412.

Jose.C. Pinheiro and Douglas.M. Bates. *Mixed-Effects Models in S and S-Plus*. Statistics and Computing. Springer, 2000. ISBN 9780387989570.

Christian Ritz and Jens C Streibig. Bioassay analysis using R. *Journal of Statistical Software*, 12(5):1–22, 2005.