

# Fitting GLMM trees in R

R scripts accompanying the webinar for University of Montreal

Marjolein Fokkema

06-05-2021

## Model formulas

The GLMM is given by:

$$E[y_{ij}|x_{ij}] = \mu_{ij}$$

$$g(\mu_{ij}) = x_{ij}^T \beta_k + z_{ij}^T b_i$$

Where  $i$  is an identifier for the level-II unit, and  $j$  is an identifier for the level-I unit.

If we have a continuous response variable with normally distributed residuals, then  $g$  is the identity function and we have:

$$y_{ij} = x_{ij}^T \beta + z_{ij}^T b_i + \epsilon_{ij}$$

We thus have a fixed-effects part ( $x_{ij}^T \beta$ ), a random-effects part ( $z_{ij}^T b_i$ ) and a residual error term ( $\epsilon_{ij}$ ).

The GLMM tree model differs in that the fixed-effect part may differ between subgroups:

$$y_{ij} = x_{ij}^T \beta_k + z_{ij}^T b_i + \epsilon_{ij}$$

Thus,  $k$  is an identifier for the subgroup. The GLMM tree algorithm finds these subgroups  $k$ , using additional covariates. These may be measured on the lowest level  $i$  (which is commonly encountered in cross-sectional multilevel data), or on the higher level  $j$  (which is commonly encountered in longitudinal data).

The subgroups  $k$ , and parameters  $\beta$  and  $b$  cannot be estimated in a single step. An iterative approach is taken, where the model-base recursive partitioning algorithm of Zeileis, Hothorn & Hornik (2008) is used to estimate the subgroups  $k$ , and the usual (restricted) maximum likelihood approach is used to estimate the fixed- and random-effects parameters, see Fokkema et al. (2018).

## Example: Treatment subgroups

For this example, we will make use of artificial data modelled after the Stratified Medicine Approaches for Treatment Selection (SMART) prediction tournament, from the Improving Access to Psychological Therapies (IAPT) project (Lucock et al., 2017). The SMART data contains data from patients receiving mental-health services in the Northern UK. Patients were (non-randomly) assigned to low intensity treatment (e.g., guided self-help, computerized cognitive behavior therapy) or high intensity treatment (e.g., face-to-face psychological therapies). The aim of the SMART tournament was to identify patients who would benefit most from HI vs. LI treatment. I do not own the data, so I generated an artificial dataset which mimics the original data, available in the file “SMART mimic data.txt”.

```
library("glmertree")
SMART <- read.table("SMART mimic data.txt", stringsAsFactors = TRUE)
names(SMART)

## [1] "recovered" "Treatment" "Age" "Gender" "Ethnicity"
## [6] "Diagnosis" "Employment" "Disability" "PHQ9_pre" "GAD7_pre"
## [11] "WSAS_pre" "Medication" "center"

trt_tree <- glmertree(recovered ~ Treatment | center | Age + PHQ9_pre +
                      GAD7_pre + WSAS_pre + Gender + Ethnicity +
                      Diagnosis + Employment + Disability + Medication,
                      data = SMART, family = binomial)
```

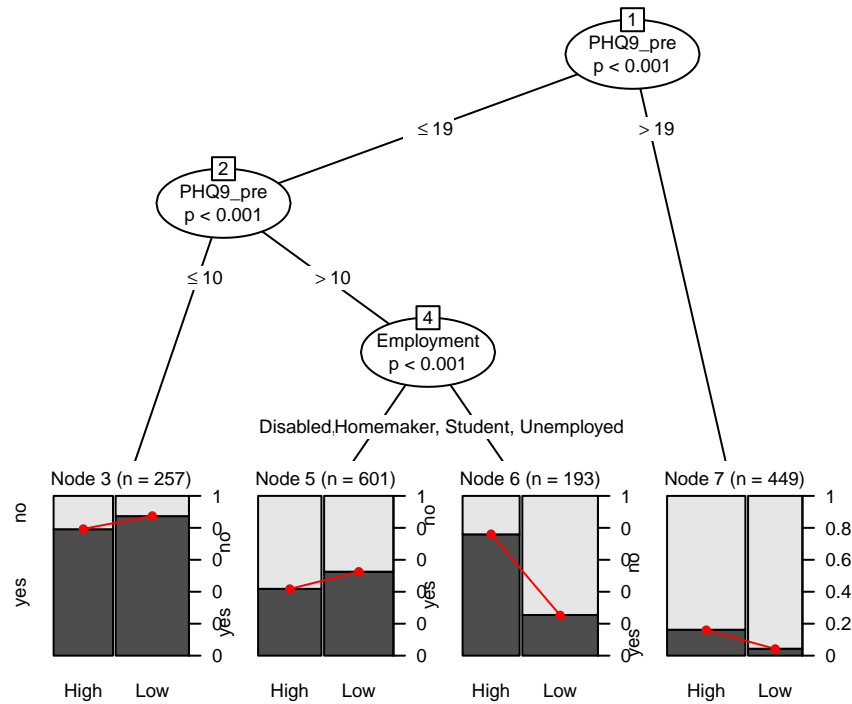
We can print and plot the results as follows:

```
trt_tree$tree

## Generalized linear model tree (family: binomial)
##
## Model formula:
## recovered ~ Treatment | Age + PHQ9_pre + GAD7_pre + WSAS_pre +
##      Gender + Ethnicity + Diagnosis + Employment + Disability +
##      Medication
##
## Fitted party:
## [1] root
## |   [2] PHQ9_pre <= 19
## |   |   [3] PHQ9_pre <= 10: n = 257
## |   |   (Intercept) TreatmentLow
## |   |   1.5255853 0.5577717
## |   |   [4] PHQ9_pre > 10
## |   |   |   [5] Employment in Disabled, Employed, Retired: n = 601
## |   |   |   (Intercept) TreatmentLow
## |   |   |   -0.2975380 0.4358343
## |   |   |   [6] Employment in Homemaker, Student, Unemployed: n = 193
## |   |   |   (Intercept) TreatmentLow
## |   |   |   1.096908 -2.359766
## |   [7] PHQ9_pre > 19: n = 449
## |   (Intercept) TreatmentLow
## |   -1.870597 -1.529347
##
```

```
## Number of inner nodes: 3
## Number of terminal nodes: 4
## Number of parameters per node: 2
## Objective function (negative log-likelihood): 699.7975
```

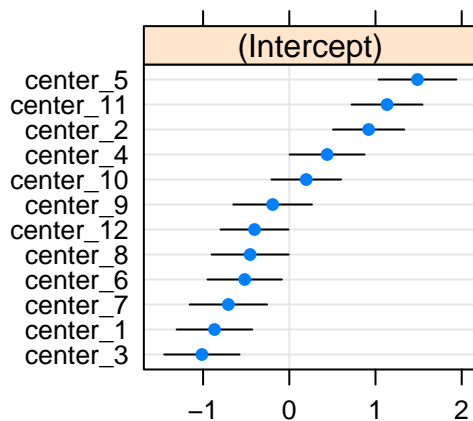
```
plot(trt_tree, which = "tree", gp = gpar(cex = .6))
```



```
plot(trt_tree, which = "ranef")
```

```
## $center
```

## center



```
VarCorr(trt_tree)
```

```
## Groups Name      Std.Dev.  
## center (Intercept) 0.82557
```

```
fixef(trt_tree)
```

```
## (Intercept) TreatmentLow  
## 3  1.5406327  0.5620038  
## 5 -0.2995581  0.4400173  
## 6  1.1087452 -2.3810215  
## 7 -1.8839567 -1.5372864
```

## Example: Alcohol trajectories

Curran, Stice, and Chassin (1997) collected data on 82 adolescents at three time points starting at age 14 to assess factors that affect teen drinking behavior. Key variables in the data set “alcohol.csv” (accessed via Singer and Willett, 2003) are as follows:

- `id` = numerical identifier for subject
- `age` = 14, 15, or 16
- `coa` = 1 if the teen is a child of an alcoholic parent; 0 otherwise
- `male` = 1 if male; 0 if female
- `peer` = a measure of peer alcohol use, taken when each subject was 14. This is the square root of the sum of two 6-point items about the proportion of friends who drink occasionally or regularly.
- `alcuse` = the primary response. Four items—(a) drank beer or wine, (b) drank hard liquor, (c) 5 or more drinks in a row, and (d) got drunk—were each scored on an 8-point scale, from 0=“not at all” to 7=“every day”. Then `alcuse` is the square root of the sum of these four items.

Primary research questions included: Do trajectories of alcohol use differ by parental alcoholism? Do trajectories of alcohol use differ by peer alcohol use?

```
alco <- read.table("alcohol.csv", header= TRUE, sep = ",",
                  stringsAsFactors = TRUE)[ , -1]
summary(alco)
```

```
##           id           age           coa           male           peer
## Min.      : 1.0    Min.    :14    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
## 1st Qu.:21.0    1st Qu.:14    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000
## Median :41.5    Median :15    Median :0.0000    Median :1.0000    Median :0.8944
## Mean   :41.5    Mean   :15    Mean   :0.4512    Mean   :0.5122    Mean   :1.0176
## 3rd Qu.:62.0    3rd Qu.:16    3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:1.5492
## Max.   :82.0    Max.   :16    Max.   :1.0000    Max.   :1.0000    Max.   :2.5298
##      alcuse
## Min.    :0.000
## 1st Qu.:0.000
## Median :1.000
## Mean   :0.922
## 3rd Qu.:1.732
## Max.   :3.606
```

```
alco$age <- alco$age - 14L
```

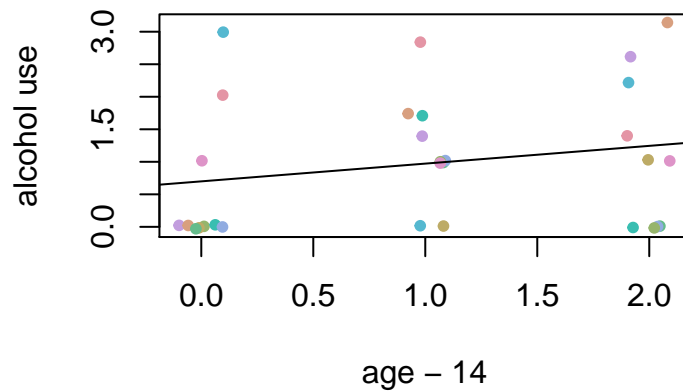
Fit and plot a fixed-effects linear model (note: I used only a subset of the data to keep plot readable) :

```
set.seed(42)
ids <- sample(unique(alco$id), 10)
cols <- colorspace::rainbow_hcl(length(ids))
alc <- alco[alco$id %in% ids, ]
LM <- lm(alcuse ~ age, data = alc)
summary(LM)
```

```
##
## Call:
```

```
## lm(formula = alcuse ~ age, data = alc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2451 -0.6993 -0.1086  0.6804  2.3007
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6993     0.2996   2.334   0.027 *
## age           0.2729     0.2321   1.176   0.249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.038 on 28 degrees of freedom
## Multiple R-squared:  0.04707,    Adjusted R-squared:  0.01304
## F-statistic: 1.383 on 1 and 28 DF,  p-value: 0.2495
```

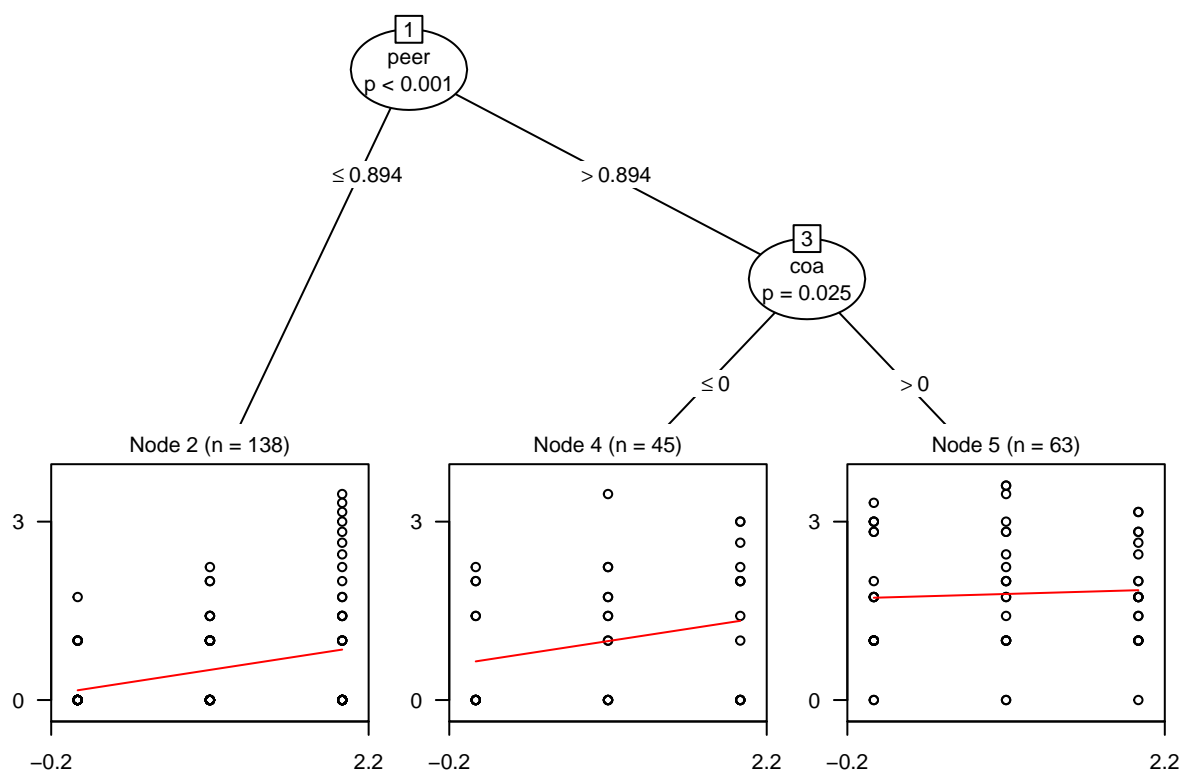
```
beta <- coef(LM)
plot(jitter(alc$age, 0.5), jitter(alc$alcuse), col = cols, pch = 20,
     xlab = "age - 14", ylab = "alcohol use")
abline(beta[1], beta[2], cex = 3)
```



Fit and plot linear model tree (note: here

I used the full dataset):

```
library("partykit")
lt <- lmtree(alcuse ~ age | coa + male + peer, data = alco)
plot(lt, gp = gpar(cex = 0.7))
```



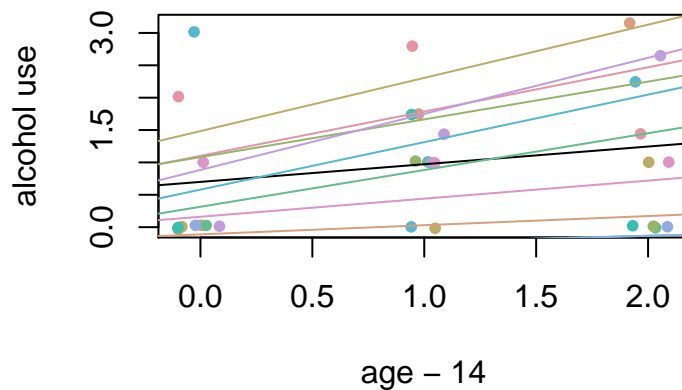
Fit and plot linear mixed-effects model (again, I used only a subset of the original data):

```
LMM <- lmer(alcuse ~ age + (age|id), data = alc)
summary(LMM)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: alcuse ~ age + (age | id)
## Data: alc
##
## REML criterion at convergence: 80.5
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -1.2732 -0.6618 -0.1832  0.6578  2.1834
##
## Random effects:
##   Groups   Name      Variance Std.Dev. Corr
##   id       (Intercept) 0.38433  0.6199
##           age         0.05946  0.2438  0.44
## Residual                0.48226  0.6945
## Number of obs: 30, groups: id, 10
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   0.6993    0.2804    2.494
## age           0.2729    0.1734    1.574
```

```
##
## Correlation of Fixed Effects:
##      (Intr)
## age -0.359

beta <- fixef(LMM)
b <- ranef(LMM)$id
plot(jitter(alc$age, 0.5), jitter(alc$alcuse), col = cols, pch = 20,
     xlab = "age - 14", ylab = "alcohol use")
abline((beta[1]), beta[2], cex = 3)
for (j in 1:length(ids)) {
  abline((b+beta)[j,1], (b+beta)[j,2], col = cols[j])
}
```



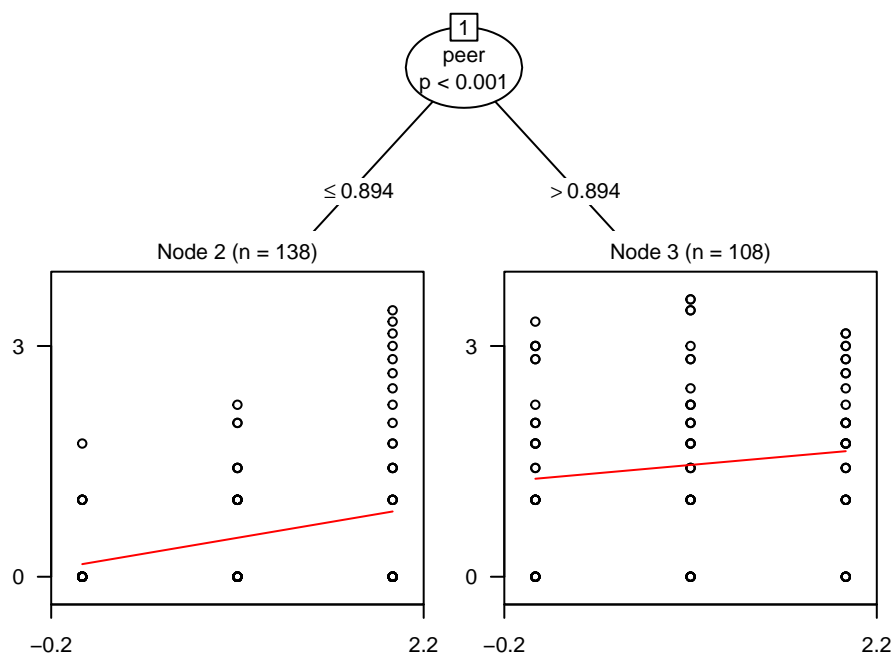
Fit and plot a linear mixed-effects model tree:

```
lt <- lmertree(alcuse ~ age | (age|id) | coa + male + peer, data = alco,
              cluster = id)
```

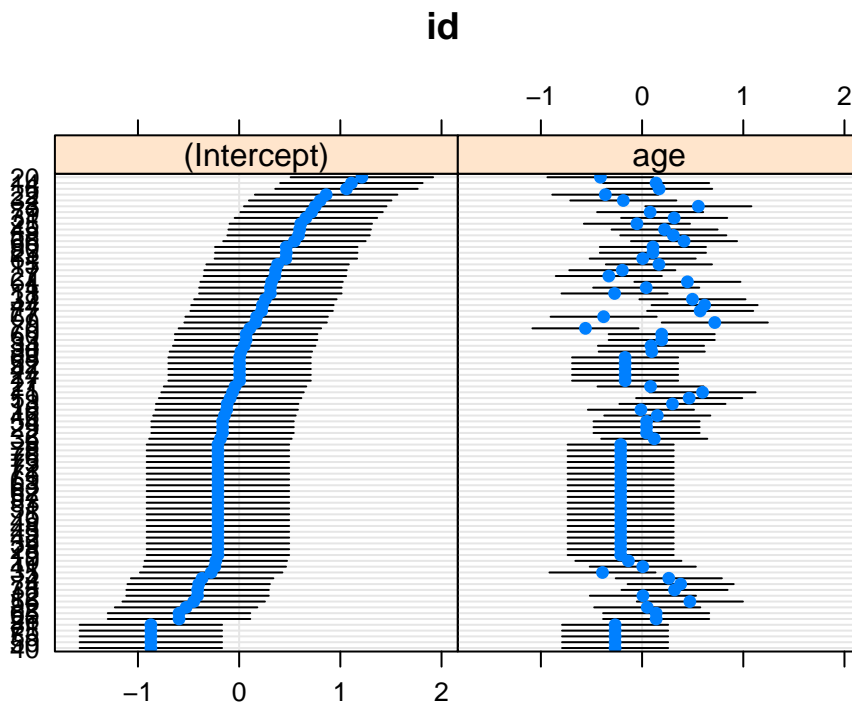
Note that because the possible partitioning variables are measured on level II, we use the `cluster` argument to specify this to the algorithm (otherwise, the parameter stability tests will assume partitioning variables are measured at level I and the tests are likely overpowered).

```
plot(lt, fitted = "marginal", gp = gpar(cex = 0.7))
```





## \$id



I specified 'fitted = "marginal", so the lines in the terminal node represent the effect of time, while fixing all other predictors (fixed and random, but not the potential partitioning variables) at their means.

The plots of the random effects look quite messy, because of the large number of study participants. We can request the (co)variances of the random effects as follows:

```
fixef(lt)
```

```
##      (Intercept)      age
## 2      0.1638493 0.3423005
## 3      1.2741616 0.1790998
```

```
VarCorr(lt)
```

```
## Groups   Name      Std.Dev. Corr
## id      (Intercept) 0.57947
##          age         0.39048 -0.127
## Residual              0.58077
```

## Example: Stage fright trajectories

Sadler and Miller (2010) studied the emotional state of musicians before performances and factors which may affect their emotional state. Data was collected among 37 undergraduate music majors from a competitive undergraduate music program. They filled out diaries prior to performances over the course of an academic year. Specifically, participants completed a Positive Affect Negative Affect Schedule (PANAS) before each performance, providing two key outcomes: negative affect (**na**, a state measure of anxiety) and positive affect (**pa**, a state measure of happiness).

Factors which were examined for their potential relationships with performance anxiety included: performance type (solo, large ensemble, or small ensemble); audience (instructor, public, students, or juried); if the piece was played from memory; age; gender; instrument (voice, orchestral, or keyboard); and years studying the instrument. In addition, the personalities of study participants were assessed at baseline through the Multidimensional Personality Questionnaire (MPQ). The MPQ provided scores for one lower-order factor (absorption) and three higher-order factors: positive emotionality (PEM); negative emotionality (NEM); and constraint.

Here, we look at trajectories of negative affect scores over the course of repeated assessments at solo performances.

```
music <- read.table("musicdata.csv", header=T, sep="," ,
                    stringsAsFactors = TRUE)[-1]
summary(music)
```

```
##          id          diary          previous          perform_type
## Min.      : 1.00   Min.      : 1.000   Min.      : 0.000   Large Ensemble:136
## 1st Qu.:10.00   1st Qu.: 4.000   1st Qu.: 3.000   Small Ensemble: 82
## Median :22.00   Median : 8.000   Median : 7.000   Solo           :279
## Mean      :22.24   Mean      : 7.781   Mean      : 6.781
## 3rd Qu.:34.00   3rd Qu.:11.000   3rd Qu.:10.000
## Max.      :43.00   Max.      :15.000   Max.      :14.000
##          memory          audience          pa          na
## Memory      :149   Instructor      :149   Min.      :11.00   Min.      :10.00
## Score        :274   Juried Recital    : 44   1st Qu.:26.00   1st Qu.:12.00
## Unspecified: 74   Public Performance:204   Median :32.00   Median :15.00
##              Student(s)      :100   Mean      :32.27   Mean      :16.21
##              3rd Qu.:38.00   3rd Qu.:19.00
##              Max.      :50.00   Max.      :35.00
##          age          gender          instrument          years_study
## Min.      :18.00   Female:339   keyboard (piano or organ): 75   Min.      : 2.000
## 1st Qu.:19.00   Male  :158   orchestral instrument      :235   1st Qu.: 5.000
## Median :20.00              voice      :187   Median : 7.000
## Mean      :19.75
## 3rd Qu.:21.00
## Max.      :22.00
##          mpqab          mpqsr          mpqpem          mpqnem
## Min.      : 7.00   Min.      : 1.00   Min.      :23.00   Min.      :11.00
## 1st Qu.:17.00   1st Qu.: 7.00   1st Qu.:41.00   1st Qu.:24.00
## Median :23.00   Median :11.00   Median :52.00   Median :28.00
## Mean      :21.99   Mean      :11.84   Mean      :48.14   Mean      :31.63
## 3rd Qu.:27.00   3rd Qu.:16.00   3rd Qu.:57.00   3rd Qu.:40.00
## Max.      :31.00   Max.      :22.00   Max.      :68.00   Max.      :50.00
##          mpqcon
## Min.      :22.00
```

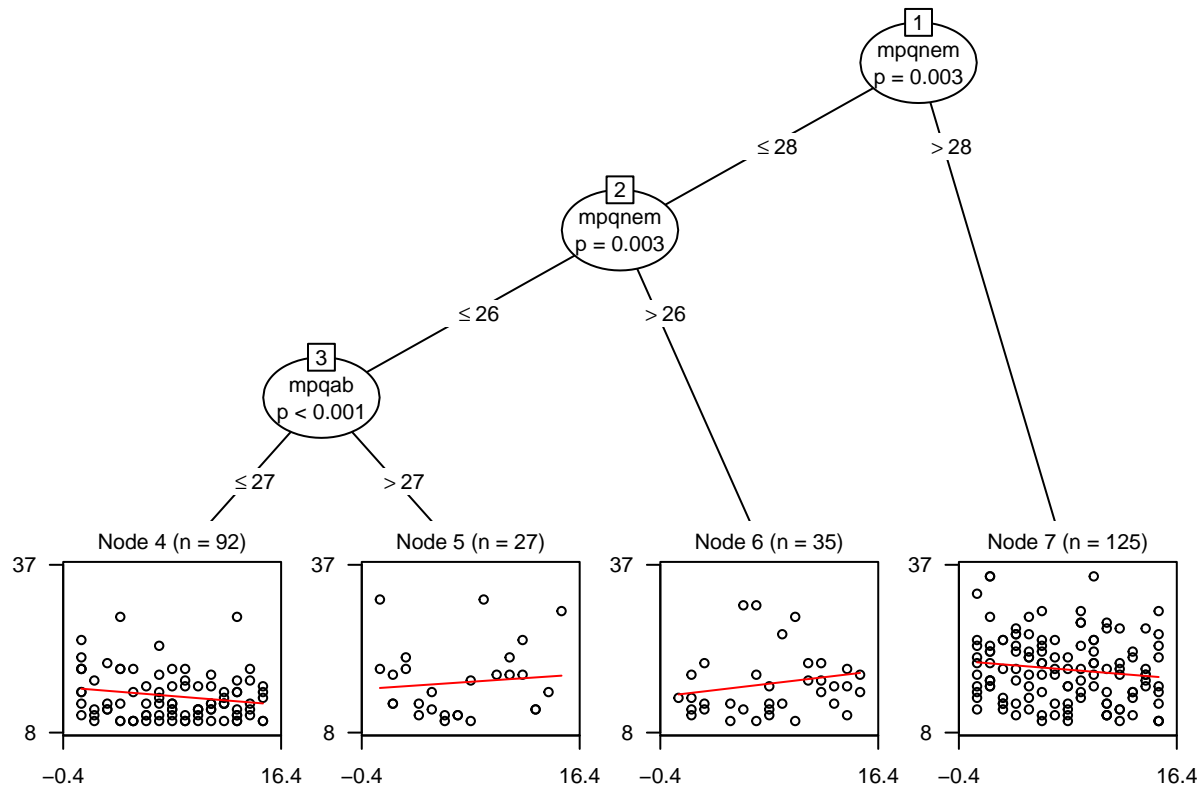
```
## 1st Qu.:39.00
## Median :52.00
## Mean   :51.83
## 3rd Qu.:64.00
## Max.   :88.00
```

```
music <- music[music$perform_type == "Solo", ]
levels(music$instrument) <- c("keyboard", "orch_instr", "voice")
```

- `id` = unique musician identification number
- `diary` = cumulative total of diaries filled out by musician (level I; timing metric)
- `audience` = who attended performance (Instructor, Public, Students, or Juried) (level I)
- `na` = negative affect score from PANAS (level I)
- `gender` = musician gender (level II)
- `instrument` = Voice, Orchestral, or Piano (level II)
- `mpqab` = absorption subscale from MPQ (level II)
- `mpqpem` = positive emotionality (PEM) composite scale from MPQ (level II)
- `mpqnem` = negative emotionality (NEM) composite scale from MPQ (level II)
- `mpqcon` = constraint scale from MPQ (level II)

Here, we fit both random intercepts and slopes of time (`diary`). These random intercepts and slopes are deviations from the subgroup-specific intercept and slope:

```
lmm1 <- lmertree(na ~ diary | (diary|id) | gender + instrument +
  mpqab + mpqpem + mpqnem + mpqcon, data = music, cluster = id)
plot(lmm1, fitted = "marginal", which = "tree", gp = gpar(cex = .7))
```



```
fixef(lmmt1, which = "tree")
```

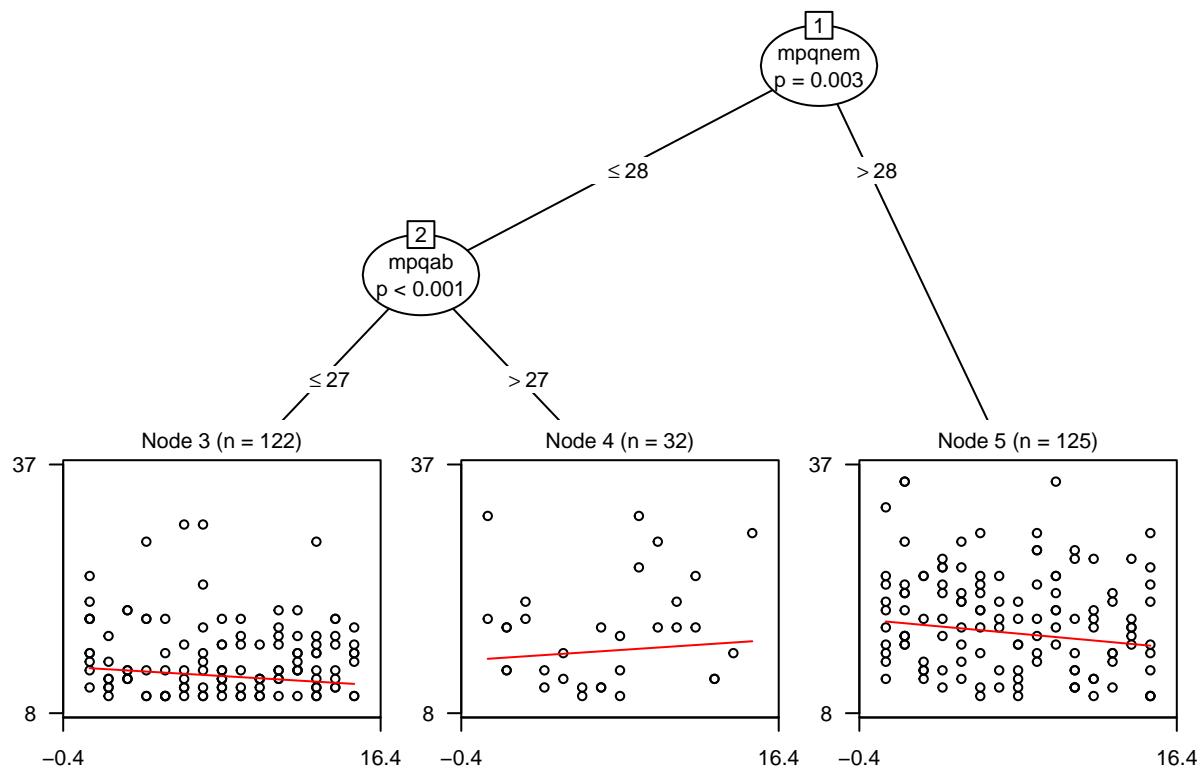
```
## (Intercept)      diary
## 4      15.76600 -0.1815890
## 5      15.56076  0.1519619
## 6      14.32766  0.2654400
## 7      20.33042 -0.1817579
```

```
VarCorr(lmmt1)
```

```
## Groups   Name      Std.Dev. Corr
## id       (Intercept) 2.60098
##         diary       0.29835 -0.728
## Residual                      4.80919
```

We can also account for time-varying covariates. For example, the variable **audience** has a strong effect on negative affect, and should be accounted for. Time-varying predictors can either be included in the subgroup-specific model, which we would do if we expect and are interested in possible between-subgroup differences, but note that this makes the resulting subgroups more difficult to interpret, because they can differ in terms of several parameters. So it might be better to globally correct/account for the effect of **audience**:

```
lmmt2 <- lmertree(na ~ diary | audience + (diary|id) | gender + instrument +
                  mpqab + mpqpem + mpqnem + mpqcon, data = music, cluster = id)
plot(lmmt2, fitted = "marginal", which = "tree", gp = gpar(cex = .7))
```



```
fixef(lmmt2, which = "tree")
```

```
##      (Intercept)      diary  
## 3      13.37823 -0.1308454  
## 4      14.19399  0.1460969  
## 5      18.89028 -0.2011977
```

```
fixef(lmmt2, which = "global")
```

```
##      audienceJuried Recital audiencePublic Performance  
##              3.983440              3.288610  
##      audienceStudent(s)  
##              4.072934
```

```
VarCorr(lmmt2)
```

```
## Groups   Name      Std.Dev. Corr  
## id      (Intercept) 2.26847  
##        diary      0.25583 -0.736  
## Residual              4.51190
```

## Further reading

The vignette of package **glmertree** provides further info on how the GLMM tree models can be further customized, and checks on model fit can be performed. You can access it on <https://cran.r-project.org/web/packages/glmertree/vignettes/glmertree.pdf> or in **R** by typing:

```
vignette("glmertree")
```

Furthermore, Fokkema, Smits, Zeileis, Hothorn & Kelderman (2018) provides an in-depth technical discussion of GLMM trees, while Fokkema, Edbrooke-Childs & Wolpert (2020) provides a less technical introduction.

## From the future: GAM trees

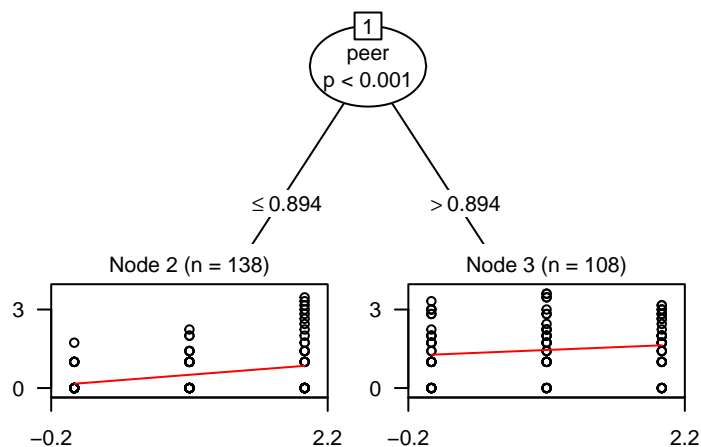
The GAM tree package allows for detecting subgroups in non-linear trajectories. However, it is still experimental. Instead of fitting GLMMs, it fit GAMs, which allows for fitting non-linear smoothing splines, as well as random effects. For an in-depth discussion of GAMs, see Wood (2017). Or Chapter 7 of James et al. (2013) for a less formal, more introductory version (you can download that book freely via <https://www.statlearning.com/>).

The use of GAM tree assumes some familiarity with **R** package **mgcv**. You can install the current development version as follows:

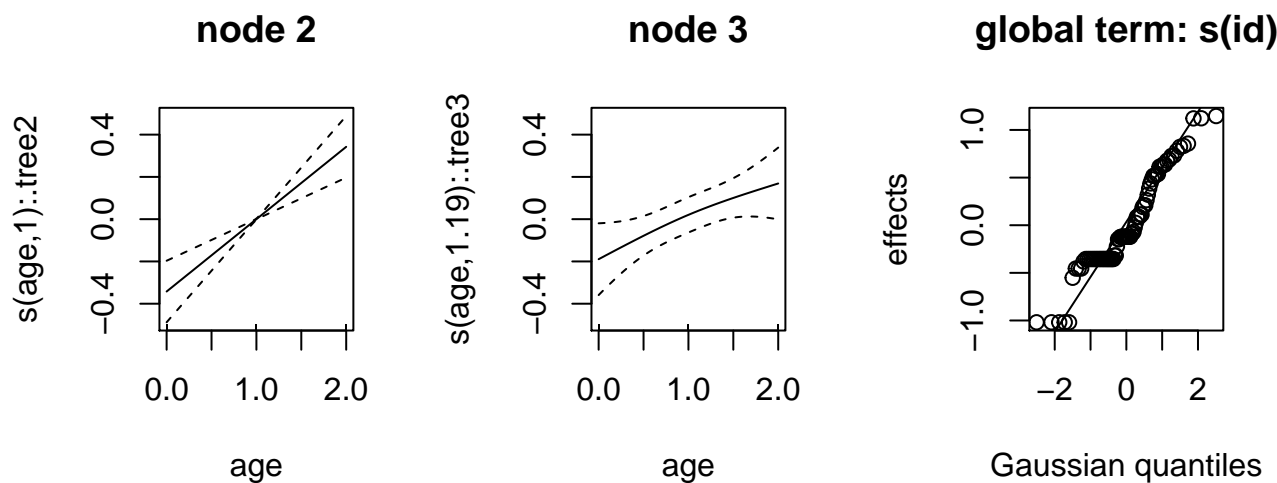
```
library("devtools")
install_github("marjoleinF/gamtree")
```

We now fit non-linear trajectories to the alcohol use data. To fit a smooth curve to a predictor, we use the `s()` function (type `?mgcv::gam` and `?mgcv::s` for more info). For fitting random effects, we use the same function, and specify that we `bs = "re"` to indicate we want to estimate a random-effect, not a non-linear curve:

```
library("gamtree")
alco$id <- factor(alco$id)
gt <- gamtree(alcuse ~ s(age, k = 3) | s(id, bs = "re") | coa + male + peer,
             data = alco, cluster = alco$id, verbose = FALSE)
plot(gt, which = "tree", treeplot_ctrl = list(gp = gpar(cex = .7)))
```



```
plot(gt, which = "nodes")
```



The fitted (marginal) curves in the tree indicate no or little non-linearity in either of the resulting subgroups (nodes). The groups seem to mostly differ in terms of the intercept and linear slope, not in terms of the shape of the effect of age. Note also the similarity to the GLMM tree we fitted earlier. Only the conditional effects plotted later indicate slight non-linearity in the node-3 subgroup.

```
summary(gt)
```

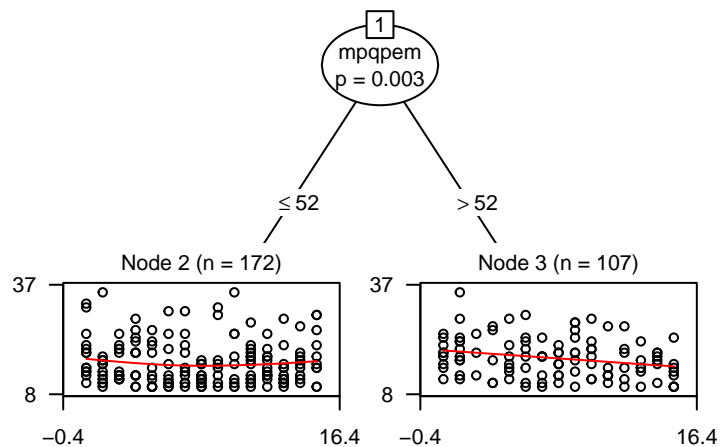
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## alcuse ~ .tree + s(age, k = 3, by = .tree) + s(id, bs = "re") -
```



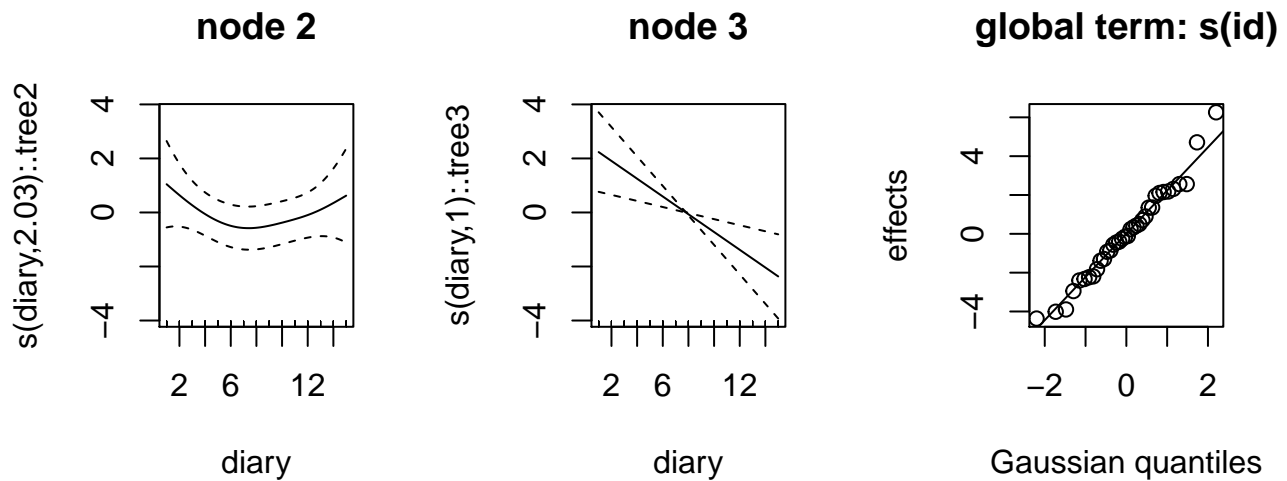
```
##      1
##
## Parametric coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## .tree2    0.5061     0.1087   4.657 6.11e-06 ***
## .tree3    1.4533     0.1229  11.828 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##      edf Ref.df      F  p-value
## s(age):.tree2  1.000  1.001 22.115 5.28e-06 ***
## s(age):.tree3  1.193  1.349  4.108  0.0501 .
## s(id)          56.095 80.000  2.347 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.569   Deviance explained = 67.3%
## -REML = 314.4   Scale est. = 0.48716    n = 246
```

We can also fit non-linear curves to the stage-fright data:

```
music$id <- factor(music$id)
gam1 <- gamtree(na ~ s(diary) | audience + s(id, bs = "re") | gender +
               instrument + mpqab + mpqpem + mpqnem + mpqcon, data = music,
               verbose = FALSE, cluster = music$id)
plot(gam1, which = "tree", treeplot_ctrl = list(gp = gpar(cex = .7)))
```



```
plot(gamt1, which = "nodes")
```



```
summary(gamt1)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## na ~ .tree + s(diary, by = .tree) + audience + s(id, bs = "re") -
##      1
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## .tree2          14.6068    0.8009  18.237 < 2e-16 ***
## .tree3          15.7690    0.8989  17.542 < 2e-16 ***
## audienceJuried Recital    3.7832    0.8879   4.261 2.90e-05 ***
## audiencePublic Performance 3.3039    0.9418   3.508 0.000536 ***
## audienceStudent(s)       4.0609    0.7366   5.513 8.87e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df      F p-value
## s(diary):.tree2  2.03  2.527 1.208 0.35772
## s(diary):.tree3  1.00  1.001 9.227 0.00264 **
## s(id)           24.07 34.000 2.794 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.346  Deviance explained = 41.9%
## -REML = 837.18  Scale est. = 20.883    n = 279
```

We see some evidence of non-linearity: Those with lower negative emotionality at baseline (node 3) show

a decrease and then an increase over the academic year. Those with higher levels of negative emotionality at baseline (node 4) show a stronger, more steady decrease of negative affect during performance, over the course of the academic year.

Note that the  $p$ -values should be taken with a (large) grain of salt, because they do not account for the searching of the subgroups.

## References

- Curran, P. J., Stice, E., & Chassin, L. (1997). The relation between adolescent alcohol use and peer alcohol use: a longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology*, 65(1), 130.
- Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, 50(5), 2016-2034. <https://doi.org/10.3758/s13428-017-0971-x>
- Fokkema, M., Edbrooke-Childs, J., & Wolpert, M. (2020). Generalized linear mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. *Psychotherapy Research*, 31(3), 329-341. <https://doi.org/10.1080/10503307.2020.1785037>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *Introduction to Statistical Learning*. New York: springer.
- Lucock, M., Barkham, M., Donohoe, G., Kellett, S., McMillan, D., Mullaney, S., . . . & Delgadillo, J. (2017). The role of Practice Research Networks (PRN) in the development and implementation of evidence: The Northern improving access to psychological therapies PRN case study. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(6), 919-931.
- Sadler, M. E., & Miller, C. J. (2010). Performance anxiety: A longitudinal study of the roles of personality and experience in musicians. *Social Psychological and Personality Science*, 1(3), 280-287.
- Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.
- Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.
- Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17(2), 492-514. <https://doi.org/10.1198/106186008X319331>