Model selection for Mixed Effects Models: Effects of fire on survival of a rare plant

In a prior demo we discussed how to implement model selection for linear Mixed Effects models. Here, we discuss model selection for Mixed Effects Models for binary data (GLM). We combine procedures described in Crawley (2007) and Zuur et al. (2009). These approaches are currently developed, examples are scarce, and documentation limited. There are several procedures in R to complete this work and their results are not always consistent (see Zuur et al. 2009; pages 323-325). We tested the outcomes for our data and results were commensurate. We encourage you to review future improvements.

We already provided evidence that number of reproductive structures of *Hypericum cumulicola* is significantly associated with plant survival. Now, we evaluate the relevance of three more fixed variables to explain survival variation in this species (Quintana-Ascencio et al. 2003). However, we decide not to include number of reproductive structures in this model because the high co-linearity between plant height and number of reproductive structures (see previous demo). As before, we evaluate the effect of height (cm), number of reproductive stems and TSF but now on *Hypericum cumulicola* survival. We use a model selection approach to assess their relative importance.

Figure 1. *Dying Hypericum cumulicola*

We prepare the data in the same way as before and include survival. The variable *fate* needs to be reorganized into *surv* to convert “rip” to zeros (dead) and everything else to ones (alive). We also centered height (lgh).

```
## read the Hypericum data from file
```

```
orig_data <- read.table("hypericum_data_94_07.txt", header=T)
dt <- subset(orig_data, !is.na(ht_init) & !is.na(st_init) & rp_init > 0 & year<1997 )
yr <- unique(dt$year)
dt$lgh <- log(dt$ht_init)
dt$lfr <- log(dt$rp_init)
dt$stems <- dt$st_init
site <- unique(dt$bald)
table(dt$bald,dt$fire_year)
dt$TSF <- 1
dt$TSF[dt$fire_year <1987] <-2
dt$TSF[dt$fire_year <1973] <-3
dt$TSF <- factor(dt$TSF)
dt$fyear <- factor(dt$year)
dt$fbald <- factor(dt$bald)
dt$surv <-1
dt$surv[dt$fate =="rip"] <- 0
table(dt$surv,dt$fate)
I <- order(dt$lgh)
lgh <- sort(dt$lgh)
table(dt$bald,dt$TSF)
tsf <- unique(dt$TSF)
tsf <- sort(tsf)
TSF <-dt$TS
boxplot(dt$lgh~dt$TSF)
summary(lm(dt$lgh~factor(dt$TSF)))
pairs(subset(dt,select=c(ht_init, rp_init,stems)))
dt$stems[dt$stems>8] <- 8
dt$stems <- factor(dt$stems)
dt$lghc <- dt$lgh- mean(dt$lgh)
```

We call the libraries of two packages that we will need during the analysis

```
library(nlme)
library(bbmle)
library(lme4)
```

We check for co-linearity and find a significant association among time-since-fire and plant height (Figure 2). There is evidence that the average height is higher in long-unburned populations than in populations more recently burned and with intermediate time-since-fire, but there is enough variation in height to proceed with our analysis.

```
boxplot(dt$ht_init~dt$TSF)
summary(lm(dt$lgh~factor(dt$TSF)))
```

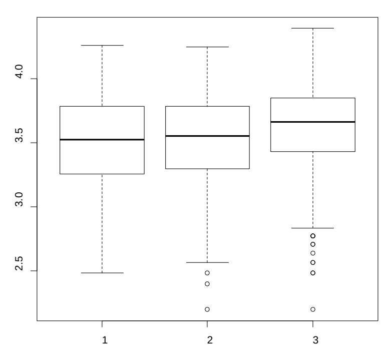

Figure 2. *Plot of height (cm) as a function of time-since-fire*

We follow Zuur et al. (2009) to evaluate the best configuration for the random factors. We evaluate a saturated model with all the fix factors. For our data this model includes all single factor, two way interactions and the three way interaction among height, stems and TSF. We propose three options for the random configuration: (i) no random effects, (ii) random intercept and (iii) random intercept and slope. We use the function *glmer* and specify the use of the binomial family. The *glmer* function requires the specification of a random term. In this occasion we were not able to identify the proper procedure to allow the comparison of the three models with REML. We use the procedure *glm* for the non-random model and, tentatively, compare the AICs of these models. This comparison indicates that the one with only random intercept is the more informative of the two mixed models. Some of these models have concerns due to not reaching convergence. We use the procedure *optimx* to address this issue.

```
require(optimx)
m1 <- glm(surv~lghc*TSF*stems,data=dt,family =binomial)
m2 <- glmer(surv~lghc*TSF*stems + (1|fbald) + (1|year),data=dt,family =binomial)
m2_nlminb <- update(m2,control=glmerControl(optimizer="optimx",
             optCtrl=list(method="nlminb")))
m3 <- glmer(surv~lghc*TSF*stems + (lghc|fbald)+ (lghc|year),data=dt,family =binomial)
m3_nlminb <- update(m3,control=glmerControl(optimizer="optimx",
             optCtrl=list(method="nlminb")))

AICtab(m1,m2_nlminb,m3_nlminb,weights=TRUE,base = TRUE))

          AIC    dAIC   df weight
m2_nlminb 2136.4    0.0 50 0.962
m3_nlminb 2142.9    6.5 54 0.038
m1        2230.6   94.2 48 <0.001
```

We proceed to evaluate the optimal fixed structure of the random structure that we just found. We fit models with the same random effects structure (Zuur et al. 2009). We compare them using AIC.

```
M11 <- glmer(surv~lghc*TSF*stems + (1|fbald) + (1|year),data=dt,family=binomial)
M11_nlminb <- update(M11,control=glmerControl(optimizer="optimx",
    optCtrl=list(method="nlminb")))
M13 <- glmer(surv~lghc+TSF*stems + (1|fbald) + (1|year),data=dt,family =binomial)
M13_nlminb <- update(M13,control=glmerControl(optimizer="optimx",
    optCtrl=list(method="nlminb")))
M14 <- glmer(surv~lghc+TSF+stems + (1|fbald) + (1|year),data=dt,family =binomial)
M15 <- glmer(surv~lghc+TSF + (1|fbald) + (1|year),data=dt,family =binomial)
M16 <- glmer(surv~lghc+stems + (1|fbald) + (1|year),data=dt,family =binomial)
M17 <- glmer(surv~lghc*stems + (1|fbald) + (1|year),data=dt,family =binomial)
M18 <- glmer(surv~lghc*stems + TSF + (1|fbald) + (1|year),data=dt,family =binomial)
M19 <- glmer(surv~lghc*TSF + stems + (1|fbald) + (1|year),data=dt,family =binomial)
M20 <- glmer(surv~lghc*TSF + stems*TSF + (1|fbald) + (1|year),data=dt,
       family =binomial)
M20_nlminb <- update(M20,control=glmerControl(optimizer="optimx",
       optCtrl=list(method="nlminb")))
```

There are three models providing significant information (M19, M13 and M20). We chose Model M20, including interactive effects of height and stems with TSF because integrates the information of the other two.

```
AICtab(M11_nlminb,M13_nlminb,M14,M15,M16,M17,M18,M19,M20_nlminb,
       weights=TRUE,base = TRUE)
```

```
           AIC    dAIC   df weight
M19        2105.3    0.0 15 0.395
M13_nlminb 2105.5    0.2 27 0.363
M20_nlminb 2106.8    1.5 29 0.185
M14        2109.9    4.6 13 0.040
M18        2111.7    6.4 20 0.016
M16        2117.3   12.0 11 <0.001
M17        2119.6   14.3 18 <0.001
M15        2131.8   26.5 6  <0.001
M11_nlminb 2136.4   31.1 50 <0.001
```

The plots of Model M20 are presented below (Figure 3). We conclude that increasing height and time-since-fire tend to decrease survival compared to recently burned populations. The effect of number of stems differentially affects survival depending of time-since-fire. There is considerable random variation by population and year.

```
> summary(M20_nlminb)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: surv ~ lghc * TSF + stems * TSF + (1 | fbald) + (1 | year)
   Data: dt
Control: glmerControl(optimizer = "optimx", optCtrl = list(method = "nlminb"))

     AIC      BIC   logLik deviance df.resid
  2106.8   2264.9  -1024.4   2048.8     1693

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7381 -0.9305  0.2943  0.8173  2.3938

Random effects:
 Groups Name        Variance Std.Dev.
 fbald  (Intercept) 0.4711   0.6864
 year   (Intercept) 0.1145   0.3383
Number of obs: 1722, groups:  fbald, 14; year, 3

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.03731    0.66237   4.586 4.53e-06 ***
lghc        -1.05429    0.49923  -2.112 0.034700 *
TSF2        -2.76853    0.73622  -3.760 0.000170 ***
TSF3        -2.46282    0.76558  -3.217 0.001296 **
stems2      -0.49470    0.58320  -0.848 0.396301
stems3      -0.83737    0.54408  -1.539 0.123789
stems4      -1.16465    0.56787  -2.051 0.040274 *
stems5      -0.86475    0.65805  -1.314 0.188809
stems6      -1.46080    0.66389  -2.200 0.027780 *
stems7      -1.79445    0.79376  -2.261 0.023777 *
stems8      -2.98296    0.74766  -3.990 6.62e-05 ***
lghc:TSF2    0.90464    0.56210   1.609 0.107531
lghc:TSF3    0.77116    0.57486   1.341 0.179760
TSF2:stems2  0.49836    0.65322   0.763 0.445505
TSF3:stems2  0.09307    0.67742   0.137 0.890722
TSF2:stems3  0.50196    0.62370   0.805 0.420928
TSF3:stems3  0.95242    0.64650   1.473 0.140698
TSF2:stems4  0.98900    0.65214   1.517 0.129381
TSF3:stems4  1.21460    0.67996   1.786 0.074052 .
TSF2:stems5  0.38886    0.76027   0.511 0.609022
TSF3:stems5  0.08770    0.76154   0.115 0.908314
```

```
TSF2:stems6  0.98569    0.82298   1.198 0.231028
TSF3:stems6  0.99886    0.80595   1.239 0.215212
TSF2:stems7  0.60744    1.00703   0.603 0.546373
TSF3:stems7  0.97142    0.93435   1.040 0.298488
TSF2:stems8  3.07252    0.90548   3.393 0.000691 ***
TSF3:stems8  1.49966    0.88469   1.695 0.090052 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

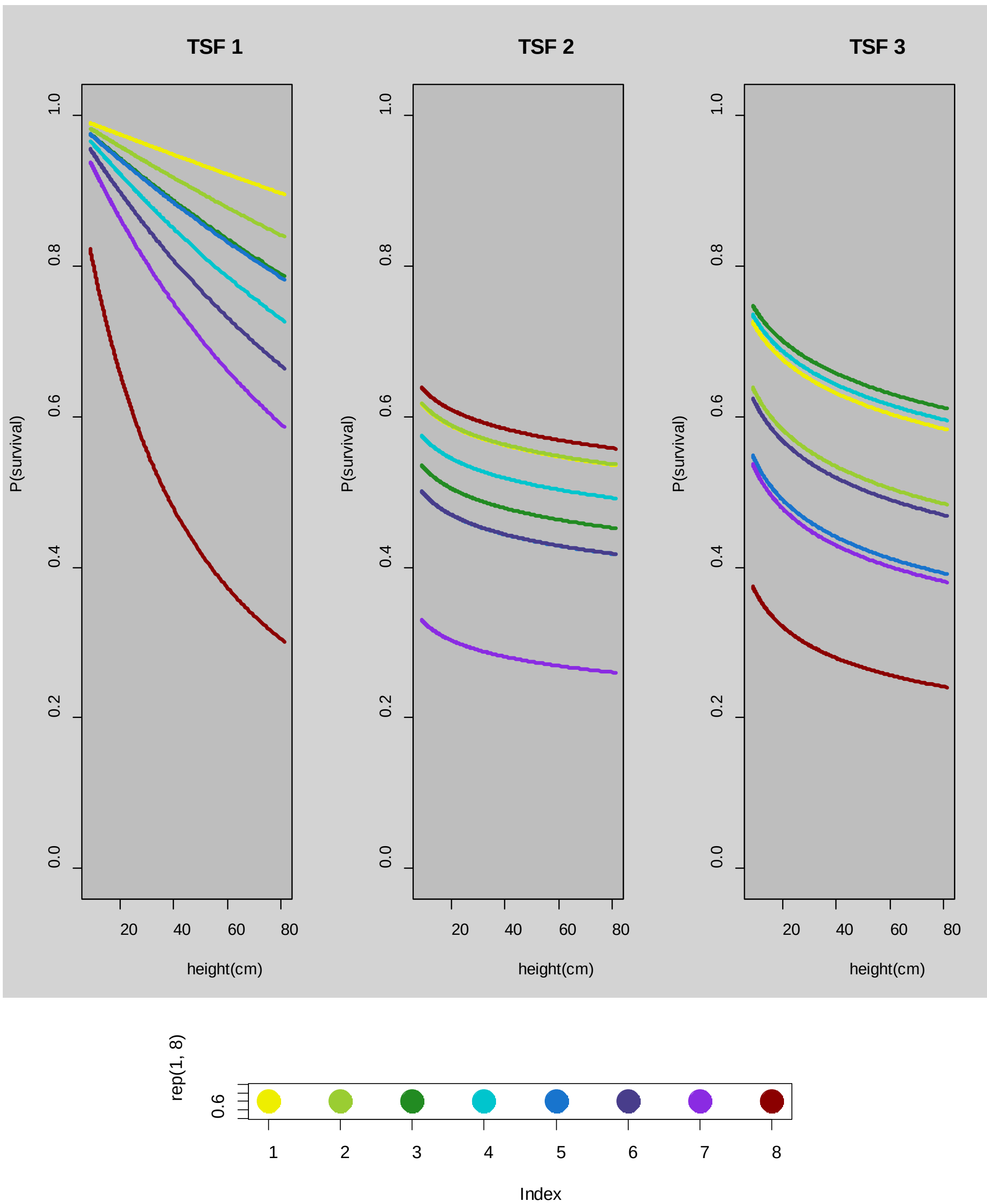

Figure 3. *Plot of survival as a function of height by plants with different number of stems and time-since-fire (x= height, y= survival, stems in different colors)*

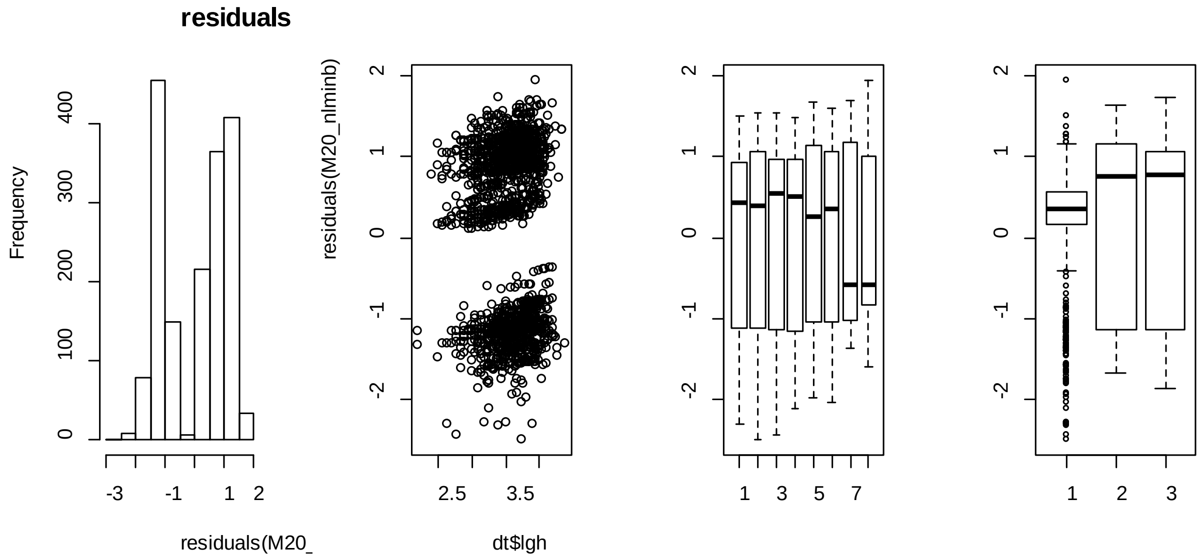

Figure 4. *Residuals of model M20*

Crawley, M. J. 2006. The R Book. Wiley.

Quintana-Ascencio, P. F., E. S. Menges, and C. Weekley. 2003. A fire-explicit population viability analysis of *Hypericum cumulicola* in Florida rosemary scrub. Conservation Biology, 17: 433-449.

Zuur, A.F., E.N. Ieno, N.J. Walker, A. Savaliev, G.M. Smith. 2009. Mixed effects models and extensions in Ecology with R. Springer.