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

In a previous demo we provided evidence that number of reproductive structures of *Hypericum cumulicola* is clearly negatively associated with plant survival. Now, we will evaluate the relevance of four variables to explain survival variation in this species (Quintana-Ascencio et al. 2003, 2018 and 2019). We will evaluate the effect of plant height (cm), number of stems, number of reproductive stems and time-since-fire on *Hypericum cumulicola* survival. This species benefits of recurrent fire that reduces competition with dominant shrubs, releases nutrients and increases water availability. We use a model selection approach to assess their relative importance. We will also estimate the random variation on survival by population and year.

Figure 1. *Dying Hypericum cumulicola*

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

```
library(rethinking)
library(rstan)
library(bbmle)
```

We prepare the data. The variable *fate* needs to be reorganized into *surv* to convert "rip" to zeros (dead) and everything else to ones (alive). We identify three time-since-fire categories (early, intermediate and late). Because these are allometric relationships we use the logarithms of height (*lgh*) and number of reproductive structures (*lfr*). We centered *lgh* to facilitate its interpretation. We identify few plants with more than 8 stems and group them together under 8 stems.

```
dt <- subset(orig_data, !is.na(ht_init) & !is.na(st_init) & rp_init > 0 & year<1997)
yr <- unique(dt$year) #studied years
dt$lgh <- log(dt$ht_init) #log of height
dt$lfr <- log(dt$rp_init) # log of reproductive structures
dt$stems <- dt$st_init # stems
site <- unique(dt$bald) # studied sites
#Survival
dt$surv <-1
dt$surv[dt$fate =="rip"] <- 0
table(dt$surv,dt$fate)
## Time-since-fire
table(dt$bald,dt$fire_year)
dt$TSF <- 1
dt$TSF[dt$fire_year <1987] <-2
dt$TSF[dt$fire_year <1973] <-3
table(dt$bald,dt$TSF)
tsf <- unique(dt$TSF)

dt$lghc <- dt$lgh- mean(dt$lgh) # centered for interpretation
dt$stems[dt$stems>8] <- 8
```

Checking for co-linearity, we find as before a clear association among number of reproductive structures and plant height (Figure 2). We decide not to include number of reproductive structures in this analysis.

```
par(mfrow=c(1,2))
plot(dt$lgh ~dt$fire_year)
summary(lm(dt$lgh~factor(dt$TSF)))
plot(dt$lfr ~dt$fire_year)
summary(lm(dt$lfr ~factor(dt$TSF)))

par(mfrow=c (1,1))
pairs(subset(dt,select=c(lghc, lfr, stems)))
```

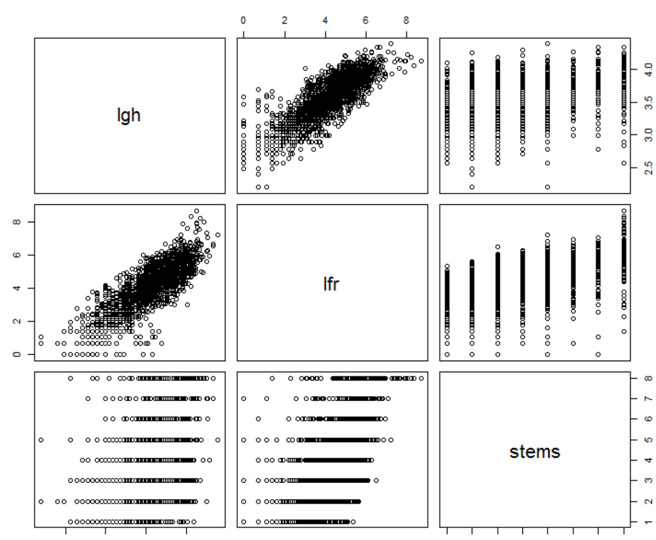


Figure 2. *Plot of height (cm) as a function of time-since-fire and correlation plots among log(height), log(number of reproductive structures) and number of stems*

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 intercepts 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.

```
dt$fyear <- factor(dt$year)
dt$fbald <- factor(dt$bald)
dt$fstems <- factor(dt$stems)
dt$fTSF <- factor(dt$TSF)
require(optimx)
m1 <- glm(surv~lghc*fTSF*fstems,data=dt,family =binomial)
m2 <- glmer(surv~lghc*fTSF*fstems + (1|fbald) + (1|fyear),data=dt,family =binomial)
m2_nlminb <-
update(m2,control=glmerControl(optimizer="optimx",optCtrl=list(method="nlminb")))
m3 <- glmer(surv~lghc*fTSF*fstems + (lghc|fbald)+ (lghc|fyear),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*fTSF*fstems + (1|fbald) + (1|fyear),data=dt,family=binomial)
M11_nlminb <-
update(M11,control=glmerControl(optimizer="optimx",optCtrl=list(method="nlminb")))
M13 <- glmer(surv~lghc+fTSF*fstems + (1|fbald) + (1|fyear),data=dt,family =binomial)
M13_nlminb <-
update(M13,control=glmerControl(optimizer="optimx",optCtrl=list(method="nlminb")))
M14 <- glmer(surv~lghc+fTSF+fstems + (1|fbald) + (1|fyear),data=dt,family =binomial)
M15 <- glmer(surv~lghc+fTSF + (1|fbald) + (1|fyear),data=dt,family =binomial)
M16 <- glmer(surv~lghc+fstems + (1|fbald) + (1|fyear),data=dt,family =binomial)
M17 <- glmer(surv~lghc*fstems + (1|fbald) + (1|fyear),data=dt,family =binomial)
M18 <- glmer(surv~lghc*fstems + TSF + (1|fbald) + (1|fyear),data=dt,family =binomial)
M19 <- glmer(surv~lghc*fTSF + fstems + (1|fbald) + (1|fyear),data=dt,family =binomial)
M20 <- glmer(surv~lghc*fTSF + fstems*fTSF + (1|fbald) + (1|fyear),data=dt,family
=binomial)
M20_nlminb <-
update(M20,control=glmerControl(optimizer="optimx",optCtrl=list(method="nlminb")))

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

There are three models providing 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. It also is a compromise for the estimates for the coefficients and their standard errors

```
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
```

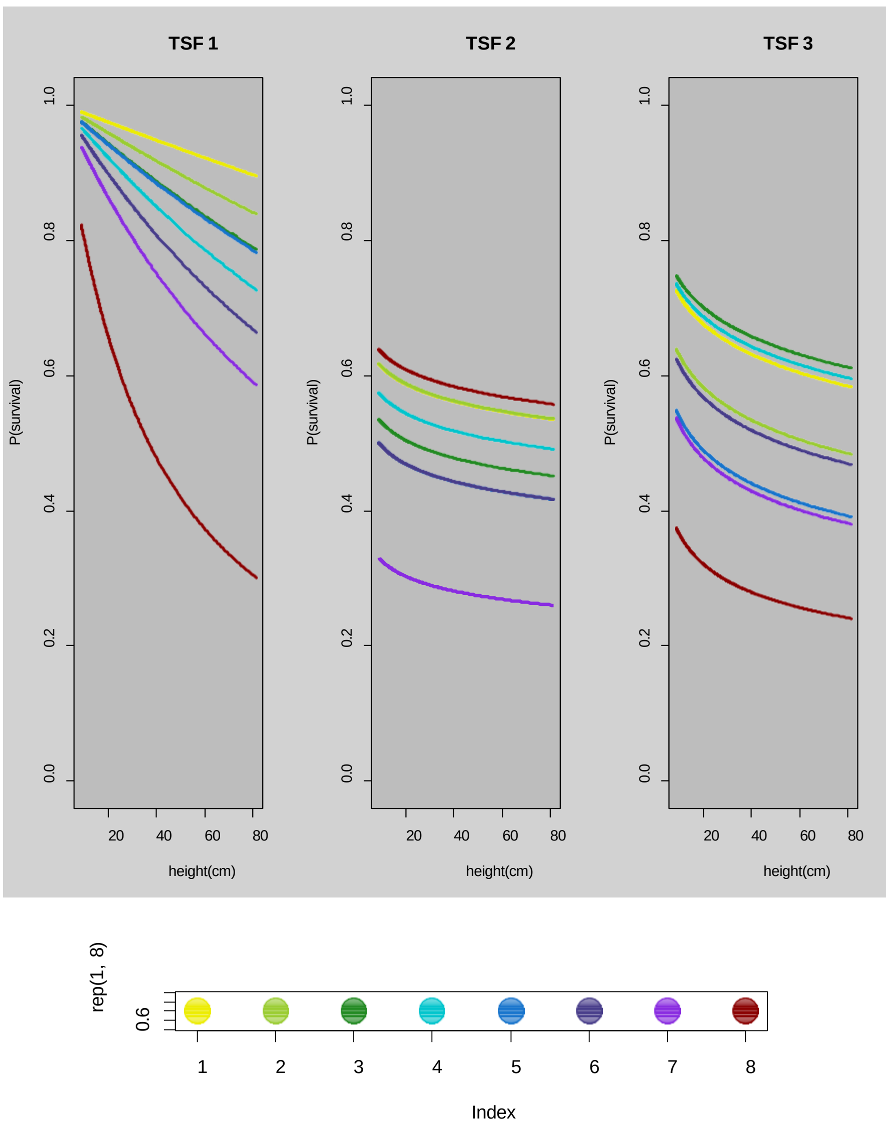


Figure 3. *Plot of survival as a function of height by plants with different number of stems and time-since-fire (x= height, y= survival, stems in different colors)*

Bayesian approach

```
m19B <- ulam(
alist(
surv ~ dbinom(1,p),
logit(p) <- a  + pop[pj] +  y[yi] + b*lghc +
            d*tsf2 + e*tsf3 + f*lghc*tsf2 + g*lghc*tsf3 +
            c2*s2 + c3*s3 + c4*s4 + c5*s5 + c6*s6 + c7*s7 +
            c8*s8,
a ~ dnorm(0,100),
b ~ dnorm(0,10),
c2 ~ dnorm(0,10),
c3 ~ dnorm(0,10),
c4 ~ dnorm(0,10),
c5 ~ dnorm(0,10),
c6 ~ dnorm(0,10),
c7 ~ dnorm(0,10),
c8 ~ dnorm(0,10),
d ~ dnorm(0,10),
e ~ dnorm(0,10),
f ~ dnorm(0,10),
g ~ dnorm(0,10),
pop[pj] ~ dnorm(0,sigmap),
y[yi] ~ dnorm(0,sigmabp),
sigmap ~ dcauchy(0,10),
sigmabp ~ dcauchy(0,1)
),
data = dt,chains =3,
iter=4000,warmup=1000
 )

precis(m19B,digits=2)
```

	Mean	StdDev	lower 0.89	upper 0.89	n_eff	Rhat
a	2.38	0.80	1.20	3.43	963	1
b	-1.62	0.46	-2.37	-0.90	4833	1
c2	-0.24	0.21	-0.58	0.07	4641	1
c3	-0.21	0.21	-0.54	0.13	4189	1
c4	-0.22	0.22	-0.56	0.14	4573	1
c5	-0.66	0.25	-1.06	-0.27	4643	1
c6	-0.60	0.29	-1.07	-0.14	4789	1
c7	-1.02	0.34	-1.58	-0.50	5795	1
c8	-1.32	0.31	-1.80	-0.82	5649	1
d	-1.97	0.64	-3.02	-1.01	3687	1
e	-1.70	0.65	-2.69	-0.66	3535	1
f	1.40	0.52	0.55	2.21	4996	1
g	1.40	0.52	0.56	2.22	5465	1
sigmap	0.83	0.23	0.49	1.14	3822	1
sigmabp	0.68	0.66	0.16	1.19	665	1

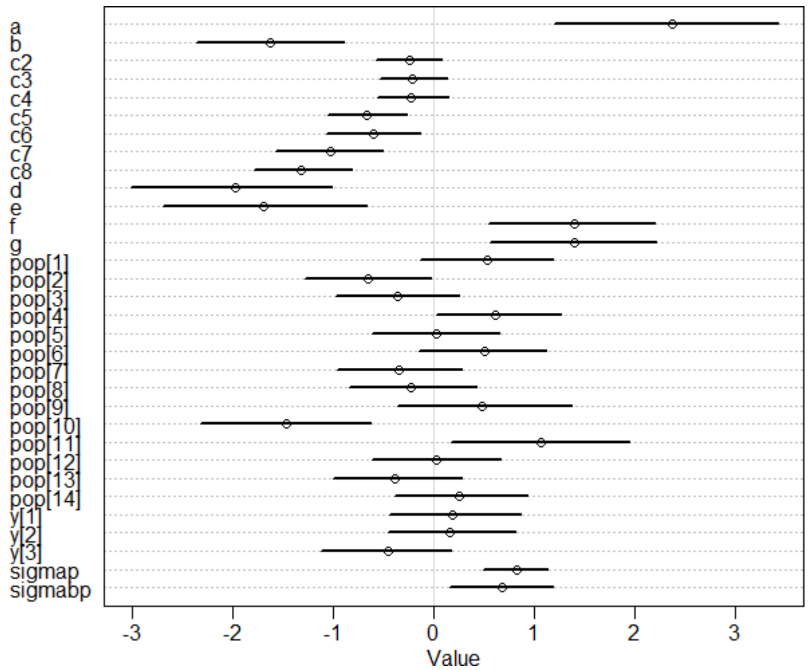
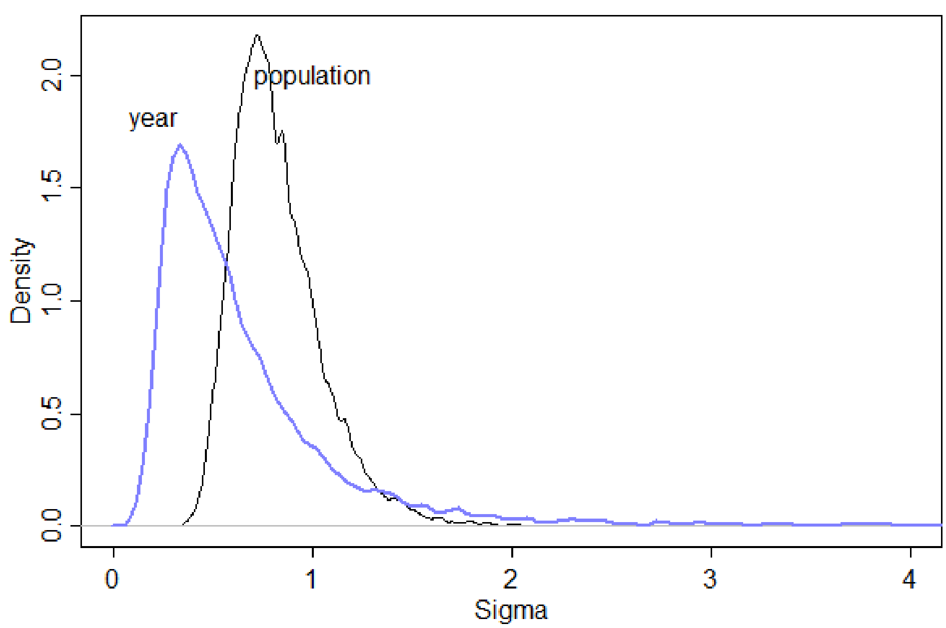


Figure 5. *Posterior means and 89% intervals for model M19B. Notice the large variation for long time-since-fire*



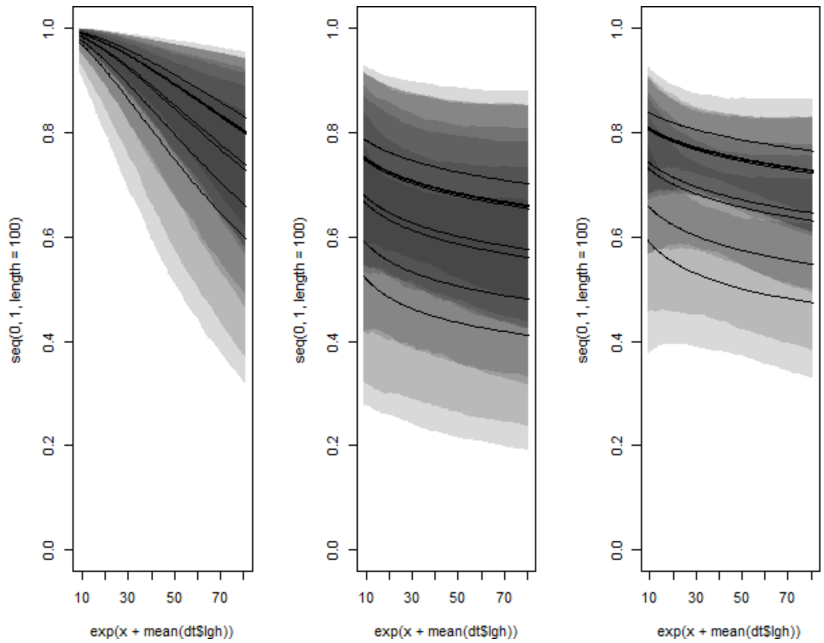
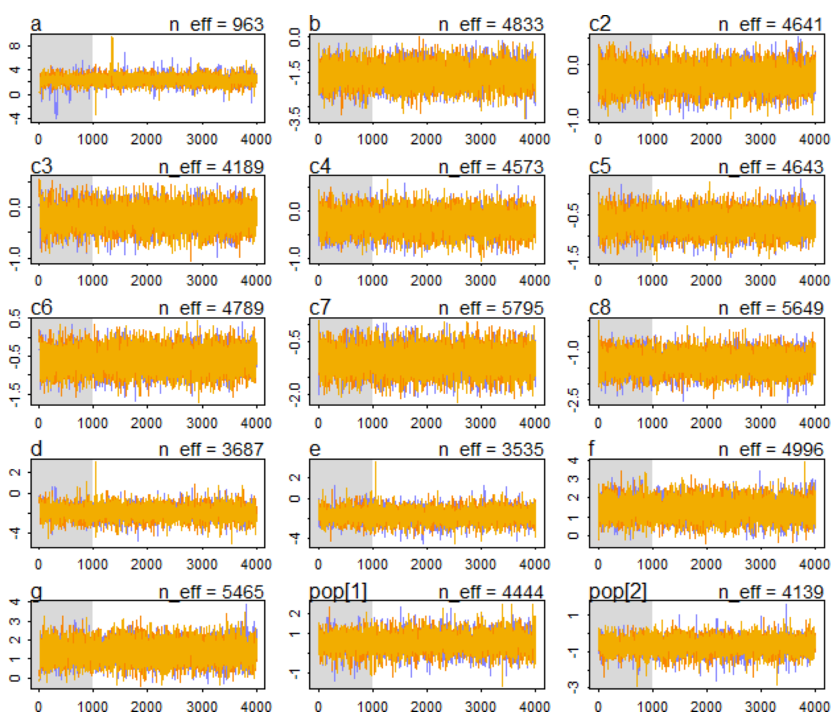
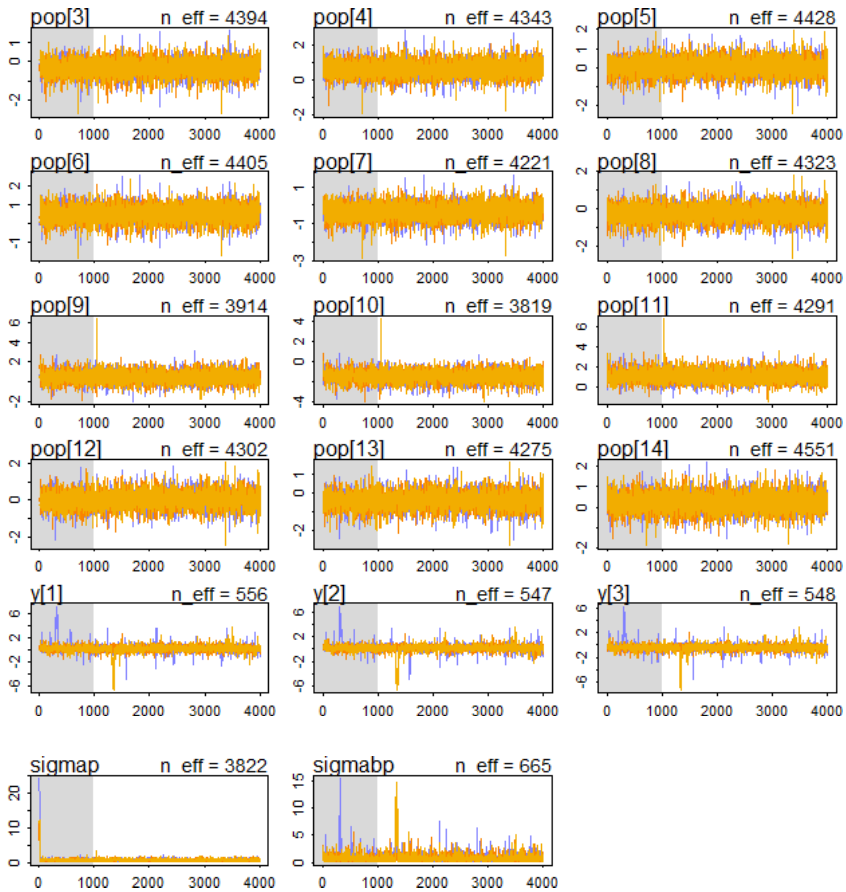
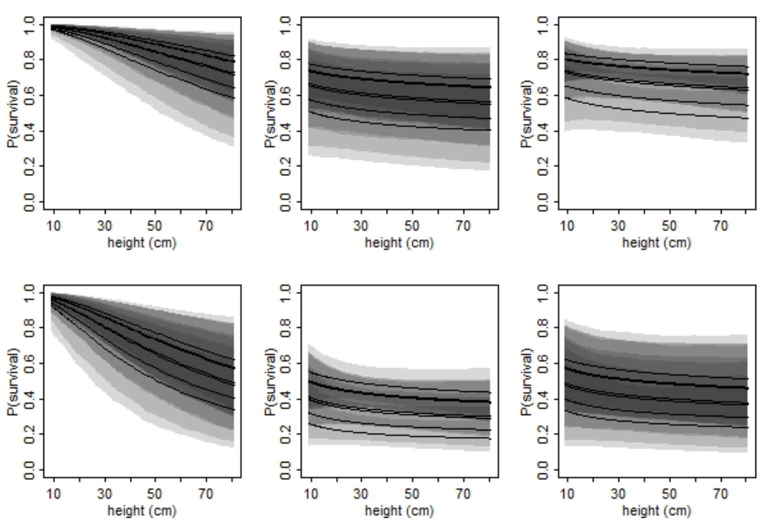


Figure 6. *Plot of survival as a function of height by plants with different number of stems and time-since-fire (x= height, y= survival, stems in different greys)*

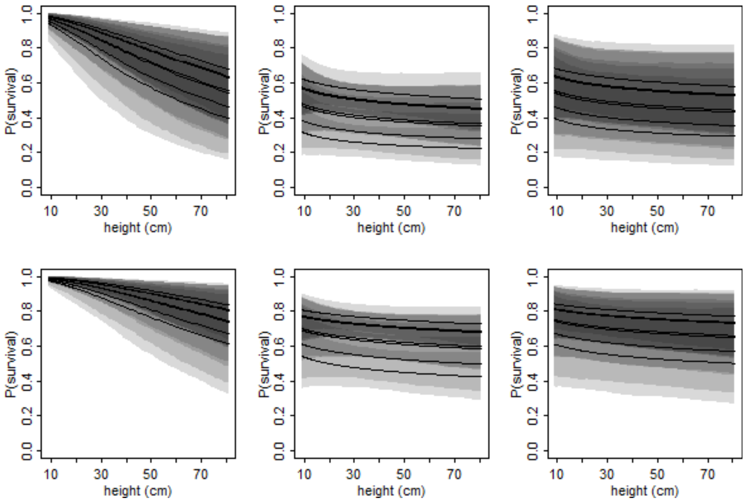




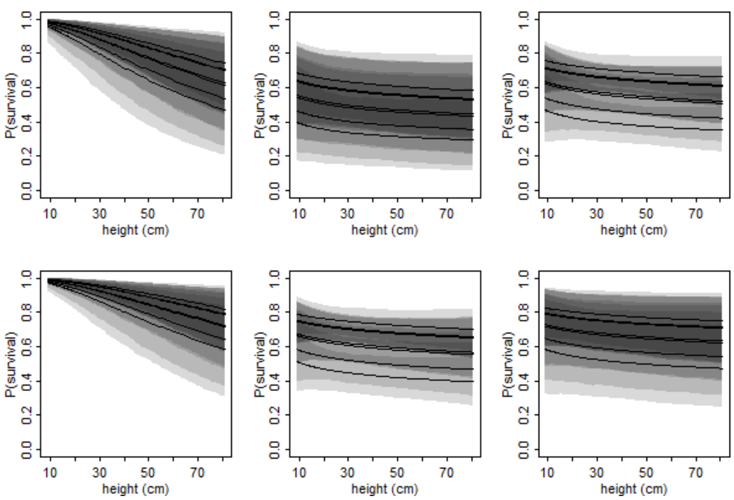
Populations 1 and 29



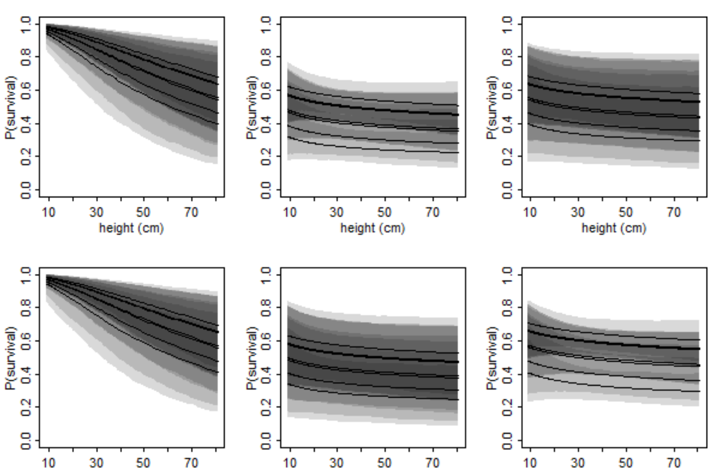
Populations 32 and 42



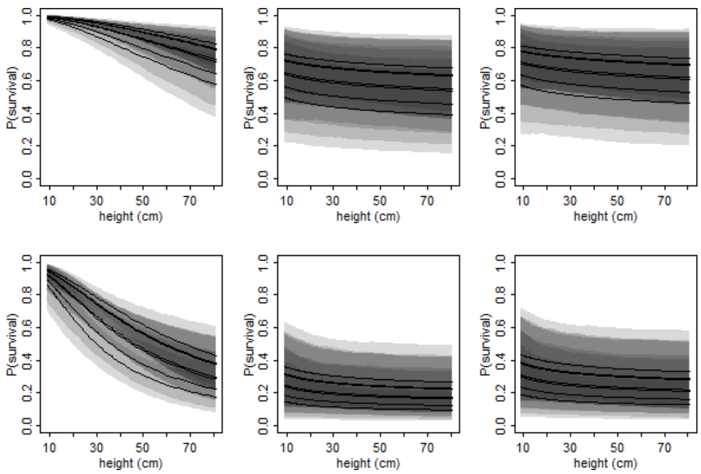
Population 50 and 57



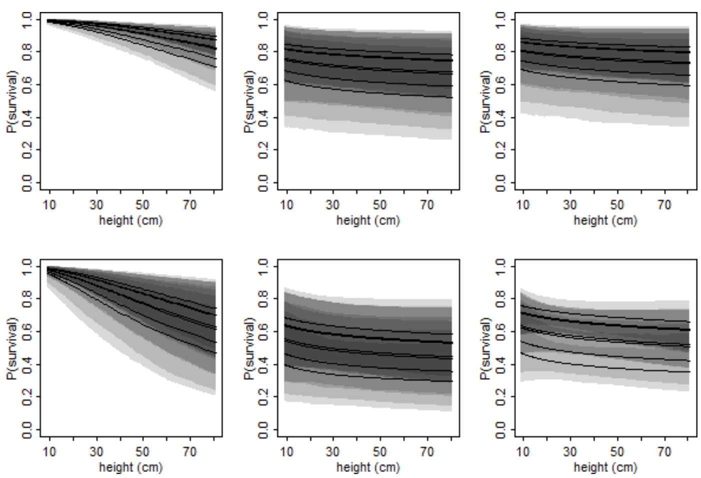
Populations 59 and 62



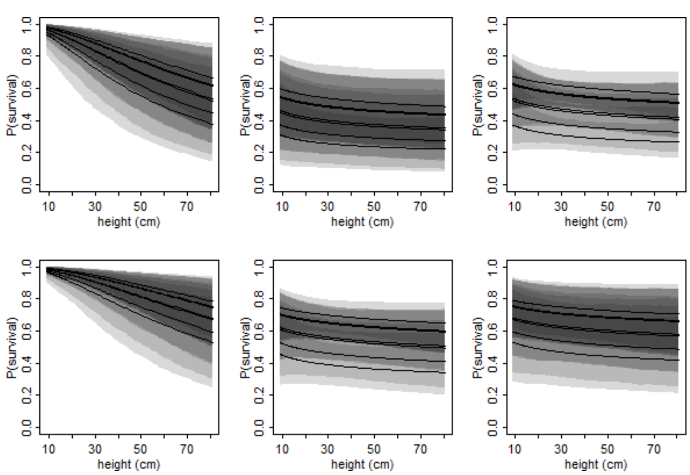
Populations 67 and 87



Populations 88 and 91



Populations 93 and 103



NOTE: all the materials for this demo can be found at:

https://sciences.ucf.edu/biology/d4lab/methods-2/

References

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

McElreath, R.M. 2016. Statistical Rethinking: a Bayesian course with examples in R and Stan. Chapman and Hall.

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.

Quintana-Ascencio, P.F. Koontz, S., Smith, V., David, A., Sclater, V. L. & E. S Menges. 2018. Predicting landscape-level distribution and abundance: Integrating demography, fire, elevation, and landscape habitat configuration. Journal of Ecology, 106: 2395-2408

Quintana-Ascencio, P.F. Koontz, S.M., Ochocki, B., Sclater, V. L., López-Borghesi, F., Li, H. & E. S Menges. 2019. Assessing the roles of seed bank, seed dispersal and historical disturbances for metapopulation persistence of a pyrogenic herb. Journal of Ecology, 107: 2760-2771.

Zuur, A, J.M. Hilbe and E N. Leno. 2015. A beginner's guide to GLM and GLMM with R. Highland Statistics, Ltd.