MORTALITY VARIATION THROUGH TIME

R. CONDIT

DESCRIPTION: The question addressed here is how tree mortality has changed with time, in particular how species differ in the changes. There are two alternative models described:

- Mortality rate changing linearly with time
- Mortality rate varying from census to census, but not following a consistent change

Unlike growth analysis, mortality rates cannot be addressed with *lmer*, due to the variation in time interval among trees and among censuses. A proper mortality analysis must take this into account using a different death probability for every individual at every census. This can only be done with a Bayesian hierarchical model. The CTFSRPackage has a function *lmerBayes* which accomplishes the analysis. The Bayesian model has the advantage of producing complete confidence limits easily; the disadvantage is that it runs slowly. Analyses here on big datasets will take a couple hours to complete. END_DESCRIPTION

1. Formatting data

The calculations start with the full plot dataframes described at `http://ctfs.arnarb.harvard.edu/Public/CTFSRPackage/help/RoutputFull.pdf`. The BCI data can be downloaded from `http://ctfs.arnarb.harvard.edu/bci/RAnalyticalTables` after filling out the request form `http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci/`.

In the CTFSRPackage, there are functions for loading data and calculating mortality rates. The function *individual_mort.table* in the *demogchange* topic collects results from several other functions to calculate mortality rates for individual trees and assemble them into a table in the format used by *lmerBayes* (which is the same format used by *lmer*. It is described in the Demography package.

The purpose is to create tables of mortality rates for every tree in every census interval. Choose any diameter category, and any set of censuses you wish to work with. Here is an example for big trees at BCI (dbh ≥ 400), with all seven censuses included.

```
> mrate=individual_mort.table(cnsdata=list(bci.full1,bci.full2,bci.full3,bci.full4,bci.full5,bc
> head(mrate)
> table(mrate$census)
```

The table *mrate* has one row per individual, and includes the tree's tag number, coordinates, species, and dbh at the start of a census interval. The column *census* gives the census interval, so 1 means census 1 to 2, and 2 means census 2 to 3, etc. The column *censusfact* is the same, but as a factor variable instead of an integer. The *time* column is the mid-point of the interval between the two censuses, centered on 1992. This is the format needed by *lmer* and *lmerBayes*. The reason for centering time is that it is much better in linear regressions to have the x-axis near zero.

The column holding survival information is *fate*, TRUE for trees that survived the census interval; FALSE for the deaths.

2. Mean forest-wide mortality

First find the mean mortality rate of all individuals in each census period using the table *mrate*. This requires counting living and dead trees in each census, and there is a function *calcMortIndivTable* that does it, found with the DemogChange topic of the CTFSRPackage.

_____
Date: May 22, 2014.

The same program also calculates the observed mortality rate for each species in each interval, saved below as *spmort*. Calls to *tapply* are also used to create a matrix of sample sizes: individuals per species at the start of each census interval, and number of survivors over each interval.

The time, meanyr, is the mean years since 1992 of each census interval.

```
> meanyr=tapply(mrate$time,mrate$census,mean)
> mort=calcMortIndivTable(mrate,by='all')
> meanMort=data.frame(meanyr,mort)
> spmort=calcMortIndivTable(mrate,by='species')
> N=tapply(mrate$fate,list(mrate$sp,mrate$census),length)
> S=tapply(mrate$fate,list(mrate$sp,mrate$census),sum,na.rm=TRUE)
```

Graph median mortality vs. time.

```
> par(cex.lab=1.5,cex.axis=1.5)
> plot(mort~meanyr,data=meanMort,type='l',xlab='years since 1992',ylab='median mortality',xlim=
```
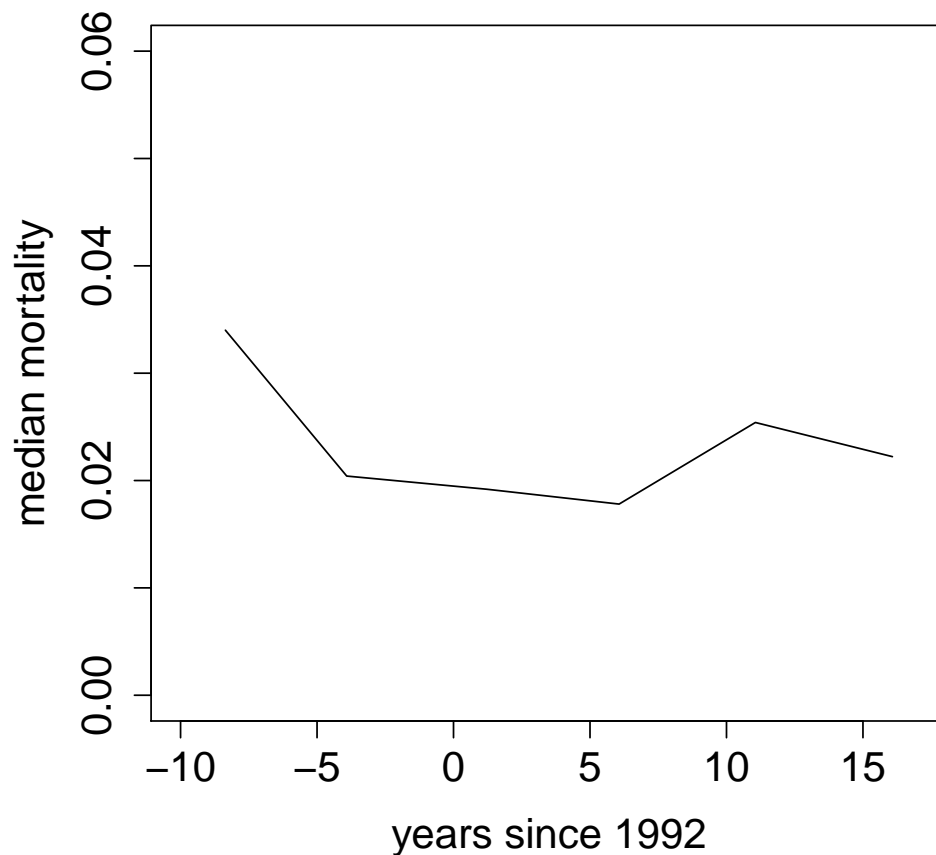

3. Mixed model to assess species variation in linear changes through time

To test for changes in mortality rate, I have a model where the *log(survival)* is linearly related to time, that is *log(survival)* ≈ *a+b*time*. The two parameters, *a* and *b*, the intercept and slope, follow a Gaussian hyperdistribution across species. The function to produce predicted survival given this model is called *lmerMortLinear* and is found with the CTFSRPackage topic on demographic change.

The function *lmerBayes* is a Bayesian version of *lmer* found with the CTFSRPackage. It can fit mortality rate as a function of time, with species as a random effect; equivalent to a hierarchical model, where a Gaussian distribution describes the variation in species responses. The Bayesian version does not require a normal error, and it can account for differences in the census interval between trees and censuses.

Based on the shift in mortality that happened after the first 3 census intervals, I follow the approach of dividing the data into two periods, with censuses 1-3 and then 3-6. This approach should not necessarily be followed in other forests. Unlike growth, there is no *tag* effect; it's not clear exactly whether there are repeated measures, since trees can only die once.

The calls to *lmerBayes* are complex, with many arguments. The data are defined in the first argument, exactly as in *lmer*. The arguments *xcol, ycol, and randcol* designate the predictors (*xcol*), the response *ycol*), and the mixed effect (*randcol*. They are quoted because they are character variables designating the names of the columns holding the required data. That's how to write a formula in *lmerBayes*: in *lmer*, it would be written *ycol   xcol + (xcol|randcol)*.

Note that there are two *xcol*s because the time interval is needed to calculate annual mortality.

Subsequent arguments are controls for *lmerBayes*:

- *start* gives starting parameters for the fixed effect; they must be valid, and it is helpful if they are reasonably close to good answers
  the vector of starting parameters should give each a name
- *model* names the function used to calculate the response from the predictors
- *startSD* is not used in a binomial (ie, mortality) model, but something must be included
- *startCov* are starting estimates for the variance among species; it is helpful if they are reasonably close to a good result
- *includeCovar=FALSE* means estimate only the variances among species' parameters, not their covariances
- *badSDparam* are not used in this model, but something needs to be included
- *steps* is how long to run the parameter search;
  - the example below has 2500 steps, which will produce good results
  - each took about 20 minutes on a good computer
  - one run will take an hour for saplings, where the sample is much larger
- *burn* is the number of steps to discard as burn-in; for runs of 4000 or more, it should be 1000
- *showstep* is how frequently to print intermediate search parameters to the screen

```
[1] 4

[1] 3

> mod1=lmerBayes(data=subset(mrate,census<=3),ycol='fate',randcol='sp',xcol=c('time','interval'
> mod2=lmerBayes(data=subset(mrate,census>=3),ycol='fate',randcol='sp',xcol=c('time','interval'
```

The coefficients of the models are the key result. The fixed effect gives the forest-wide response, which is the average mortality rate of all the species. It is thus not the same as the mean forest-wide mortality, because the fixed effect weights each species equally. The fixed effect is called *bestmu*, the error terms are in *bestsigma*, and species parameters are in *best*.

The following graph uses these to get the parameters, then the function *curve* to add them to a graph. The points are forest-wide means, as graphed above.

3.1. Fixed effects.

```
> par(cex.lab=1.5,cex.axis=1.5)
> plot(mort~meanyr,data=meanMort,pch=16,xlab='years since 1992',ylab='median mortality',xlim=c(
> par=mod1$bestmu; curve(exp(par[1]+x*par[2]),from=(-10),to=0,add=TRUE)
> par=mod2$bestmu; curve(exp(par[1]+x*par[2]),from=0,to=17,add=TRUE)
```
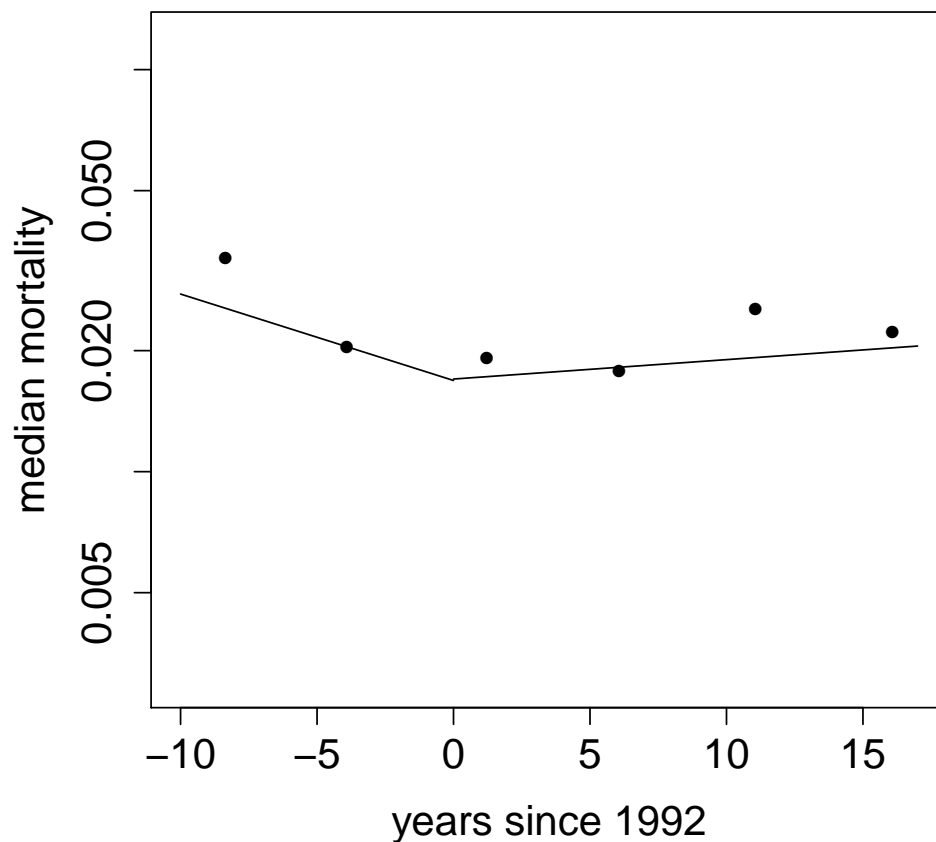
3.2. **Statistical signficance of the fixed effects.** The Bayesian models return long chains of parameter values from which confidence limits can be calculated. For the fixed effects, they are in a table *mu*. A vector *keep* is used to exclude the burn-in steps. The CTFSRPackage function *CI* calculates 95% credible intervals.

The credible intervals show that In the first time period, census intervals 1-3, mortality declined significantly through time, since the intervals do not include zero. Over intervals 3-6, though, there was no significant change.

```
> head(mod1$mu)
          inter         slope
[1,] -3.500000  0.010000000
[2,] -3.736942 -0.012429449
[3,] -3.682008  0.006553816
[4,] -3.665217  0.006338040
[5,] -3.651383  0.005737462
[6,] -3.664225  0.004720303

> head(mod1$mu[mod1$keep,])
          inter        slope
[1,] -4.092651 -0.05487119
[2,] -4.221568 -0.05621155
```

```
[3,] -4.069272 -0.05531380
[4,] -4.071846 -0.05621243
[5,] -4.042103 -0.05619094
[6,] -3.982200 -0.05970534
```

```
> apply(mod1$mu[mod1$keep,],2,CI)
          inter        slope
2.5%  -4.335693 -0.06967726
97.5% -3.866445 -0.02302401
```

```
> apply(mod2$mu[mod2$keep,],2,CI)
          inter        slope
2.5%  -4.319027 -0.008982551
97.5% -3.844331  0.026474374
```

3.3. **Individual species variation.** The error term from the Bayesian model is found in a matrix of covariance terms, *bestsigma*. In these models, only standard deviations were fitted, not the full covariance, so the off-diagonal elements are zero. Since the diagonal gives the variance, the square root is needed to show standard deviations.

Consider first the intercept in the first model. The fixed effect, or average intercept of all species, is −4.08. But species differ, with standard deviation 0.92. Recall that most results fall within 2 standard deviations of the mean, and that these are logarithms of mortality. These can be interpreted then to mean that individual species have mortality intercepts as high as 0.106 (10.6% y^{-1}), or $exp(-4.08 + 2 \times 0.92)$, and others as low as 0.0027 (0.27% y^{-1}), which is $exp(-4.08 - 2 \times 0.92)$.

But the main point has been how mortality rate varies through time, given as the slope parameters. The species standard deviation for slope is 0.046, and the fixed effect −0.049. Considering again 2 standard deviations from the mean, the entire distribution of species responses overlaps 0, that is, though most species had declining mortality, quite a few showed increases.

```
> mod1$bestsigma
          inter        slope
inter 0.8412129 0.000000000
slope 0.0000000 0.002096569
```

```
> diag(sqrt(mod1$bestsigma))
     inter      slope
0.91717662 0.04578831
```

```
> diag(sqrt(mod2$bestsigma))
     inter      slope
0.79153501 0.03554838
```

```
> exp(-4.08 + 2*0.92)
[1] 0.1064585
```

```
> exp(-4.08 - 2*0.92)
[1] 0.0026852
```

3.4. **Individual species responses.** Like *lmer*, *lmerBayes* also returns a set of coefficients for every species, and these make for direct interpretation. The coefficients are in a table *best*. Here is graph overlaying every individual species onto the graph of fixed effects, using the models with tag included. First I save the coefficients into a temporary variable called *cf* to reduce the amount of typing. There is a loop through every species, using the function curve to overlay one species' results with a gray line. Since I chose to run models for two separate intervals of time, I repeat this

for mod1 then mod2. At the end, I add the fixed effects again, just as above, plus points for the forest-wide means.

Visually it is evident first that most of the variation between species is in the intercept term, while slopes of the lines are fairly consistent.

```
> par(cex.lab=1.5,cex.axis=1.5)
> plot(mort~meanyr,data=meanMort,pch=16,xlab='years since 1992',ylab='median mortality',xlim=c(
> cf=mod1$best
> nospp=dim(cf)[1]
> for(j in 1:nospp) { par=drp(cf[j,]); curve(exp(par[1]+x*par[2]),from=(-10),to=0,add=TRUE,col=
> par=mod1$bestmu; curve(exp(par[1]+x*par[2]),from=(-10),to=0,add=TRUE)
> cf=mod2$best
> nospp=dim(cf)[1]
> for(j in 1:nospp) { par=drp(cf[j,]); curve(exp(par[1]+x*par[2]),from=0,to=17,add=TRUE,col='li
> par=mod2$bestmu; curve(exp(par[1]+x*par[2]),from=0,to=17,add=TRUE)
> points(mort~meanyr,data=meanMort,pch=16)
```
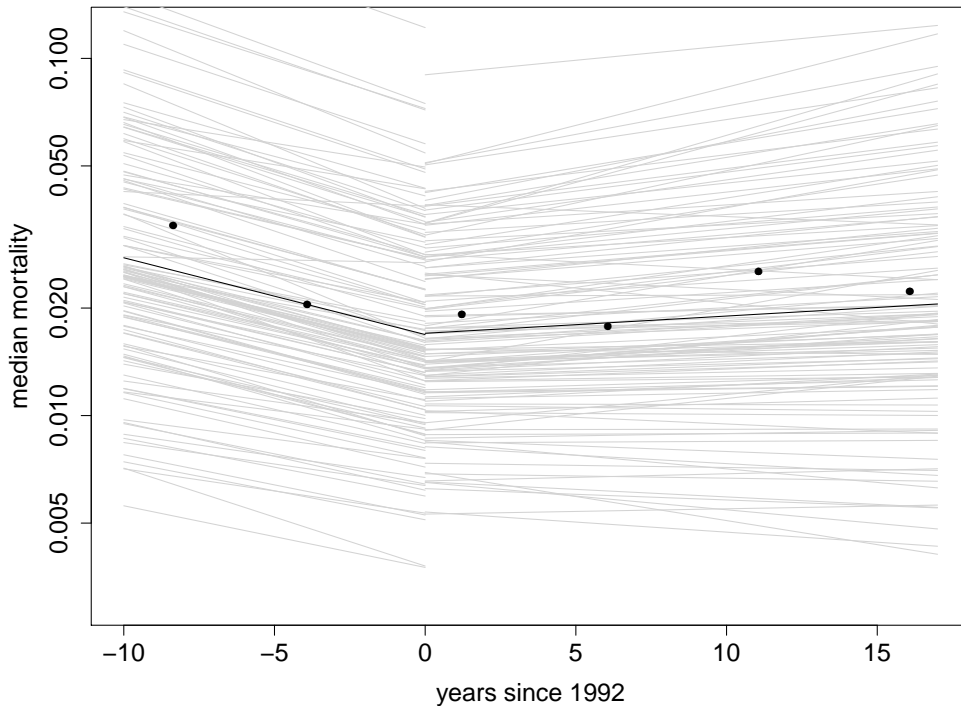

Coefficients can also be evaluated for individual species. Here are a few examples:

- Make a histogram of all mortality responses to time
- Find the species with the biggest decrease in mortality in the first interval
- Find the species with the highest mortality intercept (mortality in the very first census interval)

Other ideas to consider:

- Do species with higher initial mortality show bigger changes in mortality?
- Are species with lower mortality rates more stable (less change through time)?
- Do saplings respond similarly to big trees (ie, repeat the model for dbh range 10-50 mm, or other sizes)?

```
> par(cex.lab=1.5,cex.axis=1.5)
> cf=mod1$best
> head(cf)
```

```
          inter       slope
ade1tr -4.249996 -0.04399622
alchco -4.152601 -0.04459107
alsebl -4.878263 -0.04209767
anacex -5.584065 -0.03973522
apeihy -3.566582 -0.05328812
apeime -4.171684 -0.04548145
```

```
> hist(cf[,2],breaks=25)
> mineffect=min(cf$slope)
> subset(cf,slope<(mineffect+.01))
```

```
          inter       slope
beilpe -4.106735 -0.08015778
casear -3.269062 -0.08024912
guatdu -3.568608 -0.07294726
ocotwh -4.382122 -0.07028615
poular -3.453054 -0.07371486
sympgl -2.912138 -0.07884282
tremmi -2.593386 -0.07088900
virose -3.325908 -0.07089880
```

```
> maxmort=max(cf[,1])
> subset(cf,cf[,1]>maxmort-1)
```

```
          inter       slope
apeiti -3.037261 -0.06435901
cassel -2.634807 -0.06658077
ingama -2.103921 -0.06844740
loncla -3.002722 -0.03059099
micoar -3.015467 -0.06385746
ocotob -2.852785 -0.06411005
pterro -2.625599 -0.06234151
sympgl -2.912138 -0.07884282
tremmi -2.593386 -0.07088900
```

Histogram of cf[, 2]
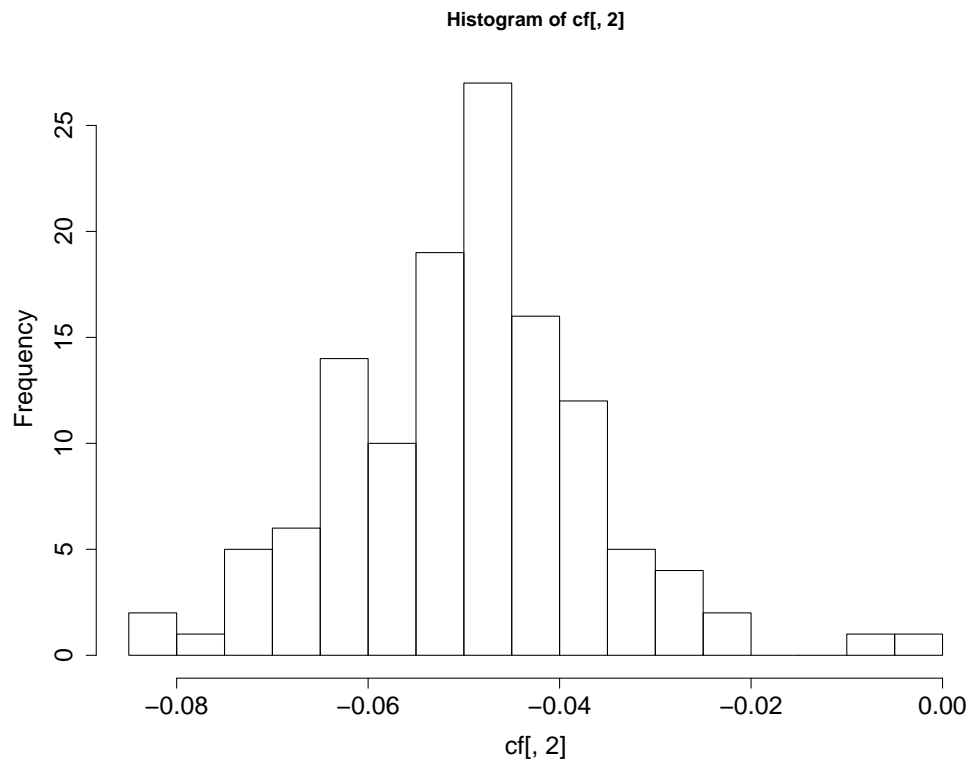
4. MIXED MODEL TO ASSESS SPECIES VARIATION AND TEMPORAL VARIATION

An alternative model of mortality change is to consider each census interval as a separate effect on mortality, but without a linear change of mortality through time. This would be reasonable if there is no a priori reason to expect steady increase (or decrease) in mortality rate with time. Mortality is allowed to vary from one census interval to the next, but can increase then decrease, etc.

The same function, lmerBayes can be used, but now time is not a numeric variable, but simply a factor. The table already created, mrate, has a column called censusfact, simply a number designating census intervals, but a factor, not an integer. Check this using R's function str.

```
[1] 2

> str(mrate)

'data.frame':         14352 obs. of  11 variables:
 $ treeID    : int   18 22 25 30 32 34 37 39 44 46 ...
 $ tag       : chr   "000001" "000005" "000008" "000013" ...
 $ sp        : chr   "quaras" "simaam" "brosal" "quaras" ...
 $ gx        : num   989 999 982 983 996 ...
 $ gy        : num   484 492 474 453 447 ...
 $ dbh       : num   1470 438 1575 570 546 ...
 $ interval  : num   4.12 4.12 4.11 4.12 4.12 ...
 $ time      : num   -8.3 -8.3 -8.3 -8.3 -8.3 ...
 $ census    : int   1 1 1 1 1 1 1 1 1 1 ...
 $ censusfact: Factor w/ 6 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ fate      : logi  TRUE FALSE TRUE TRUE TRUE TRUE ...
```

```
> modTimefact=lmerBayes(data=mrate,ycol='fate',randcol='sp',xcol=c('census','interval'),model=l
+                       start=c(yr8285=(-3.9),yr8590=(-3.9),yr9095=(-3.9),yr9500=(-3.9),yr0005=
+                       includeCovar=FALSE,badSDparam=NULL,steps=2500,burn=200,showstep=25)
>

> modTimefact$bestmu

   yr8285    yr8590    yr9095    yr9500    yr0005    yr0510
-3.508785 -4.246576 -4.106672 -4.121289 -3.942168 -4.038061

> diag(sqrt(modTimefact$bestsigma))

   yr8285    yr8590    yr9095    yr9500    yr0005    yr0510
0.7902593 1.0674807 0.6906243 0.5945213 1.1036577 0.8236507

> head(modTimefact$best,12)

          yr8285    yr8590    yr9095    yr9500    yr0005    yr0510
ade1tr -3.587128 -4.432793 -4.062271 -4.182401 -4.052210 -4.070625
alchco -3.769795 -4.042700 -4.092733 -4.132410 -4.466754 -4.064544
alsebl -4.320219 -4.929368 -4.679122 -4.709794 -4.139411 -4.412426
anacex -4.330610 -5.250803 -4.614704 -4.324225 -4.486545 -4.746223
apeihy -3.655845 -3.231356 -4.112281 -4.108328 -3.987808 -3.892533
apeime -3.663140 -4.346409 -4.076515 -4.403670 -3.856145 -4.324455
apeiti -3.086008 -3.397260 -4.137582 -4.132883 -3.899196 -4.083298
aspicr -4.049588 -4.968225 -4.441591 -4.458183 -4.851784 -4.760070
ast2gr -3.982006 -4.345425 -4.416065 -3.867244 -4.821492 -4.567105
beilpe -3.138775 -4.236013 -4.194179 -4.486422 -3.834504 -3.604205
brosal -3.530123 -4.377449 -4.300725 -3.827352 -4.638192 -3.841434
calolo -3.523083 -4.306901 -4.149346 -4.402783 -4.305933 -3.826223
```

The table *spmort* created above can also be used to overlay observed mortality for individual species. In the graph below, blue points show observed rates for *Trichilia tuberculata*, and the blue line its fitted rates. The red shows the same for *Dendropanax arborea* and the green for *Terminalia oblonga*. In the case of *Trichilia* (blue), sample size was very large, so the fitted line is very close to the observed points. In the other two, sample sizes were much smaller, and the fitted lines are drawn toward the fixed effect. That's always the way a hierarchical model works.

```
> par(cex.lab=1.5,cex.axis=1.5)
> plot(mort~meanyr,data=meanMort,pch=16,xlab='years since 1992',ylab='median mortality',xlim=c(
> cf=modTimefact$best
> nospp=dim(cf)[1]
> for(j in 1:nospp) lines(meanMort$meanyr,exp(cf[j,]),col='lightgray')
> lines(meanyr,exp(modTimefact$bestmu))
> points(mort~meanyr,data=meanMort,pch=16)
> points(spmort['tri2tu',]~meanMort$meanyr,pch=16,col='blue')
> points(spmort['dendar',]~meanMort$meanyr,pch=16,col='red')
> points(spmort['termob',]~meanMort$meanyr,pch=16,col='green')
> lines(meanMort$meanyr,exp(cf['tri2tu',]),col='blue')
> lines(meanMort$meanyr,exp(cf['dendar',]),col='red')
> lines(meanMort$meanyr,exp(cf['termob',]),col='green')
```
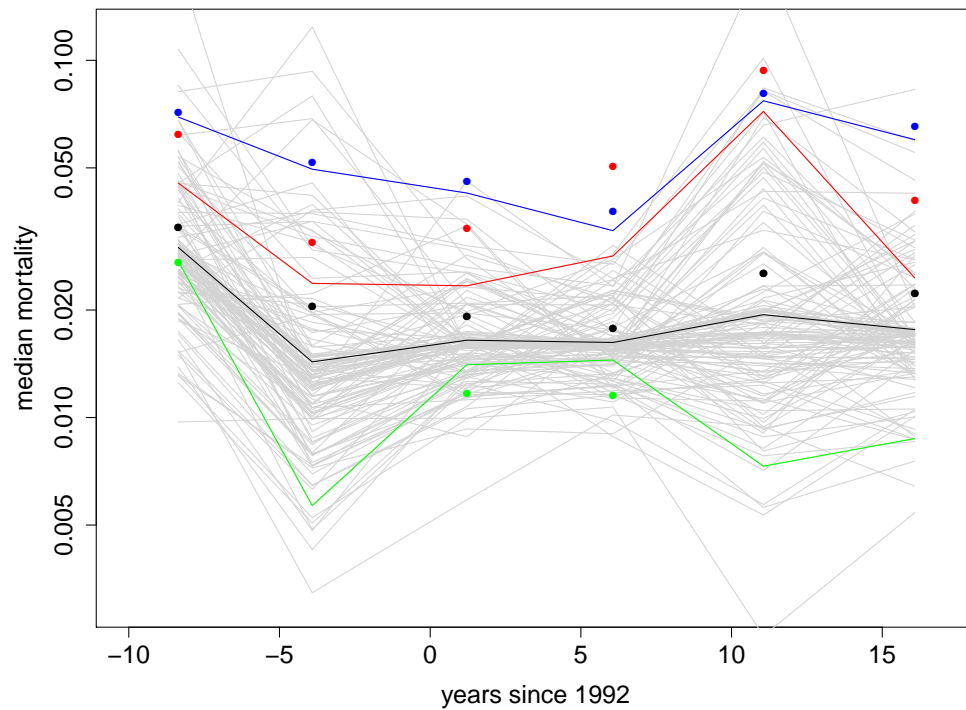
- How many species had their highest mortality in the first census interval?
- How many species had their lowest mortality in the last interval?
- Did mortality rate in the first interval correlate with mortality in the last?
- Do saplings respond similarly to big trees (ie, repeat the model for dbh range 10-50 mm, or other sizes)?

5. FILE CREATION