SimulatingHardiness

Faith Jones

27/01/2020

Winegrape winter hardiness

background

It can get really cold in some wine growing regions, sometimes pushing -30 degrees C. These types of temperatures will kill tissue unless it is protected somehow. Protecting tissue from extreme cold has a cost though, so plants dont want to be doing this unless they have to. Their solution is to get increasingly cold tollerant as the winter sets in, and then gradually lose this hardiness as winter wanes (Figure 1). Winter hardiness in generally quantified in terms of LTE50 which is teh temperature where 50% of buds die. There are a few different people trying to better understand winter hardiness in winegrapes, but we still dont have a full understanding of the process. Not all vines follow the same hardiness trajectory either, but we currently do not know how much variety matters in determening winter hardiness.

###Aim and questiom I aim to partitian the variation in winter hardiness (LTE50) into variation by year and by variety. My question is how influencial is variety in explaining/predicting cold hardiness at a given temperature. This is an interesting question because if variety doesnt show as a substantial element of the general variation in hardiness, then perhaps we can use exising data from some varieties to help predict hardiness of other varieties. I haven’t quite got a more solid question yet, nor a clear idea of what proportion of the variation i expect to come from year vs variety.

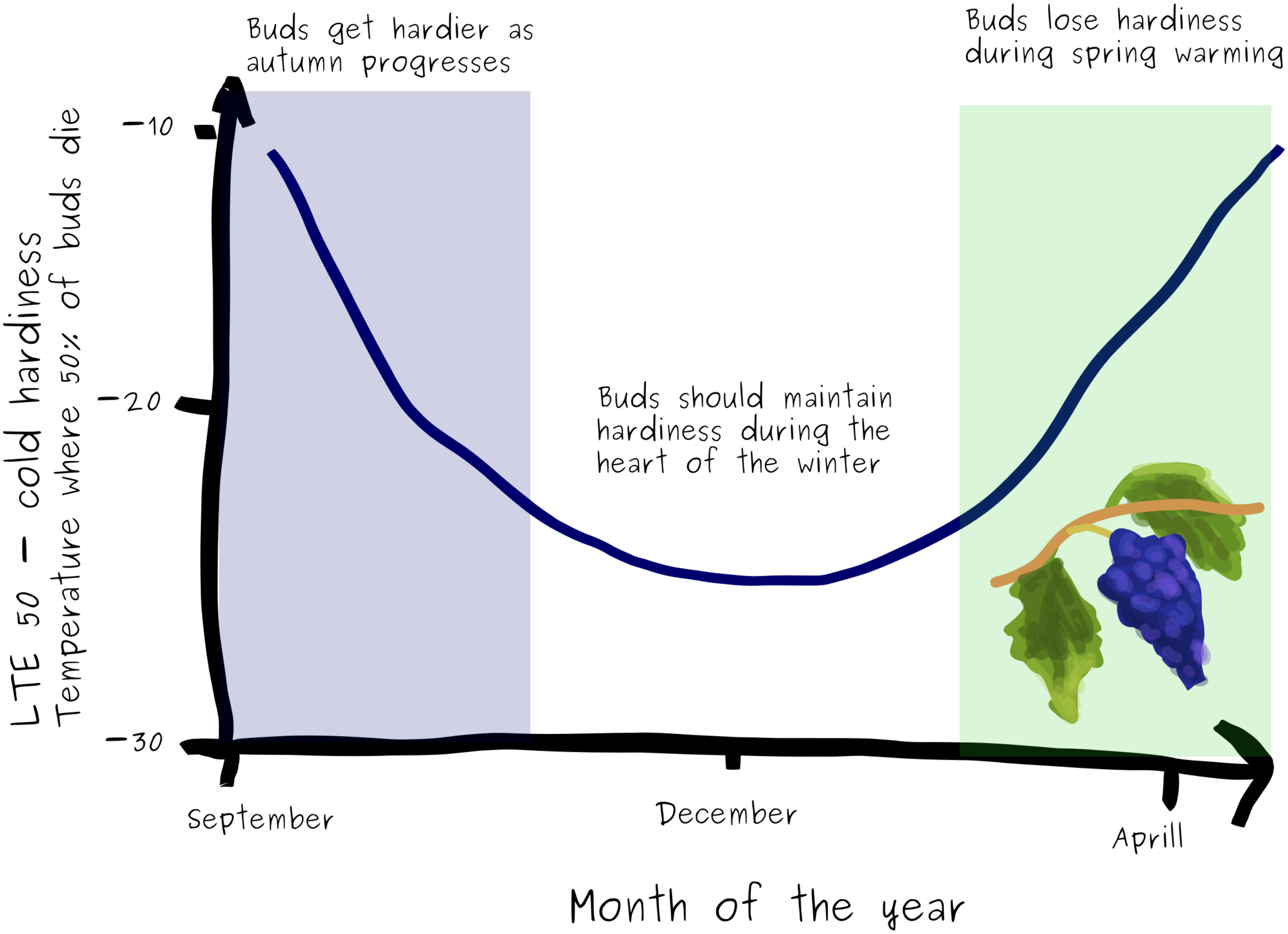

Figure 1. Conceptual diagram of winegrape winter hardiness over a year

Model structure

I intend to run a multi level model regressing LTE50 against air temperature, with variation components in the intercept from variety and year (Figure 2). I might add in a slope effect as well, we shall see. When I trialed the model in lmer the random slope model wouldn't converge. I have chosen what I hope are semi-informative priors.

$$
\begin{aligned}
LTE50_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{ijk} + \beta * x_i \\
\alpha_{ijk} &\sim \text{Normal}(\alpha_g, \sigma_{jk}) \\
\alpha_g &\sim \text{Normal}(-15, 12) \\
\sigma_{jk} &= \sigma_j + \sigma_k \\
\sigma_j &\sim \text{truncnorm}(0, 5) \\
\sigma_k &\sim \text{truncnorm}(0, 5) \\
\sigma &\sim \text{truncnorm}(0, 5)
\end{aligned}
$$

Figure 2. Sketch of the model

Simulating the data

My simulated data had simulated temperatures in the range ofof the original data’s temperatures. Is also includes 20 different varieties and 20 different years. Each variety and year combination has 30 random temperature/LTE50 observations.

Variety should effect the LTE50 intercept with a standard deviation of 0.3 Year should effect the LTE50 intercept with a standard deviation of 0.5. SO for this simulation year is more influencial than winegrape variety, but only just.

The main parameters were taken from a quick lmer mixed effect model of the real data.

As you can see from Figure 3, there is a fairly tight positive linear relationship between LTE50 and air temperture. Variety (Figure 4) and year (Figure 5) show some variation between groups, and the variation is more pronounced for year than variety.

```
nrep <- 30 # number of observationsfor each variety and year combination (days sampled i
n a year and for a variety )


#Grand model parameters
alpha <- -21.4 # grand mean
beta <- 0.52 # slope
sigma <-  0.6 #general variation

#random effects - set how many and giev them names
nvariety <- 20 # the number of vareties
varNames <- as.factor(c(1:nvariety)) # make 20 "varieties" named "1" to "20"
nyear <- 20 # there are 20 years of data
yearNames <- as.factor(1:20) #name of each year

#how many observations overall i will need
nObs <- nyear*nvariety*nrep # the number of observations  for each year and each variety
combined. The number of data points i need for temp and LTE50.

#simulating the X variable
meanTemp <- 2.03
sigmaTemp <- 4.81
simTemps <- rnorm(nObs, meanTemp,sigmaTemp) # these are the temperatures

#repeat the random effects so I have full coverage
alphavar <- rep( rnorm(nvariety, 0, 0.3), each = nyear )# random 20 draws from a normal
 distribution with mean 0 and sd 0.3, repeated 20 times
alphaVarObs <- rep(alphavar, each = nrep) # repeat each alpha for variable 30 times for
 each data observation
alphaYear <- rep(rnorm(nyear, 0, 0.5), times = nvariety) # random effect of year, 20 yea
rs and each year has a each variety in it
alphaYearObs <- rep(alphaYear, each = nrep)

#make columns for the name of the year, variety and day of the year
varNamesRep <- rep(varNames, each = nyear)
varNamesObs <- rep(varNamesRep, each = nrep)
YearNameRep <- rep(1:nyear, times = nvariety)
yearNamesObs <- rep(YearNameRep, each = nrep)

#individual data point variation
eps <- rnorm(nObs , 0, sigma) # randomly simulated variation for each point on top of ra
ndom effects

#run the simulation
simLTEVar <- alpha + alphaVarObs + alphaYearObs + beta*simTemps + eps
plot(simLTEVar ~ simTemps, xlab = "simulated temperature (degrees C)", ylab = "simulated
LTE50",
     sub = "Figure 3. Simulated slopes of winegrape hardiness")
```

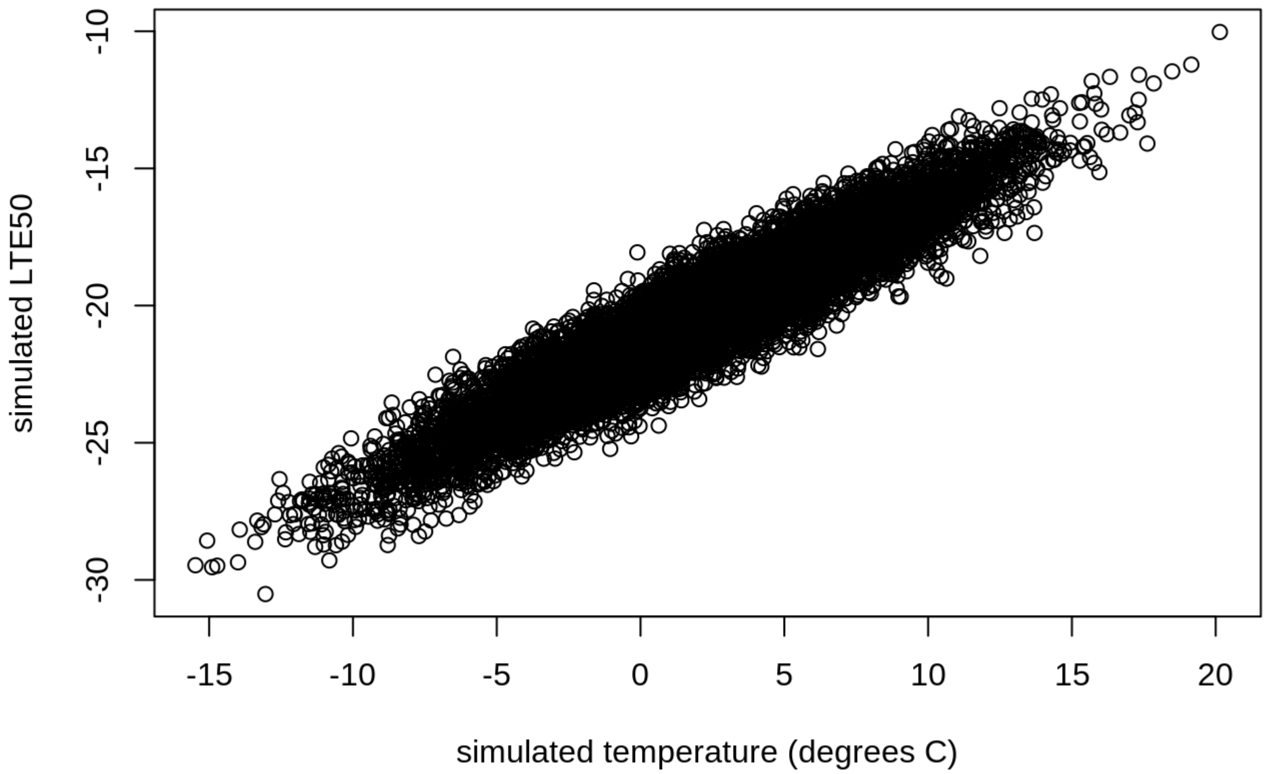

Figure 3. Simulated slopes of winegrape hardiness

```
#combine into a single data table so I can make soem more plots

simVarData <- data.frame(cbind(simTemps, varNamesObs, yearNamesObs, simLTEVar))
str(simVarData)
```

```
## 'data.frame':    12000 obs. of  4 variables:
##  $ simTemps    : num  -6.413 -0.732 0.954 5.673 -3.338 ...
##  $ varNamesObs : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ yearNamesObs: num  1 1 1 1 1 1 1 1 1 1 ...
##  $ simLTEVar   : num  -24.8 -21.5 -21.9 -18.4 -24 ...
```

```
simVarData$varNamesObs <- as.factor(simVarData$varNamesObs )
simVarData$yearNamesObs <- as.factor(simVarData$yearNamesObs )

#variety data plot
varietySimPlot <- ggplot(data = simVarData, aes(x = varNamesObs, y = simLTEVar))
varietySimPlot + geom_boxplot() +
  theme_classic()+
  ylab("Winegrape Variety") +
  xlab("Simulated LTE50") +
  labs(caption = "Figure 4. LTE50 values for each winegrape variety")
```

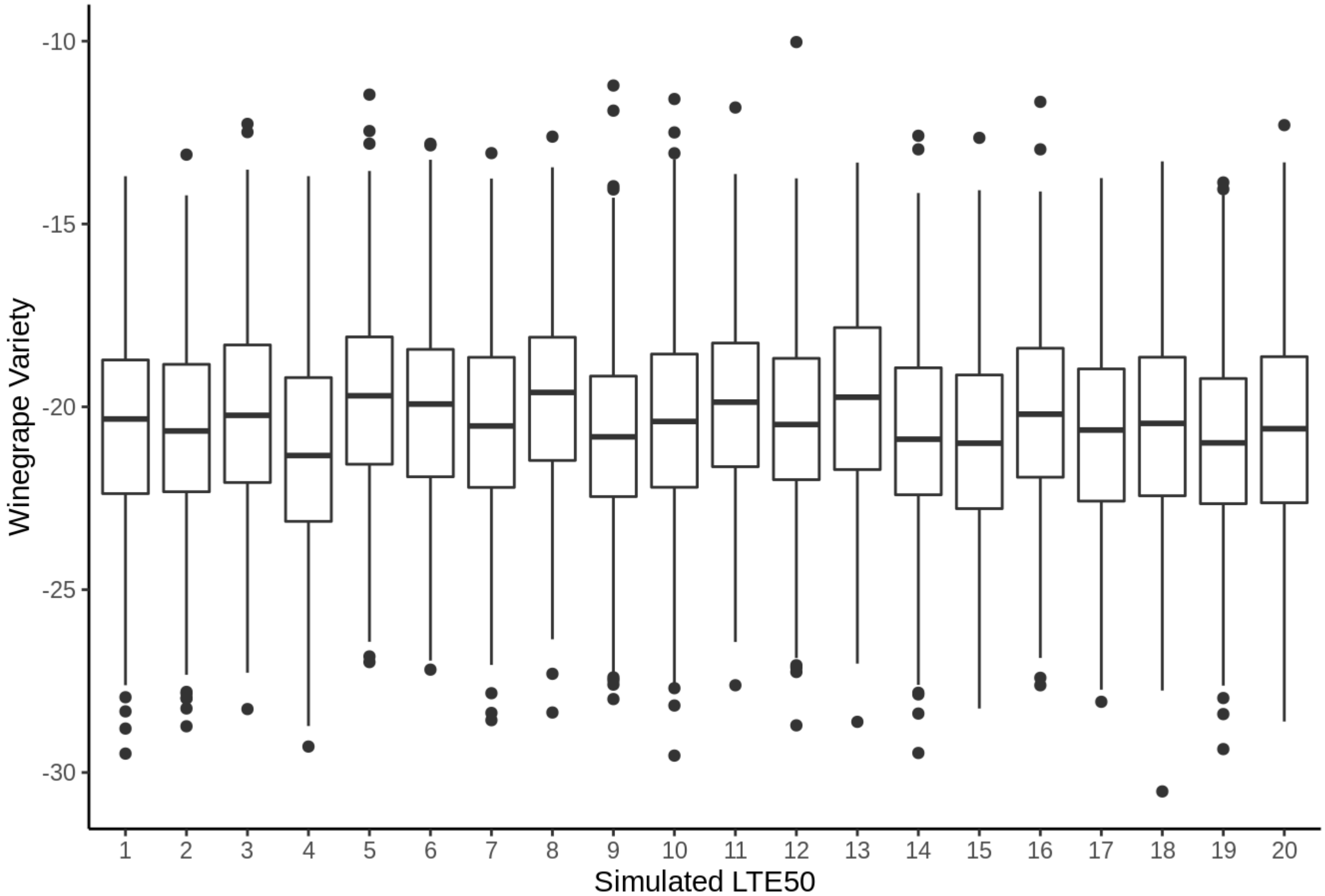

Figure 4. LTE50 values for each winegrape variety

```
#year data plot
varietySimPlot <- ggplot(data = simVarData, aes(x = yearNamesObs, y = simLTEVar))
varietySimPlot + geom_boxplot() +
  theme_classic()+
  xlab("Year of Data Collection") +
  ylab("Simulated LTE50") +
  labs(caption = "Figure 5. LTE50 values for each year of data collection")
```

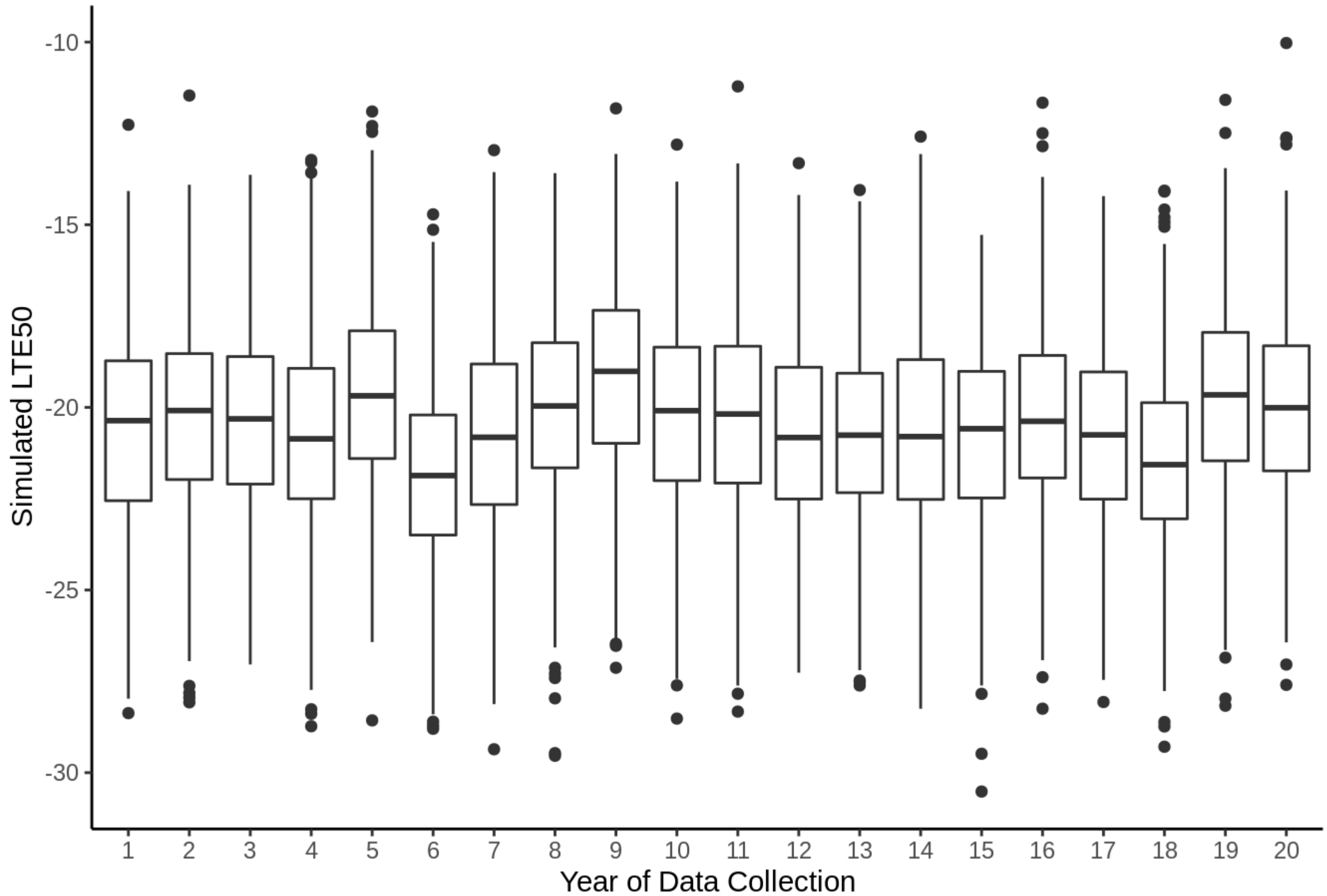

Figure 5. LTE50 values for each year of data collection

Prior predictive checks

I then ran the “model” 20 times, and plotted the main slope, the slope for the effects of year and variety seperatly, and finally how the two different random effects interact. I show the effect of a single random effect in Figure 6, which is a truncated normal distribution with a standard error of 5. Both variety and year have the same prior for their effect on the grand alpha. In Figure 7 is the effect of both variety and year on the model outputs. The suggested variation in LTE50 is quite high, but not so high that I’m losing sleep over it.

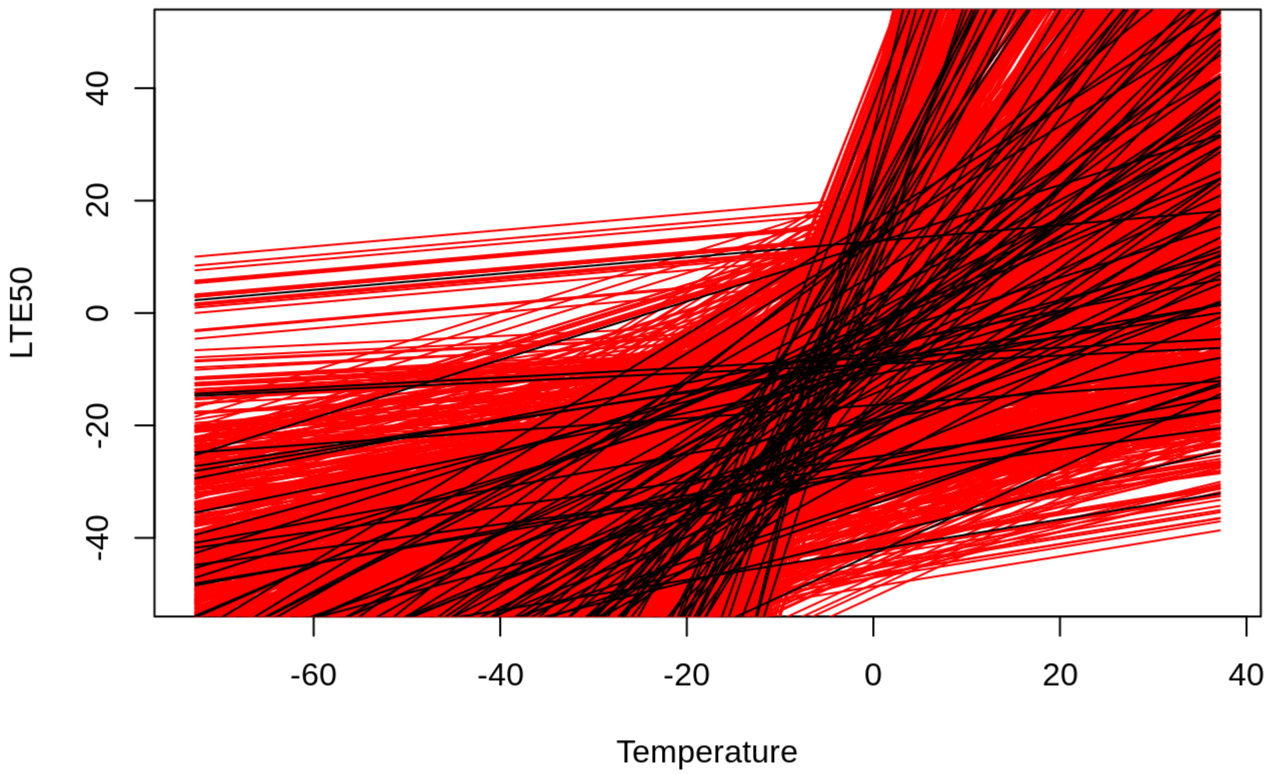

Figure 6. The distribution of slopes suggested by my model with one effect

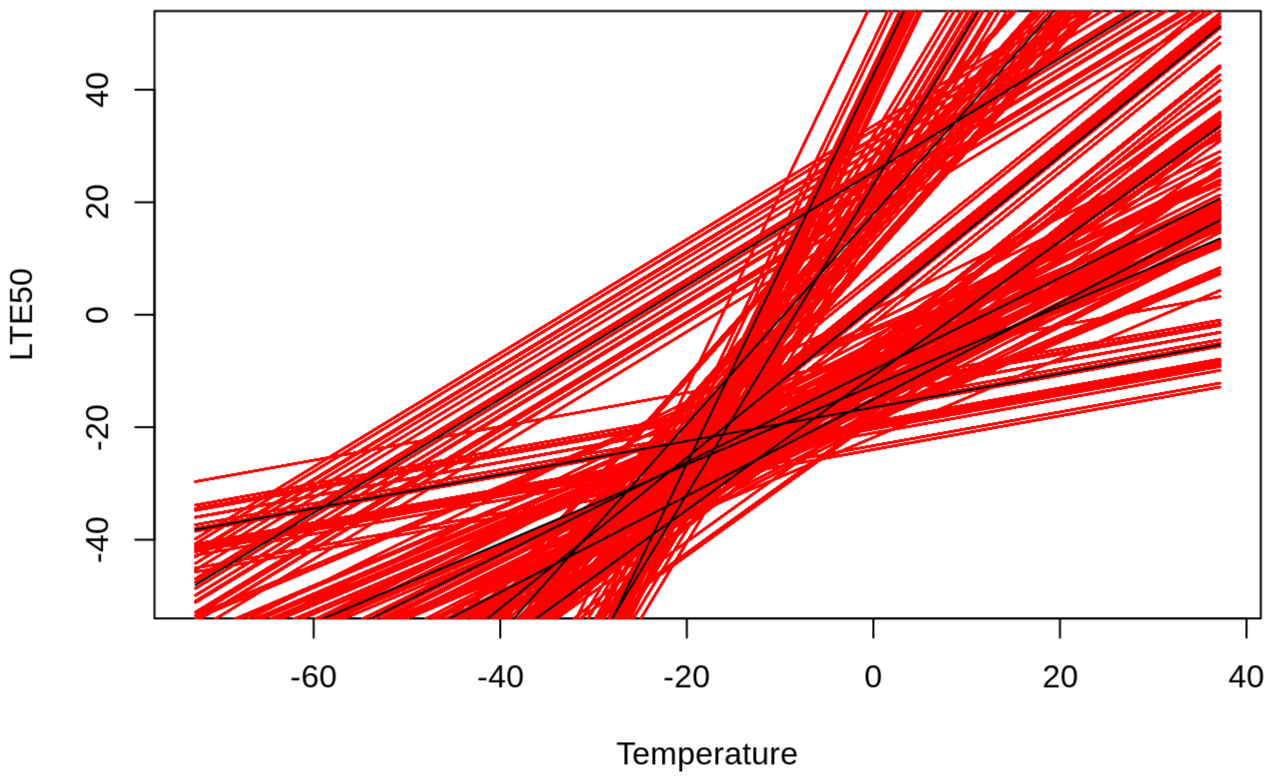

Figure 7. The distribution of slopes suggested by my model with both effects