How do you analyze non-linear relationships? Height and number of tillers in wire grass

Non-linear relationships are common in ecological studies. Here we develop frequentist and Bayesian quadratic models of an allometric relationship for a grass species. On September 8, 2013, UCF graduate students measured the height and counted the tillers of hundreds of wiregrass (*Aristida beyrichiana*, formerly called *Aristida stricta*) clumps scheduled to be transplanted to the field. The analysis presented in this demo was used as a baseline in a study that evaluates the effect of microhabitat and ridge elevation on wiregrass growth and survival, with hopes of learning how to make successful reintroductions for restoration of native ecosystems.

For this demo you will need to download two R scripts (`wire_grass.R`, `wg_priors.R`), and two data files (`wg_data.txt`, `wg_priors.txt`). You will also need to have installed the package R2OpenBUGS and the OpenBUGS software.

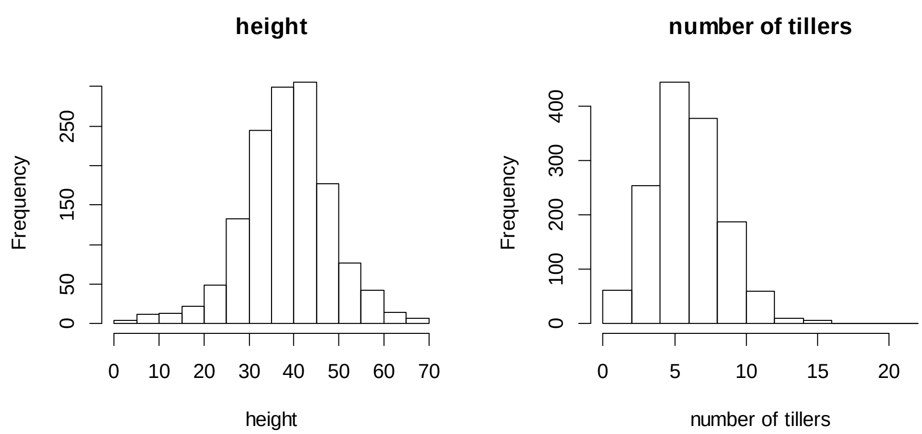

Figure 1. Histograms for plant height in cm and number of tillers of wire grass (data collected by Jennifer Navarra, UCF biology graduate students and Pedro Quintana-Ascencio).

We start in the `wire_grass.R` script by plotting the data in `wg_data.txt` (lg: height, lf: number of tillers). The histograms in Figure 1 depict the spread and distribution of height and number of tillers, while the plot in Figure 2 indicates that a quadratic relationship is a reasonable hypothesis about their association.

```
reg_data <- read.table("wg_data.txt", header=T)
par(mfrow=c(1,2))
hist(reg_data$lg)
hist(reg_data$lf)

par(mfrow=c(1,1))
plot(reg_data$lf,reg_data$lg,pch=16,cex=0.55,xlab="number of tillers",ylab="height")
```

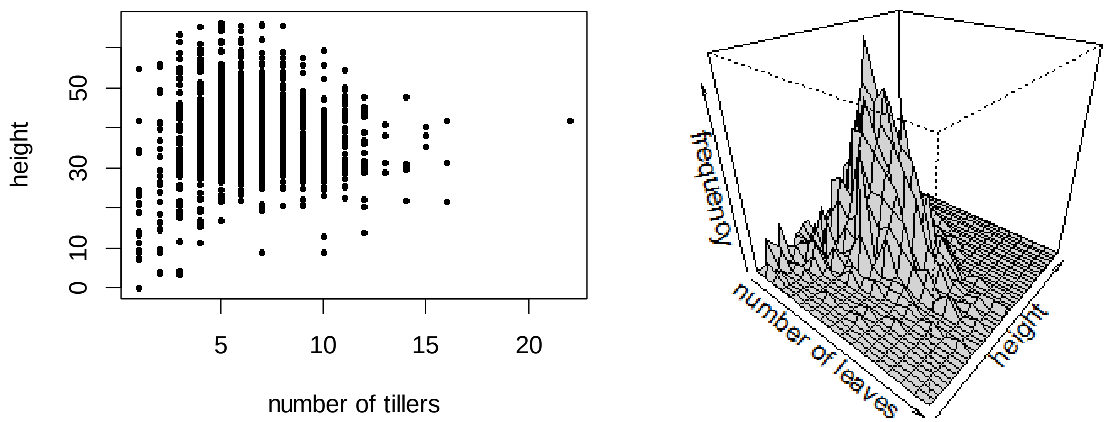

Figure 2. Plot of number of tillers vs. plant height in cm.

$$Height = \beta_1 + \beta_2 * tillers + \beta_3 * tillers^2 \qquad \epsilon \sim N(0, \sigma)$$

We also recognize that there is a potential outlier beyond 16 tillers and decide to remove the data for this individual. Then, we create the quadratic values to evaluate our hypothesis and prepare an index to order the data in the plots properly.

```
max <- 16
x<- reg_data$lf[reg_data$lf<max]
x2 <- x^2
y <- reg_data$lg[reg_data$lf<max]
x0 <- na.omit(x)
ord <- order(x0)
g <- reg_data$Observers[reg_data$lf<max]
tp <- reg_data$tp[reg_data$lf<max]
```

We evaluate six models including a linear model and five quadratic models. We also use two additional predictor variables: whether there was one plant or two plants per pot (tp), and the identity of the group of students that collected the data (g). We compare the models using AIC. Notice that when you want to include an interaction with a quadratic variable, you have to do it for both the x and the x^2 terms.

```
model1 <- lm(y ~ x)
model2 <- lm(y ~ x + x2)
model3 <- lm(y ~ x + x2 + tp)
```

```
model4 <- lm(y ~ x + x2 + tp + x:tp + x2:tp)
model5 <- lm(y ~ x + x2 + g)
model6 <- lm(y ~ x + x2 + g + tp)
AIC(model1,model2,model3,model4,model5,model6)
```

The most informative model was model 4. This is a model with quadratic relationships specific to plants in pots with either one or more than one plant (Figure 3). Below is the summary of this model. There is no pattern among its residuals that are normally distributed. The students did a good job and we do not have evidence of differences among their estimates!

$$\boldsymbol{Height = \beta_1 + \beta_2 * tillers + \beta_3 * tillers^2 + \beta_4[type] + \beta_5[type] * tillers + \beta_6[type] * tillers^2}$$

$$\boldsymbol{\epsilon \sim N(0, \sigma)}$$

```
Call:
lm(formula = y ~ x + x2 + tp + x:tp + x2:tp)

Residuals:
    Min      1Q  Median      3Q     Max 
-31.988  -5.923  -0.188   5.548  30.174 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  42.32494    4.32678   9.782  < 2e-16 ***
x             0.08122    1.36831   0.059  0.95267    
x2            0.04025    0.10953   0.367  0.71332    
tp          -11.47617    2.89623  -3.962 7.79e-05 ***
x:tp          2.87365    1.05622   2.721  0.00660 ** 
x2:tp        -0.25749    0.09400  -2.739  0.00623 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.178 on 1393 degrees of freedom
Multiple R-squared:  0.122,	Adjusted R-squared:  0.1188 
F-statistic:  38.7 on 5 and 1393 DF,  p-value: < 2.2e-16
```

Next we do some informative plots of our model (see code in the script). Figure 3 shows the data separated by the number of plants in the pot together with our two prediction lines. Figure 4 shows the distribution of the residuals for each type.

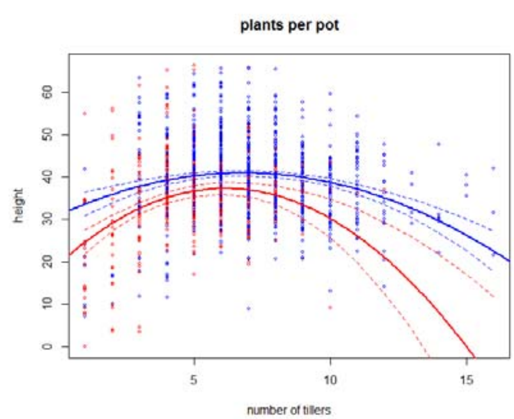

Figure 3. Plot of number of tillers vs. plant height in cm. In red plants with more than one plant per por. In blue plants with one plant per pot.

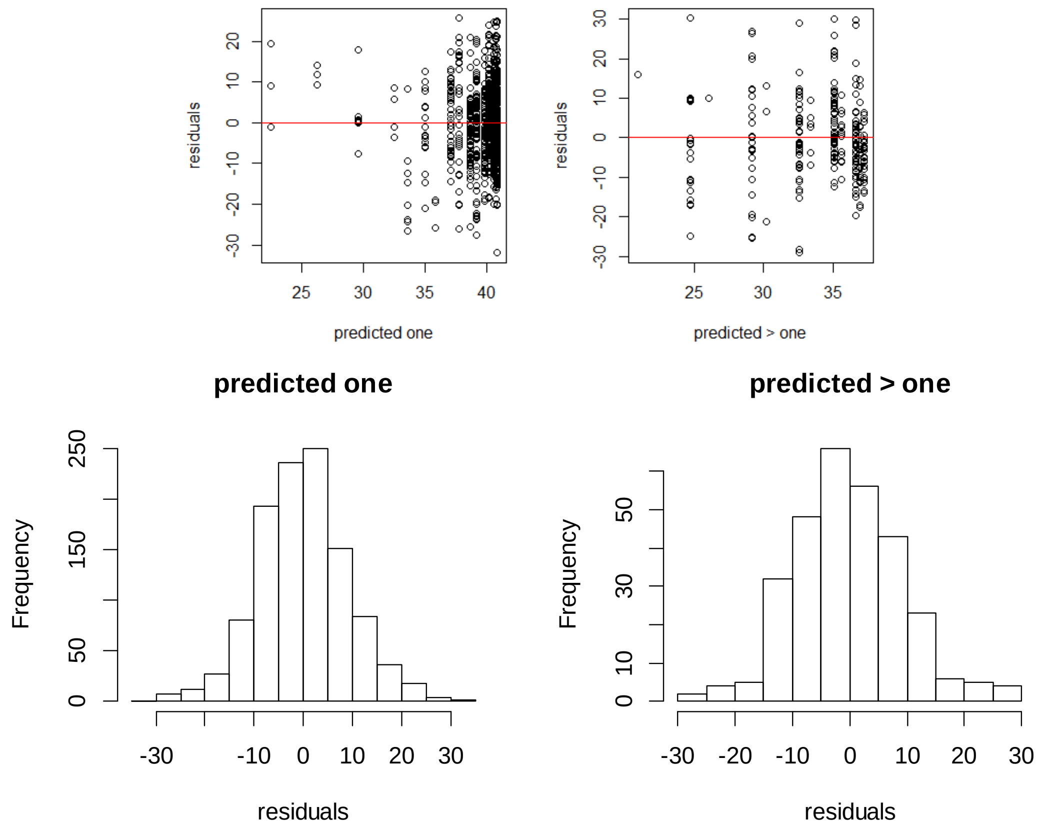

Figure 4. Plot of the residuals of model 4 by category of number individuals per pot. These residuals are presented for heuristic purposes.

We now implement model 4 under the Bayesian framework with uninformative priors and plot the results in Figure 5.

```
library(R2OpenBUGS)
n <- length(y)

# Write model
linreg<-function(){
  for (i in 1:n){
  Y[i] <- y[i]
  Y[i] ~ dnorm(mean[i], prec)
  mean[i]<-b[1]+b[2]*x[i]+b[3]*x2[i]+b[4]*tp[i]+b[5]*x[i]*tp[i]+b[6]*x2[i]*tp[i]
  }

# Uninformative priors
  for (i in 1:6){
  b[i] ~ dnorm(0, 1.0E-6)
  }
  prec ~ dgamma(0.001, 0.001)
}
write.model(linreg,"linreg.txt")

# Bundle data
win.data <- list("n","x","x2","y","tp")
```

```
# Inits function
inits<-function()
{list(b=c(runif(1),runif(1),runif(1),runif(1),runif(1),runif(1)),prec=100)}

# Parameters to estimate
params <- c("b","prec")

# MCMC settings
nc=3
ni=2000
nb=200
nt=100

# Start Gibbs sampler
out <- bugs(data=win.data,inits=inits,parameters=params,model="linreg.txt",
n.thin=nt,n.chains=nc,n.burnin=nb,n.iter=ni,codaPkg=T)
```

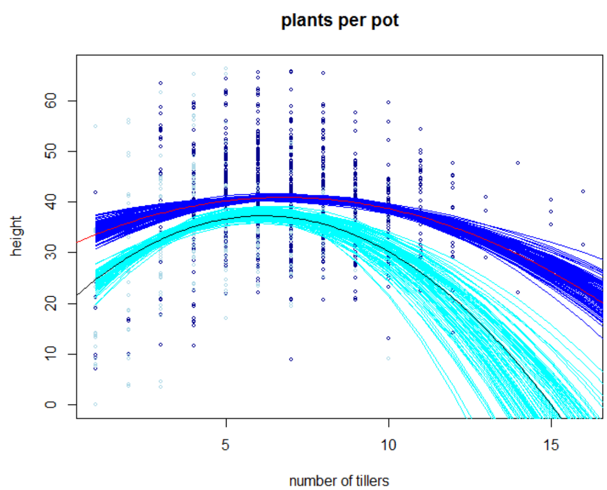

Figure 5. Plot of number of tillers vs. plant height in cm. In light blue plants with more than one plant per pot. In dark blue plants with one plant per pot. Models with uninformative priors. The frequentist predicted values are represented with the red and black lines.

The parameters of the frequentist and the Bayesian model with uninformed Bayesian priors were commensurate (Table 1). Notice that x and x^2 (the two parameters for which the null hypothesis of no difference from zero was not rejected in the frequentist framework), were the most different among models. Also observe that in the plot of the Bayesian model we do not present one line, instead we represent the distribution of the predicted models.

Table 1. Model 4 parameters and their standard errors under Frequentist and Bayesian with uninformed priors (remember your results will be slightly different).

Parameter	Frequentist	SE	Bayesian (uninformed)	SE
Intercept - b[1]	42.32	4.32	43.80	5.08
x - b[2]	0.081	1.37	-0.45	1.63
x^2 - b[3]	0.040	0.11	0.082	0.13
tp - b[4]	-11.48	2.90	-12.56	3.41
x:tp - b[5]	2.87	1.06	3.29	1.26
x^2:tp - b[6]	-0.257	0.09	-0.292	0.11

But what if we could get information from the literature to build informed priors? After all we must not be the first people to ever work with wiregrass! Gordon and Rice (1998) collected similar data on number of tillers and height of *Aristida beyrichiana* at seven different populations in Florida: Torreya State Park, Apalachicola Bluffs, Ravines Preserve, Wekiwa Springs (2 populations) and Rock Springs Run (2 populations - Figure 6). They kindly allowed us to use their data (which is summarized in Table 2). See the script called `wg_priors.R` for how we got this.

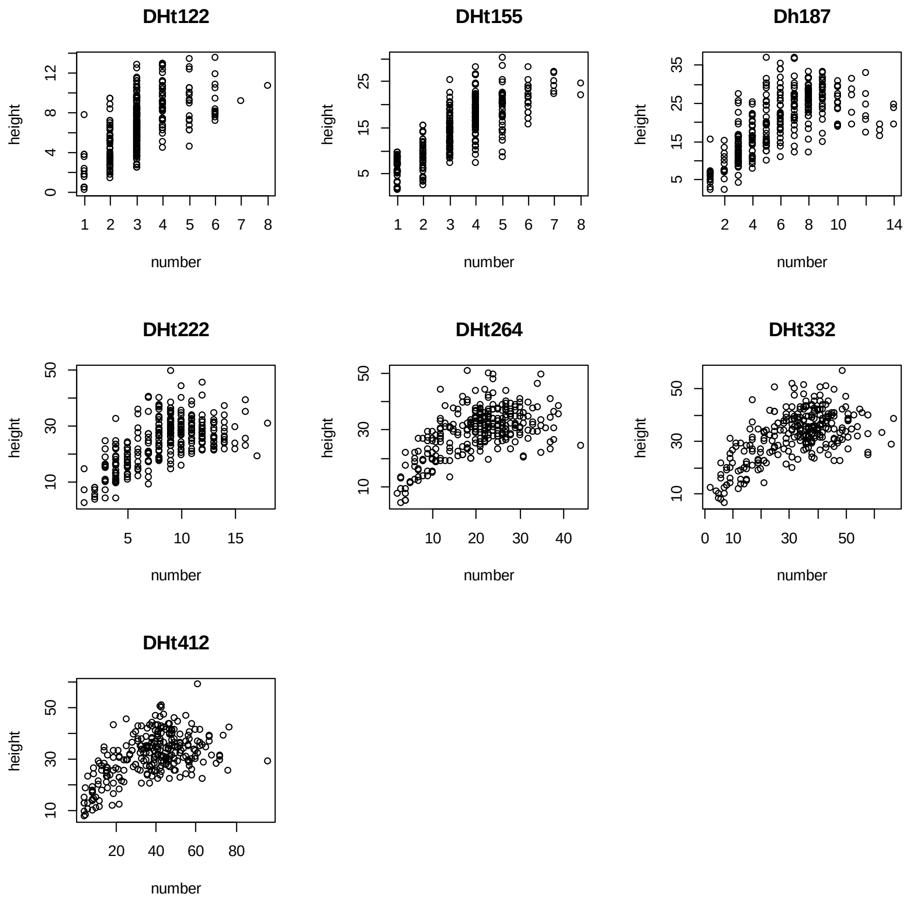

Figure 6. Plots of number of tillers vs. plant height in cm using data from Gordon and Rice (1998).

Table 2. Parameters and their standard errors calculated using Gordon and Rice (1998) data.

Populations	Intercept	x	x^2	tp	tp:x	tp:x^2
Pop 1	-1.2664	3.5446	-0.2885	NA	NA	NA
Pop 2	0.9001	5.2416	-0.2883	NA	NA	NA
Pop 3	-0.7671	5.6831	-0.2987	NA	NA	NA
Pop 4	-0.2650	4.9421	-0.2056	NA	NA	NA
Pop 5	5.8658	2.0024	-0.0356	NA	NA	NA
Pop 6	7.2630	1.3669	-0.0156	NA	NA	NA
Pop 7	11.3291	0.9112	-0.0085	NA	NA	NA
Mean	3.294	3.384	-0.163			
Variance	23.769	3.880	0.019			

We replace the priors in the above Bayesian Model with ones that reflect this new information as shown below (the rest of the code is the same). A comparison of the uncertainty of both models indicates narrower credibility intervals than the confidence intervals for the frequentist approach (Table 3). A plot of the realization of the Bayesian model with informative priors is shown in Figure 7, and a comparison among priors and posteriors in Figure 8.

```
# Informative priors
  b[1] ~ dnorm(3.294,0.042)
  b[2] ~ dnorm(3.384,0.258)
  b[3] ~ dnorm(-0.163,52.86)
    for (i in 4:6){
    b[i] ~ dnorm(0,1.0E-6)
    }
```

Table 3. Comparison between parameters of the Frequentist and the Bayesian Model with informed priors based on data from Gordon and Rice (1998).

Parameter	Frequentist	SE	Bayesian (informed)	SE
intercept	42.32	4.32	27.06	2.86
x	0.081	1.37	4.37	0.854
x^2	0.040	0.11	-0.236	0.073
tp	-11.48	2.90	-2.08	1.93
:tp	2.87	1.06	0.090	0.717
x^2:tp	-0.257	0.09	-0.069	0.069

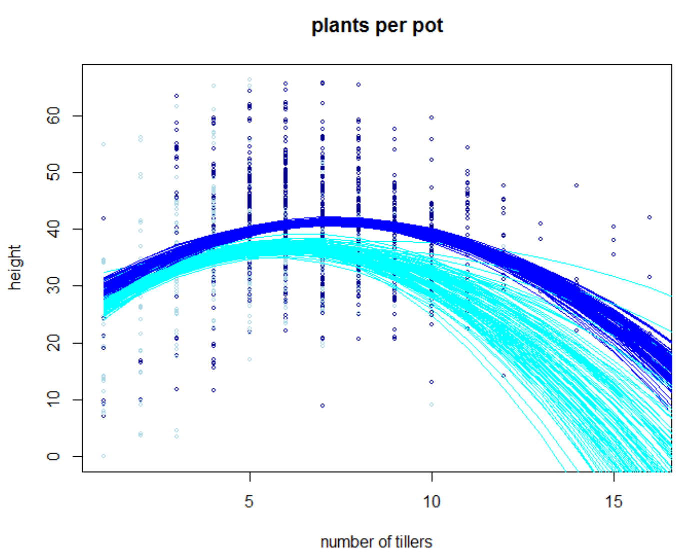

Figure 7. Plot of number of tillers vs. plant Height in cm. In light blue plants with more than one plant per pot. In dark blue plants with one plant per pot. Model with informative priors based on data from Gordon and Rice (1998).

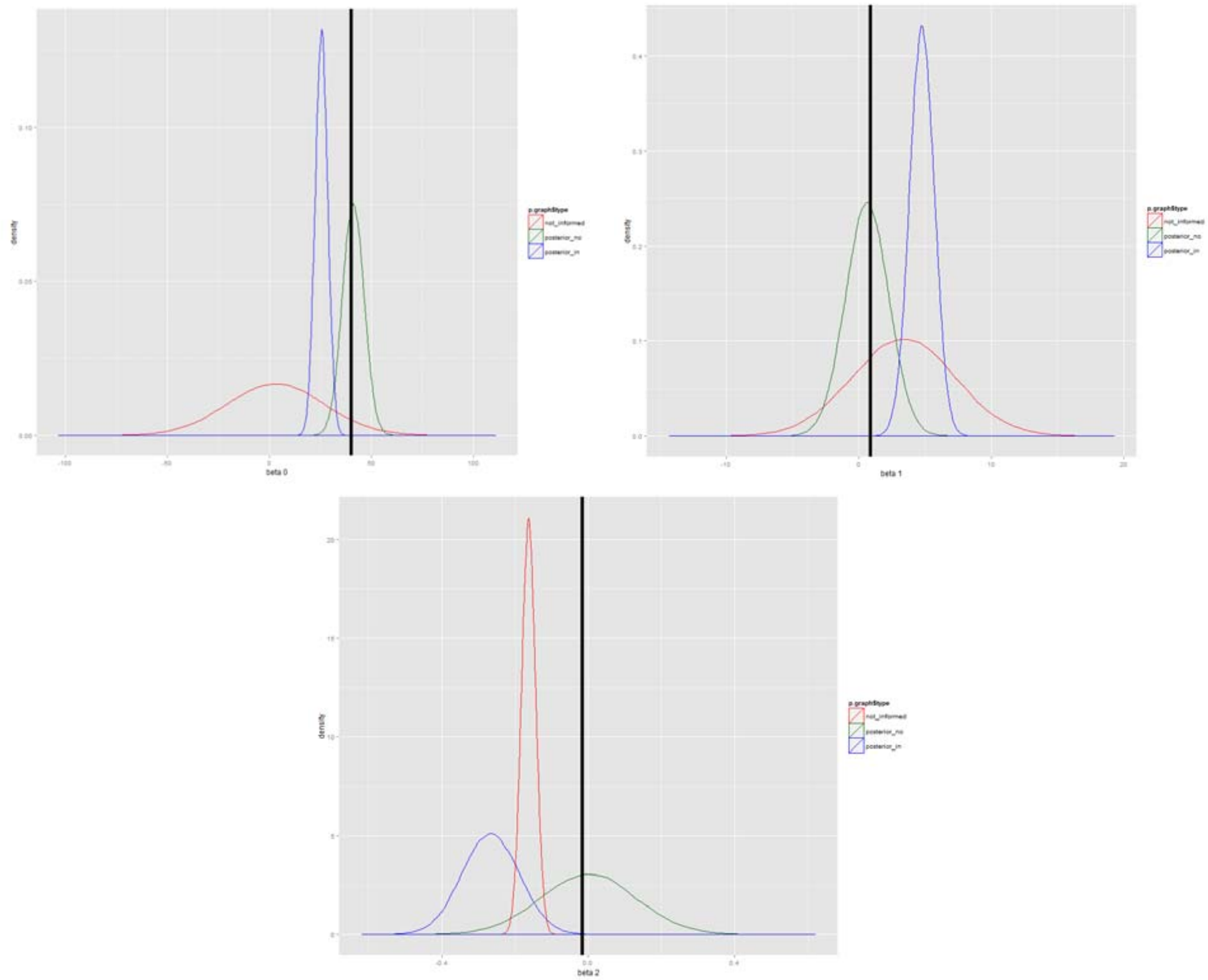

Figure 8. Informative prior distribution in red and posteriors of models with non-informative (in green) and informative priors (in blue) for parameters β_0, β_1 and β_2. Value of the frequentist estimate as a line in black. We used uninformed priors for those parameters for which there was no available information.

References

Gordon, D.R. and K. J. Rice. 1998. Patterns of differentiation in Wiregrass (*Aristida beyrichiana*): implications for restoration efforts. Restoration Ecology 6: 166-174.