Joseph Crockett

ES 207: Environmental Data Analysis

March 1, 2016

**Homework Assignment 4**

**Objective Statement:**

In order to reasonably estimate standing carbon, we will further refine our linear model between Diameter at Breast Height and Tree height, factored by either genus or project site.

**Methods:**

We determined the mean, standard deviation, and count, for the five subsetted genera by plot site. Plots at which more than one measurement was taken were retained. Of these, we randomly sampled 6 plot sites for each of the four project areas. The coefficient of variation was calculated, and an ANOVA and an TukeyHSD were run to determine whether project area was a determining factor for tree height. We developed two linear models using genus as single terms and as an interaction variable describing Height (cm) ~ DBH (cm).

**Data:**

The data are tree measurements from 4 riparian project areas in Northern California. We have previously cleansed the dataset for illogical or unknown values and subsetted for the 5 most frequent genera. Outliers were identified and removed as well.

**Code:**

rd\_lab4$projplot <- as.factor( paste(rd\_lab4$ProjCode,rd\_lab4$Plot.Name))  
  
#Generating mean, sd, and count of 5 most frequent genera at each plot.  
rd\_sum <- data.frame(cbind(tapply(rd\_lab4$Woody\_Height\_cm, rd\_lab4$projplot, mean), tapply(rd\_lab4$Woody\_Height\_cm,rd\_lab4$projplot,sd),tapply(rd\_lab4$Woody\_Height\_cm,rd\_lab4$projplot,length)))  
  
colnames(rd\_sum) <- c("htcmmn","htcmsd","plot.n")  
  
rd\_sum$projplot <- as.factor(rownames(rd\_sum))  
#needs greater than 1 measurement  
rd\_sum <- rd\_sum[rd\_sum$plot.n >1 ,]  
#new column for labels  
rd\_sum$proj <- as.factor(substr(rd\_sum$projplot,1,5))  
  
projlevels <- levels(rd\_sum$proj)  
  
for (p in 1:length(projlevels)){  
 print(summary(rd\_sum[rd\_sum$proj == projlevels[p],]))  
}

## htcmmn htcmsd plot.n projplot   
## Min. : 137.1 Min. : 42.43 Min. : 2.00 CORP 1 : 1   
## 1st Qu.: 768.7 1st Qu.: 247.74 1st Qu.: 10.00 CORP 2 : 1   
## Median :1000.0 Median : 412.96 Median : 18.00 CORP 3 : 1   
## Mean :1035.2 Mean : 439.64 Mean : 25.15 CORP 4 : 1   
## 3rd Qu.:1229.1 3rd Qu.: 615.85 3rd Qu.: 32.00 CORP 5 : 1   
## Max. :3330.0 Max. :1031.60 Max. :111.00 CORP 6 : 1   
## (Other):83   
## proj   
## CORP :89   
## HOWY : 0   
## NASO : 0   
## SRRB : 0   
##   
##   
##   
## htcmmn htcmsd plot.n projplot  
## Min. : 421.3 Min. : 68.18 Min. : 7.00 HOWY RIP01:1   
## 1st Qu.: 505.6 1st Qu.: 141.83 1st Qu.:11.00 HOWY RIP02:1   
## Median : 883.0 Median : 218.21 Median :27.00 HOWY RIP03:1   
## Mean : 964.1 Mean : 409.68 Mean :38.86 HOWY RIP04:1   
## 3rd Qu.:1292.6 3rd Qu.: 528.12 3rd Qu.:62.00 HOWY RIP05:1   
## Max. :1848.3 Max. :1241.50 Max. :92.00 HOWY RIP06:1   
## (Other) :1   
## proj   
## CORP :0   
## HOWY :7   
## NASO :0   
## SRRB :0   
##   
##   
##   
## htcmmn htcmsd plot.n projplot  
## Min. : 671.0 Min. :171.6 Min. : 4.00 NASO Crp2013\_509:1   
## 1st Qu.: 745.4 1st Qu.:472.4 1st Qu.: 6.75 NASO S02013\_200 :1   
## Median : 814.6 Median :556.3 Median :17.50 NASO S02013\_202 :1   
## Mean : 977.8 Mean :538.2 Mean :24.00 NASO S02013\_203 :1   
## 3rd Qu.:1016.4 3rd Qu.:649.9 3rd Qu.:29.25 NASO S02013\_207 :1   
## Max. :1888.3 Max. :819.5 Max. :81.00 NASO S02013\_211 :1   
## (Other) :2   
## proj   
## CORP :0   
## HOWY :0   
## NASO :8   
## SRRB :0   
##   
##   
##   
## htcmmn htcmsd plot.n projplot   
## Min. : 443.2 Min. : 57.79 Min. : 2.0 SRRB EW2013\_100: 1   
## 1st Qu.: 664.8 1st Qu.: 246.65 1st Qu.: 4.0 SRRB EW2013\_101: 1   
## Median : 977.0 Median : 572.42 Median : 7.0 SRRB EW2013\_102: 1   
## Mean :1065.4 Mean : 610.42 Mean :11.8 SRRB EW2013\_103: 1   
## 3rd Qu.:1256.6 3rd Qu.: 882.65 3rd Qu.:17.0 SRRB EW2013\_106: 1   
## Max. :2532.9 Max. :1668.77 Max. :59.0 SRRB EW2013\_110: 1   
## (Other) :38   
## proj   
## CORP : 0   
## HOWY : 0   
## NASO : 0   
## SRRB :44   
##   
##   
##

lapply(projlevels, function(x) summary(rd\_sum[rd\_sum$proj == x,]))

## [[1]]  
## htcmmn htcmsd plot.n projplot   
## Min. : 137.1 Min. : 42.43 Min. : 2.00 CORP 1 : 1   
## 1st Qu.: 768.7 1st Qu.: 247.74 1st Qu.: 10.00 CORP 2 : 1   
## Median :1000.0 Median : 412.96 Median : 18.00 CORP 3 : 1   
## Mean :1035.2 Mean : 439.64 Mean : 25.15 CORP 4 : 1   
## 3rd Qu.:1229.1 3rd Qu.: 615.85 3rd Qu.: 32.00 CORP 5 : 1   
## Max. :3330.0 Max. :1031.60 Max. :111.00 CORP 6 : 1   
## (Other):83   
## proj   
## CORP :89   
## HOWY : 0   
## NASO : 0   
## SRRB : 0   
##   
##   
##   
##   
## [[2]]  
## htcmmn htcmsd plot.n projplot  
## Min. : 421.3 Min. : 68.18 Min. : 7.00 HOWY RIP01:1   
## 1st Qu.: 505.6 1st Qu.: 141.83 1st Qu.:11.00 HOWY RIP02:1   
## Median : 883.0 Median : 218.21 Median :27.00 HOWY RIP03:1   
## Mean : 964.1 Mean : 409.68 Mean :38.86 HOWY RIP04:1   
## 3rd Qu.:1292.6 3rd Qu.: 528.12 3rd Qu.:62.00 HOWY RIP05:1   
## Max. :1848.3 Max. :1241.50 Max. :92.00 HOWY RIP06:1   
## (Other) :1   
## proj   
## CORP :0   
## HOWY :7   
## NASO :0   
## SRRB :0   
##   
##   
##   
##   
## [[3]]  
## htcmmn htcmsd plot.n projplot  
## Min. : 671.0 Min. :171.6 Min. : 4.00 NASO Crp2013\_509:1   
## 1st Qu.: 745.4 1st Qu.:472.4 1st Qu.: 6.75 NASO S02013\_200 :1   
## Median : 814.6 Median :556.3 Median :17.50 NASO S02013\_202 :1   
## Mean : 977.8 Mean :538.2 Mean :24.00 NASO S02013\_203 :1   
## 3rd Qu.:1016.4 3rd Qu.:649.9 3rd Qu.:29.25 NASO S02013\_207 :1   
## Max. :1888.3 Max. :819.5 Max. :81.00 NASO S02013\_211 :1   
## (Other) :2   
## proj   
## CORP :0   
## HOWY :0   
## NASO :8   
## SRRB :0   
##   
##   
##   
##   
## [[4]]  
## htcmmn htcmsd plot.n projplot   
## Min. : 443.2 Min. : 57.79 Min. : 2.0 SRRB EW2013\_100: 1   
## 1st Qu.: 664.8 1st Qu.: 246.65 1st Qu.: 4.0 SRRB EW2013\_101: 1   
## Median : 977.0 Median : 572.42 Median : 7.0 SRRB EW2013\_102: 1   
## Mean :1065.4 Mean : 610.42 Mean :11.8 SRRB EW2013\_103: 1   
## 3rd Qu.:1256.6 3rd Qu.: 882.65 3rd Qu.:17.0 SRRB EW2013\_106: 1   
## Max. :2532.9 Max. :1668.77 Max. :59.0 SRRB EW2013\_110: 1   
## (Other) :38   
## proj   
## CORP : 0   
## HOWY : 0   
## NASO : 0   
## SRRB :44   
##   
##   
##

#randomly selecting 6 sample plot summaries from each project site  
nsamples <- 6  
ripres <-lapply(projlevels, function(x) rd\_sum[which(rd\_sum$proj == x),][sample(nrow(rd\_sum[which(rd\_sum$proj ==x),]),nsamples),])  
  
#do.call, similar to sapply, but not simple. only works on lists (hence lapply). Combining by each value (1-4) by rows.  
ripsample <- do.call(rbind,ripres)  
summary(ripsample$proj)

## CORP HOWY NASO SRRB   
## 6 6 6 6

#Find Coefficient of variation  
ripsample$cv <- with(ripsample, htcmsd/htcmmn)  
  
#one-way ANOVA   
rip.proj.cv.aov <- aov(cv~proj, data = ripsample)  
summary(rip.proj.cv.aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## proj 3 0.2962 0.09874 2.476 0.091 .  
## Residuals 20 0.7976 0.03988   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#vs  
summary.lm(rip.proj.cv.aov)

##   
## Call:  
## aov(formula = cv ~ proj, data = ripsample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.30158 -0.10465 0.00316 0.09007 0.41011   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.39556 0.08153 4.852 9.67e-05 \*\*\*  
## projHOWY 0.02264 0.11530 0.196 0.8463   
## projNASO 0.28091 0.11530 2.436 0.0243 \*   
## projSRRB 0.13089 0.11530 1.135 0.2697   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1997 on 20 degrees of freedom  
## Multiple R-squared: 0.2708, Adjusted R-squared: 0.1614   
## F-statistic: 2.476 on 3 and 20 DF, p-value: 0.09098

#summary.lm returns a linear regression summary of cv ~ proj, including expanded coefficients, std. error, r2, residuals, p and t values. Summary of the anova returns proj as a single variable, the residuals, sum of squares, mean squares, f value and degrees of freedom.   
  
rip.aov.hsd <- TukeyHSD(rip.proj.cv.aov)  
rip.aov.hsd

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = cv ~ proj, data = ripsample)  
##   
## $proj  
## diff lwr upr p adj  
## HOWY -CORP 0.02264174 -0.30007221 0.3453557 0.9972259  
## NASO -CORP 0.28090766 -0.04180629 0.6036216 0.1021690  
## SRRB -CORP 0.13088616 -0.19182779 0.4536001 0.6727620  
## NASO -HOWY 0.25826592 -0.06444803 0.5809799 0.1466194  
## SRRB -HOWY 0.10824442 -0.21446954 0.4309584 0.7845812  
## SRRB -NASO -0.15002150 -0.47273546 0.1726925 0.5728045

#Based on the adjusted p value (last column), there is no signifant difference between sites. From the one-way ANOVA, the predictors do not have a significant impact on the response.

#Two models: First, with single terms (hcm ~ dbhcm), second with interaction (hcm ~ dbhcm \* genus). Looking first at hcm ~ dbhcm by Genus group.  
# + vs \*: + calculates main effects only while \* estimates interactions between factors.  
  
g1 <- ggplot(data = rd\_lab4) + geom\_point(aes(x = Woody\_DBH\_cm, y = Woody\_Height\_cm, color = Genus, shape = ProjCode)) + theme\_minimal() + labs(title = "", x = "DBH, cm", y = "Height, cm") + theme(legend.position = "none",axis.text.x = element\_text(angle = 90, size = 5), axis.text.y = element\_text(size = 5)) + geom\_smooth(aes(x = Woody\_DBH\_cm, y = Woody\_Height\_cm), formula = y~x, method = "lm")  
  
  
#Model 1, simple global model  
glb\_lm <- lm(Woody\_Height\_cm ~ Woody\_DBH\_cm, data = rd\_lab4)  
summary(glb\_lm)

##   
## Call:  
## lm(formula = Woody\_Height\_cm ~ Woody\_DBH\_cm, data = rd\_lab4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1866.92 -245.25 -58.96 205.18 1760.02   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 507.0323 9.5142 53.29 <2e-16 \*\*\*  
## Woody\_DBH\_cm 23.6879 0.3714 63.79 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 389.5 on 3228 degrees of freedom  
## Multiple R-squared: 0.5576, Adjusted R-squared: 0.5575   
## F-statistic: 4069 on 1 and 3228 DF, p-value: < 2.2e-16

anova(glb\_lm)

## Analysis of Variance Table  
##   
## Response: Woody\_Height\_cm  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Woody\_DBH\_cm 1 617404685 617404685 4068.7 < 2.2e-16 \*\*\*  
## Residuals 3228 489830490 151744   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Model 3, global model with interaction terms.  
glb\_lm3 <- lm(Woody\_Height\_cm ~ (Woody\_DBH\_cm + Genus + ProjCode)^2, data = rd\_lab4)  
summary(glb\_lm3)

##   
## Call:  
## lm(formula = Woody\_Height\_cm ~ (Woody\_DBH\_cm + Genus + ProjCode)^2,   
## data = rd\_lab4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1641.3 -196.8 -24.7 185.9 1928.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 309.8524 33.2026 9.332 < 2e-16 \*\*\*  
## Woody\_DBH\_cm 33.8463 1.7298 19.566 < 2e-16 \*\*\*  
## GenusFraxinus -27.7369 40.9243 -0.678 0.497972   
## GenusPopulus 585.0856 37.0798 15.779 < 2e-16 \*\*\*  
## GenusQuercus 104.0769 38.1730 2.726 0.006437 \*\*   
## GenusSalix 205.9142 39.9841 5.150 2.76e-07 \*\*\*  
## ProjCodeHOWY -208.6366 79.0416 -2.640 0.008341 \*\*   
## ProjCodeNASO 0.7029 58.5448 0.012 0.990421   
## ProjCodeSRRB -11.1135 37.8168 -0.294 0.768870   
## Woody\_DBH\_cm:GenusFraxinus -5.3683 2.4249 -2.214 0.026912 \*   
## Woody\_DBH\_cm:GenusPopulus -12.3009 1.6511 -7.450 1.19e-13 \*\*\*  
## Woody\_DBH\_cm:GenusQuercus -11.3566 1.8018 -6.303 3.32e-10 \*\*\*  
## Woody\_DBH\_cm:GenusSalix -7.8775 2.2383 -3.519 0.000438 \*\*\*  
## Woody\_DBH\_cm:ProjCodeHOWY -5.3276 1.4028 -3.798 0.000149 \*\*\*  
## Woody\_DBH\_cm:ProjCodeNASO -5.3086 1.8426 -2.881 0.003991 \*\*   
## Woody\_DBH\_cm:ProjCodeSRRB -2.8559 1.1098 -2.573 0.010118 \*   
## GenusFraxinus:ProjCodeHOWY 204.4685 149.1341 1.371 0.170459   
## GenusPopulus:ProjCodeHOWY -123.1097 86.2630 -1.427 0.153636   
## GenusQuercus:ProjCodeHOWY 459.5071 126.3871 3.636 0.000282 \*\*\*  
## GenusSalix:ProjCodeHOWY 194.2905 84.7012 2.294 0.021864 \*   
## GenusFraxinus:ProjCodeNASO 52.2826 111.2103 0.470 0.638298   
## GenusPopulus:ProjCodeNASO 278.3867 66.5190 4.185 2.93e-05 \*\*\*  
## GenusQuercus:ProjCodeNASO 27.3487 117.7794 0.232 0.816395   
## GenusSalix:ProjCodeNASO -137.0358 77.1731 -1.776 0.075878 .   
## GenusFraxinus:ProjCodeSRRB 76.5335 64.2569 1.191 0.233720   
## GenusPopulus:ProjCodeSRRB 39.2888 68.7138 0.572 0.567515   
## GenusQuercus:ProjCodeSRRB -28.7766 66.1966 -0.435 0.663799   
## GenusSalix:ProjCodeSRRB -71.1037 48.4658 -1.467 0.142450   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 333.4 on 3202 degrees of freedom  
## Multiple R-squared: 0.6786, Adjusted R-squared: 0.6759   
## F-statistic: 250.4 on 27 and 3202 DF, p-value: < 2.2e-16

anova(glb\_lm3)

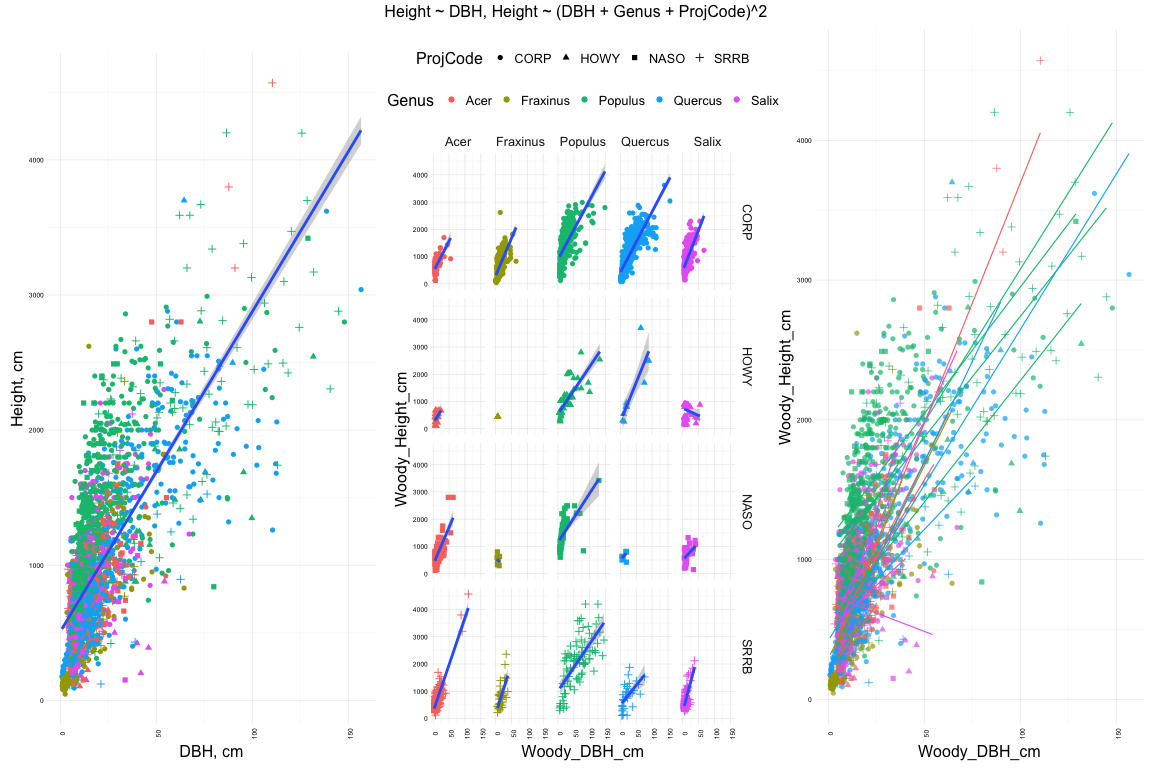
## Analysis of Variance Table  
##   
## Response: Woody\_Height\_cm  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Woody\_DBH\_cm 1 617404685 617404685 5555.9237 < 2.2e-16 \*\*\*  
## Genus 4 97049572 24262393 218.3333 < 2.2e-16 \*\*\*  
## ProjCode 3 12545974 4181991 37.6331 < 2.2e-16 \*\*\*  
## Woody\_DBH\_cm:Genus 4 8327734 2081933 18.7350 3.089e-15 \*\*\*  
## Woody\_DBH\_cm:ProjCode 3 2644977 881659 7.9339 2.861e-05 \*\*\*  
## Genus:ProjCode 12 13438445 1119870 10.0775 < 2.2e-16 \*\*\*  
## Residuals 3202 355823788 111125   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

g2 <- ggplot(data = rd\_lab4, aes(Woody\_DBH\_cm, Woody\_Height\_cm)) + geom\_point(aes(color = Genus, shape = ProjCode)) + geom\_smooth(method = "lm", formula = y ~ x ) + theme\_minimal()+ facet\_grid(ProjCode ~ Genus) + theme(legend.position = "top", legend.direction = "horizontal", axis.text.x = element\_text(angle = 90, size = 5), axis.text.y = element\_text(size = 5))   
  
g3 <-ggplot(data = rd\_lab4, aes(Woody\_DBH\_cm, Woody\_Height\_cm, color = Genus, shape = ProjCode))+ theme\_minimal() + geom\_point(, alpha = .7) + geom\_smooth(method = "lm", formula = y ~ x, se = F, size =.4) + theme(legend.position = "none",axis.text.x = element\_text(angle = 90, size = 5), axis.text.y = element\_text(size = 5))   
  
  
#Model 3 explains the most variation of the three models (R2 = .901). The second model of the single terms indicates that the Fraxinus Genus is a significant predictor; however, model 3 indicates that the Fraxinus and interaction term between Fraxinus and DBH is not significant (pr(>|t|) == .6154 and .3366, respectively).

**Results:**

Based upon our random sample of plot by project area, we found tree height does not significantly vary based upon plot within project site (p value = .297, R2 = .09, F-value = 1.315). Our initial model based upon untransformed values of Ht~ DBH (from lab 3) explained approximately 76% of the variance if a zero intercept were forced. We increased our R2 in that case by log10 transforming variables. We did not transform values for this report or force a zero intercept. The second model relating HT ~ DBH with genus as a factor and including interaction terms explained the majority of variance compared to the initial global model including single terms (R2 = 56 and.68 respectively). Analysis of interactions of covariates in the second model also indicates that several interactions are superfluous and can be removed to increase model fit (see HOWY:Salix, HOWY:Fraxinus, NASO:Quercus, and NASO:Fraxinus in Fig 1).

grid.arrange(g1,g2, g3, ncol = 3, top = "Height ~ DBH, Height ~ (DBH + Genus + ProjCode)^2" )



**Discussion:**

Our finding that project site does not have a significant effect on the relationship between height and DBH suggests that we use either height or DBH to estimate the effects of environmental conditions, rather than modeled biomass. The relationship is dependent upon genus, as fig 1 shows. Several terms can be removed from the model, including Fraxinus and several interactions with Salix. Certain interactions with Salix have a negative trend, suggesting that there is an issue with not forcing a zero intercept.

**Limitations:**

Our subset of 5 genera limits our conclusions, especially as the counts of the 6th most frequent genus was close to the 5th. It is unlikely that we would see site playing a different role on the height~dbh relationship in the other genera; however, a full report should address that. Finally, the relationships with several interactions indicate that zero intercepts must be forced to accurately depict the relationships between DBH and height