# NestZModels

In this analysis the slopes obtained by fitting SAC to each site by a ln/ln lme model is explored. The slopes were obtained using ln/ln transpose of species and area across the nests ($4m^{2}$ to $200m^{2}$). Slope was constant and intercept random with plot as the random variable. The slopes were approximately normally distributed about .25. We want to examine whether any of the abiotic variables collected can explain any of the variation in the z values.

The mixed effects model has the effect of averaging the alpha diversity of each plot across the site(?), so you could expect that the slope would be proportional to the average of abiotic site variables.

(I have already removed several variables for clarity, which I saw had no effect, altitude, area, perimeter. I also decided to restrict this analysis to the mean variables, and not look first at the standard deviations. I think this is OK, because we suspect that this z will be related to means - and when I included it the whole story was getting epic. Since I am repeating this for 5 other variables, I feel like I need to condense things a bit. And if I say that this variable could relate to the means, then it makes sense to related the standard dev of intercepts to the standard deviations?)
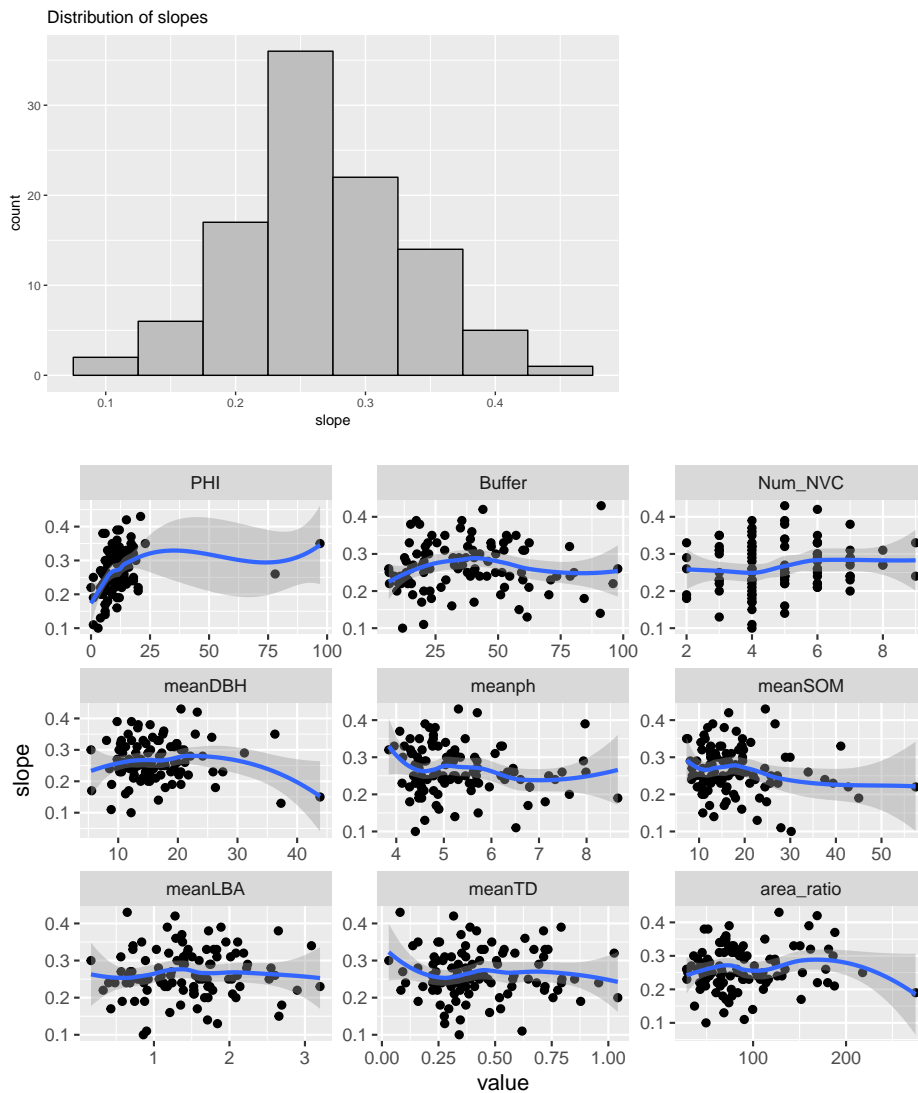
```
##           corrs
## PHI        0.38
## Num_NVC    0.21
## meanph     0.21
## Buffer     0.20
## meanSOM    0.19
## meanDBH    0.16
## area_ratio 0.16
## meanLBA    0.14
## meanTD     0.12
```

The scatter plots show the slope with the variables collected. There is a lot of scatter, and no obvious relationship. The two variables with the lowest correlation coefficients (using dcor distance correlation) to slope are meanLBA and meanTD,(Corr = 0.14,0.12)

The outliers in PHI, area_ratio make it hard to see the response of slope to these variable. Initially the outliers were removed, but the distribution of some variables, eg sd_SOM was very right skew, and removing outliers did not rectify this, therefore the data is replotted with a log transpose to the x axis
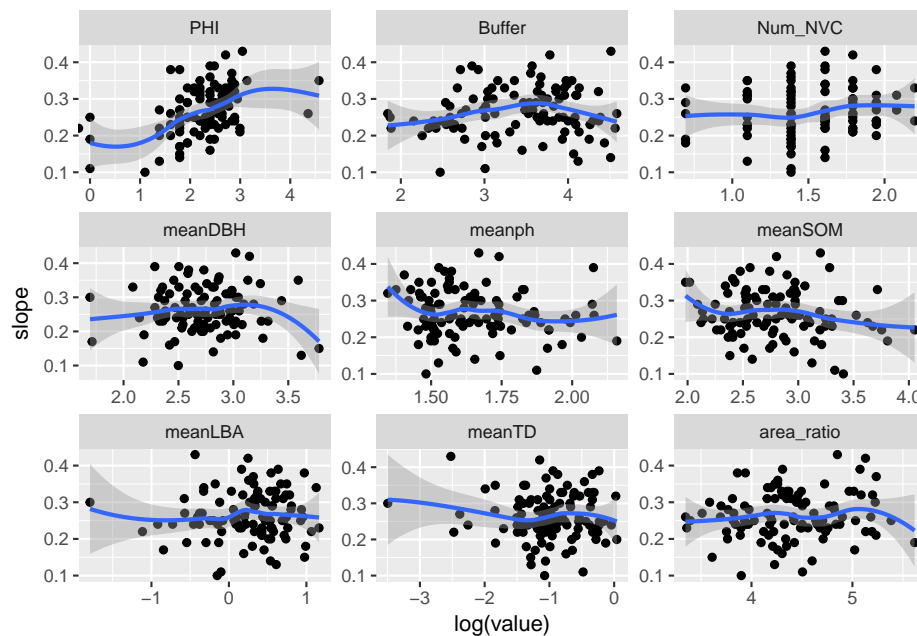

```
##           slopecor
## PHI          0.42
## Buffer       0.21
## Num_NVC      0.20
## meanph       0.20
## meanSOM      0.18
## area_ratio   0.16
## meanDBH      0.15
## meanLBA      0.15
## meanTD       0.13
```

The strongest correlation is still with PHI.

pearson correlation

kendal correlation
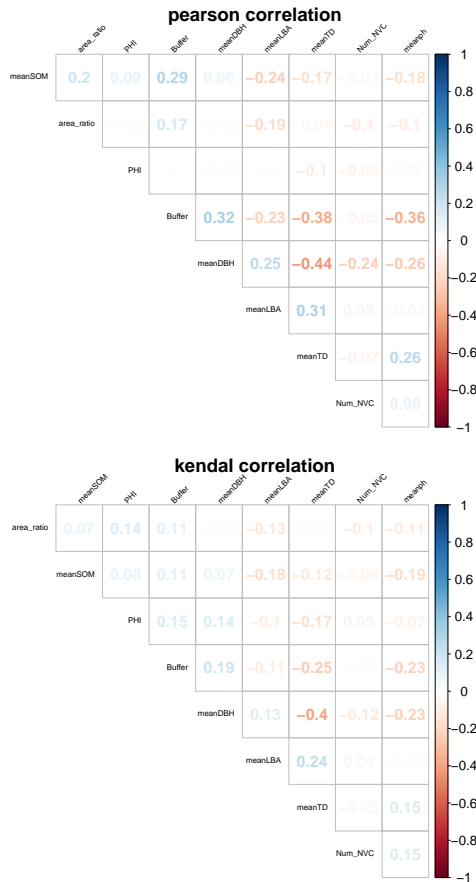
Kendall correlation was used as well as Pearson due to the outliers and the non-linearity of the data. The Pearson method show a network of correlations above/below 0.3/-.03 between buffer, meanDBH,meanTD,meanpH,meanLBA. This is not reflected in the Kendal correlation, which is -0.4 for meanTD and meanDBH. It is obvious that meanLBA, meanTD and meanDBH are correlated. The relationship with meanpH could be through additional litter dropped by large trees. Additionally, we have already seen in previous analysis the the northern woods have larger buffers, are less dense and more mature. This could explain the correlation with meanpH to these variables. It seems sensible to eliminate two from meanDBH, meanTD and meanLBA, but first the dedge method of MuMIn will be used to look at all models.

Models

```
## Fixed term is "(Intercept)"
## $`434`
##
## Call:
## lm(formula = slope ~ area_ratio + meanph + meanSOM + Num_NVC +
##     PHI + 1, data = data, na.action = "na.fail")
##
## Coefficients:
## (Intercept)    area_ratio        meanph       meanSOM       Num_NVC
##   0.2854521     0.0001995    -0.0121737    -0.0018080     0.0079558
##         PHI
##   0.0017081
```

```
## 
## 
## $`433`
## 
## Call:
## lm(formula = slope ~ meanph + meanSOM + Num_NVC + PHI + 1, data = data,
##     na.action = "na.fail")
## 
## Coefficients:
## (Intercept)      meanph       meanSOM       Num_NVC           PHI
##    0.306006    -0.012730     -0.001613      0.007396      0.001679
## 
## 
## attr(,"rank")
## function (x)
## do.call("rank", list(x))
## <environment: 0x0000000016c27d58>
## attr(,"rank")attr(,"call")
## AICc(x)
## attr(,"rank")attr(,"class")
## [1] "function"     "rankFunction"
## attr(,"beta")
## [1] "none"

## Global model call: lm(formula = slope ~ ., data = data, na.action = "na.fail")
## ---
## Model selection table
##       (Int)    are_rat      mnp        mSO      mTD  Num_NVC       PHI df
## 434 0.2855 0.0001995 -0.01217 -0.001808           0.007956 0.001708  7
## 433 0.3060             -0.01273 -0.001613           0.007396 0.001679  6
## 305 0.3385             -0.01185 -0.001637                    0.001618  5
## 306 0.3226 0.0001748 -0.01131 -0.001810                    0.001639  6
## 418 0.2183 0.0002161           -0.001569           0.007372 0.001668  6
## 498 0.2812 0.0002001 -0.01289 -0.001768 0.01388 0.008164 0.001734  8
##       logLik    AICc delta weight
## 434 146.079 -277.0   0.00  0.284
## 433 144.872 -276.9   0.11  0.269
## 305 143.072 -275.5   1.45  0.137
## 306 143.971 -275.1   1.91  0.109
## 418 143.904 -274.9   2.05  0.102
## 498 146.201 -274.9   2.11  0.099
## Models ranked by AICc(x)
```

The "best" two models use number NVC, meanpH, meanSOM, PHI and either do or do not include area_ratio. Neither use buffer, meanLBA or meanTD so this potential confounding network has been removed.

We expect meanSOM to effect pH, and that pH effects richness, with greatest richness occurring around neutral pH - so the inclusion of both terms seems unnecessary.
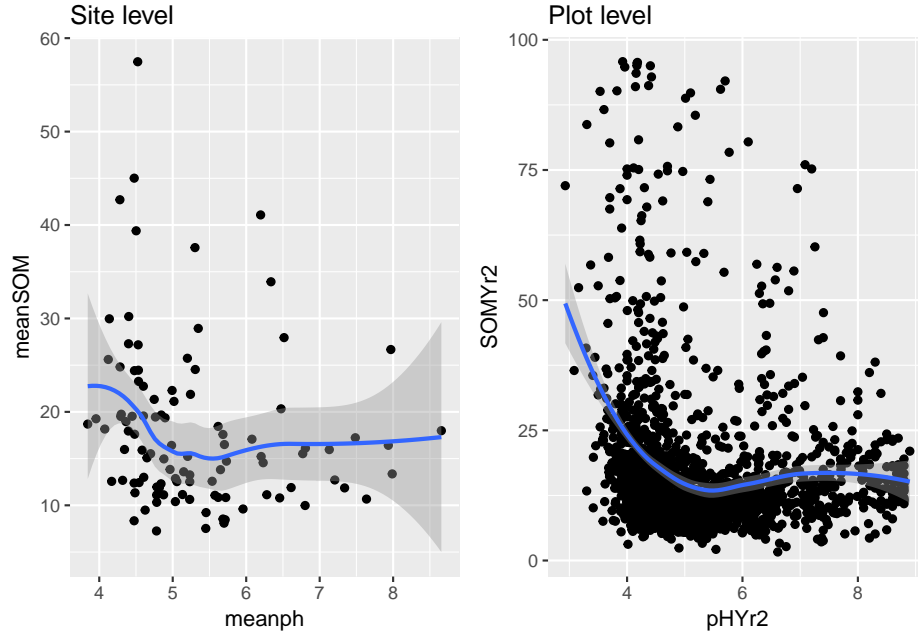
The plots suggest that in these sites meanSOM is not strongly correlated with meanpH at Site level, but at plot level there is a negative correlation below pH5, However, there is a lot of scatter.
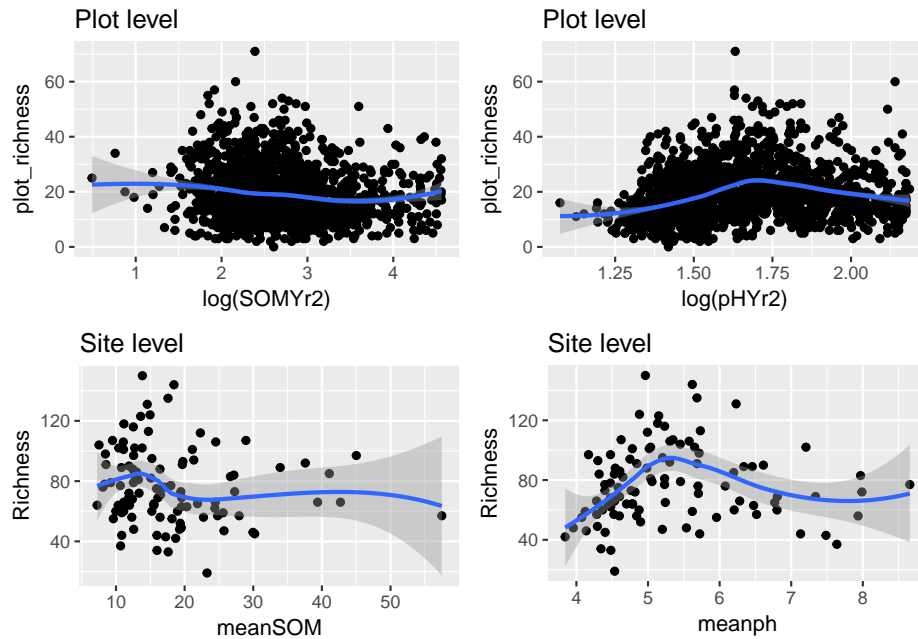

The richness at site and plot level does not have a strong linear correlation with pH, but there is the expected unimodal peak, which is stronger at site level. This would suggest that pH is a strong influencing factor on richness, and thereby z, but its effect is not linear. Therefore the coefficient of the linear term in the model could be expected to be small and less significant than a non-linear term, and its effect on z when averaged across the site will be reduced.

The richness at site or plot level does not appear to correlate strongly with SOM, and its effect on the slope would therefore be small

These plots suggest that meanpH would be a better term for predicting richness and therefore slope, than

meanSOM but that it should be non-linear.
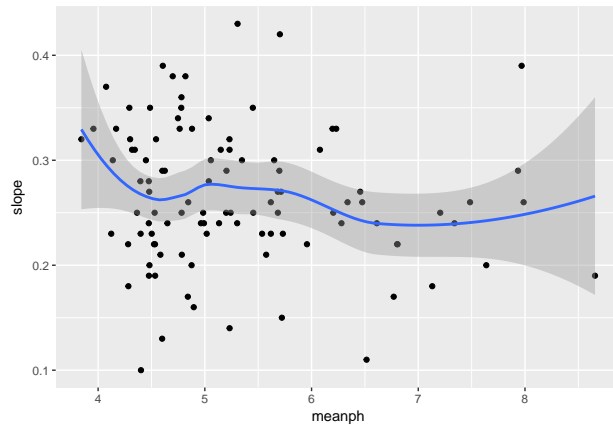
However, when we look at the plot of slope with mean pH, although there is the corresponding peak at 5.5, the curve is not unimodal,and therefore using it in the model as either a unimodal or linear term will not give a strong relationship to the actual value of the response.
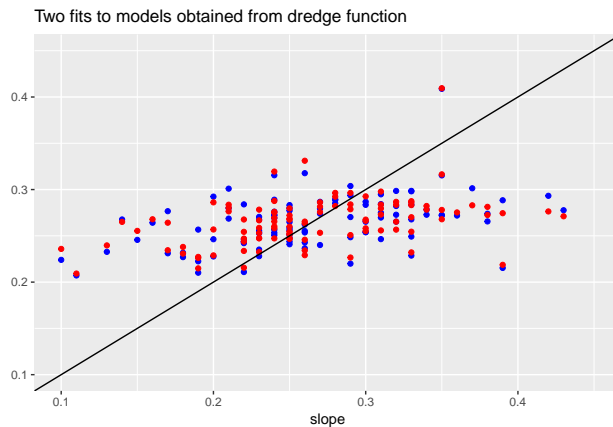
The plot of predicted slope values versus actual slope values using the variables predicted by the dredge function show that these variables do not predict the value of z successfully. The first uses slope ~ area_ratio + meanph + meanSOM + Num_NVC + PHI, the second does not use area_ratio. In both models the coefficients are significant at 5% (except area ratio in the first 13% and Num NVC in the second 6%). In both models R2 is below 0.2, as suggested by the plot. In both models PHI was the most significant coefficient.

This analysis has shown that these variables are not good predictors of the z value obtained from a ln/ln linear mixed effects model to the SAC of each site where plot was used as the random variable. Although we saw that plot was a strong predictor of richness, often equal to or greater than area (previous analysis). Therefore the standard deviations of the intercepts will be considered next since these reflect the plot to plot variation.....

...TBC!!