

Analysing the effect of abiotic factors on species richness

Petra Guy

9 May 2018

Overview:

The modelling structure, based on Burnham et al, has been to select the variables from the data set that have been shown in the literature to have an effect on species richness and then transformed only if there is a predetermined reason for doing so. In our case richness is known to have a unimodal response to pH and therefore this variable was transformed by taking the square. Correlations between covariates are considered which resulted in the data being split into two subsets. The data is then standardized, [Gelman] and fitted to a linear model. The variance inflation factors of the model parameters are considered, and the fit of the linear model is discussed using residual vs fitted values and QQ plots. The top model set was selected using a delta < 2 and these models were averaged. The parameter estimates with 95% confidence intervals were calculated from the adjusted standard errors and are plotted with variable importance. Finally a predicted richness is calculated and the r2 value of this when fitted to the empirical richness is shown.

Variables selected.

Northing, Species richness decreases with increasing latitude [Gaston,1996]

Positive Heterogeneity Index, "PHI". This is the sum of counts of site features which indicate habitat heterogeneity, but specifically where that heterogeneity has been used in the literature and shown to correlate positively with plant species richness. [Boch used logging trails as indicator, Schmidt, Paillet show forest management increases richness, Hannay uses length of rides, Philips shows mean annual tree mortality increases richness]

Buffer, This could impact richness by increasing habitat connectivity leading to a species pool for the site. . . need refs. . .

Number of major soil groups, no_MSG,[Hannay uses soil types]

number of NVC codes, no_NVC, This is another indicator of heterogeneity. Although there is the circular argument that NVC reflects richness because it describes a species assembly, here we use number of different NVC codes. this variable is therefore an expression of the heterogeneity of the site.

Soil pH, sd_pH,meanph, [Shen, unimodal response - need ref, but therefore meanph^2 was used]

Soil organic matter, sd_SOM,meanSOM,[Boch,]. Note Boch also discuss effect of beech tree soil litter and presence of beech.

Diameter at breast height, sd_meandbh,meandbh,[]

Tree density, sd_TD,meanTD [Shen]

the perimeter divided by the area, area_ratio,[Hanny, shape index]

Notes on data handling.

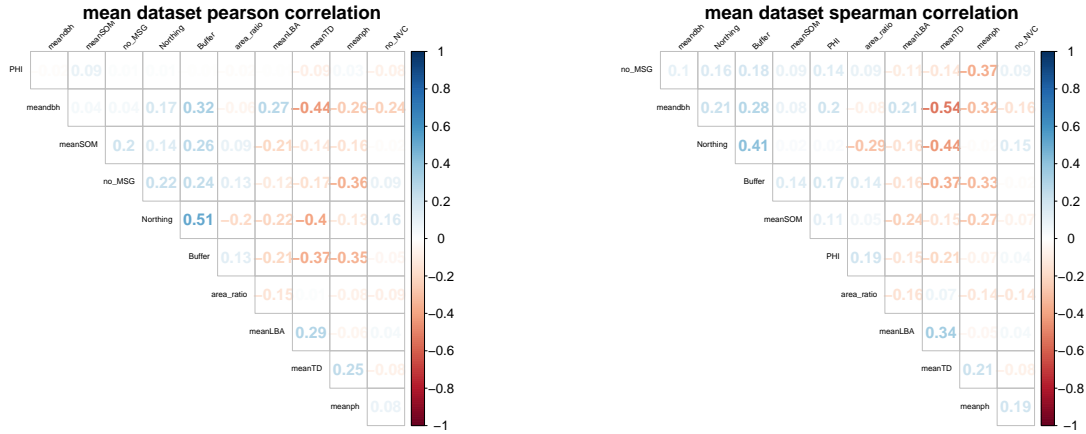
One site had no recorded site descriptors leading to PHI of NA - this has been mean imputed

All woods had areas below 100 hectares,except one, which had an area of 311 ha.This was removed

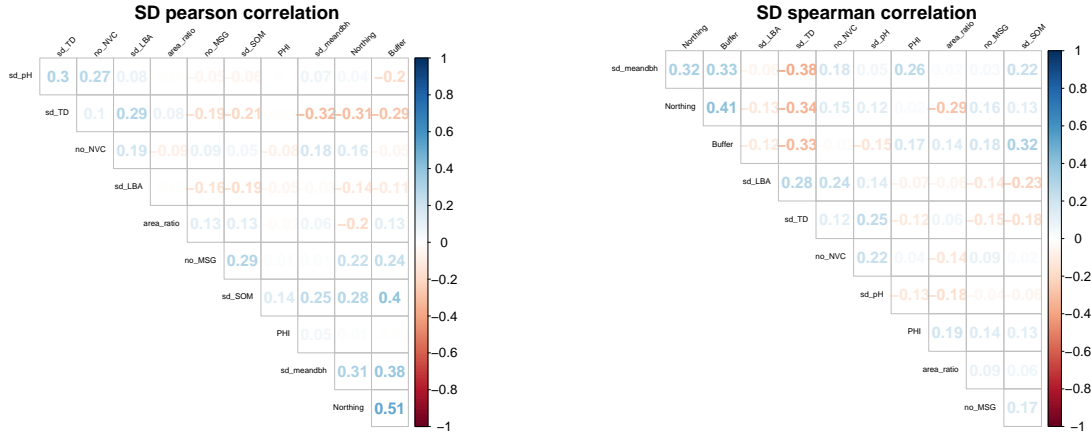
Spearman correlations of mean and sd of tree density, live basal area, SOM and DBH are above 0.66. The variables will therefore be split into two groups. One group containing the meanTD, meanLBA, meanSOM

and meanDBH along with all other variables, and the other containing the standard deviation of these factors along with all other variables. Although meanpH and sd_pH only show Spearman correlation of 0.46 it makes sense to include these variables in this split as the mean describes a site level condition whereas the SD describes a site heterogeneity, therefore we are dividing the data into two subsets, one which is more slanted to considering overall abiotic properties of the site and one which is more focused on looking at heterogeneity within the site.

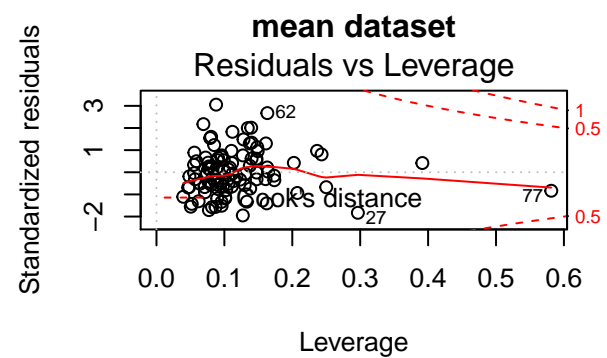
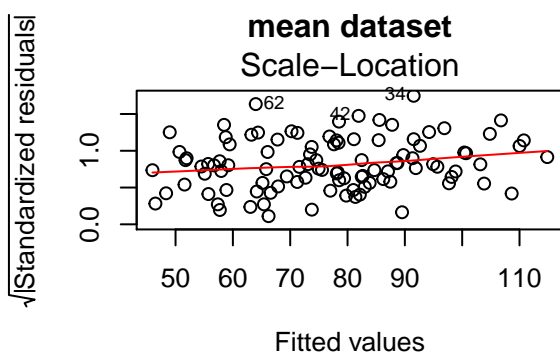
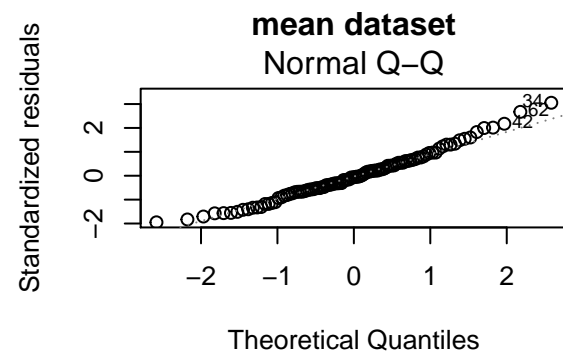
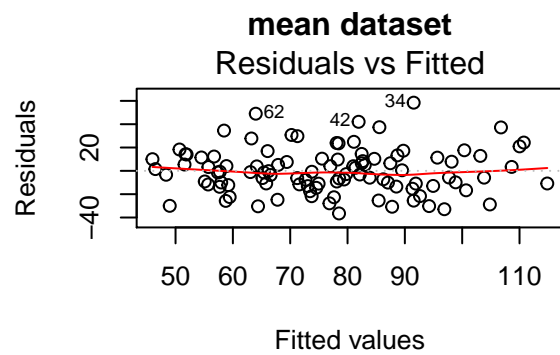
Correlations between covariates after creating two sets



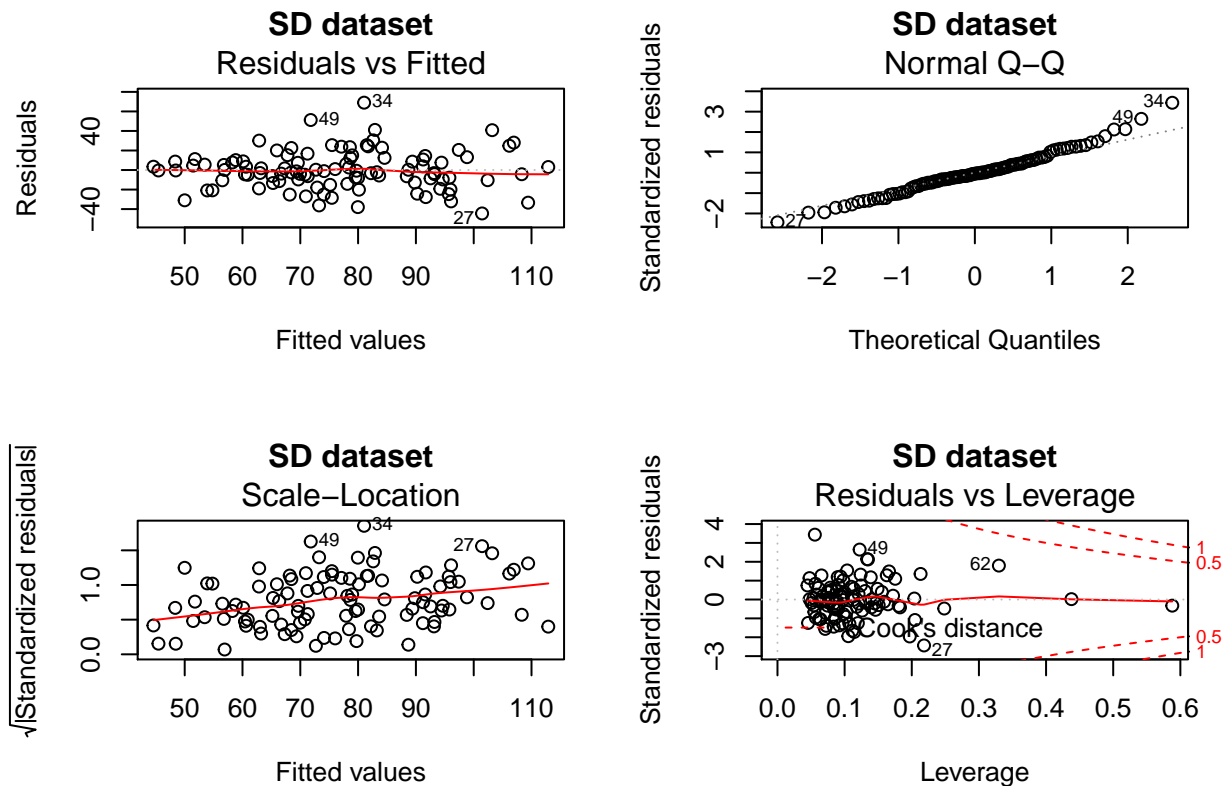
Correlations all < 0.5 , except Pearson for Nothing and buffer = 0.51 and Spearman for meanTD and meanDBH = -0.55, however correlations at this level are unlikely to effect the parameter estimates [Freckleton]



Correlations all < 0.51



The residuals versus fitted values suggest the variance is homogeneous. The QQ plot shows that the residual distribution is approximately normal. There are 3 points which represent the sites with the highest richness which fall slightly above the line.



The residuals versus fitted values suggest the variance is homogeneous. The QQ plot shows that the residual distribution is approximately normal. There are 2 points which represent the sites with the highest richness which fall slightly above the line.

```
## Northing      PHI      meandbh      meanph      Buffer      meanSOM
## 1.722833      1.048575  2.002343  1.357780  1.791938  1.136126
## meanLBA      meanTD area_ratio      no_NVC      no_MSG
## 1.695391      1.949286  1.188994  1.264155  1.250528
```

The variance inflation factors in the mean dataset are low, suggesting that correlations between covariates are low and not likely to increase the variance of the parameter estimates.

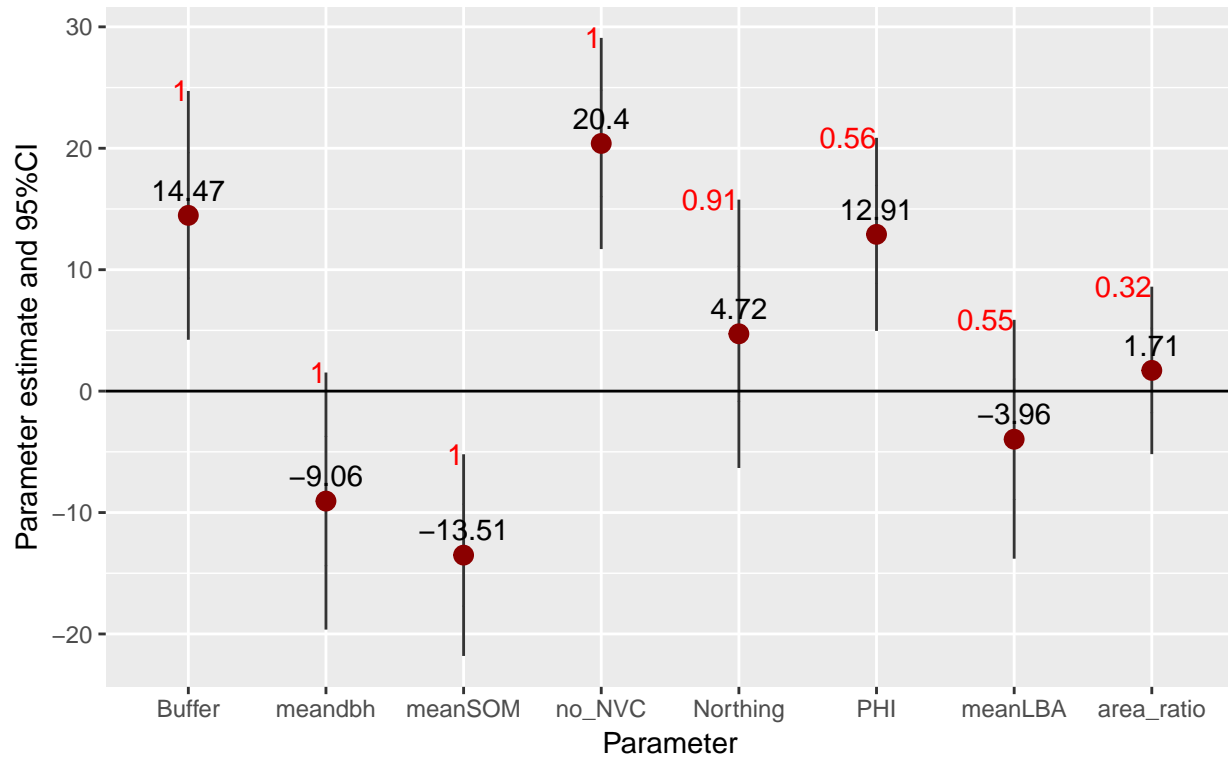
```
## Northing      PHI      Buffer      no_MSG      no_NVC      sd_pH
## 1.714511      1.042396  1.832215  1.218467  1.251549  1.286635
## sd_SOM      sd_LBA sd_meandbh      sd_TD area_ratio
## 1.349899      1.180705  1.443288  1.503652  1.192361
```

The variance inflation factors in the SD dataset are low, suggesting that correlations between covariates are low and not likely to increase the variance of the parameter estimates.

The first four models in the mean dataset, which had a delta < 2 were selected from the MuMin dredge function as the top model set. (PS a delta less than 3 was also checked, but this resulted in additional effects which all had CI which included zero). Model averaging was achieved using model.avg function of MuMin using the subset call. this means that the zero method was used. Feckleton recommends this when the purpose is to see which factors have the strongest effect on the response.

Model averaged results for delta <2, mean dataset

Numbers in red are importance



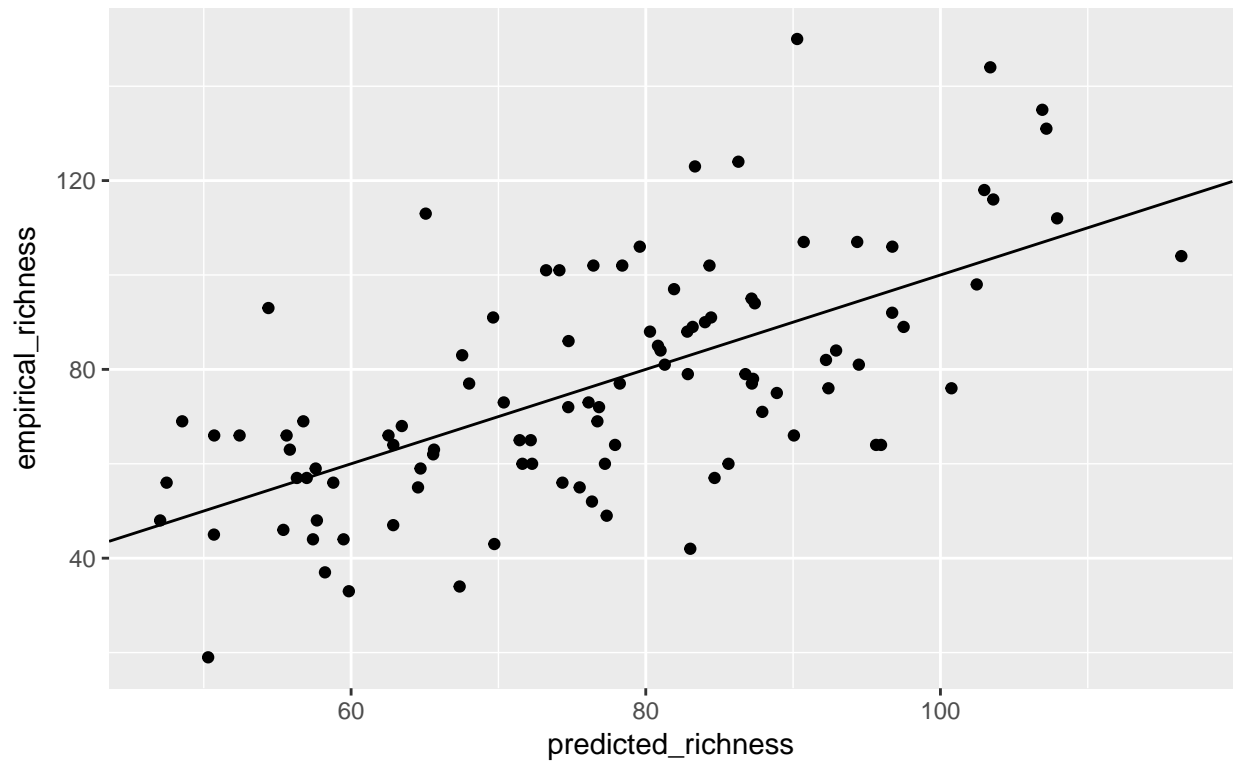
The graph shows the averaged effect sizes of the model with delta < 2. The effects which have a CI which does not include zero and therefore may influence richness [Gruber] are buffer, mean soil organic matter, number of NVC codes, and Positive Heterogeneity Index. Because the data is standardized, the effects sizes can be directly compared. Therefore the number of NVC codes can be seen to have the greatest effect, with SOM and PHI having approximately equal but opposite effects.

The first seven models from the SD dataset, which had a delta <1.99 were selected from the MuMin dredge function as the top model set. (PS a delta less than 2 was also checked, this resulted in the effect of sd_LBA with importance of 0.08 and had CI which included zero).

Using the model for prediction

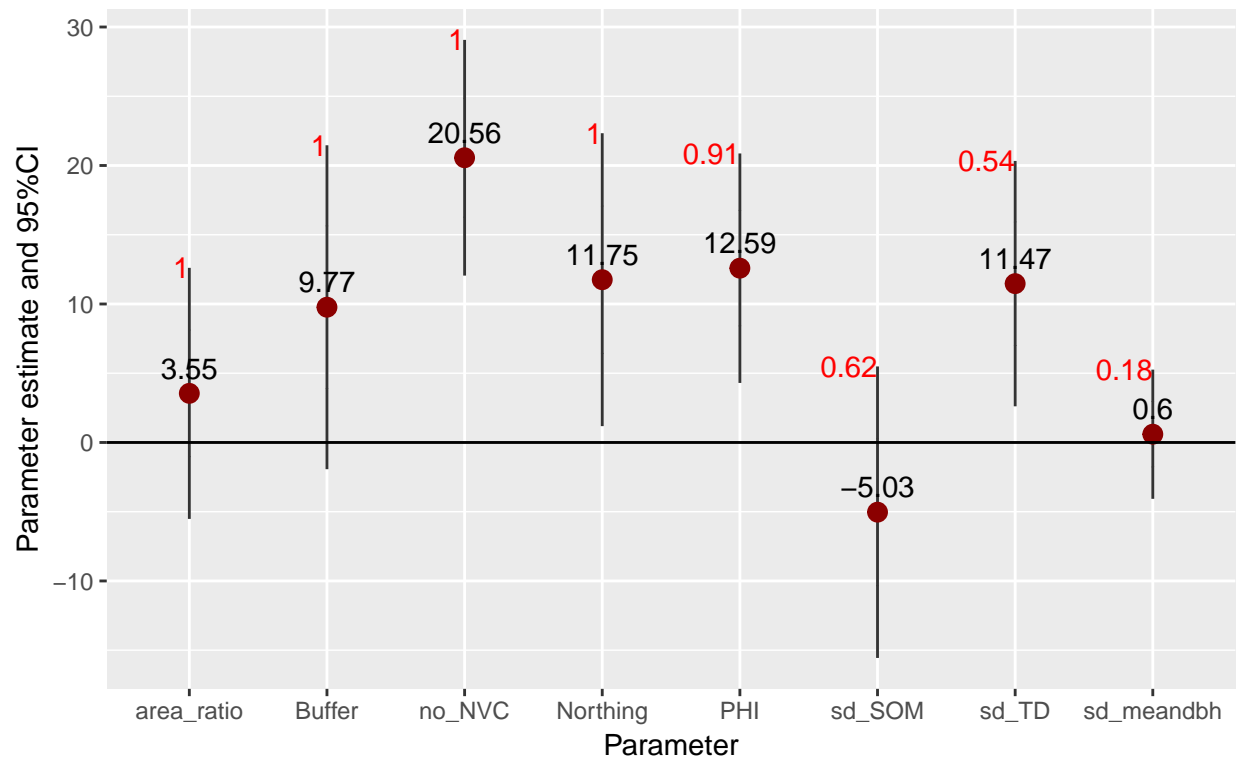
Observed versus predicted data, mean dataset

$R^2 = 0.42$



Model averaged results for delta <1.99 for SD data subset

Numbers in red are importance



The graph shows the averaged effect sizes of the model with delta < 1.99. The effect which have a CI which does not include zero are again number of NVC codes, and Positive Heterogeneity Index. in addition Northing and standard deviation of tree density have an effect on richness in this model.

Using the model for prediction

Observed versus predicted data, SD dataset

$R^2 = 0.39$

