# More Zeta Modelling

*Petra Guy*

*26 May 2018*

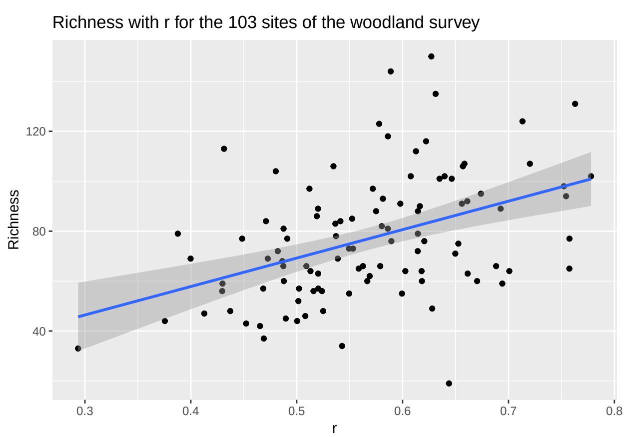
## A coefficient that might be useful??

I suggest: zeta1 and zeta2 are the average alpha and beta diversity of a site - the difference between them tells you something about the heterogeneity of the woodland.

The gradient of the zeta decline curve indicates the heterogeneity of a site. For a perfectly homogenous site, where all species occur in all plots, all the zeta values are the same and the gradient of the zeta decline curve is zero. For a perfectly heterogenous woodland, each plot has a different set of species, therefore every zeta from order 2 onwards is zero and the gradient of the zeta decline curve is zeta1/1 = zeta1.

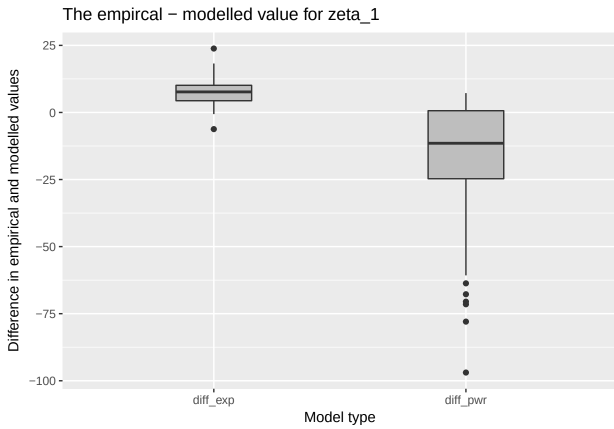
In reality, the site will have an initial gradient somewhere between, but zeta1 - zeta2 could be used as a measure of the heterogeneity of the wood. Although only using zeta1 and zeta2 omits higher orders of zeta, and therefore reduces our knowledge of the heterogeneity on a larger scale, it does include all plots within a site, and therefore considers the site as a whole. This is because the zeta diversity calculation uses every plot to calculate zeta2. The problem is that this value is not comparable between woodlands because richer woodlands will have a higher zeta1. Therefore consider the ratio r = (zeta1-zeta2)/zeta1. This varies between 1, for a perfectly heterogeneous woodland and 0 for a perfectly homogeneous woodland - regardless of the richness.

Then, if richness is increased with heterogeneity, you will see a positive correlation between site richness's and the coefficient r.




## Notes on zeta modelling and additional

When we tried to fit the exponential or power law to the empirical zeta decay we found a poor fit. Most of the error was in the fitted value for zeta1. The power law tended to overestimate this while the exponential fit underestimated it.

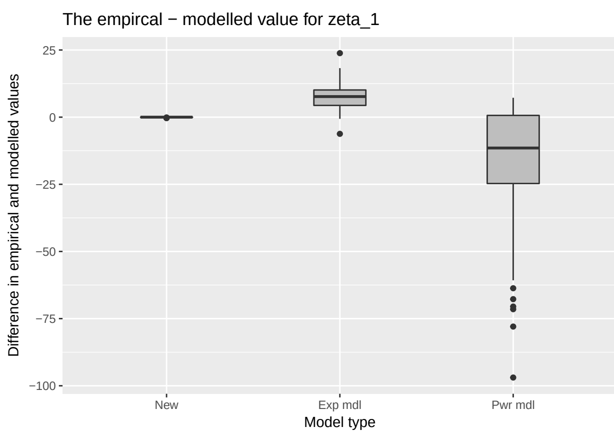



This mean that when checking the accuracy of the model by using it to predict the species richness of our 16 plots - the value was either much too low or much too high.

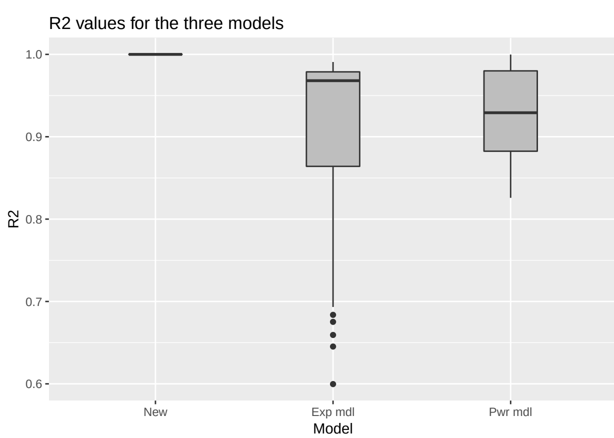
CP suggested a fit of exp(xˆ-b). This has the problem that it does not decay to zero - as our zeta values often do. But, exp(xˆ-b)*exp(-ax) does decay to zero due to the second exponential term. The first exponential term is just e for zeta = 1, and therefore with a coefficient multiplier, this expression could be used to increase the estimated value of zeta1 whilst ensuring the values can decay to zero. So I tried another model:

zeta = c*exp(xˆ-a)exp(-bx).

The model was fitted using the nls function.

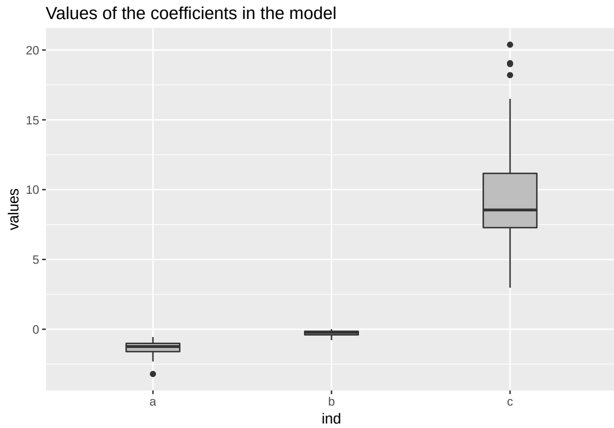



The new model is correctly estimating zeta1.



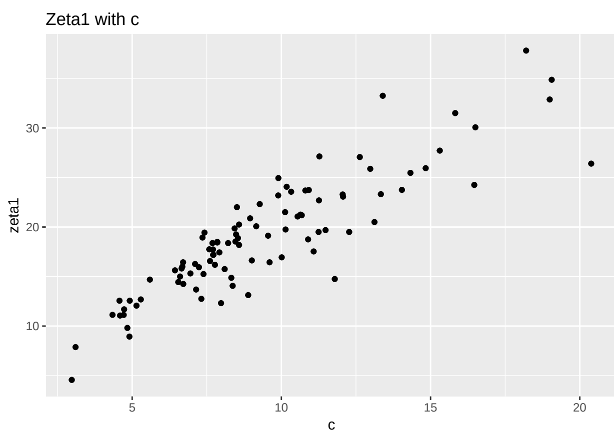
The R2 suggests the model is good at estimating the values of zeta.

What do the coefficients look like?

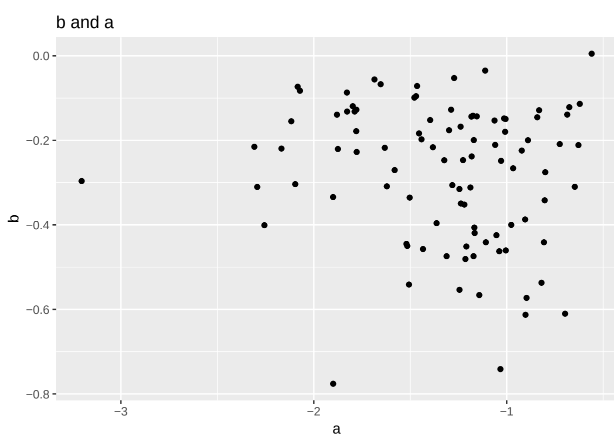



a and b are very similar across all the models. the upper and lower quartiles for a are -1.6<a<1.01 and for b, -0.4<b<-0.14. (PS - I include the negative here but the model has a negative value for a and b, I dont mean that you substitute -1.6 into exp(xˆ-b))

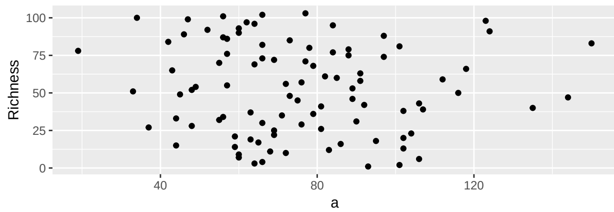
The upper and lower quartiles of c are such that 7.28<c<11.16. This large value for c is probably what allows the fit to correct estimate zeta1. For zeta1, exp(xˆ-a) is 1, exp(-bx) varies between 0.6 and 0.86, therefore by fitting an appropriate value for c, the model can easily result in accurate estimates.

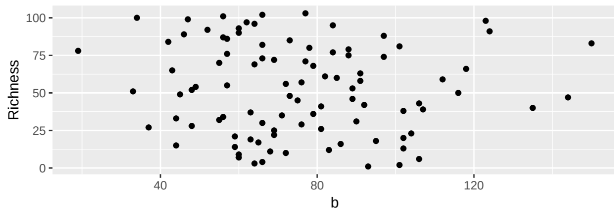



The graph shows that c increases with zeta1, such that an accurate fit will be achieved.



There doesn't appear to be any relationship between b and a. Are a or b related to richness??




Not really.

So it seems that the model zeta = cexp(xˆ-a)exp(-bx) fits the data well, with the values of a and b being very similar across all sites. (PS - 8 sites were omitted because the nls algorithm did not converge using the same starting values that were used on all the other sites). But - what does this new model mean? The power law and the exponential law had the advantage of interpretable ecological meaning; exponential model implying that the probability of occupancy across plots is equal (more homogeneous site??), whereas the power model implies a niche differentiation process (more heterogeneous site??).

But what does this new model represent? And does it only fit because it has 3 parameters instead of 2?