Richness Extrapolation

Petra Guy

2 June 2018

the summary so far of the zeta modeling is this:

The new model, $\zeta = c\exp(x^{-a})(\exp(-bx))$, is a good fit for the zeta values of 95 out of 103 sites. I.e. the non linear least squares function converges given the same starting values for 95 sites. The predicted zeta values from the model are very close to the observed values (R2 is above 0.975 for all modelled zeta decay curves). Using the modelled zetas to predict richness for the 16 plots gives three quarters of the predicted values accurate to at least 25% (where accuracy is (maximum predicted richness - minimum predicted richness)/observed richness. The maximum and minimum richness are calculated using the coefficients for the model +/- standard error, PS - need to change to 1.96 x se).

BUT, there are issues with the richness extrapolation. The richness calculation is a series sum involving the zetas and the number of combinations of sites for each zeta order. Richness = $\Sigma(-1)^{k+1}\ {}^{n}C_k\ \zeta_k$
???

So for a large site, where you could have had say 100 plots, you get nCr values many orders of magnitude larger than the zeta values. I think this results in floating point errors so that the series sum cannot lways give the right answers and large negative richness are seen. So I think that extrapolating to large orders of zeta needs to be avoided. One way to avoind it is to consider what it means if empirical zeta values are already zero at order 16

If the empirical values of zeta are already zero at order 16, (PS, we only have maximum ζ_{16} for empirical values because we have 16 plots) Then the model - if "perfect", would give a value of zero for all zetas of order 16 and above. The only reason it doesn't is that we are approximating the zeta decay to a function that can never actually reach zero. Therefore, all zeta orders used in the extrapolation above order 16 are wrong, they should be zero. It may seem that if more plots had been used, 20 rather than 16 say, then all the zeta values would change, and higher order zetas might exist. But this is unlikely. Consider adding a new plot to the already surveyed set of 16. If the new plot included many more species than any other plot, ζ_1 would change and the modelled richness would be greater. At the highest order, there would still be no shared species, because we know that at least 16 plots do not share any. Also, consider the environment we are studying, UK woodlands. Even in a very large woodland, what is the chance that a new plot will be very distint from 16 plots already surveyed. If this is thought to be likely, then perhaps this is an argument for not having totally randomised plot placement. Rather, plots should be positoined in order to ensure all habitats within the wood are surveyed.

So this means that if the above equation was used to extrapolate richness above plot 16 with the "actual" zeta values for order 17 and up - i.e. zero, - all subsequesnt terms in the series are zero. The richness for the 16 plots is then the total richness for the site. So that means you dont need to extrapolate. That must imply that if you look at the highest orde zeta values of your site - if those zeta values are zero, the species area curve is saturated. Any additional species added to the curve are due to rare species, which, by definition, there wont be many of.

This should mean that if we look at the species area curves for the woodlands with ζ_{16} equal to zero, we should see that the curve looks saturated.

Secondly **If the empirical values of zeta are NOT zero for the highest orders** then more terms would be added to the summation used to cacualte richness, the species area curve is not saturated. But, we know that because our model does not produce zero values, ever so we might produce nonsense results. (I.e using to many orders of zeta). So how far is enough? If you have a site where th highest empirical zeta values are above zero, this is suggesting that there are common species that are occuring in all plots. E.g bracken or brambles. It is also likely that if more plots were added these species, being ubiquitous, would

occur in those plots also. If a new species occured in a new plot, then this would not change the values of the higher orders of zeta - because this species is not shared between all plots. So, my suggestion is, that whatever is the highest order of zeta, use that value in the extrapolation from order 16 onwards.