Different approaches to scaling SAD and the logseries

A. J. Rominger

07 July 2017

When I look at how well the Fisher log-series works across scales, it seems as though spatial subsets are not different from random subsets. I come to this conclusion by comparing the z^2-value (which measures goodness of fit) for spatial and random subsets of the BCI and PASO datasets (Fig. 1). I define a spatial subset as all the stems within a specified area. I define a comparable random subset as a random sample of the same number of stems taken from across the plot.

I find this result surprising, as spatial and random subsets certainly have different shaped SADs. We can see this just by comparing the rare tail. In Figure 2 I compare the number of singletons in spatial and random subsets.

Clearly, random subsets diverge from spatial subsets at intermediate scale, where we see more rarity (more singletons) in random subsets. Interestingly, this pattern is very similar across the two datasets.

So how could the SADs look different, but their fit to the Fisher log series be the same? From Figure 3 it seems that across scales the log-series first accurately predicts the number of singletons at small scale (for both spatial and random subsets) but then at large scale, the log-series over-predicts the number of singletons—but importantly this over-prediction is again the same for both spatial and random subsets. So while the log-series does indeed fail at large scale, it fails the same for both spatial and random subsets.

So we're left to ask, is space what matters for the log-series, or is it just sample size. It would seem to be just sample size. That's a little disappointing given that we had cool hypotheses about local dispersal limitation facilitating a shift from null/neutral processes at small scale to other ecological/evolutionary processes at large scale. But we can still pick up a signal of dispersal limitation by looking at spatial aggregation. The short story is that most species' spatial distributions are very well described by a negative binomial across scales for spatial subsets, and not surprisingly random subsets look like Poisson distributions. I evaluate that by looking at the difference between the log likelihoods of the negative binomial versus the Poisson for spatial and random subsets (Fig. 4). Species' spatial distributions are computed as the number of individuals of each species in each random sample—either in a spatial grid, or from random samples of the same size taken arbitrarily across the plot.
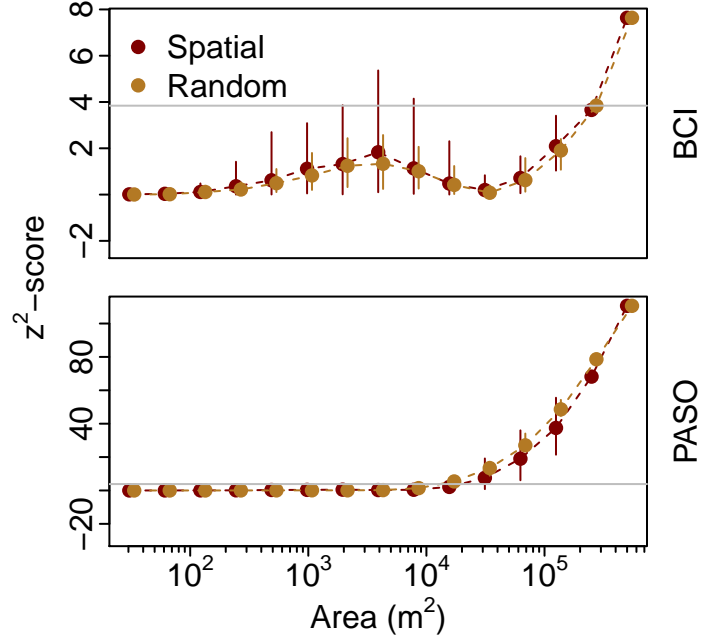
Figure 1: Relationship between z^2 value and scale. Gray dashed line corresponds to hypothesis test critical value, with points above rejecting the log-series at $\alpha = 0.05$
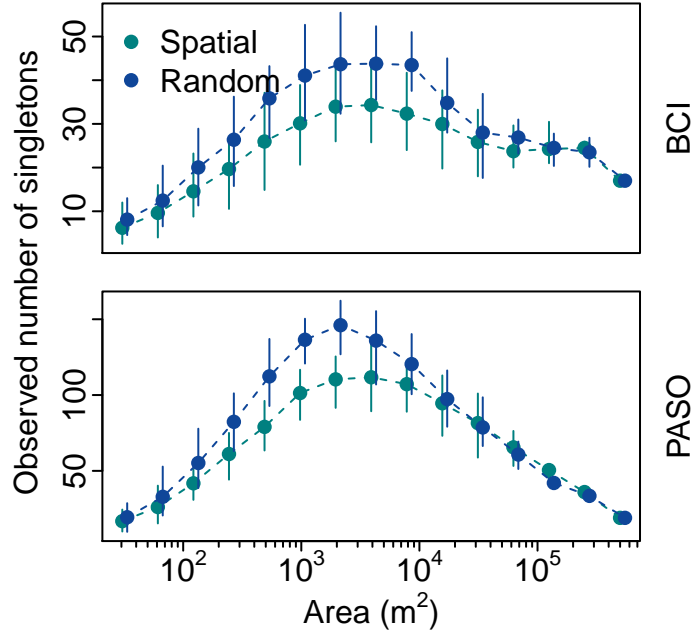

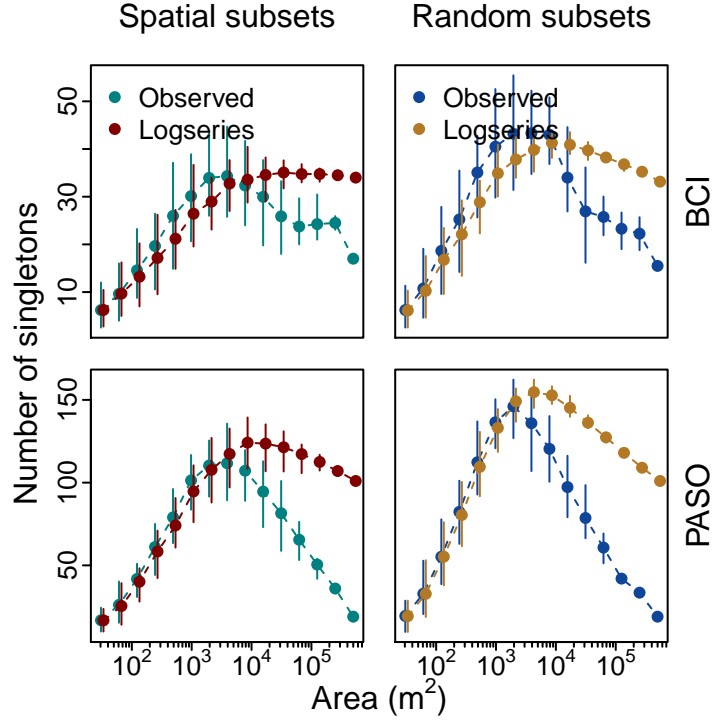Figure 2: Relationship between number of singletons and scale.

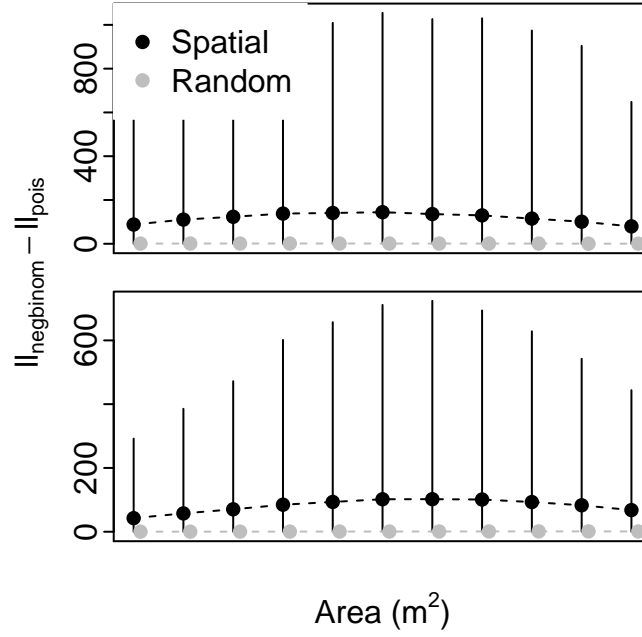Figure 3: Relationship between predicted and observed number of singletons versus scale.

Figure 4: Scaling of nagative binomial fit. When the points are greater than 0, the negative binomial is favored; when the points equal 0, there is no difference between negative binomial and Poisson.