Some thoughts on the formulation of Nee Soon IBM

June 9, 2021

Let's write down the Nee Soon individual-based model (IBM) mathematically since we will inevitably need to do so in the manuscript.

1 Transitional diagram

[To do later]

2 Fruiting incidence

For tree individual i of species j with diameter-at-breast-height, D_{ij}, in a spatial grid p with habitat index, H_p, at time point t, we model its fruiting incidence [yes or no] in the subsequent time point as a Binomial–logit process with mean $p_{\text{fruit},ijp}$, which is a linear function of the species-specific potential fruiting incidence $\pi_{0,j}$, species-specific size dependence $\pi_{D,j}$, and overall effect of habitat mismatch, π_H:

$$\Pr\left(\text{Fruiting}_{ij,p,t+1} = 1\right) \sim \text{Binomial}\left(1,\ p_{\text{fruit},ij,p,t+1}\right) \tag{1}$$

$$\text{logit}\left(p_{\text{fruit},ij,p,t+1}\right) = \pi_{0,j} + \pi_{D,j} D_{ij,p,t} + \pi_H \Delta H_{ij,p,t}\,. \tag{2}$$

The habitat mismatch, $\Delta H_{ij,p,t}$, of an individual is measured as [placeholder]

3 Fruiting volume

For each individual that fruited, we model its fruiting volume [count] as size dependent:

$$F_{ij,t+1} \sim \text{Poisson}\left(f_{ij,t+1}\right) \tag{3}$$

$$f_{ij,t+1} = \phi_{0,j} D_{ij,t}^{\phi_{D,j}} \tag{4}$$

$$\log\left(f_{ij,t+1}\right) = \log \phi_{0,j} + \phi_{D,j} \log D_{ij,t}\,. \tag{5}$$

I presume that you had a power relationship in mind. This is different from the current main document, which doesn't log-transform $D_{ij,t}$.

4 Seedling location

We allow the recruitment of each tree to follow a dispersal kernel[1], under the Janzen–Connell hypothesis. For each species, the probability of having at least one seedling recruiting in a grid

---

[1] We may want to call this 'recruitment / establishment kernel' instead, because dispersal can be misleading—some may think of dispersal kernel as that of fruits. 'Dispersal kernel' was also used synonymously for the fruits' 'dispersal kernel survey data'.

at d distance away from the parent tree, $\Pr(R_{ij,p,t+1} > 0)$, is currently the 2DT probability kernel[2]:

$$\Pr(R_{ij,p,t+1} > 0) \sim \text{Binomial}(1, \; p_{\text{recruit},ij,p,t+1}) \tag{6}$$

$$p_{\text{recruit},ij,p,t+1} = \frac{S_j}{\pi L_j (1 + \frac{d_p^2}{L_j})^{S_j+1}} \,. \tag{7}$$

In the original document, Weng Ngai cautioned that:

> ...the outcome variable in this (recruitment) dispersal kernel was (within-transect) probability, not recruitment number/density. To translate this to IBMs, we needed a way to scale recruitment volume by parent tree size.

I understand this as finding a way to convert recruitment probability to recruitment count, or more accurately, to estimate recruitment count and then distribute the recruits across space using the recruitment dispersal kernel.

In the current solution, we observed the total number of seedlings (< 40 cm height, presumably ≤ 1 year-old) per parent tree in a 100-m transect, and then estimate the recruitment count, $R_{ij,t+1}$, as size-dependent:

$$R_{ij,t+1} \sim \text{Poisson}(r_{ij,t+1}) \tag{8}$$

$$r_{ij,t+1} = \rho_{0,j} D_{ij,t}^{\rho_{D,j}} \tag{9}$$

$$\log(r_{ij,t+1}) = \log \rho_{0,j} + \rho_{D,j} \log D_{ij,t} \,. \tag{10}$$

However, this model is hard to fit because of sparse data. So

> ...a better solution here was to borrow the scaling factor from the fruiting volume model for the seedling location model, implicitly assuming that recruitment fraction is the same (for each species) between individuals of different sizes (a reasonable assumption)...

to save us from having to estimate $\rho_{D,j}$. I think this solution may not be working as it was intended, but let's first go over this solution. It assumes that $\rho_{D,j} = \phi_{D,j}$, hence we obtain the equality between Equations 4 and 9:

$$\frac{f_{ij,t+1}}{\phi_{0,j}} = \frac{r_{ij,t+1}}{\rho_{0,j}}$$

$$\therefore \frac{r_{ij,t+1}}{f_{ij,t+1}} = \frac{\rho_{0,j}}{\phi_{0,j}} \,. \tag{11}$$

Because we did not constrain $\rho_{0,j}$, there is no guarantee against $\rho_{0,j} > \phi_{0,j}$ and hence $r > f$, meaning that parent tree i produces more recruits per fruit. While this is plausible for fruits with multiple seeds, this cannot be true for a drupe species (i.e., *Prunus polystachya*?).

We also need to make sure that r and f have the same area scale, i.e., both are per-100-m^2. In the main document, sometimes a 100-m-radius area is mentioned, which is not the same as a 100-m^2 transect.

The main document also states that:

> ... $\rho_{0,j}$, which is an approximation of the (logged) germination + survival rate of the fruits of each species, can easily be calculated for each parent tree in the seedling location surveys, and then averaged across[3] the species.

---

[2]S and L also varied by transect during estimation, need to think about how transect effect fits in this equation, which is currently incomplete or misspecified.

[3]Should it be 'within'? Otherwise $\rho_{0,j}$ wouldn't have the j subscript.

With '(logged) germination + survival rate of the fruits', I think you may be thinking of $\frac{\rho_{0,j}}{\phi_{0,j}} = \eta_j$ instead of $\rho_{0,j}$. This is because by rearranging Equation 11, we obtain:

$$r_{ij,t+1} = \frac{\rho_{0,j}}{\phi_{0,j}} f_{ij,t+1} \tag{12}$$

$$= \eta_j f_{ij,t+1} . \tag{13}$$

If Equation 13 is agreeable, then we can proceed to discuss its implication (on how we may want to rewrite some of the formulae). We may want to constrain $\eta_j = \frac{\rho_{0,j}}{\phi_{0,j}} \leq 1$ as discussed above, but let's leave this to later. Here, η_j is a combination of seed survival rate $s_{\text{seed},j}$, germination rate g_j, and seedling survival rate, $s_{\text{seedling},j}$. If we want to be explicit about the transition from fruit to seedling, we can express it like many demography matrix models:

$$\eta_j = s_{\text{seed},j} \cdot g_j \cdot s_{\text{seedling},j}$$
$$r_{ij,t+1} = s_{\text{seed},j} \cdot g_j \cdot s_{\text{seedling},j} \cdot f_{ij,t+1} ,$$

but we do not have so many data so let's keep it simply with a phenomenological η_j, and call it the fruit-to-seedling transition rate.

An implication of η_j is that recruitment, r is *not* size dependent. Rather, it depends on fruit volume only, and is only related to parent size *indirectly* through the fruits. Even if we could estimate $\rho_{D,j}$ in Equation 9, it would be even more phenomenological and indirect compared to $\phi_{D,j}$. Instead of worrying about our inability to estimate $\rho_{D,j}$, I'd take this as a justification not needing to do so.

Now, can we just focus on Equation 13 and estimate η_j using fruit and seedling data collected from the same location? Do we have such data? If so, it is a simple (random) slope-only model with forced intercept through origin.

[Still need to write out growth and survival models of other life stages, as well as making sure parameterisation matches simulation, but leaving that for later.]