Comments on the July 20 draft of the thesis, focusing on papers 3 and 4

Title of chapter 2: includes weed community composition, but you only report on biodiversity measures (richness, diversity, evenness) and individual species responses. Neither is “community composition”. You have a poster in the hallway with nMDS results. Biodiversity is definitely a community aspect but it’s rarely considered community composition.

We had made an nMDS graph but decided to replace it with a stacked bar graph of individual weed species aboveground weights and stand densities for ease of composition.

Evenness: What does evenness tell you that “more rare species” doesn’t. My personal opinion is that evenness is regularly calculated and rarely relevant. A large part of the “rarely relevant” is that evenness is very dependent on sampling effort. Consider a plant population with 5 relatively common species and 95 relatively rare species. If you sample a few plants (e.g. 100) you’ll probably see all 5 common species and few of the rare species. Relatively high evenness. If you look at 1000 plants, you’ll still see the 5 common species and a lot more of the rare species, much lower evenness. Same population, different effort, different evenness.

P 96, l 5: “higher degrees of freedom” You’re talking about # sampling quadrats per eu, right? How does more subsampling give you higher degrees of freedom for the error terms that matter?

I had always equated higher degrees of freedom with larger sample size, but I get it now.

P 96, l 9: Aha – not degrees of freedom, but total sample size. Too late to change the words since published.

Chapter 2 appendix B (imputation). Why were observations missing? No surviving water hemp? Would an average of 8 quadrats or a weighted average of 8 quadrats be more relevant?

If you use a weighted average, quadrats with no plants are very good data: they contribute 0 to both numerator and denominator.

We chatted about this at the defense. I did not appreciate that soybeans had no (or 1) waterhemp in any plots. I know you talked about strong weed control in 2019 but did not realize the completeness of that. I’ll write out my concern: Multiple imputation is great at “filling in” missing observations based on what is observed. When all soybean plots are missing, the only “what is observed” is the average male and average female count. Multiple imputation accounts for the extra variability imputing missing values.

General: you repeat Figure 4.1 in at least two chapters and may repeat the text describing the study. Be careful of copyright violation. (Does Frontiers require transfer of copyright?)

Frontiers has the first authors hold the copyright.

Chapter 3: assessing variability of repro potential. What is your measure of variability?

P 147. Should describe the stages in M\_s and M\_p. You mentioned sampling two seedbank depths, but that was a long time ago. You never said (that I noticed) what the 6 plant stages are, and the connection to the 8-10 sampled cohorts is unclear. This comes back again on p 148 when you say “plant cohorts were recorded”.

Chapter 4 title: Why do you call the simulations “retrospective”? I almost asked about this in the oral, but decided not to. Prospective analyses ask “what will happen?”. Retrospective analyses look for explanation of what has already happened (e.g. a life-table-response-experiment). My sense is that the simulations are asking “what will happen when I change the fecundity?”, which sounds prospective, not retrospective.

P 148: I don’t follow the rationale for the names “seed preserving” or “neutral”, especially neutral.

Equation (4.2): As written, this equation is correct for scalar multiplication but not for matrix multiplication.

1. Be careful of the ordering and labelling of periodic matrices. The product generates B\_1 \* B\_2 … B\_H. You probably want B\_H B\_{H-1} … B\_1.

Fixed

2) You need to post multiply N, not pre multiply.

Fixed

3) Explanation below the equation. “population transition matrix during summer” implies one B\_h matrix. There are many B\_h.

P 148, bottom: I don’t follow the two sentences about crop competitiveness and mature plant size. How is this information relevant to the model?

P 152: Text in 4.3.2.8 mentions P\_{2w} and A\_2. Are these the same things? If not, what are the A’s?

The A’s are the product of all the six B\_h in one crop phase, from spring tillage through overwinter

The P\_rw’s are the product of all the A’s within one rotation x corn weed management.

P 142: How do you know plant fecundity is the most influential demographic rate?

I ran a sensitivity analysis for each simulation, but ended up excluding it because the tables are too large.

P 158 top: Here there are 6 cohorts, earlier you said 8-10 were sampled.

Clarity: If you’re focusing on 2019 rates, why tell me something about 2018?

P 158: Estimating fecundity.

1. The fecundity prediction models predicted log (seeds+1) from individual plant data (2018), right? Plant size is variable within a quadrat or rotation. Did you use the mean plant size in your equation, or did you predict individual plants, backtransform, then average? Makes a difference when the prediction is a non-linear function of plant size.

The individual log plant weight was used first in a Crop ID x corn weed management treatment specific equation of general form [Y = aX + b], then average for cohort value. Y = log(seed + 1); X = log(Weight + 0.005); a and b specific to each Crop ID x corn weed management treatment.

2) When you predicted the estimated fecundity going into the demographic model, did you predict the mean over plants or the median? They are not the same when the model predicts log fecundity. Which is the more appropriate quantity (mean or median)?

I used the back-transformed means in the model because all the rates in the six-matrix sets are on original scales. In each row of the dataframe of the estimated number of seis are calculated follows :

* log(seed + 1) = Y = a\*log(Weight + 0.005) + b
* var[log(seed + 1)] = var(Y)
* Seed = exp[Y + var(Y)/2] - 1

P 158, scenario-specific results: Clarity. On p 148, the difference between the two scenarios is a methodological difference. Here we find out they also use data from two different years.

Clarification added on pages ### to focus on the level of efficacy.

P 162: Do you show the relationship between emergence delay and decreased fecundity? This has between year and within year (between cohort) consequences, because cohort 1 emerges at different times (Table 4.1?).

A new graph is added.

P 163, Figure 4.5: This figure is confusing because you’re trying to show two things at once: 1) given a mean seed density, what is the population trajectory? That’s what most of the page shows and what the reader’s attention will be drawn to. 2) what is the seed density threshold? Those are hard to see, in both the figure panels and especially in the legend. I suggest you figure out what you want to show and plan a figure that clearly demonstrates that. One possibility is a graph of X = seed density threshold and Y = annualized lambda. Each dot is one simulation. Then highlight where that curve crosses Y = 1. You may need to think carefully how to parameterize “seed density threshold” when you manipulate more than one cohort.

Section 4.5 is more a discussion section than a conclusions section.

General: You started year 1 with all phases of all rotations. How did you start A4? Was that started the year with oats/alfalfa the year before?

All the projections started with 1000 seeds at the top and 0 seeds at the bottom (1000 for y-axis scale), then pre-planting tillage, emergence, etc., for phase 1 (corn), continued with pre-planting, emergence, etc., for phase 2, and then continued through the end of the last phase of any rotation. So A4 was started at the seed column after overwintering from O4.

All the simulations started with 10000 seeds at the top and 0 seeds at the bottom and followed the same sequence/crop phase and sequence of crop phases within a rotation. Now, the populations are kept such that λ ≈ 1. Also, 10000 seeds/m2 is closer to the average seedbank density of waterhemp (considering waterhemp proportion in the underground weed community) in the Midwest.