Reviewer 1

Code comments

Hi y'all, I'm one of the reviewers and haven't had a chance to look at the paper yet (the new one or the 1999 one; next week, I promise!). But I did run code (R scripts only) and wanted to note a couple of things to update, maybe before the next reviewer comes on board.

I dove into the code without even trying to figure out the ecological context: purely a question of "can I run all this code". And the answer is "almost entirely, yes". The way the files were organized, named, and discussed in the README made it really easy to figure out what to run and when. I appreciated the warning about the code for Fig. 3 in particular taking a long time to run. And best I can tell, the images I generated were the same as the authors'. So with a few updates, yes, the code was reproducible on my machine. Specifics follow.

Important code updates - things I had to change in the code to make it run. All notes are for `Script_Experiment`.

- need a call to `library(xtable)` or lines 143 and 144 won't run.
- extra space in line 116, towards the end of the line. There's a `> =` that needs to be `>=`. Context: `(results_exp1[dim(results_exp1)[1], 2:13])> = 0),rep(T,12)`
- similar extra spaces in lines 135 and 206: `" = = "` should be `"=="` in the `Probability = c(colSums......)` portion.

>> We thank the reviewer for reproducing the code, and noticing these typos. We have now corrected them.

Additional but less critical notes

- In the README, I believe you're referring to the `save_image()` function here, not `save.image()`: "Finally, saving the plotly plots into .pdf of .png files requires the `save.image()` function." **>> Yes. This is now corrected.**

- It would be helpful with that same line to note that the `save_image()` function requires an installation of python/miniconda, since that particular function isn't available directly in the R {plotly} package. (I didn't understand why I needed the provided installation instructions until running that line in `Script_Fig_1` threw errors.) **>> We have added instructions.**

- Same function, again - need to make sure the {reticulate} package is recent. The CRAN version that I'd installed had a bug and I had to install the github version in order to get the miniconda installation to work correctly. See [issue 1297](#) for details on that issue. Unsure when a new version will be pushed to CRAN. **>> We now mention the package versions required and used.**

- I still never got that function to work but I don't think it's the authors' job to get me through the error I couldn't get past ("NameError: name 'sys' is not defined"). I tried to do some troubleshooting but figured actually saving the images was less important than generating them. **>> We have now added an explanation to the README on how to workaround this problem (that occurs only in some configurations).**

- a couple minor typos in the README, `Code/` .R files section: `Code/Script_Exp.R` should be `Code/Script_Experiment.R`; `Code/FScript_Functions.R` should be `Code/Script_Functions.R` **>> These are now corrected.**

- For experiment 1, I got a few of these warning messages; not sure if this matters: `DLSODA- At T (=R1), too much accuracy requested for precision of machine.. See TOLSF (=R2) In above message, R1 = 8656.66, R2 = nan` **>> Unfortunately numerical resolution can fail for some time series, but these have now been removed from the summary statistics presented in the figures 5 and 7 and therefore cannot affect the results. See README/Computation times for more details.**

- Also for experiment 1, I got a few warnings along these lines; again not sure if they are concerning but want to point them out in case. `1: In lsoda(y, times, func, parms, ...) : Excessive precision requested. scale up` rtol' and `atol'` e.g by the factor 10 2: In lsoda(y,

```
times, func, parms, ...) : Returning early. Results are
accurate, as far as they go 3: In lsoda(y, times, func,
parms, ...) : an excessive amount of work (> maxsteps )
was done, but integration was not successful - increase
maxsteps.
```

- Didn't get any warnings for experiment 2.

I think that's all for now. I hope it's helpful. Thanks for letting me review this; I'm looking forward to exploring the paper and learning what all this code was for! More next week sometime.

Took a little longer than I thought to review the paper itself. After doing so, and digging back into the code, I'm even happier with how well the code was organized and documented. Good job to the authors on that. **>> Many thanks!**

Here are my additional comments; I hope they're helpful.

Additional code thoughts:

- in `Script_Fig_exp_1` - turns out I hadn't run this before submitting my first comment:
  - I ended up with 371 instead of 372, and 430 instead of 431. Other panels were the same as in the paper. **>> These slight differences may be due to using different computers (random number generation can vary).**
  - On the second row, my y-axis goes up to 60 rather than 40 (maybe because of the different outputs that were selected). **>> Exactly.**
- Fig_exp_2 came out exactly the same.
- though I was able to reproduce all figures (with the minor exceptions above for figure 5), I would recommend structuring for loop outputs differently in the future. At least in R, data frames as output for a loop are known to make loops very slow and inefficient because they have to be copied with every iteration. A list would probably work better; each iteration adds to the list and you can `rbind` all the pieces at the end. I wouldn't go so far as to say the current code should be restructured, because it's complicated and it works (albeit slowly) as-is. Just something to keep in mind for future, happier work - avoid data frames as loop outputs in R.

>> We thank the referee for testing the code, and keep in mind their comments for further work (for now it is easier to have a dataframe format to save results into well-organized and easily readable text files).

Notes on the paper:

In this article, the authors replicate Huisman and Weissing's (1999) simulations and figures of phytoplankton species abundances in a stable system with limited resources, using model parameters from the original paper. The authors then go further and perform simulations in which the the input parameters vary, and the results vary as well: to the point where the authors suggest that supersaturation is NOT common. The replication is well done and the code is well-documented. I do have misgivings though about the conclusions drawn from the new simulations, and would like to see more discussion and justification to ensure the conclusions are appropriate based on the outputs. What I'm most concerned about is calling the perturbations to the original (HW1999) parameters 'small'. I'll separate my comments into the 'replication' portion of the paper and the 'experimental' portion.

Replication portion

The paper is short and to the point. Without having the original next to me while reading this one though, it would have been hard to understand. Therefore many of my suggestions involve adding some detail and exposition to enable the current paper to stand more on its own.

- in the introduction, when discussing the different "parameter scenarios" from the original paper, would be helpful to set up the context that these are parameters of the models for competition and saturation etc. Okay to refer to original paper for specifics of parameters used, but please give some examples of how this would play out (like in the original paper when they listed things like N, P, Si, etc. as limiting resources). **>> We would rather not speculate as to the mindset of the original authors or repeat too much the original article, in our opinion it would not be fitting for a replication.**

- what was r_i in the original simulations? Looks like it starts at 1 for all species in figure 4 (this is based on the code - but I'd like to see it mentioned in the methods of the article, and then talk about what r_i values were used in the experiments) **>> We now mention that r_i values were set to 1**.

- please add citations for the {deSolve} package and R itself **>> Done.**

- the tables in results, with probabilities of species numbers - this wasn't entirely clear to me. Is this how likely each outcome was from the simulation results? Should the probabilities add up to 100%? If yes, I'd suggest saying something

like "proportion of simulated communities with X species after 10,000 days" in the caption. However the numbers do add up to more than 100. **>> We now plot the percentages as a figure. The probabilities add up to 100%, except for the supersaturated category which corresponds to > 5 species.**

- same tables would be easier to read in vertical form, with two columns **>> Now done graphically.**

Experimental portion

- The paper's introduction refers to the experimental perturbations of growth rates as both "mild" and "small". Please provide some justification and citations (if possible) for this - to me, going from "all species are assumed to have the same max growth rate" (all r_i values were 1 in HW1999's simulations) to "all species have differing growth rates, which can vary by more than 50%" (0.7 - 1.3) is far from small. It may indeed be more realistic, but is this really a small perturbation? **>> Yes, these are actually small perturbations. Using the paper provided by reviewer #2 (Edwards et al. 2015), the coefficient of variation (CV = SD/mean) across species was 0.4 for P-limited species and 0.7 for N-limited ones, while we simulated CV = 0.1. Another important point to keep in mind is that we *truncated* the distribution of growth rates to exclude values below 0.7 and above 1.3. It is only a device to exclude extreme values: these bounds are already quite unlikely values, being located at 3 sigmas of the mean. We added a sentence at the end of the last paragraph of the Methods, *"This range of values remains consistent, and even conservative, when compared to experimental values (Edwards et al. 2015), where CVs can be larger."***

- In the vein of these parameter values being more realistic - is variation in actual growth rates something that's known or has been estimated? if so, please add citations to justify these choices; if not, please explain why these numbers seemed reasonable to use (I'm not arguing; just want to see clear reasoning). **>> See the response to the above comment, this variation is an underestimation of actual variation in intrinsic growth rates between species.**

- After looking at some of the results from experiment 1, I don't think "number of species present after 10,000 days" adequately represents what happened with the species communities. If this really is the only metric that is important, it needs some justification. It seemed that a fair amount of scenarios had good stable oscillations for more than 5 species, sometimes even 8-10, and then

everything crashed with the last species additions. **>> We added two sentences to explain this "***As we are trying to challenge the robustness of the replicated simulation, we do not change any other parameters : thus we only conduct the simulations for a duration of 10000 days, as the original simulations. In order to easily visualize variation in biodiversity, we also measure the number of extant species at the times used in the article to introduce new species, i.e. after 1,000, 3,000 and 5,000 days.***"**

I'd like to see some more detailed information about these results in the paper itself. **>> We hope that the numbers of species at 1,000, 3,000 and 5,000 days add precisely that information.**

- The code was well-documented enough (again - kudos) that I could play around with it, but more information could be included in the results/discussion without the reader having to run code. A couple of ideas:
  - The simplest thing would be to include graphs from *more* random selections of the simulations, rather than two panels for each random selection. This could be personal preference, but I don't need to see a separate panel for the later-added species; just throw it all on one panel and I can get an idea of whether things are stably oscillating or not. And I think showing more outputs will be more helpful for the reader, rather than more detail for fewer outputs. **>> We have previously considered one panel for all species but it was actually difficult to see them all. We understand the need to see more simulations and we have added more of them as an Appendix.**

  - I'd love to see some information about number of species present just before each addition as well - e.g. in the tables, rather than only including proportions for the number at 10,000 days, add columns (if you take my suggestion above - otherwise it would be added rows) for number present at 1,000 days (out of 5); 3,000 days (out of 8); 5,000 days (out of 10); 10,000 days (out of 12). I realize this may be a lot of work to pull out of the results. But it would be incredibly helpful and enlightening to see. This may work better as a series of histograms or density plot(s) once all that info is added - see what works, if you tackle it. **>> This is now implemented - in graphical format, which hopefully makes it easier to see.**
  -
- description in the text (on page 7) of figure 5 says that "it is already noticeable that ..... is sometimes already unstable" - I'd recommend describing what the reader should look for regarding stability/instability. e.g. "in many scenarios, at

least a few species oscillated around stable points until additional species were introduced (e.g. figure 5, 315 and 81); but sometimes these stable oscillations were never established in the first place (e.g. top row of figure 6)." Not sure I captured that completely - but I didn't know what you meant by stable or unstable until I really dove into more code, so a verbal description would be helpful. **>> We have added a more detailed explanation of what was meant by unstable here and now write "We observe contrasted stationary endpoints, possibly with some oscillations or chaos during the invasion process. As shown in Figure 6, however, most species do not persist."**

Reviewer 2

Hi
I finally got the time to run the code and review the paper
Saving figures
 As @swmpkim I was not able to save the figures and it appears to be the same issue "NameError: name 'sys' is not defined". From my understanding, all your code works in R and you solely retuculate python code for the plotting, which to me seems a bit overkill. Especially because now reproduction requires the installation of python (and I was worried whether this installation of python may mess up with my other pyhton on computer). If not too much work, I would recommend to save these figures with R code, as both reviewers were not able to save the figures. Addtionally, for me the figures weren't shown in R itself, so I actually don't know whether they match the ones in the paper.

>> We understand this perspective. However, some of the figures to replicate have 3-dimensional panels, and we have found that plot_ly() was very handy to customize those graphs (compared to other packages) so that we could compare properly to the original article, with the same angles of 3D graphs. Therefore, we have kept the current packages but explained how to work around that issue in the README.

Running code
 I tried to run the Script_DataHW1999.R, but it appears to be stuck. Currently trying to compute code for Fig. 3a and 3b. I'm not sure how long this part of the code takes, however if it takes longer than say 5 minutes on your machine I would recommend to add a warning in text stating how long it takes on your machine (well aware that this is only little info for how long it will take on my machine). But currently I'm not sure whether it is stuck (certainly 10minutes + running for now).Even better, if possible, add an update stating how much of the progress is already done. Figure 3a and 3b appear to compute the same task for many different levels of K, so maybe add a

progress bar of how many K are already computed, this way I know that the code is not stuck somewhere.
Same applies to the experiments where a progress and an estimation of how long this takes on your computer might help to ensure the code is not stuck.

>> We had actually written that Fig. 3 took longer to run in the README as remarked by Reviewer 1. However, we had forgotten to explicitly write the runtime, which we now do. We have also inserted that information as a comment in the code for Fig. 3.

Finally, you mention in the README.md that Matrix_C and Matrix_K contain info, but what info is contained? After looking a bit at the code it appears that these are the parameter settings for the different figures. I recommend mentioning this explicitly in the README.md, which matrix is used for which simulation and what it contains (K contains the halfsaturation if I understood correctly). Additionally, I would add column and row names to these matrices

>> We added more information to the README about the info contained in Matrix_C and Matrix_K. We also changed the storage format of said matrices in order to display column and row names.

For the article

Overall, I found the article, to the point, we replicate the idea of oversaturated coexistence in pyhtoplankton communities. Interestingly, I find that the paper replicates essentially two works at the same time. The Huisman & Weising 1999 and Schippers et al 2001 (which I do not know in detail). They in principle replicate the simulations of Huisman and Weising, but then go on to replicate results of Schippers showing that Huisman and Weising, while technically reproducable, may not be very important as it is very sensitive to perturbations.

>> We thank the reviewer for this assessment of our paper. We only formally replicate Huisman & Weissing 1999, as we use a different perturbation structure than Schippers et al 2001. But it is correct that our additional simulations could be thought of as a "robustness check" of Schippers et al 2001.

I only have a few minor comment and one potentially large recommendation:
To me it is unclear how the original paper found the parameters for the super-saturated communities. If the authors have any insight on this I would appreciate such information.

>> We actually have no idea about this.

However, independent of how these parameters were found (which may very well have been brute force) I wonder how realistic these parameters are for natural communities. Or, vice versa, how likely super-saturation is for realistic communities. To test the robustness of super-saturation the authors perturbed the parameters from the paper, as reviewer 1 points out, it's unclear how realistic these values remain.

>> As explained to Reviewer #1 above, real variation in intrinsic growth rates can frequently be much larger than what we have considered. We considered CV = SD/mean = 0.1, while in reality CV of intrinsic growth rates were about 0.4 to 0.7 (and possibly larger when computing across all experiments and not simply across species). We thank the reviewer for the article linked below, which allowed us to compute growth rate variation independent of our own previous work, which already suggested CV = 0.1 to be a lower bound btw.

As a potentially quite interesting extention I would recommend to use a database of realistic phytoplankton traits, e.g. https://esapubs.org/archive/ecol/E096/202/ I happen to have a paper where I do the same (but with a more complicated model including competition for light and zooplankton predation). I would be interested in whether under these circumstances super-saturation is even a thing. I have strong suspicion that this experiment will simply show that it is not, so I would understand if the authors decline this potenially extensive additional work with potentially limited additional insight.

>> While we are very thankful for the reference, what the reviewer suggests as an extension seems to us to go way beyond the scope of our replication article of Huisman & Weissing 1999 (which already strays a little further from pure replication, by checking the robustness of the findings of Schippers et al. 2001 too). We therefore keep these ideas for future work.

Besides this I only have minor comments:
2. Paragraph, 1. sentence: At that this is true at equilibrium only, otherwise species may coexist. That is the Huisman & Weising did not find a loop-hole in the theory, but simply explored what happens if the assumptions are not met.

>> We switched the sentence to "According to many mathematical models, the number of phytoplankton species in a single homogeneous medium cannot exceed the number of separate resources available [3,4,5] if the dynamics converge to a stable fixed point. "

3. paragraph: "an hypothesis" should be changed to "a hypothesis"

>> Done.

Numerical experiments: You take the abundance at the end of the experiment. But the idea is exactly that the densities fluctuate. So a more reasonable approach would be some sort of time average of densities above a certain threshold

>> Time averaging would imply new arbitrary choices regarding the period over which to average, while chaos in fluctuating populations means there are not always clear periodicities -- thus we are not really in favour of this option. More pragmatically, the files dataresults_exp1_[simulation_number].txt store all the numerical integration results and they show that the vast majority of times where there is extinction, abundance is well below 0.001 (most of the time zero as the four digits stored have been exhausted), and when species persist -- even with fluctuations -- their abundance is usually much above 0.001. This 0.001 threshold is therefore just catching a couple of species that are still declining slowly.

Please note that following other comments by Reviewers 1 & 2 we now output the percentages of species remaining at various times, before the end of the numerical experiment.

Numerical experiments: I would find these tables in form of graphs more insight full. Reviewer 1 pointed out that your perturbations are relatively large (to which I partially agree). A really cool addition would be how these percentages change as a function of the average perturbation.

>> We replaced the tables with figures.

As explained above, the perturbations considered are actually not large (all the more that they are between species and not over time), and could even be considered rather moderate, compared to the large observed growth rate variation between species.

Our modeling philosophy for those numerical experiments is to take into account and complement the published critiques by Shippers et al. so that readers get a balanced view of what can be replicated and what is super sensitive to parameters. We therefore look at what occurs if growth rates are perturbed by a non-negligible yet modest perturbation. In our opinion, a detailed investigation of how the results change as a function of the intensity of perturbations would be original research and out of the scope of the current paper (although it might be interesting as a follow-up, and we hope that our present article might stimulate such work). Let us note in passing that in such a case, it would be great to have variation in time of those growth rates also, or some temporal rhythm in the availability of resources.