Recommendation by the Subject Editor:  
Your ms has been reviewed by two reviewers who are, like me, generally positive about this ms, and provide constructive comments to improve it. In particular, the ms needs restructuring, careful formulations and a more elaborate historical context. I am very much looking forward to the revision.

This strikes me as a positive comment and green light from the subject editor. I get this impression s/he is in favor of the paper being published.

Reviewer(s)' Comments to Author:  
Reviewer: 1  
  
Comments to the Author  
This manuscript offers six alternative empirical approaches for assessing whether species pairs can coexist. The methods are connected by the common thread of thinking in terms of the coexistence theory developed by Peter Chesson and colleagues. I found the manuscript clearly written, and (speaking as a non-empiricist) of potential value to empirical ecologists.

This comment also strikes me as positive overall.

I do have some conceptual concerns, as well as a handful of more specific comments and suggestions - please see them below.  
  
For the big picture, it is important to put the perspective of the current manuscript in an accurate historical context. The view espoused in the Introduction (that ecologists have traditionally thought niche differentiation is the main mechanism promoting coexistence until Hubbell and modern coexistence theory came along) is too simplistic. For example, In May (1973, "Stability and complexity in model ecosystems"), a coexistence condition was given in terms of how similar the K\_i values of a pair of species are. This explicitly recognizes that one can either increase niche differentiation, or bring the K\_i values closer together to achieve coexistence - it is really the same approach as the one advocated here in Eq. (1), just with a different parameterization. See also MacArthur & Levins (1967 Am Nat; also cited by the Authors) for a similar approach. But one may also mention studies like Roughgarden (1976 TPB), Slatkin (1980 Ecology), and Taper & Case (1985 Ecology). Likewise, it is also not quite correct to say that classical community ecology only considered resource partitioning as a niche mechanism - see e.g. Levin (1970 Am Nat) or Levins (1979 Am Nat).

Ok … this seems fair. I can add some text to the introduction to account for more historical context.

All of this is straightforward to modify in the Authors' manuscript, but the historical view does have bearing on how the Authors' perspective and results are interpreted. Which brings me to my main question regarding this manuscript. The empirical methods described in the manuscript are often tools for estimating the competition coefficients. Once we have those coefficients, conditions for coexistence can be evaluated. Why is it important then, in the Authors' opinion, to use Eq. (1) instead of other equivalent formulations of the Lotka-Volterra coexistence condition - e.g., that of Vandermeer (1972 Science)? What is the value of re-parameterizing in terms of ND and RFD? Do these quantities have some independent meaning which makes evaluating them worthwhile on its own? (I have argued elsewhere that this is not the case, but I am always happy to learn from my mistakes.) If so, what is this independent meaning? In general: if the empirical methods of the manuscript estimate competition coefficients, and coexistence can be surmised from those coefficients (from Eq. 1 or in other ways), then 1) why are ND and RFD in particular important; and 2) how is the current contribution going beyond classical methods of empirically evaluating coexistence (many of which do exactly what the Authors propose: estimate competition coefficients)? I believe that clarifying this point would make the manuscript much stronger: it would then not only be a catalogue of methods, but would also motivate the need for those methods better.

Hmmmm … not entirely sure how to respond to this comment. At its core, R1 seems to suggest that competition coefficients and classic LV theory are sufficient to predict coexistence, and that Chesson’s coexistence theory with ND and RD do not add anything new. S/he suggests they have argued this case already, but do not give a reference.

I don’t think we should spend time arguing the merits of Chesson’s coexistence theory with R1. Rather, we should istead cite the volumes of papers that are focused on this issue right now, and emphasize that regardless of whether R1 sees the merits in this field, it is moving forward.

I remember seeing such similar comment from last review – the meaning of ND and RFD. My opinion is that, from a mathematical viewpoint, re-parameterizing the LV model to derive ND and RFD does not add much novel insight. There are also other ways to re-parameterize the LV model, which is what the reviewer published in 2018. The reviewer’s re-parameterization method is probably more advance and might also make more sense. I suggest making argument that ND represent the geometric average, while RFD represent the geometric standard deviation, of competition coefficients as Carroll suggested in his 2011 paper.

Apart from the above broader issues, here are some more specific comments and suggestions:  
  
- One small notational suggestion: it is strictly against the rules of the mathematicians' club to denote a single quantity with multiple symbols, like ND or RFD. If possible, I would denote them differently, and with a single symbol. For example, a script capital N for ND and a script capital F for RFD would work (as was done in Spaak & De Laender, bioRxiv, doi: <https://doi.org/10.1101/482703>).

I agree that being consistent with notation is important, but I think this is being overly nitpicky and not particularly constructive. We can choose to take or ignore this point at will.

For me, the more important issue is to be consistent with the terminology and symbols that have been used in prior publications on this topic. I don’t consider an unpublished manuscript in bioRxiv to be one that sets a precedent for notation.

- l.57: Again, this is merely a terminological issue, but the doubly relative expression "relative fitness difference" is not actually ever used by Chesson and colleagues. The term is instead "average fitness differences", or "species average fitness", in case one does not take the difference between them. In fact, what is called RFD in the manuscript is not a difference, but a fitness ratio (of kappa\_i and kappa\_j; see Chesson 2018 J Ecol for details). Therefore, instead of the term "relative fitness difference", I would use the term "fitness ratio" or something to that effect.

Again, the important point is to be consistent with prior publications. Chesson 2000 (ARES) used the term ‘average fitness difference’ and ‘fitness inequalities’. Since that time, many others (e.g., Adler, Levine, Hil Ris Lambers) have used the term ‘relative fitness difference.’ I do not know where the terms originated, but I would argue that relative fitness differences are now the common usage.   
  
- l.119 (and 121): I would make the description of the alpha\_ij more clear and unambiguous: alpha\_ij is the amount of reduction in species i's intrinsic growth rate (r\_i) caused by one unit of species j's density.

Ok … simple edit.

- Eq. (6): what is eta\_i? It was not defined anywhere.

Ok … simple edit.

- l.236: It might be an overstatement to say that Carroll et al. (2011 Ecology) have shown this. Rather, they demonstrated that it recovers Eq. (1) for the two-species MacArthur consumer-resource model. Otherwise Eqs. (8) and (9) express a hypothesis, rather than a proven assertion. To my knowledge, this has not been developed further since.

R1 is incorrect. Narwani et al. provided mathematical proof for the assertion in the SI of their 2013 Ecology Letters paper.

- l.245-247: I am not sure it adds much value to have this derivation here - almost the same has been done in the Appendix of Carroll et al. (2011 Ecology). If the Authors insist on keeping it, then so be it, but otherwise I think this is something where it is sufficient to refer to earlier work (and it would also shorten the manuscript somewhat).

R1 is incredibly nit picky! I’m inclined to keep the 2-lines so that readers don’t need to reference the Appendix of Carroll.

- l.269-270: The limitations of this method are actually very nicely laid out in the original article (Carroll et al 2011, p.1160). Based on that, the main problem is not with introducing facilitation per se, but with introducing anything that creates alternative stable states and complex dynamics. (For reference, see also our analysis on why any method based on invasion criteria to assess coexistence runs into trouble when complex dynamics and alternative stable states are present, in Barabás et al. 2018 Ecol Monogr).

We can concede here, both mentioning the limitations stated in R1’s comment, and referencing his paper.

- Eqs. (16) and (17): something is not quite right - either here, or in Eqs. (14) and (15), which do not have y\_ij terms, only y\_ii. In any case, how were these equations derived in the first place?

The latest version that I have (posted on biorxix) does not have equations 16 and 17.

- l.409-411: I would not write this possibility off as being unlikely to produce coexistence. For instance, if two species can both invade when above a (low) Allee threshold, then I would say we have proper, well-groomed coexistence. It is, however, incompatible with a strict invasion picture, where species are introduced at densities below the Allee threshold. And yet, possibly a majority of populations in nature do collapse if their density drops below a threshold - either due to Allee effects, or the lack of defense from the fixation of deleterious mutations, etc.

I’m not sure about this comment, and would need to think about it more deeply. I’m inclined to give in and mention this, and cite R1’s work to try and appease him.

- l.425: Or if there is a sufficiently large region with both growth rates being positive, even if it does not extend right down to zero density (see above).

Ok.

- l.438: This is true if we confine ourselves to the ND and RFD (what Chesson 2018 J Ecol calls 1-rho and kappa\_i/kappa\_j) description of coexistence. The NFD approach might lend itself better to the alternative (and much more generally applicable) approach where community stabilization is measured by the scaled community average invasion growth rates A, and the (adjusted) average fitness differences xi\_i by the deviations from this average. See either Chesson (2019 J Ecol) or Barabás et al. (2018 Ecol Monogr) for a detailed description. This approach has the added benefit of being multispecies by default.

Ok. We can probably just add this text to the paper.

- l.445-448: I would say this is more a complication to be dealt with, rather than a limitation. In principle, I do not see any problems with measuring growth rates at various frequency levels, revealing any nonlinear dependence of growth rates on frequency (though, of course, this may be prohibitive from an empirical point of view, if too many experiments would be needed).

Ok. We can probably just add this text to the paper.

- l.685-686: There already is a multispecies framework within Chesson's original theory. It's the one using A and xi\_i instead of rho and the kappa ratio (called ND and RFD by the Authors). Second, the work of Saavedra et al. (2017 Ecol Monogr) is more of a suggestion for a multispecies definition, rather than something that organically follows from Chesson's theory - it is derived from quite a different approach. This would be good to keep in mind here.

Ok. We can probably just add this text to the paper.

Sincerely,  
Gyuri Barabás  
  
  
Reviewer: 2  
  
Comments to the Author  
In this manuscript, the authors intend to guide empiricists who wish to measure niche and fitness differences in their system by comparing different experimental methods and outlining the pros and cons of each. This would make for a very useful paper, unfortunately this manuscript falls short. I think that bridging the gap between theoretical and empirical work in ecology is very difficult -- let alone for something as complex as MCT -- so I do not want to discourage the authors, but I think that some major revisions are required for this paper to reach its goal. I have several major comments which I believe need to be addressed first before tackling any of the more minor issues, so my review focuses on those.  
  
Major Comments:  
  
1- More on the limitations of MCT when applied to empirical comms  
MCT makes several important assumptions which are commonly violated in empirical communities, and has a history of being misused and/or misunderstood. For these reasons I believe that any paper attempting to guide empiricists in applying MCT needs to make the limitations and caveats of the approach much more explicit to prevent any further misunderstandings of the theory (see Barabas 2018). Through some of the limitations of MCT are addressed throughout the manuscript, they are dispersed throughout  and some of the limitations inherent to MCT (e.g. its inability to deal with both competition and facilitation) are conflated with limitations of the methods quantifying niche and fitness diffs themselves (see refs to Bimler 2018).

This is not a very constructive comment. R2 suggests we have ignored major limitations of MCT, and then does not describe any with exception of things we already mention in the paper itself. The only thing we can do in the response is (a) argue we have made all limitations we know of as clear as possible, including those that R2 mentions here; (b) note that R2 does not support his/her point with any specific examples or information; and (c) invite R2 to offer more concrete advice or suggestions.

2- Restructuring of the paper.  
The paper is hard to follow and it is difficult to know where I should be looking for specific information. Pros and cons are split up between Part 1 and 3. Overall the structure feels messy and the manuscript is heading off into different directions, I think the authors could benefit from thinking further about their intended audience and the information they might require in order to make the mss flow more logically. I would suggest something along the following lines:  
1- applying MCT to empirical communities – general info, assumptions which empiricists should be aware of  
2- Description of the methods, their pros and cons  
3- Comparison of the methods – tradeoffs, calculations of niche and fitness diffs  
4- Selecting a method  
But most importantly information needs to be condensed and reorganised.

This is also not a very constructive comment. R2 first misrepresents the current structure of the paper, and then goes on to suggest what is essentially our current structure.

I get the impression that R2 didn’t read or understand the paper. But what concerns me is that the subject editor seems to put some weight into this comment when s/he suggests “the ms needs restructuring.”

We should talk about this, and determine if we need to write to the subject editor for clarification.

3- Clarity and revision of each method.  
It is somewhat unclear what the authors think empiricists would be using MCT for. Is it to calculate niche and fitness differences, or to predict coexistence? If it the latter, I am of the personal opinion that using MCT to predict coexistence outcomes in empirical comms is rather dubious, in large part due to the prevalence of facilitation, intransitivity and HOIs. I know that these things are mentioned but they are not discussed with enough detail. Facilitation for example is only brought up as a problem for the annual plant population model – but it is a problem associated with MCT itself (we cannot calculate niche and fit dffs with interaction coefficients of opposite signs). The plant population approach usually deals with this by forcing all interactions to be competitive. The presence of facilitation, in any case, invalidates predictions of coexistence. Though Bimler et al 2018 suggest the calculation of analogues to niche and fit diffs, they are unable to calculate coexistence so not much of a solution is given.  
In any case, more information on the pros and cons of each methods need to be given.  
It is also not very clear why the different methods give different measures of niche and fitness differences, and this should be made far more explicit. All in all, though the authors may have a very clear understanding of MCT, this is not made apparent in the mss.

FUCK YOU ASSHOLE! I hate shithead referees like this … Uuggg – I really don’t like your entire field of research, so I’m going to try and damn your paper.

Minor Comments:  
  
Introduction: a long time is spent on the history of MCT, but practically no space is dedicated to the application of MCT in empirical communities, or why it might be a useful thing. I think the introduction needs to be rewritten to be in line with the targeted audience and goal - everything up to line 70 is somewhat superfluous.

Once again … FUCK YOU ASSHOLE!

Fig1: I can’t differentiate between regular and bold typeface  
  
Fig6 could really be improved. The colour scheme is confusing, the histogram is unnecessary, and I am not sure what the numbers represent or what the top graph is trying to show. The whole figure needs to be rethought and possibly condensed.  
  
I very much like Table 1 though.